Balanced allocations under incomplete information: New settings and techniques

Dimitrios Los

St John's College,
University of Cambridge

Date of submission: *March 2023*

This dissertation is submitted for the degree of *Doctor of Philosophy*.

Declaration

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text. I further state that no substantial part of my thesis has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. This dissertation does not exceed the prescribed limit of 60,000 words.

Dimitrios Los
31 March, 2023

Abstract

Balanced allocations under incomplete information: New settings and techniques

Dimitrios Los

In the balanced allocations framework, there are m balls to be allocated into n bins with the aim of minimising the maximum load of any of the bins, or equivalently minimising the *gap*, i.e., the difference between the maximum load and the average load. In this dissertation, we focus on the *heavily-loaded case* where $m \gg n$, which tends to be more challenging to analyse.

In a decentralised setting, the simplest process is One-Choice, which allocates each ball to a bin sampled uniformly at random. It is well-known that w.h.p. $\text{Gap}(m) = \Theta(\sqrt{\frac{m}{n} \cdot \log n})$ for any $m \gg n$. A great improvement over this is the Two-Choice process [18, 99], which allocates each ball to the least loaded of *two* bins sampled uniformly at random. Berenbrink, Czumaj, Steger, and Vöcking [29] showed that w.h.p. $\text{Gap}(m) = \log_2 \log n + \Theta(1)$ for any $m \geqslant n$. This improvement is known as the "power of two choices". It has found several applications in hashing, load balancing and routing; and its importance was recently recognised in the 2020 ACM Theory and Practice Award [17].

In this dissertation, we introduce a set of techniques based on *potential functions*. These enable us to analyse (both in terms of gap and load distribution) a wide range of processes and settings in the heavily-loaded case and to establish interesting insights in the balanced allocations framework:

- We analyse variants of the Two-Choice process which trade sample efficiency, completeness of information and gap guarantees. For the $(1+\beta)$-process which mixes One-Choice and Two-Choice with probability $\beta \in (0,1]$, we prove tight bounds for small and large β, extending the results of Peres, Talwar and Wieder [152]. Another sample efficient family is that of Two-Thinning processes, which allocate to the two sampled bins in an online manner. For Two-Thinning processes that use as a decision function thresholds relative to the average load or thresholds in the rank domain, we establish tight bounds and also resolve a conjecture by Feldheim and Gurel-Gurevich [75]. We also quantify trade-offs for two-sample processes between the number of queries and the gap bound, establishing a "power of two queries" phenomenon.

- We analyse the Two-Choice process with random, adversarial and delay noise, proving tight bounds for various settings. In the adversarial setting, the adversary can decide in which of the two sampled bins the ball is allocated to, only when the two loads differ by at most g. The analysis of this setting implies bounds for settings with random noise and delay.

  For the setting where load information is updated periodically every b steps, for $b = n$ we tighten the bound of [28] to $\Theta\left(\frac{\log n}{\log \log n}\right)$ and prove that Two-Choice is optimal in this setting for any $b \in [n \cdot e^{-\log^c n}, n \log n]$ for any constant $c > 0$. For $b \in [n \log n, n^3]$, we show that Two-Choice achieves w.h.p. a $\Theta(b/n)$ gap, while surprisingly the $(1+\beta)$-process with appropriately chosen β achieves w.h.p. a $\Theta(\sqrt{\frac{b}{n} \cdot \log n})$ gap, which is optimal over a large family of processes. This proves

that in the presence of outdated information, less aggressive strategies can outperform the greedy processes (such as TWO-CHOICE), which has been empirically observed in the queuing setting for centralised processes since 2000 [58, 134], but to the best of our knowledge has not been formally proven.

- Next we analyse TWO-CHOICE in the graphical setting, where bins are vertices of a graph and each ball is allocated to the lesser loaded of the vertices adjacent to a randomly sampled edge. We extend the results of Kenthapadi and Panigrahy [100] proving that for dense expanders in the heavily-loaded case the gap is w.h.p. $\mathcal{O}(\log\log n)$. In the presence of weights, we make progress towards [152, Open Problem 1] by proving that for graphs with conductance ϕ, the gap is w.h.p. $\mathcal{O}(\log n/\phi)$.

- Further, we introduce and analyse processes which can allocate more than one balls to a sampled bin. We prove that these processes achieve w.h.p. an $\mathcal{O}(\log n)$ gap (which also applies for any d-regular graph), while still being more sample-efficient than ONE-CHOICE ("power of filling").

- For the MEMORY process that can store bins in a cache, we generalise the $\mathcal{O}(\log\log n)$ gap bound by Mitzenmacher, Prabhakar and Shah [136] to the heavily-loaded case and prove a matching lower bound. Further, in the presence of heterogeneous sampling distributions, we establish a striking difference between TWO-CHOICE (or even d-CHOICE with $d = \mathcal{O}(1)$) and MEMORY, showing that for the later the gap is bounded, while for the former it is known to diverge [176] ("power of memory").

Acknowledgements

First of all, I would like to thank God for guiding me and showing me what truly matters in life. If I did anything good over these years it was through His strength.

I am grateful for having Prof Thomas Sauerwald as my supervisor and I am thankful to him for introducing me to potential functions and balanced allocations, as well as spending countless hours discussing with me and working on drafts. I would also like to thank Dr John Sylvester for various discussions and for helping me improve my mathematical writing.

I was fortunate to be part of the Department of Computer Science and Technology, and St John's College who jointly with Cambridge Trust funded my PhD. I thank my DoS: Prof Robert Mullins, my lecturers and supervisors, as well as my supervisees, for many stimulating discussions.

As always I am thankful to my parents, my grandmother, my sister, my friends and several school teachers.

A note to the reader

The dissertation is organised mainly around the analysis techniques:

- Chapter 1 outlines these techniques, presents the processes and settings, and summarises the main results obtained.
- Chapter 2 gives the formal definitions for processes and settings.
- Chapters 3 to 6 present the main analysis techniques. Some direct applications of the relevant techniques can be found in Sections 3.2 and 5.2.2.
- Chapter 7 presents most of the results for the various processes and settings, obtained by applying the techniques in Chapters 3 to 6 mostly as a black box.

For the first read, it is recommended to read Chapter 1 in full and Sections 2.1 and 2.2 regarding the core definitions of balanced allocation processes. The rest of the definitions can be referred to, on a need-to-know basis, by clicking on the process or setting name (you may find Alt + ← useful). Table 2.4 gives a concise summary of the various settings. Then, we suggest to proceed by reading the introductory pages from each of the Chapters 3 to 6 (and perhaps the direct applications in Sections 3.2 and 5.2.2) which present each of the techniques in more detail. Finally, read applications of interest from Chapter 7, perhaps starting with the example in Section 7.1.

Tables A.1 to A.4 in Appendix A give a detailed summary of the lower and upper bounds obtained in this work and those in related work. You may also find the index useful for looking up definitions and results.

A high-level visual introduction to balanced allocations and the results obtained in this thesis can be accessed in the following link:

`dimitrioslos.com/phd-thesis`

Contents

Introduction

1.1 Balanced allocations

In the sequential balanced allocations framework, we are given m *balls* (tasks or jobs) to allocate into n *bins* (servers or machines) indexed by the set $[n] := \{1, \ldots, n\}$. Let x_i^t be the load of the i-th bin after the t-th ball has been allocated. The goal is to minimise the *maximum load* after allocating m balls, i.e., minimise $\max_{i \in [n]} x_i^m$, which is equivalent to minimising the *gap*, where $\mathrm{Gap}(m) = \max_{i \in [n]} \big(x_i^m - \frac{m}{n} \big)$ (see Fig. 1.1).

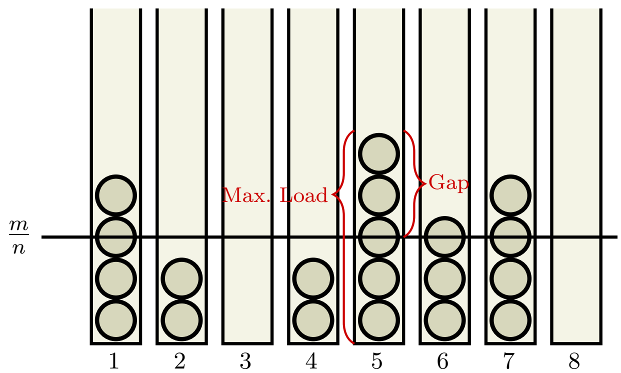


Figure 1.1: Example of a *load vector* $x^m = (4, 2, 0, 2, 5, 3, 4, 0)$ for $n = 8$ bins and $m = 20$ balls.

In a *centralised* setting, this problem is easy to solve using Round-Robin allocation, i.e., allocating each ball to the least loaded bin and breaking ties arbitrarily. However, in a *decentralised* setting, we have to consider processes that assume less coordination between the bins. This modelling assumption makes this framework applicable to hashing [177], load balancing [137] and several other areas (see Section 1.5 for further applications).

The simplest such process is the One-Choice process, where each ball is allocated to a bin chosen uniformly at random. In the *lightly-loaded case*, i.e., when $m = n$, it is well established that w.h.p.[1] the gap is $\Theta\big(\frac{\log n}{\log \log n}\big)$ [89] and in the *heavily-loaded case*, i.e., when $m \gg n$ w.h.p. the gap is $\Theta\big(\sqrt{\frac{m}{n} \cdot \log n}\big)$ [157].

Power of two choices. A great improvement over One-Choice is the Two-Choice process, where for each ball two bins are sampled at random and the ball is allocated to the lesser loaded of the two. This process has been experimentally studied in load balancing since at least 1986 in the work by Eager, Lazowska and Zahorjan [68], but was not rigorously analysed until the work by Azar, Broder, Karlin and Upfal [18] (and implicitly by Karp, Luby, and Meyer auf der Heide [99]) who showed that Two-Choice has w.h.p. a gap of $\log_2 \log n + \Theta(1)$ in the lightly-loaded case. This is exponentially better than that of One-Choice. Berenbrink, Czumaj, Steger and Vöcking [29] extended the analysis to the heavily-loaded case, proving that w.h.p. the gap remains $\log_2 \log n + \Theta(1)$. Here the improvement is even more dramatic, as the bound on the gap does not depend on m, as opposed to One-Choice. If instead, for each ball we sample d bins, then w.h.p. the gap is $\log_d \log n + \Theta(1)$. We also remark that

[1] In general, with high probability refers to probability of at least $1 - n^{-c}$ for some constant $c > 0$. For brevity, we may say *gap bound* and by this we mean a w.h.p. bound on the gap of the process in question.

the heavily-loaded case is much more challenging to analyse, because m being unbounded means that with very high probability we will encounter arbitrarily bad configurations and so the analysis needs to demonstrate that we are always able to recover from these.

The practical significance of the "power of two choices" was recognised in the 2020 ACM Paris Kanellakis Theory and Practice award [17]:

> "*[...] it is not surprising that the power of two choices that requires only a local decision rather than global coordination has led to a wide range of practical applications. These include i-Google's web index, Akamai's overlay routing network, and highly reliable distributed data storage systems used by Microsoft and Dropbox, which are all based on variants of the power of two choices paradigm. There are many other software systems that use balanced allocations as an important ingredient.*"

Motivation for variants and settings. The sequential balanced allocation setting makes several simplifying assumptions compared to real-world applications, giving rise to the following important questions:

1. What if we cannot always *access both samples*? Can we relax the *coordination* required by the two sampled bins?
2. What if the reported load of a bin at time t is *outdated*, e.g., the reported load might be as small as the load at an earlier time $t - \tau$ for some parameter τ?
3. What if the reported load of a bin at time t is subject to some *adversarial* (or *random*) *noise*, e.g., the reported load of a bin might be an adversarial (or random) perturbation from the exact load within some range g?

For these reasons, Two-Choice has been studied in various settings, and several variants of this process have been proposed. Notably, the Paris Kanellakis award mentions that practical adaptations are "based on variants of the power of two choices paradigm".

The main focus of this dissertation is to address these questions for Two-Choice and other processes, by developing the necessary analysis toolkit; and also to devise processes that overcome limitations of Two-Choice. Below, we will provide a brief overview of the variants and settings we study, and then summarise our results. Additional related work will be presented in Section 1.5.

Variants of Two-Choice. Mitzenmacher [131] introduced the OnePlusBeta–process, which with probability β it performs Two-Choice and with probability $1-\beta$ it performs One-Choice. Peres, Talwar and Wieder [152] showed that this process w.h.p. achieves an $\mathcal{O}((\log n)/\beta + \log(1/\beta)/\beta)$ gap in the heavily-loaded case, which is tight for any $\operatorname{poly}(n^{-1}) \leqslant \beta \leqslant 1-\epsilon$, for any constant $\epsilon \in (0,1)$. This process has the advantage that it does not make use of the second sample in every step, leading to the improved *sample efficiency* of $1+\beta$ samples per allocation, while maintaining a bounded gap.

Another family of processes that improve on the sample efficiency of Two-Choice is that of d-Thinning[2] processes. Here, for each ball, up to d bins may be sampled and the decision on whether to accept or reject a bin is made in an online manner. This process was empirically studied by Zhou [180] and only recently, Feldheim and Gurel-Gurevich [75] proved that the asymptotically optimal gap for Two-Thinning is $\Theta\Big(\sqrt{\frac{\log n}{\log\log n}}\Big)$ in the lightly-loaded case. They also conjectured that this bound extends

[2]We use red colour for concrete processes, blue for family of processes, and orange for settings. The distinction between the first two is often blurry. The hyperlink on the process name redirects to the formal definition of the process.

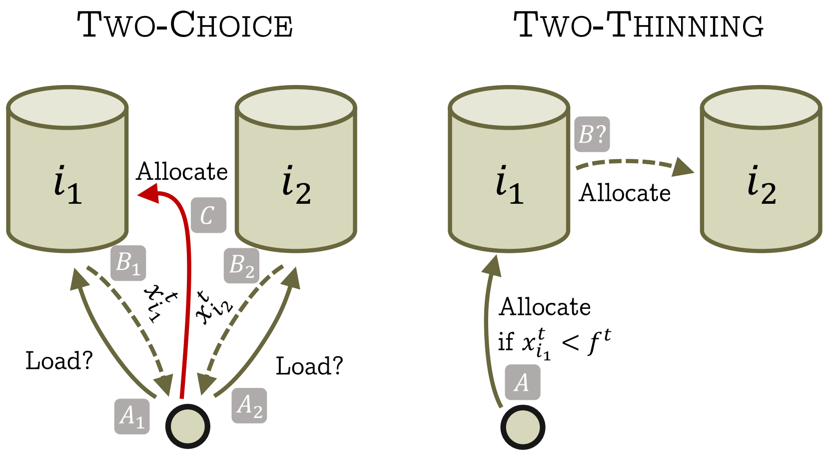


Figure 1.2: (**left**) The three stages A, B, C for allocating a request with the TWO-CHOICE process, which assumes that no other request is being allocated to the two sampled bins (so the two bins are being "held" until the allocation is complete). (**right**) The (at most) two stages for allocating a request with a TWO-THINNING process with threshold f^t. Here, none of the bins needs to be "held".

to the heavily-loaded case. Compared to TWO-CHOICE, this process has the additional advantage that it does not require to "hold" the two bins until the allocation is completed, i.e., until it receives both reported loads, and so can perform allocations independently at each server (see Fig. 1.2). Particularly attractive are the RELATIVE-THRESHOLD($f(n)$) processes, where at step $t \geqslant 0$, the decision function is a threshold $\frac{t}{n} + f(n)$, for some *offset* function $f : \mathbb{N} \to \mathbb{Z}$. This means that upon receiving a ball, the bin can decide whether to accept the ball or forward it to another random bin, requiring just an estimate for the average $\lceil t/n \rceil$, which changes only every n allocations.

A variant of d-CHOICE with memory was presented in [136, 166]. In the (d, M)-MEMORY process, in addition to the d samples taken for each ball, the process can cache M bins and use them in future allocations. In [136], for $M = d = 1$, it was shown that in the lightly-loaded case this process has w.h.p. a gap of $\log_\phi \log n + \Theta(1)$, where $\phi = \frac{1+\sqrt{5}}{2}$ is the golden ratio.

Settings. To model outdated information, Mitzenmacher [134] and Dahlin [58] introduced the b-BATCHED setting (also called *periodic bulletin model*), where the load of the bins is updated every b steps (Fig. 1.3). They used a fluid-limit approximation and empirical studies to observe that when b is large, increasing d in the d-CHOICE process leads to a worse gap. This may seem counter-intuitive given that, for $b = 1$ (no delay), d-CHOICE is the optimal out of all processes that use d samples for each allocation (Corollary 2.7). The only rigorous bound for the b-BATCHED setting was given by Berenbrink, Czumaj, Englert, Friedetzky and Nagel [28], who proved that for TWO-CHOICE and $b = n$, w.h.p. the gap is $\mathcal{O}(\log n)$ in the heavily-loaded case. A related setting where balls are again allocated in batches, but some of them are also processed/removed, was studied in [31] and in the *Repeated Balls-into-Bins setting* [23, 45–47]. The question of which processes perform well under outdated information has also been empirically investigated in the queuing setting [12, 82, 106, 175]. In particular, Whitt [175] states the following regarding optimal rules:

> "*We have shown that several natural selection rules are not optimal in various situations, but we have not identified any optimal rules. Identifying optimal rules in these situations would*

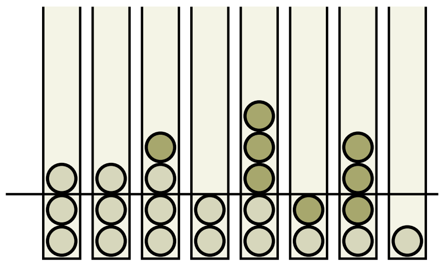


Figure 1.3: An instance of the b-BATCHED process for $b = n$. The b shaded balls are all allocated simultaneously.

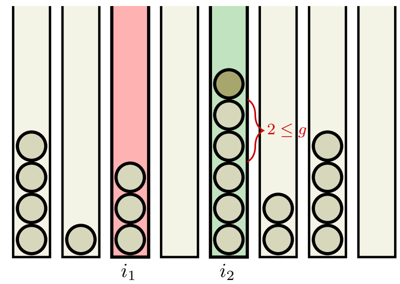


Figure 1.4: An instance of the g-BOUNDED process for $g = 3$. The two sampled bins i_1 and i_2 have a load difference of 2 and hence the adversary *can* force the (shaded) ball to be allocated to the heavier bin.

obviously be interesting, but appears to be difficult. Moreover, knowing an optimal rule might not be so useful because the optimal rule may be very complicated."

Alistarh, Brown, Kopinsky, Li and Nadiradze [7] investigated an adversarial version of TWO-CHOICE, called the g-BOUNDED process, where for each ball two bins are sampled, and if their load difference is at most g, then the ball is allocated to the heavier bin; otherwise, it is allocated to the lesser loaded bin (Fig. 1.4). This process was used to analyse relaxed concurrent data structures, such as the *multi-counter*.

Wieder [176] investigated the d-CHOICE process where the bins are sampled according to a skewed distribution. They proved that the $\mathcal{O}(\log\log n)$ gap bound remains, as long as the probability s_i of sampling a bin $i \in [n]$ satisfies $\frac{1}{an} \leqslant s_i \leqslant \frac{b}{n}$ for some constants $a := a(d)$ and $b := b(d)$, otherwise the gap may diverge. A related setting is the GRAPHICAL setting (or *graphical allocations*), where the bins are vertices of a graph G and each ball is allocated to the least loaded of the two adjacent bins of an edge sampled uniformly at random. Kenthapadi and Panigrahy [100] recovered the $\mathcal{O}(\log\log n)$ bound on expander graphs in the lightly-loaded case and [152] proved an $\mathcal{O}(\log n/\phi)$ bound for graphs with conductance ϕ in the heavily-loaded case. Bansal and Feldheim analysed a sophisticated algorithm that achieves a poly-logarithmic gap on sparse regular graphs [20], and Greenhill, Mans and Pourmiri studied graphical allocation on dynamic hypergraphs [90].

The WEIGHTED setting has also been studied for several processes. In particular, the bounds obtained for a large family of processes in [152] also hold for weights sampled from distributions with finite moment generating functions (MGFs). Talwar and Wieder [169], proved tight $o(\log n)$ bounds for d-CHOICE for a wide class of weight distributions satisfying some mild conditions on their second and fourth moments. A model with heterogeneous bin capacities was studied by Berenbrink, Brinkmann, Friedetzky and Nagel [27], who showed that the gap bound of $\log_d \log n + \mathcal{O}(1)$ for d-CHOICE continues to apply (see also [30, 32, 129]).

1.2 Main results and new processes

In this section, we state the main results of this work. We develop a set of proof techniques based on potential functions and use them to analyse numerous processes in the heavily-loaded case. These techniques allow us to prove new results for existing processes in various settings, develop new processes and new settings, and obtain insights into balanced allocations, which could also be useful in real-world settings.

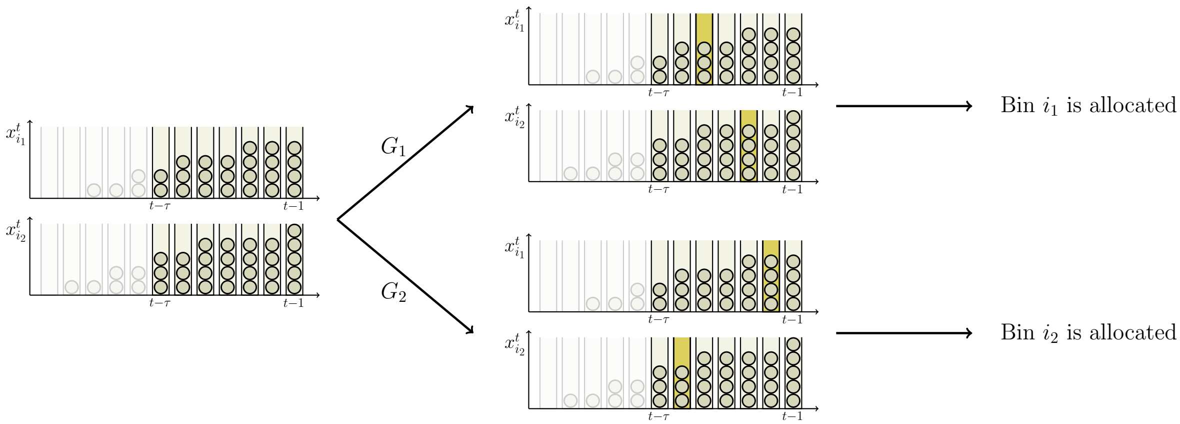


Figure 1.5: The load evolution of the two sampled bins i_1 and i_2, and two adversaries G_1 and G_2 reporting different bin loads (shown in yellow), forcing a different allocation for TWO-CHOICE in the τ-DELAY setting (with $\tau = 7$). The adversaries can report any of the non-shaded loads for the allocation of the t-th ball.

1.2.1 Outdated information settings

We begin by stating our results for settings with outdated information. For TWO-CHOICE in the b-BATCHED setting with $b = n$, we tighten the bound by [28] from $\mathcal{O}(\log n)$ to $\Theta\left(\frac{\log n}{\log\log n}\right)$, which matches the gap of ONE-CHOICE with n balls. We extend this to show that for any $b \in [n \cdot e^{-\log^c n}, n \log n]$, it follows the gap of ONE-CHOICE with b balls and, hence, it is asymptotically optimal over all processes sampling two bins for each ball (Section 7.4.3). We also show that these bounds hold in a more relaxed setting, which we call the τ-DELAY setting, where, once a bin is sampled, an adversary can choose to report the load of the bin in any of the last τ steps (see Fig. 1.5). Therefore, τ-DELAY subsumes the b-BATCHED setting for $b = \tau$ and also relaxes the requirement that all bins synchronise their loads at the same step.

On the contrary, for $b \in [\omega(n \log n), n^3]$, we show that TWO-CHOICE has w.h.p. a $\Theta\left(\frac{b}{n}\right)$ gap and that increasing the number of choices d in the d-CHOICE may lead w.h.p. to a worse gap, confirming the empirical observations of [58] and [134]. We also prove the surprising fact that the OnePlusBeta–process with an appropriate choice of parameter β, which is the mix of ONE-CHOICE and TWO-CHOICE, achieves the asymptotically optimal $\Theta\left(\sqrt{\frac{b}{n} \cdot \log n}\right)$ gap for all processes, giving a roughly quadratic improvement over TWO-CHOICE. These bounds also extend to the WEIGHTED setting and apply to a large family of processes (Section 7.5).

1.2.2 Adversarial and random noise settings

Next, we analyse the additive adversarial setting g-ADV for TWO-CHOICE, where an adversary can influence the allocation decision if the loads of the two sampled bins differ by at most g. In Section 7.4.1, we show that for any adversary, w.h.p. the gap of this process is

$$\mathcal{O}\left(g + \frac{g}{\log g} \cdot \log\log n\right).$$

To better understand this bound, we see that for $g = \Theta(1)$, we recover the $\mathcal{O}(\log\log n)$ gap bound for TWO-CHOICE and for $g = \Omega(\text{polylog}(n))$ the bound grows linearly in g. For $\omega(1) = g = o(\text{poly}(\log n))$, it grows sublinearly in g. We show this bound is tight for the g-BOUNDED process, improving on the

$\mathcal{O}(g\log(ng))$ bound by [7]. We also show that this bound is tight for a weaker version of this process, namely, the g-MYOPIC-COMP process, where the adversary allocates randomly whenever the two loads are close (Appendix C.2).

This analysis of the general g-ADV setting allows us to obtain tight (or nearly tight) bounds for a large number of interesting processes and settings, including TWO-CHOICE with random additive noise and the aforementioned b-BATCHED and τ-DELAY settings (see Section 7.4.3).

1.2.3 Thinning processes

Another main focus of our work are TWO-THINNING processes. For these processes, we disprove the conjecture by Feldheim and Gurel-Gurevich [76, Open Problem 1.3], claiming that the $\mathcal{O}\big(\sqrt{\frac{\log n}{\log\log n}}\big)$ gap bound extends to the heavily-loaded setting. We do this by showing that any TWO-THINNING process w.h.p. has an $\Omega(\sqrt{\log n})$ gap at $m=\Theta(n\sqrt{\log n})$ and an $\Omega\big(\frac{\log n}{\log\log n}\big)$ gap at least once in an interval of length $\Theta(n\log^2 n)$ (see Appendix C.3).

Further, we introduce two families of TWO-THINNING processes, namely the RELATIVE-THRESHOLD and the QUANTILE processes. In the RELATIVE-THRESHOLD($f(n)$) process, the t-th ball is allocated to the first bin sample if its load is at most $\frac{t}{n}+f(n)$, otherwise, it is allocated to the second sample. In the QUANTILE(δ) process, a ball is allocated to the first bin sample if the rank of the sampled bin in the sorted load vector is greater than δn.

For two specific instances of these processes MEAN-THINNING (= RELATIVE-THRESHOLD(0)) and MEDIAN-QUANTILE (= QUANTILE(1/2)), we prove that w.h.p. the gap is $\Theta(\log n)$ in the heavily-loaded case. As a corollary for the analysis of the MEAN-THINNING process, we obtain tight bounds for any RELATIVE-THRESHOLD($f(n)$) process with an offset $f(n)=\Omega(\log n)$ (see Section 5.2.2).

Subsequently, it was shown by Feldheim, Gurel-Gurevich and Li [76] that there is a TWO-THINNING process whose decisions depend on the entire history of the process and which w.h.p. achieves the optimal $\mathcal{O}\big(\frac{\log n}{\log\log n}\big)$ gap, matching our lower bound. We complement these results by showing that the more lightweight QUANTILE$\big(\frac{(\log\log n)^2}{\log n}\big)$ process also achieves this optimal gap (see Section 7.2.3).

Furthermore, we consider extensions of the QUANTILE and RELATIVE-THRESHOLD processes, the k-QUANTILE($\delta_1,\ldots,\delta_k$) and k-THRESHOLD($f_1,\ldots,f_k$) processes. In these processes, two bins are sampled and their loads can be distinguished if they appear in different bands of the quantile domain, i.e., $(0,\delta_1 n],\ldots,(\delta_k n,1]$, or of the load domain, i.e., $(\infty,f_1],\ldots,(f_{k-1},f_k]$ respectively (see Fig. 1.6). These can be interpreted as a version of TWO-CHOICE with incomplete information. For any $1\leqslant k=\mathcal{O}(\log\log n)$, we show that for some choice of k quantiles, the k-DENSE-QUANTILE and for some choice of k offsets, the k-DENSE-THRESHOLD process achieves w.h.p. an $\mathcal{O}(k\cdot(\log n)^{1/k})$ gap (see Section 7.2). For $k=2$, this establishes almost a quadratic improvement in the gap over $k=1$. We call this the "power of two queries" phenomenon, extending the results of [97] to the heavily-loaded case. By majorisation, we also obtain bounds for the
OnePlusBeta–process for β close to 1 (see Section 7.3.2), recover the $\mathcal{O}(\log\log n)$ bound for TWO-CHOICE, and obtain near-tight bounds for d-THINNING (see Section 7.3). Further, we obtain bounds for TWO-CHOICE in the GRAPHICAL setting, which as a special case extends the $\Theta(\log\log n)$ bound for graphs with poly(n) expansion by [100] to the heavily-loaded case.

1.2.4 Additional processes

Power of filling. We investigate processes that can allocate more than one ball at a single step. In particular, we introduce the TWINNING process, which samples one bin per step and allocates two balls if the load of the bin is at most the average (underloaded); otherwise allocates one ball. Extending this

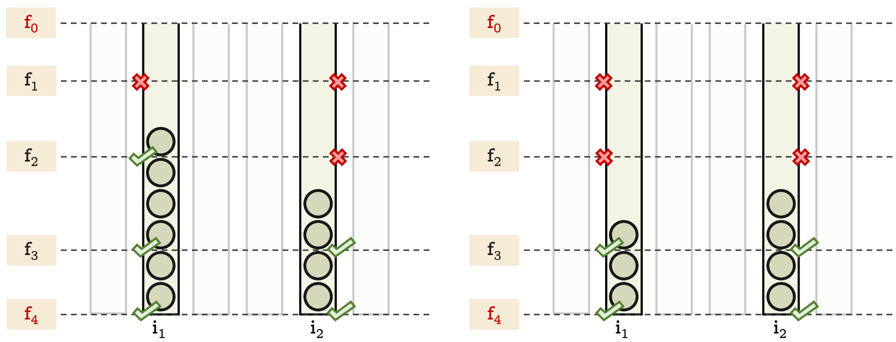


Figure 1.6: Two instances of a 3-THRESHOLD(f_1, f_2, f_3) process and its interpretation as TWO-CHOICE with incomplete information, being able to distinguish two bins iff they are in a different "load band" as defined by the thresholds. (**Left**): The loads of the two bins are in different bands and so the process allocates to the lesser loaded bin. (**Right**): The loads of the two bins are in the same band and so the process cannot distinguish between the two bins, so it allocates randomly.

idea further, we analysed the PACKING process, which allocates as many balls to underloaded bins as to make the bin overloaded (just by one ball). For both processes, we proved[3] a $\Theta(\log n)$ gap in the heavily-loaded case (Corollary 5.12), demonstrating a different way for maintaining balanced allocations, while at the same time being more sample efficient than ONE-CHOICE and also applicable to the GRAPHICAL setting for any d-regular graph.

GRAPHICAL setting. For the GRAPHICAL setting, we make progress towards [152, Open problem 1] by showing that the bounds hold even in the presence of weights sampled from a distribution with finite MGFs (see Section 7.6.1).

OnePlusBeta–process. In addition to the aforementioned bounds, for the $(1+\beta)$-process with β close to 1, we prove a tight bound for $\beta = \text{poly}(n^{-\omega(1)})$ (see Section 3.2.1). Therefore, we obtain an almost complete characterisation for the gap of this process for any $\beta \in (0, 1]$.

Memory. Finally, we analysed the MEMORY process in the heavily-loaded case and showed that it achieves a $\Theta(\log \log n)$ gap even in the presence of arbitrary constant imbalance in the bin sampling distribution $\mathcal{S}$, i.e., $\mathcal{S}$ satisfying $\frac{a}{n} \leqslant \mathcal{S}_i \leqslant \frac{b}{n}$ for each bin $i \in [n]$. This is in stark difference to the d-CHOICE process for $d = \mathcal{O}(1)$, where for some sufficiently large (constant) imbalance, the gap becomes $\Omega(\sqrt{\frac{m}{n} \cdot \log n})$ (see Section 7.7). We complement our results by proving an $\Theta(\log n)$ gap for MEMORY where the cache resets every d steps, thus demonstrating the robustness of MEMORY.[4]

1.2.5 Published papers

The work presented in this dissertation is based on the following published papers:

- Dimitrios Los and Thomas Sauerwald. 2022. Balanced Allocations with Incomplete Information: The Power of Two Queries. In *13th Innovations in Theoretical Computer Science Conference (ITCS'22)*. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, pages 103:1–103:23. [**All results**]

[3]Due to space constraints the proof of PACKING is omitted. It can be found in [117, Section 5].
[4]Due to space constraints, the proofs for MEMORY are only outlined. These can be found in [118].

- Dimitrios Los, Thomas Sauerwald, and John Sylvester. 2022. Balanced Allocations: Caching and Packing, Twinning and Thinning. In *33rd Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'22)*. SIAM, pages 1847–1874. [**Partial:** Results for MEAN-BIASED processes and proofs outlined]

- Dimitrios Los and Thomas Sauerwald. 2022. Balanced Allocations with the Choice of Noise. In *41st Annual ACM-SIGOPT Principles of Distributed Computing (PODC'22)*. ACM, pages 164–175. [**All results**]

- Dimitrios Los and Thomas Sauerwald. 2022. Balanced Allocations in Batches: Simplified and Generalized. In *34th Annual ACM Symposium on Parallel Algorithms and Architectures (SPAA'22)*. ACM, pages 389–399. [**All results**]

- Dimitrios Los and Thomas Sauerwald. 2023. Tight Bounds for Repeated Balls-Into-Bins. In *40th International Symposium on Theoretical Aspects of Computer Science (STACS'23)*, Vol. 254. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, pages 45:1–45:22. [**None**]

- Dimitrios Los, Thomas Sauerwald, and John Sylvester. 2023. Balanced Allocations with Heterogeneous Bins: The Power of Memory. In *34th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'23)*. SIAM, pages 4448–4477. [**Results and proofs outlined**]

- Dimitrios Los and Thomas Sauerwald. 2023. Balanced Allocations in Batches: The Tower of Two Choices. In *35th Annual ACM Symposium on Parallel Algorithms and Architectures (SPAA'23)*. ACM, 51–61. [**Partial**: Proofs for tighter bounds omitted]

Our paper on noisy processes [113] was awarded the "Best Student Paper Award" at PODC 2022 [6] and was one of the two papers of the conference to be invited to the *Journal of the ACM*.

1.3 A brief overview of the techniques

We obtain most of the aforementioned results by using a set of techniques on aggregate functions of the load vector, the so-called *potential functions*.

We start with some (necessary) preliminary definitions, which we revisit in more detail in the beginning of Chapter 2. Each process, based on its history $\mathfrak{F}^t$ (which includes the allocation of the t-th ball), induces a *probability allocation vector* q^t, whose i-th entry gives the probability to allocate the $(t+1)$-th ball to bin i. For some processes, such as TWO-CHOICE, the *sorted probability allocation vector* $\widetilde{q}^t$, whose i-th entry gives the probability to allocate to the i-th most loaded bin is *time-homogeneous* (modulo moving probability between bins with the same load; see Theorem 2.1). For instance, for TWO-CHOICE, the sorted probability allocation vector is given by

$$\widetilde{q}_i^t := \frac{2i-1}{n^2}, \quad \text{for any } i \in [n],$$

and for the $(1+\beta)$-process with $\beta \in [0,1]$,

$$\widetilde{q}_i^t := \beta \cdot \frac{2i-1}{n^2} + (1-\beta) \cdot \frac{1}{n}, \quad \text{for any } i \in [n].$$

A particularly "nice" family (we shall soon explain why) of $\widetilde{q}^t$ vectors are the ones that are (i) *non-decreasing* (so, there is a larger probability to allocate to lighter bins) and for which (ii) there exists a constant $\delta \in (0,1)$ and a not-necessarily constant $\epsilon \in (0,1)$ such that

$$\widetilde{q}_i^t \leqslant \frac{1-\epsilon}{n}, \quad \text{for any } i \leqslant n\delta,$$

meaning that there is a significant *probability bias* to allocate away from heavy bins. For instance, for TWO-CHOICE one can pick $\delta = 1/4$ and $\epsilon = 1/2$, and for the $(1+\beta)$-process, $\delta = 1/4$ and $\epsilon = \beta/2$.

The hyperbolic cosine potential. Peres, Talwar and Wieder [152] used the hyperbolic cosine potential, defined for some smoothing parameter $\gamma = \mathcal{O}(1)$ as

$$\Gamma^t := \Gamma^t(\gamma) := \sum_{i=1}^{n} \left(e^{\gamma(x_i^t - t/n)} + e^{-\gamma(x_i^t - t/n)} \right) = \sum_{i=1}^{n} 2 \cdot \cosh\left(\gamma \cdot (x_i^t - t/n)\right).$$

If at some step $t \geqslant 0$, we have that $\Gamma^t = \mathrm{poly}(n)$, then this implies that $\mathrm{Gap}(t) = \mathcal{O}\big(\frac{\log n}{\gamma}\big)$. In [152], they showed that the second term is necessary to deduce the following *drop inequality* when $\gamma = \mathcal{O}(\epsilon)$, for *any* step $t \geqslant 0$ (and for any history of the process $\mathfrak{F}^t$):

$$\mathbf{E}\left[\, \Gamma^{t+1} \,\middle|\, \mathfrak{F}^t \,\right] \leqslant \Gamma^t \cdot \left(1 - \frac{\kappa_1}{n}\right) + \kappa_2,$$

for some not necessarily constants $\kappa_1 := \kappa_1(\gamma, \epsilon) > 0$ and $\kappa_2 := \kappa_2(\gamma, \epsilon) > 0$. This inequality is enough to deduce (using simple induction) that for any step $t \geqslant 0$

$$\mathbf{E}\left[\, \Gamma^t \,\right] \leqslant \frac{\kappa_2}{\kappa_1} \cdot n.$$

By Markov's inequality and because each of the terms in the potential is positive, this implies w.h.p. an $\mathcal{O}\big(\frac{\log n}{\epsilon}\big)$ gap bound for these "nice" probability allocation vectors that we defined above. In particular, for TWO-CHOICE this proves an $\mathcal{O}(\log n)$ bound on the gap and for the $(1+\beta)$-process an $\mathcal{O}\big(\frac{\log n}{\beta}\big)$ bound for any $\Omega(\mathrm{poly}(n^{-1})) \leqslant \beta \leqslant 1 - \widetilde{\epsilon}$, for any constant $\widetilde{\epsilon} \in (0,1)$. For $\beta = o(\mathrm{poly}(n^{-1}))$, the bound no longer applies as $\kappa_1/\kappa_2 = \omega(\mathrm{poly}(n))$.

Now we are ready to present the main techniques that we use in this dissertation:

Technique 1: Expectation bounds for Γ. We extend the analysis in [152] so that it works for the b-BATCHED, WEIGHTED and GRAPHICAL settings. We also tighten the drop inequality to show that there exist constants $c_1, c_2 > 0$, such that for sufficiently small $\gamma > 0$ and for any $t \geqslant 0$,

$$\mathbf{E}\left[\, \Gamma^{t+1} \,\middle|\, \mathfrak{F}^t \,\right] \leqslant \Gamma^t \cdot \left(1 - \frac{c_1 \gamma \epsilon}{n}\right) + c_2 \gamma \epsilon.$$

This allows us to deduce that $\mathbf{E}\left[\, \Gamma^t \,\right] \leqslant \frac{c_2}{c_1} \cdot n$ at any step $t \geqslant 0$ (Theorem 3.2), which directly implies the tight $\mathcal{O}\big(\frac{\log n}{\beta}\big)$ bound for $(1+\beta)$-process for any $\beta \leqslant 1/2$. However, even more significant is the fact that it enables us to prove that Γ is concentrated at $\mathcal{O}(n)$. In turn, this allows us to characterise the shape of the load vector w.h.p. and gives sufficient conditions for stronger potential functions to drop in expectation, as we explain in Technique 2.

Technique 2: Analysis of super-exponential potentials. The analysis in [152] cannot be directly used to prove $o(\log n)$ bounds on the gap, since it applies only for $\gamma = \mathcal{O}(1)$. To obtain $o(\log n)$ bounds, we use *super-exponential potential functions* of the following form:

$$\Phi^t := \Phi^t(\phi, z) = \sum_{i=1}^{n} e^{\phi (x_i^t - \frac{t}{n} - z)^+}$$

for some smoothing parameter $\phi = \omega(1)$, some offset $z > 0$, and where $u^+ := \max\{u, 0\}$. These potentials may *increase* in expectation in *some* steps, even if they are large. However, we show they satisfy the conditional drop inequality (Lemma 6.2)

$$\mathbf{E}\left[\, \Phi^{t+1} \,\middle|\, \mathfrak{F}^t, \widetilde{\mathcal{K}}_\phi^t \,\right] \leqslant \Phi^t \cdot \left(1 - \frac{1}{n}\right) + 2,$$

whenever the following event holds

$$\widetilde{\mathcal{K}}_\phi^t := \left\{ \forall i \in [n]\colon\ x_i^t \geqslant \frac{t}{n} + z - 1 \ \Rightarrow\ q_i^t \leqslant \frac{1}{n} \cdot e^{-\phi} \right\}.$$

For the full details, see Chapter 6.

Technique 3: Concentration bounds for Γ. For some processes, like the Two-Choice and the k-Dense-Quantile processes, in order to show that event $\widetilde{\mathcal{K}}_\phi^t$ holds in some step t, we bound the number of bins with a load above some offset z.

If at some step $t \geqslant 0$, it holds that $\Gamma^t \leqslant cn$, then we have that the number of bins with load at least $\frac{t}{n} + z$ is at most

$$cn \cdot e^{-\gamma z}.$$

On a very high level, for Two-Choice and $\Phi := \Phi(\phi, z)$ with $\phi = \Theta(1/\sqrt{\log n})$ and $z = \Theta(\sqrt{\log n})$, we can show that the potential drops in expectation and by waiting sufficiently long, Φ becomes $\mathcal{O}(n)$ implying an $\mathcal{O}(\sqrt{\log n})$ bound on the gap. The details of these derivations can be found in Chapter 7.

This motivates us to obtain concentration bounds for Γ (and other super-exponential potentials). To achieve this, we use two potential functions $\Gamma_1 := \Gamma(\gamma_1)$ and $\Gamma_2 := \Gamma_2(\gamma_2)$ with smoothing parameters satisfying $\gamma_2 = \Theta(\gamma_1)$ and $\gamma_2 < \gamma_1$. The interplay is such that when $\Gamma_1 = \mathrm{poly}(n)$ (which we obtain by Markov's inequality), we also get that Γ_2^t cannot change much over the next step, i.e., $\left|\Gamma_2^{t+1} - \Gamma_2^t\right| = \mathcal{O}(n^{\widetilde{\epsilon}})$, for some constant $\widetilde{\epsilon} \in (0,1)$. These conditional bounded differences allow us to apply a concentration inequality with a bad event to obtain that Γ_2 is concentrated (see Chapter 3).

Technique 4: Layered induction over super-exponential potentials. To establish an $o(\log n)$ bound on the gap, we define a series of exponential and super-exponential potential functions Φ_j with smoothing parameters $\phi_1 \leqslant \ldots \leqslant \phi_k$. By using Techniques 1 and 3, we obtain that $\Phi_1 = \mathcal{O}(n)$, and conditioning on that event for an $\mathcal{O}(n \cdot \mathrm{polylog}(n))$ interval, it follows that $\widetilde{\mathcal{K}}_{\phi_{j+1}}$ also holds. So eventually, $\Phi_{j+1} = \Phi_{j+1}(\phi_{j+1}, z_{j+1})$ becomes $\mathcal{O}(n)$, for an appropriately chosen offset z_{j+1}. Through this layered induction, which is over super-exponential potentials (see Fig. 1.7), we conclude that $\Phi_k^m = \mathcal{O}(n)$, thus the gap is $\mathcal{O}\big(z_k + \frac{\log n}{\phi_k}\big)$.

This technique (with some small variations) can be used to deduce tight gaps for several processes, including k-Dense-Quantile, k-Dense-Threshold, $(1+\beta)$, Two-Choice in the g-Adv setting and a few more (see Chapter 7).

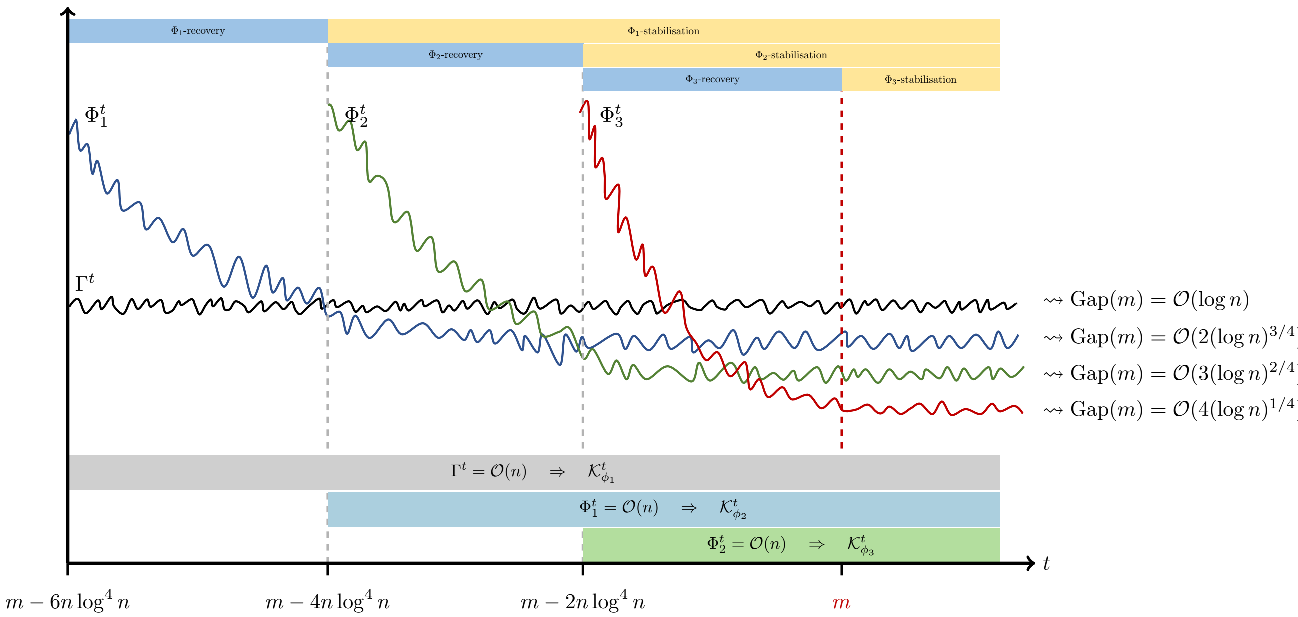


Figure 1.7: *Layered induction* for the k-DENSE-QUANTILE process for $k = 4$, showing that when the potential at layer j satisfies $\Phi_j^t = \mathcal{O}(n)$, the the drop condition at layer $j+1$, i.e., $\widetilde{\mathcal{K}}_{\phi_{j+1}}^t$ is implied. Then, after a *recovery phase* the potential Φ_{j+1} *stabilises* at $\mathcal{O}(n)$ and implies a tighter bound on the gap.

Technique 5: Interplay between absolute value and quadratic potentials. For the MEAN-THINNING, TWINNING and TWO-CHOICE in the g-ADV setting (for $g \leqslant \log n$), the hyperbolic cosine potential $\Gamma := \Gamma(\gamma)$ (for constant $\gamma > 0$) may not satisfy the drop inequality in every step.

For example, in the MEAN-THINNING process, if almost all bins' loads exceed the average, then the potential Γ may increase in expectation. In more technical terms, if δ^t is the quantile of the average, then Γ satisfies the drop inequality if at step $t \geqslant 0$, we have $\delta^t \in (\epsilon, 1-\epsilon)$ and we call such step, *good*. So, the task becomes to show that, in a sufficiently long interval, there are sufficiently many good steps.

To show this, we use an interplay between the *absolute value* and the *quadratic* potential, which are defined as

$$\Delta^t := \sum_{i=1}^{n} \left| x_i^t - \frac{t}{n} \right| \quad \text{and} \quad \Upsilon^t := \sum_{i=1}^{n} \left(x_i^t - \frac{t}{n} \right)^2.$$

In particular, we show that in any step $t \geqslant 0$,

$$\mathbf{E}\left[\Upsilon^{t+1} \,\middle|\, \mathfrak{F}^t \right] \leqslant \Upsilon^t - \frac{\Delta^t}{n} + 1. \tag{1.1}$$

Thus, whenever $\Delta^t = \Omega(n)$, the quadratic potential decreases in expectation. So, there cannot be "too many steps" with $\Delta^t = \Omega(n)$. In steps with $\Delta^t = \mathcal{O}(n)$, we show that there is a constant fraction of good steps in the next $\Theta(n)$ steps, which on aggregate implies the drop inequality for Γ. For TWINNING and g-ADV, the parameters in Eq. (1.1) are slightly different. The details can be found in Chapter 5. In [116], we also used a similar interplay between the quadratic potential and the number of empty bins to analyse the Repeated Balls-into-Bins setting, but the details are not included in this dissertation.

Technique 6: A reallocation argument This is a technique for analysing a process $\mathcal{P}_{\mathcal{A}}$ with allocation vector $q_{\mathcal{A}}^t$, by relating it to a process $\mathcal{P}_{\mathcal{B}}$, whose allocation vector $q_{\mathcal{B}}^t$ is obtained by moving "small amounts" of probability between bins whose loads are "close". We use this technique in two parts in the analysis of the g-ADV setting: (i) for bounding the change of the hyperbolic cosine potential (Section 3.3) and (ii) for bounding the change of the quadratic potential (Section 5.3); and also to formally relate the RESET-MEMORY process to the MEMORY process (see [118, Section 4]).

Fig. 1.8 gives a diagrammatic overview of how these techniques are used to prove the bounds described above.

1.4 Organisation

The dissertation is organised as follows:

In Chapter 2, we give general definitions for the balanced allocations processes and then proceed to define rigorously the various processes and settings. The second part can be skipped on the first read and used as a reference for the remaining of the dissertation.

In Chapter 3, we present the refined analysis for the bounds on the expectation of the hyperbolic potential function used in [152] (Technique 1). Then, we apply the refined analysis to obtain nearly tight gap bounds for a large family of processes in the b-BATCHED setting in the presence of weights. Further, we show how to obtain tight bounds for the $(1+\beta)$-process for small β and for the WEIGHTED GRAPHICAL setting. Finally, we present the reallocation technique (Technique 6) for obtaining an $\mathcal{O}(g \log(ng))$ gap bound for the g-ADV setting.

In Chapter 4, we obtain the high probability concentration of the hyperbolic cosine potential, by using the interplay between two hyperbolic cosine potentials (Technique 3).

In Chapter 5, we present the interplay between the absolute value and the quadratic potentials (Technique 5) for analysing the MEAN-THINNING and TWINNING processes, and the g-ADV setting for $g \geqslant \log n$, giving full details only for the later.

In Chapter 6, we prove a drop inequality for the expectation of the super-exponential potentials and prove a concentration inequality (Techniques 2 and 3).

In Chapter 7, using the lemmas from the last four chapters, we obtain $o(\log n)$ bounds for several processes and settings, including the k-DENSE-QUANTILE, the QUANTILE(δ^*) process, the $(1+\beta)$-process, the k-DENSE-THRESHOLD process, the d-THINNING process and GRAPHICAL setting on dense expanders (Technique 4).

In Chapter 8, we conclude with a summary of results and some open problems.

Finally, Appendix A gives a summary of results in table form, Appendix B lists several standard tools that we used, Appendix C gives some techniques for proving lower bounds and (oftentimes tight) lower bounds for many of the aforementioned processes, and Appendix D contains some of the details of omitted proofs. In Appendix E, we conclude we add some empirical results.

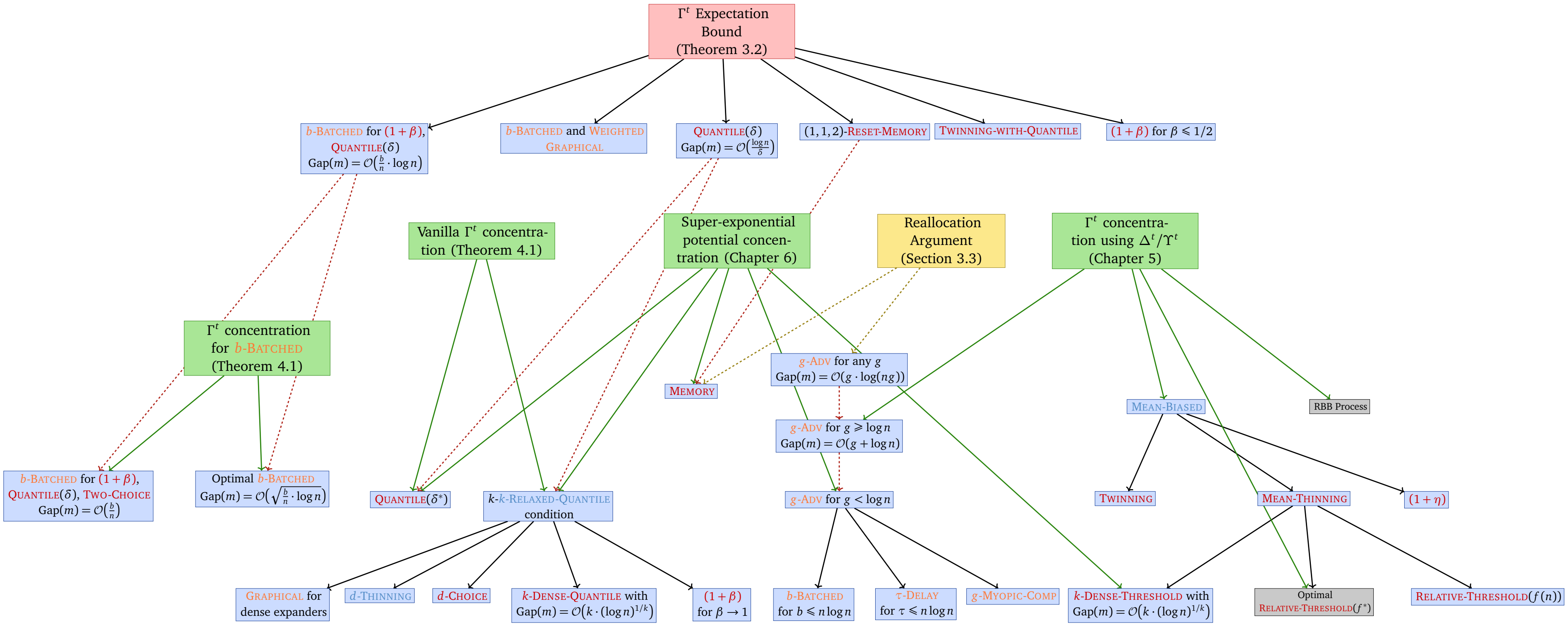


Figure 1.8: Diagram showing the relations between the main theorems and their applications: (i) in red, the tighter analysis for the hyperbolic cosine potential in expectation (Technique 1), (ii) in green, the different concentration bounds for the hyperbolic cosine potential and super-exponential potentials (Techniques 2, 3 and 5), (iii) in yellow, the re-allocation argument (Technique 6), (iv) in blue, the results for the various processes and settings, obtained using Techniques 1-6 and (v) in grey, results not presented in this dissertation.

1.5 Further related work

Randomisation in algorithms has been studied at least since the early years of computing. There are several standard books [67, 138, 141] covering randomised algorithms, their numerous applications and rich theory.

Balanced allocations have been studied under several names, including *occupancy problems*, *balls-into-bins* and *urn processes*. A comprehensive account of early classical occupancy results is provided in several textbooks [59, 98, 102] and some more recent results are included in [103, 124, 151]. Recent surveys [137, 177] in balanced allocations cover several of the results within the power of two choices paradigm.

Further processes. Several other variants of the d-Choice processes have been studied. In the $\textsc{Left}_d$ process, the bins are split into d groups and each of the d samples is uniform and independent from a different group. By breaking ties favouring bins from groups with smaller ids, Vöcking [172] showed that this process achieves a $\frac{\log\log n}{d \cdot \log(\phi_d)} + \Theta(1)$ gap, where ϕ_d is the limit of the root of the generalised Fibonacci sequence.

Relating to d-Thinning, Czumaj and Stemann [56] analysed a large family of processes in the lightly-loaded case, in which given a threshold for each of the steps, a bin is sampled until the load of the sampled bin is below the threshold. Berenbrink, Khodamoradi, Sauerwald and Stauffer [33] analysed the special case where the threshold is always $\lceil \frac{m}{n} + 1 \rceil$, meaning that the process samples a new bin until one with load equal or below the (final) average is found.

Augustine, Moses, Redlich and Upfal [15] introduced $\textsc{First-Diff}_d$ where bins are sampled until the first bin is found whose load differs from the first one or $2^{\Theta(d)}$ bins have been sampled. Redlich [158] analysed the Unfair process, which always performs the opposite decisions to that of Two-Choice. Park [146] analysed the (k,d)-Choice process, which allocates balls to k of the d least loaded bins, and has been used in [144].

Several other variants have been studied in the balanced allocations literature, such as allocations with feedback [49, 66], bins being points in a metric space [35, 43], bins having limited bits to represent their load [11, 26], bins being of bounded size (including applications to Cuckoo hashing) [61, 63, 73, 84, 101, 135, 145], bins being multi-dimensional vectors [13, 19, 40, 53, 143], chains-into-bins [21, 22, 74, 150, 162], balls being allocated in parallel and communicating in rounds to finalise allocation [3, 72, 109, 110], allocations with deletions [20, 54] and allocations with local search [37, 38]. Related versions of these processes have also been studied in the queuing setting [8, 42, 120, 121, 173]. Another line of research has focused on analysing balanced allocation processes with explicit hash function families [1, 48, 50, 57, 64, 149].

Techniques. In the lightly-loaded case, there are several techniques for analysing the Two-Choice process, many of which were later applied to analyse other processes.

Azar, Broder, Karlin and Upfal [18] used the *layered induction* approach where inductively, in layer j out of the total of $\log_2 \log n + \Theta(1)$, the number of bins with load at least j are upper bounded. In the end, a Chernoff bound is used to deduce that all bins have load at most $\log_2 \log n + \mathcal{O}(1)$. This approach has been used to prove tight lower and upper bounds for various processes [3, 54, 56, 138].

The *witness tree technique* [128, 165] is an encoding argument [34] which associates a bad event (e.g., the maximum load being at least $\log_2 \log n + \Omega(1)$) with a combinatorial object. Then by bounding the expected number of such objects, an upper bound on the probability for the bad event is obtained. This technique has also been applied to analyse the $\textsc{Left}_d$ process [172], balls-into-bins with deletions [54] and graphical allocations [90].

Mitzenmacher [133] demonstrated the use of *differential equations* in analysing balanced allocation processes. This technique is related to Wormald's method [174, 178] and the fluid limit method, and has been applied to several processes [15, 122, 130, 158].

Another approach to analysing balanced allocation processes is through the connection to *k-orientable graphs*. A graph is k-orientable if there is an assignment of each edge (u, v) to either u or v such that no vertex is assigned more than k edges. The analysis of $\log_2 \log n$-orientable graphs gives bounds for Two-Choice [44, 80] and other processes [81, 86, 108].

In the heavily-loaded case (especially prior to this dissertation), there is a smaller set of approaches to choose from.

The *Markov-chain based approach* by [29] shows that Two-Choice has a short memory in the sense that after a sufficiently long interval the initial configuration is "forgotten" and then uses a layered-induction style of proof to analyse the process for $\mathrm{poly}(n)$ steps. This was later shown to work for other settings with weights or heterogeneous bins [169, 176].

Potential functions, and more specifically, exponential potential functions, have been employed for analysing a wide range of processes in settings with weights, or in graphical allocations [33, 152].

Finally, the *Poissonisation technique* has been commonly used to analyse One-Choice [102, 138] and, more recently, to analyse some Two-Thinning processes [75, 76].

Applications. Balanced allocations have numerous theoretical and real-world applications. Theoretical applications include those in combinatorial optimisation (matching [78, 95, 140], bin-packing [4, 179], travelling salesman problem [164], online carpooling [91]), private computation [83, 93, 94, 105, 148, 153], stabilising consensus [65], space-efficient population protocols [5, 8, 139], routing and random walks algorithms [10, 16, 87, 156], local computation [125, 159], data structures [36, 60, 161], distributed data structures [7, 9, 88, 147, 155, 160], cake cutting [69, 70], selfish routing [104, 127, 154], controlled random graph processes [2, 24, 25] and group testing [41].

Real-world applications include: Two-Choice as for looking up IPs [39, 92], garbage collection [171], duplicates detection [168], simulation on shared RAMs [55, 62, 99, 128], searchable encryption [14, 167], and routing in interconnection networks [54, 123].

Notation, processes and settings

In Section 2.1, we introduce the notation used in all chapters of this dissertation and prove some basic properties, which are often used in other works, but are not explicitly proven. In Section 2.2, we define the general family of processes sampling d bins for each allocation and in Section 2.3, we introduce a few more important tools and concepts, such as composition of processes and majorisation. In Section 2.4, we introduce numerous settings including weights, batches, delays, noise and sampling with biased distributions. Finally, in Section 2.5, we define the remaining processes that we will be using in the coming chapters.

Sections 2.3 to 2.5 can be skipped when reading for the first time and definitions can be referred to when analysing the relevant processes and settings.

2.1 General notation

In this section, we introduce the notation used to define and analyse balanced allocation processes in the various settings, including weights, noise and outdated information. For convenience, a *setting* is a function that transforms a process into another process. The motivation for doing this is that we only need to define settings once for a wide range of processes, rather than redefining them separately for every process. This way in the coming chapters of the dissertation we will always be analysing processes.

2.1.1 Preliminary mathematical notation

We start with some basic mathematical notation.

- The set of natural numbers is denoted by $\mathbb{N} = \{0, 1, 2, \ldots\}$, the set of positive natural numbers by $\mathbb{N}_+ = \{1, 2, \ldots\}$ and the finite set $[n] = \{1, 2, \ldots, n\}$, for $n \in \mathbb{N}_+$.
- For a vector $v \in \mathbb{R}^n$, we denote the i-th component as v_i, so $v = (v_1, v_2, \ldots, v_n)$.
- The **standard vectors** $\mathbf{e}_k^n \in \mathbb{R}^n$ for $n \in \mathbb{N}_+$ and $k \in [n]$ defined as

$$(\mathbf{e}_k^n)_i = \begin{cases} 1 & \text{if } i = k, \\ 0 & \text{otherwise,} \end{cases}$$

  for all $i \in [n]$. The superscript n is omitted if it can be inferred from the context. The **zero vector** is denoted by $\mathbf{0} = (0, \ldots, 0) \in \mathbb{R}^n$.
- The vector $u = \left(\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}\right)$ is called the **uniform vector**.
- A **probability vector** $p \in \mathbb{R}^n$ is a vector such that $\sum_{i=1}^n p_i = 1$ and $p_i \in [0, 1]$ for any $i \in [n]$.
- We use the shorthand $u \in_{\mathcal{U}} [a, b]$ to denote sampling from the continuous uniform distribution $\mathcal{U}[a, b]$. Similarly, for a finite set S, $i \in_{\mathcal{U}} S$ denotes sampling with probability $1/|S|$ any of the elements of S. More generally, $u \in_p S$ denotes sampling element $i \in S$ with probability p_i for any probability vector p.
- For random variables Y, Z we say that Y is **stochastically smaller** than Z (or equivalently, Y is stochastically dominated by Z) if $\mathbf{Pr}[\, Y \geqslant x \,] \leqslant \mathbf{Pr}[\, Z \geqslant x \,]$ for all $x \in \mathbb{R}$.
- We write $\operatorname{rem}(u, d)$ for the reminder of u divided by d, for any $u \in \mathbb{N}$ and $d \in \mathbb{N}_+$.
- We define the function $\operatorname{arg\,small}_{i \in S}^{M} f(i)$ to return the M items of the ordered set S that achieve the M smallest values of the function $f : S \to \mathbb{R}$, favouring smaller items in case of ties.

2.1.2 Balanced allocations notation

We define balanced allocations processes so that they always allocate balls sequentially. However, as we will see in Sections 2.4.3 and 2.4.4, we can (and will use) these definitions to capture processes and settings where balls are allocated in "parallel".

For a sequential balanced allocation process SEQUENTIAL(q^t) over n bins, we maintain the loads of the bins at **step** (or **time**) $t \in \mathbb{N}$ in a **load vector** $x^t \in \mathbb{R}^n$, where x_i^t gives the load of the bin $i \in [n]$. Initially, we start with the empty load vector $x^0 = \mathbf{0} \in \mathbb{R}^n$. The t-th **ball** for $t \in \{1, 2, \ldots\}$, is allocated using the information from steps $0, 1, \ldots, t-1$, (e.g., using the loads at step $t-1$), to obtain the new load vector $x^t := x^{t-1} + \mathbf{e}_{i^t}$, where $i^t \in [n]$ is the **allocated bin** for the t-th ball (see Fig. 2.1).

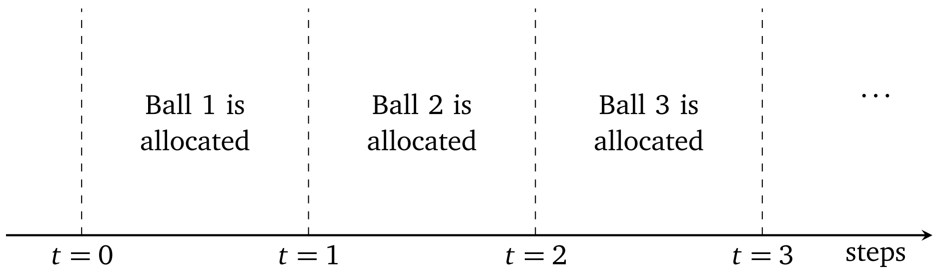


Figure 2.1: Illustration of the relation between steps and balls. In particular, the t-th ball for $t \in \{1, 2, \ldots\}$ is allocated using information in $\mathfrak{F}^{t-1}$ (which includes the load vector x^{t-1} at step $t-1$).

The allocated bin i^t is chosen according to the **probability allocation vector** (or just **allocation vector**)

$$q^{t-1} = q^{t-1}(\mathfrak{F}^{t-1}) = \left(q_1^{t-1}, \ldots, q_n^{t-1}\right),$$

where q_i^{t-1} gives the probability to allocate to bin i and $\mathfrak{F}^{t-1} \in \mathcal{F}^{t-1}$ is the **filtration** of the process, which includes the full history of the process for steps $\{0, 1, \ldots, t-1\}$. The filtration could include more than just the **allocation history** $\mathcal{H}^t = \{i^1, \ldots, i^{t-1}\} \in [n]^{t-1}$. For example, some THINNING processes in [76] make decisions based on the number of times a particular bin has been sampled.

In short, a general sequential allocation process is defined as follows:

SEQUENTIAL(q^t) Process:
Parameter: A function $q : \mathcal{F}^t \to \mathbb{R}^n$ producing a probability vector in $\mathbb{R}^n$.
Initialise: The load vector $x^0 = \mathbf{0} \in \mathbb{R}^n$.
Iteration: At step $t \geqslant 0$, to allocate ball $t+1$:
1. Determine the probability allocation vector $q^t = q^t(\mathfrak{F}^t)$, using the full history $\mathfrak{F}^t \in \mathcal{F}^t$ of the process including step t.
2. Sample the allocated bin $i^{t+1} \in_{q^t} [n]$.
3. Set $x^{t+1} := x^t + \mathbf{e}_{i^{t+1}}$.

Note that most processes for which the next allocation depends on the entire history of the process are impractical to implement. However, this general definition will allow us to capture large families of processes and various interesting settings.

The **normalised load vector** $y^t \in \mathbb{R}^n$ is defined as $y_i^t := x_i^t - \overline{x}^t$ for any $i \in [n]$, where $\overline{x}^t := \frac{1}{n} \cdot \sum_{j=1}^n x_j^t$ is the **average load** at step t. For the **unweighted case**, we have that $\overline{x}^t = t$, but in arbitrary weighted settings this may not hold (see Section 2.4.1). We call a bin $i \in [n]$ **overloaded** at step t, if $y_i^t \geqslant 0$, otherwise we call it **underloaded**. Also, we define the set of **overloaded bins** B_+^t at step t as

$$B_+^t := \left\{ i \in [n] : y_i^t \geqslant 0 \right\},$$

and the set of **underloaded bins** $B_-^t := [n] \setminus B_+^t$.

Any permutation $\ell : [n] \to [n]$, we call a **labelling** of the bins. We will overload notation and allow applying a permutation on a vector $z \in \mathbb{R}^n$ such that $\ell(z) \in \mathbb{R}^n$ and $\ell(z)_j = z_{\ell(j)}$.

We define the labelling $s^t : [n] \to [n]$ that sorts the load vector x^t, i.e., $x_{s^t(1)}^t \geqslant x_{s^t(2)}^t \geqslant \ldots \geqslant x_{s^t(n)}^t$, favouring bins with lower indices in case of a tie. We say that $s^t(j)$ is the j-th most loaded bin. We write $\widetilde{x}^t = s^t(x^t)$ for the **sorted load vector** and $\widetilde{y}^t = s^t(y^t)$ for the **sorted normalised load vector**. We refer to the inverse of s^t as Rank^t, so that $\widetilde{x}_{\text{Rank}^t(i)}^t = x_i^t$, for any bin $i \in [n]$.

The **maximum load** of the process at step t is defined as

$$\max_{i \in [n]} x_i^t = x_{s^t(1)}^t = \widetilde{x}_1^t,$$

and the **minimum load** as

$$\min_{i \in [n]} x_i^t = x_{s^t(n)}^t = \widetilde{x}_n^t.$$

The **gap** of the process at step t is defined as the normalised maximum load, i.e., the difference of the maximum to the average load,

$$\text{Gap}(t) := \max_{i \in [n]} y_i^t = \max_{i \in [n]} \left(x_i^t - \overline{x}^t \right) = \widetilde{y}_1^t.$$

We call a process **Markovian** if $q^t(\mathfrak{F}^t) = q^t(x^t)$, i.e., the probability allocation vector depends only on the current load vector.

A Markovian process is **index-independent** if for any load vector x^t and any labelling ℓ

$$q^t(\ell(x^t)) = \ell(q^t(x^t)).$$

For example, if we define the TWO-CHOICE process to compare the loads of the two randomly sampled bins and to use random tie-breaking, then it is an index-independent process, because bins with the same load get allocated the ball with the same probability. However, if we define TWO-CHOICE to tie break by favouring the bin with the smaller index, then the process is no longer index-independent. We will show that these two versions are equivalent in the following sense: Two process $\mathcal{P}$ and $\mathcal{Q}$ are called **load-vector indistinguishable**, if for any load vector $x \in \mathbb{R}^n$ at any step $m \geqslant 0$,

$$\mathbf{Pr}\left[\, \widetilde{x}_{\mathcal{P}}^m = x \,\right] = \mathbf{Pr}\left[\, \widetilde{x}_{\mathcal{Q}}^m = x \,\right].$$

A **load probability preserving transformation** is a function $f : (\mathbb{R}^n, \mathbb{R}^n) \to \mathbb{R}^n$, transforming the probability vector q^t into $\widehat{q}^t = f(x^t, q^t)$, such that for any load value $v \in \mathbb{R}$, we have

$$\sum_{i \in [n] : x_i^t = v} q_i^t = \sum_{i \in [n] : x_i^t = v} \widehat{q}_i^t.$$

Intuitively, the function f "re-allocates" probability between bins of the same load, so the load and probability vector pairs form equivalence classes under these transformations. For instance, for the load vector $x^t = (3, 4, 3, 4, 6)$, we have that $q_1^t = (0.1, 0.2, 0.15, 0.2, 0.45)$ and $q_2^t = (0.08, 0.15, 0.17, 0.25, 0.45)$ are equivalent, as one can be transformed to the other via a load probability preserving transformation that reallocates a probability of 0.02 from bin 1 to 3 (since $x_1^t = x_3^t = 3$) and of 0.05 from bin 2 to 4 (since $x_2^t = x_4^t = 4$).

Theorem 2.1. *Consider a* <u>*Markovian*</u> <u>*index-independent*</u> *process* $\mathcal{P} = \textsc{Sequential}(q^t)$ *with allocation vector* q^t *and process* $\mathcal{Q} = \textsc{Sequential}(\widehat{q}^t)$ *whose allocation vector* $\widehat{q}^t$ *satisfies* $\widehat{q}^t = f^t(x_{\mathcal{Q}}^t, q^t(x_{\mathcal{Q}}^t))$ *for some* <u>*load probability preserving transformation*</u> f^t *at any step* $t \geqslant 0$*. Then,* $\mathcal{P}$ *and* $\mathcal{Q}$ *are* <u>*load-vector indistinguishable*</u>*.*

Proof. We will define a coupling between the processes $\mathcal{P}$ and $\mathcal{Q}$, so that the sorted load vectors of the two processes are the same, i.e., $\widetilde{x}_{\mathcal{P}}^t = \widetilde{x}_{\mathcal{Q}}^t$ for each step t. At step $t = 0$, this clearly holds as both load vectors are empty, i.e., $\widetilde{x}_{\mathcal{P}}^0 = \widetilde{x}_{\mathcal{Q}}^0 = \mathbf{0}$.

Consider an arbitrary step $t \geqslant 0$ and assume that $\widetilde{x}_{\mathcal{P}}^t = \widetilde{x}_{\mathcal{Q}}^t$. Then consider any labelling ℓ of the bins such that $\ell(x_{\mathcal{P}}^t) = x_{\mathcal{Q}}^t$. Because $\mathcal{P}$ is <u>index-independent</u>, its allocation vector is given by $\ell(q^t(x_{\mathcal{P}}^t))$ after the relabelling ℓ. By the load probability-preserving property, i.e., $\widehat{q}^t = f^t(x_{\mathcal{Q}}^t, q^t(\ell(x_{\mathcal{P}}^t)))$ and so for any load value $v \in \mathbb{R}$

$$r_v := \sum_{i \in [n] : (x_{\mathcal{P}}^t)_{\ell(i)} = v} q_{\ell(i)}^t = \sum_{i \in [n] : (x_{\mathcal{Q}}^t)_i = v} \widehat{q}_i^t.$$

Now, we define the coupling between the two processes:

1. Sample the load value of the bin to increment (not the bin itself), from the (at most) n possible values $v \in \mathbb{R}$ each with probability r_v.

2. Conditional on the load of the bin, for each of the two processes, sample the bin i^{t+1} to increment. This might be different for each process, but their load values will be the same.

Hence, after incrementing a bin with the same load value, the load entries in $x_{\mathcal{P}}^{t+1}$ and $x_{\mathcal{Q}}^{t+1}$ will remain the same, and so $\widetilde{x}_{\mathcal{P}}^{t+1} = \widetilde{x}_{\mathcal{Q}}^{t+1}$. Hence, by induction, we get the claim for all steps. □

We also make the following simple observation.

Observation 2.2. *For any* <u>*index-independent*</u> *process, it is* <u>*load-vector indistinguishable*</u> *regardless of how ties are broken when sorting the load vector.*

We define $\widetilde{q}^t$ to be the **sorted probability allocation vector** at step t, where $\widetilde{q}_i^t$ gives the probability to allocate to the i-th most loaded bin. A process is said to have a **time-homogeneous probability allocation vector** if $\widetilde{q}^t = p(n)$, i.e., the sorted allocation vector does not depend on the time nor on the load vector. These processes are by definition index-independent.

<u>$\textsc{Time-Homogeneous}(p)$</u> Process: Parameters: A probability vector $p \in \mathbb{R}^n$. Iteration: At step $t \geqslant 0$, the allocation vector q^t satisfies $$q_i^t(x^t) = p_{\mathrm{Rank}^t(i)}, \qquad \text{for any } i \in [n].$$

The **averaging transformation** $A : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ is a special <u>load probability preserving transformation</u> which spreads evenly the probability across bins with the same load:

$$(A(x^t, q^t))_i := \frac{1}{\left|\{j \in [n] : x_j^t = x_i^t\}\right|} \cdot \sum_{j \in [n] : x_j^t = x_i^t} q_j^t, \quad \text{for any } i \in [n]. \tag{2.1}$$

This may be viewed as some kind of randomised tie-breaking between bins with the same load. This gives rise to the $\textsc{Time-Homogeneous-with-Rand-Tie-Breaks}(p)$ process which is <u>load-vector indistinguishable</u> from the respective $\textsc{Time-Homogeneous}(p)$ process in the unit weight case (as shown in Theorem 2.1), but in parallel settings (e.g., the b-$\textsc{Batched}$ setting) it may not be (see Section 2.4.3).

TIME-HOMOGENEOUS-WITH-RAND-TIE-BREAKS(p) Process:
Parameters: A probability vector $p \in \mathbb{R}^n$.
Iteration: At step $t \geqslant 0$, the allocation vector q^t is given by the averaging transformation,

$$q_i^t(x^t) = (A(\widetilde{x}^t, p))_{\mathrm{Rank}^t(i)}, \quad \text{for any } i \in [n].$$

The simplest TIME-HOMOGENEOUS(p) process is perhaps the ROUND-ROBIN process, which has $p_i = \mathbf{1}_{i=n}$, meaning that it allocates to a bin with minimum load and so it is the optimal process in the unit weight case.

ROUND-ROBIN Process:
Iteration: At step $t \geqslant 0$, allocate to the bin with minimum load, breaking ties arbitrarily.
[time homo. prob. vector $p_i = \mathbf{1}_{i=n}$, index independent]

2.2 Sample-based processes

2.2.1 d-SAMPLE processes

We now look at a general family of processes which includes the d-CHOICE process. In each step, it samples d bins independently and uniformly at random (called **bin samples** or just **samples**) and allocates to one of these using a general decision function Q^t.

d-SAMPLE(Q^t) Process:
Parameters:
- d: The *number of bins* sampled independently and uniformly at random in each step.
- $Q^t : \mathcal{F}^t \times [n]^d \times [d] \to [0,1]$: The *decision function*, which gives the probability to allocate to one of the d samples.
  For any $\mathfrak{F}^t \in \mathcal{F}^t$ and any samples $j_1, \ldots, j_d \in [n]$, it must satisfy:
$$\sum_{i \in [d]} Q^t(\mathfrak{F}^t, j_1, \ldots j_d, i) = 1.$$

Iteration: At step $t \geqslant 0$, the allocation vector induced by the decision function is given by

$$q_i^t(\mathfrak{F}^t) = \frac{1}{n^d} \cdot \sum_{j_1,\ldots,j_d \in [n]} \sum_{k \in [d] : j_k = i} Q^t(\mathfrak{F}^t, j_1, \ldots, j_d, k).$$

For example, the TWO-CHOICE process is a TWO-SAMPLE(Q^t) process with decision function:

$$Q^t(x^t, j_1, j_2, i) = \begin{cases} \frac{1}{2} & \text{if } x_{j_1}^t = x_{j_2}^t, \\ \mathbf{1}_{x_{j_1}^t < x_{j_2}^t} & \text{else if } i = 1, \\ \mathbf{1}_{x_{j_1}^t > x_{j_2}^t} & \text{otherwise.} \end{cases} \tag{2.2}$$

Corollary 2.3. *The* TWO-CHOICE $=$ TWO-SAMPLE(Q^t) *process is* <u>*load-vector indistinguishable*</u> *from the* TIME-HOMOGENEOUS(p) *process with* $p_i = \frac{2i-1}{n^2}$ *for any* $i \in [n]$.

Proof. We define the $\widehat{\mathcal{Q}} =$ TWO-SAMPLE$(\widehat{Q}^t)$ process, with

$$\widehat{Q}^t(x^t, j_1, j_2, i) = \begin{cases} \mathbf{1}_{i=1} & \text{if } j_1 = j_2, \\ \mathbf{1}_{\mathrm{Rank}^t(j_1) > \mathrm{Rank}^t(j_2)} & \text{else if } x^t_{j_1} = x^t_{j_2}, \\ \mathbf{1}_{x^t_{j_1} < x^t_{j_2}} & \text{else if } i = 1, \\ \mathbf{1}_{x^t_{j_1} > x^t_{j_2}} & \text{otherwise.} \end{cases}$$

Compared to the update function Q^t of TWO-CHOICE defined in Eq. (2.2), this decision function allocates differently iff the two sampled bins have the same load. Hence, there is a load probability preserving transformation between q^t of TWO-CHOICE and $\widehat{q}^t$ of $\widehat{\mathcal{Q}}$. In order to allocate to bin with rank i, we need to sample that bin and any of the $i-1$ bins with higher rank, so

$$\widehat{q}^t_{s^t(i)}(\mathfrak{F}^t) = \frac{1}{n^2} \cdot \sum_{j_1 \in [n]} Q^t(\mathfrak{F}^t, j_1, s^t(i), 2) + \frac{1}{n^2} \cdot \sum_{j_2 \in [n]} Q^t(\mathfrak{F}^t, s^t(i), j_2, 1) = \frac{i-1}{n^2} + \frac{i}{n^2} = \frac{2i-1}{n^2}.$$

Both processes are index-independent, hence they are load-vector indistinguishable by Theorem 2.1. □

2.2.2 d-CHOICE process

We start by defining the d-CHOICE process informally for any $d \in \mathbb{N}_+$.

> d-CHOICE Process:
> Iteration: At step $t \geqslant 0$, sample d bins $j_1, \ldots, j_d \in [n]$ independently and uniformly at random. Allocate ball $t+1$ to a bin $j_{\min} \in \{j_1, \ldots, j_d\}$ satisfying $x^t_{j_{\min}} = \min_{k \in [d]} x^t_{j_k}$, breaking ties arbitrarily.

Now, we make this definition formal by defining it as a d-SAMPLE process. This slightly more complicated definition will allow us to easily define the process in different settings including weighted balls, heterogeneous sampling distributions, noisy and outdated information (see Section 2.4).

> d-CHOICE ($\subseteq$ d-SAMPLE(Q^t)) Process:
> Iteration: At step $t \geqslant 0$, the decision function to allocate ball $t+1$ is given by
>
> $$Q^t(\mathfrak{F}^t, j_1, \ldots, j_d, i) := Q(x^t_{j_1}, \ldots, x^t_{j_d}, i) := \frac{\mathbf{1}_{x^t_{j_i} = v_{\min}}}{\sum_{k \in [d]} \mathbf{1}_{x^t_{j_k} = v_{\min}}},$$
>
> where $v_{\min} := \min_{k \in [d]} x^t_{j_k}$.
>
> **Special cases:** ONE-CHOICE (for $d = 1$), TWO-CHOICE (for $d = 2$)
> [time homo. prob. vector $p_i = \left(\frac{i}{n}\right)^d - \left(\frac{i-1}{n}\right)^d$, index independent, d-SAMPLE]

We can also define d-CHOICE as the TIME-HOMOGENEOUS(p) with $p_i = \left(\frac{i}{n}\right)^d - \left(\frac{i-1}{n}\right)^d$, for any $i \in [n]$. As shown in Corollary 2.3, for $d = 2$ this process is load-vector indistinguishable from the TWO-SAMPLE definition. This definition has the benefit of defining d-CHOICE for any $d \in \mathbb{R}_{\geqslant 1}$ (not just for $d \in \mathbb{N}_+$). In [176], this process was called GREEDY$[1+\epsilon]$, where $\epsilon = d - 1$.

2.2.3 d-Sample-with-Memory-M processes

More generally, we define d-Sample processes that can store up to M bin indices in a cache. In each step, the process is allowed to allocate to any of the M cached bins in addition to the d sampled bins, using some decision function Q^t. The contents of the cache are $c_1^t, \ldots, c_M^t \in [n] \cup \{0\}$, where 0 denotes an empty cache and at the end of each step they are updated through the function U^t.

d-Sample-with-Memory-$M(Q^t, U^t)$ Process:
Parameters:

- d: The *number of bins sampled* in each step.
- M: The *cache size*, i.e., the number of stored bin indices that can be used across steps.
- $Q^t : \mathcal{F}^t \times ([n] \cup \{0\})^M \times [n]^d \times [M+d] \to [0,1]$: The *decision function*, which gives the probability to allocate to one of the caches or one of the samples.
  For any $\mathfrak{F}^t \in \mathcal{F}^t$, any sampled bins $j_1, \ldots, j_d \in [n]$ and any memory contents $c_1^t, \ldots, c_M^t \in [n] \cup \{0\}$, it must satisfy:
  $$\sum_{i \in [M+d]} Q^t(\mathfrak{F}^t, c_1^t, \ldots, c_M^t, j_1, \ldots j_d, i) = 1,$$
  and for any $i \in [M]$ with $c_i^t = 0$,
  $$Q^t(\mathfrak{F}^t, c_1^t, \ldots, c_M^t, j_1, \ldots j_d, i) = 0,$$
  meaning that we cannot allocate to an empty cache.
- $U^t : \mathbb{R}^n \times ([n] \cup \{0\})^M \times [n]^d \to ([n] \cup \{0\})^M$: The *update function*, which given the contents of the caches c^t and the sampled bins (after the allocation), decides which bins to keep in the cache. In particular, for any $i \in [M]$,
  $$(U^t(x^{t+1}, c_1^t, \ldots, c_M^t, j_1, \ldots, j_d))_i \subseteq \{c_1^t, \ldots, c_M^t, j_1, \ldots, j_d\}.$$

Initialise: The cache $c^0 = \mathbf{0} \in ([n] \cup \{0\})^M$, where 0 denotes an empty cache.
Iteration: At step $t \geqslant 0$,

- The allocation vector q^t induced by the decision function is given by
  $$\begin{aligned} q_i^t(\mathfrak{F}^t) = \frac{1}{n^d} \cdot & \sum_{j_1, \ldots, j_d \in [n]} \sum_{k \in [M] : c_k^t = i} Q^t(\mathfrak{F}^t, c_1^t, \ldots, c_M^t, j_1, \ldots, j_d, k) \\ & + \frac{1}{n^d} \cdot \sum_{j_1, \ldots, j_d \in [n]} \sum_{k \in [d] : j_k = i} Q^t(\mathfrak{F}^t, c_1^t, \ldots, c_M^t, j_1, \ldots, j_d, M+k). \end{aligned}$$
- The cache is updated using
  $$c^{t+1} := U^t(x^{t+1}, c_1^t, \ldots, c_M^t, j_1, \ldots, j_d).$$

A **load comparison** d-Sample-with-Memory-M process has a decision function which depends only on the loads of the cached and sampled bins, i.e.,

$$Q^t(\mathfrak{F}^t, c_1^t, \ldots, c_M^t, j_1, \ldots, j_d, i) := Q(x_{c_1^t}^t, \ldots, x_{c_M^t}^t, x_{j_1}^t, \ldots, x_{j_d}^t, i).$$

An **average-aware load comparison** d-Sample-with-Memory-M process has decision function which may also depend on the average load

$$Q^t(\mathfrak{F}^t, c_1^t, \ldots, c_M^t, j_1, \ldots, j_d, i) := Q(\overline{x}^t, x_{c_1^t}^t, \ldots, x_{c_M^t}^t, x_{j_1}^t, \ldots, x_{j_d}^t, i).$$

Maintaining an estimate for the average is a less strict requirement than knowing the entire load vector. Both these two definitions apply for d-SAMPLE processes (by setting $M = 0$).

2.2.4 MEMORY process

The following process was introduced in [136]. It allocates to the least loaded of the sampled and cached bins, and then updates the cache to store the M least loaded of these.

(d, M)-MEMORY Process ($\subseteq$ d-SAMPLE-WITH-MEMORY-$M$$(Q^t, U^t)$):
Parameters:

- d: The *number of bins sampled* in each step.
- M: The *cache size*.

Iteration: At step $t \geqslant 0$, the decision function Q^t is given by

$$Q(x_{c_1^t}^t, \ldots, x_{c_M^t}^t, x_{j_1}^t, \ldots, x_{j_d}^t, i) = \frac{1}{\sum_{k \in [M]} \mathbf{1}_{x_{c_k^t}^t = v_{\min}} + \sum_{k \in [d]} \mathbf{1}_{x_{j_k}^t = v_{\min}}} \cdot \begin{cases} \mathbf{1}_{x_{c_k^t}^t = v_{\min}} & \text{if } i \leqslant M, \\ \mathbf{1}_{x_{j_{i-M}}^t = v_{\min}} & \text{otherwise,} \end{cases}$$

where $v_{\min} := \min\left\{x_{c_1^t}^t, \ldots, x_{c_M^t}^t, x_{j_1}^t, \ldots, x_{j_d}^t\right\}$ (with $x_0^t = \infty$ for convenience) is the minimum load out of the samples and cached bins at step t.

The update function, stores the M smallest of the cached and sampled bins, breaking ties by comparing bin indices,

$$U^t(x^{t+1}, c_1^t, \ldots, c_M^t, j_1, \ldots, j_d) = \arg \mathrm{small}^M_{k \in \{c_1^t, \ldots, c_M^t, j_1, \ldots, j_d\}} x_k^{t+1}.$$

[load comparison d-SAMPLE-WITH-MEMORY-M]

A load-vector indistinguishable version of the special case with $d = M = 1$ is given explicitly below.

MEMORY Process:
Iteration: At step $t \geqslant 0$, sample a uniform bin i, and update its load (or of cached bin c_1^t):

$$\begin{cases} x_i^{t+1} = x_i^t + 1 & \text{if } x_i^t < x_{c_1^t}^t \quad \text{(also update cache } c_1^t = i\text{)}, \\ x_i^{t+1} = x_i^t + 1 & \text{if } x_i^t = x_{c_1^t}^t, \\ x_{c_1^t}^{t+1} = x_{c_1^t}^t + 1 & \text{if } x_i^t > x_{c_1^t}^t. \end{cases}$$

We introduce the following variant of the (d, M)-MEMORY process, where the cache resets every r steps.

(d, M, r)-RESET-MEMORY Process ($\subseteq$ d-SAMPLE-WITH-MEMORY-$M$$(Q^t, U^t)$):
Parameters:

- d: The *number of bins sampled* in each step.
- M: The *cache size*.

- r: The *reset period*.

Iteration: At step $t \geqslant 0$, the decision function Q^t is given as in (d, M)-MEMORY and the update function U^t is given by

$$U^t = \begin{cases} \mathbf{0}^M & \text{if } r \mid t, \\ \arg\text{small}^M_{k \in \{c_1^t, \ldots, c_M^t, j_1, \ldots, j_d\}} x_k^{t+1} & \text{otherwise.} \end{cases}$$

The special case of $(1, 1, 2)$-RESET-MEMORY can be seen as a sample efficient $(1+\beta)$-process with $\beta = 1/2$, in the sense that it makes exactly one sample per allocation instead of $1 + 1/2$ samples (in expectation), while still maintaining the same asymptotic gap bound of $\mathcal{O}(\log n)$ (Theorem 7.49).

2.3 More notation and concepts

2.3.1 Composition of processes

In [132] (and also [152]), the $(1+\beta)$-process was defined as the process where in step t, with probability $\beta \in [0, 1]$ we do TWO-CHOICE and with $1 - \beta$ we do ONE-CHOICE.

$(1+\beta)$-Process ($= \beta$-MIXED(TWO-CHOICE, ONE-CHOICE)):

Parameter: A probability $\beta \in [0, 1]$.

Iteration: At step $t \geqslant 0$, with probability β allocate via the TWO-CHOICE process, otherwise allocate via the ONE-CHOICE process.

[time homo. prob. vector $p_i = \beta \cdot \frac{2i-1}{n} + (1-\beta) \cdot \frac{1}{n}$, TWO-SAMPLE]

Here, we generalise this way of mixing processes. More concretely for any two processes $\mathcal{P} =$ SEQUENTIAL(q^t) and $\mathcal{Q} =$ SEQUENTIAL$(\widetilde{q}^t)$, we define the process β-MIXED$(\mathcal{P}, \mathcal{Q})$ as the SEQUENTIAL(r^t) process with

$$r^t(\mathfrak{F}^t) := \beta \cdot q^t(\mathfrak{F}^t) + (1-\beta) \cdot \widetilde{q}^t(\mathfrak{F}^t).$$

In particular, when $\mathcal{P}$ and $\mathcal{Q}$ have time-homogeneous allocation vectors p and $\widehat{p}$, then their composition β-MIXED$(\mathcal{P}, \mathcal{Q})$ is the TIME-HOMOGENEOUS(r) with $r = \beta \cdot p + (1-\beta) \cdot \widehat{p}$.

More generally, for N processes $\mathcal{P}_1 =$ SEQUENTIAL$(q_1^t), \ldots, \mathcal{P}_N =$ SEQUENTIAL(q_N^t) and a probability vector $(\beta_1^t, \ldots, \beta_N^t)$, we define the $(\beta_1^t, \ldots, \beta_{N-1}^t)$-MIXED$(\mathcal{P}_1, \ldots, \mathcal{P}_N)$ process as the SEQUENTIAL(r^t) process with

$$r^t(\mathfrak{F}^t) := \beta_1^t \cdot q_1^t(\mathfrak{F}^t) + \ldots + \beta_N^t \cdot q_N^t(\mathfrak{F}^t).$$

2.3.2 Folding of a process

In some cases, it makes sense to look at a sequence of consecutive ball allocations as a single "**round**" (see for example, the b-BATCHED setting or the PACKING process). Starting at step t, we define a function $f : \mathcal{F} \times \mathbb{N} \times \mathbb{N} \to \{0, 1\}$ where $f(\mathfrak{F}^t, s_0, t)$ for $t \geqslant s_0$ determines whether the current round which started at s_0 should end at step t. This way, all steps are assigned to a particular round. The function $T_{s_0} : \mathbb{N} \to \mathbb{N}$ converts steps to rounds for any step $t \geqslant s_0$.

For convenience, we use the notation $x^{r,1}, x^{r,2}, \ldots$ to denote the load vector at substeps $1, 2, \ldots$ of round r. The load vector $x^{r,0}$ corresponds to that at the end of round $r - 1$.

2.3.3 Majorisation

We say that a vector $u \in \mathbb{R}^n$ **majorises** another vector $v \in \mathbb{R}^n$, and denote it by $u \succeq v$ iff for all $k \in [n]$,

$$\sum_{i=1}^{k} u_i \geqslant \sum_{i=1}^{k} v_i.$$

We recall that a function $f : \mathbb{R}^n \to \mathbb{R}$ is called **Schur convex** [126] if for any $u, v \in \mathbb{R}^n$,

$$u \succeq v \Rightarrow f(u) \geqslant f(v).$$

In particular, any convex function is Schur-convex, so when $\tilde{x}_{\mathcal{P}}^t \succeq \tilde{x}_{\mathcal{Q}}^t$ it also implies that:

1. The maximum load of $\mathcal{P}$ is at least that of $\mathcal{Q}$.
2. The minimum load of $\mathcal{P}$ is at most that of $\mathcal{Q}$.
3. Convex potential functions, such as the quadratic, the exponential and the hyperbolic cosine potential are at least as large for $\mathcal{P}$ than for $\mathcal{Q}$.

The following result was shown in [152, Theorem 3.1].

Theorem 2.4. *Consider any $\mathcal{P} =$ TIME-HOMOGENEOUS(p_1) and $\mathcal{Q} =$ TIME-HOMOGENEOUS(p_2) processes, such that $p_1 \succeq p_2$. Then, there is a coupling between the two processes such that for any step $t \geqslant 0$,*

$$\tilde{x}_{\mathcal{P}}^t \succeq \tilde{x}_{\mathcal{Q}}^t.$$

We will prove a slightly stronger version of this theorem where majorisation holds for the sorted allocation vectors at any step $t \geqslant 0$, but they need not be time-homogeneous.

Theorem 2.5. *Consider any two processes $\mathcal{P} =$ SEQUENTIAL(q_1^t) and $\mathcal{Q} =$ SEQUENTIAL(q_2^t) such that for any step $t \geqslant 0$ and any filtrations $\mathfrak{F}_{\mathcal{P}}^t$ and $\mathfrak{F}_{\mathcal{Q}}^t$,*

$$\tilde{q}_1^t(\mathfrak{F}_{\mathcal{P}}^t) \succeq \tilde{q}_2^t(\mathfrak{F}_{\mathcal{Q}}^t).$$

Then, there is a coupling such that $\tilde{x}_{\mathcal{P}}^t \succeq \tilde{x}_{\mathcal{Q}}^t$ for any step $t \geqslant 0$.

Usually, when we apply this theorem at least one of the two processes is TIME-HOMOGENEOUS. In order to prove this theorem, we will make use of the following lemma.

Lemma 2.6. *For any process $\mathcal{P} =$ SEQUENTIAL(q^t), at any step $t \geqslant 0$,*

$$\left| i \in [n] \colon \tilde{x}_i^{t+1} \neq \tilde{x}_i^t \right| = 1.$$

Note that this statement does not necessarily hold in the case of non-unit weights (see Fig. 2.3).

Proof. Let $i = i^{t+1}$ be the allocated bin. Then, the sorted vector $\tilde{x}^{t+1}$ is obtained from $\tilde{x}^t$ by incrementing the load of the bin with rank j, where j is smallest index such that $\tilde{x}_j^t = \tilde{x}_i^t$ (see Fig. 2.2). $\square$

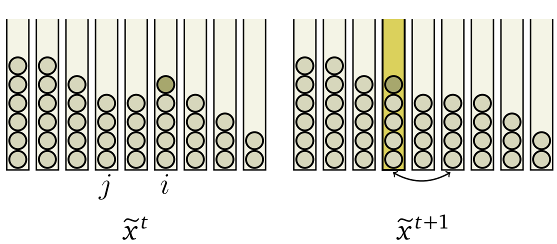


Figure 2.2: Illustration of one bin difference between $\tilde{x}^t$ and $\tilde{x}^{t+1}$ for the unit weights case.

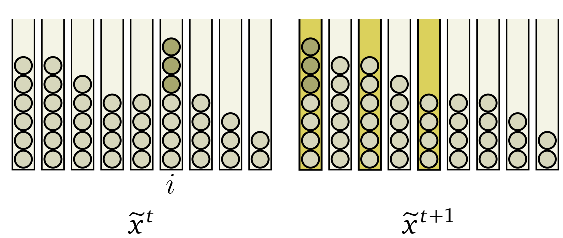


Figure 2.3: Illustration that Lemma 2.6 does not necessarily hold for non-unit weights.

Proof of Theorem 2.5. We will define a coupling between the processes $\mathcal{P}$ and $\mathcal{Q}$, so that the sorted load vectors of the two processes are the same, i.e., $\widetilde{x}_{\mathcal{P}}^t = \widetilde{x}_{\mathcal{Q}}^t$ for each step t. At step $t = 0$, this clearly holds as both load vectors are empty, i.e., $\widetilde{x}_{\mathcal{P}}^0 = \widetilde{x}_{\mathcal{Q}}^0 = \mathbf{0}$.

Consider an arbitrary step $t \geqslant 0$ and assume that $\widetilde{x}_{\mathcal{P}}^t = \widetilde{x}_{\mathcal{Q}}^t$. We define the coupling between the two processes, so that ball $t+1$ is allocated as follows:

1. Sample $u \in_{\mathcal{U}} [0,1]$.
2. For process $\mathcal{P}$, allocate to $s_{\mathcal{P}}^t(i_1)$, where $i_1 \in [n]$ is the smallest rank such that
$$\sum_{i=1}^{i_1} (\widetilde{q}_1^t)_i \geqslant u.$$
Similarly, for process $\mathcal{Q}$ allocate to $s_{\mathcal{Q}}^t(i_2)$, where $i_2 \in [n]$ is the smallest rank such that
$$\sum_{i=1}^{i_2} (\widetilde{q}_2^t)_i \geqslant u.$$

The probability to allocate to the i-th most loaded bin in $\mathcal{P}$ is $(\widetilde{q}_1^t)_i$ and to the i-th most loaded bin in $\mathcal{Q}$ is $(\widetilde{q}_2^t)_i$. Hence, the coupling is valid. Also, because $\widetilde{q}_1^t \succeq \widetilde{q}_2^t$, it follows that $i_1 \leqslant i_2$. We will now show that for $\widetilde{x}_{\mathcal{P}}^{t+1} := \widetilde{x}_{\mathcal{P}}^t + \mathbf{e}_{i_1}$ and $\widetilde{x}_{\mathcal{Q}}^{t+1} := \widetilde{x}_{\mathcal{Q}}^t + \mathbf{e}_{i_2}$, we have $\widetilde{x}_{\mathcal{P}}^{t+1} \succeq \widetilde{x}_{\mathcal{Q}}^{t+1}$.

By Lemma 2.6, only one bin changes between $\widetilde{x}^t$ and $\widetilde{x}^{t+1}$. Let j_1 be that bin for process $\mathcal{P}$ and j_2 for process $\mathcal{Q}$.

Case 1 [$j_1 \leqslant j_2$]: The prefix sums that change (by $+1$) for $\mathcal{P}$ are a superset of those in $\mathcal{Q}$, so $\widetilde{x}_{\mathcal{P}}^{t+1} \succeq \widetilde{x}_{\mathcal{Q}}^{t+1}$.

Case 2 [$j_1 > j_2$]: Now, we need to verify that majorisation still holds for all indices in $[j_2, j_1]$. Assume that this is not the case. Then, there exists $k \in [j_2, j_1]$ such that
$$\sum_{i=1}^{k} (\widetilde{x}_{\mathcal{P}}^t)_i = \sum_{i=1}^{k} (\widetilde{x}_{\mathcal{Q}}^t)_i,$$
and $(\widetilde{x}_{\mathcal{P}}^t)_k \leqslant (\widetilde{x}_{\mathcal{Q}}^t)_k$. Also, by Lemma 2.6, we have that for $\ell \in (j_2, i_2]$, $(x_{\mathcal{Q}}^t)_\ell = z$ for some fixed value z. But then we must have $(\widetilde{x}_{\mathcal{P}}^t)_{k+1} = (\widetilde{x}_{\mathcal{Q}}^t)_{k+1} = z$, as otherwise, $(\widetilde{x}_{\mathcal{P}}^t)_{k+1} < (\widetilde{x}_{\mathcal{Q}}^t)_{k+1}$ and consequently
$$\sum_{i=1}^{k+1} (\widetilde{x}_{\mathcal{P}}^t)_i = \sum_{i=1}^{k} (\widetilde{x}_{\mathcal{P}}^t)_i + (\widetilde{x}_{\mathcal{P}}^t)_{k+1} = \sum_{i=1}^{k} (\widetilde{x}_{\mathcal{Q}}^t)_i + (\widetilde{x}_{\mathcal{P}}^t)_{k+1} < \sum_{i=1}^{k+1} (\widetilde{x}_{\mathcal{Q}}^t)_i,$$
which would be a contradiction. Hence, $(\widetilde{x}_{\mathcal{P}}^t)_\ell = z$ for all $\ell \in [k, i_1]$, so $j_1 \leqslant k$ and so the prefix sums match. □

The following corollary states that Two-Choice is optimal over all Two-Sample processes in the unit weight case. As we shall show in Section 7.5, this is not true e.g., in the b-Batched setting.

Corollary 2.7. *Any* Two-Sample(Q^t) *process majorises the* Two-Choice *process.*

In general, any d-Sample process majorises the d-Choice, but for simplicity we look at the $d = 2$ case.

Proof. Consider an arbitrary Two-Sample(Q^t) process and let $q_{\mathcal{P}}^t$ be its allocation vector. Also, let $\widetilde{Q}$ be the decision function of the Two-Choice process and $q_{\mathcal{Q}}^t$ its allocation vector.

We define the probability allocation vectors $p_0 =: \widetilde{q}_{\mathcal{P}}^t, p_1, \ldots, p_{n^2}$, where p_i is derived from p_{i-1} by changing the decision for a pair of bins (j_1, j_2) from $Q(\mathfrak{F}^t, j_1, j_2, j)$ to $\widetilde{Q}(\mathfrak{F}^t, j_1, j_2, j)$ (for $j = 1$ and $j = 2$).

In the probability allocation vectors, this corresponds to (possibly) "reallocating" up to $1/n^2$ probability from heavier to lighter bins, so $p_0 \succeq p_1 \succeq \ldots \succeq p_{n^2} = \widetilde{q}_{\mathcal{Q}}$ and hence by transitivity of majorisation $\widetilde{q}^t \succeq \widehat{q}^t$. □

2.3.4 Conditions on probability vectors

The uniform probability vector $p = (\frac{1}{n}, \ldots, \frac{1}{n})$ corresponds to the ONE-CHOICE process, which has w.h.p. $\mathrm{Gap}(m) = \Omega(\sqrt{\frac{m}{n} \log n})$ for large m. In [152], the authors show that if a process has a bias of ϵ/n to place away from the $n/4$ heavier bins and towards the $n/4$ lighter bins, then the process has an upper bound on the gap that does not depend on m. Here, we generalise this condition and in Chapter 3 we will provide a tighter analysis for this family of processes.

A probability vector p is said to be **(ϵ, δ)-biased** for $\epsilon := \epsilon(n) \in (0,1)$ and quantile $\delta := \delta(n) \in \{1/n, \ldots, 1\}$, if p is non-decreasing and

$$p_{\delta n} \leqslant \frac{1-\epsilon}{n}.$$

For $\delta = 1/4$, this condition implies the condition in [152].

Observation 2.8. *For any (ϵ, δ)-biased probability vector $p \in \mathbb{R}^n$, we have that*

$$p_i \leqslant \frac{1-\epsilon}{n}, \text{ for any } i \leqslant \delta n, \quad \text{and} \quad p_i \geqslant \frac{1+\widetilde{\epsilon}}{n}, \text{ for any } i \geqslant \delta n + 1,$$

where $\widetilde{\epsilon} = \frac{\delta}{1-\delta} \cdot \epsilon$.

The following observation gives the worst-case (ϵ, δ)-biased probability vector and this simplifies the analysis in Chapter 3.

Observation 2.9. *For any (ϵ, δ)-biased probability vector $p \in \mathbb{R}^n$, we have that for probability vector $q \in \mathbb{R}^n$, whose entries are given by*

$$q_i = \begin{cases} \frac{1-\epsilon}{n} & \text{if } i \leqslant \delta n, \\ \frac{1+\widetilde{\epsilon}}{n} & \text{otherwise,} \end{cases}$$

q majorises p, i.e., $q \succeq p$.

We now generalise the (ϵ, δ)-biased condition to prefix sums. This will allow us to apply our analysis to the GRAPHICAL setting (Section 2.4.7 and Lemma 7.40).

- **Condition $\mathcal{C}_1$:** There exist constant[1] $\delta \in (0,1)$ and (not necessarily constant) $\epsilon \in (0,1)$, such that for any $1 \leqslant k \leqslant \delta n$,

$$\sum_{i=1}^{k} p_i \leqslant (1-\epsilon) \cdot \frac{k}{n},$$

  and, similarly, for any $\delta n + 1 \leqslant k \leqslant n$,

$$\sum_{i=k}^{n} p_i \geqslant \left(1 + \epsilon \cdot \frac{\delta}{1-\delta}\right) \cdot \frac{n-k+1}{n}.$$

- **Condition $\mathcal{C}_2$:** There exists $C > 1$, such that $\max_{i \in [n]} p_i \leqslant \frac{C}{n}$.
- **Condition $\mathcal{C}_3$:** There exists $C > 1$, such that $\max_{i \in [n]} \left| p_i - \frac{1}{n} \right| \leqslant \frac{C-1}{n}$.

Observation 2.10. *Any (ϵ, δ)-biased probability vector satisfies condition $\mathcal{C}_1$ with the same δ and ϵ.*

[1] Here constant means that the quantile $\delta \in (\delta_1, \delta_2)$ with constant $\delta_1, \delta_2 \in (0,1)$.

Proof. Since $p_{\delta n} \leqslant \frac{1-\epsilon}{n}$ and p is non-decreasing, it follows that $p_i \leqslant \frac{1-\epsilon}{n}$ for all $1 \leqslant i \leqslant \delta n$, and thus the prefix sum condition of $\mathcal{C}_1$ holds, i.e., for any $1 \leqslant k \leqslant \delta n$,

$$\sum_{i=1}^{k} p_i \leqslant \frac{1-\epsilon}{n} \cdot k = (1-\epsilon) \cdot \frac{k}{n}.$$

By Observation 2.8, it holds that $p_i \geqslant \frac{1+\widetilde{\epsilon}}{n}$ for any $i \geqslant \delta n + 1$. So, we obtain the suffix sum condition of $\mathcal{C}_1$, i.e., for any $\delta n + 1 \leqslant k \leqslant n$,

$$\sum_{i=k}^{n} p_i \geqslant \frac{1+\widetilde{\epsilon}}{n} \cdot (n-k+1) = (1+\widetilde{\epsilon}) \cdot \frac{n-k+1}{n}. \qquad \square$$

Using this observation, it is easy to verify that Two-Choice, the $(1+\beta)$-process and Quantile(δ) satisfy the two conditions $\mathcal{C}_1$ and $\mathcal{C}_2$.

Proposition 2.11. *For any $\beta \in (0,1]$, the $(1+\beta)$-process satisfies condition $\mathcal{C}_1$ with $\delta = \frac{1}{4}$ and $\epsilon = \frac{\beta}{2}$ and condition $\mathcal{C}_2$ with $C = 2$. Further, for any constant $\delta \in (0,1)$, the Quantile(δ) process satisfies condition $\mathcal{C}_1$ with δ and $\epsilon = 1-\delta$, and condition $\mathcal{C}_2$ with $C = 2$.*

2.3.5 Naming and notational conventions

Variable naming conventions:

- Bins are indexed by variables $i, j, k \in [n]$.
- Time and ball indices are t, s, u. When we are aiming to prove a guarantee for a specific fixed time, this will be denoted by m, i.e., after the m-th ball has been allocated.
- Processes without a name are denoted by $\mathcal{P}, \mathcal{Q}, \mathcal{R}$.

Notation conventions:

- When it is not clear from the context which process $\mathcal{P}$ we analyse, we will write $x_{\mathcal{P}}^t$ for the load vector (and $(x_{\mathcal{P}}^t)_i$ for the i-th entry) and $\text{Gap}_{\mathcal{P}}(t)$ for the gap.

2.4 Settings

Recall that a **setting** is a function that transforms a process into another process. Table 2.4 gives an overview of the settings, along with the set of processes to which they can be applied. In the coming chapters, whenever we do not specify the process $\mathcal{P}$, it should be assumed that we are working with $\mathcal{P} = \textsc{Two-Choice}$.

Setting	Input process constraints	Brief description
$\textsc{Weighted}(\mathcal{P}, \mathcal{W}^t)$	$\mathcal{P} = \textsc{Sequential}(q^t)$	Balls have weights following $\mathcal{W}^t$.
$\textsc{Heterogeneous}(\mathcal{P}, \mathcal{S}^t)$	$\mathcal{P} = d\textsc{-Sample-with-Memory-}M(Q^t, U^t)$	Bins are sampled using $\mathcal{S}^t$.
$b\textsc{-Batched}(\mathcal{P})$	$\mathcal{P} = \textsc{Sequential}(q^t)$	Balls are allocated in batches of size b.
$\tau\textsc{-Delay}(\mathcal{P}, G^t)$	Markovian $\mathcal{P} = \textsc{Sequential}(q^t)$	Bin loads can be outdated by at most τ steps.
$g\textsc{-Adv-Load}(\mathcal{P}, G^t)$	Markovian $\mathcal{P} = \textsc{Sequential}(q^t)$	Each bin load can be perturbed by an additive amount g by an adversary G^t.
$g\textsc{-Adv-Comp}(\mathcal{P}, G^t)$	$\mathcal{P} = \textsc{Two-Sample}(Q^t)$	Decisions for bins with load difference at most g can be altered by an adversary G^t.
$\rho\textsc{-Noisy-Load}(\mathcal{P})$	Markovian $\mathcal{P} = \textsc{Sequential}(q^t)$	Bin loads are perturbed by random amount sampled from ρ^t.
$\rho\textsc{-Noisy-Comp}(\mathcal{P})$	$\mathcal{P} = \textsc{Two-Sample}(Q^t)$	$\rho^t(\delta)$ gives the probability of a correct comparison between two sampled bins with load difference of δ.
$\textsc{Graphical}(\mathcal{P}, G)$	$\mathcal{P} = \textsc{Two-Sample}(Q^t)$	The two sampled bins are adjacent vertices of an edge sampled uniformly at random from graph $G = ([n], E)$.

Table 2.4: Overview of the different settings considered.

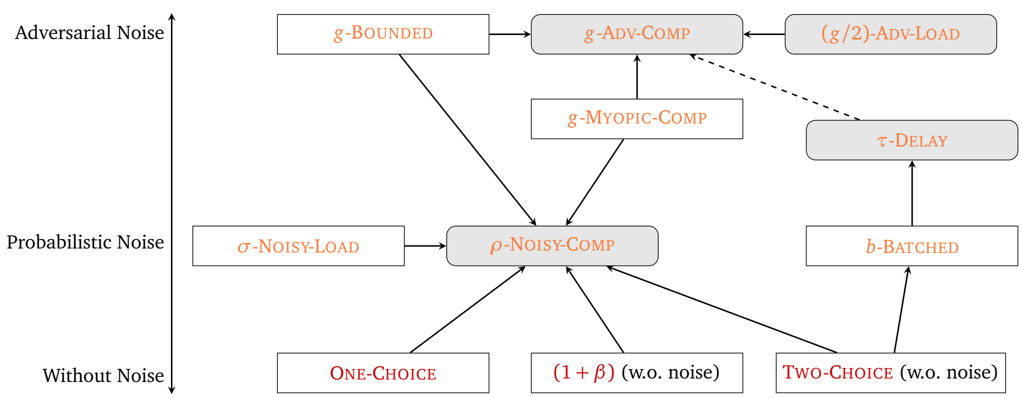


Figure 2.5: Overview of settings (rounded rectangles) and processes (rectangles) for $\textsc{Two-Choice}$. A directed arrow from setting (process) A to setting (process) B means that B is stronger than A (that is, for each process in A, there is a load-vector indistinguishable process in B). For $\tau\textsc{-Delay}$, a dashed arrow is used for the connection to $g\textsc{-Adv-Comp}$, as the relation is slightly more involved (Section 7.4.2).

2.4.1 Weighted setting

We define the weighted setting for any $\textsc{Sequential}(q^t)$ process.

WEIGHTED($\mathcal{P}, \mathcal{W}^t$) Setting:
Parameters:
- $\mathcal{P}$: A sequential allocation process SEQUENTIAL(q^t).
- $\mathcal{W}^{t+1} := \mathcal{W}^{t+1}(\mathfrak{F}^t, i^{t+1})$: The *weight distribution*.

Initialise: The load vector $x^0 = \mathbf{0} \in \mathbb{R}^n$.
Iteration: At step $t \geqslant 0$, to allocate ball $t+1$
1. Determine the probability allocation vector $q^t = q^t(\mathfrak{F}^t)$.
2. Sample the bin $i^{t+1} \in_{q^t} [n]$ for allocation.
3. Determine the distribution for the weights $\mathcal{W}^{t+1} = \mathcal{W}^{t+1}(\mathfrak{F}^t, i^{t+1})$.
4. Sample the weight $w^{t+1} \in \mathcal{W}^{t+1}$.
5. Set $x^{t+1} := x^t + w^{t+1} \cdot \mathbf{e}_{i^{t+1}}$.

The **total weight** of the allocated balls until step t is defined as

$$W^t := \sum_{s=1}^{t} w^s.$$

So, the average $\overline{x}^t = W^t/n$. We also give names to the following special cases:

- In the **unweighted case** (or **unit weight case**), $w^t = 1$ and so $W^t = t$.
- The **independent weights case**, where the weights are independent, i.e.,

  $$\mathcal{W}^{t+1}(\mathfrak{F}^t) = \mathcal{W},$$

  for some fixed distribution $\mathcal{W} := \mathcal{W}(n)$.
- The **load-dependent weights case**, where the weights depend on the bin being allocated, i.e.,

  $$\mathcal{W}^{t+1}(\mathfrak{F}^t) = \mathcal{W}^{t+1}(t, x^t_{i^{t+1}}).$$

  For example, the TWINNING process samples a bin i and allocates two balls to bins with $x_i^t < \frac{t}{n}$ and one ball otherwise. Also, PACKING allocates one ball to overloaded bins and $\lceil y_i^t + 1 \rceil$ balls to underloaded bins (see Section 2.5.5).

Weights with finite moments FINITE-MGF

In the FINITE-MGF(ζ) setting, the weight of each ball is drawn independently from a fixed distribution $\mathcal{W}$ over $[0, \infty)$. Following [152], we assume that the distribution $\mathcal{W}$ satisfies:

- $\mathbf{E}[\, \mathcal{W} \,] = 1$.
- $\mathbf{E}[\, e^{\zeta \mathcal{W}} \,] < \infty$ for some $\zeta > 0$.

Specific examples of distributions satisfying the above conditions (after scaling) are the geometric, exponential, binomial and Poisson distributions.

Similar to the arguments in [152], the above two assumptions can be used to prove that:

Lemma D.4 (Restated, page 227). *There exists $S := S(\zeta) \geqslant \max\{1, 1/\zeta\}$, such that for any $\gamma \in (0, \min\{\zeta/2, 1\})$ and any $\kappa \in [-1, 1]$,*

$$\mathbf{E}\Big[\, e^{\gamma \cdot \kappa \cdot \mathcal{W}} \,\Big] \leqslant 1 + \gamma \cdot \kappa + S\gamma^2 \cdot \kappa^2.$$

As this parameter is used in most of the upper bounds involving the FINITE-MGF(ζ) weights, we often refer to the setting as FINITE-MGF(S) or FINITE-MGF(ζ, S).

2.4.2 Heterogeneous sampling distributions

In Section 2.2, we defined the d-Sample-with-Memory-M to implicitly have the uniform sampling vector $\mathcal{S} = \left(\frac{1}{n}, \ldots, \frac{1}{n}\right)$ meaning that each d-tuple of indices had exactly the same probability $1/n^d$ of being sampled.

In [176], the setting of heterogeneous sampling distributions $\mathcal{S}$ was introduced for the d-Choice process. We extend this to any d-Sample-with-Memory-M process.

Heterogeneous($\mathcal{P}, \mathcal{S}^t$) Setting:

Parameters:

- $\mathcal{P}$: A d-Sample-with-Memory-M(Q^t, U^t) process.
- $\mathcal{S}^t$: A probability vector for sampling the bins.

Iteration: At step $t \geqslant 0$, allocate ball $t+1$ using the allocation vector

$$q_i^t(\mathfrak{F}^t) = \sum_{j_1,\ldots,j_d \in [n]} \sum_{k \in [M] : c_k^t = i} Q^t(\mathfrak{F}^t, c_1^t, \ldots, c_M^t, j_1, \ldots, j_d, k) \cdot \prod_{\ell=1}^{d} \mathcal{S}_{j_\ell}^t$$
$$+ \sum_{j_1,\ldots,j_d \in [n]} \sum_{k \in [d] : j_k = i} Q^t(\mathfrak{F}^t, c_1^t, \ldots, c_M^t, j_1, \ldots, j_d, M+k) \cdot \prod_{\ell=1}^{d} \mathcal{S}_{j_\ell}^t.$$

And update the cache using the function U^t.

The only difference with the definition of the d-Sample-with-Memory-M process is that the $1/n^d$ term is replaced by $\prod_{\ell=1}^{d} \mathcal{S}_{j_\ell}^t$, meaning that the probability to sample a bin may not be uniform.

2.4.3 b-Batched setting

In [28], a parallel setting was considered for Two-Choice, where balls are allocated in batches of size n. Here, we generalise this to arbitrary batch size $b \in \mathbb{N}_+$ and for any process $\mathcal{P}$.

b-Batched($\mathcal{P}$) Setting:

Parameters:

- $b \in \mathbb{N}_+$: The *batch size*, i.e., the number of balls allocated in the same batch.
- $\mathcal{P}$: A Sequential($\widehat{q}^t$) process.

Iteration: At step $t \geqslant 0$, allocate ball $t+1$ using the allocation vector

$$q^t := \widehat{q}^{t-\text{rem}(t,b)}(\mathfrak{F}^{t-\text{rem}(t,b)}),$$

where $t - \text{rem}(t,b)$ is the starting step of the batch containing step t.

Remark 2.12. *In the* b-Batched *setting for* $b > 1$*, it could be that two load-vector indistinguishable processes* $\mathcal{P}$ *and* $\mathcal{Q}$ *are no longer load-vector indistinguishable. For instance, in the first batch the* Two-Choice *process with random tie-breaks has allocation vector*

$$q_1 = \left(\frac{1}{n}, \ \ldots, \ \frac{1}{n}\right),$$

while with the TIME-HOMOGENEOUS *definition, the process has a allocation vector*

$$q_2 = \left(\frac{1}{n^2}, \ \ldots, \ \frac{2i-1}{n^2}, \ \ldots, \ \frac{2}{n} - \frac{1}{n^2} \right).$$

Hence, with q_1 all bins are allocated b/n balls in expectation, while with q_2 there are bins that are allocated approximately $2b/n$ balls in expectation.

2.4.4 τ-DELAY setting

A shortcoming of batching is that it requires the batches to be synchronised, i.e., all bin loads need to be updated at the same step. We relax this for any Markovian process $\mathcal{P}$ with allocation vector $\widehat{q}^t = \widehat{q}^t(x^t)$. We define the τ-DELAY$(\mathcal{P}, G^t)$ setting where in each step t, an adversary G^t transforms the load vector x^t into $\widehat{x}^t$ under the constraint that $\widehat{x}_i^t \in [x_i^{t-\tau+1}, x_i^t]$, for each $i \in [n]$. More formally:

τ-DELAY$(\mathcal{P}, G^t)$ Setting:

Parameters:

- $\mathcal{P}$: A Markovian SEQUENTIAL$(\widehat{q}^t)$ process with $\widehat{q}^t := \widehat{q}^t(x^t)$.
- $G^t : \mathcal{F}^t \to \mathbb{R}^n$: A function $\widehat{x}^t = G^t(\mathfrak{F}^t)$, such that for any filtration $\mathfrak{F}^t$ (corresponding to load vector x^t),

$$\widehat{x}_i^t \in [x_i^{t-\tau+1}, x_i^t], \qquad \text{for all } i \in [n].$$

Iteration: At step $t \geqslant 0$, allocate ball $t+1$ using the allocation vector

$$q^t(\mathfrak{F}^t) := \widehat{q}^t(G^t(\mathfrak{F}^t)).$$

Note that the process τ-DELAY$(\mathcal{P}, G^t)$ is no longer Markovian, as an allocation could depend on the last τ steps.

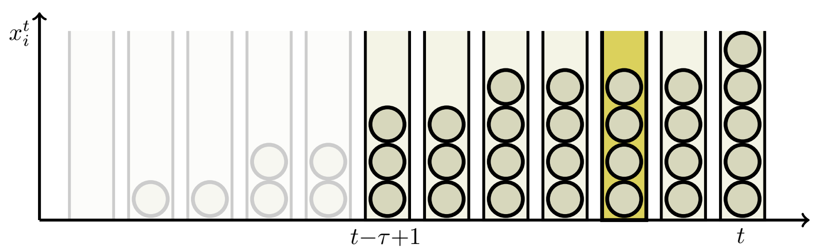


Figure 2.6: The evolution of the load of a fixed bin $i \in [n]$. In τ-DELAY, the adversary can choose to report any load of the last τ steps (shown highlighted), when allocating ball $t+1$.

2.4.5 g-ADV setting

For any Markovian process $\mathcal{P} = $ SEQUENTIAL$(\widehat{q}^t)$, we define the g-ADV-LOAD$(\mathcal{P}, G^t)$ process, where the adversary G^t can modify each of the loads by at most an additive value g. Formally, for any step $t \geqslant 0$, the resulting process has an allocation vector

$$q^t = \widehat{q}^t(\widehat{x}^t),$$

where $\widehat{x}^t = G(\mathfrak{F}^t)$ is the perturbed load vector satisfying $\widehat{x}_i^t \in [x_i^t - g, x_i^t + g]$ for all $i \in [n]$.

g-ADV-LOAD$(\mathcal{P}, G^t)$ Setting:
Parameters:

- $\mathcal{P}$: A Markovian SEQUENTIAL$(\widehat{q}^t)$ process with $\widehat{q}^t := \widehat{q}^t(x^t)$.
- $G^t : \mathcal{F}^t \to \mathbb{R}^n$: A function $\widehat{x}^t = G^t(\mathfrak{F}^t)$ satisfying for any load vector $x^t \in \mathbb{R}^n$ that

$$\widehat{x}_i^t \in [x_i^t - g, x_i^t + g], \quad \text{for all } i \in [n].$$

Iteration: At step $t \geqslant 0$, allocate ball $t+1$ using the allocation vector

$$q^t := q^t(\widehat{x}^t) := \widehat{q}^t(G^t(\mathfrak{F}^t)).$$

For any average-aware load comparison TWO-SAMPLE process $\mathcal{P}$, we also define the process g-ADV-COMP$(\mathcal{P}, G^t)$, where the adversary can influence the comparison between pairs of bins (j_1, j_2) with load difference $\left|x_{j_1}^t - x_{j_2}^t\right| \leqslant g$. Formally,

g-ADV-COMP$(\mathcal{P}, G^t)$ Setting ($\subseteq$ TWO-SAMPLE(Q^t)):
Parameters:

- $\mathcal{P}$: An average-aware load comparison TWO-SAMPLE$(\widehat{Q}^t)$ process with $\widehat{Q}^t := \widehat{Q}^t(\overline{x}^t, x_{j_1}^t, x_{j_2}^t, i)$.
- $G^t : \mathcal{F}^t \times [n]^2 \times [2] \to [0,1]$: A function $G^t(\mathfrak{F}^t, j_1, j_2, i)$ giving the decisions of the adversary.

Iteration: At step $t \geqslant 0$, allocate ball $t+1$ using the modified decision function

$$Q^t(\mathfrak{F}^t, j_1, j_2, i) := \begin{cases} G^t(\mathfrak{F}^t, j_1, j_2, i) & \text{if } |x_{j_1}^t - x_{j_2}^t| \leqslant g, \\ \widehat{Q}^t(\overline{x}^t, x_{j_1}^t, x_{j_2}^t, i) & \text{otherwise.} \end{cases}$$

Special cases: g-BOUNDED, g-MYOPIC-COMP

We will also look at two particular instances of these settings. The first one is where the adversary always allocates to the heavier bin. This setting was introduced in [7] for the TWO-CHOICE process under the name g-BOUNDED process. Here, we extend the definition to any average-aware load comparison process $\mathcal{P}$, such that g-BOUNDED$(\mathcal{P})$ is the g-ADV-COMP$(\mathcal{P}, G^t)$ process with

$$G^t(\mathfrak{F}^t, x_{j_1}^t, x_{j_2}^t, i) = \begin{cases} \mathbf{1}_{x_{j_1}^t > x_{j_2}^t} & \text{if } i = 1, \\ \mathbf{1}_{x_{j_1}^t \leqslant x_{j_2}^t} & \text{otherwise.} \end{cases}$$

We also define the g-MYOPIC-COMP$(\mathcal{P})$ process where the adversary decides randomly when the two bin loads differ by at most g, i.e., it is the g-ADV-COMP$(\mathcal{P}, G^t)$ with

$$G^t(\mathfrak{F}^t, x_{j_1}^t, x_{j_2}^t, i) = \frac{1}{2}.$$

2.4.6 ρ-NOISY setting

For a Markovian process $\mathcal{P} = $ SEQUENTIAL$(\widehat{q}^t)$, we define the ρ-NOISY-LOAD$(\mathcal{P})$ process, where in step t, the load of bin i is perturbed by $\widehat{\rho}_i^t \sim \rho$, i.e., $\widehat{x}^t = x^t + \widehat{\rho}^t$ and so its probability allocation vector is given by $q^t = \widehat{q}^t(\widehat{x}^t)$.

ρ-NOISY-LOAD($\mathcal{P}$) Setting:
Parameters:
- A density function ρ^t.
- A Markovian process $\mathcal{P} = \text{SEQUENTIAL}(\widehat{q}^t)$ with $\widehat{q}^t := \widehat{q}^t(x^t)$.

Iteration: At step $t \geqslant 0$, allocate ball $t+1$ using the allocation vector

$$q^t := \widehat{q}^t(\widehat{x}^t),$$

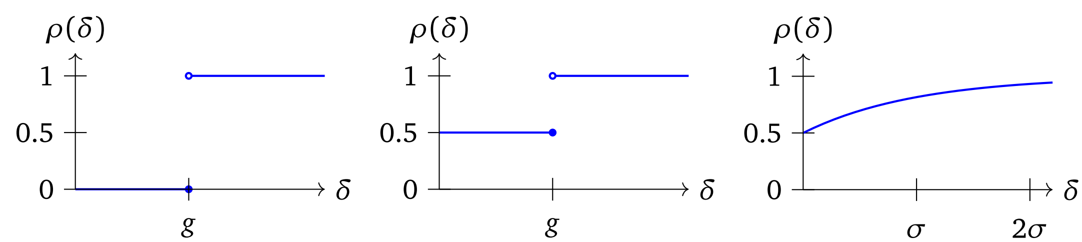
where $\widehat{x}_i^t = x_i^t + \widehat{\rho}_i^t$ with $\widehat{\rho}_i^t \sim \rho^t(\mathfrak{F}^t)$, for each $i \in [n]$.

Special cases: σ-NOISY-LOAD($\mathcal{P}$) (for $\rho = \sigma = \mathcal{N}(0, \sigma^2)$)

For any average-aware load comparison $\mathcal{P} = \text{TWO-SAMPLE}(\widehat{Q}^t)$, we also define the ρ-NOISY-COMP ($\mathcal{P}$) process, where $\rho : \mathbb{N} \to [0,1]$ gives the probability of a correct comparison, whose decision function Q^t is given by

$$Q^t(\overline{x}^t, x_{j_1}^t, x_{j_2}^t, i) = \widehat{Q}^t(\overline{x}^t, x_{j_1}^t, x_{j_2}^t, i) \cdot \rho(|x_{j_1}^t - x_{j_2}^t|) + (1 - \widehat{Q}^t(\overline{x}^t, x_{j_1}^t, x_{j_2}^t, i)) \cdot (1 - \rho(|x_{j_1}^t - x_{j_2}^t|)).$$

Special cases: $(1+\beta)$-process (for TWO-CHOICE with $\rho = \beta$), g-BOUNDED(TWO-CHOICE) (for $\rho(\delta) = \mathbf{1}_{\delta > g}$), g-MYOPIC-COMP(TWO-CHOICE) (for $\rho(\delta) = \mathbf{1}_{\delta > g} + \frac{1}{2} \cdot \mathbf{1}_{\delta \leqslant g}$).



(a) g-BOUNDED (b) g-MYOPIC-COMP (c) σ-NOISY-LOAD

Figure 2.7: In the graphs above, $\delta = \left|x_{j_1}^t - x_{j_2}^t\right|$ is the load difference among the two sampled bins and $\rho(\delta)$ is the probability that the load comparison is correct.

A special case of this is the σ-NOISY-LOAD setting, which is based on the following idea. When sampled at step t, a bin i reports an unbiased load estimate $\widehat{x}_i^t = x_i^t + Z_i^t$, where Z_i^t has a normal distribution $\mathcal{N}(0, \sigma^2)$ (and all $\{Z_i^t\}_{i \in [n], t \geqslant 0}$ are mutually independent). Then the process allocates a ball to the bin that reports the smallest load estimate. Thus if $x_{j_1}^t - x_{j_2}^t = \delta > 0$, the probability for allocating into the smaller bin can be computed as follows:

$$\begin{aligned}\mathbf{Pr}\left[\, \widehat{x}_{j_1}^{t-1} \leqslant \widehat{x}_{j_2}^{t-1} \,\right] &= \mathbf{Pr}\left[\, Z_{j_1}^{t-1} - Z_{j_2}^{t-1} \leqslant \delta \,\right] = 1 - \mathbf{Pr}\left[\, \mathcal{N}(0, 2\sigma^2) > \delta \,\right] = 1 - \mathbf{Pr}\left[\, \mathcal{N}(0,1) > \delta/(\sqrt{2}\sigma) \,\right] \\ &= 1 - \Phi(\delta/(\sqrt{2}\sigma)).\end{aligned}$$

Note that $\Phi(z) = 1/2$ for $z = 0$, and $\Phi(z)$ is increasing in z. As shown in [96, page 17],

$$\frac{1}{\sqrt{2\pi}} \cdot \frac{2}{\sqrt{z^2+4}+z} \cdot e^{-z^2/2} \leqslant 1 - \Phi(z) \leqslant \frac{1}{\sqrt{2\pi}} \cdot \frac{2}{\sqrt{z^2+2}+z} \cdot e^{-z^2/2}.$$

Hence ignoring the linear term in $1/z$, setting $z := \delta/(\sqrt{2}\sigma)$ and re-scaling σ, we can define σ-NOISY-LOAD as the setting which satisfies for all steps t and samples j_1, j_2 with $x_{j_2}^t - x_{j_1}^t = \delta > 0$ that

$$\mathbf{Pr}\left[\, \widehat{x}_{j_1}^t \leqslant \widehat{x}_{j_2}^t \,\right] = 1 - \frac{1}{2} \cdot \exp\left(-\left(\frac{\delta}{\sigma}\right)^2\right), \tag{2.3}$$

meaning that the correct comparison probability exhibits a Gaussian tail behaviour.

2.4.7 GRAPHICAL setting

We define the GRAPHICAL$(G, \mathcal{P})$ setting for any TWO-SAMPLE process $\mathcal{P}$ on the graph $G = (V = [n], E)$. In each step, an edge e is sampled randomly from the graph and the decision function of $\mathcal{P}$ is used to determine to which of the two nodes adjacent to e we are going to allocate the ball. In a sense, this extends the HETEROGENEOUS setting by relaxing the independence assumption of the two samples.

GRAPHICAL$(\mathcal{P}, G)$

Parameter:

- $G = (V, E)$: An undirected connected *graph*.
- $\mathcal{P}$: A TWO-SAMPLE(Q^t) process with decision function Q^t.

Iteration: At step $t \geqslant 0$, allocate ball $t+1$ using

$$q_i^t(\mathfrak{F}^t) = \sum_{(j_1,j_2)\in E} \sum_{k\in[2]:j_k=i} Q^t(\mathfrak{F}^t, j_1, j_2, k) \cdot \frac{1}{|E|}.$$

2.5 Additional processes

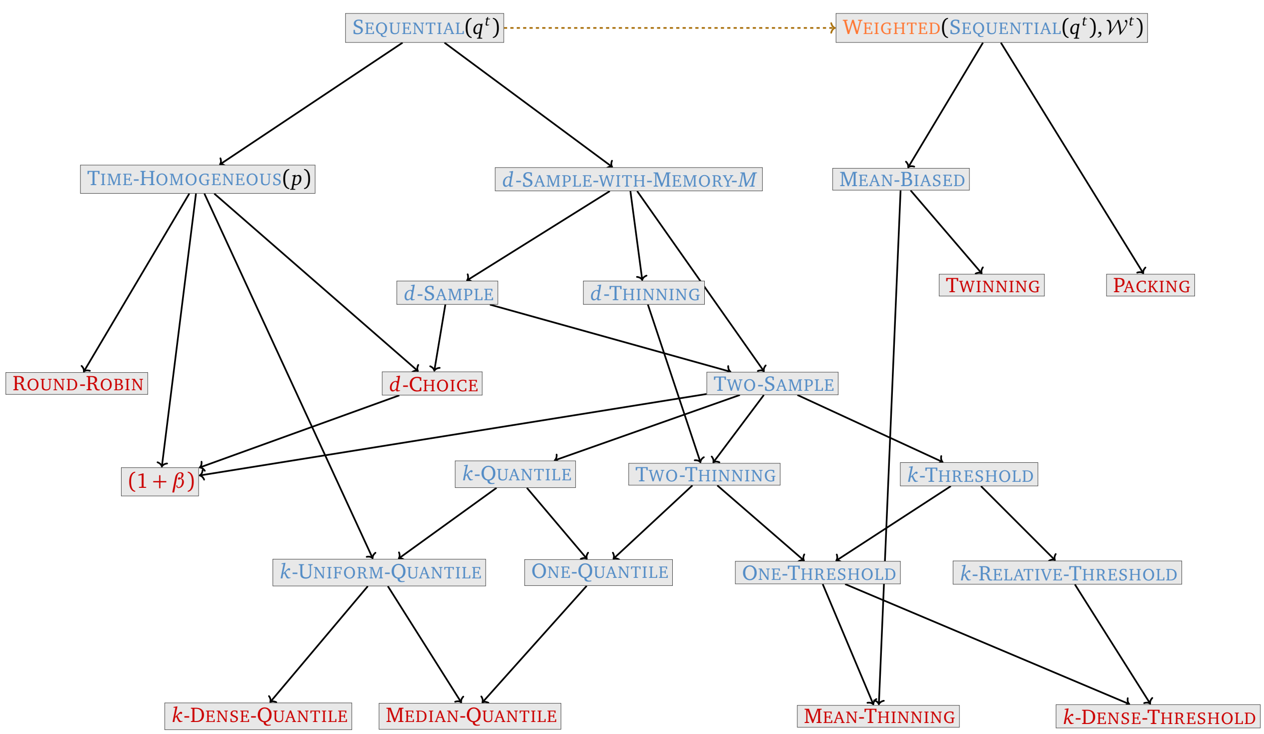


Figure 2.8: Hierarchy between some of processes and families of processes considered.

2.5.1 Thinning processes

> d-Thinning$(f_1^t, \ldots, f_{d-1}^t)$ Process ($\subseteq$ d-Sample(Q^t)):
>
> Parameters: Decision functions $f_1^t, \ldots, f_{d-1}^t$, where $f_k^t : \mathcal{F}^t \times [n]^k \to \{0,1\}$ decides whether to accept the k-th sample.
>
> Iteration: At step $t \geqslant 0$, sample d bins $j_1, \ldots, j_d \in [n]$, and allocate ball $t+1$ using the decision function
>
> $$Q^t(\mathfrak{F}^t, j_1, \ldots, j_d, i) = f_i^t(\mathfrak{F}^t, j_1, \ldots, j_i) \cdot \prod_{k=1}^{i-1} \big(1 - f_k^t(\mathfrak{F}^t, j_1, \ldots, j_k)\big).$$
>
> **Special cases:** One-Threshold, One-Quantile $\subseteq$ Two-Thinning
>
> [d-Sample]

For a d-Thinning process, we define the **number of samples** S^t taken up to step t, as

$$S^t := t + \sum_{s=1}^{t} \prod_{k=1}^{d-1} \big(1 - f_k^s(\mathfrak{F}^s, j_1^s, \ldots, j_k^s)\big),$$

and the **sample-efficiency** η^t as

$$\eta^t := \frac{\sum_{i=1}^n x_i^t}{S^t}.$$

2.5.2 Threshold processes

A special case of Two-Sample processes are the k-Threshold processes which sample two bins and send k non-adaptive queries to the bins $j \in \{j_1, j_2\}$ in the form "Is $x_j^t > f_i^t$?". The allocator gathers the responses and allocates to the bin that was witnessed to be smaller (see Fig. 1.6).

> k-Threshold$(f_1^t, f_2^t, \ldots, f_k^t)$ Process ($\subseteq$ Two-Sample(Q^t)):
>
> Parameters: Threshold functions $f_1^t, \ldots, f_k^t$, where $f_i^t : \mathcal{F}^t \to \mathbb{R}$ such that for any filtration $\mathfrak{F}^t \in \mathcal{F}^t$,
>
> $$\infty = f_0^t(\mathfrak{F}^t) > f_1^t(\mathfrak{F}^t) > f_2^t(\mathfrak{F}^t) > \ldots > f_k^t(\mathfrak{F}^t).$$
>
> Iteration: At step $t \geqslant 0$, we define the tightest witnessed upper bound for a bin as
>
> $$\ell(\mathfrak{F}^t, i) := \max\{u \in [k] \cup \{0\} : f_u^t(\mathfrak{F}^t) > x_i^t\},$$
>
> and define the decision function Q^t that decides between the two bin samples j_1 and j_2 as follows,
>
> $$Q^t(\mathfrak{F}^t, j_1, j_2, i) = \begin{cases} \frac{1}{2} & \text{if } \ell(\mathfrak{F}^t, j_1) = \ell(\mathfrak{F}^t, j_2), \\ \mathbf{1}_{\ell(\mathfrak{F}^t, j_1) > \ell(\mathfrak{F}^t, j_2)} & \text{else if } i = 1, \\ \mathbf{1}_{\ell(\mathfrak{F}^t, j_1) < \ell(\mathfrak{F}^t, j_2)} & \text{otherwise.} \end{cases}$$
>
> **Special cases:** One-Relative-Threshold $\subseteq$ Two-Threshold, k-Relative-Threshold
>
> [Two-Sample]

A particularly attractive family of k-THRESHOLD processes are the k-RELATIVE-THRESHOLD processes, where the thresholds are at a fixed offset from the average, i.e., $f_j^t = \overline{x}^t + r_j(n)$, where $r_j : \mathbb{N} \to \mathbb{N}$ and $1 \leqslant j \leqslant k$. This means that the thresholds can be computed with having just (an estimate for) the average (so they are average-aware load comparison processes). The MEAN-THINNING process is a special case with $k = 1$ and $r_1(n) = 0$ for all $n \in \mathbb{N}$. We call η-MIXED(MEAN-THINNING, ONE-CHOICE) the $(1+\eta)$-process. These will be analysed in Sections 5.2 and 5.2.2.

Note that ONE-THRESHOLD(f^t) is equivalent to the TWO-THINNING processes with $\widehat{f}^t(\mathfrak{F}^t, j) := \mathbf{1}_{f^t \leqslant x_j^t}$.

Observation 2.13. *For any* ONE-THRESHOLD(f^t) *process, there exists a load-vector indistinguishable* TWO-THINNING$(\widehat{f}^t)$ *process.*

However, for certain load vectors (where several bins have the same load), the THRESHOLD process cannot provide fine-grained control.

Remark 2.14. *There exists a* TWO-THINNING(f^t) *process that has no load-vector indistinguishable* ONE-THRESHOLD *equivalent.*

We can however add randomness to THRESHOLD processes obtaining the k-RANDOMISED-THRESHOLD processes that are load-vector indistinguishable to k-QUANTILE (as we shall see in Lemma 2.18). In Section 7.2.2, we analyse the k-DENSE-THRESHOLD process with the following thresholds,

$$f_j^t := \begin{cases} \frac{t}{n} & \text{if } j = 1, \\ \frac{t}{n} + \left\lceil \frac{3}{\gamma_2} \cdot j \cdot (\log n)^{1/k} \right\rceil - 2 & \text{if } 2 \leqslant j \leqslant k, \end{cases}$$

where $\gamma_2 > 0$ is a constant to be defined in Section 7.2.2.

MEAN-THINNING and RELATIVE-THRESHOLD

By the analysis of MEAN-THINNING, we can obtain upper bounds for RELATIVE-THRESHOLD$(f(n))$ with any non-negative threshold function $f(n)$.

Lemma 2.15. *Let $f(n)$ be any non-negative function. Let Gap_0 and $\mathrm{Gap}_{f(n)}$ be the gaps of* MEAN-THINNING *and* RELATIVE-THRESHOLD$(f(n))$*. Then $\mathrm{Gap}_{f(n)}$ is stochastically smaller than $\mathrm{Gap}_0 + f(n)$.*

Before proving the lemma, we need the following domination result:

Lemma 2.16. *Let $\mathcal{P}$ be the* THRESHOLD$(\frac{t}{n} + f(n))$ *process where $f(n)$ is non-negative, starting with an empty load vector $x_{\mathcal{P}}^0 = \mathbf{0}$. Further, let $\mathcal{Q}$ be the* THRESHOLD$(\frac{t}{n} + f(n))$ *process with initial load vector $(x_{\mathcal{Q}}^0)_1 = (x_{\mathcal{Q}}^0)_2 = \cdots = (x_{\mathcal{Q}}^0)_n = f(n)$. Then, there is a coupling so that at any step $t \geqslant 0$, it holds that $(x_{\mathcal{P}}^t)_i \leqslant (x_{\mathcal{Q}}^t)_i$, for any bin $i \in [n]$.*

Proof of Lemma 2.16. Let $j_1 = j_1^t$ and $j_2 = j_2^t$ be the two bins sampled at step $t \geqslant 0$, which are uniform and independent over $[n]$. We consider a coupling between $\mathcal{P}$ and $\mathcal{Q}$, where these random bin samples are identical, and prove inductively that for any $t \geqslant 0$ and any $i \in [n]$,

$$(x_{\mathcal{P}}^t)_i \leqslant (x_{\mathcal{Q}}^t)_i.$$

The base case $t = 0$ holds by definition. For the induction step, we make a case distinction:

Case 1 $[(x_{\mathcal{P}}^t)_{j_1} < t/n + f(n)]$. In this case, $\mathcal{P}$ allocates a ball to j_1. If $(x_{\mathcal{Q}}^t)_{j_1} < t/n + f(n)$, then $\mathcal{Q}$ also allocates a ball to j_1; otherwise, we have $(x_{\mathcal{Q}}^t)_{j_1} \geqslant t/n + f(n)$, and hence $(x_{\mathcal{Q}}^t)_{j_1} > (x_{\mathcal{P}}^t)_{j_1}$, i.e., $(x_{\mathcal{Q}}^t)_{j_1} \geqslant (x_{\mathcal{P}}^t)_{j_1} + 1$. This implies

$$(x_{\mathcal{Q}}^{t+1})_{j_1} = (x_{\mathcal{Q}}^t)_{j_1} \geqslant (x_{\mathcal{P}}^t)_{j_1} + 1 = (x_{\mathcal{P}}^{t+1})_{j_1},$$

and the inductive step follows from this and the induction hypothesis.

Case 2 $[(x_{\mathcal{P}}^t)_{j_1} \geqslant t/n + f(n)]$. In this case, $\mathcal{P}$ allocates a ball to j_1. By induction hypothesis, $(x_{\mathcal{P}}^t)_{j_1} \leqslant (x_{\mathcal{Q}}^t)_{j_1}$, which implies $\mathcal{Q}$ also allocates a ball to j_1. Thus we have

$$(x_{\mathcal{Q}}^{t+1})_{j_1} = (x_{\mathcal{Q}}^t)_{j_1} + 1 \geqslant (x_{\mathcal{P}}^t)_{j_1} + 1 = (x_{\mathcal{P}}^{t+1})_{j_1},$$

and the inductive step is complete. Since in both cases all other bins remain unchanged the proof is complete. $\square$

Lemma 2.17. *Let $\mathcal{P}$ be the* THRESHOLD$(\frac{t}{n} + f(n))$ *process starting with the initial load vector $(x_{\mathcal{P}}^0)_1 = \ldots = (x_{\mathcal{P}}^0)_n = f(n)$ and $\mathcal{Q}$ be the* MEAN-THINNING *process with initial load vector $x_{\mathcal{Q}}^0 = \mathbf{0}$. Then, there is a coupling so that $x_{\mathcal{P}}^t = x_{\mathcal{Q}}^t + f(n)$ for any step $t \geqslant 0$.*

Proof. In the execution of the process $\mathcal{P}$, we start the process at step $t = 0$ from an initial load of $f(n)$ balls in each bin. Since the threshold is $t/n + f(n)$, that is the process $\mathcal{P}$ does not have a threshold relative to the actual average load, the effect of adding these balls is to reduce the threshold of $\mathcal{P}$ by exactly $f(n)$. Thus, $\mathcal{P}$ is operating with a threshold of $t/n + f(n) - f(n) = t/n$, which is equivalent to MEAN-THINNING process, i.e., $\mathcal{Q}$. So, we obtain a coupling such that $x_{\mathcal{P}}^t = x_{\mathcal{Q}}^t + f(n)$ for any step $t \geqslant 0$. $\square$

We can now complete the proof of Lemma 2.15.

Proof of Lemma 2.15. We define the following processes:

- $\mathcal{P}_1$: The RELATIVE-THRESHOLD$(f(n))$ process (starting from the empty load vector).
- $\mathcal{P}_2$: The THRESHOLD$(\frac{t}{n} + f(n))$ starting with $(x_{\mathcal{P}_2}^0)_1 = \ldots = (x_{\mathcal{P}_2}^0)_n = f(n)$.
- $\mathcal{P}_3$: The MEAN-THINNING process (starting from the empty load vector).

By Lemma 2.16, there exists a coupling such that $x_{\mathcal{P}_2}^t$ pointwise majorises $x_{\mathcal{P}_1}^t$ for any step $t \geqslant 0$. By Lemma 2.17, there exists a coupling such that $x_{\mathcal{P}_2}^t = x_{\mathcal{P}_3}^t + f(n)$. Hence, we deduce that there is a coupling between the three processes such that

$$\mathrm{Gap}_{f(n)}(t) = \mathrm{Gap}_{\mathcal{P}_1}(t) = \max_{i\in[n]}(x_{\mathcal{P}_1}^t)_i - \frac{t}{n} \leqslant \max_{i\in[n]}(x_{\mathcal{P}_2}^t)_i - \frac{t}{n} = \max_{i\in[n]}(x_{\mathcal{P}_3}^t)_i + f(n) - \frac{t}{n} = \mathrm{Gap}_0(t) + f(n).$$ $\square$

2.5.3 QUANTILE processes

k-QUANTILE$(\delta_1^t, \delta_2^t, \ldots, \delta_k^t)$ Process ($\subseteq$ TWO-SAMPLE(Q^t)):

Parameter: Quantile functions $\delta_1^t, \ldots, \delta_k^t$ (for $k \leqslant n$), where $\delta_i^t : \mathcal{F}^t \to \{1/n, 2/n, \ldots, 1\}$ such that for any filtration $\mathfrak{F}^t$,

$$\delta_0 = 0 < \delta_1^t(\mathfrak{F}^t) < \delta_2^t(\mathfrak{F}^t) < \ldots < \delta_k^t(\mathfrak{F}^t).$$

Iteration: At step $t \geqslant 0$, we define the tightest witnessed upper bound for a bin

$$\ell(\mathfrak{F}^t, i) := \max\{j \in [k] \cup \{0\} : n \cdot \delta_j^t(\mathfrak{F}^t) < \mathrm{Rank}^t(i)\},$$

and define the decision function Q^t as follows

$$Q^t(\mathfrak{F}^t, j_1, j_2, i) = \begin{cases} \frac{1}{2} & \text{if } \ell(\mathfrak{F}^t, j_1) = \ell(\mathfrak{F}^t, j_2), \\ \mathbf{1}_{\ell(\mathfrak{F}^t, j_1) > \ell(\mathfrak{F}^t, j_2)} & \text{else if } i = 1, \\ \mathbf{1}_{\ell(\mathfrak{F}^t, j_1) < \ell(\mathfrak{F}^t, j_2)} & \text{otherwise.} \end{cases}$$

Special cases: k-UNIFORM-QUANTILE, TWO-CHOICE
[TWO-SAMPLE]

In the special case where the quantile functions do not depend on time nor load vector we call this a k-UNIFORM-QUANTILE process, i.e., $\delta_j^t := \delta_j(n)$ for $j \in [k]$. A process of this family is also TIME-HOMOGENEOUS(p), with probability vector p defined as

$$p_i = \begin{cases} \frac{\delta_1}{n}, & \text{if } i \leqslant \delta_1 n, \\ \frac{\delta_1 + \delta_2}{n}, & \text{if } \delta_1 n < i \leqslant \delta_2 n, \\ \vdots & \\ \frac{\delta_{k-1} + \delta_k}{n} & \text{if } \delta_{k-1} n < i \leqslant \delta_k n, \\ \frac{1 + \delta_k}{n} & \text{if } i > \delta_k n. \end{cases}$$

Lemma 2.18. *For every k-QUANTILE process $\mathcal{P}$, there is a k-RANDOMISED-THRESHOLD process $\mathcal{Q}$, such that $\mathcal{P}$ and $\mathcal{Q}$ are load-vector indistinguishable.*

Note that since QUANTILE(δ^t) is indistinguishable from TWO-THINNING(f^t) with $f^t(\mathfrak{F}^t, j) := \mathbf{1}_{\mathrm{Rank}^t(j) > \delta^t n}$, we have that:

Observation 2.19. *For any QUANTILE(δ^t) process, there is a load-vector indistinguishable TWO-THINNING (f^t) process.*

In Section 7.2.1, we will prove for the following concrete k-UNIFORM-QUANTILE process that w.h.p. $\mathrm{Gap}(t) = \mathcal{O}(k(\log n)^{1/k})$ for any $1 \leqslant k \leqslant \mathcal{O}(\log \log n)$.

k-DENSE-QUANTILE Process ($=$ k-UNIFORM-QUANTILE$(\delta_1, \delta_2, \ldots, \delta_k)$):
Parameters: We define

$$\widetilde{\delta}_i = \begin{cases} \frac{1}{2} & \text{if } i = k, \\ e^{-\frac{1}{4}(\log n)^{(k-i)/k}} & \text{if } 1 \leqslant i < k\,, \end{cases} \tag{2.4}$$

and the quantiles $\delta_1, \ldots, \delta_k$ as the $\widetilde{\delta}_i$'s rounded up to the nearest multiple of $1/n$.

Special cases: MEDIAN-QUANTILE (for $k = 1$)
[time homo. prob. vector, TWO-SAMPLE]

2.5.4 Relation between THRESHOLD and QUANTILE processes

We state the following results connecting k-THRESHOLD and k-QUANTILE processes and defer their proofs to Appendix D.1.

Lemma D.1 (Restated, page 225). *Any* THRESHOLD$(f_1^t, \ldots, f_k^t)$ *process can be simulated by a* QUANTILE $(\delta_1^t, \ldots, \delta_k^t)$ *process.*

Lemma D.2 (Restated, page 225). *For any* QUANTILE$(\delta_1^t, \ldots, \delta_k^t)$ *process, there exist thresholds* $f_1^t, \ldots, f_k^t$ *and probability vector* $(\beta_1^t, \ldots, \beta_k^t)$ *such that* $(\beta_1^t, \ldots, \beta_k^t)$—MIXED(THRESHOLD$(f_1^t), \ldots,$ THRESHOLD(f_k^t)).

Lemma D.3 (Restated, page 226). *Any* QUANTILE$(\delta_1, \ldots, \delta_k)$ *process can be simulated by an adaptive (and randomised)* $(2k)$-THINNING *process.*

2.5.5 TWINNING and PACKING processes

The process WEIGHTED(ONE-CHOICE, $\mathcal{W}^t$) where $\mathcal{W}^t(W^t, x_{i^{t+1}}^t) = 1 + \mathbf{1}_{y_{i^{t+1}}^t < 0}$ was presented in [117] (see Fig. 2.9). Here, we also give an explicit definition:

TWINNING Process:
Iteration: At step $t \geqslant 0$, sample a bin $i \in [n]$ uniformly at random, and update its load:

$$x_i^{t+1} = \begin{cases} x_i^t + 1 & \text{if } x_i^t \geqslant \frac{W^t}{n}, \\ x_i^t + 2 & \text{if } x_i^t < \frac{W^t}{n}. \end{cases}$$

[ONE-SAMPLE, WEIGHTED(ONE-CHOICE, $\mathcal{W}^t$)]

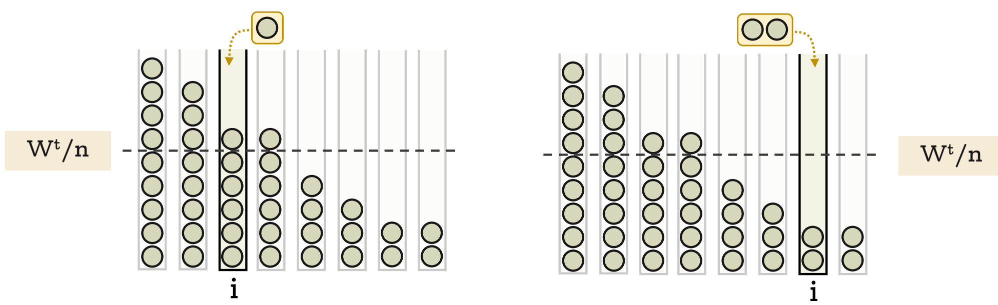


Figure 2.9: The two cases for the TWINNING process: (i) allocating to an overloaded bin and (ii) allocating to an underloaded bin.

Remark 2.20. *This process can be implemented in any d-regular graph (even the cycle), by sampling an edge randomly and then choosing one of its endpoints randomly. As we show in Corollary 5.12,* TWINNING *achieves w.h.p. an* $\mathcal{O}(\log n)$ *gap, which is much better than the observed* poly(n) *gap of* TWO-CHOICE *(e.g., on the cycle).*

The following process is a more extreme version of TWINNING, which "fills up" underloaded bins until they become overloaded. In particular it is the process WEIGHTED(ONE-CHOICE, $\mathcal{W}^t$) where $\mathcal{W}^t(W^t, x_{i^{t+1}}^t) = 1 + \mathbf{1}_{y_{i^{t+1}}^t < 0} \cdot \lceil -y_{i^{t+1}}^t \rceil$ (see Fig. 2.10).

PACKING Process:
Iteration: At step $t \geqslant 0$, sample a bin $i \in [n]$ uniformly at random, and update its load:

$$x_i^{t+1} = \begin{cases} x_i^t + 1 & \text{if } x_i^t \geqslant \frac{W^t}{n}, \\ \lceil \frac{W^t}{n} \rceil + 1 & \text{if } x_i^t < \frac{W^t}{n}. \end{cases}$$

[ONE-SAMPLE, WEIGHTED(ONE-CHOICE, $\mathcal{W}^t$)]

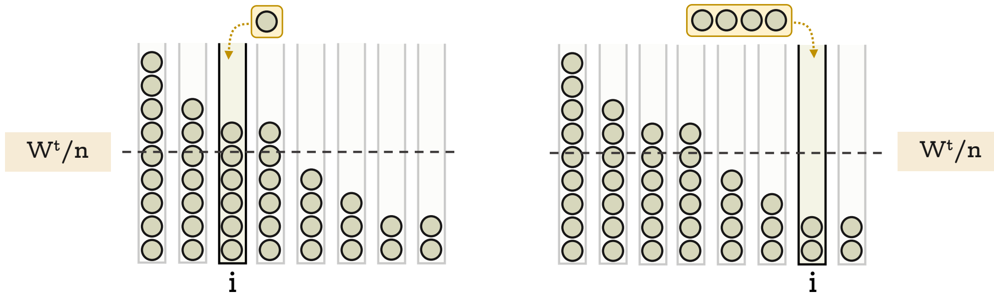


Figure 2.10: The two cases for the PACKING process: (i) allocating to an overloaded bin and (ii) allocating to an underloaded bin.

Since every ONE-SAMPLE process is trivially a ONE-THINNING process, sample efficiency is defined (and is of interest) for both TWINNING and PACKING.

2.5.6 MEAN-BIASED processes

In this section, we define the MEAN-BIASED processes that captures both the TWINNING and MEAN-THINNING, as well as several other processes. First, we define $q_+^t := \max_{i \in B_+^t} q_i^t$ and $q_-^t := \min_{i \in B_-^t} q_i^t$, as the largest (smallest) probability for allocating to a fixed overloaded (underloaded) bin at step t, respectively. We define,

Condition $\mathcal{P}_2$: At any step $t \geqslant 0$, the probability allocation vector q^t must satisfy $q_+^t \leqslant \frac{1}{n} \leqslant q_-^t$.

Condition $\mathcal{W}_2$: At any step $t \geqslant 0$, if $i := i^{t+1}$ is chosen for allocation,

- If $y_i^t < 0$, then allocate w_- balls to bin i,
- If $y_i^t \geqslant 0$, then allocate w_+ balls to bin i,

where $1 \leqslant w_+ \leqslant w_-$ are constant integers.

Both conditions are natural, but on their own they are not sufficient to establish a good bound on the gap, as the ONE-CHOICE process satisfies both conditions with equalities. Thus, we will require processes to satisfy at least one of two stronger versions of $\mathcal{P}_2$ and $\mathcal{W}_2$:

Condition $\mathcal{P}_3$: This is as Condition $\mathcal{P}_2$, but additionally, there are time-independent constants $k_1 \in (0,1], k_2 \in (0,1]$ such that for any step $t \geqslant 0$:

$$
\begin{aligned}
q_+^t &\leqslant \frac{1-k_1}{n} + \frac{k_1 \cdot |B_+^t|}{n^2} = \frac{1}{n} - \frac{k_1 \cdot (1-\delta^t)}{n}, \\
q_-^t &\geqslant \frac{1}{n} + \frac{k_2 \cdot |B_+^t|}{n^2} = \frac{1}{n} + \frac{k_2 \cdot \delta^t}{n}.
\end{aligned}
$$

Condition $\mathcal{W}_3$: This is as Condition $\mathcal{W}_2$, but additionally we have the strict inequality: $w_+ < w_-$. Also, we assume that for each $t \geqslant 0$, allocation vector $\widetilde{q}_i^t$ is non-decreasing in i.

The reason we attach the non-decreasing property of $\widetilde{q}^t$ to $\mathcal{W}_3$ and not to $\mathcal{P}_2$ is to make our main result slightly stronger. We call MEAN-BIASED processes the ones that satisfy $\mathcal{P}_2$ and $\mathcal{W}_3$ or $\mathcal{P}_3$ and $\mathcal{W}_2$.

The rationale behind condition $\mathcal{P}_3$ is that we wish to slightly bias the allocation vector q^t towards underloaded bins at each step t. However, it is natural to assume that this influence is limited by a process that samples, say, at most two bins uniformly and independently, and then allocates balls to the least loaded of the two. Concretely, if a process takes *two* independent and uniform bin samples at each step, the probability of picking two overloaded bins equals $\left(\frac{|B_+^t|}{n}\right)^2$. Hence by averaging, there must be a bin $i \in B_+^t$ such that

$$q_+^t \geqslant q_i^t \geqslant \left(\frac{|B_+^t|}{n}\right)^2 \cdot \frac{1}{|B_+^t|} = \frac{|B_+^t|}{n^2}.$$

The relaxation of the first constraint in $\mathcal{P}_3$ by taking a strict convex combination of $\frac{1}{n}$ and $\frac{|B_+^t|}{n^2}$ ensures some slack, for instance, it allows the framework to cover the
OnePlusEta-process, a "noisy" version of Mean-Thinning, where at each step, it performs a One-Choice allocation with some constant probability $\eta \in (0,1)$, and otherwise we perform an allocation following the Mean-Thinning process (see Lemma D.5 for details). Similarly, for any process which takes at most two uniform samples, by averaging, there must be a bin $j \in B_-^t$ such that

$$q_-^t \leqslant q_j^t \leqslant \frac{1 - \frac{|B_+^t|^2}{n^2}}{|B_-^t|} = \frac{(n - |B_+^t|) \cdot (n + |B_+^t|)}{n^2 |B_-^t|} = \frac{1}{n} + \frac{|B_+^t|}{n^2}.$$

Finally, we remark that $\mathcal{P}_3$ resembles the framework of [152, Equation 2], where $p_i^t = p_i$ is non-decreasing in i and $p_{n/3} \leqslant \frac{1-4\epsilon}{n}$ and $p_{2n/3} \geqslant \frac{1+4\epsilon}{n}$ holds for some $0 < \epsilon < 1/4$ (not necessarily constant). In contrast to that, for constants $k_1, k_2 > 0$, the conditions in $\mathcal{P}_3$ are relaxed as they only imply such a bias if δ^t is bounded away from 0 and 1, which may not hold in all steps. It is straightforward that Twinning fits into the Mean-Biased framework.

Lemma 2.21. *The Twinning process is a Mean-Biased process.*

Proof. The Twinning process satisfies $\mathcal{W}_3$, since $w_- = 2 > 1 = w_+$ and $\mathcal{P}_2$, since bins are sampled uniformly at random in each step. □

Mean-Thinning also fits into the framework, but satisfying a different set of conditions.

Lemma 2.22. *The Mean-Thinning process is a Mean-Biased process.*

Proof. The probability of allocating to any overloaded bin $i \in B_+^t$ is $q_i^t = \frac{\delta^t}{n} = \frac{1}{n} - \frac{1 \cdot (1-\delta^t)}{n}$, so we can choose $k_1 := 1$. For any underloaded bin $i \in B_-^t$, $q_i^t = \frac{1+\delta^t}{n} = \frac{1}{n} + \frac{1 \cdot \delta^t}{n}$, so we can choose $k_2 := 1$, and $\mathcal{P}_3$ holds. Condition $\mathcal{W}_2$ is trivially satisfied, since $w_+ = w_- = 1$. □

The $(1+\eta)$-process also fits into the framework:

Lemma D.5 (Restated, page 228). *For any constant $\eta > 0$, the $(1+\eta)$-process is a Mean-Biased process.*

Finally, we just remark that although the $(1+\beta)$-process is not a Mean-Biased process it can be majorised by $(1+\eta)$. So, the upper bound on the Mean-Biased processes also applies to $(1+\beta)$.

Lemma D.6 (Restated, page 228). *For any constant $\beta \in (0,1]$, the $(1+\eta)$-process with $\eta = \beta$ majorizes $(1+\beta)$-process at each step.*

Expectation of the hyperbolic cosine potential

Peres, Talwar and Wieder [152] analysed the hyperbolic cosine potential for a large family of processes. In this chapter,[1] we will present a refined analysis for the expectation of the hyperbolic cosine potential which also holds for a large family of processes including weights and outdated information (see discussion below for details of the refinements). Recall that the *hyperbolic cosine potential* with smoothing parameter $\gamma > 0$ is defined as

$$\Gamma^t := \Gamma^t(\gamma) := \Phi^t + \Psi^t := \sum_{i=1}^n e^{\gamma y_i^t} + \sum_{i=1}^n e^{-\gamma y_i^t}, \tag{3.1}$$

where $\Phi^t := \Phi^t(\gamma)$ is the *overload exponential potential* and $\Psi^t := \Psi^t(\gamma)$ is the *underload exponential potential*. We also decompose Γ^t by defining

$$\Gamma_i^t := \Phi_i^t + \Psi_i^t, \quad \text{where } \Phi_i^t := e^{\gamma y_i^t} \text{ and } \Psi_i^t := e^{-\gamma y_i^t} \quad \text{for any bin } i \in [n].$$

Further, we use the following shorthands to denote the changes in the potentials over one step $\Delta\Phi_i^{t+1} := \Phi_i^{t+1} - \Phi_i^t$, $\Delta\Psi_i^{t+1} := \Psi_i^{t+1} - \Psi_i^t$ and $\Delta\Gamma_i^{t+1} := \Gamma_i^{t+1} - \Gamma_i^t$.

The following theorem was proven in [152, Section 2].

Theorem 3.1 ([152, Section 2]). *Consider any* Time-Homogeneous(p) *process with non-decreasing p, satisfying for some $\epsilon \in (0, 1/4)$ that*

$$p_i \leqslant \frac{1-4\epsilon}{n}, \quad \textit{for any } i \leqslant \frac{n}{3}, \quad \textit{and} \quad p_i \geqslant \frac{1+4\epsilon}{n}, \quad \textit{for any } i \geqslant \frac{2n}{3}.$$

Further, consider the Weighted *setting with weights from a* Finite-MGF(ζ, S) *distribution with $S \geqslant 1$. Then, for $\Gamma := \Gamma(\gamma)$ with $\gamma := \min\left\{\frac{\epsilon}{6S}, \frac{\zeta}{2}\right\}$, there exists $c = \operatorname{poly}(1/\epsilon)$ such that for any step $t \geqslant 0$,*

$$\mathbf{E}\left[\, \Delta\Gamma^{t+1} \,\middle|\, \mathfrak{F}^t \,\right] \leqslant -\Gamma^t \cdot \frac{\gamma\epsilon}{4n} + c.$$

Recall from Section 1.3 (see also Lemma B.1 (ii)), that when a process satisfies this drop inequality it also satisfies $\mathbf{E}\left[\, \Gamma^t \,\right] \leqslant \frac{4c}{\gamma\epsilon} \cdot n$ for every step $t \geqslant 0$.

In order to state our generalised version of the theorem, we first recall the conditions $\mathcal{C}_1$ and $\mathcal{C}_2$ for a probability vector p as defined in Section 2.3.4:

- **Condition $\mathcal{C}_1$:** There exist constant $\delta \in (0, 1)$ and (not necessarily constant) $\epsilon \in (0, 1)$, such that for any $1 \leqslant k \leqslant \delta \cdot n$,

  $$\sum_{i=1}^k p_i \leqslant (1-\epsilon) \cdot \frac{k}{n},$$

  and similarly for any $\delta \cdot n + 1 \leqslant k \leqslant n$,

  $$\sum_{i=k}^n p_i \geqslant \left(1 + \epsilon \cdot \frac{\delta}{1-\delta}\right) \cdot \frac{n-k+1}{n}.$$

  It follows by Observation 2.10, that any process that satisfies the precondition in Theorem 3.1 with $\epsilon' \in (0, 1/4)$ also satisfies condition $\mathcal{C}_1$ with $\epsilon = 4\epsilon'$ and $\delta = 1/3$.

[1] After the submission of this thesis, we compiled the content of this chapter in [115], where we also included some simpler-to-verify conditions for the main theorem.

- **Condition** $\mathcal{C}_2$: There exists $C > 1$, such that $\max_{i\in[n]} p_i \leqslant \frac{C}{n}$.

  Recall that any d-SAMPLE process satisfies this condition with $C = d$.

As we shall describe shortly, the main theorem (Theorem 3.2) applies to a variety of settings. However, in order to more precisely highlight the differences to Theorem 3.1, we first state a corollary for TIME-HOMOGENEOUS processes in the non-batched WEIGHTED setting with weights from a FINITE-MGF distribution. The two main differences are: (i) that p satisfies preconditions $\mathcal{C}_1$ and $\mathcal{C}_2$, and (ii) the additive term which changes from $\mathrm{poly}(1/\epsilon)$ to $\mathcal{O}(\gamma\epsilon)$.

Corollary 3.6 (Of Theorem 3.2 – Restated, page 64). *Consider any* TIME-HOMOGENEOUS(p) *process with p satisfying condition $\mathcal{C}_1$ for some constant $\delta \in (0,1)$ and some $\epsilon \in (0,1)$, and condition $\mathcal{C}_2$ for some $C > 1$.*

Further, consider the WEIGHTED *setting with weights from a* FINITE-MGF(S) *distribution with $S \geqslant 1$. Then, there exists a constant $c := c(\delta) > 0$, such that for $\Gamma := \Gamma(\gamma)$ with any $\gamma \in \left(0, \frac{\epsilon\delta}{16CS}\right]$ and for any step $t \geqslant 0$,*

$$\mathbf{E}\left[\, \Delta\Gamma^{t+1} \,\middle|\, \mathfrak{F}^t \,\right] \leqslant -\Gamma^t \cdot \frac{\gamma\epsilon\delta}{8n} + c\gamma\epsilon.$$

Now we state the main theorem, where the preconditions are expressed in terms of the expected change of the overload and underload potentials for a folding of a sequential allocation process. Note that in the following theorem the probability vector p need not be the probability allocation vector q of the process being considered. When the rounds consist of multiple steps, then this probability vector expresses some kind of "average number" of balls allocated to the i-th bin.

Theorem 3.2. *Consider any folding of a $\mathcal{P} =$ SEQUENTIAL(q^t) process and a probability vector p^t satisfying condition $\mathcal{C}_1$ for some constant $\delta \in (0,1)$ and some $\epsilon \in (0,1)$ at every round $t \geqslant 0$. Further assume that there exist $K > 0$, $\gamma \in \left(0, \min\left\{1, \frac{\epsilon\delta}{8K}\right\}\right]$ and $R > 0$, such that for any round $t \geqslant 0$, process $\mathcal{P}$ satisfies for potentials $\Phi := \Phi(\gamma)$ and $\Psi := \Psi(\gamma)$ that for bins sorted in non-increasing order of their loads,*

$$\sum_{i=1}^{n} \mathbf{E}\left[\, \Delta\Phi_i^{t+1} \,\middle|\, \mathfrak{F}^t \,\right] \leqslant \sum_{i=1}^{n} \Phi_i^t \cdot \left(\left(p_i^t - \frac{1}{n}\right) \cdot R \cdot \gamma + K \cdot R \cdot \frac{\gamma^2}{n}\right),$$

and,

$$\sum_{i=1}^{n} \mathbf{E}\left[\, \Delta\Psi_i^{t+1} \,\middle|\, \mathfrak{F}^t \,\right] \leqslant \sum_{i=1}^{n} \Psi_i^t \cdot \left(\left(\frac{1}{n} - p_i^t\right) \cdot R \cdot \gamma + K \cdot R \cdot \frac{\gamma^2}{n}\right).$$

Then, there exists a constant $c := c(\delta) > 0$, such that for $\Gamma := \Gamma(\gamma)$ and any round $t \geqslant 0$,

$$\mathbf{E}\left[\, \Delta\Gamma^{t+1} \,\middle|\, \mathfrak{F}^t \,\right] \leqslant -\Gamma^t \cdot R \cdot \frac{\gamma\epsilon\delta}{8n} + R \cdot c\gamma\epsilon,$$

and

$$\mathbf{E}\left[\, \Gamma^t \,\right] \leqslant \frac{8c}{\delta} \cdot n.$$

This theorem is a refinement of Theorem 3.1 in the following ways:

- When rounds consist of a single step and the allocation vector q coincides with the probability vector p, we relax the preconditions on p, requiring that it satisfies conditions $\mathcal{C}_1$ and $\mathcal{C}_2$.

  This allows us to apply the theorem for TWO-CHOICE on the WEIGHTED, GRAPHICAL setting for d-regular expanders as q satisfies $\mathcal{C}_1$ and $\mathcal{C}_2$ (as we shall see in Lemma 7.40). Note that the majorisation argument in [152, Section 3] only applies to the unit weights case.

- It has two parts: (i) proving an upper bound on the expected change of the Φ and Ψ potentials and (ii) combining these bounds to deduce a bound on the expected change of Γ.

  This split allows us to apply the theorem for processes that allocate balls to more than one bins, such as TWINNING-WITH-QUANTILE (Section 3.2.1) and $(1,1,2)$-RESET-MEMORY, and also to the b-BATCHED setting (Sections 3.2.2 and 3.2.3).

- We show that $\Gamma := \Gamma(\gamma)$ satisfies the following drop inequality for some *constants* $c_1, c_2 > 0$,

$$\mathbf{E}\left[\, \Delta\Gamma^{t+1} \,\middle|\, \mathfrak{F}^t \,\right] \leqslant -\Gamma^t \cdot \frac{c_1 \gamma \epsilon}{n} + c_2 \gamma \epsilon.$$

  This allows us to deduce that for any round $t \geqslant 0$, it holds that $\mathbf{E}\left[\, \Gamma^t \,\right] \leqslant \frac{c_2}{c_1} \cdot n$ (Lemma B.1 (ii)) and this directly implies the tight $\mathcal{O}\big(\frac{\log n}{\beta}\big)$ bound on the $(1+\beta)$-process for all $\beta \leqslant 1-\epsilon'$, for any constant $\epsilon' > 0$ (Theorem 3.7). Furthermore, it allows us to prove, in Chapter 4, that Γ concentrates at $\mathcal{O}(n)$ which we later use to prove the tighter gap bounds for the b-BATCHED setting, improving $\mathcal{O}\big(\frac{b}{n} \cdot \log n\big)$ to $\mathcal{O}\big(\frac{b}{n}\big)$ (Section 7.5), and for the QUANTILE(δ^*) process, improving $\mathcal{O}\big(\big(\frac{\log n}{\log\log n}\big)^2\big)$ to $\mathcal{O}\big(\frac{\log n}{\log\log n}\big)$ (Section 7.2.3).

The key lemma that we use to prove Theorem 3.2 is a drift inequality that is agnostic of balanced allocation processes and is essentially an inequality involving Φ, Ψ, Γ over an arbitrary (load) vector x (with $y = x - \overline{x}$ being its normalised version) and a probability vector p satisfying condition $\mathcal{C}_1$.

Lemma 3.3 (Key Lemma). *Consider any probability vector p satisfying condition $\mathcal{C}_1$ for constant $\delta \in (0,1)$ and (not necessarily constant) $\epsilon \in (0,1)$, and any sorted load vector $x \in \mathbb{R}^n$ with $\Phi := \Phi(\gamma)$, $\Psi := \Psi(\gamma)$ and $\Gamma := \Gamma(\gamma)$ for any smoothing parameter $\gamma \in (0,1]$. Further define,*

$$\Delta\overline{\Phi} := \sum_{i=1}^{n} \Delta\overline{\Phi}_i := \sum_{i=1}^{n} \Phi_i \cdot \left(p_i - \frac{1}{n}\right) \cdot \gamma, \quad \textit{and} \quad \Delta\overline{\Psi} := \sum_{i=1}^{n} \Delta\overline{\Psi}_i := \sum_{i=1}^{n} \Psi_i \cdot \left(\frac{1}{n} - p_i\right) \cdot \gamma.$$

Then, there exists a constant $c := c(\delta) > 0$, such that

$$\Delta\overline{\Gamma} := \sum_{i=1}^{n} \Delta\overline{\Gamma}_i := \sum_{i=1}^{n} \Delta\overline{\Phi}_i + \Delta\overline{\Psi}_i \leqslant -\Gamma \cdot \frac{\gamma\epsilon\delta}{4n} + c\gamma\epsilon.$$

Organisation. The remainder of this chapter is structured as follows. In Section 3.1, we prove the key lemma and the main theorem. Then, in Section 3.2.1 we prove Corollary 3.6 for the non-batched setting and apply it to get bounds for the QUANTILE(δ), $(1+\beta)$ and TWINNING-WITH-QUANTILE processes. In Section 3.2.2, for the b-BATCHED setting, we show that processes whose allocation vectors satisfy conditions $\mathcal{C}_1$ and $\mathcal{C}_2$, satisfy the preconditions of the main theorem, allowing us to deduce an $\mathcal{O}\big(\frac{b}{n} \cdot \log n\big)$ bound on their gap. Finally, in Section 3.2.3 for processes whose allocation vectors satisfy condition $\mathcal{C}_3$ (i.e., they are *close* to the uniform distribution), we deduce an $\mathcal{O}\big(\sqrt{\frac{b}{n}} \cdot \log n\big)$ bound on their gap.

Later, in Section 7.5 we will improve the $\mathcal{O}\big(\frac{b}{n} \cdot \log n\big)$ bound to $\mathcal{O}\big(\frac{b}{n}\big)$, which in Appendix C.1 we show to be tight. Similarly in Section 7.5, we will also improve the $\mathcal{O}(\sqrt{\frac{b}{n}} \cdot \log n)$ bound to $\mathcal{O}(\sqrt{\frac{b}{n} \cdot \log n})$, which again in Appendix C.1, we show to be tight.

3.1 Proof of Theorem 3.2

We start by proving the key lemma. Note that this holds for any probability vector p satisfying condition $\mathcal{C}_1$ and any sorted load vector x (with y being its normalised load vector). Before presenting the proof, we outline the key ideas in the proof:

1. It suffices to analyse $\Delta\overline{\Gamma} = \Delta\overline{\Phi} + \Delta\overline{\Psi}$ for the simplified probability vector,

$$r_i := \begin{cases} \frac{1-\epsilon}{n} & \text{if } i \leqslant \delta n, \\ \frac{1+\widetilde{\epsilon}}{n} & \text{otherwise}, \end{cases} \tag{3.2}$$

where $\widetilde{\epsilon} := \epsilon \cdot \frac{\delta}{1-\delta}$, as r maximises the terms $\Delta\overline{\Phi}$ and $\Delta\overline{\Psi}$, over all probability vectors satisfying condition $\mathcal{C}_1$ for a given δ and ϵ.

2. For any bin $i \in [n]$, there is one dominant term in $\Gamma_i = \Phi_i + \Psi_i = e^{\gamma y_i} + e^{-\gamma y_i}$: for overloaded bins ($y_i \geqslant 0$) it is $\Phi_i = e^{\gamma y_i}$ (and $\Psi_i = e^{-\gamma y_i} \leqslant 1$) and for underloaded bins ($y_i < 0$) it is Ψ_i (and $\Phi_i \leqslant 1$). In Claim 3.4, we show that the contribution of the non-dominant term in $\Delta\overline{\Gamma}$ is subsumed by the additive term, i.e., $c\gamma\epsilon$.

3. Any overloaded bin $i \in [n]$ with $i \leqslant \delta n$, satisfies $r_i = \frac{1-\epsilon}{n}$ and so $\Delta\overline{\Phi}_i = -\Phi_i \cdot \frac{\gamma\epsilon}{n}$. We call these the set $\mathcal{G}_+$ of *good overloaded bins*, as their dominant term decreases in expectation. The rest of the overloaded bins are the *bad overloaded bins* $\mathcal{B}_+$, as these satisfy $\Delta\overline{\Phi}_i = +\Phi_i \cdot \frac{\gamma\widetilde{\epsilon}}{n}$.

   Similarly, *good underloaded bins* $\mathcal{G}_-$ with $i > \delta n$, satisfy $\Delta\overline{\Psi}_i = -\Psi_i \cdot \frac{\gamma\widetilde{\epsilon}}{n}$ and *bad underloaded bins* $\mathcal{B}_-$ satisfy $\Delta\overline{\Psi}_i = +\Psi_i \cdot \frac{\gamma\epsilon}{n}$.

Set	Load	Index	r_i	Contribution $\Delta\overline{\Gamma}_i$
$\mathcal{G}_+$	$y_i \geqslant 0$	$i \leqslant \delta n$	$\frac{1-\epsilon}{n}$	$-\Phi_i \cdot \frac{\gamma\epsilon}{n} + \Psi_i \cdot \frac{\gamma\epsilon}{n}$
$\mathcal{B}_+$	$y_i \geqslant 0$	$i > \delta n$	$\frac{1+\widetilde{\epsilon}}{n}$	$+\Phi_i \cdot \frac{\gamma\widetilde{\epsilon}}{n} - \Psi_i \cdot \frac{\gamma\widetilde{\epsilon}}{n}$
$\mathcal{G}_-$	$y_i < 0$	$i > \delta n$	$\frac{1+\widetilde{\epsilon}}{n}$	$+\Phi_i \cdot \frac{\gamma\widetilde{\epsilon}}{n} - \Psi_i \cdot \frac{\gamma\widetilde{\epsilon}}{n}$
$\mathcal{B}_-$	$y_i < 0$	$i \leqslant \delta n$	$\frac{1-\epsilon}{n}$	$-\Phi_i \cdot \frac{\gamma\epsilon}{n} + \Psi_i \cdot \frac{\gamma\epsilon}{n}$

Table 3.1: The definition of the four sets of bins and the contribution term of each bin to $\Delta\overline{\Gamma}$. The *dominant term* is coloured. The sign of the dominant term determines if a bin is good (negative sign/decrease) or bad (positive sign/increase).

4. We can either have $\mathcal{B}_+ \neq \emptyset$ or $\mathcal{B}_- \neq \emptyset$ (see Fig. 3.2).

   The handling of one case is symmetric to the other due to the symmetric nature of $\Delta\overline{\Phi}$ and $\Delta\overline{\Psi}$ (with δ being replaced by $1-\delta$). So, from here on we only consider the case with $\mathcal{B}_+ \neq \emptyset$ (and $\mathcal{B}_- = \emptyset$).

5. **Case A.1:** When the number of bad overloaded bins is small (i.e., $1 \leqslant |\mathcal{B}_+| \leqslant \frac{n}{2} \cdot (1-\delta)$), the positive contribution of the bins in $\mathcal{B}_+$ is counteracted by the negative contribution of the bins in $\mathcal{G}_+$ (Fig. 3.3). We prove this by making the worst-case assumption that all bad bins are equal to $y_{\delta n}$. All underloaded bins are good, so on aggregate we get a decrease.

6. **Case A.2:** Consider the case when the number of bad overloaded bins is large $|\mathcal{B}_+| > \frac{n}{2} \cdot (1-\delta)$. The positive contribution of the first $\frac{n}{2} \cdot (1-\delta)$ of the bins $\mathcal{B}_+$, call them $\mathcal{B}_1$, is counteracted by the negative contribution of the bins in $\mathcal{G}_+$ as in Case A.1. The positive contribution of the remaining bad bins $\mathcal{B}_2$ is counteracted by a fraction of the negative contribution of the bins in $\mathcal{G}_-$. This is because the number of "holes" (empty ball slots in the underloaded bins) in the bins of $\mathcal{G}_-$ are significantly more than the number of balls in $\mathcal{B}_2$. Hence, again on aggregate we get a decrease (Fig. 3.4).

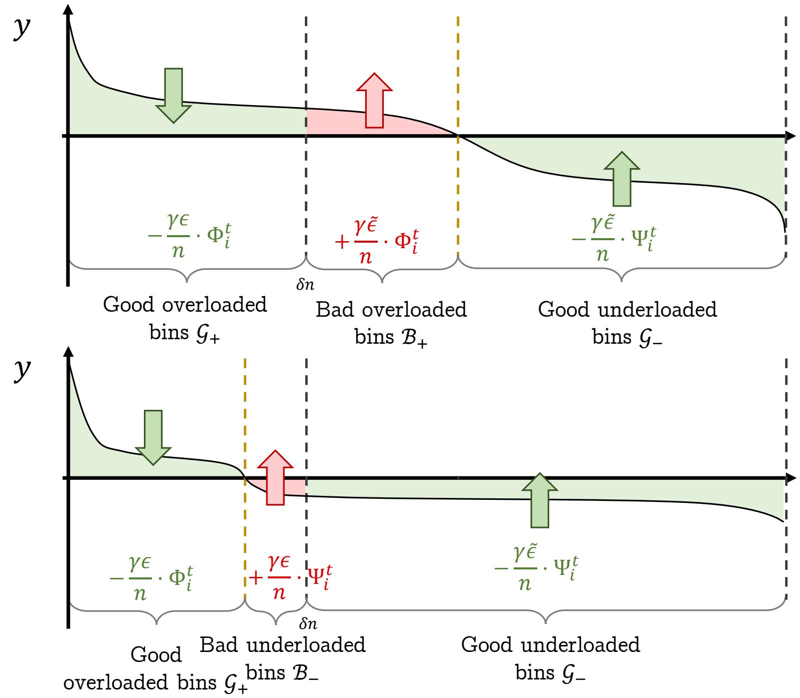


Figure 3.2: The two cases of bad bins ($\mathcal{B}_+ \neq \emptyset$ and $\mathcal{B}_- \neq \emptyset$ respectively) in a load vector and their dominant terms in $\Delta\overline{\Gamma}$ for each of the set of bins. The dominant terms that are decreasing are shown in green and the dominant terms that are increasing are shown in red.

We proceed with a simple claim for bounding the contributions of the non-dominant terms:

Claim 3.4. *Consider the probability vector r as defined in Eq.* (3.2)*. For any bin $i \in [n]$ with $y_i \geqslant 0$, we have that*

$$\Delta\overline{\Psi}_i \leqslant -\Psi_i \cdot \frac{\gamma\epsilon\delta}{4n} + \frac{2\gamma\epsilon}{n},$$

and for any bin $i \in [n]$ with $y_i < 0$, we have that

$$\Delta\overline{\Phi}_i \leqslant -\Phi_i \cdot \frac{\gamma\epsilon\delta}{4n} + \frac{2\gamma\widetilde{\epsilon}}{n}.$$

Proof. For any bin $i \in [n]$ with $y_i \geqslant 0$, we have that

$$\Delta\overline{\Psi}_i \leqslant \max\left\{+\Psi_i \cdot \frac{\gamma\epsilon}{n}, -\Psi_i \cdot \frac{\gamma\widetilde{\epsilon}}{n}\right\} = -\Psi_i \cdot \frac{\gamma\epsilon\delta}{4n} + \Psi_i \cdot \frac{\gamma}{n} \cdot \left(\frac{\epsilon\delta}{4} + \epsilon\right) \leqslant -\Psi_i \cdot \frac{\gamma\epsilon\delta}{4n} + \Psi_i \cdot \frac{2\gamma\epsilon}{n},$$

using that $\delta \leqslant 1$.

Similarly, for any bin $i \in [n]$ with $y_i < 0$, we have that

$$\Delta\overline{\Phi}_i \leqslant \max\left\{+\Phi_i \cdot \frac{\gamma\widetilde{\epsilon}}{n}, -\Phi_i \cdot \frac{\gamma\epsilon}{n}\right\} = -\Phi_i \cdot \frac{\gamma\epsilon\delta}{4n} + \Phi_i \cdot \frac{\gamma}{n} \cdot \left(\frac{\epsilon\delta}{4} + \widetilde{\epsilon}\right) \leqslant -\Phi_i \cdot \frac{\gamma\epsilon\delta}{4n} + \Phi_i \cdot \frac{2\gamma\widetilde{\epsilon}}{n},$$

using that $\epsilon\delta \leqslant \frac{\epsilon\delta}{1-\delta} = \widetilde{\epsilon}$. □

We now turn to the proof of Lemma 3.3.

Proof of Lemma 3.3. Fix a labelling of the bins so that they are sorted non-increasingly according to their load in x. Let p be the probability vector satisfying condition $\mathcal{C}_1$ for some $\epsilon \in (0,1)$ and $\delta \in \{1/n, \ldots, 1\}$. Recall that the probability vector r was defined as,

$$r_i := \begin{cases} \frac{1-\epsilon}{n} & \text{if } i \leqslant \delta n, \\ \frac{1+\widetilde{\epsilon}}{n} & \text{otherwise,} \end{cases} \tag{3.3}$$

where $\widetilde{\epsilon} := \epsilon \cdot \frac{\delta}{1-\delta}$. Thanks to the definition of $\widetilde{\epsilon}$, it is clear that r is also a probability vector. Further, for any $1 \leqslant k \leqslant \delta n$, due to condition $\mathcal{C}_1$,

$$\sum_{i=1}^{k} p_i \leqslant (1-\epsilon) \cdot \frac{k}{n} = \sum_{i=1}^{k} r_i,$$

and any $\delta \cdot n + 1 \leqslant k \leqslant n$,

$$\sum_{i=k}^{n} p_i \geqslant \left(1 + \epsilon \cdot \frac{\delta}{1-\delta}\right) \cdot \frac{n-k+1}{n} = \sum_{i=k}^{n} r_i.$$

This implies that p is majorised by r. Since Φ_i (and Ψ_i) are non-increasing (and non-decreasing) in $i \in [n]$, using Lemma B.2, the terms

$$\Delta\overline{\Phi} = \sum_{i=1}^{n} \Phi_i \cdot \left(p_i - \frac{1}{n}\right) \cdot \gamma, \quad \text{and} \quad \Delta\overline{\Psi} = \sum_{i=1}^{n} \Psi_i \cdot \left(\frac{1}{n} - p_i\right) \cdot \gamma,$$

are at least as large for r than for p. Hence, from now on, we will be working with $p = r$.

Recall that we partition overloaded bins i with $y_i \geqslant 0$ into *good overloaded bins* $\mathcal{G}_+$ with $p_i = \frac{1-\epsilon}{n}$ and into *bad overloaded bins* $\mathcal{B}_+$ with $p_i = \frac{1+\widetilde{\epsilon}}{n}$ (see Table 3.1). These are called good bins, because any bin $i \in \mathcal{G}_+$ satisfies $\Delta\overline{\Phi}_i = -\Phi_i \cdot \frac{\gamma\epsilon}{n}$ and since $\Psi_i \leqslant 1$ for overloaded bins, this implies overall a drop in expectation for Γ_i.

In Case A, we consider the case where $\mathcal{B}_+ \neq \emptyset$. Further, we partition $\mathcal{B}_+$ into $\mathcal{B}_1 := \mathcal{B}_+ \cap \{i \in [n] : i \leqslant \frac{n}{2} \cdot (1+\delta)\}$ and $\mathcal{B}_2 := \mathcal{B}_+ \setminus \mathcal{B}_1$. In Case A.1 we handle the case where $\mathcal{B}_2 = \emptyset$ and in Case A.2 the case where $\mathcal{B}_2 \neq \emptyset$. Finally, in Case B we handle the case where $\mathcal{B}_+ \neq \emptyset$ by a symmetry argument.

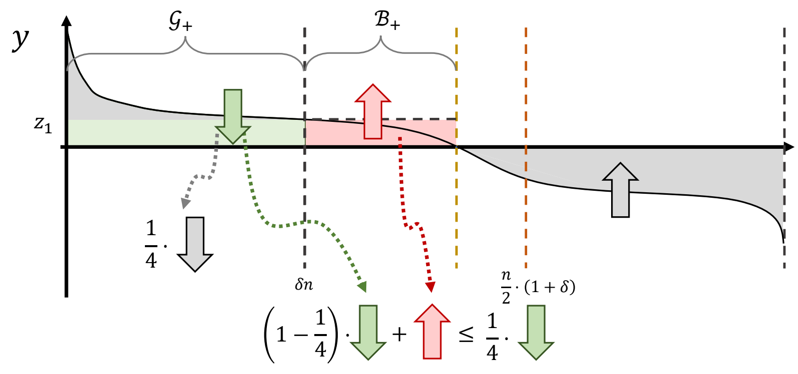


Figure 3.3: Case A.1: The dominant (positive) contribution of bins in $\mathcal{B}_+$ (shown in red) is counteracted by a fraction of the (negative) contribution term of the good bins $\mathcal{G}_+$ (shown in green). In grey, the dominant decrease terms for $\mathcal{G}_+$ and $\mathcal{G}_-$ (which do not contribute to counteracting the increase).

Case A.1 $[1 \leqslant |\mathcal{B}_+| \leqslant \frac{n}{2} \cdot (1-\delta)]$**:** Intuitively, in this case there are not many bad bins in $\mathcal{B}_+$, so their (positive) contribution is counteracted by the (negative) contribution of good bins in $\mathcal{G}_+$ (see Fig. 3.3).

To formalise this, let $z_1 := y_{\delta n}$ (by assumption, $|\mathcal{B}_+| \geqslant 1$ and so $z_1 \geqslant 0$). Then, $y_i \geqslant z_1$ for any bin $i \in \mathcal{G}_+$, and $y_i \leqslant z_1$, for any $i \in \mathcal{B}_+$. With some foresight, we use $\mathcal{B}_1$ instead of $\mathcal{B}_+$, since in this case $\mathcal{B}_1 = \mathcal{B}_+$ and it will also allow us to use Eq. (3.6) in Case A.2. Hence,

$$\begin{aligned}
\sum_{i \in \mathcal{B}_1} \Delta\overline{\Phi}_i &= \sum_{i \in \mathcal{B}_1} \Phi_i \cdot \frac{\gamma\widetilde{\epsilon}}{n} = -\sum_{i \in \mathcal{B}_1} \Phi_i \cdot \frac{\gamma\epsilon\delta}{4n} + \sum_{i \in \mathcal{B}_1} \Phi_i \cdot \frac{\gamma}{n} \cdot \left(\widetilde{\epsilon} + \frac{\epsilon\delta}{4}\right) \\
&\overset{(a)}{\leqslant} -\sum_{i \in \mathcal{B}_1} \Phi_i \cdot \frac{\gamma\epsilon\delta}{4n} + \sum_{i \in \mathcal{B}_1} \Phi_i \cdot \frac{3\gamma\widetilde{\epsilon}}{2n} \\
&\overset{(b)}{\leqslant} -\sum_{i \in \mathcal{B}_1} \Phi_i \cdot \frac{\gamma\epsilon\delta}{4n} + \frac{n}{2} \cdot (1-\delta) \cdot e^{\gamma z_1} \cdot \frac{3\gamma\widetilde{\epsilon}}{2n} \\
&\overset{(c)}{=} -\sum_{i \in \mathcal{B}_1} \Phi_i \cdot \frac{\gamma\epsilon\delta}{4n} + e^{\gamma z_1} \cdot \frac{3\gamma\epsilon\delta}{4},
\end{aligned} \tag{3.4}$$

using in (a) that $\epsilon\delta \leqslant \widetilde{\epsilon}$ and in (b) that $y_i \leqslant z_1$ for $i \in \mathcal{B}_1$ (and so $\Phi_i \leqslant e^{\gamma z_1}$) and $|\mathcal{B}_1| \leqslant \frac{n}{2} \cdot (1-\delta)$, and in (c) that $\widetilde{\epsilon} = \frac{\epsilon\delta}{1-\delta}$. For bins in $\mathcal{G}_+$,

$$\begin{aligned}
\sum_{i \in \mathcal{G}_+} \Delta\overline{\Phi}_i &= -\sum_{i \in \mathcal{G}_+} \Phi_i \cdot \frac{\gamma\epsilon}{n} = -\sum_{i \in \mathcal{G}_+} \Phi_i \cdot \frac{\gamma\epsilon}{4n} - \sum_{i \in \mathcal{G}_+} \Phi_i \cdot \frac{3\gamma\epsilon}{4n} \\
&\overset{(a)}{\leqslant} -\sum_{i \in \mathcal{G}_+} \Phi_i \cdot \frac{\gamma\epsilon}{4n} - \sum_{i \in \mathcal{G}_+} e^{\gamma z_1} \cdot \frac{3\gamma\epsilon}{4n} \overset{(b)}{=} -\sum_{i \in \mathcal{G}_+} \Phi_i \cdot \frac{\gamma\epsilon}{4n} - e^{\gamma z_1} \cdot \frac{3\gamma\epsilon\delta}{4},
\end{aligned} \tag{3.5}$$

using in (a) that $y_i \geqslant z_1$ for any $i \in \mathcal{G}_+$ and in (b) that $|\mathcal{G}_+| = \delta n$, since $|\mathcal{B}_1| \geqslant 1$.

Hence, combining Eq. (3.4) and Eq. (3.5), the contribution of overloaded bins to Φ is given by

$$\sum_{i \in \mathcal{G}_+ \cup \mathcal{B}_1} \Delta\overline{\Phi}_i \leqslant -\sum_{i \in \mathcal{G}_+} \Phi_i \cdot \frac{\gamma\epsilon}{4n} - \sum_{i \in \mathcal{B}_1} \Phi_i \cdot \frac{\gamma\epsilon\delta}{4n} \leqslant -\sum_{i \in \mathcal{G}_+ \cup \mathcal{B}_1} \Phi_i \cdot \frac{\gamma\epsilon\delta}{4n}. \tag{3.6}$$

Therefore, aggregating the contributions to $\Delta\overline{\Gamma}$ as described above, we get that

$$\begin{aligned}
\Delta\overline{\Gamma} &= \sum_{i \in \mathcal{G}_+} \Delta\overline{\Gamma}_i + \sum_{i \in \mathcal{B}_1} \Delta\overline{\Gamma}_i + \sum_{i \in \mathcal{G}_-} \Delta\overline{\Gamma}_i \\
&= \sum_{i \in \mathcal{G}_+ \cup \mathcal{B}_1} \Delta\overline{\Phi}_i + \sum_{i \in \mathcal{G}_-} \Delta\overline{\Psi}_i + \sum_{i \in \mathcal{G}_+ \cup \mathcal{B}_1} \Delta\overline{\Psi}_i + \sum_{i \in \mathcal{G}_-} \Delta\overline{\Phi}_i \\
&\overset{(3.6)}{\leqslant} -\sum_{i \in \mathcal{G}_+ \cup \mathcal{B}_1} \Phi_i \cdot \frac{\gamma\epsilon\delta}{4n} - \sum_{i \in \mathcal{G}_-} \Psi_i \cdot \frac{\gamma\widetilde{\epsilon}}{n} + \sum_{i \in \mathcal{G}_+ \cup \mathcal{B}_1} \Delta\overline{\Psi}_i + \sum_{i \in \mathcal{G}_-} \Delta\overline{\Phi}_i \\
&\overset{(a)}{\leqslant} -\sum_{i \in \mathcal{G}_+ \cup \mathcal{B}_1} \Phi_i \cdot \frac{\gamma\epsilon\delta}{4n} - \sum_{i \in \mathcal{G}_-} \Psi_i \cdot \frac{\gamma\epsilon\delta}{4n} - \sum_{i \in \mathcal{G}_+ \cup \mathcal{B}_1} \Psi_i \cdot \frac{\gamma\epsilon\delta}{4n} - \sum_{i \in \mathcal{G}_-} \Phi_i \cdot \frac{\gamma\epsilon\delta}{4n} + \sum_{i=1}^{n} \frac{2\gamma}{n} \cdot \max\{\epsilon, \widetilde{\epsilon}\} \\
&\overset{(b)}{=} -\sum_{i=1}^{n} \Gamma_i \cdot \frac{\gamma\epsilon\delta}{4n} + 2\gamma \cdot \max\{\epsilon, \widetilde{\epsilon}\} = -\Gamma \cdot \frac{\gamma\epsilon\delta}{4n} + 2\gamma \cdot \max\{\epsilon, \widetilde{\epsilon}\},
\end{aligned}$$

using in (a) that $\epsilon\delta \leqslant \widetilde{\epsilon}$ and Claim 3.4 to bound the contributions of the non-dominant terms and in (b) that $\Gamma_i = \Phi_i + \Psi_i$ for any bin $i \in [n]$.

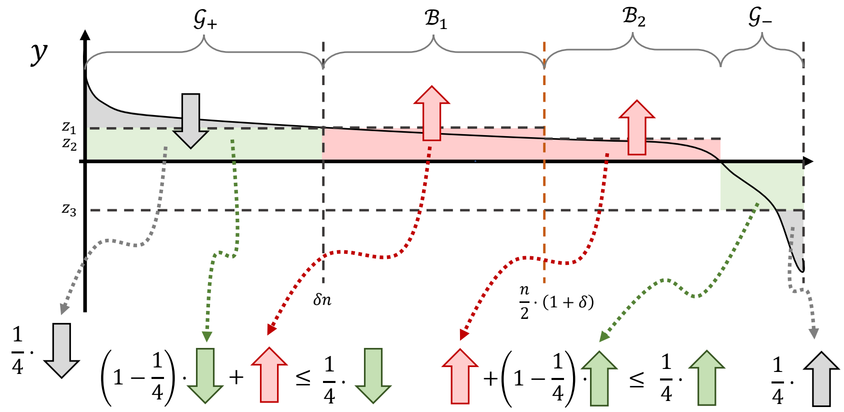


Figure 3.4: Case A.2: The dominant (positive) contribution of the bins in $\mathcal{B}_1$ (shown in red) is counteracted by a fraction of the dominant (negative) contribution of the bins in $\mathcal{G}_+$ (shown in green) as in Case A.1. The dominant (negative) contribution of the bins in $\mathcal{B}_2$ is counteracted by a fraction of the dominant (negative) contribution of the bins in $\mathcal{G}_-$ (shown in green), when z_2 is sufficiently large. In grey, the dominant decrease terms for $\mathcal{G}_+$ and $\mathcal{G}_-$ (which do not contribute to counteracting the increase).

Case A.2 [$|\mathcal{B}_+| > \frac{n}{2}\cdot(1-\delta)$]**:** Recall that we partitioned the bins $\mathcal{B}_+$ into $\mathcal{B}_1 := \mathcal{B}_+ \cap \{i \in [n] : i \leqslant \frac{n}{2}\cdot(1+\delta)\}$ and $\mathcal{B}_2 := \mathcal{B}_+ \setminus \mathcal{B}_1$. We will counteract the positive contribution $\Delta\overline{\Phi}_i$ for bins $i \in \mathcal{B}_1$ by the negative contribution of the bins in $\mathcal{G}_+$ as in Eq. (3.6) in Case A.1. For that of bins in $\mathcal{B}_2$ we will consider two cases based on $z_2 := y_{\frac{n}{2}\cdot(1+\delta)} > 0$, the load of the heaviest bin in $\mathcal{B}_2$. Similarly to Claim 3.4, we obtain a bound for the dominant contribution of the bins in $\mathcal{B}_2$

$$\begin{aligned}
\sum_{i\in\mathcal{B}_2}\Delta\overline{\Phi}_i &= \sum_{i\in\mathcal{B}_2}\Phi_i\cdot\frac{\gamma\widetilde{\epsilon}}{n} = -\sum_{i\in\mathcal{B}_2}\Phi_i\cdot\frac{\gamma\epsilon\delta}{4n} + \sum_{i\in\mathcal{B}_2}\Phi_i\cdot\frac{\gamma}{n}\cdot\left(\frac{\epsilon\delta}{4}+\widetilde{\epsilon}\right)\\
&\leqslant -\sum_{i\in\mathcal{B}_2}\Phi_i\cdot\frac{\gamma\epsilon\delta}{4n} + \sum_{i\in\mathcal{B}_2}\Phi_i\cdot\frac{2\gamma\widetilde{\epsilon}}{n}\\
&\overset{(a)}{\leqslant} -\sum_{i\in\mathcal{B}_2}\Phi_i\cdot\frac{\gamma\epsilon\delta}{4n} + |\mathcal{B}_2|\cdot e^{\gamma z_2}\cdot\frac{2\gamma\widetilde{\epsilon}}{n}\\
&\overset{(b)}{\leqslant} -\sum_{i\in\mathcal{B}_2}\Phi_i\cdot\frac{\gamma\epsilon\delta}{4n} + e^{\gamma z_2}\cdot(\gamma\epsilon\delta),
\end{aligned} \tag{3.7}$$

using in (a) that $y_i \leqslant z_2$ for $i \in \mathcal{B}_2$ and in (b) that $|\mathcal{B}_2| \leqslant \frac{n}{2}\cdot(1-\delta)$ and $\widetilde{\epsilon} = \frac{\epsilon\delta}{1-\delta}$.

Case A.2.1 [$z_2 \leqslant \frac{1}{\gamma}\cdot\frac{1-\delta}{2\delta}\cdot\log(8/3)$]**:** In this case, the loads of the bins in $\mathcal{B}_2$ are small enough for their contribution to be counteracted by the additive term. More precisely, we get that

$$\sum_{i\in\mathcal{B}_2}\Delta\overline{\Phi}_i \overset{(3.7)}{\leqslant} -\sum_{i\in\mathcal{B}_2}\Phi_i\cdot\frac{\gamma\epsilon\delta}{4n} + e^{\gamma z_2}\cdot(\gamma\epsilon\delta) \leqslant -\sum_{i\in\mathcal{B}_2}\Phi_i\cdot\frac{\gamma\epsilon\delta}{4n} + e^{\frac{1-\delta}{2\delta}\cdot\log(8/3)}\cdot(\gamma\epsilon\delta). \tag{3.8}$$

Hence, we can now aggregate the contributions as follows

$$\begin{aligned}
\Delta\overline{\Gamma} &= \sum_{i\in\mathcal{G}_+}\Delta\overline{\Gamma}_i + \sum_{i\in\mathcal{B}_1}\Delta\overline{\Gamma}_i + \sum_{i\in\mathcal{B}_2}\Delta\overline{\Gamma}_i + \sum_{i\in\mathcal{G}_-}\Delta\overline{\Gamma}_i\\
&= \sum_{i\in\mathcal{G}_+\cup\mathcal{B}_1}\Delta\overline{\Phi}_i + \sum_{i\in\mathcal{B}_2}\Delta\overline{\Phi}_i + \sum_{i\in\mathcal{G}_-}\Delta\overline{\Psi}_i + \sum_{i\in\mathcal{G}_+\cup\mathcal{B}_+}\Delta\overline{\Psi}_i + \sum_{i\in\mathcal{G}_-}\Delta\overline{\Phi}_i
\end{aligned}$$

$$\overset{(a)}{\leqslant} - \sum_{i \in \mathcal{G}_+ \cup \mathcal{B}_1} \Phi_i \cdot \frac{\gamma\epsilon\delta}{4n} - \sum_{i \in \mathcal{B}_2} \Phi_i \cdot \frac{\gamma\epsilon\delta}{4n} + e^{\frac{1-\delta}{2\delta} \cdot \log(8/3)} \cdot (\gamma\epsilon\delta) - \sum_{i \in \mathcal{G}_-} \Psi_i \cdot \frac{\gamma\epsilon\delta}{4n}$$
$$- \sum_{i \in \mathcal{G}_+ \cup \mathcal{B}_+} \Psi_i \cdot \frac{\gamma\epsilon\delta}{4n} - \sum_{i \in \mathcal{G}_-} \Phi_i \cdot \frac{\gamma\epsilon\delta}{4n} + \sum_{i=1}^{n} \frac{2\gamma}{n} \cdot \max\{\epsilon, \widetilde{\epsilon}\}$$
$$\leqslant - \sum_{i=1}^{n} \Gamma_i \cdot \frac{\gamma\epsilon\delta}{4n} + 4\gamma \cdot \max\left\{\epsilon, \widetilde{\epsilon}, \epsilon\delta \cdot e^{\frac{1-\delta}{2\delta} \cdot \log(8/3)}\right\},$$

using in (a): (i) Eq. (3.6) for bounding the contribution of bins in $\mathcal{B}_1$, (ii) the Eq. (3.8) for bounding the contribution of bins in $\mathcal{B}_2$ and (iii) the Claim 3.4 for bounding the contributions of the non-dominant terms.

Case A.2.2 [$z_2 > \frac{1}{\gamma} \cdot \frac{1-\delta}{2\delta} \cdot \log(8/3)$]: In this case, z_2 being large means that there are substantially more holes (ball slots below the average line) in the underloaded bins than balls in the overloaded bins of $\mathcal{B}_1$. Hence, as we will prove below the negative contribution $\Delta\overline{\Psi}$ for bins in $\mathcal{G}_-$ counteracts the positive contribution of $\Delta\overline{\Phi}$ for $\mathcal{B}_1$ (Fig. 3.4).

Next note that because Ψ_i is non-decreasing in $i \in [n]$, the term $\sum_{i \in \mathcal{G}_-} \Psi_i$ is minimised when all underloaded bins are equal to the same load $-z_3 < 0$, i.e., $\sum_{i \in \mathcal{G}_-} \Psi_i \geqslant |\mathcal{G}_-| \cdot e^{\gamma z_3}$. Further, note that $z_3 \geqslant \frac{z_2 \cdot (|\mathcal{G}_+| + |\mathcal{B}_+|)}{|\mathcal{G}_-|} \geqslant \frac{z_2 \cdot \frac{n}{2} \cdot (1+\delta)}{|\mathcal{G}_-|}$ by the assumption $|\mathcal{B}_+| > \frac{n}{2} \cdot (1-\delta)$ and $|\mathcal{G}_+| = \delta n$, and therefore,

$$\sum_{i \in \mathcal{G}_-} \Psi_i \geqslant |\mathcal{G}_-| \cdot e^{\gamma \cdot \frac{z_2 \cdot \frac{n}{2} \cdot (1+\delta)}{|\mathcal{G}_-|}} =: g(|\mathcal{G}_-|),$$

where we seek to lower bound the function $g : [1, n] \to \mathbb{R}$. To this end, we will first upper bound $|\mathcal{G}_-|$, using the assumption for Case A.2,

$$|\mathcal{G}_-| = n - |\mathcal{G}_+| - |\mathcal{B}_+| \leqslant n - n\delta - \frac{n}{2} \cdot (1-\delta) = \frac{n}{2} \cdot (1-\delta).$$

Further, $\frac{n}{2} \cdot (1-\delta) \leqslant \gamma \cdot z_2 \cdot \frac{n}{2} \cdot (1+\delta) =: M$, by definition of z_2 (and that $2\delta \leqslant 1+\delta$), and so we also have $|\mathcal{G}_-| \leqslant M$. By Lemma B.3, the function $f(x) = x \cdot e^{k/x}$ is decreasing for $0 < x \leqslant k$, and so g is decreasing for $1 < |\mathcal{G}_-| \leqslant \frac{n}{2} \cdot (1-\delta) \leqslant M$. Hence, $g(|\mathcal{G}_-|)$ is minimised by $|\mathcal{G}_-| = \frac{n}{2} \cdot (1-\delta)$. Therefore,

$$\sum_{i \in \mathcal{G}_-} \Delta\overline{\Psi}_i = - \sum_{i \in \mathcal{G}_-} \Psi_i \cdot \frac{\gamma\widetilde{\epsilon}}{n} \geqslant - \min_{|\mathcal{G}_-| \in [1, \frac{n}{2} \cdot (1-\delta)]} |\mathcal{G}_-| \cdot e^{\frac{M}{|\mathcal{G}_-|}} \cdot \frac{\gamma\widetilde{\epsilon}}{n} = -\frac{n}{2} \cdot (1-\delta) \cdot e^{\frac{M}{\frac{n}{2} \cdot (1-\delta)}} \cdot \frac{\gamma\widetilde{\epsilon}}{n}. \tag{3.9}$$

We can lower bound the exponent of the last term as follows,

$$\frac{M}{\frac{n}{2} \cdot (1-\delta)} = \frac{\gamma z_2 \cdot (1+\delta)}{1-\delta} = \gamma z_2 + \gamma z_2 \cdot \frac{2\delta}{1-\delta} \geqslant \gamma z_2 + \log(8/3),$$

using the assumption that $z_2 > \frac{1}{\gamma} \cdot \frac{1-\delta}{2\delta} \cdot \log(8/3)$.

Now we will split the contributions of the bins in $\mathcal{G}_-$,

$$\sum_{i \in \mathcal{G}_-} \Delta\overline{\Psi}_i = - \sum_{i \in \mathcal{G}_-} \Psi_i \cdot \frac{\gamma\widetilde{\epsilon}}{n} = - \sum_{i \in \mathcal{G}_-} \Psi_i \cdot \frac{\gamma\widetilde{\epsilon}}{4n} - \sum_{i \in \mathcal{G}_-} \Psi_i \cdot \frac{3\gamma\widetilde{\epsilon}}{4n}$$
$$\overset{(3.9)}{\leqslant} - \sum_{i \in \mathcal{G}_-} \Psi_i \cdot \frac{\gamma\widetilde{\epsilon}}{4n} - \frac{n}{2} \cdot (1-\delta) \cdot e^{\gamma z_2 + \log(8/3)} \cdot \frac{3\gamma\widetilde{\epsilon}}{4n}$$

$$= -\sum_{i \in \mathcal{G}_-} \Psi_i \cdot \frac{\gamma \widetilde{\epsilon}}{4n} - e^{\gamma z_2} \cdot (\gamma \epsilon \delta), \tag{3.10}$$

using in the last equality that $\widetilde{\epsilon} = \frac{\epsilon\delta}{1-\delta}$.

We will now show that the dominant increase for bins in $\mathcal{B}_2$ is counteracted by a fraction of the dominant decrease of those in $\mathcal{G}_-$. Combining Eq. (3.7) and Eq. (3.10)

$$\begin{aligned}
\sum_{i \in \mathcal{B}_2} \Delta\overline{\Phi}_i + \sum_{i \in \mathcal{G}_-} \Delta\overline{\Psi}_i &\leqslant -\sum_{i \in \mathcal{B}_2} \Phi_i \cdot \frac{\gamma\epsilon\delta}{4n} + e^{\gamma z_2} \cdot (\gamma\epsilon\delta) - \sum_{i \in \mathcal{G}_-} \Psi_i \cdot \frac{\gamma\widetilde{\epsilon}}{4n} - \sum_{i \in \mathcal{G}_-} \Psi_i \cdot \frac{3\gamma\widetilde{\epsilon}}{4n} \\
&\overset{(3.10)}{\leqslant} -\sum_{i \in \mathcal{B}_2} \Phi_i \cdot \frac{\gamma\epsilon\delta}{4n} - \sum_{i \in \mathcal{G}_-} \Psi_i \cdot \frac{\gamma\widetilde{\epsilon}}{4n} \leqslant -\sum_{i \in \mathcal{B}_2} \Phi_i \cdot \frac{\gamma\epsilon\delta}{4n} - \sum_{i \in \mathcal{G}_-} \Psi_i \cdot \frac{\gamma\epsilon\delta}{4n}.
\end{aligned} \tag{3.11}$$

Finally, overall the contributions are given by

$$\begin{aligned}
\Delta\overline{\Gamma} &= \sum_{i \in \mathcal{G}_+} \Delta\overline{\Gamma}_i + \sum_{i \in \mathcal{B}_1} \Delta\overline{\Gamma}_i + \sum_{i \in \mathcal{B}_2} \Delta\overline{\Gamma}_i + \sum_{i \in \mathcal{G}_-} \Delta\overline{\Gamma}_i \\
&= \sum_{i \in \mathcal{G}_+ \cup \mathcal{B}_1} \Delta\overline{\Phi}_i + \left(\sum_{i \in \mathcal{B}_2} \Delta\overline{\Phi}_i + \sum_{i \in \mathcal{G}_-} \Delta\overline{\Psi}_i \right) + \sum_{i \in \mathcal{G}_+ \cup \mathcal{B}_+} \Delta\overline{\Psi}_i + \sum_{i \in \mathcal{G}_-} \Delta\overline{\Phi}_i \\
&\overset{(a)}{\leqslant} - \sum_{i \in \mathcal{G}_+ \cup \mathcal{B}_+} \Phi_i \cdot \frac{\gamma\epsilon\delta}{4n} - \sum_{i \in \mathcal{B}_2} \Phi_i \cdot \frac{\gamma\epsilon\delta}{4n} - \sum_{i \in \mathcal{G}_-} \Psi_i \cdot \frac{\gamma\epsilon\delta}{4n} - \sum_{i \in \mathcal{G}_+ \cup \mathcal{B}_+} \Psi_i \cdot \frac{\gamma\epsilon\delta}{4n} - \sum_{i \in \mathcal{G}_-} \Phi_i \cdot \frac{\gamma\epsilon\delta}{4n} \\
&\quad + \sum_{i=1}^{n} \frac{2\gamma}{n} \cdot \max\{\epsilon, \widetilde{\epsilon}\} \\
&= -\Gamma \cdot \frac{\gamma\epsilon\delta}{4n} + 2\gamma \cdot \max\{\epsilon, \widetilde{\epsilon}\},
\end{aligned} \tag{3.12}$$

using in (a) that (i) Eq. (3.6) for bounding the contribution of the bins in $\mathcal{B}_1$, (ii) the Eq. (3.11) for bounding the contribution of the bins in $\mathcal{B}_2 \cup \mathcal{G}_-$ and (iii) Claim 3.4 for bounding the non-dominant terms.

Case B [$\mathcal{B}_- \neq \emptyset$]: This case is symmetric to Case A, by interchanging Φ with Ψ, δ with $1-\delta$, $\epsilon\delta$ with $\widetilde{\epsilon}(1-\delta)$, and negating and sorting the normalised load vector. In particular, the three sub-cases are:

- **Case B.1** $[1 \leqslant |\mathcal{B}_-| \leqslant \frac{n}{2} \cdot \delta]$
- **Case B.2.1** $[|\mathcal{B}_-| > \frac{n}{2} \cdot \delta, z_2' \leqslant \frac{1}{\gamma} \cdot \frac{\delta}{2\cdot(1-\delta)} \cdot \log(8/3)]$ where $z_2' := y_{\frac{n}{2}\cdot\delta}$
- **Case B.2.2** $[|\mathcal{B}_-| > \frac{n}{2} \cdot \delta, z_2' > \frac{1}{\gamma} \cdot \frac{\delta}{2\cdot(1-\delta)} \cdot \log(8/3)]$

Combining the Case A and Case B, we get that

$$\Delta\overline{\Gamma} \leqslant -\Gamma \cdot \frac{\gamma\epsilon\delta}{8n} + c\gamma\epsilon,$$

where $c := 4 \cdot \max\left\{\frac{1-\delta}{\delta}, e^{\frac{1-\delta}{2\delta}\cdot\log(8/3)} \cdot \frac{\delta}{1-\delta}, \delta \cdot e^{\frac{\delta}{2\cdot(1-\delta)}\cdot\log(8/3)}\right\}$, recalling that $\widetilde{\epsilon} := \epsilon \cdot \frac{\delta}{1-\delta}$. □

By scaling the quantities $\Delta\overline{\Phi}$ and $\Delta\overline{\Psi}$ in Lemma 3.3 by some $R > 0$ (usually the number of steps in each round, e.g., $R := b$ for b-BATCHED) and selecting a sufficiently small smoothing parameter γ, we obtain the main theorem.

Theorem 3.2 (Restated). *Consider any folding of a* $\mathcal{P} = \textsc{Sequential}(q^t)$ *process and a probability vector* p^t *satisfying condition* $\mathcal{C}_1$ *for some constant* $\delta \in (0,1)$ *and some* $\epsilon \in (0,1)$ *at every round* $t \geqslant 0$*. Further assume that there exist* $K > 0$, $\gamma \in \left(0, \min\left\{1, \frac{\epsilon\delta}{8K}\right\}\right]$ *and* $R > 0$*, such that for any round* $t \geqslant 0$*, process* $\mathcal{P}$ *satisfies for potentials* $\Phi := \Phi(\gamma)$ *and* $\Psi := \Psi(\gamma)$ *that for bins sorted in non-increasing order of their loads,*

$$\sum_{i=1}^{n} \mathbf{E}\left[\, \Delta\Phi_i^{t+1} \,\middle|\, \mathfrak{F}^t \,\right] \leqslant \sum_{i=1}^{n} \Phi_i^t \cdot \left(\left(p_i^t - \frac{1}{n} \right) \cdot R \cdot \gamma + K \cdot R \cdot \frac{\gamma^2}{n} \right),$$

and,

$$\sum_{i=1}^{n} \mathbf{E}\left[\, \Delta\Psi_i^{t+1} \,\middle|\, \mathfrak{F}^t \,\right] \leqslant \sum_{i=1}^{n} \Psi_i^t \cdot \left(\left(\frac{1}{n} - p_i^t \right) \cdot R \cdot \gamma + K \cdot R \cdot \frac{\gamma^2}{n} \right).$$

Then, there exists a constant $c := c(\delta) > 0$*, such that for* $\Gamma := \Gamma(\gamma)$ *and any round* $t \geqslant 0$*,*

$$\mathbf{E}\left[\, \Delta\Gamma^{t+1} \,\middle|\, \mathfrak{F}^t \,\right] \leqslant -\Gamma^t \cdot R \cdot \frac{\gamma\epsilon\delta}{8n} + R \cdot c\gamma\epsilon,$$

and

$$\mathbf{E}\left[\, \Gamma^t \,\right] \leqslant \frac{8c}{\delta} \cdot n.$$

Proof. Consider a labelling of the bins so that they are sorted in non-increasing order of the loads at step t. Applying Lemma 3.3 for the current load vector x^t and the quantities

$$\Delta\overline{\Phi} := \sum_{i=1}^{n} \Phi_i^t \cdot \left(p_i - \frac{1}{n} \right) \cdot \gamma \quad \text{and} \quad \Delta\overline{\Psi} := \sum_{i=1}^{n} \Psi_i^t \cdot \left(\frac{1}{n} - p_i \right) \cdot \gamma,$$

we get that

$$\Delta\overline{\Phi} + \Delta\overline{\Psi} \leqslant -\frac{\gamma\epsilon\delta}{4n} \cdot \Gamma^t + c\gamma\epsilon. \tag{3.13}$$

By the assumptions,

$$\mathbf{E}\left[\, \Delta\Gamma^{t+1} \,\middle|\, \mathfrak{F}^t \,\right] = \mathbf{E}\left[\, \Delta\Phi^{t+1} \,\middle|\, \mathfrak{F}^t \,\right] + \mathbf{E}\left[\, \Delta\Psi^{t+1} \,\middle|\, \mathfrak{F}^t \,\right] \leqslant R \cdot \left(\Delta\overline{\Phi} + \Delta\overline{\Psi} + K \cdot \frac{\gamma^2}{n} \cdot \Gamma^t \right). \tag{3.14}$$

Hence, combining Eq. (3.13) and Eq. (3.14), we get

$$\mathbf{E}\left[\, \Delta\Gamma^{t+1} \middle| \mathfrak{F}^t \,\right] \leqslant R \cdot \left(-\frac{\gamma\epsilon\delta}{4n} \cdot \Gamma^t + c\gamma\epsilon + K \cdot \frac{\gamma^2}{n} \cdot \Gamma^t \right) \leqslant -R \cdot \frac{\gamma\epsilon\delta}{8n} \cdot \Gamma^t + R \cdot c\gamma\epsilon,$$

using that $\gamma \leqslant \frac{\epsilon\delta}{8K}$.

Finally, by Lemma B.1 (ii), the second statement follows. □

3.2 Applications

3.2.1 The non-batched setting

We start by verifying the preconditions of Theorem 3.2 for the non-batched setting.

Lemma 3.5. *Consider any* SEQUENTIAL(q^t) *process with probability allocation vector q^t satisfying condition $\mathcal{C}_2$ for $C > 1$ at every step $t \geqslant 0$. Further, consider the* WEIGHTED *setting with weights from a* FINITE-MGF(ζ, S) *distribution with $S \geqslant 1$. Then, for the potentials $\Phi := \Phi(\gamma)$ and $\Psi := \Psi(\gamma)$ with any smoothing parameter $\gamma \in \big(0, \min\{\zeta/2, 1\}\big]$, for any step $t \geqslant 0$,*

$$\mathbf{E}\big[\, \Delta\Phi^{t+1} \,\big|\, \mathfrak{F}^t \,\big] \leqslant \sum_{i=1}^{n} \Phi_i^t \cdot \left(\left(q_i^t - \frac{1}{n} \right) \cdot \gamma + 2CS \cdot \frac{\gamma^2}{n} \right),$$

and

$$\mathbf{E}\big[\, \Delta\Psi^{t+1} \,\big|\, \mathfrak{F}^t \,\big] \leqslant \sum_{i=1}^{n} \Psi_i^t \cdot \left(\left(\frac{1}{n} - q_i^t \right) \cdot \gamma + 2CS \cdot \frac{\gamma^2}{n} \right).$$

Proof. Consider an arbitrary bin $i \in [n]$. Then, for the overload potential we have that

$$\begin{aligned}
\mathbf{E}\big[\, \Phi_i^{t+1} \,\big|\, \mathfrak{F}^t \,\big] &= \Phi_i^t \cdot \mathbf{E}\big[\, e^{\gamma W(1-1/n)} \,\big|\, \mathfrak{F}^t \,\big] \cdot q_i^t + \Phi_i^t \cdot \mathbf{E}\big[\, e^{-\gamma W/n} \,\big|\, \mathfrak{F}^t \,\big] \cdot (1 - q_i^t) \\
&\overset{(a)}{\leqslant} \Phi_i^t \cdot \left(1 + \gamma \cdot \left(1 - \frac{1}{n} \right) + S\gamma^2 \right) \cdot q_i^t + \Phi_i^t \cdot \left(1 - \frac{\gamma}{n} + S\frac{\gamma^2}{n^2} \right) \cdot (1 - q_i^t) \\
&= \Phi_i^t \cdot \left(1 + \gamma \cdot \left(q_i^t - \frac{1}{n} \right) + S\gamma^2 \cdot q_i^t + S\frac{\gamma^2}{n^2} \cdot (1 - q_i^t) \right) \\
&\overset{(b)}{\leqslant} \Phi_i^t \cdot \left(1 + \gamma \cdot \left(q_i^t - \frac{1}{n} \right) + 2CS \cdot \frac{\gamma^2}{n} \right),
\end{aligned}$$

using in (a) Lemma D.4 twice with $\kappa = 1 - \frac{1}{n}$ and with $\kappa = -\frac{1}{n}$ respectively (and that $(1-1/n)^2 \leqslant 1$), and in (b) that $q_i \leqslant \frac{C}{n}$ with $C > 1$ by condition $\mathcal{C}_2$. Similarly, for the underloaded potential we have that

$$\begin{aligned}
\mathbf{E}\big[\, \Psi_i^{t+1} \,\big|\, \mathfrak{F}^t \,\big] &= \Psi_i^t \cdot \mathbf{E}\big[\, e^{-\gamma W(1-1/n)} \,\big|\, \mathfrak{F}^t \,\big] \cdot q_i^t + \Psi_i^t \cdot \mathbf{E}\big[\, e^{\gamma W/n} \,\big|\, \mathfrak{F}^t \,\big] \cdot (1 - q_i^t) \\
&\overset{(a)}{\leqslant} \Psi_i^t \cdot \left(1 - \gamma \cdot \left(1 - \frac{1}{n} \right) + S\gamma^2 \right) \cdot q_i^t + \Psi_i^t \cdot \left(1 + \frac{\gamma}{n} + S\frac{\gamma^2}{n^2} \right) \cdot (1 - q_i^t) \\
&= \Psi_i^t \cdot \left(1 + \gamma \cdot \left(\frac{1}{n} - q_i^t \right) + S\gamma^2 \cdot q_i^t + S\frac{\gamma^2}{n^2} \cdot (1 - q_i^t) \right) \\
&\overset{(b)}{\leqslant} \Psi_i^t \cdot \left(1 + \gamma \cdot \left(\frac{1}{n} - q_i^t \right) + 2CS \cdot \frac{\gamma^2}{n} \right),
\end{aligned}$$

using in (a) Lemma D.4 with $\kappa = -(1 - \frac{1}{n})$ and with $\kappa = \frac{1}{n}$ respectively (and that $(1-1/n)^2 \leqslant 1$) and in (b) that $q_i \leqslant \frac{C}{n}$ with $C > 1$ by condition $\mathcal{C}_2$. This completes the proof. $\square$

Combining Lemma 3.5 with Theorem 3.2, for the identity folding, we obtain:

Corollary 3.6. *Consider any* TIME-HOMOGENEOUS(p) *process with p satisfying condition $\mathcal{C}_1$ for some constant $\delta \in (0,1)$ and some $\epsilon \in (0,1)$, and condition $\mathcal{C}_2$ for some $C > 1$.*

Further, consider the WEIGHTED *setting with weights from a* FINITE-MGF(S) *distribution with $S \geqslant 1$. Then, there exists a constant $c := c(\delta) > 0$, such that for $\Gamma := \Gamma(\gamma)$ with any $\gamma \in \big(0, \frac{\epsilon\delta}{16CS}\big]$ and for any step $t \geqslant 0$,*

$$\mathbf{E}\big[\, \Delta\Gamma^{t+1} \,\big|\, \mathfrak{F}^t \,\big] \leqslant -\Gamma^t \cdot \frac{\gamma\epsilon\delta}{8n} + c\gamma\epsilon.$$

$(1+\beta)$-process for small β

Next we improve the upper bound on the gap for the $(1+\beta)$-process from previous work for very small β in the unit weights setting. In [152, Corollary 2.12], it was shown that this gap is $\mathcal{O}(\log n/\beta + \log(1/\beta)/\beta)$. For $\beta = n^{-\omega(1)}$, the second term dominates. We improve this gap bound to $\mathcal{O}(\log n/\beta)$. This is tight up to multiplicative constants for any $\beta \leqslant 1/2$, due to a lower bound of $\Omega(\log n/\beta)$ shown in [152, Section 4].

Theorem 3.7. *Consider the $(1+\beta)$-process for any $\beta \in (0,1]$. Then, there exists a constant $\kappa > 0$, such that for any step $m \geqslant 0$,*

$$\mathbf{Pr}\left[\, \mathrm{Gap}(m) \leqslant \kappa \cdot \frac{\log n}{\beta} \,\right] \geqslant 1 - n^{-2}.$$

Proof. By Proposition 2.11, the $(1+\beta)$-process satisfies conditions $\mathcal{C}_1$ for $\epsilon = \frac{\beta}{4}$ and $\delta = \frac{1}{4}$ and $\mathcal{C}_2$ for $C = 2$. Hence, by Corollary 3.6 (with $S = 1$ since we are in the unit weights setting), there exists a constant $c := c(\delta) > 0$, such that for $\Gamma := \Gamma(\gamma)$ with $\gamma := \frac{\epsilon\delta}{16CS} = \frac{\beta}{512}$, for any step $m \geqslant 0$,

$$\mathbf{E}\left[\, \Gamma^m \,\right] \leqslant \frac{8c}{\delta} \cdot n.$$

Hence, using Markov's inequality

$$\mathbf{Pr}\left[\, \Gamma^m \leqslant \frac{8c}{\delta} \cdot n^3 \,\right] \geqslant 1 - n^{-2}.$$

When the event $\{\Gamma^m \leqslant \frac{8c}{\delta} \cdot n^3\}$ holds, we deduce the desired bound on the gap

$$\mathrm{Gap}(m) \leqslant \frac{1}{\gamma} \cdot \left(\log\left(\frac{8c}{\delta}\right) + 3 \cdot \log n \right) \leqslant \frac{4}{\gamma} \cdot \log n = 4 \cdot \frac{512}{\beta} \cdot \log n = \mathcal{O}\left(\frac{\log n}{\beta}\right). \qquad \square$$

Quantile(δ) process

Recall from Section 2.5.3 that the Quantile(δ) process is equivalent to the Time-Homogeneous(p) process with

$$p_i = \begin{cases} \frac{\delta}{n} & \text{for } i \leqslant \delta n, \\ \frac{1+\delta}{n} & \text{otherwise.} \end{cases}$$

Note that for any constant $\delta \in (0,1)$, the Quantile(δ) process satisfies condition $\mathcal{C}_1$ at quantile δ and for $\epsilon = 1 - \delta$, and condition $\mathcal{C}_2$ for $C = 2$, as it is a Two-Sample process. Hence, by Corollary 3.6, we obtain an $\mathcal{O}(\log n)$ bound on the gap.

For non-constant δ, we cannot directly apply Theorem 3.2 because condition $\mathcal{C}_1$ is not satisfied. We use the following majorisation lemma to "move" the bias to a constant quantile, e.g., $1/3$.

Lemma 3.8. *For any quantile $\delta \leqslant 1/3$, the Quantile(δ) process is majorised by the Time-Homogeneous(p) process with*

$$p_i = \begin{cases} \frac{1-\epsilon}{n} & \text{if } i \leqslant n/3, \\ \frac{1+\widetilde{\epsilon}}{n} & \text{otherwise,} \end{cases}$$

where $\epsilon = 2\delta$ and $\widetilde{\epsilon} = \delta$.

Proof. The prefix sums for the two probability vectors at $i = n/3$ are equal since,

$$\sum_{i=1}^{n/3} \widetilde{q}_i = \delta^2 + \left(\frac{n}{3} - \delta n\right) \cdot \frac{1+\delta}{n} = \frac{1-2\delta}{3}, \quad \text{and} \quad \sum_{i=1}^{n/3} p_i = \frac{n}{3} \cdot \frac{1-\epsilon}{n} = \frac{1-\epsilon}{3}.$$

Since the first $n/3$ entries in p are equal and $\widetilde{q}$ is non-decreasing, p majorises $\widetilde{q}$ for all indices $i \leqslant n/3$. Similarly, this argument applies to the last $2n/3$ entries. □

Hence, by Corollary 3.6, QUANTILE(δ) satisfies $\mathcal{C}_1$ for $\epsilon = \Theta(\delta)$ and $\delta = 1/3$, and $\mathcal{C}_2$ for $C = 2$, so we deduce an $\mathcal{O}\big(\frac{\log n}{\delta}\big)$ bound on the gap.

Corollary 3.9. *Consider any* QUANTILE(δ) *process with any quantile $\delta \leqslant 1/2$. Then, there exists a constant $\kappa > 0$ such that for any step $m \geqslant 0$,*

$$\mathbf{Pr}\left[\, \mathrm{Gap}(m) \leqslant \kappa \cdot \frac{\log n}{\delta} \,\right] \geqslant 1 - n^{-2}.$$

TWINNING-WITH-QUANTILE

To demonstrate the power of Theorem 3.2, we also analyse a variant of the TWINNING process, which is based on quantiles. Note that this process allocates $1 \cdot \delta^2 + 2 \cdot (1-\delta^2) = 2 - \delta^2$ balls per sample in expectation, so it is more sample efficient than ONE-CHOICE. We will now show that this process also has an $\mathcal{O}(\log n)$ gap.

TWINNING-WITH-QUANTILE(δ) Process:
Iteration: At step $t \geqslant 0$, sample a bin $i \in [n]$ uniformly at random, and update its load:

$$x_i^{t+1} = \begin{cases} x_i^t + 1 & \text{if } \mathrm{Rank}^t(i) \leqslant \delta n, \\ x_i^t + 2 & \text{otherwise.} \end{cases}$$

We will now show that this process satisfies the preconditions for Theorem 3.2 for the sorted allocation vector p of the QUANTILE(δ) process, given by

$$p_i = \begin{cases} \frac{\delta}{n} & \text{if } i \leqslant \delta n, \\ \frac{1+\delta}{n} & \text{otherwise.} \end{cases}$$

Lemma 3.10. *Consider the* TWINNING-WITH-QUANTILE(δ) *process for any constant quantile $\delta \in (0,1)$ and the overload potential $\Phi := \Phi(\gamma)$ with any $\gamma \in (0, 1/2]$. Then, for any step $t \geqslant 0$, we have that*

$$\mathbf{E}\left[\, \Delta\Phi^{t+1} \,\middle|\, \mathfrak{F}^t \,\right] \leqslant \sum_{i=1}^{n} \Phi_i^t \cdot \left(\gamma \cdot \left(p_i - \frac{1}{n}\right) + 5 \cdot \frac{\gamma^2}{n}\right),$$

where p is the sorted allocation vector of QUANTILE(δ).

Proof. We will analyse the expected change of Φ over an arbitrary step t. We consider two cases based on the rank of a bin $i \in [n]$, splitting them into heavy ($\mathrm{Rank}^t(i) \leqslant \delta n$) and light ($\mathrm{Rank}^t(i) > \delta n$):

Case 1 [$\text{Rank}^t(i) \leqslant \delta n$]**:** If we sample this heavy bin i, then we allocate one ball to it, so

$$\mathbf{E}\left[\, \Phi_i^{t+1} \,\middle|\, \mathfrak{F}^t \,\right] = \underbrace{\Phi_i^t \cdot e^{\gamma(1-1/n)} \cdot \frac{1}{n}}_{\text{Allocate one ball to bin } i} + \underbrace{\Phi_i^t \cdot e^{-\gamma/n} \cdot \left(\delta - \frac{1}{n}\right)}_{\text{Allocate one ball to any other heavy bin}} + \underbrace{\Phi_i^t \cdot e^{-2\gamma/n} \cdot (1-\delta)}_{\text{Allocate to any light bin}}.$$

Now, we proceed to bound this quantity,

$$\begin{aligned}
\mathbf{E}\left[\, \Phi_i^{t+1} \,\middle|\, \mathfrak{F}^t \,\right] &= \Phi_i^t \cdot \left(1 + (e^{\gamma(1-1/n)} - 1) \cdot \frac{1}{n} + (e^{-\gamma/n} - 1) \cdot \left(\delta - \frac{1}{n}\right) + (e^{-2\gamma/n} - 1) \cdot (1-\delta)\right) \\
&\overset{(a)}{\leqslant} \Phi_i^t \cdot \left(1 + \left(\gamma \cdot \left(1 - \frac{1}{n}\right) + \gamma^2\right) \cdot \frac{1}{n} + \left(-\frac{\gamma}{n} + \frac{\gamma^2}{n^2}\right) \cdot \left(\delta - \frac{1}{n}\right) + \left(-\frac{2\gamma}{n} + \frac{4\gamma^2}{n^2}\right) \cdot (1-\delta)\right) \\
&\leqslant \Phi_i^t \cdot \left(1 + \gamma \cdot \left(\frac{\delta}{n} - \frac{1}{n}\right) + 3 \cdot \frac{\gamma^2}{n}\right),
\end{aligned}$$

using in (a) the Taylor estimate $e^v \leqslant 1 + v + v^2$ for any $|v| \leqslant 1$ (and that $0 < \gamma \leqslant 1/2$).

Case 2 [$\text{Rank}^t(i) > \delta n$]**:** If we sample this light bin i, then we allocate two balls to it, so

$$\mathbf{E}\left[\, \Phi_i^{t+1} \,\middle|\, \mathfrak{F}^t \,\right] = \underbrace{\Phi_i^t \cdot e^{\gamma(2-2/n)} \cdot \frac{1}{n}}_{\text{Allocate two balls to bin } i} + \underbrace{\Phi_i^t \cdot e^{-\gamma/n} \cdot \delta}_{\text{Allocate one ball to any heavy bin}} + \underbrace{\Phi_i^t \cdot e^{-2\gamma/n} \cdot \left(1 - \delta - \frac{1}{n}\right)}_{\text{Allocate to any other light bin}}.$$

Similarly to Case 1, we proceed to bound this quantity,

$$\begin{aligned}
\mathbf{E}\left[\, \Phi_i^{t+1} \,\middle|\, \mathfrak{F}^t \,\right] &= \Phi_i^t \cdot \left(1 + (e^{\gamma(2-2/n)} - 1) \cdot \frac{1}{n} + (e^{-\gamma/n} - 1) \cdot \delta + (e^{-2\gamma/n} - 1) \cdot \left(1 - \delta - \frac{1}{n}\right)\right) \\
&\overset{(a)}{\leqslant} \Phi_i^t \cdot \left(1 + \left(\gamma \cdot \left(2 - \frac{2}{n}\right) + 4\gamma^2\right) \cdot \frac{1}{n} + \left(-\frac{\gamma}{n} + \frac{\gamma^2}{n^2}\right) \cdot \delta + \left(-\frac{2\gamma}{n} + \frac{4\gamma^2}{n^2}\right) \cdot \left(1 - \delta - \frac{1}{n}\right)\right) \\
&\leqslant \Phi_i^t \cdot \left(1 + \gamma \cdot \frac{\delta}{n} + 5 \cdot \frac{\gamma^2}{n}\right),
\end{aligned}$$

using in (a) the Taylor estimate $e^v \leqslant 1 + v + v^2$ for any $|v| \leqslant 1$ (and that $0 < \gamma \leqslant 1/2$). Hence, combining the two cases, we have that

$$\mathbf{E}\left[\, \Phi^{t+1} \,\middle|\, \mathfrak{F}^t \,\right] \leqslant \sum_{i=1}^{n} \Phi_i^t \cdot \left(1 + \gamma \cdot \left(p_i - \frac{1}{n}\right) + 5 \cdot \frac{\gamma^2}{n}\right). \qquad \square$$

In a similar manner, we also obtain the bounds for the change of the underload potential.

Lemma 3.11. *Consider the* Twinning-with-Quantile(δ) *process for any constant* $\delta \in (0,1)$ *and the underload potential* $\Psi := \Psi(\gamma)$ *with any* $\gamma \in (0, 1/2]$*. Then, for any step* $t \geqslant 0$*, we have that*

$$\mathbf{E}\left[\, \Delta\Psi^{t+1} \,\middle|\, \mathfrak{F}^t \,\right] \leqslant \sum_{i=1}^{n} \Psi_i^t \cdot \left(\gamma \cdot \left(\frac{1}{n} - p_i\right) + 5 \cdot \frac{\gamma^2}{n}\right),$$

where p *is the sorted allocation vector of* Quantile(δ).

Lemmas 3.10 and 3.11 verify the preconditions of Theorem 3.2, so applying the theorem for the identity folding, we get an $\mathcal{O}(\log n)$ bound on the gap.

Corollary 3.12. *For the* Twinning-with-Quantile(δ) *process for any constant* $\delta \in (0,1)$*, there exists a constant* $\kappa := \kappa(\delta) > 0$*, such that for any step* $m \geqslant 0$*,*

$$\mathbf{Pr}\left[\, \text{Gap}(m) \leqslant \kappa \cdot \log n \,\right] \geqslant 1 - n^{-2}.$$

3.2.2 b-Batched setting: The $\mathcal{O}\big(\frac{b}{n} \cdot \log n\big)$ upper bound

In this section we derive an upper bound of $\mathcal{O}\big(\frac{b}{n} \cdot \log n\big)$ for the weighted batched setting for a family of processes. For the $(1+\beta)$-process with constant $\beta \in (0,1)$, this upper bound is tight for $b = \Theta(n)$, as we will show in Appendix C. This will also serve as the base case for the tighter analysis for $b = \omega(n)$ in Section 7.5.

The main goal is to derive the preconditions of Theorem 3.2 and apply it for $R := b$ over the batches (not individual time steps).

Lemma 3.13. *Consider any* Sequential(q^t) *process with q^t satisfying condition $\mathcal{C}_2$ for some $C > 1$ at every step $t \geqslant 0$. Further, consider the* Weighted b-Batched *setting with batch size $b \geqslant n$ with weights from a* Finite-MGF(S) *distribution with $S \geqslant 1$. Then, for $\Phi := \Phi(\gamma)$ and $\Psi := \Psi(\gamma)$ with any $\gamma \in \big(0, \frac{n}{2CSb}\big]$ and for any step $t \geqslant 0$ being a multiple of b,*

$$\mathbf{E}\big[\, \Phi^{t+b} \,\big|\, \mathfrak{F}^t \,\big] \leqslant \sum_{i=1}^n \Phi_i^t \cdot \left(1 + \left(q_i^t - \frac{1}{n}\right) \cdot b \cdot \gamma + \frac{5C^2S^2b}{n} \cdot b \cdot \frac{\gamma^2}{n}\right), \tag{3.15}$$

and

$$\mathbf{E}\big[\, \Psi^{t+b} \,\big|\, \mathfrak{F}^t \,\big] \leqslant \sum_{i=1}^n \Psi_i^t \cdot \left(1 + \left(\frac{1}{n} - q_i^t\right) \cdot b \cdot \gamma + \frac{5C^2S^2b}{n} \cdot b \cdot \frac{\gamma^2}{n}\right). \tag{3.16}$$

Proof. Consider an arbitrary step t being a multiple of b and for convenience let $q = q^t$. Consider an arbitrary bin $i \in [n]$. Let $Z \in \{0,1\}^b$ be the indicator vector, where Z_j indicates whether the j-th ball was allocated to bin i. The expected change for the overload potential Φ_i^t, is given by

$$\mathbf{E}\big[\, \Phi_i^{t+b} \,\big|\, \mathfrak{F}^t \,\big] = \Phi_i^t \cdot \sum_{z \in \{0,1\}^b} \mathbf{Pr}\left[\, Z = z \,\right] \cdot \mathbf{E}\left[\, e^{\gamma \sum_{j=1}^b \left(z_j w^{t+j} - \frac{w^{t+j}}{n}\right)} \,\middle|\, \mathfrak{F}^t, Z = z \,\right].$$

In the following, let us upper bound the factor of Φ_i^t:

$$\begin{aligned}
&\sum_{z \in \{0,1\}^b} \mathbf{Pr}\left[\, Z = z \,\right] \cdot \mathbf{E}\left[\, e^{\gamma \sum_{j=1}^b \left(z_j w^{t+j} - \frac{w^{t+j}}{n}\right)} \,\middle|\, \mathfrak{F}^t, Z = z \,\right] \\
&\stackrel{(a)}{=} \sum_{z \in \{0,1\}^b} \prod_{j=1}^b (q_i)^{z_j} (1-q_i)^{1-z_j} (\mathbf{E}[\, e^{\gamma W (1 - \frac{1}{n})} \,])^{z_j} (\mathbf{E}[\, e^{-\gamma W / n} \,])^{1-z_j} \\
&\stackrel{(b)}{\leqslant} \sum_{z \in \{0,1\}^b} \prod_{j=1}^b \left(q_i \cdot \left(1 + \gamma \cdot \left(1 - \frac{1}{n}\right) + S\gamma^2\right)\right)^{z_j} \cdot \left((1-q_i) \cdot \left(1 - \frac{\gamma}{n} + \frac{S\gamma^2}{n^2}\right)\right)^{1-z_j} \\
&\stackrel{(c)}{=} \left(q_i \cdot \left(1 + \gamma \cdot \left(1 - \frac{1}{n}\right) + S\gamma^2\right) + (1-q_i) \cdot \left(1 - \frac{\gamma}{n} + \frac{S\gamma^2}{n^2}\right)\right)^b \\
&= \left(1 + \gamma \cdot \left(q_i - \frac{1}{n}\right) + q_i \cdot S\gamma^2 + (1-q_i) \cdot \frac{S\gamma^2}{n^2}\right)^b \qquad (3.17) \\
&\stackrel{(d)}{\leqslant} \left(1 + \gamma \cdot \left(q_i - \frac{1}{n}\right) + 2 \cdot \frac{C}{n} \cdot S\gamma^2\right)^b, \qquad (3.18)
\end{aligned}$$

using in (a) that the weights are independent given $\mathfrak{F}^t$, in (b) Lemma D.4 twice with $\kappa = 1 - \frac{1}{n}$ and with $\kappa = -\frac{1}{n}$ respectively (and that $(1-1/n)^2 \leqslant 1$), in (c) the binomial theorem and in (d) that $q_i \leqslant \frac{C}{n}$ for $C > 1$. Let us define

$$u := \left(q_i - \frac{1}{n}\right) \cdot \gamma + 2 \cdot \frac{C}{n} \cdot S\gamma^2.$$

We first claim that $|u \cdot b| \leqslant 1$, which holds indeed since

$$|u \cdot b| = \left|\left(q_i - \frac{1}{n}\right) \cdot b \cdot \gamma\right| + 2 \cdot \frac{C}{n} \cdot b \cdot S\gamma^2 \leqslant \frac{C}{n} \cdot b \cdot \gamma + 2 \cdot \frac{C}{n} \cdot b \cdot S\gamma^2 \leqslant \frac{2CS}{n} \cdot \gamma \cdot b \leqslant 1,$$

where we used[2] $q_i \leqslant \frac{C}{n}$ for $C > 1$ and that $\gamma \leqslant \frac{n}{2CSb} \leqslant \frac{1}{2}$.

Then,

$$\begin{aligned}
\mathbf{E}\left[\, \Phi_i^{t+b} \,\middle|\, \mathfrak{F}^t \,\right] &\overset{(a)}{\leqslant} \Phi_i^t \cdot e^{u \cdot b} \\
&\overset{(b)}{\leqslant} \Phi_i^t \cdot \left(1 + u \cdot b + (u \cdot b)^2\right) \\
&= \Phi_i^t \cdot \left(1 + \left(q_i - \frac{1}{n}\right) \cdot b \cdot \gamma + 2 \cdot \frac{C}{n} \cdot S\gamma^2 \cdot b + \left(\left(q_i - \frac{1}{n}\right) \cdot b \cdot \gamma + 2 \cdot \frac{C}{n} \cdot b \cdot S\gamma^2\right)^2\right),
\end{aligned}$$

using in (a) that $1 + v \leqslant e^v$ for any v, and in (b) that $e^v \leqslant 1 + v + v^2$ for $v \leqslant 1.75$. Since $q_i \leqslant \frac{C}{n}$ for all $i \in [n]$ and $S \geqslant 1$, we conclude

$$\begin{aligned}
\mathbf{E}\left[\, \Phi_i^{t+b} \,\middle|\, \mathfrak{F}^t \,\right] &\leqslant \Phi_i^t \cdot \left(1 + \left(q_i - \frac{1}{n}\right) \cdot b \cdot \gamma + 2 \cdot \frac{CS}{n} \cdot \gamma^2 \cdot b + \left(\frac{2CS}{n} \cdot b \cdot \gamma\right)^2\right) \\
&\leqslant \Phi_i^t \cdot \left(1 + \left(q_i - \frac{1}{n}\right) \cdot b \cdot \gamma + \frac{5C^2S^2b}{n} \cdot b \cdot \frac{\gamma^2}{n}\right).
\end{aligned}$$

Similarly, for the underloaded potential Ψ^t, for any bin $i \in [n]$,

$$\mathbf{E}\left[\, \Psi_i^{t+b} \,\middle|\, \mathfrak{F}^t \,\right] = \Psi_i^t \cdot \sum_{z \in \{0,1\}^b} \mathbf{Pr}\left[\, Z = z \,\right] \cdot \mathbf{E}\left[\, e^{-\gamma \sum_{j=1}^b \left(z_j w^{t+j} - \frac{w^{t+j}}{n}\right)} \,\middle|\, \mathfrak{F}^t, Z = z \,\right].$$

As before, we will upper bound the factor of Ψ_i^t:

$$\begin{aligned}
&\sum_{z \in \{0,1\}^b} \mathbf{Pr}\left[\, Z = z \,\right] \cdot \mathbf{E}\left[\, e^{-\gamma \sum_{j=1}^b \left(z_j w^{t+j} - \frac{w^{t+j}}{n}\right)} \,\middle|\, \mathfrak{F}^t, Z = z \,\right] \\
&\overset{(a)}{=} \sum_{z \in \{0,1\}^b} \prod_{j=1}^b (q_i)^{z_j} (1 - q_i)^{1 - z_j} (\mathbf{E}[\, e^{-\gamma W \cdot (1 - \frac{1}{n})} \,])^{z_j} (\mathbf{E}[\, e^{\gamma W / n} \,])^{1 - z_j} \\
&\overset{(b)}{\leqslant} \sum_{z \in \{0,1\}^b} \prod_{j=1}^b \left(q_i \cdot \left(1 - \gamma \cdot \left(1 - \frac{1}{n}\right) + S\gamma^2\right)\right)^{z_j} \cdot \left((1 - q_i) \cdot \left(1 + \frac{\gamma}{n} + \frac{S\gamma^2}{n^2}\right)\right)^{1 - z_j} \\
&\overset{(c)}{=} \left(q_i \cdot \left(1 - \gamma \cdot \left(1 - \frac{1}{n}\right) + S\gamma^2\right) + (1 - q_i) \cdot \left(1 + \frac{\gamma}{n} + \frac{S\gamma^2}{n^2}\right)\right)^b \\
&= \left(1 + \left(\frac{1}{n} - q_i\right) \cdot \gamma + q_i \cdot S\gamma^2 + (1 - q_i) \cdot \frac{S\gamma^2}{n^2}\right)^b \qquad (3.19) \\
&\leqslant \left(1 + \left(\frac{1}{n} - q_i\right) \cdot \gamma + 2 \cdot \frac{C}{n} \cdot S\gamma^2\right)^b, \qquad (3.20)
\end{aligned}$$

using in (a) that the weights W are independent given $\mathfrak{F}^t$, in (b) Lemma D.4 twice with $\kappa = -\left(1 - \frac{1}{n}\right)$ and with $\kappa = \frac{1}{n}$ respectively and in (c) the binomial theorem. So,

$$\mathbf{E}[\, \Psi_i^{t+b} \mid \mathfrak{F}^t \,] \overset{(a)}{\leqslant} \Psi_i^t \cdot e^{(\frac{1}{n} - q_i) \cdot b \cdot \gamma + 2 \cdot \frac{C}{n} \cdot S\gamma^2 \cdot b}$$

[2]There is some slack in this bound, especially for $C = 1 + o(1)$. We will exploit this in Lemma 3.16.

$$\overset{(b)}{\leqslant} \Psi_i^t \cdot \left(1 + \left(\frac{1}{n} - q_i\right) \cdot b \cdot \gamma + 2 \cdot \frac{C}{n} \cdot S\gamma^2 \cdot b + \left(\left(\frac{1}{n} - q_i\right) \cdot b \cdot \gamma + 2 \cdot \frac{C}{n} \cdot S\gamma^2 \cdot b\right)^2\right)$$
$$\overset{(c)}{\leqslant} \Psi_i^t \cdot \left(1 + \left(\frac{1}{n} - q_i\right) \cdot b \cdot \gamma + 2 \cdot \frac{CS}{n} \cdot \gamma^2 \cdot b + \left(\frac{2CS}{n} \cdot b \cdot \gamma\right)^2\right)$$
$$\leqslant \Psi_i^t \cdot \left(1 + \left(\frac{1}{n} - q_i\right) \cdot b \cdot \gamma + \frac{5C^2S^2b}{n} \cdot b \cdot \frac{\gamma^2}{n}\right),$$

using in (a) that $1 + v \leqslant e^v$ for any v, in (b) that $e^v \leqslant 1 + v + v^2$ for any $v \leqslant 1.75$ and that $(\frac{1}{n} - q_i) \cdot b \cdot \gamma + 2 \cdot q_i \cdot S\gamma^2 \cdot b \leqslant \frac{1}{n} \cdot b \cdot \gamma + 2 \cdot \frac{C}{n} \cdot S\gamma^2 \cdot b \leqslant 2CS \cdot \gamma \cdot \frac{b}{n} \leqslant 1$, since $\gamma \leqslant \frac{n}{2CSb} \leqslant \frac{1}{2}$ and $S \geqslant 1$. $\square$

We are now ready to apply Theorem 3.2 for $R := b$, folding every b steps, meaning that each round consists of b consecutive allocations.

Theorem 3.14. *Consider any* SEQUENTIAL(q^t) *process with q^t satisfying condition $\mathcal{C}_1$ for constant $\delta \in (0,1)$ and (not necessarily constant) $\epsilon \in (0,1)$ as well as condition $\mathcal{C}_2$ for some constant $C > 1$ at every step $t \geqslant 0$. Further, consider the* WEIGHTED b-BATCHED *setting with any $b \geqslant n$ and weights from a* FINITE-MGF(S) *distribution with constant $S \geqslant 1$. Then, there exists a constant $\kappa := \kappa(\delta, C, S) > 0$, such that for any step $m \geqslant 0$ being a multiple of b,*

$$\mathbf{Pr}\left[\max_{i \in [n]} |y_i^m| \leqslant \kappa \cdot \frac{1}{\epsilon} \cdot \frac{b}{n} \cdot \log n\right] \geqslant 1 - n^{-2}.$$

Remark 3.15. *The same upper bound in Theorem 3.14 also holds for* TIME-HOMOGENEOUS-WITH-RAND-TIE-BREAKS(p) *processes with a probability vector p satisfying the preconditions of Lemma 3.13. The reason for this is that (i) averaging probabilities in Eq.* (2.1) *can only reduce the maximum entry in the allocation vector q^t, i.e. $\max_{i \in [n]} q_i^t(x^t) \leqslant \max_{i \in [n]} p_i$, so it still satisfies $\mathcal{C}_2$ and (ii) moving probability between bins i, j with $x_i^t = x_j^t$ (and thus $\Phi_i^t = \Phi_j^t$ and $\Psi_i^t = \Psi_j^t$), implies that the aggregate upper bounds in* (3.15) *and* (3.16) *remain the same.*

Proof of Theorem 3.14. By Lemma 3.13, the preconditions of Theorem 3.2 are satisfied for $p := q$, $K := \frac{5C^2S^2b}{n}$, $R := b$ and $\gamma := \frac{\epsilon\delta}{8K} \leqslant \frac{n}{2CSb}$ (as $\epsilon \leqslant 1$, $\delta \leqslant 1$ and $C \geqslant 1$, $S \geqslant 1$). Hence, there exists a constant $c := c(\delta) > 0$ such that for any step $m \geqslant 0$ being multiple of b,

$$\mathbf{E}[\,\Gamma^m\,] \leqslant \frac{8c}{\delta} \cdot n.$$

Hence, by Markov's inequality

$$\mathbf{Pr}\left[\Gamma^m \leqslant \frac{8c}{\delta} \cdot n^3\right] \geqslant 1 - n^{-2}.$$

To prove the claim, note that when $\{\Gamma^m \leqslant \frac{8c}{\delta} \cdot n^3\}$ holds, then also,

$$\max_{i \in [n]} |y_i^m| \leqslant \frac{1}{\gamma} \cdot \left(\log\left(\frac{8c}{\delta}\right) + 3 \cdot \log n\right) \leqslant \frac{4}{\gamma} \cdot \log n = 4 \cdot \frac{8 \cdot 5C^2S^2}{\epsilon\delta} \cdot \frac{b}{n} \cdot \log n. \qquad \square$$

3.2.3 b-BATCHED setting: The $\mathcal{O}\left(\sqrt{\frac{b}{n}} \cdot \log n\right)$ upper bound

We now refine our analysis from Section 3.2.2, allowing us to obtain tighter bounds for the $(1+\beta)$-process. These bounds show that for batch sizes $b = \omega(n \log n)$, the $(1+\beta)$-process outperforms the TWO-CHOICE process.

The key idea is to redo the later part of the analysis in Lemma 3.13 for probability allocation vectors q that satisfy the following condition $\mathcal{C}_3$:

- There exists $C > 1$, such that for any bin $i \in [n]$,

$$\left| q_i - \frac{1}{n} \right| \leqslant \frac{C-1}{n}.$$

Note that this condition implies condition $\mathcal{C}_2$, i.e., that $q_i \leqslant \frac{C}{n}$ (with the same C).

This condition is satisfied by the $(1+\beta)$-process for $C = 1+\beta$, since

$$\left| q_i - \frac{1}{n} \right| = \left| \frac{2i-1}{n^2} \cdot \beta + \frac{1}{n} \cdot (1-\beta) - \frac{1}{n} \right| \leqslant \max\left\{ q_n - \frac{1}{n}, \frac{1}{n} - q_1 \right\} = \frac{\beta}{n} - \frac{\beta}{n^2} \leqslant \frac{\beta}{n}.$$

It is also satisfied by the β-Mixed(Quantile$(1/2)$, One-Choice) process.

Lemma 3.16 (cf. Lemma 3.13). *Consider any* Sequential(q^t) *process with probability allocation vector q^t satisfying condition $\mathcal{C}_3$ for some $C \in (1, 1.9)$ at every step $t \geqslant 0$. Further, consider the* Weighted *b-*Batched *setting with weights from a* Finite-MGF(S) *distribution with constant $S \geqslant 1$ and a batch size $b \geqslant \frac{2CS}{(C-1)^2} \cdot n$. Then for $\Phi := \Phi(\gamma)$ and $\Psi := \Psi(\gamma)$ with any smoothing parameter $\gamma \in (\frac{n}{2(C-1) \cdot b}]$ and any step $t \geqslant 0$ being a multiple of b,*

$$\mathbf{E}\left[\Phi^{t+b} \,\middle|\, \mathfrak{F}^t \right] \leqslant \sum_{i=1}^{n} \Phi_i^t \cdot \left(1 + \left(q_i^t - \frac{1}{n} \right) \cdot b \cdot \gamma + \frac{5(C-1)^2 b}{n} \cdot b \cdot \frac{\gamma^2}{n} \right),$$

and

$$\mathbf{E}\left[\Psi^{t+b} \,\middle|\, \mathfrak{F}^t \right] \leqslant \sum_{i=1}^{n} \Psi_i^t \cdot \left(1 + \left(\frac{1}{n} - q_i^t \right) \cdot b \cdot \gamma + \frac{5(C-1)^2 b}{n} \cdot b \cdot \frac{\gamma^2}{n} \right).$$

Proof. Consider an arbitrary step $t \geqslant 0$ being a multiple of b and for convenience let $q = q^t$. First note that the given assumptions $\gamma \leqslant \frac{n}{2(C-1) \cdot b}$ and $b \geqslant \frac{2CS}{(C-1)^2} \cdot n$ imply that

$$\gamma \leqslant \frac{n}{2(C-1) \cdot b} \leqslant \frac{C-1}{4CS}.$$

Consider an arbitrary bin $i \in [n]$. Using Eq. (3.17) in Lemma 3.13,

$$\mathbf{E}\left[\Phi_i^{t+b} \,\middle|\, \mathfrak{F}^t \right] \leqslant \Phi_i^t \cdot \left(1 + \gamma \cdot \left(q_i - \frac{1}{n} \right) + q_i \cdot S\gamma^2 + (1-q_i) \cdot \frac{S\gamma^2}{n^2} \right)^b \leqslant \Phi_i^t \cdot \left(1 + \gamma \cdot \left(q_i - \frac{1}{n} \right) + 2 \cdot q_i \cdot S\gamma^2 \right)^b,$$

using that $\left| q_i - \frac{1}{n} \right| \leqslant \frac{C-1}{n}$ for $C \in (1, 1.9)$. Let us define

$$u := \left(q_i - \frac{1}{n} \right) \cdot \gamma + 2 \cdot q_i \cdot S\gamma^2. \tag{3.21}$$

Compared to the proof in Lemma 3.13, we will aim for a tighter bound. More specifically, we will now show that $|u \cdot b| \leqslant 2(C-1) \cdot b \cdot \frac{\gamma}{n} \leqslant 1$, which holds indeed since

$$\begin{aligned}
|u \cdot b| &= \left| \left(q_i - \frac{1}{n} \right) \cdot b \cdot \gamma + 2 \cdot q_i \cdot b \cdot S\gamma^2 \right| \\
&\leqslant \left| \left(q_i - \frac{1}{n} \right) \cdot b \cdot \gamma \right| + 2 \cdot q_i \cdot b \cdot S\gamma^2 \\
&\overset{(a)}{\leqslant} \frac{C-1}{n} \cdot b \cdot \gamma + 2 \cdot \frac{C}{n} \cdot b \cdot S\gamma^2
\end{aligned}$$

$$
\begin{aligned}
&= (C-1+2CS\gamma)\cdot b\cdot\frac{\gamma}{n}\\
&\overset{(b)}{\leqslant} 2(C-1)\cdot b\cdot\frac{\gamma}{n} && (3.22)\\
&\overset{(c)}{\leqslant} 1, && (3.23)
\end{aligned}
$$

using in (a) that $\left|q_i-\frac{1}{n}\right|\leqslant\frac{C-1}{n}$ by condition $\mathcal{C}_3$, in (b) that $\gamma\leqslant\frac{C-1}{2CS}$ and in (c) that $\gamma\leqslant\frac{n}{2(C-1)\cdot b}$.

Then,

$$
\begin{aligned}
\mathbf{E}\left[\,\Phi_i^{t+b}\,\middle|\,\mathfrak{F}^t\,\right] &\overset{(a)}{\leqslant} \Phi_i^t\cdot e^{u\cdot b}\\
&\overset{(b)}{\leqslant} \Phi_i^t\cdot\left(1+u\cdot b+(u\cdot b)^2\right)\\
&\overset{(3.21)}{=} \Phi_i^t\cdot\left(1+\left(q_i-\frac{1}{n}\right)\cdot b\cdot\gamma+2\cdot q_i\cdot b\cdot S\gamma^2+(u\cdot b)^2\right)\\
&\overset{(3.22)}{\leqslant} \Phi_i^t\cdot\left(1+\left(q_i-\frac{1}{n}\right)\cdot b\cdot\gamma+2\cdot q_i\cdot b\cdot S\gamma^2+\left(2(C-1)\cdot b\cdot\frac{\gamma}{n}\right)^2\right)\\
&\overset{(c)}{\leqslant} \Phi_i^t\cdot\left(1+\left(q_i-\frac{1}{n}\right)\cdot b\cdot\gamma+\frac{5(C-1)^2b}{n}\cdot b\cdot\frac{\gamma^2}{n}\right),
\end{aligned}
$$

using in (a) that $1+v\leqslant e^v$ for any v, in (b) that $e^v\leqslant 1+v+v^2$ for $v\leqslant 1.75$ and Eq. (3.23), and in (c) that $\frac{(C-1)^2b}{n}\cdot b\cdot\frac{\gamma^2}{n}\geqslant 2\cdot\frac{C}{n}\cdot b\cdot S\gamma^2\geqslant 2\cdot q_i\cdot b\cdot S\gamma^2$, since $b\geqslant\frac{2CS}{(C-1)^2}\cdot n$.

Similarly, for the underloaded potential Ψ^t, using Eq. (3.19), for any bin $i\in[n]$

$$
\mathbf{E}\left[\,\Psi_i^{t+b}\,\middle|\,\mathfrak{F}^t\,\right]\leqslant\Psi_i^t\cdot\left(1+\left(\frac{1}{n}-q_i\right)\cdot\gamma+q_i\cdot S\gamma^2+(1-q_i)\cdot\frac{S\gamma^2}{n^2}\right)^b\leqslant\Psi_i^t\cdot\left(1+\left(\frac{1}{n}-q_i\right)\cdot\gamma+2\cdot q_i\cdot S\gamma^2\right)^b,
$$

using that $\left|q_i-\frac{1}{n}\right|\leqslant\frac{C-1}{n}$ for $C\in(1,1.9)$. Let us define

$$
\widetilde{u}:=\left(\frac{1}{n}-q_i\right)\cdot\gamma+2\cdot q_i\cdot S\gamma^2. \tag{3.24}
$$

Similarly, to Eq. (3.23), we get that

$$
\begin{aligned}
|\widetilde{u}b| &\leqslant \left|\left(\frac{1}{n}-q_i\right)\cdot b\cdot\gamma\right|+2\cdot q_i\cdot b\cdot S\gamma^2\leqslant 2(C-1)\cdot b\cdot\frac{\gamma}{n} && (3.25)\\
&\leqslant 1. && (3.26)
\end{aligned}
$$

So,

$$
\begin{aligned}
\mathbf{E}\left[\,\Psi_i^{t+b}\,\middle|\,\mathfrak{F}^t\,\right] &\overset{(a)}{\leqslant} \Psi_i^t\cdot e^{\widetilde{u}b}\\
&\overset{(b)}{\leqslant} \Psi_i^t\cdot\left(1+\widetilde{u}b+(\widetilde{u}b)^2\right)\\
&\overset{(3.24)}{=} \Psi_i^t\cdot\left(1+\left(\frac{1}{n}-q_i\right)\cdot b\cdot\gamma+2\cdot q_i\cdot S\gamma^2\cdot b+(\widetilde{u}\cdot b)^2\right)\\
&\overset{(3.25)}{\leqslant} \Psi_i^t\cdot\left(1+\left(\frac{1}{n}-q_i\right)\cdot b\cdot\gamma+2\cdot q_i\cdot b\cdot\gamma^2+\left(2(C-1)\cdot b\cdot\frac{\gamma}{n}\right)^2\right)\\
&\overset{(c)}{\leqslant} \Psi_i^t\cdot\left(1+\left(\frac{1}{n}-q_i\right)\cdot b\cdot\gamma+\frac{5(C-1)^2b}{n}\cdot b\cdot\frac{\gamma^2}{n}\right),
\end{aligned}
$$

using in (a) that $1+v\leqslant e^v$ for any v, in (b) that $e^v\leqslant 1+v+v^2$ for $v\leqslant 1.75$ and Eq. (3.26), and in (c) that $\frac{(C-1)^2b}{n}\cdot b\cdot\frac{\gamma^2}{n}\geqslant 2\cdot\frac{C}{n}\cdot b\cdot S\gamma^2\geqslant 2\cdot q_i\cdot b\cdot S\gamma^2$, since $b\geqslant\frac{2CS}{(C-1)^2}\cdot n$. $\square$

Theorem 3.17. *Consider any* SEQUENTIAL(q^t) *process with* q^t *satisfying condition* $\mathcal{C}_1$ *for constant* $\delta \in (0,1)$ *and (not necessarily constant)* $\epsilon \in (0,1)$ *as well as condition* $\mathcal{C}_3$ *for some* $C > 1$*, at every step* $t \geqslant 0$*. Further, consider the* WEIGHTED b-BATCHED *setting with weights from a* FINITE-MGF(S) *distribution with constant* $S \geqslant 1$ *and a batch size* $b \geqslant \frac{2CS}{(C-1)^2} \cdot n$*. Then, there exists a constant* $\kappa := \kappa(\delta) > 0$*, such that for any step* $m \geqslant 0$ *being a multiple of* b*,*

$$\mathbf{Pr}\left[\max_{i \in [n]} |y_i^m| \leqslant \kappa \cdot \frac{(C-1)^2}{\epsilon} \cdot \frac{b}{n} \cdot \log n \right] \geqslant 1 - n^{-2}.$$

Proof. Consider the folding of the b-BATCHED process at steps that are a multiple of b. By Lemma 3.16, the preconditions of Theorem 3.2 are satisfied for $K := 5 \cdot (C-1)^2 \cdot \frac{b}{n}$, $R := b$ and $\gamma := \frac{\epsilon\delta}{8K} = \frac{\epsilon\delta}{40 \cdot (C-1)^2 \cdot \frac{b}{n}} \leqslant \frac{n}{2(C-1) \cdot b}$, since $\epsilon \leqslant C - 1$ and also $\gamma \leqslant 1$ since $b \geqslant \frac{2CS}{(C-1)^2} \cdot n$, $C > 1$ and $S \geqslant 1$. Hence, there exists a constant $c := c(\delta) > 0$ such that for any step $m \geqslant 0$ which is a multiple of b,

$$\mathbf{E}[\, \Gamma^m \,] \leqslant \frac{8c}{\delta} \cdot n.$$

Therefore, by Markov's inequality

$$\mathbf{Pr}\left[\Gamma^m \leqslant \frac{8c}{\delta} \cdot n^3 \right] \geqslant 1 - n^{-2}.$$

To prove the claim, note that when $\left\{\Gamma^m \leqslant \frac{8c}{\delta} \cdot n^3\right\}$ holds, then also,

$$\max_{i \in [n]} |y_i^m| \leqslant \frac{1}{\gamma} \cdot \left(\log\left(\frac{8c}{\delta}\right) + 3 \cdot \log n \right) \leqslant 4 \cdot \frac{\log n}{\gamma} \leqslant 4 \cdot \frac{8 \cdot 5 \cdot (C-1)^2}{\epsilon\delta} \cdot \frac{b}{n} \cdot \log n. \qquad \square$$

Recall that the $(1+\beta)$-process satisfies condition $\mathcal{C}_1$ with $\epsilon = \frac{\beta}{4}$ and $\delta = \frac{1}{4}$, and conditions $\mathcal{C}_2$ and $\mathcal{C}_3$ with $C = 1 + \beta$.

In particular, by considering $\beta = \Theta(\sqrt{n/b})$ we get a process that is asymptotically better than TWO-CHOICE. As we show in Observation C.4, this is just a $\sqrt{\log n}$ factor from the optimal $\Omega(\sqrt{b/n \cdot \log n})$ bound, which holds for any d-SAMPLE process with constant d (in the unit weights setting).

Corollary 3.18. *Let* $b \geqslant n \log n$ *and consider the* WEIGHTED b-BATCHED *setting with weights from a* FINITE-MGF(S) *distribution with* $S \geqslant 1$*. Then, there exists* $\kappa := \kappa(S) > 0$ *such that for the* $(1+\beta)$*-process with* $\beta = \sqrt{4S \cdot \frac{n}{b}}$ *and for any step* $m \geqslant 0$ *being a multiple of* b*,*

$$\mathbf{Pr}\left[\mathrm{Gap}(m) \leqslant \kappa \cdot \sqrt{\frac{b}{n} \cdot \log n} \right] \geqslant 1 - n^{-2}.$$

3.3 A re-allocation argument

In this section, we exploit the intuitive idea that for TWO-SAMPLE processes, it should not matter much if we change decisions between bins with a small (e.g., constant) load difference. We will handle such changes with a potential function argument which bounds the expected change of the potential function for the "noisy" process by relating it to that of the "original" process.

In this section we will prove the $\mathcal{O}(g \log(ng))$ gap bound for the g-ADV-COMP setting for the TWO-CHOICE process (with $g \geqslant 1$ being arbitrary) using this *re-allocation argument*, also recovering the $\mathcal{O}(g \log(ng))$ gap bound for the g-BOUNDED process proven in [142].

By Lemma 3.5, we have for any TWO-SAMPLE process (as it satisfies condition $\mathcal{C}_2$ for $C = 2$),

Lemma 3.19. *Consider any* Two-Sample *process with probability allocation vector* q^t *and the potential* $\Gamma := \Gamma(\gamma)$ *with any* $\gamma \in (0,1]$. *Then, for any step* $t \geqslant 0$,

$$\mathbf{E}\left[\, \Delta\Gamma^{t+1} \,\middle|\, y^t \,\right] \leqslant \frac{4\gamma^2}{n} \cdot \Gamma^t + \sum_{i=1}^{n} \left[e^{\gamma y_i^t} \cdot \left(q_i^t - \frac{1}{n} \right) \cdot \gamma + e^{-\gamma y_i^t} \cdot \left(\frac{1}{n} - q_i^t \right) \cdot \gamma \right]. \tag{3.27}$$

For convenience, we rewrite Eq. (3.27) by decomposing the upper bound into the components that are independent of the probability allocation vector q^t and those that are not, i.e.,

$$\mathbf{E}\left[\, \Delta\Gamma^{t+1} \,\middle|\, y^t \,\right] \leqslant h(y^t) + \sum_{i=1}^{n} q_i^t \cdot f(y_i^t), \tag{3.28}$$

where $h(y^t) := \frac{4\gamma^2}{n} \cdot \Gamma^t - \frac{\gamma}{n} \cdot \sum_{i=1}^n e^{\gamma y_i^t} + \frac{\gamma}{n} \cdot \sum_{i=1}^n e^{-\gamma y_i^t}$ and $f(y_i^t) := e^{\gamma y_i^t} \cdot \gamma - e^{-\gamma y_i^t} \cdot \gamma$.

Recalling that Two-Choice without noise satisfies condition $\mathcal{C}_1$ with $\epsilon = \frac{1}{2}$, $\delta = \frac{1}{4}$ and condition $\mathcal{C}_2$ with $C = 2$, by Lemma 3.3 the expectation of Γ satisfies the following drop inequality:

Lemma 3.20. *Consider the* Two-Choice *process without noise with probability allocation vector* p *and the potential* $\Gamma := \Gamma(\gamma)$ *with any* $\gamma \in (0,1]$. *Then, there exists a constant* $c > 0$, *such that for any step* $t \geqslant 0$,

$$\mathbf{E}\left[\, \Delta\Gamma^{t+1} \,\middle|\, y^t \,\right] \leqslant h(\tilde{y}^t) + \sum_{i=1}^{n} p_i \cdot f(\tilde{y}_i^t) \leqslant -\frac{\gamma}{32n} \cdot \Gamma^t + \frac{4\gamma^2}{n} \cdot \Gamma^t + c.$$

We will analyze $\Delta\Gamma^{t+1}$ for the g-Adv-Comp setting by relating it to the change $\Delta\Gamma^{t+1}$ for the Two-Choice process without noise. To this end, it will be helpful to define all pairs of ranks of bins (of unequal load), whose comparison is under the control of the adversary:

$$R^t := \left\{ (i,j) \in [n] \times [n] \colon \tilde{y}_j^t < \tilde{y}_i^t \leqslant \tilde{y}_j^t + g \right\}. \tag{3.29}$$

So for each pair $(i,j) \in R^t$, the adversary determines the outcome of the load comparison assuming $\{i,j\}$ are the two bin samples in step $t+1$, which happens with probability $2 \cdot \frac{1}{n} \cdot \frac{1}{n} = \frac{2}{n^2}$. This can be seen as moving a probability of up to $\frac{2}{n^2}$ from bin j to bin i, if we relate the probability allocation vector p (of Two-Choice without noise) to the sorted allocation vector $\tilde{q}^t = \tilde{q}^t(\mathfrak{F}^t)$ of Two-Choice with noise.

Theorem 3.21. *Consider the* g-Adv-Comp *setting with any* $g \geqslant 1$, *and the potential* $\Gamma := \Gamma(\gamma)$ *with* $\gamma := -\log(1 - \frac{1}{17 \cdot 32})/g < \frac{1}{16 \cdot 32}$. *Then, there exist constants* $c_1 \geqslant 1, c_2 > 0, c_3 \geqslant 2$, *such that the following three statements hold for all steps* $t \geqslant 0$:

$$\begin{aligned} (i) \qquad & \mathbf{E}\left[\, \Delta\Gamma^{t+1} \,\middle|\, y^t \,\right] \leqslant -\frac{\gamma}{64n} \cdot \Gamma^t + c_1, \\ (ii) \qquad & \mathbf{E}\left[\, \Gamma^t \,\right] \leqslant c_2 n g, \\ (iii) \qquad & \mathbf{Pr}\left[\, \max_{i \in [n]} \left| y_i^t \right| \leqslant c_3 g \log(ng) \,\right] \geqslant 1 - (ng)^{-14}. \end{aligned}$$

Proof. First statement. Consider the sorted allocation vector $\tilde{q}^t$ at step t in the g-Adv-Comp setting. By Eq. (3.28) we have

$$\mathbf{E}\left[\, \Delta\Gamma^{t+1} \,\middle|\, y^t \,\right] \leqslant h(\tilde{y}^t) + \sum_{i=1}^{n} \tilde{q}_i^t \cdot f(\tilde{y}_i^t).$$

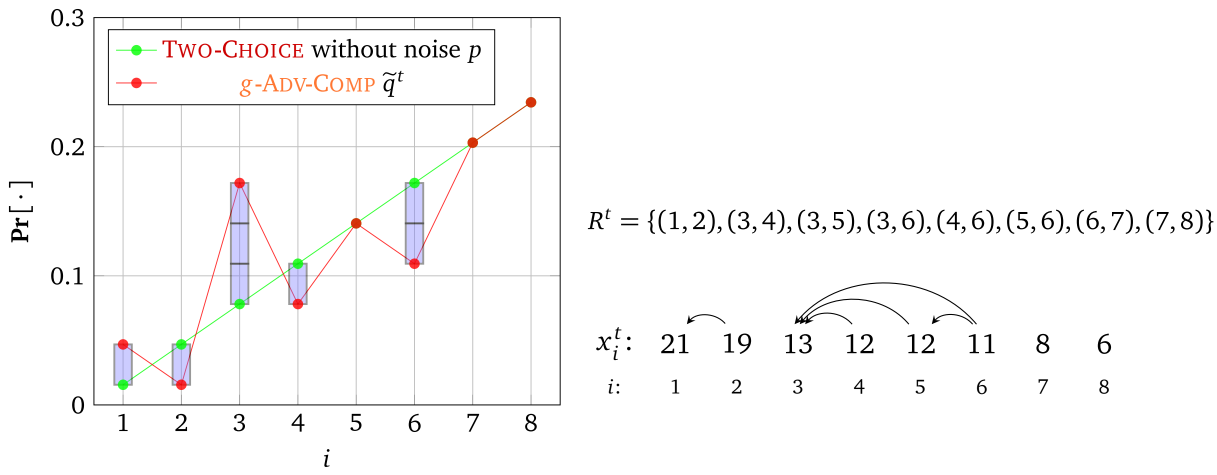


Figure 3.5: Illustration of the set R^t and the change in the sorted allocation vector from p to $\widetilde{q}^t$, where $n = 8$ and $g = 3$. In the example, each directed arrow moves a probability of $\frac{2}{n^2}$ (indicated by the blue rectangles) from a bin j to a heavier bin $i < j$. Note that in this example, the adversary decides not to reverse some of the comparisons, e.g., between bins 7 and 8.

Recall that p is the probability allocation vector of Two-Choice without noise. Then,

$$\widetilde{q}^t := p + \sum_{(i,j)\in R^t} \big(\mathbf{e}_i - \mathbf{e}_j\big) \cdot \gamma_{i,j}^t + \sum_{(i,j)\in[n]\times[n]\colon\, \widetilde{y}_i^t = \widetilde{y}_j^t} \big(\mathbf{e}_i - \mathbf{e}_j\big) \cdot \gamma_{i,j}^t,$$

where $\mathbf{e}_i$ is the i-th unit vector, and $\gamma_{i,j}^t$ is a number in $\big[0, \frac{2}{n^2}\big]$. Hence,

$$\begin{aligned}
\mathbf{E}\big[\, \Delta\Gamma^{t+1} \,\big|\, y^t \,\big] &\leqslant h(\widetilde{y}^t) + \sum_{i=1}^n p_i \cdot f(\widetilde{y}_i^t) + \sum_{(i,j)\in R^t} \gamma_{i,j}^t \cdot \big(f(\widetilde{y}_i^t) - f(\widetilde{y}_j^t)\big) \\
&\qquad + \sum_{(i,j)\in[n]\times[n]\colon\, \widetilde{y}_i^t = \widetilde{y}_j^t} \gamma_{i,j}^t \cdot \big(f(\widetilde{y}_i^t) - f(\widetilde{y}_j^t)\big) \\
&= h(\widetilde{y}^t) + \sum_{i=1}^n p_i \cdot f(\widetilde{y}_i^t) + \sum_{(i,j)\in R^t} \gamma_{i,j}^t \cdot \big(f(\widetilde{y}_i^t) - f(\widetilde{y}_j^t)\big) \\
&\leqslant -\frac{\gamma}{32n} \cdot \Gamma^t + \frac{4\gamma^2}{n} \cdot \Gamma^t + c + \sum_{(i,j)\in R^t} \gamma_{i,j}^t \cdot \big(f(\widetilde{y}_i^t) - f(\widetilde{y}_j^t)\big),
\end{aligned} \tag{3.30}$$

using in the last inequality that by Lemma 3.20 there exists such a constant $c > 0$ for the Two-Choice process without noise and for the same γ.

For any $(i,j) \in R^t$, we define

$$\xi_{i,j}^t := \gamma_{i,j}^t \cdot \big(f(\widetilde{y}_i^t) - f(\widetilde{y}_j^t)\big) \leqslant \frac{2\gamma}{n^2} \cdot \big(e^{\gamma \widetilde{y}_i^t} - e^{-\gamma \widetilde{y}_i^t} - e^{\gamma \widetilde{y}_j^t} + e^{-\gamma \widetilde{y}_j^t}\big),$$

and proceed to upper bound $\xi_{i,j}^t$, using the following lemma, which is based on a case distinction and Taylor estimates for $\exp(\cdot)$.

Lemma 3.22. *For any pair of indices $(i,j) \in R^t$, we have $\xi_{i,j}^t \leqslant \frac{\gamma}{128n^2} \cdot (\Gamma_i^t + \Gamma_j^t) + \frac{16}{n^2}$.*

Proof of Lemma 3.22. Recall that for any $(i,j) \in R^t$ we have that $\tilde{y}_j^t < \tilde{y}_i^t \leqslant \tilde{y}_j^t + g$. So, now we consider the following three disjoint cases:

Case 1 [$\tilde{y}_i^t > g$]: In this case, we also have that $\tilde{y}_j^t > 0$, so

$$\xi_{i,j}^t \overset{(a)}{\leqslant} \frac{2\gamma}{n^2} \cdot \left(e^{\gamma \tilde{y}_i^t} - e^{\gamma \tilde{y}_j^t} + 1\right) \overset{(b)}{\leqslant} \frac{2\gamma}{n^2} \cdot \left(e^{\gamma \tilde{y}_i^t} - e^{\gamma(\tilde{y}_i^t - g)} + 1\right) = \frac{2\gamma}{n^2} \cdot \left(e^{\gamma \tilde{y}_i^t} \cdot (1 - e^{-\gamma g}) + 1\right),$$

using in (a) that $e^{-\gamma \tilde{y}_k^t} \leqslant 1$ for any bin $k \in [n]$ with $\tilde{y}_k^t > 0$, and in (b) that $\tilde{y}_i^t \leqslant \tilde{y}_j^t + g$.

Case 2 [$\tilde{y}_j^t < -g$]: In this case, we also have that $\tilde{y}_i^t < 0$,

$$\xi_{i,j}^t \overset{(a)}{\leqslant} \frac{2\gamma}{n^2} \cdot \left(-e^{-\gamma \tilde{y}_i^t} + e^{-\gamma \tilde{y}_j^t} + 1\right) \overset{(b)}{\leqslant} \frac{2\gamma}{n^2} \cdot \left(e^{-\gamma(\tilde{y}_j^t + g)} - e^{-\gamma \tilde{y}_j^t} + 1\right) = \frac{2\gamma}{n^2} \cdot \left(e^{-\gamma \tilde{y}_j^t} \cdot (e^{-\gamma g} - 1) + 1\right),$$

using in (a) that $e^{\gamma \tilde{y}_k^t} \leqslant 1$ for any bin $k \in [n]$ with $\tilde{y}_k^t < 0$ and in (b) that $\tilde{y}_i^t \leqslant \tilde{y}_j^t + g$.

Case 3 [$\max\{|\tilde{y}_i^t|, |\tilde{y}_j^t|\} \leqslant g$]: In this case we have that

$$\xi_{i,j}^t \leqslant \frac{2\gamma}{n^2} \cdot (\Gamma_i^t + \Gamma_j^t) \leqslant \frac{2}{n^2} \cdot (2 \cdot 2 \cdot e^{\gamma g}) \leqslant \frac{2\gamma}{n^2} \cdot 8,$$

using that by definition of γ, we have $e^{\gamma g} = e^{-\log(1 - \frac{1}{17 \cdot 32})} < 2$.

Combining the upper bounds for the three cases, we have that for any $(i,j) \in R^t$,

$$\begin{aligned}
\xi_{i,j}^t &\leqslant \frac{2\gamma}{n^2} \cdot \left(e^{\gamma \tilde{y}_i^t} \cdot (1 - e^{-\gamma g}) + e^{-\gamma \tilde{y}_j^t} \cdot (1 - e^{-\gamma g}) + 8\right) \\
&\leqslant \frac{2\gamma}{n^2} \cdot \left(\Gamma_i^t \cdot (1 - e^{-\gamma g}) + \Gamma_j^t \cdot (1 - e^{-\gamma g}) + 8\right) \\
&\leqslant \frac{2\gamma}{n^2} \cdot \left(\Gamma_i^t + \Gamma_j^t\right) \cdot \frac{1}{4 \cdot 32} + \frac{16}{n^2},
\end{aligned}$$

using in the last inequality that $\gamma = -\log(1 - \frac{1}{17 \cdot 32})/g$ and $\gamma \leqslant 1$. $\square$

We continue with the proof of Theorem 3.21. By Lemma 3.22 and Eq. (3.30), we have

$$\begin{aligned}
\mathbf{E}\left[\, \Delta\Gamma^{t+1} \,\middle|\, y^t \,\right] &\leqslant -\frac{\gamma}{32n} \cdot \Gamma^t + \frac{4\gamma^2}{n} \cdot \Gamma^t + c + \sum_{(i,j) \in R^t} \left(\frac{\gamma}{4 \cdot 32 n^2} \cdot (\Gamma_i^t + \Gamma_j^t) + \frac{16}{n^2}\right) \\
&\overset{(a)}{\leqslant} -\frac{\gamma}{32n} \cdot \Gamma^t + \frac{4\gamma^2}{n} \cdot \Gamma^t + c + \frac{\gamma}{4 \cdot 32n} \cdot \Gamma^t + 16 \\
&\overset{(b)}{\leqslant} -\frac{\gamma}{32n} \cdot \Gamma^t + \frac{\gamma}{2 \cdot 32n} \cdot \Gamma^t + c + 16 \\
&\leqslant -\frac{\gamma}{64n} \cdot \Gamma^t + c_1, \qquad (3.31)
\end{aligned}$$

for $c_1 := c + 16 \geqslant 1$, where (a) holds since if $(i,j) \in R^t$ then $(j,i) \notin R^t$, so every bin $k \in [n]$ appears at most n times in R^t and in (b) that $\gamma \leqslant \frac{1}{16 \cdot 32}$. This concludes the proof of the first statement.

Second statement. By Lemma B.1 (ii) (for $a = 1 - \frac{\gamma}{64n}$ and $b = c_1$), since $\Gamma^0 = 2n \leqslant \frac{64c_1}{\gamma} \cdot n$ (as $c_1 \geqslant 1$ and $\gamma \leqslant 1$) and Eq. (3.31) hold, it follows that

$$\mathbf{E}\left[\, \Gamma^t \,\right] \leqslant \frac{64c_1}{\gamma} \cdot n =: c_2 n g.$$

Third statement. Using Markov's inequality, for any step $t \geqslant 0$,

$$\mathbf{Pr}\left[\, \Gamma^t \leqslant c_2 \cdot (ng)^{15} \,\right] \geqslant 1 - (ng)^{-14}.$$

When the event $\left\{\Gamma^t \leqslant c_2 \cdot (ng)^{15}\right\}$ holds, we have that

$$\max_{i \in [n]} \left|y_i^t\right| \leqslant \frac{1}{\gamma} \cdot (\log c_2 + 15 \log(ng)) \leqslant \frac{16 \log(ng)}{\gamma} =: c_3 g \log(ng),$$

for sufficiently large n and for the constant

$$c_3 := \frac{16}{\gamma g} = \frac{16}{-\log(1 - \frac{1}{17 \cdot 32})} \geqslant 2. \tag{3.32}$$

Therefore we conclude that,

$$\mathbf{Pr}\left[\, \max_{i \in [n]} \left|y_i^t\right| \leqslant c_3 g \log(ng) \,\right] \geqslant 1 - (ng)^{-14}. \qquad \square$$

Next we will also state a simple corollary that starting with a "small" gap, in any future step, w.h.p. the gap will be small. This corollary will be used in obtaining the tighter $\mathcal{O}(g + \log n)$ gap bound in Section 5.3.

Corollary 3.23. *Consider the* g-Adv-Comp *setting for any* $g \geqslant 1$. *Then, for any steps* $t_0 \geqslant 0$ *and* $t_1 \geqslant t_0$, *we have that*

$$\mathbf{Pr}\left[\, \max_{i \in [n]} \left|y_i^{t_1}\right| \leqslant 2g(\log(ng))^2 \,\middle|\, \mathfrak{F}^{t_0}, \max_{i \in [n]} \left|y_i^{t_0}\right| \leqslant g(\log(ng))^2 \,\right] \geqslant 1 - (ng)^{14}.$$

Proof. We will be using the hyperbolic cosine potential $\Gamma := \Gamma(\gamma)$ with smoothing parameter $\gamma := -\log(1 - \frac{1}{17 \cdot 32})/g$ as we did in the proof of Theorem 3.21.

Consider an arbitrary step t_0 with $\max_{i \in [n]} \left|y_i^{t_0}\right| \leqslant g(\log(ng))^2$. Then, it follows that

$$\Gamma^{t_0} \leqslant 2n \cdot e^{\gamma g (\log(ng))^2}.$$

By Theorem 3.21 (i), there exists a constant $c_1 \geqslant 1$ such that

$$\mathbf{E}\left[\, \Gamma^{t+1} \,\middle|\, \mathfrak{F}^t \,\right] \leqslant \left(1 - \frac{\gamma}{64n}\right) + c_1,$$

and using Lemma B.1 (i) (for $a := 1 - \frac{\gamma}{64n}$ and $b := c_1$) at step $t_1 \geqslant t_0$, we have that

$$\mathbf{E}\left[\, \Gamma^{t_1} \,\middle|\, \mathfrak{F}^{t_0}, \Gamma^{t_0} \leqslant 2n \cdot e^{\gamma g (\log(ng))^2} \,\right] \leqslant 2n \cdot e^{\gamma g (\log(ng))^2} \cdot \left(1 - \frac{\gamma}{64n}\right)^{t_1 - t_0} + \frac{64 c_1 n}{\gamma} \leqslant 4n \cdot e^{\gamma g (\log(ng))^2},$$

recalling that $\gamma = \Theta\left(\frac{1}{g}\right)$. Hence, by Markov's inequality, we have that

$$\begin{aligned}
&\mathbf{Pr}\left[\, \Gamma^{t_1} \leqslant 4n \cdot e^{\gamma g (\log(ng))^2} \cdot (ng)^{14} \,\middle|\, \mathfrak{F}^{t_0}, \max_{i \in [n]} \left|y_i^{t_0}\right| \leqslant g(\log(ng))^2 \,\right] \\
&\quad \geqslant \mathbf{Pr}\left[\, \Gamma^{t_1} \leqslant 4n \cdot e^{\gamma g (\log(ng))^2} \cdot (ng)^{14} \,\middle|\, \mathfrak{F}^{t_0}, \Gamma^{t_0} \leqslant 2n \cdot e^{\gamma g (\log(ng))^2} \,\right] \geqslant 1 - (ng)^{-14}.
\end{aligned}$$

When the event $\left\{\Gamma^{t_1} \leqslant 4n \cdot e^{\gamma g (\log(ng))^2} \cdot (ng)^{14}\right\}$ holds, then it also follows that

$$\max_{i \in [n]} \left|y_i^{t_1}\right| \leqslant \frac{\log \Gamma^{t_1}}{\gamma} \leqslant g(\log(ng))^2 + \frac{1}{\gamma} \cdot (\log(4n) + 14 \log(ng)) \leqslant 2g(\log(ng))^2,$$

for sufficiently large n and using that $\gamma = \Theta\left(\frac{1}{g}\right)$. Hence, we get the conclusion. $\square$

Chapter $3^{1/2}$: Physical balanced allocations

In the spirit of the Galton board [85], we can construct physical automata that "sample" distributions of various balanced allocation processes. The simplest process to simulate is the ONE-CHOICE process, where balls flow from the top of the board and then they encounter a binary branching structure, where at each junction the ball is assumed to randomly pick one of the two disjoint balanced subsets of the bins. After encountering $\log_2 n$ such junctions, it reaches one of the n bins. See Fig. 3.6 for the blueprints and laser cut version of the automaton.

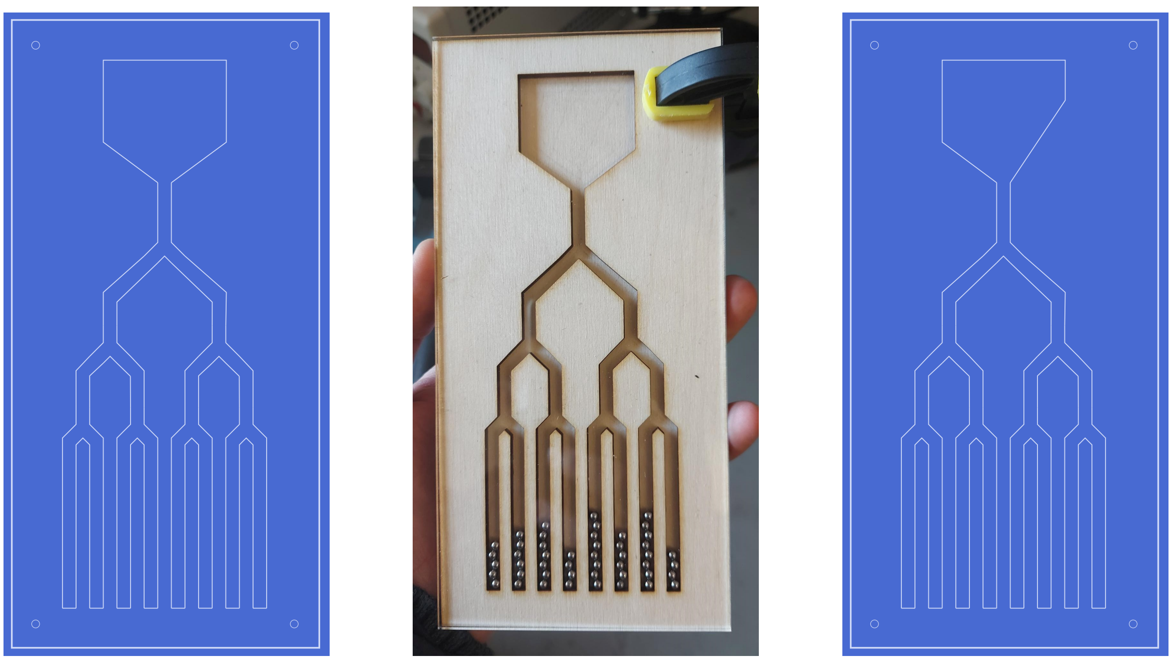


Figure 3.6: On the left is the blueprint for the ONE-CHOICE process, in the middle is the actual implementation of this blueprint and on the right is the blueprint of a modification that prevents balls being stuck in the top junction.

For more complicated processes, these automata become more difficult to construct. Fig. 3.7 shows the blueprint for a general TWO-THINNING process, where the decision function is encoded using a knob for each bin, which determines if that bin should accept a ball at a first allocation (and the ball is directed through the dashed lines to the respective bin) or whether it should be allocated to a second random sample, which is done by passing through a second ONE-CHOICE structure.

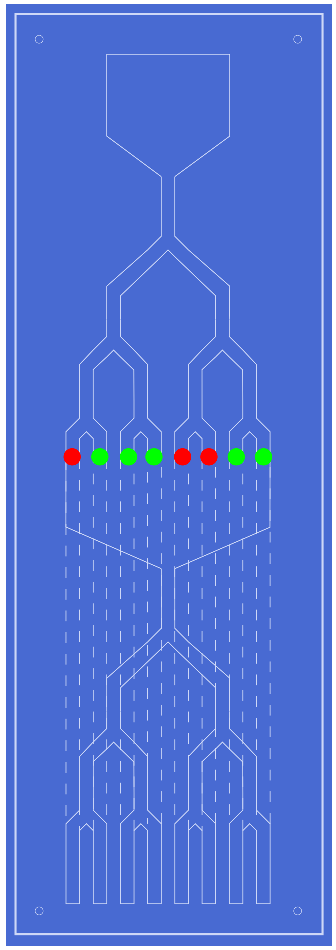


Figure 3.7: Blueprint for the TWO-THINNING processes, where the red/green circles correspond to knobs directing the incoming balls either directly to the bins below (dashed path) or to a second ONE-CHOICE allocation (solid path), corresponding to allocating to the second sample.

Concentration of the hyperbolic cosine potential

Recall the hyperbolic cosine potential $\Gamma := \Gamma(\gamma)$ was defined in Eq. (3.1) as

$$\Gamma^t := \Gamma^t(\gamma) := \sum_{i=1}^{n} e^{\gamma y_i^t} + \sum_{i=1}^{n} e^{-\gamma y_i^t}.$$

In Chapter 3, we saw that for a wide family of processes the hyperbolic cosine potential Γ with a suitable smoothing parameter $\gamma \leqslant 1$ is $\mathcal{O}(n)$ in expectation. By Markov's inequality this expectation bound implies that w.h.p. $\Gamma^t = \mathrm{poly}(n)$, for any step t. This is enough for showing that w.h.p. the gap is $\mathcal{O}(\log n/\gamma)$, for $\gamma \leqslant 1$. The main focus of this chapter is to strengthen this, by obtaining concentration bounds for the hyperbolic cosine potential function, i.e., proving that w.h.p. $\Gamma^t = \mathcal{O}(n)$ for any step t.

The reason why high probability bounds are useful was outlined in Section 1.3. We will repeat the motivating factors here and make connections to Chapter 6. For some processes (e.g., Two-Choice, k-Dense-Quantile, etc), our goal is to prove w.h.p. a tighter $o(\log n)$ gap. In order to do this, we use exponential potentials with smoothing parameters $\gamma = \omega(1)$, the so-called *super-exponential potentials*. Unlike the hyperbolic cosine potential these potentials may not always decrease in expectation even if large. However, in Chapter 6 we show that in steps t where the number of bins $N_{\geqslant v}^t$ with normalised load at least some value v is small, then these potentials decrease in expectation. In order to bound $N_{\geqslant v}^t$, we use the concentration of Γ, i.e., that w.h.p. $\Gamma^t = \mathcal{O}(n)$.[1] More specifically, in any step t with $\Gamma^t \leqslant cn$, we can deduce that

$$N_{\geqslant v}^t \leqslant cn \cdot e^{-\gamma v}.$$

In Chapter 7, we show how these can be applied to obtain bounds for the Quantile(δ^*), k-Dense-Quantile, k-Dense-Threshold and g-Adv(Two-Choice, G^t) processes.[2]

We start by stating the main theorem for the concentration of the hyperbolic cosine potential. The rest of the chapter is devoted to proving this theorem.

Theorem 4.1. *Consider any balanced allocation process $\mathcal{P}$, where in each step at most $d \in \mathbb{N}_+$ balls are allocated (where d is not necessarily constant) and consider an arbitrary constant $\kappa \geqslant 6$. Further, assume for this process that for the hyperbolic potential functions $\Gamma_1 := \Gamma_1(\gamma_1)$ and $\Gamma_2 := \Gamma_2(\gamma_2)$ with smoothing parameters $\gamma_1 \in (0, 1/(2d))$ and $\gamma_2 \leqslant \frac{\gamma_1}{12\kappa}$ respectively, there exists an $\epsilon > 0$ (with $\gamma_2 \epsilon \geqslant n^{-1/6}$) and constants $c_1, c_2 > 0$ (with $c_1 \leqslant c_2$), such that for any step $t \geqslant 0$,*

$$\mathbf{E}\left[\, \Gamma_1^{t+1} \,\middle|\, \mathfrak{F}^t \,\right] \leqslant \Gamma_1^t \cdot \left(1 - \frac{c_1 \gamma_1 \epsilon}{n}\right) + c_2 \gamma_1 \epsilon, \tag{4.1}$$

and

$$\mathbf{E}\left[\, \Gamma_2^{t+1} \,\middle|\, \mathfrak{F}^t \,\right] \leqslant \Gamma_2^t \cdot \left(1 - \frac{c_1 \gamma_2 \epsilon}{n}\right) + c_2 \gamma_2 \epsilon. \tag{4.2}$$

Then, for $c := 2 \cdot \frac{c_2}{c_1} \geqslant 2$ and for any step $t \geqslant 0$,

$$\mathbf{Pr}\left[\, \Gamma_2^t \leqslant 3cn \,\right] \geqslant 1 - n^{-\kappa}.$$

[1] See Section 7.1.1 for a discussion about some results that can be deduced using just Markov's inequality with a weaker probability guarantee.

[2] For the last two processes, the layered induction over super-exponential potentials is very similar to that of k-Dense-Quantile. However, the concentration for the hyperbolic cosine potential is a bit more involved, making use of the absolute value and quadratic potentials, and so will be the main focus of Chapter 5.

4.1 Proof outline of Theorem 4.1

In this section, we will outline the proof of Theorem 4.1, giving some intuition for the requirement/choice of the two potential functions Γ_1 and Γ_2.

Our goal is to show that w.h.p. $\Gamma_2^t \leqslant 3cn$, for any given $t \geqslant 0$. We will prove this by analysing the steps in the interval $[t - T_r, t]$, where $T_r := \left\lceil 2 \cdot \frac{4/3+2\kappa}{c_1\gamma_2\epsilon} \cdot n \log n \right\rceil$. In particular, in this interval, which we call the *recovery interval*, we will show that w.h.p. $\Gamma_2^r \leqslant cn$ for at least *one* step $r \in [t - T_r, t]$ and then we will show that it *stabilises*, i.e., remains small, for *all* steps in $[r, t]$.

Now, we will give a few more details for the steps in the proof (see Fig. 4.1). By the expectation bound proven in Chapter 3, we have that $\mathbf{E}\left[\, \Gamma_1^t \,\right] \leqslant cn$ for any step $t \geqslant 0$. So, by Markov's inequality w.h.p. $\Gamma_1^s \leqslant cn^{2\kappa+1}$ for all $s \in [t - T_r, t]$.

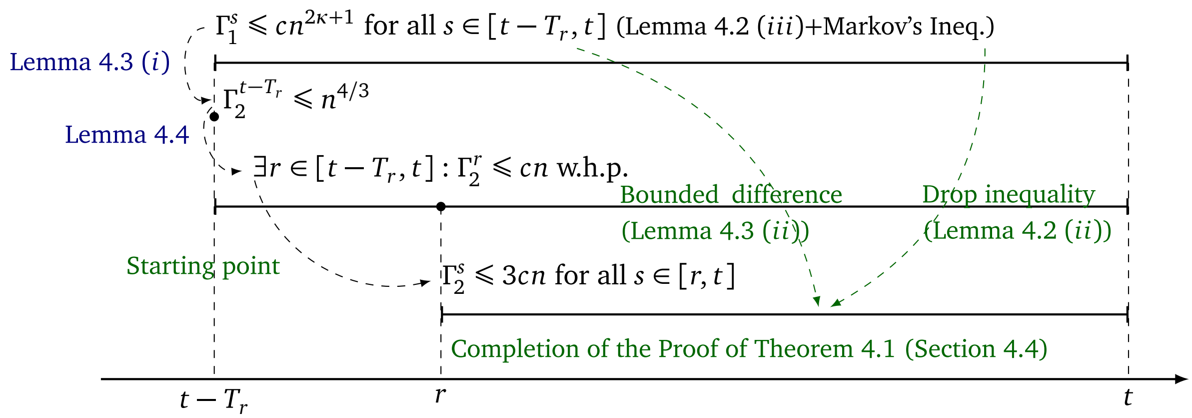


Figure 4.1: Outline for the proof of Theorem 4.1. Results in green are used in the application of Azuma's concentration inequality for super-martingales (Lemma B.10) in Theorem 4.1.

By the choice of $\gamma_2 \leqslant \frac{\gamma_1}{12\kappa}$, we will show that when $\Gamma_1^s \leqslant cn^{2\kappa+1}$, then we also have (i) that $\Gamma_2^s \leqslant n^{4/3}$ and (ii) that $\left|\Gamma_2^{s+1} - \Gamma_2^s\right| \leqslant n^{1/3}$ (Lemma 4.3). The first condition will be useful for proving the *recovery*, i.e., that $\Gamma_2^r \leqslant cn$ for at least one step $r \in [t - T_r, t]$ (Lemma 4.4). Then, starting from this step r and using the second condition allows us to use a concentration inequality to deduce that Γ_2 *stabilises*, i.e., that $\Gamma_2^s \leqslant 3cn$ for all steps $s \in [r, t]$ (Lemma 4.5).

4.2 Auxiliary lemmas

In this section, we will prove some auxiliary lemmas for the potentials $\Gamma_1 := \Gamma_1(\gamma_1)$ and $\Gamma_2 := \Gamma_2(\gamma_2)$ as defined in Theorem 4.1.

Lemma 4.2. *Consider any balanced allocation process $\mathcal{P}$ satisfying the preconditions of Theorem 4.1. Then, for any step $t \geqslant 0$,*

$$
\begin{aligned}
(i) \quad & \mathbf{E}\left[\, \Gamma_1^{t+1} \,\middle|\, \mathfrak{F}^t, \Gamma_1^t > cn \,\right] \leqslant \Gamma_1^t \cdot \left(1 - \frac{c_1\gamma_1\epsilon}{2n}\right), \\
(ii) \quad & \mathbf{E}\left[\, \Gamma_2^{t+1} \,\middle|\, \mathfrak{F}^t, \Gamma_2^t > cn \,\right] \leqslant \Gamma_2^t \cdot \left(1 - \frac{c_1\gamma_2\epsilon}{2n}\right), \\
(iii) \quad & \mathbf{E}\left[\, \Gamma_1^t \,\right] \leqslant cn.
\end{aligned}
$$

Proof. First Statement. Recall that $c = 2 \cdot \frac{c_2}{c_1} \geqslant 2$. For the first statement, by the assumptions

$$\begin{aligned}\mathbf{E}\left[\, \Gamma_1^{t+1} \,\middle|\, \mathfrak{F}^t, \Gamma_1^t > cn \,\right] &\leqslant \Gamma_1^t \cdot \left(1 - \frac{c_1\gamma_1\epsilon}{n}\right) + c_2\gamma_1\epsilon \\ &= \Gamma_1^t \cdot \left(1 - \frac{c_1\gamma_1\epsilon}{2n}\right) - \Gamma_1^t \cdot \frac{c_1\gamma_1\epsilon}{2n} + c_2\gamma_1\epsilon \\ &\leqslant \Gamma_1^t \cdot \left(1 - \frac{c_1\gamma_1\epsilon}{2n}\right) - 2 \cdot \frac{c_2}{c_1} \cdot n \cdot \frac{c_1\gamma_1\epsilon}{2n} + c_2\gamma_1\epsilon = \Gamma_1^t \cdot \left(1 - \frac{c_1\gamma_1\epsilon}{2n}\right).\end{aligned}$$

Second Statement. Similarly, we obtain the second statement for Γ_2.

Third statement. By Lemma B.1 (ii) for $a = 1 - \frac{c_1\gamma_1\epsilon}{n}$ and $b = c_2\gamma_1\epsilon$, since $\Gamma_1^0 = 2n \leqslant 2 \cdot \frac{c_2}{c_1} \cdot n = cn$, it follows that $\mathbf{E}[\, \Gamma^t \,] \leqslant cn$, for any step $t \geqslant 0$. □

Lemma 4.3. *Consider any process $\mathcal{P}$ satisfying the preconditions of Theorem 4.1. For any step $t \geqslant 0$ where $\Gamma_1^t \leqslant cn^{2\kappa+1}$ holds, we have that*

$$\begin{aligned} &(i) \qquad \Gamma_2^t \leqslant n^{4/3}, \\ &(ii) \qquad \left|\Gamma_2^{t+1} - \Gamma_2^t\right| \leqslant n^{1/3}. \end{aligned}$$

Proof. Consider an arbitrary step t where $\Gamma_1^t \leqslant cn^{2\kappa+1}$. We start by proving the following bound on the normalised load y_i^t for any bin $i \in [n]$,

$$\Gamma_1^t \leqslant cn^{2\kappa+1} \Rightarrow e^{\gamma_1 \cdot y_i^t} + e^{-\gamma_1 \cdot y_i^t} \leqslant cn^{2\kappa+1} \Rightarrow y_i^t \leqslant \frac{3\kappa}{\gamma_1} \cdot \log n \wedge -y_i^t \leqslant \frac{3\kappa}{\gamma_1} \cdot \log n,$$

where in the second implication we used $\log c + \frac{2\kappa+1}{\gamma_1} \log n \leqslant \frac{3\kappa}{\gamma_1} \log n$, for sufficiently large n as c is a constant and $\kappa \geqslant 6 \geqslant 1$.

First Statement. Recall that $\gamma_2 \leqslant \gamma_1/12\kappa$. By the definition of Γ_2^t and the bound on each normalised bin load, we get that

$$\Gamma_2^t \leqslant 2 \cdot \sum_{i=1}^{n} \exp\left(\gamma_2 \cdot \frac{3\kappa}{\gamma_1} \cdot \log n\right) = 2n \cdot n^{1/4} \leqslant n^{4/3}.$$

Second Statement. Consider Γ_2^{t+1} as a sum over $2n$ exponentials, which is obtained from Γ_2^t by slightly changing the values of the $2n$ exponents. The total ℓ_1-change in the exponents is upper bounded by $4d$, as we will increment d entries in the load vector x^t (and each of these entries appear twice), and we will also increment the average load by $\frac{d}{n}$ in all $2n$ exponents. Since $\exp(\cdot)$ is convex, the largest change is upper bounded by the (hypothetical) scenario in which the largest exponent increases by $4d$ and all others remain the same,

$$\left|\Gamma_2^{t+1} - \Gamma_2^t\right| \leqslant \exp\left(\gamma_2 \cdot \left(4d + \max_{i \in [n]} \left|y_i^t\right|\right)\right) \leqslant e^{4d\gamma_2} \cdot \exp\left(\gamma_2 \cdot \frac{3\kappa}{\gamma_1} \cdot \log n\right) = e^{4d\gamma_2} \cdot n^{1/4} \leqslant n^{1/3},$$

using that $\gamma_2 \leqslant \frac{\gamma_1}{12\kappa}$ and that $\gamma_2 \leqslant \gamma_1 \leqslant 1/(2d)$. □

4.3 Recovery and stabilisation

Using the second and third statements in Lemma 4.2, we will now prove a weaker statement of Theorem 4.1, showing that $\Gamma_2^r \leqslant cn$ for *at least one* step $r \in [t - T_r, t]$, where T_r is the length of the recovery interval

$$T_r := \left\lceil 2 \cdot \frac{4/3 + 2\kappa}{c_1\gamma_2\epsilon} \cdot n \log n \right\rceil. \tag{4.3}$$

Before we do this, we proceed by defining an auxiliary process.

Auxiliary process. Let $\mathcal{P}$ be the process satisfying the preconditions of Theorem 4.1. We want to condition that $\mathcal{P}$ has $\Gamma^s \leqslant cn^{2\kappa+1}$ for every step s in an interval of poly(n) length, so that we can deduce it satisfies the bounded difference condition (Lemma 4.2) and then apply Azuma's inequality (Lemma B.10).

To this end, we will define an auxiliary process $\widetilde{\mathcal{P}}_{t_0} := \widetilde{\mathcal{P}}_{t_0}(\mathcal{P})$ for some arbitrary step $t_0 \geqslant 0$. Let $\sigma := \inf\{s \geqslant t_0 : \Gamma^s > \frac{1}{2}cn^{2\kappa+1}\}$. Then, we define $\widetilde{\mathcal{P}}_{t_0}$ so that

- in steps $[0, \sigma)$ it makes the same allocations as $\mathcal{P}$, and
- in steps $[\sigma, \infty)$ it allocates to the currently least loaded bin, i.e., it uses the sorted probability allocation vector $\widetilde{q}^s = (0, \ldots, 0, 1)$.

Let $y^s_{\widetilde{\mathcal{P}}}$ be the normalised load vector of $\widetilde{\mathcal{P}}_{t_0}$ at step $s \geqslant 0$. By Lemma 4.2 (iii), Markov's inequality and the union bound, it follows that for any interval $[t_0, m]$ with $m - t_0 \leqslant T_r$, with high probability the two processes agree,

$$\mathbf{Pr}\left[\bigcap_{s\in[t_0,m]}\left\{y^s_{\widetilde{\mathcal{P}}} = y^s\right\}\right] \geqslant \mathbf{Pr}\left[\bigcap_{s\in[t_0,m]}\left\{\Gamma_1^s \leqslant cn^{2\kappa+1}\right\}\right] \geqslant 1 - n^{-2\kappa} \cdot T_r. \tag{4.4}$$

The process $\widetilde{\mathcal{P}}_{t_0}$ is defined in this way to satisfy the following property:

- (**Property 1**) The $\widetilde{\mathcal{P}}_{t_0}$ process satisfies the drop inequalities for the potential functions $\Gamma_{1,\widetilde{\mathcal{P}}}$ and $\Gamma_{2,\widetilde{\mathcal{P}}}$ (preconditions (4.1) and (4.2)) for any step $s \geqslant 0$. This holds because for any step $s < \sigma$, the process follows $\mathcal{P}$. For any step $s \geqslant \sigma$, the process allocates to the currently least loaded bin and therefore minimises the potential $\Gamma^{s+1}_{1,\widetilde{\mathcal{P}}}$ given any $\mathfrak{F}^s$, which means that $\Gamma^{s+1}_{1,\widetilde{\mathcal{P}}} \leqslant \mathbf{E}\left[\,\Gamma_1^{s+1} \mid \mathfrak{F}^s\,\right]$ and so it trivially satisfies any drop inequality (and similarly for $\Gamma_{2,\widetilde{\mathcal{P}}}$).

Further, we define the event that the potential Γ_1 is small at step t_0, as

$$\mathcal{Z}^{t_0} := \left\{\Gamma^{t_0}_{1,\widetilde{\mathcal{P}}} \leqslant \frac{1}{2}cn^{2\kappa+1}\right\}, \tag{4.5}$$

where $c \geqslant 1$ is the constant defined in Theorem 4.1. When the event $\mathcal{Z}^{t_0}$ holds, then the process $\widetilde{\mathcal{P}}_{t_0}$ also satisfies the following property (which "implements" the conditioning that $\Gamma^s_{1,\widetilde{\mathcal{P}}} \leqslant cn^{2\kappa+1}$):

- (**Property 2**) For any step $s \geqslant t_0$, it follows that
$$\Gamma^s_{1,\widetilde{\mathcal{P}}} \leqslant cn^{2\kappa+1}.$$
At any step $s \in [t_0, \sigma)$, this holds by the definition of σ. For any step $s \geqslant \sigma$, a ball will never be allocated to a bin with $y_i^s > 0$ and in every n steps the at most n bins with load equal to the minimum load (at step s) will be allocated at least one ball each. Hence, over any n steps the maximum absolute normalised load does not increase and in the steps in between this can be larger by at most d and hence,
$$\Gamma^s_{1,\widetilde{\mathcal{P}}} \leqslant e^{\gamma_1 d} \cdot \Gamma^\sigma_{1,\widetilde{\mathcal{P}}} \leqslant e^{\gamma_1 d} \cdot \frac{1}{2}cn^{2\kappa+1} \leqslant cn^{2\kappa+1}.$$

Lemma 4.4 (Recovery). *Consider any step $t \geqslant 0$ and the auxiliary process $\widetilde{\mathcal{P}}_{t-T_r} := \widetilde{\mathcal{P}}_{t-T_r}(\mathcal{P})$ for any $\mathcal{P}$ satisfying the preconditions of Theorem 4.1 and with $\mathcal{Z}^{t-T_r}$ being the event defined in Eq.* (4.5). *For any step $t \geqslant 0$,*

$$\mathbf{Pr}\left[\bigcup_{r\in[t-T_r,t]}\left\{\Gamma_2^r \leqslant cn\right\} \,\middle|\, \mathfrak{F}^{t-T_r}, \mathcal{Z}^{t-T_r}\right] \geqslant 1 - 2n^{-2\kappa-1}.$$

Proof. When $t < T_r$, then deterministically $\Gamma_2^r = 2n \leqslant cn$ for $r = 0$ and so the statement follows. Otherwise, by the condition $\mathcal{Z}^{t-T_r}$, we have that $\{\Gamma_1^{t-T_r} \leqslant cn^{2\kappa+1}\}$ holds. By Lemma 4.3 (i), this implies that $\{\Gamma_2^{t-T_r} \leqslant n^{4/3}\}$ also holds.

By Lemma 4.2 (ii), for any step $s \geqslant 0$,

$$\mathbf{E}\left[\, \Gamma_2^{s+1} \,\middle|\, \mathfrak{F}^s, \Gamma_2^s > cn \,\right] \leqslant \Gamma_2^s \cdot \left(1 - \frac{c_1\gamma_2\epsilon}{2n}\right). \tag{4.6}$$

Next, we define the "killed" potential function at steps $s \geqslant t - T_r$ as

$$\widehat{\Gamma}_2^s := \Gamma_2^s \cdot \mathbf{1}_{\bigcap_{r\in[t-T_r,s]}\{\Gamma_2^r > cn\}}.$$

Note that when $\{\Gamma_1^s \leqslant cn\}$ then also $\{\widehat{\Gamma}_1^s = 0\}$ and $\{\widehat{\Gamma}_1^{s+1} = 0\}$. Therefore, the $\widehat{\Gamma}$ potential unconditionally satisfies the inequality of Eq. (4.6), that is for any $s \geqslant t - T_r$

$$\mathbf{E}\left[\, \widehat{\Gamma}_2^{s+1} \,\middle|\, \mathfrak{F}^s \,\right] \leqslant \widehat{\Gamma}_2^s \cdot \left(1 - \frac{c_1\gamma_2\epsilon}{2n}\right).$$

Inductively applying this for T_r steps, starting with $\widehat{\Gamma}_2^{t-T_r} \leqslant \Gamma_2^{t-T_r} \leqslant n^{4/3}$, we get

$$\mathbf{E}\left[\, \widehat{\Gamma}_2^t \,\middle|\, \mathfrak{F}^{t-T_r}, \mathcal{Z}^{t-T_r} \,\right] \leqslant \mathbf{E}\left[\, \widehat{\Gamma}_2^t \,\middle|\, \mathfrak{F}^{t-T_r}, \Gamma_2^{t-T_r} \leqslant n^{4/3} \,\right] \leqslant n^{4/3} \cdot \left(1 - \frac{c_1\gamma_2\epsilon}{2n}\right)^{T_r} \overset{(a)}{\leqslant} n^{4/3} \cdot e^{-\frac{1}{2}c_1\gamma_2\epsilon\cdot\frac{T_r}{n}} \overset{(b)}{\leqslant} n^{-2\kappa},$$

using in (a) that $1 + u \leqslant e^u$ (for any u) and in (b) that $T_r = \left\lceil 2 \cdot \frac{4/3+2\kappa}{c_1\gamma_2\epsilon} \cdot n\log n \right\rceil$. So, by Markov's inequality,

$$\mathbf{Pr}\left[\, \widehat{\Gamma}_2^t \leqslant n \,\middle|\, \mathfrak{F}^{t-T_r}, \mathcal{Z}^{t-T_r} \,\right] \geqslant 1 - n^{-2\kappa-1}.$$

Since deterministically $\Gamma_2^s \geqslant 2n$ at any step s, we conclude that when $\{\widehat{\Gamma}_2^t \leqslant n\}$, then also $\{\widehat{\Gamma}_2^t = 0\}$, and so

$$\mathbf{1}_{\bigcap_{r\in[t-T_r,t]}\{\Gamma_2^r > cn\}} = 0, \text{ implying that } \neg\bigcap_{r\in[t-T_r,t]}\{\Gamma_2^r > cn\} \text{ holds}$$

with probability at least $1 - 2n^{-2\kappa}$, concluding the claim. □

We will now show that whenever $\Gamma_2^r \in [cn, 2cn]$ holds for some step $r \in [t - T_r, t]$, then with high probability (i) it remains *small* until step t, i.e., $\Gamma_2^s \leqslant 3cn$ for *all* $s \in [r, t]$ or (ii) it remains small until some step $s \leqslant t$ where it becomes *very small*, i.e., $\Gamma_2^s \leqslant cn$.

Lemma 4.5 (Stabilisation). *Consider any step $t \geqslant 0$ and the auxiliary process $\widetilde{\mathcal{P}}_{t-T_r} := \widetilde{\mathcal{P}}_{t-T_r}(\mathcal{P})$ for any $\mathcal{P}$ satisfying the preconditions of Theorem 4.1 and with $\mathcal{Z}^{t-T_r}$ being the event defined in Eq.* (4.5). *Then, for any step $r \in [t - T_r, t]$ for T_r as defined in Eq.* (4.3),

$$\mathbf{Pr}\left[\, \bigcap_{s\in[r,t]} \{\Gamma_2^s \leqslant 3cn\} \cup \bigcup_{s\in(r,t]} \left(\bigcap_{u\in[r,s]} \{\Gamma_2^u \leqslant 3cn\} \cap \{\Gamma_2^s \leqslant cn\} \right) \,\middle|\, \mathcal{Z}^{t-T_r}, \mathfrak{F}^r, \Gamma_2^r \in [cn, 2cn] \,\right] \geqslant 1 - n^{-\frac{4}{3}\kappa}.$$

Proof. Consider an arbitrary step $r \in [t - T_r, t]$ such that $\Gamma_2^r \in [cn, 2cn]$. We define the stopping time

$$\tau := \inf\left\{\widetilde{r} > r : \Gamma_2^{\widetilde{r}} \leqslant cn\right\},$$

and the sequence X_r^s for any step $s \in [r, t]$,

$$X_r^s := \Gamma_2^{s\wedge\tau}.$$

We defined $(X_r^s)_{s\in[r,t]}$ this way so that it forms a super-martingale. To see this, note that by Lemma 4.2 (ii), for any step $s<\tau$ we have that

$$\mathbf{E}\left[\, X_r^{s+1} \,\middle|\, \mathfrak{F}^s, s<\tau \,\right] \leqslant \mathbf{E}\left[\, \Gamma_2^{s+1} \,\middle|\, \mathfrak{F}^s, \Gamma_2^s \geqslant cn \,\right] \leqslant \Gamma_2^s, \tag{4.7}$$

and for any step $s\geqslant\tau$,

$$\mathbf{E}\left[\, X_r^{s+1} \,\middle|\, \mathfrak{F}^s, s\geqslant\tau \,\right] = X_r^s. \tag{4.8}$$

Recall that when $\mathcal{Z}^{t-T_r}$ holds, then by Property 2 (see Section 4.3), it holds that $\Gamma_1^s \leqslant cn^{2\kappa+1}$ for every step $s\geqslant t-T_r$. So, by Lemma 4.3 (ii) it also holds that $\left|\Gamma_2^{s+1}-\Gamma_2^s\right| \leqslant n^{1/3}$.

Hence, applying Azuma's inequality (Lemma B.10) for any $s\in[r,t]$ gives

$$\mathbf{Pr}\left[\, X_r^s \geqslant X_r^r + cn \,\middle|\, \mathcal{Z}^{t-T_r}, \mathfrak{F}^r, \Gamma_2^r \in [cn, 2cn] \,\right] \leqslant \exp\left(-\frac{c^2n^2}{2\cdot T_r\cdot (n^{1/3})^2}\right) \leqslant 2\cdot T_r\cdot n^{-2\kappa},$$

using that $T_r=\mathcal{O}(n\cdot n^{1/6}\cdot \log n)$. Also recall that at the starting point r, it holds that $X_r^r=\Gamma_2^r\leqslant 2cn$. Hence, we can conclude that

$$\mathbf{Pr}\left[\, X_r^s > 3cn \,\middle|\, \mathcal{Z}^{t-T_r}, \mathfrak{F}^r, \Gamma_2^r \in [cn, 2cn] \,\right] \leqslant 2\cdot T_r\cdot n^{-2\kappa}.$$

By taking the union bound over all steps $s\in[r,t]$, we get

$$\mathbf{Pr}\left[\, \bigcap_{s=r}^{t} \left\{X_r^s \leqslant 3cn\right\} \,\middle|\, \mathcal{Z}^{t-T_r}, \mathfrak{F}^r, \Gamma_2^r \in [cn, 2cn] \,\right] \geqslant 1-3\cdot T_r^2\cdot n^{-2\kappa} \geqslant 1-n^{-\kappa},$$

using that $\kappa\geqslant 6$. Now, assuming that $\bigcap_{s=r}^t \left\{X_r^s\leqslant 3cn\right\}$ holds, we consider the following cases based on the stopping time τ:

- **Case 1 $[\tau>t]$:** Then for all steps $u\in[r,t]$, we have that $\Gamma_2^u=X_r^u\leqslant 3cn$.
- **Case 2 $[\tau\leqslant t]$:** Then for all steps $u\in[r,\tau]$, we have that $\Gamma_2^u=X_r^u\leqslant 3cn$ and $\Gamma_2^\tau\leqslant cn$. So the following event holds for $s=\tau>r$,

$$\bigcup_{s\in(r,t]}\left(\bigcap_{u\in[r,s]}\left\{\Gamma_2^u\leqslant 3cn\right\}\cap\left\{\Gamma_2^s\leqslant cn\right\}\right).$$

Hence, this concludes the claim. $\square$

4.4 Completing the proof of Theorem 4.1

Before we complete the proof of Theorem 4.1, we first prove the statement for the auxiliary process $\widetilde{\mathcal{P}}_{t-T_r}$.

Lemma 4.6. *Consider any step $t\geqslant 0$ and the auxiliary process $\widetilde{\mathcal{P}}_{t-T_r}:=\widetilde{\mathcal{P}}_{t-T_r}(\mathcal{P})$ for any $\mathcal{P}$ satisfying the preconditions of Theorem 4.1 and with $\mathcal{Z}^{t-T_r}$ being the event defined in Eq.* (4.5). *Then, for $c:=2\cdot\frac{c_2}{c_1}\geqslant 2$ and for any step $t\geqslant 0$,*

$$\mathbf{Pr}\left[\, \Gamma_2^t \leqslant 3cn \,\middle|\, \mathcal{Z}^{t-T_r} \,\right] \geqslant 1-\frac{1}{2}n^{-\kappa}.$$

Proof. The proof will be concerned with steps $\in [t - T_r, t]$. First, by applying Lemma 4.4, it holds that

$$\mathbf{Pr}\left[\bigcup_{r_0 \in [t-T_r, t]} \left\{ \Gamma_2^{r_0} \leqslant cn \right\} \,\middle|\, \mathfrak{F}^{t-T_r}, \mathcal{Z}^{t-T_r} \right] \geqslant 1 - 2n^{-2\kappa - 1}. \tag{4.9}$$

Consider now an arbitrary step $r_0 \in [t - T_r, t]$ and assume that $\Gamma_2^{r_0} \leqslant cn$. We partition the time-steps $s \in [r_0, t]$ into red and green phases (see Fig. 4.2):

1. **Red Phase:** The step s is in a red phase if $\Gamma_2^s > cn$.
2. **Green Phase:** Otherwise, the process is in a green phase.

Note that by the choice of r_0, the process is at a green phase at time r_0. Then each green phase may be preceded by a red phase. Trivially, for each step s in a green phase, we have $\Gamma_2^s \leqslant cn$. Also, when s is the first step of a red phase after a green phase, it follows that $\Gamma_2^s \leqslant e^{\gamma_2 d} \cdot \Gamma_2^{s-1} \leqslant 2 \cdot \Gamma_2^{s-1} \leqslant 2cn$, since $0 < \gamma_2 < 1/(2d)$.

We now let $\mathcal{R}^s$ denote the event that step s is the first step of a red phase. Additionally, let $\mathcal{A}^s$ be the event that all steps $u \in [s, t]$ in the same phase as s, satisfy $\Gamma_2^u \leqslant 3cn$. By Lemma 4.5, we have that

$$\mathbf{Pr}\left[\mathcal{A}^s \,\middle|\, \mathcal{Z}^{t-T_r}, \mathfrak{F}^s, \mathcal{R}^s \right] \geqslant 1 - n^{-\frac{4}{3}\kappa}.$$

For any events $\mathcal{E}_1 \neq \emptyset$ and $\mathcal{E}_2$, we have that $\mathbf{Pr}[\, \mathcal{E}_2 \cup \neg\mathcal{E}_1 \,] \geqslant 1 - \mathbf{Pr}[\, \neg\mathcal{E}_2 \mid \mathcal{E}_1 \,]$ and hence

$$\mathbf{Pr}\left[\mathcal{A}^s \cup \neg\mathcal{R}^s \,\middle|\, \mathcal{Z}^{t-T_r}, \mathfrak{F}^s \right] \geqslant 1 - n^{-\frac{4}{3}\kappa}.$$

By taking the union-bound over all steps s in $[r_0, t]$, we have that

$$\mathbf{Pr}\left[\bigcap_{s \in [r_0, t]} (\mathcal{A}^s \cup \neg\mathcal{R}^s) \,\middle|\, \mathcal{Z}^{t-T_r}, \mathfrak{F}^{r_0}, \Gamma_2^{r_0} \leqslant cn \right] \geqslant 1 - n^{-\frac{4}{3}\kappa} \cdot T_r \geqslant 1 - \frac{1}{4} n^{-\kappa}.$$

When $\cap_{s \in [r_0, t]} (\mathcal{A}^s \cup \neg\mathcal{R}^s)$ holds, all steps u in all red phases satisfy $\Gamma_2^u \leqslant 3cn$. Thus, since steps in green phases are good by definition, we have that

$$\mathbf{Pr}\left[\bigcap_{s \in [r_0, t]} \left\{ \Gamma_2^s \leqslant 3cn \right\} \,\middle|\, \mathcal{Z}^{t-T_r}, \mathfrak{F}^{r_0}, \Gamma_2^{r_0} \leqslant cn \right] \geqslant 1 - n^{-\frac{4}{3}\kappa} \cdot T_r \geqslant 1 - \frac{1}{4} n^{-\kappa}.$$

Hence, by defining the stopping time $\rho := \inf\{r_0 \geqslant t - T_r : \Gamma_2^{r_0} \leqslant cn\}$, we have

$$\begin{aligned}
\mathbf{Pr}\left[\Gamma_2^t \leqslant 3cn \right] &\geqslant \sum_{r_0 = t - T_r}^{t} \mathbf{Pr}\left[\Gamma_2^t \leqslant 3cn \,\middle|\, \mathcal{Z}^{t-T_r}, \rho = r_0 \right] \cdot \mathbf{Pr}\left[\rho = r_0 \,\middle|\, \mathcal{Z}^{t-T_r} \right] \\
&\geqslant \sum_{r_0 = t - T_r}^{t} \mathbf{Pr}\left[\bigcap_{s \in [r_0, t]} \left\{ \Gamma_2^s \leqslant 3cn \right\} \,\middle|\, \mathcal{Z}^{t-T_r}, \mathfrak{F}^{r_0}, \Gamma_2^{r_0} \leqslant cn \right] \cdot \mathbf{Pr}\left[\rho = r_0 \,\middle|\, \mathcal{Z}^{t-T_r} \right] \\
&\geqslant \left(1 - \frac{1}{4} n^{-\kappa}\right) \cdot \sum_{r_0 = t - T_r}^{t} \mathbf{Pr}\left[\rho = r_0 \,\middle|\, \mathcal{Z}^{t-T_r} \right] \\
&= \left(1 - \frac{1}{4} n^{-\kappa}\right) \cdot \mathbf{Pr}\left[\rho \leqslant t \right] \\
&\overset{(a)}{\geqslant} \left(1 - \frac{1}{4} n^{-\kappa}\right) \cdot \left(1 - 2n^{-2\kappa}\right) \geqslant 1 - \frac{1}{2} n^{-\kappa},
\end{aligned}$$

using Eq. (4.9) in (a) and so the conclusion follows. $\square$

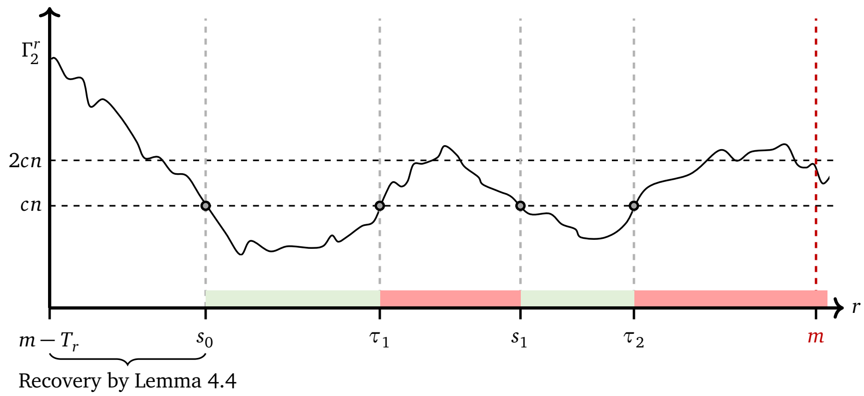


Figure 4.2: Green phases indicate steps where Γ_2^r is small and red phases indicate steps for which the potential is large and drops (in expectation). In Lemma 4.5, we show that $\Gamma_2 \leqslant 3cn$ at every point within a red phase using a concentration inequality.

Now we complete the proof of Theorem 4.1.

Proof of Theorem 4.1. Consider the auxiliary process $\widetilde{\mathcal{P}}_{t-T_r}$ and let $\Gamma_{2,\widetilde{\mathcal{P}}}$ be its Γ_2 potential. Then, by Lemma 4.6 we have that

$$\mathbf{Pr}\left[\, \Gamma_{2,\widetilde{\mathcal{P}}}^t \leqslant 3cn \,\middle|\, \mathcal{Z}^{t-T_r} \,\right] \geqslant 1 - \frac{1}{2} n^{-\kappa}. \tag{4.10}$$

By Lemma 4.2 and Markov's inequality, since $\widetilde{\mathcal{P}}_{t-T_r}$ and $\mathcal{P}$ agree for every step $s \leqslant t - T_r$, we have that

$$\mathbf{Pr}\left[\, \mathcal{Z}^{t-T_r} \,\right] = \mathbf{Pr}\left[\, \Gamma_1^{t-T_r} \leqslant \frac{1}{2} cn^{2\kappa+1} \,\right] \geqslant 1 - 2n^{-2\kappa}. \tag{4.11}$$

Hence, by combining Eq. (4.10) and Eq. (4.11), we have that

$$\mathbf{Pr}\left[\, \Gamma_{2,\widetilde{\mathcal{P}}}^t \leqslant 3cn \,\right] \geqslant \left(1 - \frac{1}{2} n^{-\kappa}\right) \cdot \left(1 - 2n^{-2\kappa}\right) \geqslant 1 - \frac{3}{4} n^{-\kappa}.$$

As shown in Eq. (4.4), w.h.p. the process $\widetilde{\mathcal{P}}$ agrees with $\mathcal{P}$ in all steps in $[t - T_r, t]$, and hence

$$\mathbf{Pr}\left[\, \Gamma_2^t \leqslant 3cn \,\right] \geqslant \mathbf{Pr}\left[\left\{ \Gamma_{2,\widetilde{\mathcal{P}}}^t \leqslant 3cn \right\} \cap \bigcap_{s \in [t-T_r, t]} \left\{ y^s = y_{\widetilde{\mathcal{P}}}^s \right\} \right] \geqslant 1 - \frac{3}{4} n^{-\kappa} - n^{-2\kappa} \cdot T_r \geqslant 1 - n^{-\kappa},$$

using that $T_r = \mathcal{O}(n \cdot n^{1/6} \cdot \log n)$. $\square$

Chapter $4^{1/2}$: Disks-into-bins

A load balancing process related to balls-into-bins is the *disks-into-bins* process. At step $t \geqslant 0$, this process:

- Samples one point (x, y) uniformly at random from the square $[0,1]^2$, i.e., $x \sim \mathcal{U}[0,1]$ and $y \sim \mathcal{U}[0,1]$.
- Places a disk of radius r at that position.

The load of each point in the $[-r, 1+r]^2$ square is the number of times that a disk covers it. What is the *maximum load* over all points for this process?

The equivalent *coupon collector* question, i.e., the number of disks required to cover all points, can be perhaps answered using slight modifications of the classical argument.

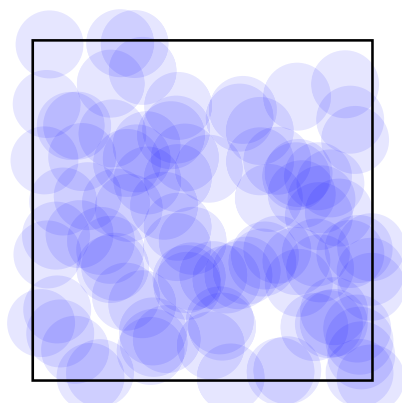


Figure 4.3: Example of the ONE-CHOICE process in the disks-into-bins setting with $n = 100$ balls and radius $r = 0.1$. The maximum load here is 8.

Another reasonable question is whether a process similar to TWO-CHOICE would produce a lower maximum load than ONE-CHOICE. In particular, consider the process, which at step $t \geqslant 0$:

- Samples two points P_1 and P_2 uniformly at random from the square $[0,1]^2$.
- Computes the maximum load of the points covered by placing a disk at P_1 and at P_2.
- Places to the disk at the point where this value is smaller (breaking ties randomly).

Some empirical results in Table 4.4 for various values of radius r and number of disks m, suggest that this TWO-CHOICE process is superior to ONE-CHOICE. However, it is not clear whether TWO-CHOICE is optimal over all TWO-SAMPLE processes or even if the tie-breaking rule is important.

n\r	0.05	0.1	0.2	0.3
100	5/3.4	8.9/6.7	21.2/17.5	36.7/31.1
200	7.5/4.9	15.7/11	38.2/31.7	69.4/59.2
300	9.4/6.2	20.5/15.3	54.4/44.7	100.5/88.2
500	12.7/8.9	29.6/22.9	84.1/72.6	160.8/141.9

Table 4.4: Maximum loads for the ONE-CHOICE/TWO-CHOICE processes for various values of n and r.

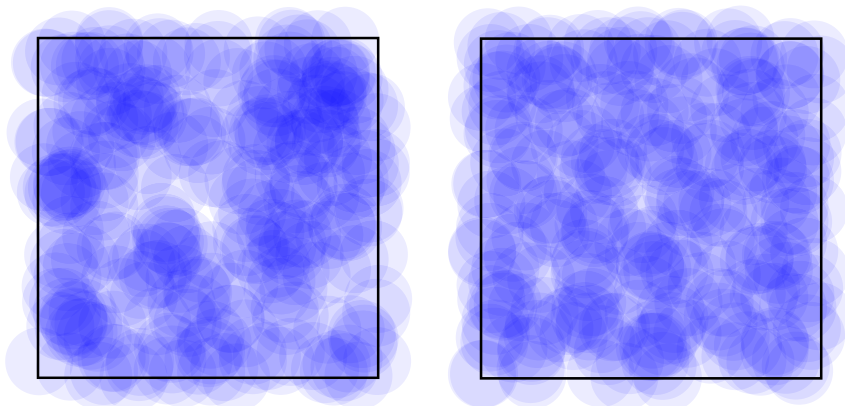


Figure 4.5: Example of disk distribution for ONE-CHOICE and TWO-CHOICE for $n = 300$ and $r = 0.1$. Note that in the TWO-CHOICE distribution the loads are more balanced.

Interplay between the absolute value and quadratic potentials

In this section, we present the interplay between two lower-order potential functions, namely the *absolute value potential* defined as the ℓ_1-norm of the normalised load vector

$$\Delta^t := \sum_{i=1}^{n} \left| y_i^t \right|,$$

and the *quadratic potential* defined as the sum of the squares of normalised loads

$$\Upsilon^t := \sum_{i=1}^{n} (y_i^t)^2.$$

This interplay allows us to analyse:

1. The g-Adv-Comp setting for the Two-Choice process showing an $\mathcal{O}(g+\log n)$ bound in Section 5.3 and establishing the base case for the $\mathcal{O}\big(\frac{g}{\log g} \cdot \log\log n\big)$ bound for any $g = o(\log n)$ to be proven in Section 7.4.1. This upper bound also applies to g-Bounded, g-Adv-Load and g-Myopic-Comp settings for Two-Choice, so we refer to g-Adv-Comp as g-Adv setting for short.

2. The Mean-Thinning process showing an $\mathcal{O}(\log n)$ bound in Section 5.2. The analysis works for the family of Mean-Biased processes, so we obtain bounds for the Twinning, $(1+\eta)$ and Relative-Threshold$(f(n))$ processes with $f(n) \geqslant 0$ as corollaries in Section 5.2.2.

In Section 5.1, we outline the limitations of using a single hyperbolic cosine potential and explain how the interplay between the absolute value and the quadratic potential overcomes these limitations. In Section 5.2, we present a detailed outline for the analysis of the Mean-Biased processes, but proofs are deferred to [117] (and [119]). In Section 5.3, we give the full details for obtaining the $\mathcal{O}(g+\log n)$ in the g-Adv setting for Two-Choice. Finally, in Section 5.4, we prove *strong stabilisation*, i.e., that a variant of the hyperbolic cosine potential is w.h.p. $\mathcal{O}(n)$, which we later use for the tighter bounds for $g = o(\log n)$ in Section 7.4.1.

5.1 Motivation

5.1.1 For the Mean-Thinning process

Recall that the Mean-Thinning (= Relative-Threshold(0)) process is the Two-Thinning process which uses the mean (average) t/n as a threshold decision for whether or not to allocate a ball to the first bin sample or not. For Mean-Thinning, we aim to prove an $\mathcal{O}(\log n)$ bound on the gap.

Let δ^t be the quantile of the average, i.e., $\delta^t := |B_+^t|/n$, where $B_+^t := \{i \in [n] : y_i^t \geqslant 0\}$. At any step $t \geqslant 0$ the sorted allocation vector is the same as that of Quantile(δ^t), i.e.,

$$\tilde{q}^t := \Bigg(\underbrace{\frac{\delta^t}{n}, \ldots, \frac{\delta^t}{n}}_{n\delta^t \text{ entries}}, \underbrace{\frac{1+\delta^t}{n}, \ldots, \frac{1+\delta^t}{n}}_{n\cdot(1-\delta^t) \text{ entries}} \Bigg).$$

The problem here is that when $\delta^t = o(1)$ or when $\delta^t = 1-o(1)$, then the hyperbolic cosine potential $\Gamma := \Gamma(\gamma)$ with any constant smoothing parameter $\gamma > 0$ might increase in expectation. A concrete such load vector is given in [117, Claim B.2]. In particular, when $\delta^t = \frac{1}{n}$, the process is equivalent to the process that just avoids allocating to the maximum load (if possible), i.e.,

$$\widetilde{q}^t := \left(\frac{1}{n^2}, \frac{1}{n} + \frac{1}{n^2}, \ldots, \frac{1}{n} + \frac{1}{n^2}\right).$$

For processes with such small biases, we can only show that the hyperbolic cosine potential $\Gamma_0 := \Gamma_0(\gamma_0)$ with smoothing parameter $\gamma_0 = \Theta(1/n)$ is linear in expectation, from which we get an $\mathcal{O}(n \log n)$ bound on the gap.

However, we need to bound $\Gamma := \Gamma(\gamma)$ for constant $\gamma > 0$ to get the $\mathcal{O}(\log n)$ bound. By the analysis in Chapter 3, when $\delta^t \in (\epsilon, 1-\epsilon)$ for some constant $\epsilon \in (0,1)$, we have that $\Gamma := \Gamma(\gamma)$ with (sufficiently small) constant $\gamma := \gamma(\epsilon) > 0$ drops in expectation.

Thus, our goal becomes to show that $\delta^t \in (\epsilon, 1-\epsilon)$ in a large constant fraction of the steps, so that the potential Γ drops overall in expectation. We do this using an interplay between the absolute value potential and the quadratic potential.

More specifically, we show that the expected value of the quadratic potential for Mean-Thinning satisfies the following drop inequality at any step $t \geqslant 0$,

$$\mathbf{E}\left[\, \Upsilon^{t+1} \,\middle|\, \mathfrak{F}^t \,\right] \leqslant \Upsilon^t - \frac{\Delta^t}{n} + 1. \tag{5.1}$$

This implies that when $\Delta^t = \Omega(n)$ (even if the quantile δ^t of the average is $o(1)$ or $1-o(1)$), then the quadratic potential drops in expectation. By looking at a sufficiently long interval we can deduce that a large constant fraction of the steps t satisfy $\Delta^t = \mathcal{O}(n)$. Then by a One-Choice argument (Lemma 5.5), we can deduce that for *many steps* s in $[t, t+\Theta(n)]$, the quantile stabilises, i.e., $\delta^s \in (\epsilon, 1-\epsilon)$. On aggregate, using an adjusted version of the hyperbolic cosine potential, we show that Γ is w.h.p. $\mathcal{O}(n)$ every $\mathcal{O}(n \log n)$ steps, implying by smoothness an $\mathcal{O}(\log n)$ gap at every step.

5.1.2 For the g-Adv setting

In the g-Adv setting for Two-Choice, we have a similar problem: There exist configurations where the hyperbolic cosine potential for constant smoothing parameter may increase in expectation even when large. For instance, when all bins have almost the same load, i.e., x^t with $\max_{i\in[n]} x_i^t - \min_{i\in[n]} x_i^t \leqslant g$, then the adversary G^t can "force" the sorted allocation vector $\widetilde{q}^t$ to be worse than that of One-Choice, e.g., the reverse of the Two-Choice vector,

$$\widetilde{q}^t := \left(\frac{2}{n} - \frac{1}{n^2}, \ldots, \frac{2(n-i+1)-1}{n^2}, \ldots, \frac{1}{n^2}\right).$$

Again, by investigating the expected change of the quadratic potential, we obtain the following interplay with Δ^t

$$\mathbf{E}\left[\, \Upsilon^{t+1} \,\middle|\, \mathfrak{F}^t \,\right] \leqslant \Upsilon^t - \frac{\Delta^t}{n} + 2g + 1.$$

This is very similar to the drop inequality Eq. (5.1) for Mean-Thinning, with the difference that the additive term may be non-constant. This inequality implies that when $\Delta^t > Dng$ (for $D = 365$), then the quadratic potential drops in expectation. So again by looking at a sufficiently long interval we can deduce that a large constant fraction of the steps satisfy $\Delta^t \leqslant Dng$. Next note that when $\Delta^t \leqslant Dng$, then using a first moment argument, the number of bins with load at least $\frac{3}{2}Dg$ is at most $\frac{n}{3}$, and so there is a constant bias $\epsilon = 1/3$ to allocate away from bins with load at least $2Dg$. Similarly, for bins with load at most $-2Dg$, there is a constant bias to allocate to them (see Fig. 5.1). Hence, in Lemma 5.21 using a similar analysis to that in Chapter 3, we obtain that the hyperbolic cosine potential Λ *with an offset* of $c_4 g = 2Dg$,

$$\Lambda^t := \Lambda^t(\alpha, c_4 g) = \sum_{i=1}^{n} \Lambda_i^t := \sum_{i=1}^{n} \left(e^{\alpha \cdot (y_i^t - c_4 g)^+} + e^{\alpha \cdot (-y_i^t - c_4 g)^+} \right),$$

has w.h.p. $\Lambda^t = \mathcal{O}(n)$ at an arbitrary step t and so we can deduce that $\mathrm{Gap}(t) = \mathcal{O}(g + \log n)$.

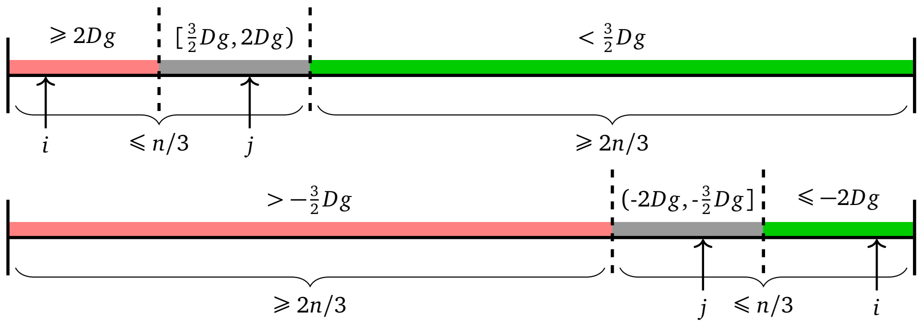


Figure 5.1: Consider any step t with $\Delta^t \leqslant Dng$, then:
(**Top**) There are at most $\frac{n}{3}$ bins with $y_i^t \geqslant \frac{3}{2}Dg$, so in the g-Adv-Comp, we can distinguish between the red and green bins, so the probability to allocate to a red bin is at most $\frac{2}{3n}$.
(**Bottom**) There are at most $\frac{n}{3}$ bins with $y_i^t \leqslant -\frac{3}{2}Dg$, so in the g-Adv-Comp, we can distinguish between the red and green bins, so the probability to allocate to a green bin is at least $\frac{4}{3n}$.

5.2 Mean-Biased processes

Recall that a Mean-Biased process is any process that satisfies conditions $\mathcal{P}_3$ and $\mathcal{W}_2$ or $\mathcal{P}_2$ and $\mathcal{W}_3$ defined in Section 2.5.6, where we also verified that Mean-Thinning, Twinning and some other processes satisfy them. For convenience, we repeat these conditions here:

Condition $\mathcal{P}_2$: At any step $t \geqslant 0$, the probability allocation vector q^t must satisfy $q_+^t \leqslant \frac{1}{n} \leqslant q_-^t$.

Condition $\mathcal{W}_2$: At any step $t \geqslant 0$, if $i := i^{t+1}$ is chosen for allocation,

- If $y_i^t < 0$, then allocate w_- balls to bin i,
- If $y_i^t \geqslant 0$, then allocate w_+ balls to bin i,

where $1 \leqslant w_+ \leqslant w_-$ are constant integers.

Condition $\mathcal{P}_3$: This is as Condition $\mathcal{P}_2$, but additionally, there are time-independent constants $k_1 \in (0,1], k_2 \in (0,1]$ such that for any step $t \geqslant 0$:

$$q_+^t \leqslant \frac{1-k_1}{n} + \frac{k_1 \cdot |B_+^t|}{n^2} = \frac{1}{n} - \frac{k_1 \cdot (1-\delta^t)}{n},$$
$$q_-^t \geqslant \frac{1}{n} + \frac{k_2 \cdot |B_+^t|}{n^2} = \frac{1}{n} + \frac{k_2 \cdot \delta^t}{n}.$$

Condition $\mathcal{W}_3$: This is as Condition $\mathcal{W}_2$, but additionally we have the strict inequality: $w_+ < w_-$. Also, we assume that for each $t \geqslant 0$, allocation vector $\widetilde{q}_i^t$ is non-decreasing in i.

5.2.1 Analysis outline

In this section, we will give the statements for the key lemmas used in the proof of the following theorem. All omitted proofs can be found in [117].

Theorem 5.1 ([117, Theorem 4.15]). *For any* MEAN-BIASED *process, there exists a constant* $\kappa > 0$ *such that for any step* $m \geqslant 0$*,*

$$\mathbf{Pr}\left[\max_{i\in[n]}\left|x_i^m - \frac{W^m}{n}\right| \leqslant \kappa \log n\right] \geqslant 1 - n^{-3};$$

so in particular, $\mathbf{Pr}[\,\mathrm{Gap}(m) \leqslant \kappa \log n\,] \geqslant 1 - n^{-3}$.

The following basic, yet crucial result follows from the preconditions in Theorem 5.1:

Lemma 5.2. *For any* MEAN-BIASED *process, there exists a constant* $c_1 := c_1(k_1, k_2, w_-, w_+) > 0$*, so that for any step* $t \geqslant 0$*,*

$$q_-^t \cdot w_- - q_+^t \cdot w_+ \geqslant \frac{c_1}{n}.$$

Proof. First assume $\mathcal{P}_2$ and $\mathcal{W}_3$ holds. In this case, $\mathcal{P}_2$ implies $q_-^t \geqslant \frac{1}{n} \geqslant q_+^t$, and thus

$$q_-^t \cdot w_- - q_+^t \cdot w_+ \geqslant \frac{1}{n} \cdot w_- - \frac{1}{n} \cdot w_+ \geqslant \frac{c_1}{n},$$

for $c_1 := w_- - w_+ > 0$, since by $\mathcal{W}_3$ the weights are constants satisfying $w_- > w_+$.

Next assume $\mathcal{P}_3$ and $\mathcal{W}_2$ holds. In this case, $\mathcal{W}_2$ implies $w_- \geqslant w_+ \geqslant 1$. Using $\mathcal{P}_2$, for $c_1 := \min\{k_1, k_2\}$,

$$q_-^t \cdot w_- - q_+^t \cdot w_+ \geqslant q_-^t \cdot w_+ - q_+^t \cdot w_+ \geqslant (q_-^t - q_+^t) \cdot 1 \geqslant \left(\frac{1}{n} + \frac{k_2 \cdot \delta^t}{n} - \frac{1}{n} + \frac{k_1 \cdot (1-\delta^t)}{n}\right) \geqslant \frac{c_1}{n}. \qquad \square$$

A large portion of the analysis is devoted to derive a *weaker* quantile condition, that is, we prove that for sufficiently many steps, the quantile δ^t is in the range $(\epsilon, 1-\epsilon)$ for some (small) constant ϵ. For this analysis, the inequality in Lemma 5.2 will be useful when we establish a connection between the absolute and quadratic potential function (Lemma 5.4). We now list the 9 key steps in the analysis along with some of the important lemmas/theorems.

1. (**Weak gap bound**) In any step, there is at least one bin with load at least the average, so the MEAN-BIASED process always has a bias not to allocate to that bin. By using a hyperbolic cosine potential $\Gamma_0 := \Gamma_0(\gamma_0)$ with $\gamma_0 = \Theta(1/n)$, we get w.h.p. at any step $t \geqslant 0$, the weak bound of $\mathrm{Gap}(t) = \mathcal{O}(n \log n)$ and also a weak bound for $\Gamma := \Gamma(\gamma)$ with any constant $\gamma \in (0,1)$.

Lemma 5.3 (Lemma 8.10 in [117]). *For any* MEAN-BIASED *process, there exists a constant* $c_6 > 0$ *such that for any step* $m \geqslant 0$*,*

$$\mathbf{Pr}[\,\mathrm{Gap}(m) \leqslant c_6 \cdot n \log n\,] \geqslant 1 - n^{-12},$$

and so, for $\Gamma := \Gamma(\gamma)$ *with any (constant) smoothing parameter* $0 < \gamma < 1$*,*

$$\mathbf{Pr}[\,\Gamma^m \leqslant \exp(2c_6 n \log n)\,] \geqslant 1 - n^{-12}.$$

2. (**Absolute value/Quadratic potential interplay**) Next, we establish the following interplay between the absolute value potential and the quadratic potential. For MEAN-BIASED processes, this implies that when $\Delta^t = \Omega(n)$, then the quadratic potential drops in expectation.

Lemma 5.4 (Lemma 6.2 in [117]). *Consider any allocation process satisfying $\mathcal{P}_2$ and $\mathcal{W}_2$. Then for any step $t \geqslant 0$, the quadratic potential satisfies*

$$\mathbf{E}\Big[\,\Upsilon^{t+1}\,\Big|\,\mathfrak{F}^t\,\Big] \leqslant \Upsilon^t - (q_-^t \cdot w_- - q_+^t \cdot w_+) \cdot \Delta^t + 4 \cdot (w_-)^2.$$

Hence for any MEAN-BIASED *process, this implies by Lemma 5.2 that there exist constants $c_1, c_2 > 0$ such that for any step $t \geqslant 0$,*

$$\mathbf{E}\Big[\,\Upsilon^{t+1}\,\Big|\,\mathfrak{F}^t\,\Big] \leqslant \Upsilon^t - \frac{c_1}{n} \cdot \Delta^t + c_2.$$

3. (**Quantile stabilisation**) For every step t with $\Delta^t \leqslant Cn$ (for sufficiently large constant $C := C(c_1, c_2) > 0$), we have that in the next $\Theta(n)$ steps, there is a large constant fraction of steps with $\delta^t \in (\epsilon, 1-\epsilon)$ for some constant $\epsilon > 0$. This follows from the fact that $\Delta^t \leqslant Cn$ implies a constant fraction of the bins have constant normalised load $|y_i^t| \leqslant \tilde{c}$ and that MEAN-BIASED processes sample uniformly among the bins. If there are too many overloaded bins, then because $q_+^t \leqslant \frac{1}{n}$, by a ONE-CHOICE argument w.h.p. in the next $\tilde{c}n$ steps a constant fraction of them will not be chosen, so they will become underloaded. If there are too many underloaded bins with constant load, then because $q_-^t \geqslant \frac{1}{n}$, by a ONE-CHOICE argument w.h.p. a constant fraction of them will be allocated at least $2\tilde{c}$ balls to exceed the average and become overloaded.

Lemma 5.5 (Mean Quantile Stabilisation (Lemma 6.1 in [117])). *Consider any allocation process satisfying $\mathcal{P}_2$ and $\mathcal{W}_2$. Then, for any integer constant $C \geqslant 1$, there exists some $\epsilon := \epsilon(C) > 0$ such that for any integers $t_0 \geqslant 0$ and $t_1 := t_0 + \left\lceil \frac{2Cn}{w_+} \right\rceil + \left\lceil \frac{n}{w_+} \right\rceil + \left\lceil \frac{n}{10w_-} \right\rceil$ we have*

$$\mathbf{Pr}\left[\,\left|\left\{t \in [t_0, t_1] \colon \delta^t \in (\epsilon, 1-\epsilon)\right\}\right| \geqslant \epsilon \cdot n \;\middle|\; \mathfrak{F}^{t_0}, \Delta^{t_0} \leqslant C \cdot n\,\right] \geqslant 1 - e^{-\epsilon \cdot n}.$$

4. (**Many steps with $\delta^t \in (\epsilon, 1-\epsilon)$**) Using the interplay between the absolute value and the quadratic potential, we can deduce that in sufficiently long intervals, a large constant fraction of the steps s satisfy $\Delta^s \leqslant Cn$, for some constant $C > 0$. Then, using the mean quantile stabilisation lemma, we get a large constant fraction of steps with $\delta^s \in (\epsilon, 1-\epsilon)$. We have two versions of this theorem, one for the *recovery phase* (where the starting point is a weak bound on Γ) and one for the *stabilisation phase* (where the starting point is a linear bound on Γ). In the statements below, we let $G_{t_0}^{t_1}$ be the number of steps $t \in [t_0, t_1]$ with $\Delta^t \leqslant Cn$.

Lemma 5.6 (Lemma 9.3 in [117]). *Consider any* MEAN-BIASED *process and the potential $\Gamma := \Gamma(\gamma)$ for any constant $\gamma \in (0,1)$. Then, for $\epsilon := \epsilon(C)$ as in Lemma 5.5, $r := \min\left\{\frac{\epsilon}{20C}, \frac{1}{2}\right\}$, and for any steps t_0 and t_1 with $t_1 \geqslant t_0 + n^3 \log^3 n$, it holds that*

$$\mathbf{Pr}\left[\, G_{t_0}^{t_1} > r \cdot (t_1 - t_0) \;\middle|\; \mathfrak{F}^{t_0}, \Gamma^{t_0} \leqslant \exp(2c_6 n \log n)\,\right] \geqslant \frac{1}{2} - (t_1 - t_0) \cdot e^{-\epsilon n},$$

where $c_6 > 0$ is the constant from Lemma 5.3.

Lemma 5.7 (Lemma 9.4 in [117]). *Consider any* MEAN-BIASED *process and the potential $\Gamma := \Gamma(\gamma)$ with any constant $\gamma \in (0,1)$. Then, for $\epsilon := \epsilon(C)$ as in Lemma 5.5, $r := \min\left\{\frac{\epsilon}{20C}, \frac{1}{2}\right\}$, for any constants $\kappa_1, \kappa_2 > 0$ and for any steps t_0 and t_1 satisfying $t_1 := t_0 + \kappa_2 \cdot n \log n$, it holds that*

$$\mathbf{Pr}\left[\, G_{t_0}^{t_1} > r \cdot (t_1 - t_0) \;\middle|\; \mathfrak{F}^{t_0}, \Gamma^{t_0} \leqslant \kappa_1 \cdot n\,\right] \geqslant 1 - 3 \cdot n^{-12}.$$

5. (**Good/bad steps for Γ**) For the expectation of the hyperbolic cosine potential $\Gamma := \Gamma(\gamma)$ with constant $\gamma > 0$, we prove a drop inequality in a *good step* (where $\delta^t \in (\epsilon, 1-\epsilon)$) and a weaker bound for a *bad step*.

Corollary 5.8. *Consider any* MEAN-BIASED *process, let $\epsilon \in (0,1)$ be any constant and the potential $\Gamma := \Gamma(\gamma)$ with $\gamma := \gamma(\epsilon)$ as defined in [117, Lemma 7.4]. Choose $c := \max\{\frac{3(c_3+w_-)\cdot e^{2w_-}}{\gamma c_3}, \frac{2}{\gamma^2}\} > 1$, for some constant $c_3 := c_3(\epsilon)$. Then, for any step $t \geqslant 0$,*

$$\mathbf{E}\left[\,\Gamma^{t+1} \,\middle|\, \mathfrak{F}^t, \{\delta^t \in (\epsilon, 1-\epsilon)\}, \Gamma^t \geqslant c \cdot n\,\right] \leqslant \Gamma^t \cdot \left(1 - \frac{c_3\gamma}{n}\right).$$

More generally, for any step $t \geqslant 0$, and $c_4 := 3w_- \cdot e^{2w_-} > 0$, we have

$$\mathbf{E}\left[\,\Gamma^{t+1} \,\middle|\, \mathfrak{F}^t, \Gamma^t \geqslant c \cdot n\,\right] \leqslant \Gamma^t \cdot \left(1 + \frac{\gamma^2 c_4}{n}\right).$$

6. (**Adjusted hyperbolic potential**) In order to show that the hyperbolic cosine potential $\Gamma := \Gamma(\gamma)$ becomes small, we use the *adjusted hyperbolic cosine potential*, defined as $\widetilde{\Gamma}_{t_0}^{t_0}(c_3, \gamma, \xi) := \Gamma^{t_0}(\gamma)$ and, for any step $s > t_0$ as

$$\widetilde{\Gamma}_{t_0}^{s}(c_3, \gamma, \xi) := \Gamma^s \cdot \mathbf{1}_{\mathcal{E}_{t_0}^{s-1}} \cdot \exp\left(-\frac{c_3\gamma\xi}{n} \cdot B_{t_0}^{s-1}\right) \cdot \exp\left(+\frac{c_3\gamma}{n} \cdot G_{t_0}^{s-1}\right), \tag{5.2}$$

where $\xi := \frac{r}{2(1-r)}$.

Lemma 5.9 (Lemma 9.1 in [117]). *The sequence $(\widetilde{\Gamma}_{t_0}^{s})_{s \geqslant t_0}$ forms a super-martingale.*

7. (**Recovery phase**) Starting with the weak $\mathcal{O}(n \log n)$ bound on the gap at step t_0 and using the adjusted hyperbolic cosine potential, we can deduce that for any sufficiently long interval w.h.p. there exists *some* step t with $\Gamma^t = \mathcal{O}(n)$.

Lemma 5.10 (Recovery (Lemma 9.5 in [117])). *Consider any* MEAN-BIASED *process and the potential $\Gamma := \Gamma(\gamma)$ with $\gamma := \gamma(\epsilon)$ as defined in [117, Lemma 7.4]. Then, for the constant $c > 1$ as defined in Corollary 5.8, for any step $m \geqslant 0$,*

$$\mathbf{Pr}\left[\bigcup_{s \in [m - 40n^3 \log^4 n, m]} \{\Gamma^s < cn\}\right] \geqslant 1 - n^{-10}.$$

8. (**Stabilisation phase**) Again, using the adjusted hyperbolic cosine potential, we show that once Γ is small w.h.p. it becomes small again every $\mathcal{O}(n \log n)$ steps.

Lemma 5.11 (Stabilisation (Lemma 9.7 in [117])). *Consider any* MEAN-BIASED *process and the potential $\Gamma := \Gamma(\gamma)$ with $\gamma := \gamma(\epsilon)$ as defined in [117, Lemma 7.4]. Then, for the constant $c > 1$ as defined in Corollary 5.8, there exists a constant $c_s > 0$, such that for any step $t_0 \geqslant 0$,*

$$\mathbf{Pr}\left[\bigcup_{t \in [t_0, t_0 + c_s n \log n - 1]} \left\{\Gamma^t < cn\right\} \,\middle|\, \mathfrak{F}^{t_0}, \Gamma^{t_0} \in [cn, 2cn]\right] \geqslant 1 - \frac{1}{2} \cdot n^{-7}.$$

9. (**Gap deduction**) Finally, by a *smoothness argument* since the gap can decrease by at most $\mathcal{O}(w_- \cdot \log n)$ in $\Theta(n \log n)$ steps, we deduce the $\mathcal{O}(\log n)$ bound on the gap at every step.

5.2.2 Applications

Thanks to the reductions in Lemmas 2.21, 2.22 and D.6, by Theorem 5.1 we also deduce:

Corollary 5.12. *For* MEAN-THINNING, TWINNING *and the* $(1+\beta)$*-process for any constant* $\beta \in (0,1]$*, there exists a constant* $\kappa > 0$ *such that for any step* $m \geqslant 0$*,*

$$\mathbf{Pr}\left[\, \mathrm{Gap}(m) \leqslant \kappa \log n \,\right] \geqslant 1 - n^{-3}.$$

As we shall show in Corollary C.24, these bounds are tight for MEAN-THINNING, TWINNING and the $(1+\beta)$-process for constant $\beta \in (0,1)$. By Lemma 2.15, we also obtain the following corollary.

Corollary 5.13. *For any* RELATIVE-THRESHOLD$(f(n))$ *process with* $f(n) > 0$*, there exists a constant* $\kappa > 0$ *such that for any step* $m \geqslant 0$*,*

$$\mathbf{Pr}\left[\, \mathrm{Gap}(m) \leqslant \kappa \log n + f(n) \,\right] \geqslant 1 - n^{-3}.$$

We will show in Lemma C.25 that for $f(n) \geqslant \log n$, this bound is asymptotically tight.

5.3 g-ADV setting: Stabilisation

In this section we give the proof for the $\mathcal{O}(g + \log n)$ gap bound, as stated in the theorem below. For any $g = \Omega(\log n)$, this matches the lower bound for the g-MYOPIC-COMP process in Proposition C.8 up to multiplicative constants.

Lemma 5.14 (Simplified, page 113). *Consider the* g-ADV-COMP *setting for any* $g \geqslant 1$*. Then, there exists a constant* $\kappa > 0$*, such that for any step* $m \geqslant 0$*,*

$$\mathbf{Pr}\left[\, \max_{i \in [n]} \left| y_i^m \right| \leqslant \kappa \cdot (g + \log n) \,\right] \geqslant 1 - 2 \cdot (ng)^{-9}.$$

5.3.1 Proof outline of Theorem 5.26

The proof of this theorem is considerably more involved than Theorem 3.21, requiring the interplay between an exponential potential, the absolute value potential and the quadratic potential, similar to the one for the MEAN-BIASED processes.

1. (**Weak gap bound**) By Theorem 3.21, w.h.p. $\mathrm{Gap}(t) = \mathcal{O}(g \log(ng))$ for any step $t \geqslant 0$.

2. (**Absolute value/Quadratic potential interplay**) We establish the following interplay between the absolute value and the quadratic potential. This is similar to the one for the MEAN-BIASED processes, but the additive term may be super-constant here.

Lemma 5.17 (Restated, page 100). *Consider the* g-ADV-COMP *setting for any* $g \geqslant 1$*. Then, for any step* $t \geqslant 0$*,*

$$\mathbf{E}\left[\, \Upsilon^{t+1} \,\middle|\, y^t \,\right] \leqslant \Upsilon^t - \frac{\Delta^t}{n} + 2g + 1.$$

3. (**Many steps with $\Delta^t \leqslant Dng$ for $D = 365$**) Using this interplay, we can establish that starting with a bound of T on the quadratic potential $\Upsilon^{t_0} \leqslant T$, then in the next $\Theta(T/g)$ steps, a large constant fraction of the steps satisfy $\Delta^t \leqslant Dng$ (*good steps*). In the recovery, stabilisation and strong stabilisation phases, we instantiate this lemma with different values for T (depending on the strength of the bound). For convenience, we let $G_{t_0}^{t_1}$ be the number of good steps in $[t_0, t_1]$.

Lemma 5.18 (Restated, page 101). *Consider the g-ADV-COMP setting for any $g \geqslant 1$ and let $r := \frac{6}{6+\epsilon}$, $\epsilon := 1/12$ and $D := 365$. Then, for any constant $\hat{c} \geqslant 1$ and any $T \in [ng^2, n^2 g^3/\hat{c}]$, we have for any steps $t_0 \geqslant 0$ and $t_1 := t_0 + \hat{c} \cdot T \cdot g^{-1} - 1$,*

$$\mathbf{Pr}\left[G_{t_0}^{t_1}(D) \geqslant r \cdot (t_1 - t_0 + 1) \,\middle|\, \mathfrak{F}^{t_0}, \Upsilon^{t_0} \leqslant T, \max_{i \in [n]} \left| y_i^{t_0} \right| \leqslant g(\log(ng))^2 \right] \geqslant 1 - 2 \cdot (ng)^{-12}.$$

3. (**Good/bad steps for Λ**) Consider the following variant of the hyperbolic cosine potential with a constant smoothing parameter $\alpha := \frac{1}{18}$ and with an offset of $c_4 g = 2Dg$,

$$\Lambda^t := \Lambda^t(\alpha, c_4 g) = \sum_{i=1}^n \Lambda_i^t := \sum_{i=1}^n \left(e^{\alpha \cdot (y_i^t - c_4 g)^+} + e^{\alpha \cdot (-y_i^t - c_4 g)^+} \right), \tag{5.3}$$

where $v^+ := \max\{v, 0\}$. We prove the following drop inequality for the expectation of Λ over a *good step* (where $\Delta^t \leqslant Dng$) and a weaker bound over a *bad step*.

Lemma 5.14 (Simplified versions of Lemmas 5.20 and 5.21). *Consider the g-ADV-COMP setting with $g \geqslant 1$ and let $\epsilon := 1/12$. Then, for any step $t \geqslant 0$,*

$$\mathbf{E}\left[\Lambda^{t+1} \,\middle|\, \mathfrak{F}^t, \Delta^t \leqslant Dng \right] \leqslant \Lambda^t \cdot \left(1 - \frac{2\alpha\epsilon}{n} \right) + 18\alpha.$$

More generally, for any step $t \geqslant 0$,

$$\mathbf{E}\left[\Lambda^{t+1} \,\middle|\, \mathfrak{F}^t \right] \leqslant \Lambda^t \cdot \left(1 + \frac{3\alpha}{n} \right).$$

4. (**Adjusted hyperbolic cosine potential**) In order to show that Λ drops in expectation, we use the *adjusted hyperbolic cosine potential*, defined as $\widetilde{\Lambda}_{t_0}^{t_0} := \Lambda^{t_0}$ and, for any step $s > t_0$ as

$$\widetilde{\Lambda}_{t_0}^s := \Lambda^s \cdot \mathbf{1}_{\mathcal{E}_{t_0}^{s-1}} \cdot \exp\left(-\frac{3\alpha}{n} \cdot B_{t_0}^{s-1} \right) \cdot \exp\left(+\frac{\alpha\epsilon}{n} \cdot G_{t_0}^{s-1} \right), \tag{5.4}$$

where $G_{t_0}^{s-1}$ (and $B_{t_0}^{s-1}$) is the number of good (and bad) steps in $[t_0, s)$.

5. (**Recovery phase**) Using the weak bound as a starting point t_0 and the adjusted hyperbolic cosine potential, we show that after $\Theta(ng \cdot (\log(ng))^2)$ steps, w.h.p. the potential Λ becomes $\mathcal{O}(n)$.

Lemma 5.23 (Recovery – Simplified version, page 108). *Consider the g-ADV-COMP setting with $g \geqslant 1$. Then, there exists a constant $c > 0$, such that for any step $t_0 \geqslant 0$ and $\Delta_r := \Theta(ng \cdot (\log(ng))^2)$,*

$$\mathbf{Pr}\left[\bigcup_{t \in [t_0, t_0 + \Delta_r]} \{\Lambda^t \leqslant cn\} \right] \geqslant 1 - (ng)^{-11}.$$

6. (**Stabilisation phase**) Again, using the adjusted hyperbolic cosine potential, we show that once $\Lambda = \mathcal{O}(n)$, w.h.p. it becomes small again every $\mathcal{O}(n \cdot \max\{\log n, g\})$ steps.

Lemma 5.24 (Stabilisation – Simplified version, page 110). *Consider the g-ADV-COMP setting with $g \geqslant 1$. For any step $t_0 \geqslant 0$ and $\Delta_s := \frac{60 c_s}{\alpha \epsilon r} \cdot n \cdot \max\{\log n, g\}$, we have that*

$$\mathbf{Pr}\left[\bigcup_{t \in [t_0, t_0 + \Delta_s]} \{\Lambda^t \leqslant cn\} \,\middle|\, \mathfrak{F}^{t_0}, \Lambda^{t_0} \leqslant 2cn \right] \geqslant 1 - (ng)^{-11}.$$

7. (**Gap deduction**) Hence w.h.p. there is a step $t \in [m, m + \Theta(n \cdot (g + \log n))]$ with $\mathrm{Gap}(t) = \mathcal{O}(g + \log n)$, which by smoothness implies $\mathrm{Gap}(m) = \mathcal{O}(g + \log n)$ (Section 5.3.7).

In this section we give the proof of the $\mathcal{O}(g + \log n)$ gap bound, as stated in the theorem below. For $g = \Omega(\log n)$, this matches the lower bound for the g-Myopic-Comp setting in Proposition C.8 up to multiplicative constants.

Theorem 5.26 (Simplified version, page 113). *Consider the g-Adv-Comp setting for any $g \geqslant 1$. Then, there exists a constant $\kappa > 0$, such that for any step $m \geqslant 0$,*

$$\mathbf{Pr}\left[\max_{i \in [n]} \left|y_i^m\right| \leqslant \kappa \cdot (g + \log n)\right] \geqslant 1 - 2 \cdot (ng)^{-9}.$$

5.3.2 Absolute value and quadratic potentials

Recall that the *absolute value potential* is defined as

$$\Delta^t := \sum_{i=1}^{n} \left|y_i^t\right|, \tag{5.5}$$

and the *quadratic potential* is defined as

$$\Upsilon^t := \sum_{i=1}^{n} (y_i^t)^2. \tag{5.6}$$

We will upper bound the expected change of the quadratic potential $\mathbf{E}\left[\, \Delta\Upsilon^{t+1} \,\middle|\, y^t \,\right]$ in the g-Adv-Comp setting by relating it to the change of the quadratic potential for Two-Choice without noise, starting with the same load vector y^t at step t.

We will first analyse the expected change of the quadratic potential for the Two-Choice process without noise. We will make use of the following general lemma, which provides a formula for the change of the quadratic potential:

Lemma 5.15. *Consider any Sequential(r^t) process. Then, for any step $t \geqslant 0$, (i) it holds that*

$$\mathbf{E}\left[\, \Delta\Upsilon^{t+1} \,\middle|\, y^t \,\right] = \sum_{i=1}^{n} 2 \cdot r_i^t \cdot y_i^t + 1 - \frac{1}{n} \leqslant \sum_{i=1}^{n} 2 \cdot r_i^t \cdot y_i^t + 1,$$

and (ii) it holds that

$$\left|\Delta\Upsilon^{t+1}\right| \leqslant 4 \cdot \max_{i \in [n]} \left|y_i^t\right| + 2.$$

Proof. First statement. For any bin $i \in [n]$, its expected contribution to Υ^{t+1} is given by,

$$\begin{aligned}
\mathbf{E}\left[\, \Upsilon_i^{t+1} \,\middle|\, y^t \,\right] &= \left(y_i^t + 1 - \frac{1}{n}\right)^2 \cdot r_i^t + \left(y_i^t - \frac{1}{n}\right)^2 \cdot (1 - r_i^t) \\
&= (y_i^t)^2 + 2 \cdot \left(1 - \frac{1}{n}\right) \cdot y_i^t \cdot r_i^t - 2 \cdot \frac{1}{n} \cdot y_i^t \cdot (1 - r_i^t) + \left(1 - \frac{1}{n}\right)^2 \cdot r_i^t + \frac{1}{n^2} \cdot (1 - r_i^t) \\
&= (y_i^t)^2 + 2 \cdot \left(r_i^t - \frac{1}{n}\right) \cdot y_i^t + \left(1 - \frac{1}{n}\right)^2 \cdot r_i^t + \frac{1}{n^2} \cdot (1 - r_i^t).
\end{aligned}$$

Hence, by aggregating over all bins we get,

$$
\begin{aligned}
\mathbf{E}\left[\, \Upsilon^{t+1} \,\middle|\, y^t \,\right] &= \sum_{i=1}^{n} \left[(y_i^t)^2 + 2 \cdot \left(r_i^t - \frac{1}{n} \right) \cdot y_i^t + \left(1 - \frac{1}{n}\right)^2 \cdot r_i^t + \frac{1}{n^2} \cdot (1 - r_i^t) \right] \\
&= \Upsilon^t + \sum_{i=1}^{n} 2 \cdot \left(r_i^t - \frac{1}{n} \right) \cdot y_i^t + \left(1 - \frac{1}{n}\right)^2 + \frac{1}{n} - \frac{1}{n^2} \\
&= \Upsilon^t + \sum_{i=1}^{n} 2 \cdot \left(r_i^t - \frac{1}{n} \right) \cdot y_i^t + 1 - \frac{1}{n} \\
&= \Upsilon^t + \sum_{i=1}^{n} 2 \cdot r_i^t \cdot y_i^t - \sum_{i=1}^{n} \frac{2}{n} \cdot y_i^t + 1 - \frac{1}{n} \\
&\overset{(a)}{=} \Upsilon^t + \sum_{i=1}^{n} 2 \cdot r_i^t \cdot y_i^t + 1 - \frac{1}{n} \\
&\leqslant \Upsilon^t + \sum_{i=1}^{n} 2 \cdot r_i^t \cdot y_i^t + 1,
\end{aligned}
$$

using in (a) that $\sum_{i=1}^{n} y_i^t = 0$. Therefore, by subtracting Υ^t, statement (i) follows.

Second statement. Let $M := \max_{i \in [n]} |y_i^t|$. We will upper bound the change $\Delta\Upsilon_i^{t+1}$ for an arbitrary bin $i \in [n]$, by considering the following two cases:

Case 1: Ball at step $t+1$ is allocated to bin i. So,

$$
\left|\Delta\Upsilon_i^{t+1}\right| = \left| \left(y_i^t + 1 - \frac{1}{n} \right)^2 - (y_i^t)^2 \right| = \left| 2 \cdot \left(1 - \frac{1}{n}\right) \cdot y_i^t + \left(1 - \frac{1}{n}\right)^2 \right| \leqslant 2M + 1.
$$

Case 2: Ball at step $t+1$ is not allocated to bin i. So,

$$
\left|\Delta\Upsilon_i^{t+1}\right| = \left| \left(y_i^t - \frac{1}{n} \right)^2 - (y_i^t)^2 \right| = \left| -\frac{2}{n} \cdot y_i^t + \frac{1}{n^2} \right| \leqslant \frac{2M}{n} + \frac{1}{n}.
$$

Aggregating over all bins $i \in [n]$ yields

$$
\left|\Delta\Upsilon^{t+1}\right| \leqslant \sum_{i=1}^{n} \left|\Delta\Upsilon_i^{t+1}\right| \leqslant 2M + 1 + (n-1) \cdot \left(\frac{2M}{n} + \frac{1}{n} \right) \leqslant 4M + 2. \qquad \square
$$

We now use the general formula in Lemma 5.15 (i) to obtain an expression for the expected change of the quadratic potential for Two-Choice without noise.

Lemma 5.16. *Consider the* Two-Choice = Time-Homogeneous(p) *process without noise, where* $p_i = \frac{2i-1}{n^2}$ *for any* $i \in [n]$. *Then, for any step* $t \geqslant 0$, *it holds that*

$$
\mathbf{E}\left[\, \Delta\Upsilon^{t+1} \,\middle|\, y^t \,\right] \leqslant \sum_{i=1}^{n} 2 \cdot p_i \cdot \tilde{y}_i^t + 1 \leqslant -\frac{\Delta^t}{n} + 1.
$$

Proof. Applying Lemma 5.15 (i) to Two-Choice yields

$$
\mathbf{E}\left[\, \Delta\Upsilon^{t+1} \,\middle|\, y^t \,\right] \leqslant \sum_{i=1}^{n} 2 \cdot p_i \cdot \tilde{y}_i^t + 1.
$$

Let $B_+^t := \{i \in [n] \colon \widetilde{y}_i^t \geqslant 0\}$ be the set of overloaded bins at step t and $B_-^t := \{i \in [n] \colon \widetilde{y}_i^t < 0\}$, the set of underloaded bins. The TWO-CHOICE process allocates a ball to the set of overloaded bins with probability $|B_+^t|^2/n^2$, and thus the average allocation probability across overloaded bins is $p_+^t = |B_+^t|/n^2$. Consequently, TWO-CHOICE allocates to the set of underloaded bins with probability $1 - |B_+^t|^2/n^2$, and thus the average allocation probability across underloaded bins is

$$p_-^t = \frac{1}{|B_-^t|} \cdot \left(1 - \frac{|B_+^t|^2}{n^2}\right) = \frac{1}{n - |B_+^t|} \cdot \frac{(n + |B_+^t|) \cdot (n - |B_+^t|)}{n^2} = \frac{1}{n} + \frac{|B_+^t|}{n^2}.$$

By splitting the sum $\sum_{i=1}^n 2 \cdot p_i \cdot \widetilde{y}_i^t$ into underloaded and overloaded bins, we get

$$\sum_{i=1}^n 2 \cdot p_i \cdot \widetilde{y}_i^t = \sum_{i \in B_+^t} 2 \cdot p_i \cdot \widetilde{y}_i^t + \sum_{i \in B_-^t} 2 \cdot p_i \cdot \widetilde{y}_i^t.$$

Since p_i is non-decreasing, we have $\sum_{i=1}^j p_i \leqslant \sum_{i=1}^j p_+^t$ for all $1 \leqslant j \leqslant |B_+^t|$. Further, since $\widetilde{y}^t$ is non-increasing over the overloaded bins, by Lemma B.2 we have

$$\sum_{i \in B_+^t} 2 \cdot p_i \cdot \widetilde{y}_i^t \leqslant \sum_{i \in B_+^t} 2 \cdot p_+^t \cdot \widetilde{y}_i^t = 2 \cdot p_+^t \cdot \sum_{i \in B_+^t} \widetilde{y}_i^t = \frac{|B_+^t|}{n^2} \cdot \Delta^t,$$

since $\sum_{i \in B_-^t} \widetilde{y}_i^t = -\sum_{i \in B_+^t} \widetilde{y}_i^t$ and thus $\sum_{i \in B_+^t} \widetilde{y}_i^t = \frac{1}{2}\Delta^t$. Analogously, since $\widetilde{y}^t$ is non-increasing over the underloaded bins,

$$\sum_{i \in B_-^t} 2 \cdot p_i \cdot \widetilde{y}_i^t \leqslant \sum_{i \in B_-^t} 2 \cdot p_-^t \cdot \widetilde{y}_i^t = 2 \cdot p_-^t \cdot \sum_{i \in B_-^t} \widetilde{y}_i^t = \left(\frac{1}{n} + \frac{|B_+^t|}{n^2}\right) \cdot \Delta^t.$$

Combining these we get

$$\mathbf{E}\left[\, \Delta\Upsilon^{t+1} \,\middle|\, y^t \,\right] \leqslant \sum_{i=1}^n 2 \cdot p_i \cdot \widetilde{y}_i^t + 1 \leqslant \frac{|B_+^t|}{n^2} \cdot \Delta^t - \left(\frac{1}{n} + \frac{|B_+^t|}{n^2}\right) \cdot \Delta^t + 1 = -\frac{\Delta^t}{n} + 1. \qquad \square$$

Now we relate the change of the quadratic potential for the g-ADV-COMP setting to the change of the quadratic potential for TWO-CHOICE without noise, using that the adversary can determine (and possibly revert) a load comparison between y_i^t and y_j^t only if $|y_i^t - y_j^t| \leqslant g$.

Lemma 5.17. *Consider the g-ADV-COMP setting for any $g \geqslant 1$. Then, for any step $t \geqslant 0$,*

$$\mathbf{E}\left[\, \Upsilon^{t+1} \,\middle|\, y^t \,\right] \leqslant \Upsilon^t - \frac{\Delta^t}{n} + 2g + 1.$$

Proof. By Lemma 5.15 (i), for the g-ADV-COMP probability allocation vector q^t we have,

$$\mathbf{E}\left[\, \Delta\Upsilon^{t+1} \,\middle|\, y^t \,\right] \leqslant \sum_{i=1}^n 2 \cdot q_i^t \cdot y_i^t + 1 = \sum_{i=1}^n 2 \cdot \widetilde{q}_i^t \cdot \widetilde{y}_i^t + 1.$$

This sorted allocation vector $\widetilde{q}^t$ is obtained from the probability vector p of TWO-CHOICE without noise by moving a probability of up to $\frac{2}{n^2}$ from any bin j to a bin i with $\widetilde{y}_j^t < \widetilde{y}_i^t \leqslant \widetilde{y}_j^t + g$. Recalling that $R^t := \{(i,j) \in [n] \times [n] \colon \widetilde{y}_j^t < \widetilde{y}_i^t \leqslant \widetilde{y}_j^t + g\}$,

$$\mathbf{E}\left[\, \Delta\Upsilon^{t+1} \,\middle|\, y^t \,\right] \leqslant \sum_{i=1}^n 2 \cdot p_i \cdot \widetilde{y}_i^t + 1 + 2 \cdot \sum_{(i,j) \in R^t} \frac{2}{n^2} \cdot (\widetilde{y}_i^t - \widetilde{y}_j^t)$$

$$\leqslant \sum_{i=1}^{n} 2 \cdot p_i \cdot \widetilde{y}_i^t + 1 + 2 \cdot \sum_{(i,j) \in R^t} \frac{2}{n^2} \cdot g$$
$$\leqslant \sum_{i=1}^{n} 2 \cdot p_i \cdot \widetilde{y}_i^t + 1 + 2g,$$

using that $|R^t| < \frac{1}{2}n^2$. Hence, using Lemma 5.16, we conclude that

$$\mathbf{E}\big[\, \Delta\Upsilon^{t+1} \,\big|\, y^t \,\big] \leqslant -\frac{\Delta^t}{n} + 2g + 1. \qquad \square$$

5.3.3 Constant fraction of good steps

We define a step $s \geqslant 0$ to be a *good step* if $\mathcal{G}^s := \{\Delta^s \leqslant Dng\}$ holds, for $D := 365$. Further, $G_{t_0}^{t_1} := G_{t_0}^{t_1}(D)$ denotes the number of good steps in $[t_0, t_1]$. Later, in Section 5.3.4 we will show that in a good step, the exponential potential Λ with any sufficiently small constant α drops in expectation.

In the following lemma we show that at least a constant fraction r of the steps are good in a sufficiently long interval. We will apply this lemma with two different values for T: (i) in the recovery phase, to prove that there exists a step $s \in [m - \Theta(ng \cdot (\log(ng))^2), m]$ with $\Lambda^s = \mathcal{O}(n)$ and (ii) in the stabilisation phase, to prove that every $\mathcal{O}(n \cdot (g + \log n))$ steps there exists a step s with $\Lambda^s = \mathcal{O}(n)$. In the analysis below, we pick $r := \frac{6}{6+\epsilon}$, where $\epsilon := \frac{1}{12}$.

Lemma 5.18. *Consider the g-*Adv-Comp* setting for any $g \geqslant 1$ and let $r := \frac{6}{6+\epsilon}$, $\epsilon := 1/12$ and $D := 365$. Then, for any constant $\hat{c} \geqslant 1$ and any $T \in [ng^2, n^2g^3/\hat{c}]$, we have for any steps $t_0 \geqslant 0$ and $t_1 := t_0 + \hat{c} \cdot T \cdot g^{-1} - 1$,*

$$\mathbf{Pr}\left[G_{t_0}^{t_1}(D) \geqslant r \cdot (t_1 - t_0 + 1) \,\middle|\, \mathfrak{F}^{t_0}, \Upsilon^{t_0} \leqslant T, \max_{i \in [n]} \left| y_i^{t_0} \right| \leqslant g(\log(ng))^2 \right] \geqslant 1 - 2 \cdot (ng)^{-12}.$$

Proof. We define the sequence $(Z^t)_{t \geqslant t_0}$ with $Z^{t_0} := \Upsilon^{t_0}$ and for any $t > t_0$,

$$Z^t := \Upsilon^t + \sum_{s=t_0}^{t-1} \left(\frac{\Delta^s}{n} - 2g - 1 \right).$$

This sequence forms a super-martingale since by Lemma 5.17,

$$\mathbf{E}\big[\, Z^{t+1} \,\big|\, \mathfrak{F}^t \,\big] = \mathbf{E}\left[\Upsilon^{t+1} + \sum_{s=t_0}^{t} \left(\frac{\Delta^s}{n} - 2g - 1 \right) \,\middle|\, \mathfrak{F}^t \right]$$
$$\leqslant \Upsilon^t - \frac{\Delta^t}{n} + 2g + 1 + \sum_{s=t_0}^{t} \left(\frac{\Delta^s}{n} - 2g - 1 \right)$$
$$= \Upsilon^t + \sum_{s=t_0}^{t-1} \left(\frac{\Delta^s}{n} - 2g - 1 \right) = Z^t.$$

Further, let $\tau := \inf\left\{ t \geqslant t_0 \colon \max_{i \in [n]} |y_i^t| > 2g(\log(ng))^2 \right\}$ and consider the stopped random variable

$$\widetilde{Z}^t := Z^{t \wedge \tau},$$

which is then also a super-martingale. Applying Corollary 3.23 and the union bound over steps $[t_0, t_1]$, we get

$$\mathbf{Pr}\left[\tau \leqslant t_1 \,\middle|\, \mathfrak{F}^{t_0}, \Upsilon^{t_0} \leqslant T, \max_{i\in[n]} \left|y_i^{t_0}\right| \leqslant g(\log(ng))^2 \right] \leqslant (\hat{c} \cdot T \cdot g^{-1}) \cdot (ng)^{-14} \leqslant (ng)^{-12}, \tag{5.7}$$

using that $T \leqslant n^2 g^3/\hat{c}$. This means that the maximum absolute normalised load does not increase above $2g(\log(ng))^2$ in any of the steps in $[t_0, t_1]$ w.h.p.

To prove concentration of $\widetilde{Z}^{t_1+1}$, we will now derive an upper bound on the difference $|\widetilde{Z}^{t+1} - \widetilde{Z}^t|$:

Case 1 $[t \geqslant \tau]$: In this case, $\widetilde{Z}^{t+1} = Z^{(t+1)\wedge\tau} = Z^\tau$, and similarly, $\widetilde{Z}^t = Z^{t\wedge\tau} = Z^\tau$, so $|\widetilde{Z}^{t+1} - \widetilde{Z}^t| = 0$.

Case 2 $[t < \tau]$: In this case, we have $\max_{i\in[n]} |y_i^t| \leqslant 2g(\log(ng))^2$ and by Lemma 5.15 (ii), we have that $|\Delta\Upsilon^{t+1}| \leqslant 8c_3 g(\log(ng))^2 + 2$. This implies that

$$|\widetilde{Z}^{t+1} - \widetilde{Z}^t| \leqslant |\Delta\Upsilon^{t+1}| + \left|\frac{\Delta^t}{n} - 2g - 1\right| \leqslant 8g(\log(ng))^2 + 2 + (2g(\log(ng))^2 - 2g - 1) \leqslant 10g(\log(ng))^2.$$

Combining the two cases above, we conclude that for all $t \geqslant t_0$,

$$|\widetilde{Z}^{t+1} - \widetilde{Z}^t| \leqslant 10g(\log(ng))^2.$$

Using Azuma's inequality for super-martingales (Lemma B.10) for $\lambda = T$ and $a_i = 10g(\log(ng))^2$,

$$\begin{aligned}
&\mathbf{Pr}\left[\widetilde{Z}^{t_1+1} - \widetilde{Z}^{t_0} \geqslant T \,\middle|\, \mathfrak{F}^{t_0}, \Upsilon^{t_0} \leqslant T, \max_{i\in[n]} \left|y_i^{t_0}\right| \leqslant g(\log(ng))^2 \right] \\
&\quad \leqslant \exp\left(-\frac{T^2}{2 \cdot \sum_{t=t_0}^{t_1} (10g(\log(ng))^2)^2} \right) \\
&\quad = \exp\left(-\frac{T^2}{\hat{c} \cdot T \cdot g^{-1} \cdot 200 \cdot g^2 \cdot (\log(ng))^4} \right) \\
&\quad = \exp\left(-\frac{T}{200 \cdot \hat{c} \cdot g \cdot (\log(ng))^4} \right) \\
&\quad \overset{(a)}{\leqslant} \exp\left(-\frac{ng^2}{200 \cdot \hat{c} \cdot g \cdot (\log(ng))^4} \right) = (ng)^{-\omega(1)},
\end{aligned}$$

where in (a) we used that $T \geqslant ng^2$. Hence, we conclude that

$$\mathbf{Pr}\left[\widetilde{Z}^{t_1+1} < \widetilde{Z}^{t_0} + T \,\middle|\, \mathfrak{F}^{t_0}, \Upsilon^{t_0} \leqslant T, \max_{i\in[n]} \left|y_i^{t_0}\right| \leqslant g(\log(ng))^2 \right] \geqslant 1 - (ng)^{-\omega(1)}.$$

Thus by taking the union bound with Eq. (5.7) we have

$$\mathbf{Pr}\left[Z^{t_1+1} < Z^{t_0} + T \,\middle|\, \mathfrak{F}^{t_0}, \Upsilon^{t_0} \leqslant T, \max_{i\in[n]} \left|y_i^{t_0}\right| \leqslant g(\log(ng))^2 \right] \geqslant 1 - 2 \cdot (ng)^{-12}.$$

For the sake of a contradiction, assume now that more than an $(1-r)$ fraction of the steps $t \in [t_0, t_1]$ satisfy $\Delta^t > Dng$. This implies that

$$\sum_{t=t_0}^{t_1} \frac{\Delta^t}{n} > Dg \cdot (1-r) \cdot (t_1 - t_0 + 1) = D \cdot (1-r) \cdot \hat{c} \cdot T, \tag{5.8}$$

using that $t_1 - t_0 + 1 = \hat{c} \cdot T \cdot g^{-1}$. When $\{Z^{t_1+1} < Z^{t_0} + T\}$ and $\{\Upsilon^{t_0} \leqslant T\}$ hold, then we have

$$\Upsilon^{t_1+1} + \sum_{t=t_0}^{t_1} \frac{\Delta^t}{n} - (2g+1) \cdot (t_1 - t_0 + 1) < \Upsilon^{t_0} + T \leqslant 2T.$$

By rearranging this leads to a contradiction as

$$\begin{aligned} 0 \leqslant \Upsilon^{t_1+1} &< 2T - \sum_{t=t_0}^{t_1} \frac{\Delta^t}{n} + (2g+1) \cdot (t_1 - t_0 + 1) \\ &\overset{(a)}{=} 2T - \sum_{t=t_0}^{t_1} \frac{\Delta^t}{n} + \hat{c}(2g+1) \cdot T \cdot g^{-1} \\ &\overset{(b)}{\leqslant} - \sum_{t=t_0}^{t_1} \frac{\Delta^t}{n} + 5\hat{c} \cdot T \\ &\overset{(5.8)}{<} -D \cdot (1-r) \cdot \hat{c} \cdot T + 5\hat{c} \cdot T \\ &\overset{(c)}{=} 0 \end{aligned}$$

using in (a) that $t_1 - t_0 + 1 = \hat{c} \cdot T \cdot g^{-1}$, in (b) that $\hat{c} \geqslant 1$ and $g \geqslant 1$, and in (c) that $D = \frac{5}{1-r} = 365$ (as $r = \frac{6}{6+1/12}$).

We conclude that when $\{Z^{t_1+1} < Z^{t_0} + T\}$ holds, then at least an r fraction of the steps $t \in [t_0, t_1]$ satisfy $\Delta^t \leqslant Dng$, and thus,

$$\mathbf{Pr}\left[G_{t_0}^{t_1} \geqslant r \cdot (t_1 - t_0 + 1) \;\middle|\; \mathfrak{F}^{t_0}, \Upsilon^{t_0} \leqslant T, \max_{i \in [n]} \left| y_i^{t_0} \right| \leqslant g(\log(ng))^2 \right] \geqslant 1 - 2 \cdot (ng)^{-12}. \qquad \square$$

The following lemma provides two ways of upper bounding the quadratic potential using the exponential potential Λ. These will be used in the recovery and stabilisation lemmas, to obtain the starting point condition ($\Upsilon^{t_0} \leqslant T$) for Lemma 5.18.

Lemma 5.19. *Consider the potential $\Lambda := \Lambda(\alpha, c_4 g)$ for any constant $\alpha \in (0,1)$, any $g \geqslant 1$ and any constant $c_4 > 0$. Then (i) for any constant $\hat{c} > 0$, there exists a constant $c_s := c_s(\alpha, c_4, \hat{c}) \geqslant 1$, such that for any step $t \geqslant 0$ with $\Lambda^t \leqslant \hat{c} \cdot n$,*

$$\Upsilon^t \leqslant c_s \cdot ng^2.$$

Furthermore, (ii) there exists a constant $c_r := c_r(\alpha, c_4) \geqslant 1$, such that for any step $t \geqslant 0$,

$$\Upsilon^t \leqslant c_r \cdot n \cdot \left(g^2 + (\log \Lambda^t)^2 \right).$$

Proof. First statement. We begin by proving some basic inequalities between exponential, quadratic and linear terms. Let $\hat{u} := (4/\alpha) \cdot \log(4/\alpha)$. Note that $e^u \geqslant u$ (for any $u \geqslant 0$) and hence for any $u \geqslant \hat{u}$,

$$e^{\alpha u/2} = e^{\alpha u/4} \cdot e^{\alpha u/4} \geqslant \frac{\alpha u}{4} \cdot e^{\alpha \hat{u}/4} = \frac{\alpha u}{4} \cdot \frac{4}{\alpha} = u,$$

and $e^{\alpha u} = e^{\alpha u/2} \cdot e^{\alpha u/2} \geqslant u \cdot u = u^2$. Therefore, for every $u \geqslant 0$,

$$u^2 \leqslant \max\left\{ \hat{u}^2, e^{\alpha u} \right\}. \tag{5.9}$$

Recall that for any bin $i \in [n]$, $\Lambda_i^t := e^{\alpha \cdot (y_i^t - c_4 g)^+} + e^{\alpha \cdot (-y_i^t - c_4 g)^+}$. Hence,

$$\left((y_i^t - c_4 g)^+\right)^2 + \left((-y_i^t - c_4 g)^+\right)^2 \overset{(a)}{\leqslant} \max\left\{2\hat{u}^2, \Lambda_i^t\right\} \overset{(b)}{\leqslant} \max\left\{2\hat{u}^2 \cdot \Lambda_i^t, \Lambda_i^t\right\} \overset{(c)}{=} 2\hat{u}^2 \cdot \Lambda_i^t, \tag{5.10}$$

where in (a) we used Eq. (5.9) first with $u = (y_i^t - c_4 g)^+$ and then with $u = (-y_i^t - c_4 g)^+$, in (b) that $\Lambda_i^t \geqslant 1$ for any $i \in [n]$ and in (c) that $\hat{u} \geqslant 1$, since $\alpha \in (0, 1)$.

We now proceed to upper bound the quadratic potential,

$$\begin{aligned}
\Upsilon^t &\leqslant \sum_{i=1}^{n} \Big[\left((y_i^t - c_4 g)^+ + c_4 g\right)^2 + \left((-y_i^t - c_4 g)^+ + c_4 g\right)^2 \Big] \\
&\overset{(a)}{\leqslant} 2 \cdot \sum_{i=1}^{n} \Big[\left((y_i^t - c_4 g)^+\right)^2 + \left((-y_i^t - c_4 g)^+\right)^2 + 2 \cdot (c_4 g)^2 \Big] \\
&\overset{(5.10)}{\leqslant} 4\hat{u}^2 \cdot \Lambda^t + 4c_4^2 \cdot n g^2 \\
&\overset{(b)}{\leqslant} 4\hat{c}\hat{u}^2 \cdot n + 4c_4^2 \cdot n g^2 \\
&\leqslant (4\hat{c}\hat{u}^2 + 4c_4^2) \cdot n g^2,
\end{aligned}$$

using in (a) that $(a+b)^2 \leqslant 2 \cdot (a^2 + b^2)$ (for any a, b) and in (b) that $\Lambda^t \leqslant \hat{c} \cdot n$. Therefore, for the constant $c_s := c_s(\alpha, c_4, \hat{c}) := 4\hat{c}\hat{u}^2 + 4c_4^2$, we get the first statement.

Second statement. For any bin $i \in [n]$ we have,

$$|y_i^t| \leqslant c_4 g + \frac{1}{\alpha} \log \Lambda^t.$$

Hence, using that $(a+b)^2 \leqslant 2 \cdot (a^2 + b^2)$,

$$(y_i^t)^2 \leqslant \left(c_4 g + \frac{1}{\alpha} \log \Lambda^t\right)^2 \leqslant 2 \cdot \left(c_4^2 g^2 + \frac{1}{\alpha^2} \cdot (\log \Lambda^t)^2\right) \leqslant c_r \cdot \left(g^2 + (\log \Lambda^t)^2\right),$$

for some constant $c_r := c_r(\alpha, c_4) = \max\left\{2c_4^2, \frac{2}{\alpha^2}\right\} \geqslant 1$. By aggregating the contributions over all bins, we get the second statement. □

5.3.4 Hyperbolic cosine potential

We now prove bounds on the expected change of the Λ potential function over one step. Note that these hold for any sufficiently small constant $\alpha > 0$.

We start with a relatively weak bound which holds at any step.

Lemma 5.20. *Consider any* SEQUENTIAL(r^t) *process such that* $\max_{i \in [n]} r_i^t \leqslant \frac{2}{n}$ *for any step* $t \geqslant 0$. *Further, consider the potential* $\Lambda := \Lambda(\alpha, c_4 g)$ *for any* $\alpha \in \left(0, \frac{1}{2}\right]$, *any* $g \geqslant 1$ *and any* $c_4 > 0$. *Then, for any step* $t \geqslant 0$, (i) *for every bin* $i \in [n]$ *it holds that*

$$\mathbf{E}\left[\, \Lambda_i^{t+1} \,\middle|\, \mathfrak{F}^t \,\right] \leqslant \Lambda_i^t \cdot \left(1 + \frac{3\alpha}{n}\right).$$

Furthermore, by aggregating over all bins, (ii) *it holds that*

$$\mathbf{E}\left[\, \Lambda^{t+1} \,\middle|\, \mathfrak{F}^t \,\right] \leqslant \Lambda^t \cdot \left(1 + \frac{3\alpha}{n}\right).$$

Proof. Consider an arbitrary bin $i \in [n]$. We upper bound the change of the overloaded component, i.e., $e^{\alpha \cdot (y_i^t - c_4 g)^+}$, by placing one ball in bin $i \in [n]$ with probability $r_i^t \leqslant \frac{2}{n}$ and ignoring the change of the average load. Also, we upper bound the change for the underloaded component, i.e., $e^{\alpha \cdot (-y_i^t - c_4 g)^+}$, by considering only the change of the average load by $1/n$. Hence,

$$\begin{aligned}
\mathbf{E}\left[\, \Lambda_i^{t+1} \,\middle|\, \mathfrak{F}^t \,\right] &\leqslant e^{\alpha \cdot (y_i^t - c_4 g)^+} \cdot ((1 - r_i^t) + r_i^t \cdot e^{\alpha}) + e^{\alpha \cdot (-y_i^t - c_4 g)^+} \cdot ((1 - r_i^t) \cdot e^{\alpha/n} + r_i^t \cdot e^{\alpha/n}) \\
&= e^{\alpha \cdot (y_i^t - c_4 g)^+} \cdot (1 + r_i^t \cdot (e^{\alpha} - 1)) + e^{\alpha \cdot (-y_i^t - c_4 g)^+} \cdot e^{\alpha/n} \\
&\overset{(a)}{\leqslant} e^{\alpha \cdot (y_i^t - c_4 g)^+} \cdot \Big(1 + \frac{2}{n} \cdot 1.5\alpha\Big) + e^{\alpha \cdot (-y_i^t - c_4 g)^+} \cdot \Big(1 + 1.5\frac{\alpha}{n}\Big) \\
&\leqslant \Lambda_i^t \cdot \Big(1 + \frac{3\alpha}{n}\Big),
\end{aligned}$$

using in (a) that $e^u \leqslant 1 + 1.5u$ (for any $0 \leqslant u \leqslant 0.7$), $\alpha \leqslant 1/2$ and $r_i^t \leqslant \frac{2}{n}$. $\square$

Now we improve this bound for any good step t, i.e., when $\Delta^t \leqslant Dng$ holds.

Lemma 5.21. *Consider the g-*Adv-Comp* setting for any $g \geqslant 1$ and the potential $\Lambda := \Lambda(\alpha, c_4 g)$ for any $\alpha \in \big(0, \frac{1}{18}\big]$, $c_4 := 2D$ and $D := 365$. Then, for any step $t \geqslant 0$, (i) for $\epsilon := \frac{1}{12}$, it holds that*

$$\mathbf{E}\left[\, \Lambda^{t+1} \,\middle|\, \mathfrak{F}^t, \Delta^t \leqslant Dng \,\right] \leqslant \Lambda^t \cdot \Big(1 - \frac{2\alpha\epsilon}{n}\Big) + 18\alpha.$$

Furthermore, this also implies that (ii) for $c := \frac{18}{\epsilon}$, it holds that

$$\mathbf{E}\left[\, \Lambda^{t+1} \,\middle|\, \mathfrak{F}^t, \Delta^t \leqslant Dng, \Lambda^t > cn \,\right] \leqslant \Lambda^t \cdot \Big(1 - \frac{\alpha\epsilon}{n}\Big).$$

Proof. First statement. Consider an arbitrary step $t \geqslant 0$ with $\Delta^t \leqslant Dng$. We bound the expected change of Λ over one step, by considering the following cases for each bin $i \in [n]$:

Case 1 [$y_i^t \in (-c_4 g - 2, c_4 g + 2)$]: Using Lemma 5.20 (i),

$$\begin{aligned}
\mathbf{E}\left[\, \Lambda_i^{t+1} \,\middle|\, \mathfrak{F}^t \,\right] &\leqslant \Lambda_i^t \cdot \Big(1 + \frac{3\alpha}{n}\Big) \\
&= \Lambda_i^t \cdot \Big(1 - \frac{\alpha}{6n}\Big) + \Lambda_i^t \cdot \Big(\frac{\alpha}{6n} + \frac{3\alpha}{n}\Big) \\
&\overset{(a)}{\leqslant} \Lambda_i^t \cdot \Big(1 - \frac{\alpha}{6n}\Big) + (2 \cdot e^{2\alpha}) \cdot \Big(\frac{1}{6} + 3\Big) \cdot \frac{\alpha}{n} \\
&\overset{(b)}{\leqslant} \Lambda_i^t \cdot \Big(1 - \frac{\alpha}{6n}\Big) + \frac{18\alpha}{n},
\end{aligned}$$

using in (a) that $\Lambda_i^t \leqslant e^{2\alpha} + 1 \leqslant 2 \cdot e^{2\alpha}$ by the assumption that $y_i^t \in (-c_4 g - 2, c_4 g + 2)$ and in (b) that $(2 \cdot e^{2\alpha}) \cdot (\frac{1}{6} + 3) \leqslant 18$, since $\alpha \leqslant 1/4$.

Case 2 [$y_i^t \geqslant c_4 g + 2$]: By the condition $\Delta^t \leqslant Dng$, the number of bins j with $y_j^t \geqslant \frac{3}{2}Dg$ is at most $\frac{1}{2}\Delta^t \cdot \frac{2}{3Dg} \leqslant \frac{Dng}{3Dg} = \frac{n}{3}$ (see Fig. 5.1 (**top**)). We allocate to bin $i \in [n]$ with $y_i^t \geqslant c_4 g + 2 = 2Dg + 2 \geqslant \frac{3}{2}Dg + g$ if we sample bin i and a bin j with $y_j^t \geqslant \frac{3}{2}Dg$. Hence,

$$q_i^t \leqslant 2 \cdot \frac{1}{n} \cdot \frac{1}{3} = \frac{2}{3n}.$$

By assumption, bin i deterministically satisfies $y_i^{t+1} \geqslant c_4 g + 2 - 1/n > c_4 g$, so

$$\mathbf{E}\left[\, \Lambda_i^{t+1} \,\middle|\, \mathfrak{F}^t, \Delta^t \leqslant Dng \,\right]$$

$$
\begin{aligned}
&= e^{\alpha\cdot(y_i^t-c_4g)^+}\cdot\Big((1-q_i^t)\cdot e^{-\alpha/n}+q_i^t\cdot e^{\alpha(1-1/n)}\Big)+1\\
&\overset{(a)}{\leqslant} e^{\alpha\cdot(y_i^t-c_4g)^+}\cdot\left((1-q_i^t)\cdot\Big(1-\frac{\alpha}{n}+\frac{\alpha^2}{n^2}\Big)+q_i^t\cdot\Big(1+\alpha\cdot\Big(1-\frac{1}{n}\Big)+\alpha^2\Big)\right)+1\\
&= e^{\alpha\cdot(y_i^t-c_4g)^+}\cdot\Big(1+\alpha\cdot\Big(q_i^t-\frac{1}{n}\Big)+(1-q_i^t)\cdot\frac{\alpha^2}{n^2}+q_i^t\cdot\alpha^2\Big)+1\\
&\overset{(b)}{\leqslant} e^{\alpha\cdot(y_i^t-c_4g)^+}\cdot\Big(1-\frac{\alpha}{3n}+\frac{4\alpha^2}{3n}\Big)+1\\
&\overset{(c)}{\leqslant} e^{\alpha\cdot(y_i^t-c_4g)^+}\cdot\Big(1-\frac{\alpha}{6n}\Big)+1\\
&= e^{\alpha\cdot(y_i^t-c_4g)^+}\cdot\Big(1-\frac{\alpha}{6n}\Big)+1\cdot\Big(1-\frac{\alpha}{6n}\Big)+\frac{\alpha}{6n}\\
&= \Lambda_i^t\cdot\Big(1-\frac{\alpha}{6n}\Big)+\frac{\alpha}{6n},
\end{aligned}
$$

using in (a) that $e^u \leqslant 1+u+u^2$ for $u<1.75$, $\alpha\leqslant 1$ and $(1-1/n)^2\leqslant 1$, in (b) that $q_i^t \leqslant \frac{2}{3n}$ and $(1-q_i^t)\cdot\frac{\alpha^2}{n^2}\leqslant\frac{2\alpha^2}{3n}$ for $n\geqslant 2$, and in (c) that $\alpha\leqslant\frac{1}{8}$, so $\frac{4\alpha^2}{3n}\leqslant\frac{\alpha}{6n}$.

Case 3 $[y_i^t \leqslant -c_4g-2]$**:** The number of bins j with $y_j^t\leqslant -\frac{3}{2}Dg$ is at most $\frac{1}{2}\Delta^t\cdot\frac{2}{3Dg}\leqslant\frac{Dng}{3Dg}=\frac{n}{3}$ and the number of bins j with $y_j^t > -\frac{3}{2}Dg$ is at least $\frac{2n}{3}$ (see Fig. 5.1 (**bottom**)). Similarly to Case 2, we can allocate to a bin $i\in[n]$ with load $y_i^t\leqslant -c_4g-2$ only if we sample i and a bin j with $y_j^t > -\frac{3}{2}Dg$. Hence,

$$
q_i^t \geqslant 2\cdot\frac{1}{n}\cdot\frac{2}{3}=\frac{4}{3n}.
$$

By assumption, bin i deterministically satisfies $y_i^{t+1}\leqslant -c_4g-2+1<-c_4g$, so

$$
\begin{aligned}
&\mathbf{E}\big[\,\Lambda_i^{t+1}\,\big|\,\mathfrak{F}^t,\Delta^t\leqslant Dng\,\big]\\
&\quad= 1+e^{\alpha(-y_i^t-c_4g)^+}\cdot\Big((1-q_i^t)\cdot e^{\alpha/n}+q_i^t\cdot e^{-\alpha(1-1/n)}\Big)\\
&\quad\overset{(a)}{\leqslant} 1+e^{\alpha(-y_i^t-c_4g)^+}\cdot\left((1-q_i^t)\cdot\Big(1+\frac{\alpha}{n}+\frac{\alpha^2}{n^2}\Big)+q_i^t\cdot\Big(1-\alpha\cdot\Big(1-\frac{1}{n}\Big)+\alpha^2\Big)\right)\\
&\quad= 1+e^{\alpha(-y_i^t-c_4g)^+}\cdot\Big(1+\alpha\cdot\Big(\frac{1}{n}-q_i^t\Big)+(1-q_i^t)\cdot\frac{\alpha^2}{n^2}+q_i^t\cdot\alpha^2\Big)\\
&\quad\overset{(b)}{\leqslant} 1+e^{\alpha(-y_i^t-c_4g)^+}\cdot\Big(1-\frac{\alpha}{3n}+\frac{5\alpha^2}{2n}\Big)\\
&\quad\overset{(c)}{\leqslant} 1+e^{\alpha(-y_i^t-c_4g)^+}\cdot\Big(1-\frac{\alpha}{6n}\Big)\\
&\quad= \Big(1-\frac{\alpha}{6n}\Big)+e^{\alpha(-y_i^t-c_4g)^+}\cdot\Big(1-\frac{\alpha}{6n}\Big)+\frac{\alpha}{6n}\\
&\quad= \Lambda_i^t\cdot\Big(1-\frac{\alpha}{6n}\Big)+\frac{\alpha}{6n},
\end{aligned}
$$

using in (a) that $e^u\leqslant 1+u+u^2$ for $u<1.75$, $\alpha\leqslant 1$ and $(1-1/n)^2\leqslant 1$, in (b) that $q_i^t\in[\frac{4}{3n},\frac{2}{n}]$ and $(1-q_i^t)\cdot\frac{\alpha^2}{n^2}\leqslant\frac{\alpha^2}{2n}$ for $n\geqslant 2$ and in (c) that $\alpha\leqslant\frac{1}{15}$, so $\frac{5\alpha^2}{2n}\leqslant\frac{\alpha}{6n}$.

Combining these three cases and letting $\epsilon := \frac{1}{12}$, we conclude that

$$
\mathbf{E}\big[\,\Lambda^{t+1}\,\big|\,\mathfrak{F}^t,\Delta^t\leqslant Dng\,\big]\leqslant\sum_{i=1}^n\Big(\Lambda_i^t\cdot\Big(1-\frac{\alpha}{6n}\Big)+\frac{18\alpha}{n}\Big)=\Lambda^t\cdot\Big(1-\frac{2\alpha\epsilon}{n}\Big)+18\alpha.
$$

Second statement. Letting $c := \frac{18}{\epsilon} = 18 \cdot 12$, it follows that

$$\begin{aligned}\mathbf{E}\left[\, \Lambda^{t+1} \,\middle|\, \mathfrak{F}^t, \Delta^t \leqslant Dng, \Lambda^t > cn \,\right] &\leqslant \Lambda^t \cdot \left(1 - \frac{2\alpha\epsilon}{n}\right) + 18\alpha \\ &= \Lambda^t \cdot \left(1 - \frac{\alpha\epsilon}{n}\right) - \Lambda^t \cdot \frac{\alpha\epsilon}{n} + 18\alpha \\ &\leqslant \Lambda^t \cdot \left(1 - \frac{\alpha\epsilon}{n}\right).\end{aligned}$$

□

5.3.5 Adjusted hyperbolic cosine potential

In Lemma 5.21, we proved that in a good step t with $\Lambda^t > cn$ (for $c := 18 \cdot 12$), the potential drops in expectation by a multiplicative factor. Our goal will be to show that w.h.p. $\Lambda^t \leqslant cn$ at a single step (**recovery**) and then show that it becomes small at least once every $\mathcal{O}(n \cdot (g + \log n))$ steps (**stabilisation**). Since we do not have an expected drop in every step, but only at a constant fraction r of the steps, we will define an *adjusted hyperbolic cosine potential function*. First, for any step t_0, and any step $s \geqslant t_0$, we define the following event:

$$\mathcal{E}_{t_0}^s := \bigcap_{t \in [t_0, s]} \left\{\Lambda^t > cn\right\}.$$

Next, we define the sequence $(\widetilde{\Lambda}_{t_0}^s)_{s \geqslant t_0} := (\widetilde{\Lambda}_{t_0}^s)_{s \geqslant t_0}(\alpha, c_4 g, \epsilon)$ as $\widetilde{\Lambda}_{t_0}^{t_0} := \Lambda^{t_0}(\alpha, c_4 g)$ and, for any $s > t_0$,

$$\widetilde{\Lambda}_{t_0}^s := \Lambda^s(\alpha, c_4 g) \cdot \mathbf{1}_{\mathcal{E}_{t_0}^{s-1}} \cdot \exp\left(-\frac{3\alpha}{n} \cdot B_{t_0}^{s-1}\right) \cdot \exp\left(+\frac{\alpha\epsilon}{n} \cdot G_{t_0}^{s-1}\right), \tag{5.11}$$

recalling that $G_{t_0}^{t_1}$ is the number of good steps in $[t_0, t_1]$, i.e., steps where the event $\mathcal{G}^s := \{\Delta^s \leqslant Dng\}$ holds, and $B_{t_0}^{t_1} := [t_0, t_1] \setminus G_{t_0}^{t_1}$ is the number of bad steps in $[t_0, t_1]$.

In the next lemma, we prove that by its definition and Lemmas 5.20 and 5.21, the sequence $(\widetilde{\Lambda}_{t_0}^s)_{s \geqslant t_0}$ is a super-martingale.

Lemma 5.22. *Consider the g-Adv-Comp setting for any $g \geqslant 1$ and the sequence $(\widetilde{\Lambda}_{t_0}^s)_{s \geqslant t_0} := (\widetilde{\Lambda}_{t_0}^s)_{s \geqslant t_0}(\alpha, c_4 g, \epsilon)$ for any starting step $t_0 \geqslant 0$, any $\alpha \in (0, \frac{1}{18}]$ and $\epsilon, c_4 > 0$ as defined in Lemma 5.21. Then, for any step $s \geqslant t_0$,*

$$\mathbf{E}\left[\, \widetilde{\Lambda}_{t_0}^{s+1} \,\middle|\, \mathfrak{F}^s \,\right] \leqslant \widetilde{\Lambda}_{t_0}^s.$$

Proof. Recalling the definition of $\widetilde{\Lambda}$ in Eq. (5.11),

$$\begin{aligned}&\mathbf{E}\left[\, \widetilde{\Lambda}_{t_0}^{s+1} \,\middle|\, \mathfrak{F}^s \,\right] \\ &= \mathbf{E}\left[\, \Lambda^{s+1} \cdot \mathbf{1}_{\mathcal{E}_{t_0}^s} \,\middle|\, \mathfrak{F}^s \,\right] \cdot \exp\left(-\frac{3\alpha}{n} \cdot B_{t_0}^s\right) \cdot \exp\left(\frac{\alpha\epsilon}{n} \cdot G_{t_0}^s\right) \\ &= \mathbf{E}\left[\, \Lambda^{s+1} \cdot \mathbf{1}_{\mathcal{E}_{t_0}^s} \,\middle|\, \mathfrak{F}^s \,\right] \cdot \exp\left(\frac{\alpha\epsilon}{n} \cdot \mathbf{1}_{\mathcal{G}^s} - \frac{3\alpha}{n} \cdot \mathbf{1}_{\neg\mathcal{G}^s}\right) \cdot \exp\left(-\frac{3\alpha}{n} \cdot B_{t_0}^{s-1}\right) \cdot \exp\left(\frac{\alpha\epsilon}{n} \cdot G_{t_0}^{s-1}\right).\end{aligned}$$

Thus, to prove the statement, it suffices to show that

$$\mathbf{E}\left[\, \Lambda^{s+1} \cdot \mathbf{1}_{\mathcal{E}_{t_0}^s} \,\middle|\, \mathfrak{F}^s \,\right] \cdot \exp\left(\frac{\alpha\epsilon}{n} \cdot \mathbf{1}_{\mathcal{G}^s} - \frac{3\alpha}{n} \cdot \mathbf{1}_{\neg\mathcal{G}^s}\right) \leqslant \Lambda^s \cdot \mathbf{1}_{\mathcal{E}_{t_0}^{s-1}}. \tag{5.12}$$

To show Eq. (5.12), we consider two cases based on whether $\mathcal{G}^s$ holds.

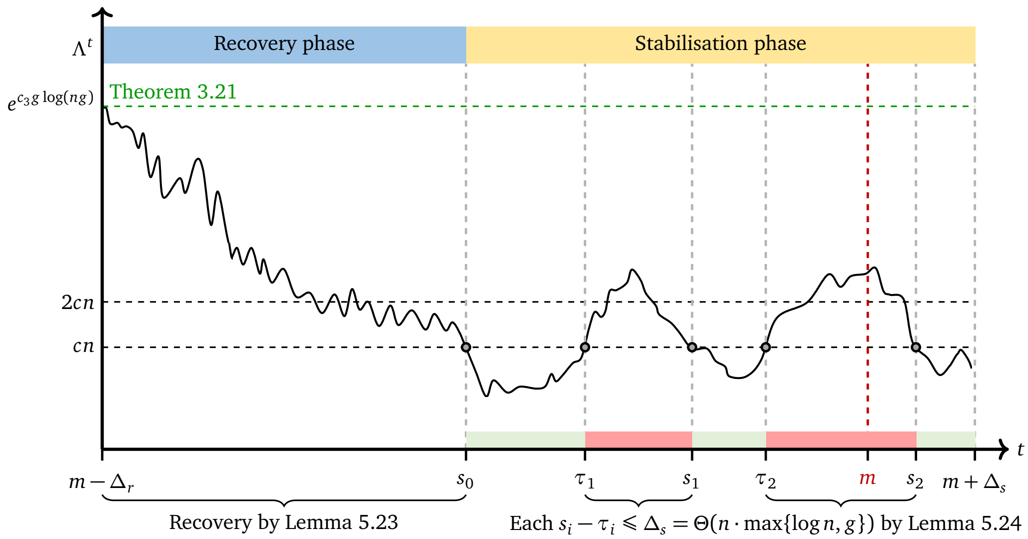


Figure 5.2: Visualisation of the recovery (Lemma 5.23) and stabilisation (Lemma 5.24) phases. Note that the red intervals w.h.p. will have length $\leqslant \Delta_s$, implying $\text{Gap}(m) = \mathcal{O}(g + \log n)$ by smoothness.

Case 1 [$\mathcal{G}^s$ **holds**]**:** Recall that when $\mathcal{G}^s$ holds, then $\Delta^s \leqslant Dng$. Further, when $\Lambda^s \leqslant cn$ holds (for $c > 0$ the constant in Lemma 5.21), then $\mathbf{1}_{\mathcal{E}_{t_0}^s} = 0$. Thus, using Lemma 5.21 (ii),

$$\mathbf{E}\left[\, \Lambda^{s+1} \cdot \mathbf{1}_{\mathcal{E}_{t_0}^s} \,\middle|\, \mathfrak{F}^s, \mathcal{G}^s \,\right] \leqslant \Lambda^s \cdot \mathbf{1}_{\mathcal{E}_{t_0}^{s-1}} \cdot \left(1 - \frac{\alpha\epsilon}{n}\right) \leqslant \Lambda^s \cdot \mathbf{1}_{\mathcal{E}_{t_0}^{s-1}} \cdot \exp\left(-\frac{\alpha\epsilon}{n}\right).$$

Hence, since in this case $\mathbf{1}_{\mathcal{G}^s} = 1$, the left hand side of Eq. (5.12) is equal to

$$\mathbf{E}\left[\, \Lambda^{s+1} \cdot \mathbf{1}_{\mathcal{E}_{t_0}^s} \,\middle|\, \mathfrak{F}^s, \mathcal{G}^s \,\right] \cdot \exp\left(\frac{\alpha\epsilon}{n}\right) \leqslant \left(\Lambda^s \cdot \mathbf{1}_{\mathcal{E}_{t_0}^{s-1}} \cdot \exp\left(-\frac{\alpha\epsilon}{n}\right)\right) \cdot \exp\left(\frac{\alpha\epsilon}{n}\right) = \Lambda^s \cdot \mathbf{1}_{\mathcal{E}_{t_0}^{s-1}}.$$

Case 2 [$\mathcal{G}^s$ **does not hold**]**:** By Lemma 5.20 (ii), we get

$$\mathbf{E}\left[\, \Lambda^{s+1} \cdot \mathbf{1}_{\mathcal{E}_{t_0}^s} \,\middle|\, \mathfrak{F}^s, \neg\mathcal{G}^s \,\right] \leqslant \Lambda^s \cdot \mathbf{1}_{\mathcal{E}_{t_0}^{s-1}} \cdot \left(1 + \frac{3\alpha}{n}\right) \leqslant \Lambda^s \cdot \mathbf{1}_{\mathcal{E}_{t_0}^{s-1}} \cdot \exp\left(\frac{3\alpha}{n}\right).$$

Hence, since in this case $\mathbf{1}_{\mathcal{G}^s} = 0$, the left hand side of Eq. (5.12) is equal to

$$\mathbf{E}\left[\, \Lambda^{s+1} \cdot \mathbf{1}_{\mathcal{E}_{t_0}^s} \,\middle|\, \mathfrak{F}^s, \neg\mathcal{G}^s \,\right] \cdot \exp\left(-\frac{3\alpha}{n}\right) \leqslant \left(\Lambda^s \cdot \mathbf{1}_{\mathcal{E}_{t_0}^{s-1}} \cdot \exp\left(\frac{3\alpha}{n}\right)\right) \cdot \exp\left(-\frac{3\alpha}{n}\right) = \Lambda^s \cdot \mathbf{1}_{\mathcal{E}_{t_0}^{s-1}}.$$

Since Eq. (5.12) holds in either case, we deduce that $(\widetilde{\Lambda}_{t_0}^s)_{s \geqslant t_0}$ forms a super-martingale. □

5.3.6 Recovery and stabilisation

We are now ready to prove the *recovery*, i.e., that Λ becomes *small* at least once every $\mathcal{O}(ng(\log(ng))^2)$ steps.

Lemma 5.23 (Recovery). *Consider the* g-Adv-Comp *setting for any* $g \geqslant 1$ *and the potential* $\Lambda := \Lambda(\alpha, c_4 g)$ *with* $\alpha = \frac{1}{18}$, *and* $c_4 > 0$ *as defined in Lemma 5.21. Further, let the constants* $c, \epsilon > 0$ *be as defined in*

Lemma 5.21, $c_r := c_r(\alpha, c_4) \geqslant 1$ as in Lemma 5.19 (ii), $r \in (0,1)$ as in Lemma 5.18 and $c_3 \geqslant 2$ as in Theorem 3.21. Then, for any step $t_0 \geqslant 0$, (i) for $\Delta_r := \Delta_r(g) := \frac{60c_3^2 c_r}{\alpha\epsilon r} \cdot ng \cdot (\log(ng))^2$, it holds that

$$\mathbf{Pr}\left[\bigcup_{t \in [t_0, t_0+\Delta_r]} \{\Lambda^t \leqslant cn\} \;\middle|\; \mathfrak{F}^{t_0}, \max_{i \in [n]} \left|y_i^{t_0}\right| \leqslant c_3 g \log(ng) \right] \geqslant 1 - 3 \cdot (ng)^{-12}.$$

Further, (ii) for any step $t_0 \geqslant 0$, it holds that,

$$\mathbf{Pr}\left[\bigcup_{t \in [t_0, t_0+\Delta_r]} \{\Lambda^t \leqslant cn\} \right] \geqslant 1 - (ng)^{-11}.$$

Proof. First statement. Consider an arbitrary step t_0 with $\max_{i \in [n]} |y_i^{t_0}| \leqslant c_3 g \log(ng)$. Our aim is to show that w.h.p. $\Lambda^t \leqslant cn$ for some step $t \in [t_0, t_0+\Delta_r]$. We will do this by first showing that w.h.p. there is a significant number of good steps, i.e., $G_{t_0}^{t_1} \geqslant r \cdot \Delta_r$, and when this happens w.h.p. $\widetilde{\Lambda}_{t_0}^{t_1} = 0$, which implies the conclusion.

We start by upper bounding Λ^{t_0} as follows,

$$\Lambda^{t_0} \leqslant 2n \cdot e^{\alpha c_3 g \log(ng)} \leqslant e^{c_3 g \log(ng)} =: \lambda, \tag{5.13}$$

since $\alpha \leqslant \frac{1}{18}$ and $c_3 \geqslant 2$. Hence, by Lemma 5.19 (ii), there exists a constant $c_r := c_r(\alpha, c_4)$, such that

$$\Upsilon^{t_0} \leqslant c_r \cdot n \cdot (g^2 + (\log \Lambda^{t_0})^2) \leqslant 2c_r \cdot n \cdot (c_3 g \log(ng))^2 =: T.$$

Let $t_1 := t_0 + \Delta_r$. Applying Lemma 5.18 with $T = 2c_r \cdot n \cdot (c_3 g \log(ng))^2 = o(n^2 g^3)$ (and $\geqslant ng^2$) and $\hat{c} = \frac{\Delta_r \cdot g}{T} \geqslant \frac{30}{\alpha\epsilon r} \geqslant 1$ as $\alpha, \epsilon, r \leqslant 1$, we get

$$\begin{aligned}
&\mathbf{Pr}\left[G_{t_0}^{t_1-1} \geqslant r \cdot \Delta_r \;\middle|\; \mathfrak{F}^{t_0}, \Lambda^{t_0} \leqslant \lambda, \max_{i \in [n]} \left|y_i^{t_0}\right| \leqslant c_3 g \log(ng) \right] \\
&\quad \geqslant \mathbf{Pr}\left[G_{t_0}^{t_1-1} \geqslant r \cdot \Delta_r \;\middle|\; \mathfrak{F}^{t_0}, \Upsilon^{t_0} \leqslant T, \max_{i \in [n]} \left|y_i^{t_0}\right| \leqslant g(\log(ng))^2 \right] \\
&\quad \geqslant 1 - 2 \cdot (ng)^{-12}.
\end{aligned} \tag{5.14}$$

By Lemma 5.22, $(\widetilde{\Lambda}_{t_0}^t)_{t \geqslant t_0}$ is a super-martingale, so $\mathbf{E}[\, \widetilde{\Lambda}_{t_0}^{t_1} \mid \mathfrak{F}^{t_0} \,] \leqslant \widetilde{\Lambda}_{t_0}^{t_0} = \Lambda^{t_0}$. Hence, using Markov's inequality we get $\mathbf{Pr}\big[\, \widetilde{\Lambda}_{t_0}^{t_1} > \Lambda^{t_0} \cdot (ng)^{12} \,\big|\, \mathfrak{F}^{t_0}, \Lambda^{t_0} \leqslant \lambda \,\big] \leqslant (ng)^{-12}$. Thus, by the definition of $\widetilde{\Lambda}_{t_0}^{t_1}$ in Eq. (5.11), we have

$$\mathbf{Pr}\left[\Lambda^{t_1} \cdot \mathbf{1}_{\mathcal{E}_{t_0}^{t_1-1}} \leqslant \Lambda^{t_0} \cdot (ng)^{12} \cdot \exp\left(\frac{3\alpha}{n} \cdot B_{t_0}^{t_1-1} - \frac{\alpha\epsilon}{n} \cdot G_{t_0}^{t_1-1} \right) \;\middle|\; \mathfrak{F}^{t_0}, \Lambda^{t_0} \leqslant \lambda \right] \geqslant 1 - (ng)^{-12}. \tag{5.15}$$

Further, if in addition to the two events $\{\widetilde{\Lambda}_{t_0}^{t_1} \leqslant \Lambda^{t_0} \cdot (ng)^{12}\}$ and $\{\Lambda^{t_0} \leqslant \lambda\}$, also the event $\{G_{t_0}^{t_1-1} \geqslant r \cdot \Delta_r\}$ holds, then

$$\begin{aligned}
\Lambda^{t_1} \cdot \mathbf{1}_{\mathcal{E}_{t_0}^{t_1-1}} &\leqslant \Lambda^{t_0} \cdot (ng)^{12} \cdot \exp\left(\frac{3\alpha}{n} \cdot B_{t_0}^{t_1-1} - \frac{\alpha\epsilon}{n} \cdot G_{t_0}^{t_1-1} \right) \\
&\leqslant e^{c_3 g \log(ng)} \cdot (ng)^{12} \cdot \exp\left(\frac{3\alpha}{n} \cdot (1-r) \cdot \Delta_r - \frac{\alpha\epsilon}{n} \cdot r \cdot \Delta_r \right)
\end{aligned}$$

$$
\begin{aligned}
&\overset{(a)}{=} e^{c_3 g \log(ng)} \cdot (ng)^{12} \cdot \exp\left(\frac{\alpha\epsilon}{n} \cdot \frac{r}{2} \cdot \Delta_r - \frac{\alpha\epsilon}{n} \cdot r \cdot \Delta_r\right) \\
&= e^{c_3 g \log(ng)} \cdot (ng)^{12} \cdot \exp\left(-\frac{\alpha\epsilon}{n} \cdot \frac{r}{2} \cdot \Delta_r\right) \\
&= e^{c_3 g \log(ng)} \cdot (ng)^{12} \cdot \exp\left(-\frac{\alpha\epsilon}{n} \cdot \frac{r}{2} \cdot \frac{60 c_3^2 c_r}{\alpha\epsilon r} \cdot ng \cdot (\log(ng))^2\right) \\
&\overset{(b)}{\leqslant} e^{c_3 g \log(ng)} \cdot (ng)^{12} \cdot \exp(-30 c_3 g \log(ng)) \\
&\leqslant 1,
\end{aligned}
$$

where we used in (a) that $r = \frac{6}{6+\epsilon}$ implies $\frac{3\alpha}{n} \cdot (1-r) = \frac{3\alpha}{n} \cdot \frac{\epsilon}{6+\epsilon} = \frac{\alpha\epsilon}{n} \cdot \frac{r}{2}$, in (b) that $g \geqslant 1$, $c_r \geqslant 1$ and $c_3 \geqslant 1$. By the definition of Λ, we have that $\Lambda^{t_1} \geqslant n$ holds deterministically, and so we can deduce from the above inequality that $\mathbf{1}_{\mathcal{E}_{t_0}^{t_1-1}} = 0$, that is,

$$
\mathbf{Pr}\left[\neg\mathcal{E}_{t_0}^{t_1-1} \,\middle|\, \mathfrak{F}^{t_0},\ \ \tilde{\Lambda}_{t_0}^{t_1} \leqslant \Lambda^{t_0} \cdot (ng)^{12},\ \ \Lambda^{t_0} \leqslant \lambda,\ \ G_{t_0}^{t_1-1} \geqslant r \cdot \Delta_r\right] = 1.
$$

Recalling the definition of $\mathcal{E}_{t_0}^{t_1-1} := \bigcap_{t \in [t_0, t_1-1]} \{\Lambda^t > cn\}$ and taking the union bound over Eq. (5.14) and Eq. (5.15) yields

$$
\mathbf{Pr}\left[\bigcup_{t \in [t_0, t_0+\Delta_r]} \{\Lambda^t \leqslant cn\} \,\middle|\, \mathfrak{F}^{t_0}, \Lambda^{t_0} \leqslant \lambda\right] \geqslant 1 - 2 \cdot (ng)^{-12} - (ng)^{-12} = 1 - 3 \cdot (ng)^{-12}. \tag{5.16}
$$

Second statement. Using Theorem 3.21 (iii), for the constant $c_3 \geqslant 2$, it holds that,

$$
\mathbf{Pr}\left[\max_{i \in [n]} \left|y_i^{t_0}\right| \leqslant c_3 g \log(ng)\right] \geqslant 1 - (ng)^{-14}.
$$

Recall by Eq. (5.13), that this implies that

$$
\mathbf{Pr}\left[\Lambda^{t_0} \leqslant \lambda\right] \geqslant 1 - (ng)^{-14}.
$$

Hence, combining with Eq. (5.16), we conclude that

$$
\mathbf{Pr}\left[\bigcup_{t \in [t_0, t_0+\Delta_r]} \{\Lambda^t \leqslant cn\}\right] \geqslant \left(1 - 3 \cdot (ng)^{-12}\right) \cdot \left(1 - (ng)^{-14}\right) \geqslant 1 - (ng)^{-11}. \qquad \square
$$

The derivation of the lemma below is similar to that of Lemma 5.23, with the main difference being the tighter condition that $\Lambda^{t_0} \leqslant 2cn$, which allows us to choose a slightly shorter time window of $\Theta(n \cdot \max\{\log n, g\})$ steps (see Fig. 5.2).

Lemma 5.24 (Stabilisation). *Consider the g-Adv-Comp setting for any $g \geqslant 1$ and the potential $\Lambda := \Lambda(\alpha, c_4 g)$ with $\alpha = \frac{1}{18}$, and $c_4 > 0$ as defined in Lemma 5.21. Further, let the constants $c, \epsilon > 0$ be as defined in Lemma 5.21, $c_s := c_s(\alpha, c_4, 2c) \geqslant 1$ as in Lemma 5.19 (i) and $r \in (0,1)$ as in Lemma 5.18. Then, for $\Delta_s := \Delta_s(g) := \frac{60 c_s}{\alpha\epsilon r} \cdot n \cdot \max\{\log n, g\}$, we have that for any step $t_0 \geqslant 0$,*

$$
\mathbf{Pr}\left[\bigcup_{t \in [t_0, t_0+\Delta_s]} \{\Lambda^t \leqslant cn\} \,\middle|\, \mathfrak{F}^{t_0}, \Lambda^{t_0} \leqslant 2cn\right] \geqslant 1 - (ng)^{-11}.
$$

Proof. By Lemma 5.19 (i), $\Lambda^{t_0} \leqslant 2cn$ implies that deterministically $\Upsilon^{t_0} \leqslant c_s n g^2$ for constant $c_s := c_s(\alpha, c_4, 2c) \geqslant 1$ and

$$\max_{i\in[n]} \left| y_i^{t_0} \right| \leqslant c_4 g + \frac{1}{\alpha} \cdot \log(2cn) \leqslant g(\log(ng))^2,$$

for sufficiently large n using that $c_4, \alpha, c > 0$ are constants. Let $t_1 := t_0 + \Delta_s$. By Lemma 5.18 with $T := c_s n g \cdot \max\{\log n, g\} \geqslant c_s n g^2$ (and $T = o(n^2 g^3)$) and $\hat{c} := \frac{\Delta_s \cdot g}{T} = \frac{60}{\alpha\epsilon r} \geqslant 1$ as $\alpha, \epsilon, r \leqslant 1$, we have that

$$\begin{aligned}
&\mathbf{Pr}\left[\, G_{t_0}^{t_1-1} \geqslant r \cdot \Delta_s \,\middle|\, \mathfrak{F}^{t_0}, \Lambda^{t_0} \leqslant 2cn \,\right] \\
&\geqslant \mathbf{Pr}\left[\, G_{t_0}^{t_1-1} \geqslant r \cdot \Delta_s \,\middle|\, \mathfrak{F}^{t_0}, \Upsilon^{t_0} \leqslant T, \max_{i\in[n]} \left| y_i^{t_0} \right| \leqslant g(\log(ng))^2 \,\right] \\
&\geqslant 1 - 2 \cdot (ng)^{-12}.
\end{aligned} \tag{5.17}$$

By Lemma 5.22, $(\widetilde{\Lambda}_{t_0}^{t})_{t \geqslant t_0}$ is a super-martingale, so $\mathbf{E}[\, \widetilde{\Lambda}_{t_0}^{t_1} \mid \mathfrak{F}^{t_0}, \Lambda^{t_0} \leqslant 2cn \,] \leqslant \widetilde{\Lambda}_{t_0}^{t_0} = \Lambda^{t_0}$. Hence, using Markov's inequality we get $\mathbf{Pr}[\, \widetilde{\Lambda}_{t_0}^{t_1} > \Lambda^{t_0} \cdot (ng)^{12} \mid \mathfrak{F}^{t_0}, \Lambda^{t_0} \leqslant 2cn \,] \leqslant (ng)^{-12}$. Thus, by the definition of $\widetilde{\Lambda}_{t_0}^{t_1}$ in Eq. (5.11), we have

$$\mathbf{Pr}\left[\, \Lambda^{t_1} \cdot \mathbf{1}_{\mathcal{E}_{t_0}^{t_1-1}} \leqslant \Lambda^{t_0} \cdot (ng)^{12} \cdot \exp\left(\frac{3\alpha}{n} \cdot B_{t_0}^{t_1-1} - \frac{\alpha\epsilon}{n} \cdot G_{t_0}^{t_1-1} \right) \,\middle|\, \mathfrak{F}^{t_0}, \Lambda^{t_0} \leqslant 2cn \,\right] \geqslant 1 - (ng)^{-12}. \tag{5.18}$$

Further, if in addition to the two events $\{\widetilde{\Lambda}_{t_0}^{t_1} \leqslant \Lambda^{t_0} \cdot (ng)^{12}\}$ and $\{\Lambda^{t_0} \leqslant 2cn\}$, also the event $\{G_{t_0}^{t_1-1} \geqslant r \cdot \Delta_s\}$ holds, then

$$\begin{aligned}
\Lambda^{t_1} \cdot \mathbf{1}_{\mathcal{E}_{t_0}^{t_1-1}} &\leqslant \Lambda^{t_0} \cdot (ng)^{12} \cdot \exp\left(\frac{3\alpha}{n} \cdot B_{t_0}^{t_1-1} - \frac{\alpha\epsilon}{n} \cdot G_{t_0}^{t_1-1} \right) \\
&\leqslant 2cn \cdot (ng)^{12} \cdot \exp\left(\frac{3\alpha}{n} \cdot (1-r) \cdot \Delta_s - \frac{\alpha\epsilon}{n} \cdot r \cdot \Delta_s \right) \\
&\overset{(a)}{=} 2cn \cdot (ng)^{12} \cdot \exp\left(-\frac{\alpha\epsilon}{n} \cdot \frac{r}{2} \cdot \Delta_s \right) \\
&= 2cn \cdot (ng)^{12} \cdot \exp\left(-\frac{\alpha\epsilon}{n} \cdot \frac{r}{2} \cdot \frac{60 c_s}{\alpha\epsilon r} \cdot n \cdot \max\{\log n, g\} \right) \\
&\overset{(b)}{\leqslant} 2cn \cdot (ng)^{12} \cdot \exp\left(-30 \cdot \max\{\log n, g\} \right) \\
&\leqslant 2cn \cdot (ng)^{12} \cdot \exp\left(-15 \cdot \log(ng) \right) \\
&\leqslant 1,
\end{aligned}$$

where we used in (a) that $r = \frac{6}{6+\epsilon}$ implies $\frac{3\alpha}{n} \cdot (1-r) = \frac{3\alpha}{n} \cdot \frac{\epsilon}{6+\epsilon} = \frac{\alpha\epsilon}{n} \cdot \frac{r}{2}$ and in (b) that $c_s \geqslant 1$ and $\alpha \leqslant 1/2$. Also $\Lambda^{t_1} \geqslant 2n$ holds deterministically, so we can deduce from the above inequality that $\mathbf{1}_{\mathcal{E}_{t_0}^{t_1-1}} = 0$, that is,

$$\mathbf{Pr}\left[\, \neg\mathcal{E}_{t_0}^{t_1-1} \,\middle|\, \mathfrak{F}^{t_0},\ \ \widetilde{\Lambda}_{t_0}^{t_1} \leqslant \Lambda^{t_0} \cdot (ng)^{12},\ \ \Lambda^{t_0} \leqslant 2cn,\ \ G_{t_0}^{t_1-1} \geqslant r \cdot \Delta_s \,\right] = 1.$$

Recalling the definition of $\mathcal{E}_{t_0}^{t_1-1} := \bigcap_{t\in[t_0, t_1-1]} \{\Lambda^t > cn\}$, and taking the union bound over Eq. (5.17) and Eq. (5.18) yields

$$\mathbf{Pr}\left[\, \bigcup_{t\in[t_0, t_0+\Delta_s]} \{\Lambda^t \leqslant cn\} \,\middle|\, \mathfrak{F}^{t_0}, \Lambda^{t_0} \leqslant 2cn \,\right] \geqslant 1 - 2 \cdot (ng)^{-12} - (ng)^{-12} \geqslant 1 - (ng)^{-11}. \qquad \square$$

5.3.7 Completing the proof of Theorem 5.26

We will now prove that starting with $\Lambda^{t_0} \leqslant 2cn$ implies that for any step $t_1 \in [t_0, t_0 + (ng)^2]$, we have w.h.p. $\mathrm{Gap}(t_1) = \mathcal{O}(g + \log n)$.

Lemma 5.25. *Consider the* g-ADV-COMP *setting for any* $g \geqslant 1$*, the potential* $\Lambda := \Lambda(\alpha, c_4 g)$ *with* $\alpha = \frac{1}{18}$*,* $c_4 > 0$ *as defined in Lemma 5.21 and* $\Delta_s > 0$ *as defined in Lemma 5.24. Then, there exists a constant* $\kappa \geqslant \frac{1}{\alpha}$ *such that for any steps* $t_0 \geqslant 0$ *and* $t_1 \in (t_0, t_0 + (ng)^2]$*,*

$$\mathbf{Pr}\left[\max_{i\in[n]} \left| y_i^{t_1} \right| \leqslant \kappa \cdot (g + \log n) \;\middle|\; \mathfrak{F}^{t_0}, \Lambda^{t_0} \leqslant cn \right] \geqslant 1 - (ng)^{-9}.$$

Proof. Consider any step t_0 with $\Lambda^{t_0} \leqslant cn$. We define the event for any step $t_1 \geqslant t_0$

$$\mathcal{M}_{t_0}^{t_1} := \{\text{for all } t \in [t_0, t_1] \text{ there exists } s \in [t, t + \Delta_s] \text{ such that } \Lambda^s \leqslant cn\},$$

that is, if $\mathcal{M}_{t_0}^{t_1}$ holds then we have $\Lambda^s \leqslant cn$ at least once every $\Delta_s := \frac{60c_s}{\alpha\epsilon r} \cdot n \cdot \max\{\log n, g\}$ steps.

Assume now that $\mathcal{M}_{t_0}^{t_1}$ holds. We will show that

$$\max_{i\in[n]} \left| y_i^{t_1} \right| \leqslant \kappa \cdot (g + \log n).$$

Choosing $t = t_1$, implies that there exists $s \in [t_1, t_1 + \Delta_s]$ such that $\Lambda^s \leqslant cn$, which in turn implies by definition of Λ that $\max_{i\in[n]} |y_i^s| \leqslant \frac{1}{\alpha} \cdot \log(cn) + c_4 g < \frac{2}{\alpha} \cdot \log n + c_4 g$. Clearly, any y_i^t can decrease by at most $1/n$ in each step, and from this it follows that if $\mathcal{M}_{t_0}^{t_1}$ holds, then

$$\max_{i\in[n]} y_i^{t_1} \leqslant \max_{i\in[n]} \left| y_i^s \right| + \frac{\Delta_s}{n} \leqslant \kappa \cdot (g + \log n),$$

for the constant

$$\kappa := \frac{2}{\alpha} + c_4 + \frac{\Delta_s}{n \cdot \max\{\log n, g\}} = \frac{2}{\alpha} + c_4 + \frac{60c_s}{\alpha\epsilon r} > 0. \tag{5.19}$$

If $t_1 \geqslant t_0 + \Delta_s$ and $\mathcal{M}_{t_0}^{t_1}$ holds, then choosing $t = t_1 - \Delta_s$, there exists $s \in [t_1 - \Delta_s, t_1]$ such that $\Lambda^s \leqslant cn$. (In case $t_1 < t_0 + \Delta_s$, then we arrive at the same conclusion by choosing $s = t_0$ and using the precondition $\Lambda^{t_0} \leqslant cn$). This in turn implies $\max_{i\in[n]} |y_i^s| \leqslant \frac{2}{\alpha} \cdot \log n + c_4 g$. Hence

$$\min_{i\in[n]} y_i^{t_1} \geqslant -\max_{i\in[n]} \left| y_i^s \right| - \frac{\Delta_s}{n} \geqslant -\kappa \cdot (g + \log n).$$

Hence, $\mathcal{M}_{t_0}^{t_1}$ together with the precondition on $\Lambda^{t_0} \leqslant cn$ implies that $\max_{i\in[n]} \left| y_i^{t_1} \right| \leqslant \kappa \cdot (g + \log n)$. It remains to bound $\mathbf{Pr}\left[\neg\mathcal{M}_{t_0}^{t_1} \;\middle|\; \mathfrak{F}^{t_0}, \Lambda^{t_0} \leqslant cn \right]$.

Note that if for some step j_1 we have that $\Lambda^{j_1} \leqslant cn$ and for some $j_2 \geqslant j_1$ that $\Lambda^{j_2} > 2cn$, then there must exist $j \in (j_1, j_2)$ such that $\Lambda^j \in (cn, 2cn]$, since for every $t \geqslant 0$ it holds that $\Lambda^{t+1} \leqslant \Lambda^t \cdot e^{\alpha} \leqslant 2\Lambda^t$, as $\alpha \leqslant 1/2$. Let $t_0 < \tau_1 < \tau_2 < \cdots$ and $t_0 =: s_0 < s_1 < \cdots$ be two interlaced sequences defined recursively for $i \geqslant 1$ by

$$\tau_i := \inf\{\tau > s_{i-1} : \Lambda^{\tau} \in (cn, 2cn]\} \qquad \text{and} \qquad s_i := \inf\{s > \tau_i : \Lambda^s \leqslant cn\}.$$

Thus we have

$$t_0 = s_0 < \tau_1 < s_1 < \tau_2 < s_2 < \cdots,$$

and since $\tau_i > \tau_{i-1}$ we have $\tau_{t_1-t_0} \geqslant t_1$. Therefore, if the event $\cap_{i=1}^{t_1-t_0}\{s_i - \tau_i \leqslant \Delta_s\}$ holds, then also $\mathcal{M}_{t_0}^{t_1}$ holds.

Recall that by Lemma 5.24 we have for any $i = 1, 2, \ldots, t_1 - t_0$ and any $\tau = t_0 + 1, \ldots, t_1$

$$\mathbf{Pr}\left[\bigcup_{t \in [\tau_i, \tau_i + \Delta_s]} \{\Lambda^t \leqslant cn\} \;\middle|\; \mathfrak{F}^\tau,\ \Lambda^\tau \in (cn, 2cn], \tau_i = \tau \right] \geqslant 1 - (ng)^{-11},$$

and by negating and the definition of s_i,

$$\mathbf{Pr}\left[s_i - \tau_i > \Delta_s \;\middle|\; \mathfrak{F}^\tau, \Lambda^\tau \in (cn, 2cn], \tau_i = \tau \right] \leqslant (ng)^{-11}.$$

Since the above bound holds for any $i \geqslant 1$ and $\mathfrak{F}^\tau$, with $\tau_i = \tau$, it follows by the union bound over all $i = 1, 2, \ldots, t_1 - t_0$, as $t_1 - t_0 \leqslant (ng)^2$,

$$\mathbf{Pr}\left[\neg\mathcal{M}_{t_0}^{t_1} \;\middle|\; \mathfrak{F}^{t_0}, \Lambda^{t_0} \leqslant cn \right] \leqslant (t_1 - t_0) \cdot (ng)^{-11} \leqslant (ng)^{-9}. \qquad \square$$

Finally, we deduce that for the g-Adv-Comp setting, for an arbitrary step m w.h.p. $\mathrm{Gap}(m) = \mathcal{O}(g + \log n)$.

Theorem 5.26. *Consider the g-Adv-Comp setting for any $g \geqslant 1$, the constant $\kappa \geqslant \frac{1}{\alpha}$ defined in Eq.* (5.19) *in Lemma 5.25 and $\alpha = \frac{1}{18}$. Then, for any step $m \geqslant 0$,*

$$\mathbf{Pr}\left[\max_{i \in [n]} \left| y_i^m \right| \leqslant \kappa \cdot (g + \log n) \right] \geqslant 1 - 2 \cdot (ng)^{-9}.$$

Proof. Consider an arbitrary step $m \geqslant 0$ and recall that $\Delta_r := \frac{60 c_3^2 c_r}{\alpha \epsilon r} \cdot ng \cdot (\log(ng))^2$. If $m < \Delta_r$, then the claim follows by Lemma 5.25 as $\Lambda^0 = 2n \leqslant cn$ and $\Delta_r < (ng)^2$.

Otherwise, let $t_0 := m - \Delta_r$. Firstly, by the recovery lemma (Lemma 5.23 (ii)), we get

$$\mathbf{Pr}\left[\bigcup_{t \in [t_0, t_0 + \Delta_r]} \{\Lambda^t \leqslant cn\} \right] \geqslant 1 - (ng)^{-11}. \tag{5.20}$$

Hence for $\tau := \inf\{s \geqslant t_0 \colon \Lambda^s \leqslant cn\}$ we have $\mathbf{Pr}[\, \tau \leqslant m \,] \geqslant 1 - (ng)^{-11}$, as $t_0 + \Delta_r = m$.

Secondly, using Lemma 5.25, there exists a constant $\kappa := \kappa(\alpha, \epsilon) > 0$ such that for any step $s \in [t_0, m]$,

$$\mathbf{Pr}\left[\max_{i \in [n]} \left| y_i^m \right| \leqslant \kappa \cdot (g + \log n) \;\middle|\; \mathfrak{F}^s, \Lambda^s \leqslant cn \right] \geqslant 1 - (ng)^{-9}. \tag{5.21}$$

Combining the two inequalities from above, we conclude the proof

$$\begin{aligned}
\mathbf{Pr}\left[\max_{i \in [n]} \left| y_i^m \right| \leqslant \kappa \cdot (g + \log n) \right] &\geqslant \sum_{s=t_0}^{m} \mathbf{Pr}\left[\max_{i \in [n]} \left| y_i^m \right| \leqslant \kappa \cdot (g + \log n) \;\middle|\; \tau = s \right] \cdot \mathbf{Pr}[\, \tau = s \,] \\
&\geqslant \sum_{s=t_0}^{m} \mathbf{Pr}\left[\max_{i \in [n]} \left| y_i^m \right| \leqslant \kappa \cdot (g + \log n) \;\middle|\; \mathfrak{F}^s, \Lambda^s \leqslant cn \right] \cdot \mathbf{Pr}[\, \tau = s \,] \\
&\overset{(5.21)}{\geqslant} \left(1 - (ng)^{-9}\right) \cdot \mathbf{Pr}[\, \tau \leqslant m \,] \\
&\overset{(5.20)}{\geqslant} \left(1 - (ng)^{-9}\right) \cdot \left(1 - (ng)^{-11}\right) \geqslant 1 - 2 \cdot (ng)^{-9}. \qquad \square
\end{aligned}$$

5.4 g-Adv setting: Strong stabilisation

In this section, we will obtain for the g-Adv-Comp setting a stronger guarantee for a variant of the Λ potential used in Section 5.3. The precise upper bound that we need on g is $g \leqslant c_6 \log n$, where $c_6 > 0$ is a sufficiently small constant defined as

$$c_6 := \frac{r}{9 \cdot 20 \cdot \widetilde{c}_s \cdot \log(2ce^{2\alpha_1})} \leqslant 1, \tag{5.22}$$

where $r \in (0,1)$ is as defined in Lemma 5.18, $\widetilde{c}_s := \widetilde{c}_s(\alpha_1, c_4, e^{2\alpha_1}c) \geqslant 1$ as defined in Lemma 5.19 (for $c_4 := 730$), $c > 0$ as defined in Lemma 5.21 and $\alpha_1 := \frac{1}{6\kappa} \leqslant \frac{\alpha}{6}$, for $\kappa > 0$ the constant defined in Eq. (5.19) and $\alpha := \frac{1}{18}$ used in Section 5.3.

In Section 5.3, we showed that w.h.p. $\Lambda^t = \mathcal{O}(n)$ at least *once* every $\mathcal{O}(n \cdot (g + \log n))$ steps. Here, we will show that w.h.p. for any step t, we have for *all* steps $s \in [t, t + n\log^5 n]$ that $\Psi_0^s = \mathcal{O}(n)$. This will serve as the base case for the layered induction in Theorem 7.25.

We start by defining the potential function $V := V(\alpha_1, c_4 g)$ which is a variant of the $\Lambda := \Lambda(\alpha, c_4 g)$ potential function (defined in Eq. (5.3)), with the same offset $c_4 g = 2Dg = 730g$, but with a smaller smoothing parameter $\alpha_1 \leqslant \frac{\alpha}{6}$,

$$V^t := V^t(\alpha_1, c_4 g) := \sum_{i=1}^n V_i^t := \sum_{i=1}^n \Big[e^{\alpha_1 (y_i^t - c_4 g)^+} + e^{\alpha_1 (-y_i^t - c_4 g)^+} \Big]. \tag{5.23}$$

In Section 5.3, we proved that w.h.p. every $\mathcal{O}(n \cdot (g + \log n))$ steps the potential Λ satisfies $\Lambda^s \leqslant cn$. In this section, we will strengthen this to show that every $\mathcal{O}(ng)$ steps (for $g = \mathcal{O}(\log n)$) the potential V satisfies $V^s \leqslant e^{\mathcal{O}(\alpha_1 g)} \cdot n$. We will derive Lemma 5.29, which implies the base case of the layered induction in Section 7.4.1, i.e., that for *all* steps $s \in [m - n\log^5 n, m]$, $\Psi_0^s \leqslant Cn$ for $C := 2e^{2\alpha_1} \cdot c + 1$ and recalling that $\Psi_0 := \Psi_0(\alpha_1, c_5 g)$ for some sufficiently large constant $c_5 > 0$ (to be defined in Eq. (5.36) in Lemma 5.33) is given by

$$\Psi_0^t = \sum_{i=1}^n \exp\Big(\alpha_1 \cdot (y_i^s - z_0)^+\Big) = \sum_{i=1}^n \exp\Big(\alpha_1 \cdot (y_i^s - c_5 g)^+\Big).$$

We also define

$$\Phi_0^s := \Phi_0^s(\alpha_2, z_0) := \sum_{i=1}^n \Phi_{0,i}^s := \sum_{i=1}^n \exp\Big(\alpha_2 \cdot (y_i^s - z_0)^+\Big), \tag{5.24}$$

where $\alpha_2 := \frac{\alpha_1}{84}$.

The proof follows along the lines of Theorem 5.26 in Section 5.3, but it further conditions on the gap being $\mathcal{O}(g + \log n)$ at every step of the analysis. In particular, by conditioning on $\max_{i \in [n]} |y_i^t| \leqslant \kappa \cdot (g + \log n)$, we obtain that $|\Delta V^{t+1}| = \mathcal{O}(n^{1/3})$ (Lemma 5.30), which allows us to apply Azuma's inequality (Lemma B.10) to deduce that w.h.p. V remains small. This bounded difference condition is similar to the one used in Chapter 4.

5.4.1 A modified process

Let $\mathcal{P}$ be the process in the g-Adv-Comp setting (with arbitrary $1 \leqslant g \leqslant c_6 \log n$) that we want to analyse. We would like to condition on the event that $\mathcal{P}$ satisfies $\max_{i \in [n]} |y_i^t| \leqslant \kappa \cdot (g + \log n)$, for every step t in an interval of $2n\log^5 n$ steps, which holds w.h.p., as implied by Theorem 5.26.

We implement this conditioning by defining a *modified process* $\mathcal{Q}_{g,r_0} := \mathcal{Q}_{g,r_0}(\mathcal{P})$ for the same g and some arbitrary step r_0. Consider the stopping time $\sigma := \inf\{s \geqslant r_0 : \max_{i \in [n]} |y_i^s| > \kappa \cdot (g + \log n)\}$, then the process $\mathcal{Q}_{g,r_0}$ is defined so that

- in steps $s \in [0, \sigma)$ makes the same allocations as $\mathcal{P}$, and
- in steps $s \in [\sigma, \infty)$ allocates to the currently least loaded bin, i.e., it uses the probability allocation vector $r^s = (0, 0, \ldots, 0, 1)$.

Let $y_{\mathcal{Q}}^s$ be the normalised load vector of $\mathcal{Q}_{g,r_0}$ at step $s \geqslant 0$. By Theorem 5.26, it follows that for any interval $[r_0, m]$ with $m - r_0 \leqslant n^2$, with high probability the two processes agree, i.e.,

$$\begin{aligned} \mathbf{Pr}\Bigg[\bigcap_{s \in [r_0, m]} \{ y_{\mathcal{Q}}^s = y^s \} \Bigg] &\geqslant \mathbf{Pr}\Bigg[\bigcap_{s \in [r_0, m]} \Big\{ \max_{i \in [n]} \big| y_i^s \big| \leqslant \kappa \cdot (g + \log n) \Big\} \Bigg] \\ &\geqslant 1 - 2 \cdot (ng)^{-9} \cdot n^2 \geqslant 1 - 2n^{-7}. \end{aligned} \tag{5.25}$$

The process $\mathcal{Q}_{g,r_0}$ is defined in a way to satisfy the following property:

- (**Property 1**) The $\mathcal{Q}_{g,r_0}$ process satisfies the drop inequalities for the potential functions $\Lambda_{\mathcal{Q}}$, $V_{\mathcal{Q}}$ and $\Upsilon_{\mathcal{Q}}$ (Lemmas 5.17, 5.20 (ii) and 5.21) for any step $s \geqslant 0$. This holds because for any step $s < \sigma$, the process follows $\mathcal{P}$ and so it is an instance of the g-Adv-Comp setting. For any step $s \geqslant \sigma$, the process allocates to the currently least loaded bin and therefore minimises the potential $\Lambda_{\mathcal{Q}}^{s+1}$ given any $\mathfrak{F}^s$, which means that $\Lambda_{\mathcal{Q}}^{s+1} \leqslant \mathbf{E}[\, \Lambda^{s+1} \mid \mathfrak{F}^s \,]$ and so it trivially satisfies any drop inequality of the original process (and similarly for $V_{\mathcal{Q}}$ and $\Upsilon_{\mathcal{Q}}$).

Further, we define the event that the maximum normalised load in absolute value is small at step r_0 as,

$$\mathcal{Z}^{r_0} := \Big\{ \max_{i \in [n]} \big| y_{\mathcal{Q},i}^{r_0} \big| \leqslant \min\{\kappa \cdot (g + \log n), c_3 g \log(ng)\} \Big\}, \tag{5.26}$$

where $c_3 \geqslant 2$ is the constant defined in Eq. (3.32). We are primarily interested in the $\kappa \cdot (g + \log n)$ bound on the gap and the second bound is only needed for very small values of $g = \mathcal{O}(1)$. When the event $\mathcal{Z}^{r_0}$ holds, then the process $\mathcal{Q}_{g,r_0}$ also satisfies the following property (which “implements” the conditioning that the gap is $\mathcal{O}(g + \log n)$):

- (**Property 2**) For any step $s \geqslant r_0$, it follows that

  $$\max_{i \in [n]} \big| y_i^s \big| \leqslant \kappa \cdot (g + \log n) + 1 \leqslant 2\kappa \log n,$$

  using that $g \leqslant c_6 \log n$ with $c_6 \leqslant \frac{1}{4}$ by Eq. (5.22). At any step $s \in [r_0, \sigma)$, this holds by the definition of σ. For any step $s \geqslant \sigma$, a ball will never be allocated to a bin with $y_i^s > 0$ and in every n steps the at most n bins with load equal to the minimum load (at step s) will receive at least one ball each. Hence, over any n steps the maximum absolute normalised load does not increase and in the steps in between this can be larger by at most 1.

5.4.2 Preliminaries

We now define the adjusted potential $\widetilde{V}$, analogously to $\widetilde{\Lambda}$ in Eq. (5.11). Note that Lemma 5.21 with constants $\epsilon = \frac{1}{12}, c = 12 \cdot 18$ also applies to the potential V, since V has the same form as Λ but a smaller smoothing parameter $\alpha_1 \leqslant \alpha$. Next, we define the sequence $(\widetilde{V}_{t_0}^s)_{s \geqslant t_0} := (\widetilde{V}_{t_0}^s)_{s \geqslant t_0}(\alpha_1, c_4 g, \epsilon)$ as $\widetilde{V}_{t_0}^{t_0} := V^{t_0}(\alpha_1, c_4 g)$ and, for any $s > t_0$,

$$\widetilde{V}_{t_0}^s := V^s(\alpha_1, c_4 g) \cdot \mathbf{1}_{\widetilde{\mathcal{E}}_{t_0}^{s-1}} \cdot \exp\Big(- \frac{3\alpha_1}{n} \cdot B_{t_0}^{s-1} \Big) \cdot \exp\Big(+ \frac{\alpha_1 \epsilon}{n} \cdot G_{t_0}^{s-1} \Big), \tag{5.27}$$

where G_t^{s-1} (and B_t^{s-1}) is the number of good (bad) steps in $[t_0, s-1]$ (as defined in Section 5.3.5) and

$$\widetilde{\mathcal{E}}_{t_0}^s := \widetilde{\mathcal{E}}_{t_0}^s(V, c) := \bigcap_{t \in [t_0, s]} \left\{ V^t > cn \right\}.$$

Similarly, to $\widetilde{\Lambda}$ in Section 5.3, we have that $\widetilde{V}$ is a super-martingale.

Lemma 5.27 (cf. Lemma 5.22). *Consider the $\mathcal{Q}_{g,r_0}$ process for any $g \geqslant 1$, any step $r_0 \geqslant 0$, the sequence $(\widetilde{V}_{t_0}^s)_{s \geqslant t_0} := (\widetilde{V}_{t_0}^s)_{s \geqslant t_0}(\alpha_1, c_4 g, \epsilon)$ for any $t_0 \geqslant r_0$ with $\alpha_1 > 0$ as defined in Eq. (7.4) and $\epsilon, c_4 > 0$ as defined in Lemma 5.21. For any step $s \geqslant t_0$, we have that,*

$$\mathbf{E}[\, \widetilde{V}_{t_0}^{s+1} \mid \mathfrak{F}^s \,] \leqslant \widetilde{V}_{t_0}^s.$$

Proof. The proof is analogous to that of Lemma 5.22, by substituting Λ with V and $\widetilde{\Lambda}$ with $\widetilde{V}$. The drop inequalities follow from Lemma 5.20 and Lemma 5.21, since V has the same form as Λ and a smaller smoothing parameter $\alpha_1 \leqslant \alpha$. The process $\mathcal{Q}_{g,r_0}$ also satisfies the drop inequalities by Property 1 (see Section 5.4.1). □

The next lemma is a simple smoothness argument for the potential V defined in Eq. (5.23).

Lemma 5.28. *Consider the potential $V := V(\alpha_1, c_4 g)$ for any $\alpha_1 > 0$, any $c_4 > 0$ and any $g \geqslant 1$. Then, (i) for any step $t \geqslant 0$, we have that*

$$e^{-\alpha_1} \cdot V^t \leqslant V^{t+1} \leqslant e^{\alpha_1} \cdot V^t.$$

Further, (ii) for any $\hat{c} > 0$, for any integer $T > 0$ and any step $t \geqslant 0$, for which there exist steps $s_0 \in [t-T, t]$ and $s_1 \in [t, t+T]$, such that $V^{s_0} \leqslant \hat{c}n$ and $V^{s_1} \leqslant \hat{c}n$, we have that

$$V^t \leqslant e^{\alpha_1 \frac{T}{n}} \cdot 2\hat{c}n.$$

Proof. First statement. In each step the normalised load of any bin can change by at most 1, i.e., $|y_i^{t+1} - y_i^t| \leqslant 1$ and so $e^{-\alpha_1} \cdot V_i^t \leqslant V_i^{t+1} \leqslant e^{\alpha_1} \cdot V_i^t$. By aggregating over all bins, we get the claim.

Second statement. For any bin $i \in [n]$, in T steps the normalised load can decrease by at most T/n, i.e., $y_i^{s_1} \geqslant y_i^t - \frac{T}{n}$. So, the overload term is bounded by

$$e^{\alpha_1 (y_i^t - c_4 g)^+} \leqslant e^{\alpha_1 \frac{T}{n}} \cdot e^{\alpha_1 (y_i^{s_1} - c_4 g)^+} \leqslant e^{\alpha_1 \frac{T}{n}} \cdot V_i^{s_1}.$$

Similarly, $y_i^t \geqslant y_i^{s_0} - \frac{T}{n}$, and so the underload term is bounded by

$$e^{\alpha_1 (-y_i^t - c_4 g)^+} \leqslant e^{\alpha_1 \frac{T}{n}} \cdot e^{\alpha_1 (-y_i^{s_0} - c_4 g)^+} \leqslant e^{\alpha_1 \frac{T}{n}} \cdot V_i^{s_0}.$$

Hence, by aggregating over all bins and using the preconditions $V^{s_0} \leqslant \hat{c}n$ and $V^{s_1} \leqslant \hat{c}n$,

$$V^t = \sum_{i=1}^n \left[e^{\alpha_1 (y_i^t - c_4 g)^+} + e^{\alpha_1 (-y_i^t - c_4 g)^+} \right] \leqslant e^{\alpha_1 \frac{T}{n}} \cdot \sum_{i=1}^n \left(V_i^{s_1} + V_i^{s_0} \right) = e^{\alpha_1 \frac{T}{n}} \cdot (V^{s_1} + V^{s_0}) \leqslant e^{\alpha_1 \frac{T}{n}} \cdot 2\hat{c}n. \quad \square$$

The following lemma shows that by choosing a large enough offset $c_5 > 0$ in the potential $\Psi_0 := \Psi_0(\alpha_1, c_5 g)$ (defined in Eq. (7.10)), when $V^t = e^{\mathcal{O}(\alpha_1 g)} \cdot cn$, then $\Psi_0^t = \mathcal{O}(n)$.

Lemma 5.29. *Consider any $c, \hat{c} > 0$ and the potential $V := V(\alpha_1, c_4 g)$ for any $\alpha_1 > 0$, any $c_4 > 0$ and any $g \geqslant 1$. Further, consider the potential $\Psi_0 := \Psi_0(\alpha_1, c_5 g)$ with offset $c_5 := 2 \cdot \max\{c_4, \hat{c}\}$ and $C := 2e^{2\alpha_1} \cdot c + 1$. Then, for any step $t \geqslant 0$ with $V^t \leqslant e^{\alpha_1 \cdot \hat{c} \cdot g} \cdot 2e^{2\alpha_1} cn$, it holds that $\Psi_0^t \leqslant Cn$.*

Proof. We start by upper bounding Ψ_0^t,

$$\begin{aligned}
\Psi_0^t &= \sum_{i=1}^{n} e^{\alpha_1 (y_i^t - c_5 g)^+} \\
&= \sum_{i \in [n] \colon y_i^t \geqslant c_5 g} e^{\alpha_1 (y_i^t - c_5 g)} + \sum_{i \in [n] \colon y_i^t < c_5 g} e^{0} \\
&\leqslant e^{-\alpha_1 (c_5/2) g} \sum_{i \in [n] \colon y_i^t \geqslant c_5 g} e^{\alpha_1 (y_i^t - (c_5/2) g)} + n \\
&\overset{(a)}{\leqslant} e^{-\alpha_1 (c_5/2) g} \sum_{i \in [n] \colon y_i^t \geqslant c_4 g} e^{\alpha_1 (y_i^t - c_4 g)} + n \\
&\leqslant e^{-\alpha_1 (c_5/2) g} \sum_{i=1}^{n} e^{\alpha_1 (y_i^t - c_4 g)^+} + n \\
&= e^{-\alpha_1 (c_5/2) g} \cdot V^t + n,
\end{aligned}$$

where in (a) we used that $c_5/2 \geqslant c_4$. Now, using the precondition of the lemma that $V^t \leqslant e^{\alpha_1 \cdot \hat{c} \cdot g} \cdot 2e^{2\alpha_1} cn$ and $c_5 \geqslant 2 \cdot \hat{c}$, we conclude

$$\Psi_0^t \leqslant e^{-\alpha_1 (c_5/2) g} \cdot V^t + n \leqslant 2e^{2\alpha_1} \cdot cn + n = Cn. \qquad \square$$

Compared to Section 5.3, where we proved stabilisation over an interval of $\Omega(n \cdot \max\{\log n, g\})$ steps, here we will be using a shorter interval of length

$$\widetilde{\Delta}_s := \frac{20 \cdot \widetilde{c}_s \cdot \log(2ce^{2\alpha_1})}{\alpha_1 \epsilon r} \cdot ng, \tag{5.28}$$

where constants $\epsilon := \frac{1}{12}, r := \frac{6}{6+\epsilon} > 0$ are as defined in Section 5.3 and $\widetilde{c}_s := \widetilde{c}_s(\alpha_1, c_4, e^{2\alpha_1} c) \geqslant 1$ is defined in Lemma 5.19.

We now prove the bounded difference condition for the $\widetilde{V}$ potential. This follows from the second property of $\mathcal{Q}_{g,r_0}$ that the maximum normalised load in absolute value is $\mathcal{O}(g + \log n)$ for any step $s \geqslant r_0$.

Lemma 5.30. *Consider the $\mathcal{Q}_{g,r_0}$ process for any $g \in [1, c_6 \log n]$ with $c_6 > 0$ as defined in Eq. (5.22), any step $r_0 \geqslant 0$, and $\mathcal{Z}^{r_0}$ as defined in Eq. (5.26). Further, consider the sequence $(\widetilde{V}_{t_0}^s)_{s \geqslant t_0} := (\widetilde{V}_{t_0}^s)_{s \geqslant t_0}(\alpha_1, c_4 g, \epsilon)$ for any step $t_0 \geqslant r_0$ with $\alpha_1 > 0$ as defined in Eq. (7.4) and $\epsilon, c_4 > 0$ as defined in Lemma 5.21. Then, for any step $s \geqslant t_0 \geqslant r_0$ we have that $\widetilde{V}_{t_0}^{s+1} = 0$ or*

$$\left(\left| \Delta \widetilde{V}_{t_0}^{s+1} \right| \mid \mathcal{Z}^{r_0}, \mathfrak{F}^s \right) \leqslant 16 \cdot e^{\alpha_1 \epsilon \cdot \frac{s - t_0}{n}} \cdot n^{1/3}.$$

Proof. Consider an arbitrary step $s \geqslant r_0$ and assume that the event $\mathcal{Z}^{r_0}$ holds. By Property 2 of the $\mathcal{Q}_{g,r_0}$ process (see Section 5.4.1), we have that

$$\max_{i \in [n]} \left| y_i^s \right| \leqslant \kappa \cdot (g + \log n) + 1,$$

which also implies for $c_4 := 730 > 0$, since $g \leqslant c_6 \log n \leqslant \log n$ and $\kappa > 1$ that

$$\max_{i \in [n]} \left\{ (y_i^s - c_4 g)^+, (-y_i^s - c_4 g)^+ \right\} \leqslant \kappa \cdot (g + \log n) + 1 - c_4 g \leqslant 2\kappa \log n. \tag{5.29}$$

We will now show that $|\Delta V^{s+1}| \leqslant 5n^{1/3}$ and then use this to bound $|\Delta \widetilde{V}^{s+1}|$. By Eq. (5.29) for any bin $i \in [n]$,

$$V_i^s \leqslant 2 \cdot e^{2\alpha_1 \kappa \log n} = 2 \cdot e^{\frac{1}{3} \cdot \log n} = 2n^{1/3},$$

using that $\alpha_1 := \frac{1}{6\kappa}$. Hence, by aggregating over all bins, $V^s \leqslant 2n^{4/3}$. Furthermore, if the ball at step $s+1$ is allocated to bin $j \in [n]$, then

$$\Delta V^{s+1} \leqslant e^{\alpha_1/n} \cdot V^s + e^{\alpha_1} \cdot V_j^s - V^s \leqslant \frac{2\alpha_1}{n} V^s + 2 \cdot 2n^{1/3} \leqslant \frac{2\alpha_1}{n} \cdot (2n^{4/3}) + 4n^{1/3} \leqslant 5n^{1/3},$$

using that $e^{\alpha_1/n} \leqslant 1 + 2 \cdot \frac{\alpha_1}{n}$ and $e^{\alpha_1} \leqslant 2$, which both hold as $\alpha_1 \leqslant 1/4$. Similarly,

$$\Delta V^{s+1} \geqslant e^{-\alpha_1/n} \cdot V^s - e^{\alpha_1} \cdot V_j^s - V^s \geqslant -\frac{\alpha_1}{n} V^s - 2 \cdot 2n^{1/3} \geqslant -\frac{\alpha_1}{n} \cdot (2n^{4/3}) - 4n^{1/3} \geqslant -5n^{1/3},$$

using that $e^{-\alpha_1/n} \geqslant 1 - \frac{\alpha_1}{n}$ and $e^{\alpha_1} \leqslant 2$, as $\alpha_1 \leqslant 1/4$.

Now, we turn to upper bounding $|\Delta \widetilde{V}_{t_0}^{s+1}|$ by proving lower and upper bounds on $\widetilde{V}_{t_0}^{s+1}$. If $\widetilde{V}_{t_0}^{s+1} = 0$, then the conclusion follows. Otherwise, since $\widetilde{V}_{t_0}^{s+1} > 0$, we have that $\mathbf{1}_{\widetilde{\mathcal{E}}_{t_0}^s} = \mathbf{1}_{\widetilde{\mathcal{E}}_{t_0}^{s-1}} = 1$, so by definition of $\widetilde{V}$ in Eq. (5.27),

$$\widetilde{V}_{t_0}^s = V^s \cdot \exp\Big(-\frac{3\alpha_1}{n} \cdot B_{t_0}^{s-1}\Big) \cdot \exp\Big(+\frac{\alpha_1 \epsilon}{n} \cdot G_{t_0}^{s-1}\Big) \leqslant 2 \cdot e^{\alpha_1 \epsilon \cdot \frac{s-t_0}{n}} \cdot n^{4/3}, \tag{5.30}$$

using that $G_{t_0}^{s-1} \leqslant s - t_0$ and $V^s \leqslant 2n^{4/3}$.

Now, we upper bound $\widetilde{V}_{t_0}^{s+1}$, recalling that $\mathbf{1}_{\widetilde{\mathcal{E}}_{t_0}^s} = 1$,

$$\begin{aligned}
\widetilde{V}_{t_0}^{s+1} &= V^{s+1} \cdot \exp\Big(-\frac{3\alpha_1}{n} \cdot B_{t_0}^s\Big) \cdot \exp\Big(+\frac{\alpha_1 \epsilon}{n} \cdot G_{t_0}^s\Big) \\
&\leqslant (V^s + 5n^{1/3}) \cdot \exp\Big(-\frac{3\alpha_1}{n} \cdot B_{t_0}^s\Big) \cdot \exp\Big(+\frac{\alpha_1 \epsilon}{n} \cdot G_{t_0}^s\Big) \\
&\leqslant (V^s + 5n^{1/3}) \cdot \exp\Big(-\frac{3\alpha_1}{n} \cdot B_{t_0}^{s-1}\Big) \cdot \exp\Big(+\frac{\alpha_1 \epsilon}{n} \cdot G_{t_0}^{s-1}\Big) \cdot \exp\Big(\frac{\alpha_1 \epsilon}{n}\Big) \\
&= \widetilde{V}_{t_0}^s \cdot \exp\Big(\frac{\alpha_1 \epsilon}{n}\Big) + 5 \cdot n^{1/3} \cdot \exp\Big(-\frac{3\alpha_1}{n} \cdot B_{t_0}^{s-1}\Big) \cdot \exp\Big(+\frac{\alpha_1 \epsilon}{n} \cdot G_{t_0}^{s-1}\Big) \cdot \exp\Big(\frac{\alpha_1 \epsilon}{n}\Big) \\
&\overset{(a)}{\leqslant} \widetilde{V}_{t_0}^s \cdot \Big(1 + \frac{3\alpha_1}{n}\Big) + 5 \cdot n^{1/3} \cdot e^{\alpha_1 \epsilon \cdot \frac{s-t_0}{n}} \cdot 2 \\
&\overset{(b)}{\leqslant} \widetilde{V}_{t_0}^s + (2 \cdot e^{\alpha_1 \epsilon \cdot \frac{s-t_0}{n}} \cdot n^{4/3}) \cdot \frac{3\alpha_1}{n} + 10 \cdot n^{1/3} \cdot e^{\alpha_1 \epsilon \cdot \frac{s-t_0}{n}} \\
&\leqslant \widetilde{V}_{t_0}^s + 16 \cdot e^{\alpha_1 \epsilon \cdot \frac{s-t_0}{n}} \cdot n^{1/3},
\end{aligned} \tag{5.31}$$

using in (a) that $e^{\alpha_1 \epsilon / n} \leqslant 1 + \frac{3\alpha_1}{n}$ as $\alpha_1 \leqslant 1$ and $\epsilon = \frac{1}{12}$, $G_{t_0}^{s-1} \leqslant s - t_0$, and $e^{\alpha_1 \epsilon / n} \leqslant 2$ and in (b) using Eq. (5.30).

Similarly, we lower bound $\widetilde{V}_{t_0}^{s+1}$,

$$\begin{aligned}
\widetilde{V}_{t_0}^{s+1} &= V^{s+1} \cdot \exp\Big(-\frac{3\alpha_1}{n} \cdot B_{t_0}^s\Big) \cdot \exp\Big(+\frac{\alpha_1 \epsilon}{n} \cdot G_{t_0}^s\Big) \\
&\geqslant (V^s - 5n^{1/3}) \cdot \exp\Big(-\frac{3\alpha_1}{n} \cdot B_{t_0}^s\Big) \cdot \exp\Big(+\frac{\alpha_1 \epsilon}{n} \cdot G_{t_0}^s\Big)
\end{aligned}$$

$$
\begin{aligned}
&\overset{(a)}{\geqslant} (V^s - 5n^{1/3}) \cdot \exp\Big(-\frac{3\alpha_1}{n} \cdot B_{t_0}^{s-1}\Big) \cdot \exp\Big(+\frac{\alpha_1 \epsilon}{n} \cdot G_{t_0}^{s-1}\Big) \cdot \exp\Big(-\frac{3\alpha_1}{n}\Big) \\
&= \widetilde{V}_{t_0}^s \cdot \exp\Big(-\frac{3\alpha_1}{n}\Big) - 5 \cdot n^{1/3} \cdot \exp\Big(-\frac{3\alpha_1}{n} \cdot B_{t_0}^{s-1}\Big) \cdot \exp\Big(+\frac{\alpha_1 \epsilon}{n} \cdot G_{t_0}^{s-1}\Big) \cdot \exp\Big(-\frac{3\alpha_1}{n}\Big) \\
&\overset{(b)}{\geqslant} \widetilde{V}_{t_0}^s \cdot \Big(1 - \frac{3\alpha_1}{n}\Big) - 5 \cdot n^{1/3} \cdot e^{\alpha_1 \epsilon \cdot \frac{s-t_0}{n}} \\
&\overset{(c)}{\geqslant} \widetilde{V}_{t_0}^s - (2 \cdot e^{\alpha_1 \epsilon \cdot \frac{s-t_0}{n}} \cdot n^{4/3}) \cdot \frac{3\alpha_1}{n} - 5 \cdot n^{1/3} \cdot e^{\alpha_1 \epsilon \cdot \frac{s-t_0}{n}} \\
&\geqslant \widetilde{V}_{t_0}^s - 16 \cdot e^{\alpha_1 \epsilon \cdot \frac{s-t_0}{n}} \cdot n^{1/3},
\end{aligned}
\tag{5.32}
$$

using in (a) that $V^s \geqslant 2n \geqslant 5n^{1/3}$ holds deterministically, in (b) that $e^{-3\alpha_1/n} \geqslant 1 - \frac{3\alpha_1}{n}$ and $G_{t_0}^{s-1} \leqslant s - t_0$ and in (c) using Eq. (5.30).

Hence, combining the two upper bounds in Eq. (5.31) and Eq. (5.32), we conclude that $|\Delta \widetilde{V}_{t_0}^{s+1}| \leqslant 16 \cdot e^{\alpha_1 \epsilon \cdot \frac{s-t_0}{n}} \cdot n^{1/3}$. □

5.4.3 Strong stabilisation

We will now prove the following slightly stronger version of Lemma 5.24, meaning that stabilisation is over intervals of length $\Theta(ng)$ instead of $\Theta(n \cdot (g + \log n))$.

Lemma 5.31 (Strong stabilisation). *Consider the $\mathcal{Q}_{g,r_0}$ process for any $g \in [1, c_6 \log n]$ for $c_6 > 0$ as defined in Eq.* (5.22)*, any step $r_0 \geqslant 0$ and $\mathcal{Z}^{r_0}$ as defined in Eq.* (5.26)*. Then, for the potential $V := V(\alpha_1, c_4 g)$ with $\alpha_1 > 0$ as defined in Eq.* (7.4)*, $c_4, c > 0$ as defined in Lemma 5.21 and $\widetilde{\Delta}_s > 0$ as defined in Eq.* (5.28)*, it holds that for any step $t_0 \geqslant r_0$,*

$$
\mathbf{Pr}\left[\bigcup_{s \in [t_0, t_0 + \widetilde{\Delta}_s]} \{V^s \leqslant e^{\alpha_1} cn\} \;\middle|\; \mathcal{Z}^{r_0}, \mathfrak{F}^{t_0}, e^{\alpha_1} cn < V^{t_0} \leqslant e^{2\alpha_1} cn \right] \geqslant 1 - n^{-11}.
$$

Proof. The proof of this lemma proceeds similarly to that of Lemma 5.24, but we will apply Azuma's inequality for $\widetilde{V}_{t_0}^s$ instead of Markov's inequality. However, we cannot directly apply concentration to $\widetilde{V}_{t_0}^s$ because the bounded difference condition (Lemma 5.30) holds only when $\widetilde{V}_{t_0}^s$ is positive. So instead we apply it to a stopped random variable $X_{t_0}^s$ to be defined in a way that ensures it is always positive.

Let $t_1 := t_0 + \widetilde{\Delta}_s$. We define the stopping time $\tau := \inf\{s \geqslant t_0 \colon V^s \leqslant e^{\alpha_1} cn\}$ and for any $s \in [t_0, t_1]$,

$$
X_{t_0}^s := \widetilde{V}_{t_0}^{s \wedge \tau}.
$$

We will now verify that $X_{t_0}^s > 0$ for all $s \in [t_0, t_1]$. Firstly, consider any $s < \tau$. Since $V^s > e^{\alpha_1} cn$, by Lemma 5.28 (i), we have that $V^{s+1} \geqslant V^s \cdot e^{-\alpha_1} > cn$ and hence $X_{t_0}^{s+1} = \widetilde{V}_{t_0}^{s+1} > 0$. Secondly, for any $s \geqslant \tau$, it trivially holds that $X_{t_0}^{s+1} = X_{t_0}^s > 0$.

We proceed to verify the preconditions of Azuma's inequality for super-martingales (Lemma B.10) for the sequence $(X_{t_0}^s)_{s \in [t_0, t_1]}$. Firstly, using Lemma 5.27 it forms a super-martingale, i.e., that $\mathbf{E}[X_{t_0}^s \mid \mathfrak{F}^{s-1}] \leqslant X_{t_0}^{s-1}$. Secondly, by Lemma 5.30, since $X_{t_0}^s > 0$, for any filtration $\mathfrak{F}^{s-1}$ where $\mathcal{Z}^{r_0}$ holds, we have that

$$
\begin{aligned}
\Big(\big|X_{t_0}^s - X_{t_0}^{s-1}\big| \mid \mathcal{Z}^{r_0}, \mathfrak{F}^{s-1}\Big) &\leqslant \Big(\big|\widetilde{V}_{t_0}^s - \widetilde{V}_{t_0}^{s-1}\big| \mid \mathcal{Z}^{r_0}, \mathfrak{F}^{s-1}\Big) \\
&\leqslant 16 \cdot e^{\alpha_1 \epsilon \cdot \frac{\widetilde{\Delta}_s}{n}} \cdot n^{1/3}
\end{aligned}
$$

$$\overset{(a)}{=} (16 \cdot e^{20\cdot\widetilde{c}_s \cdot r^{-1}\cdot \log(2ce^{2\alpha_1})g}) \cdot n^{1/3}$$
$$\overset{(b)}{\leqslant} 16 \cdot e^{20\cdot\widetilde{c}_s \cdot r^{-1}\cdot \log(2ce^{2\alpha_1})\cdot c_6 \log n} \cdot n^{1/3}$$
$$= 16 \cdot n^{4/9},$$

using in (a) that $\widetilde{\Delta}_s := \frac{20\cdot\widetilde{c}_s\cdot\log(2ce^{2\alpha_1})}{\alpha_1 \epsilon r} \cdot ng$ and in (b) that $g \leqslant c_6 \log n$ and $c_6 := \frac{r}{9\cdot 20\cdot\widetilde{c}_s\cdot\log(2ce^{2\alpha_1})}$.

Hence, applying Lemma B.10 for $\lambda := n$, $N := \widetilde{\Delta}_s$ and $c_i := 16 \cdot n^{4/9}$,

$$\mathbf{Pr}\left[\, X_{t_0}^{t_1} \geqslant X_{t_0}^{t_0} + n \;\middle|\; \mathcal{Z}^{r_0}, \mathfrak{F}^{t_0}, e^{\alpha_1} cn < V^{t_0} \leqslant e^{2\alpha_1} cn \,\right] \leqslant \exp\left(-\frac{n^2}{2 \cdot \widetilde{\Delta}_s \cdot (16 \cdot n^{4/9})^2} \right) \leqslant n^{-\omega(1)}, \quad (5.33)$$

using that $\widetilde{\Delta}_s = \mathcal{O}(ng) = \mathcal{O}(n \log n)$.

As we condition on $\{V^{t_0} \leqslant e^{2\alpha_1} cn\}$, we have that

$$\max_{i \in [n]} \left| y_i^{t_0} \right| \leqslant c_4 g + \frac{\log(e^{2\alpha_1} cn)}{\alpha_1} \leqslant g(\log(ng))^2,$$

for sufficiently large n, using that $\alpha_1, c_5, c > 0$ are constants. Further, by Lemma 5.19 (i) (since V has the same form as Λ), we have that $\Upsilon^{t_0} \leqslant \widetilde{c}_s ng^2$, for some constant $\widetilde{c}_s := \widetilde{c}_s(\alpha_1, c_4, e^{2\alpha_1}c) \geqslant 1$.

Applying Lemma 5.18 for $T := \widetilde{c}_s ng^2$ (since $T \in [ng^2, o(n^2 g^3)]$) and $\hat{c} := \frac{\widetilde{\Delta}_s \cdot g}{T} \geqslant \frac{20}{\alpha_1 \epsilon r} \geqslant 1$ (since $\alpha_1, \epsilon, r \leqslant 1$),

$$\begin{aligned}
&\mathbf{Pr}\left[\, G_{t_0}^{t_1 - 1} \geqslant r \cdot \widetilde{\Delta}_s \;\middle|\; \mathcal{Z}^{r_0}, \mathfrak{F}^{t_0}, e^{\alpha_1} cn < V^{t_0} \leqslant e^{2\alpha_1} cn \,\right] \\
&\qquad \geqslant \mathbf{Pr}\left[\, G_{t_0}^{t_1 - 1} \geqslant r \cdot \widetilde{\Delta}_s \;\middle|\; \mathfrak{F}^{t_0}, \Upsilon^{t_0} \leqslant T, \max_{i \in [n]} \left| y_i^{t_0} \right| \leqslant g(\log(ng))^2 \,\right] \geqslant 1 - 2 \cdot n^{-12}. \qquad (5.34)
\end{aligned}$$

Taking the union bound over Eq. (5.33) and Eq. (5.34), we have

$$\mathbf{Pr}\left[\, \left\{ X_{t_0}^{t_1} < X_{t_0}^{t_0} + n \right\} \cap \left\{ G_{t_0}^{t_1 - 1} \geqslant r \cdot \widetilde{\Delta}_s \right\} \;\middle|\; \mathcal{Z}^{r_0}, \mathfrak{F}^{t_0}, e^{\alpha_1} cn < V^{t_0} \leqslant e^{2\alpha_1} cn \,\right] \geqslant 1 - n^{-11}.$$

Assume that $\{X_{t_0}^{t_1} < X_{t_0}^{t_0} + n\}$ and $\{G_{t_0}^{t_1 - 1} \geqslant r \cdot \widetilde{\Delta}_s\}$ hold. Then, we consider two cases based on whether the stopping time occurred before t_1.

Case 1 [$\tau \leqslant t_1$]: Here, clearly there is an $s \in [t_0, t_1]$, namely $s = \tau$, such that $V^s \leqslant e^{\alpha_1} cn$ and the conclusion follows.

Case 2 [$\tau > t_1$]: Here, using that $\{\tau > t_1\}$ and $\{X_{t_0}^{t_1} < X_{t_0}^{t_0} + n\}$ both hold, it follows that

$$\widetilde{V}_{t_0}^{t_1} < \widetilde{V}_{t_0}^{t_0} + n,$$

and so, by definition of $\widetilde{V}_{t_0}^{t_1}$ (Eq. (5.27)),

$$V^{t_1} \cdot \mathbf{1}_{\widetilde{\mathcal{E}}_{t_0}^{t_1 - 1}} \cdot \exp\left(-\frac{3\alpha_1}{n} \cdot B_{t_0}^{t_1 - 1} \right) \cdot \exp\left(+\frac{\alpha_1 \epsilon}{n} \cdot G_{t_0}^{t_1 - 1} \right) < V^{t_0} + n.$$

By re-arranging and using that $\{G_{t_0}^{t_1 - 1} \geqslant r \cdot \widetilde{\Delta}_s\}$ holds, we have that

$$\begin{aligned}
V^{t_1} \cdot \mathbf{1}_{\widetilde{\mathcal{E}}_{t_0}^{t_1 - 1}} &< (V^{t_0} + n) \cdot \exp\left(\frac{3\alpha_1}{n} \cdot B_{t_0}^{t_1 - 1} - \frac{\alpha_1 \epsilon}{n} \cdot G_{t_0}^{t_1 - 1} \right) \\
&\leqslant (e^{2\alpha_1} cn + n) \cdot \exp\left(\frac{3\alpha_1}{n} \cdot (1 - r) \cdot \widetilde{\Delta}_s - \frac{\alpha_1 \epsilon}{n} \cdot r \cdot \widetilde{\Delta}_s \right)
\end{aligned}$$

$$
\begin{aligned}
&\overset{(a)}{=} 2e^{2\alpha_1} \cdot cn \cdot \exp\Big(\frac{\alpha_1\epsilon}{n} \cdot \frac{r}{2} \cdot \widetilde{\Delta}_s - \frac{\alpha_1\epsilon}{n} \cdot r \cdot \widetilde{\Delta}_s\Big) \\
&= 2e^{2\alpha_1} \cdot cn \cdot \exp\Big(-\frac{\alpha_1\epsilon}{n} \cdot \frac{r}{2} \cdot \widetilde{\Delta}_s\Big) \\
&= 2e^{2\alpha_1} \cdot cn \cdot \exp\Big(-\frac{\alpha_1\epsilon}{n} \cdot \frac{r}{2} \cdot \frac{20 \cdot \widetilde{c}_s \cdot \log(2ce^{2\alpha_1})}{\alpha_1 \epsilon r} \cdot ng\Big) \\
&\overset{(b)}{\leqslant} 2e^{2\alpha_1} \cdot cn \cdot \exp\big(-10\log(2ce^{2\alpha_1})\big) \\
&\overset{(c)}{\leqslant} n/2,
\end{aligned}
$$

where we used in (a) that $r = \frac{6}{6+\epsilon}$ implies $\frac{3\alpha_1}{n} \cdot (1-r) = \frac{3\alpha_1}{n} \cdot \frac{\epsilon}{6+\epsilon} = \frac{\alpha_1\epsilon}{n} \cdot \frac{r}{2}$ and $c \geqslant 1$, in (b) that $g \geqslant 1$, $c \geqslant 1$ and $\widetilde{c}_s \geqslant 1$ and in (c) that $c = 18 \cdot 12$. Since deterministically we have that $V^{t_1} \geqslant n$, it must be that $\mathbf{1}_{\widetilde{\mathcal{E}}_{t_0}^{t_1-1}} = 0$, implying that there exists $s \in [t_0, t_1)$ such that $V^s \leqslant cn \leqslant e^{\alpha_1} cn$. □

We will now show that the potential V becomes small every $\Theta(ng)$ steps. The proof proceeds similarly to the proof of Lemma 5.25.

Lemma 5.32. *Consider the $\mathcal{Q}_{g,r_0}$ process for any $g \in [1, c_6 \log n]$ for $c_6 > 0$ as defined in Eq.* (5.22) *and any step $r_0 \geqslant 0$. Then, for the potential $V := V(\alpha_1, c_4 g)$ with α_1 as defined in Eq.* (7.4), *$c_4, c > 0$ as defined in Lemma 5.21 and $\widetilde{\Delta}_s > 0$ as defined in Eq.* (5.28), *it holds that for any step $t_0 \geqslant r_0$ and t_1 such that $t_0 < t_1 \leqslant t_0 + 2n \log^5 n$,*

$$
\mathbf{Pr}\left[\bigcap_{t \in [t_0, t_1]} \bigcup_{s \in [t, t+\widetilde{\Delta}_s]} \big\{V^s \leqslant e^{2\alpha_1} cn\big\} \;\middle|\; \mathcal{Z}^{r_0}, \mathfrak{F}^{t_0}, V^{t_0} \leqslant cn \right] \geqslant 1 - n^{-9},
$$

Proof. Analogously to the proof of Lemma 5.25, we begin by defining the event

$$
\widetilde{\mathcal{M}}_{t_0}^{t_1} = \big\{\text{for all } t \in [t_0, t_1] \text{ there exists } s \in [t, t + \widetilde{\Delta}_s] \text{ such that } V^s \leqslant e^{2\alpha_1} cn\big\},
$$

that is, if $\widetilde{\mathcal{M}}_{t_0}^{t_1}$ holds then we have that $V^s \leqslant e^{2\alpha_1} cn$ at least once every $\widetilde{\Delta}_s$ steps and so the claim follows.

Note that if for some step j_1 we have that $V^{j_1} \leqslant e^{\alpha_1} cn$ and for some $j_2 \geqslant j_1$ that $V^{j_2} > e^{2\alpha_1} cn$, then there must exist $j \in (j_1, j_2)$ such that $V^j \in (e^{\alpha_1} cn, e^{2\alpha_1} cn]$, since for every $t \geqslant 0$ it holds that $V^{t+1} \leqslant e^{\alpha_1} \cdot V^t$ (Lemma 5.28 (i)). Let $t_0 < \tau_1 < \tau_2 < \cdots$ and $t_0 =: s_0 < s_1 < \cdots$ be two interlaced sequences defined recursively for $i \geqslant 1$ by

$$
\tau_i := \inf\{\tau > s_{i-1} : V^\tau \in (e^{\alpha_1} cn, e^{2\alpha_1} cn]\} \qquad \text{and} \qquad s_i := \inf\{s > \tau_i : V^s \leqslant e^{\alpha_1} cn\}.
$$

Thus we have

$$
t_0 = s_0 < \tau_1 < s_1 < \tau_2 < s_2 < \cdots,
$$

and since $\tau_i > \tau_{i-1}$ we have $\tau_{t_1 - t_0} \geqslant t_1 - t_0$. Therefore, if the event $\cap_{i=1}^{t_1-t_0} \{s_i - \tau_i \leqslant \widetilde{\Delta}_s\}$ holds, then also $\widetilde{\mathcal{M}}_{t_0}^{t_1}$ holds.

Recall that by the strong stabilisation (Lemma 5.31) we have for any $i = 1, 2, \ldots, t_1 - t_0$ and any $\tau = t_0 + 1, \ldots, t_1$,

$$
\mathbf{Pr}\left[\bigcup_{t \in [\tau_i, \tau_i + \widetilde{\Delta}_s]} \big\{V^t \leqslant e^{\alpha_1} cn\big\} \;\middle|\; \mathcal{Z}^{r_0}, \mathfrak{F}^\tau,\, e^{\alpha_1} cn < V^\tau \leqslant e^{2\alpha_1} cn, \tau_i = \tau \right] \geqslant 1 - n^{-11},
$$

and by negating and the definition of s_i,

$$\mathbf{Pr}\Big[\, s_i - \tau_i > \widetilde{\Delta}_s \;\Big|\; \mathcal{Z}^{r_0}, \mathfrak{F}^{\tau}, e^{\alpha_1} cn < V^{\tau} \leqslant e^{2\alpha_1} cn, \tau_i = \tau \,\Big] \leqslant n^{-11}.$$

Since the above bound holds for any $i \geqslant 1$ and $\mathfrak{F}^{\tau}$, with $\tau_i = \tau$, it follows by the union bound over all $i = 1, 2, \ldots, t_1 - t_0$, as $t_1 - t_0 \leqslant 2n \log^5 n$,

$$\mathbf{Pr}\Big[\, \neg \widetilde{\mathcal{M}}_{t_0}^{t_1} \;\Big|\; \mathcal{Z}^{r_0}, \mathfrak{F}^{t_0}, V^{t_0} \leqslant cn \,\Big] \leqslant (t_1 - t_0) \cdot n^{-11} \leqslant n^{-9}. \qquad \square$$

5.4.4 Completing the proof of Theorem 5.34

In this section, we will complete the proof of the base case of the layered induction used in Section 7.4.1, using the stronger stabilisation for the V potential (see Lemma 5.31) and upper bounding Ψ_0 using V (see Lemma 5.29). We will first prove the result for the modified process and then relate the results to the g-ADV-COMP setting.

Lemma 5.33. *Consider the $\mathcal{Q}_{g,r_0}$ process for any $g \in [1, c_6 \log n]$ and any step $r_0 \geqslant -\Delta_r - n \log^5 n$, where $c_6 > 0$ is as defined in Eq. (5.22) and $\Delta_r := \Delta_r(g) > 0$ is as in Lemma 5.23. Further, let $c > 0$ be as defined in Lemma 5.21 and $\alpha_1 > 0$ as in Eq. (7.4), then for the constant $C := 2e^{2\alpha_1} \cdot c + 1 \geqslant 8$ and the potential $\Psi_0 := \Psi_0(\alpha_1, c_5 g)$ with the constant integer $c_5 > 0$ (to be defined in Eq. (5.36)), we have that,*

$$\mathbf{Pr}\left[\bigcap_{s \in [r_0 + \Delta_r, r_0 + \Delta_r + n \log^5 n]} \left\{ \Psi_0^s \leqslant Cn \right\} \;\middle|\; \mathcal{Z}^{r_0}, \mathfrak{F}^{r_0} \right] \geqslant 1 - n^{-8}.$$

Proof. This proof proceeds similarly to that of Theorem 5.26 having a recovery and a stabilisation phase to show that the potential $V := V(\alpha_1, c_4 g)$ stabilises at $\leqslant e^{\mathcal{O}(\alpha_1 g)} \cdot cn$. By Lemma 5.29, this implies that $\Psi_0 := \Psi_0(\alpha_1, c_5 g)$ stabilises at $\leqslant Cn$ for sufficiently large constants $c_5, C > 0$.

For the recovery phase, we will use the potential function $\Lambda := \Lambda(\alpha, c_4 g)$ defined in Eq. (5.3). More specifically, we have

$$\begin{aligned}
\mathbf{Pr}\left[\bigcup_{t \in [r_0, r_0 + \Delta_r]} \left\{ V^t \leqslant cn \right\} \;\middle|\; \mathcal{Z}^{r_0}, \mathfrak{F}^{r_0} \right] &\overset{(a)}{\geqslant} \mathbf{Pr}\left[\bigcup_{t \in [r_0, r_0 + \Delta_r]} \left\{ \Lambda^t \leqslant cn \right\} \;\middle|\; \mathcal{Z}^{r_0}, \mathfrak{F}^{r_0} \right] \\
&\overset{(b)}{\geqslant} \mathbf{Pr}\left[\bigcup_{t \in [r_0, r_0 + \Delta_r]} \left\{ \Lambda^t \leqslant cn \right\} \;\middle|\; \mathfrak{F}^{r_0}, \max_{i \in [n]} \left| y_i^{r_0} \right| \leqslant c_3 g \log(ng) \right] \\
&\overset{(c)}{\geqslant} 1 - 3 \cdot (ng)^{-12} \geqslant 1 - n^{-11},
\end{aligned}$$

using in (a) that $V^t \leqslant \Lambda^t$ for any step $t \geqslant 0$, as V and Λ have the same form, but V has a smoothing parameter $\alpha_1 < \alpha$, in (b) recalling that $\mathcal{Z}^{r_0} := \{\max_{i \in [n]} |y_i^{r_0}| \leqslant \min\{\kappa \cdot (g + \log n), c_3 g \log(ng)\}\}$ for $c_3 \geqslant 2$ the constant defined in Eq. (3.32) and in (c) applying Lemma 5.23 (i).

Therefore, for the stopping time $\tau := \inf\{t \geqslant r_0 \colon V^t \leqslant cn\}$, it holds that

$$\mathbf{Pr}[\, \tau \leqslant r_0 + \Delta_r \mid \mathcal{Z}^{r_0}, \mathfrak{F}^{r_0} \,] \geqslant 1 - n^{-11}. \tag{5.35}$$

Consider any $t_0 \in [r_0, r_0 + \Delta_r]$ (t_0 will play the role of a concrete value of τ). By Lemma 5.32 (for $t_0 := t_0$ and $t_1 := r_0 + \Delta_r + n \log^5 n$, since $t_1 - t_0 \leqslant n \log^5 n + \Delta_r \leqslant 2n \log^5 n$), we have that

$$\mathbf{Pr}\left[\bigcap_{t \in [t_0, r_0 + \Delta_r + n \log^5 n]} \bigcup_{s \in [t, t + \widetilde{\Delta}_s]} \left\{ V^s \leqslant e^{2\alpha_1} cn \right\} \;\middle|\; \mathcal{Z}^{r_0}, \mathfrak{F}^{t_0}, V^{t_0} \leqslant cn \right] \geqslant 1 - n^{-9}.$$

When the above event holds, then for every $t \in [t_0, r_0 + \Delta_r + n \log^5 n]$, there exists $s_0 \in [t - \widetilde{\Delta}_s, t]$ and $s_1 \in [t, t + \widetilde{\Delta}_s]$ such that $V^{s_0} \leqslant e^{2\alpha_1} cn$ and $V^{s_1} \leqslant e^{2\alpha_1} cn$, using that for any $t \in [t_0, t_0 + \widetilde{\Delta}_s]$ we can set $s_0 := t_0$ since by the conditioning $V^{t_0} \leqslant cn \leqslant e^{2\alpha_1} cn$. So, by Lemma 5.28 (ii) (for $\hat{c} := e^{2\alpha_1} c$ and $T := \widetilde{\Delta}_s$), it follows that

$$\mathbf{Pr}\left[\bigcap_{t \in [t_0, r_0 + \Delta_r + n \log^5 n]} \left\{ V^t \leqslant e^{\alpha_1 \frac{\widetilde{\Delta}_s}{n}} \cdot 2e^{2\alpha_1} cn \right\} \;\middle|\; \mathcal{Z}^{r_0}, \mathfrak{F}^{t_0}, V^{t_0} \leqslant cn \right] \geqslant 1 - n^{-9}.$$

Next, adjusting the range of the big intersection using that $t_0 \leqslant r_0 + \Delta_r$, it follows that

$$\mathbf{Pr}\left[\bigcap_{t \in [r_0 + \Delta_r, r_0 + \Delta_r + n \log^5 n]} \left\{ V^t \leqslant e^{\alpha_1 \frac{\widetilde{\Delta}_s}{n}} \cdot 2e^{2\alpha_1} cn \right\} \;\middle|\; \mathcal{Z}^{r_0}, \mathfrak{F}^{t_0}, V^{t_0} \leqslant cn \right] \geqslant 1 - n^{-9}.$$

By Lemma 5.29 (for $\hat{c} := \lceil \frac{\widetilde{\Delta}_s}{ng} \rceil$), we conclude that for $\Psi_0 := \Psi_0(\alpha_1, c_5 g)$ with constant integer

$$c_5 := 2 \cdot \max\left\{ c_4, \left\lceil \frac{\widetilde{\Delta}_s}{ng} \right\rceil \right\} = 2 \cdot \max\left\{ c_4, \left\lceil \frac{20 \cdot \widetilde{c}_s \cdot \log(2ce^{2\alpha_1})}{\alpha_1 \epsilon r} \right\rceil \right\}, \tag{5.36}$$

and for the constant $C := 2e^{2\alpha_1} \cdot c + 1$, it holds that

$$\mathbf{Pr}\left[\bigcap_{t \in [r_0 + \Delta_r, r_0 + \Delta_r + n \log^5 n]} \left\{ \Psi_0^t \leqslant Cn \right\} \;\middle|\; \mathcal{Z}^{r_0}, \mathfrak{F}^{t_0}, V^{t_0} \leqslant cn \right] \geqslant 1 - n^{-9}. \tag{5.37}$$

Finally,

$$\begin{aligned}
&\mathbf{Pr}\left[\bigcap_{t \in [r_0 + \Delta_r, r_0 + \Delta_r + n \log^5 n]} \left\{ \Psi_0^t \leqslant Cn \right\} \;\middle|\; \mathcal{Z}^{r_0}, \mathfrak{F}^{r_0} \right] \\
&\quad \geqslant \sum_{t_0 = r_0}^{r_0 + \Delta_r} \mathbf{Pr}\left[\bigcap_{t \in [r_0 + \Delta_r, r_0 + \Delta_r + n \log^5 n]} \left\{ \Psi_0^t \leqslant Cn \right\} \;\middle|\; \mathcal{Z}^{r_0}, \mathfrak{F}^{r_0}, \tau = t_0 \right] \cdot \mathbf{Pr}\left[\tau = t_0 \mid \mathcal{Z}^{r_0}, \mathfrak{F}^{r_0} \right] \\
&\quad \geqslant \sum_{t_0 = r_0}^{r_0 + \Delta_r} \mathbf{Pr}\left[\bigcap_{t \in [r_0 + \Delta_r, r_0 + \Delta_r + n \log^5 n]} \left\{ \Psi_0^t \leqslant Cn \right\} \;\middle|\; \mathcal{Z}^{r_0}, \mathfrak{F}^{t_0}, V^{t_0} \leqslant cn \right] \cdot \mathbf{Pr}\left[\tau = t_0 \mid \mathcal{Z}^{r_0}, \mathfrak{F}^{r_0} \right] \\
&\quad \overset{(5.37)}{\geqslant} \left(1 - n^{-9}\right) \cdot \mathbf{Pr}\left[\tau \leqslant r_0 + \Delta_r \mid \mathcal{Z}^{r_0}, \mathfrak{F}^{r_0} \right] \\
&\quad \overset{(5.35)}{\geqslant} \left(1 - n^{-9}\right) \cdot \left(1 - n^{-11}\right) \geqslant 1 - n^{-8},
\end{aligned}$$

This concludes the claim for $r_0 \geqslant 0$.

If $r_0 < 0$, then deterministically $\tau = 0$, since $V^\tau = n \leqslant cn$. Hence, the rest of the proof follows for an interval of length at most $2n \log^5 n$. $\square$

Theorem 5.34 (Base case). *Consider the g-ADV-COMP setting for any $g \in [1, c_6 \log n]$, where $c_6 > 0$ is as defined in Eq.* (5.22)*. For constant $C := 2e^{2\alpha_1} \cdot c + 1 \geqslant 8$ with $\alpha_1 > 0$ as defined in Eq.* (7.4) *and $c > 0$*

as in Lemma 5.21, and the potential $\Phi_0 := \Phi_0(\alpha_2, c_5 g)$ *with* $\alpha_2 > 0$ *as in Eq.* (7.5)*, and constant integer* $c_5 > 0$ *as in Eq.* (5.36)*, we have that for any step* $m \geqslant 0$*,*

$$\mathbf{Pr}\left[\bigcap_{s\in[m-n\log^5 n, m]} \left\{\Phi_0^s \leqslant Cn\right\}\right] \geqslant 1-n^{-4}.$$

Proof. Let $\mathcal{P}$ be the original process in the g-Adv-Comp setting. Consider the modified process $\mathcal{Q}_{g,r_0}$ for $r_0 := m - n\log^5 n - \Delta_r$ where $\Delta_r := \Theta(ng\cdot(\log(ng))^2)$ is as defined in Lemma 5.23. Let $y_{\mathcal{Q}}$ be its normalised load vector and $\Psi_{\mathcal{Q},0}$ be its Ψ_0 potential function. Recall from Eq. (5.25) that the two processes $\mathcal{Q}_{g,r_0}$ and $\mathcal{P}$ agree at every step w.h.p., since $m - r_0 \leqslant 2n\log^5 n$,

$$\mathbf{Pr}\left[\bigcap_{s\in[r_0, m]} \left\{y_{\mathcal{Q}}^s = y^s\right\}\right] \geqslant 1-2n^{-7}. \tag{5.38}$$

Taking the union bound of the conclusions in Theorem 5.26 and Theorem 3.21 (iii), we have that

$$\mathbf{Pr}\left[\,\mathcal{Z}^{r_0}\,\right] = \mathbf{Pr}\left[\max_{i\in[n]} \left|y_i^{r_0}\right| \leqslant \min\{\kappa\cdot(g+\log n), c_3 g\log(ng)\}\right] \geqslant 1-3\cdot(ng)^{-9}. \tag{5.39}$$

By Lemma 5.33 we have that,

$$\mathbf{Pr}\left[\bigcap_{s\in[m-n\log^5 n, m]} \left\{\Psi_{\mathcal{Q},0}^s \leqslant Cn\right\} \,\middle|\, \mathfrak{F}^{r_0}, \mathcal{Z}^{r_0}\right] \geqslant 1-n^{-8}.$$

By combining with Eq. (5.39),

$$\mathbf{Pr}\left[\bigcap_{s\in[m-n\log^5 n, m]} \left\{\Psi_{\mathcal{Q},0}^s \leqslant Cn\right\}\right] \geqslant \left(1-n^{-8}\right)\cdot\left(1-3\cdot(ng)^{-9}\right) \geqslant 1-n^{-7}.$$

Taking the union bound with Eq. (5.38), we get that

$$\mathbf{Pr}\left[\bigcap_{s\in[r_0, m]} \left\{y_{\mathcal{Q}}^s = y^s\right\} \cap \bigcap_{s\in[m-n\log^5 n, m]} \left\{\Psi_{\mathcal{Q},0}^s \leqslant Cn\right\}\right] \geqslant 1-n^{-7}-2n^{-7} \geqslant 1-n^{-6}.$$

When this event holds we have that $\Psi_{\mathcal{Q},0}^s = \Psi_0^s$ for every step $s \in [r_0, m]$ and hence we can deduce for the original process $\mathcal{P}$ that

$$\mathbf{Pr}\left[\bigcap_{s\in[m-n\log^5 n, m]} \left\{\Psi_0^s \leqslant Cn\right\}\right] \geqslant 1-n^{-6}.$$

Finally, since $\alpha_2 \leqslant \alpha_1$, we have that $\Phi_0^s \leqslant \Psi_0^s$ for any step $s \geqslant 0$ and hence, the conclusion follows. $\square$

Super-exponential potentials

In this chapter, we will analyse super-exponential potential functions, i.e., exponential potentials with smoothing parameters $\geqslant 1$. In Section 6.1, we define the specific form of super-exponential potentials that we will be working with and give an overview of the theorems we prove. In Section 6.2, we give two sufficient conditions for the expectation of a super-exponential potential to satisfy a drop inequality over one step, and in Section 6.3, we prove the concentration theorem for super-exponential potentials.

In Chapter 7, super-exponential potentials are used to analyse a large number of processes. More specifically, a layered induction argument is applied over a (not-necessarily constant) number of super-exponential potentials with increasing smoothing parameters and using the concentration theorem at one layer to establish the sufficient condition for the next one. In [163], a single super-exponential potential was used to analyse load balancing of tokens in graphs. In Section 7.5, we prove similar concentration theorems for the b-Batched setting, where the proof is slightly more involved.

6.1 Outline

We start by defining the specific form of *super-exponential potentials* that we will be using.

Definition 6.1. *The* super-exponential potential function *with smoothing parameter $\phi \geqslant 1$ and integer offset $z := z(n) > 0$ is defined for any step $t \geqslant 0$ as*

$$\Phi^t := \Phi^t(\phi, z) := \sum_{i=1}^{n} \Phi_i^t := \sum_{i=1}^{n} e^{\phi \cdot (y_i^t - z)^+},$$

where $u^+ = \max\{u, 0\}$.

Note that if for some step $t \geqslant 0$, we have that $\Phi^t = \mathcal{O}(\operatorname{poly}(n))$, then

$$\operatorname{Gap}(t) = \mathcal{O}\left(z + \frac{\log n}{\phi}\right).$$

There are two differences in the *form* compared to the hyperbolic cosine potential Γ that we used in Chapters 3 to 5: (i) there is no underloaded component, as w.h.p., its contribution would be $\omega(n)$ for any d-Sample process (with $d = \mathcal{O}(1)$)[1] and (ii) there is the $(\ldots)^+$ operation which is not essential, but simplifies some of the derivations.

However, the *main* difference is that for processes allocating one ball in each step, there exist load vectors where the potential *could increase* in expectation, even when sufficiently large.[2] We will show that in each step where the probability to allocate to a bin with load at least $z-1$ is sufficiently small, the potential function Φ drops in expectation over one step. More specifically, we require that the following condition holds at step $s \geqslant 0$,

$$\mathcal{K}^s := \mathcal{K}_\phi^s(q^s) := \left\{ \left(\widetilde{q}_1^s, \ldots, \widetilde{q}^s_{|\mathcal{B}^s_{\geqslant z-1}|} \right) \preceq \left(\frac{1}{n} \cdot e^{-\phi}, \ldots, \frac{1}{n} \cdot e^{-\phi} \right) \right\},$$

[1] It follows by a coupon collector's argument that the minimum load is w.h.p. $\frac{t}{n} - \Omega(\log n)$, for sufficiently large t.

[2] An example of such a configuration for the Two-Choice process is one where $n-1$ bins have normalised loads $z + \Omega(1)$ and one bin is underloaded.

where $\mathcal{B}^s_{\geqslant z-1} := \{i \in [n] : y_i^s \geqslant z-1\}$ and $\widetilde{q}^s$ is the sorted allocation vector used by the process at step s.

In Chapter 7, for most applications we make use of the following simpler condition which implies $\mathcal{K}^s$,

$$\widetilde{\mathcal{K}}^s := \widetilde{\mathcal{K}}^s_{\phi}(q^s) := \left\{ \forall i \in [n]\colon \ y_i^s \geqslant z-1 \ \Rightarrow \ q_i^s \leqslant \frac{1}{n} \cdot e^{-\phi} \right\}.$$

Lemma 6.2 (General drop inequality). *Consider any $\mathcal{P} = \textsc{Sequential}(q^s)$ process and any super-exponential potential $\Phi := \Phi(\phi, z)$ with $\phi \in [4, n]$. At any step $s \geqslant 0$, satisfying condition $\mathcal{K}^s := \mathcal{K}^s_{\phi}(q^s)$, we have that*

$$\mathbf{E}\left[\, \Phi^{s+1} \,\middle|\, \mathfrak{F}^s, \mathcal{K}^s \,\right] \leqslant \Phi^s \cdot \left(1 - \frac{1}{n}\right) + 2.$$

We will try to establish that this condition $\mathcal{K}^s$ holds for a sufficiently long interval and then show that in this interval the potential becomes small. In most of our applications in Chapter 7, this condition will arise from the concentration of the hyperbolic cosine potential Γ or from a super-exponential potential with smaller smoothing parameter.

Now we are ready to state the main theorem of this chapter.

Theorem 6.9 (Restated, page 133). *Consider any $\mathcal{P} = \textsc{Sequential}(q^s)$ process for which there exist super-exponential potential functions $\Phi_1 := \Phi_1(\phi_1, z)$ and $\Phi_2 := \Phi_2(\phi_2, z)$ with integer offset $z := z(n) > 0$ and smoothing parameters $\phi_1, \phi_2 \in (0, (\log n)/6]$ with $\phi_2 \leqslant \frac{\phi_1}{84}$, such that they satisfy for any step $s \geqslant 0$,*

$$\mathbf{E}\left[\, \Phi_1^{s+1} \,\middle|\, \mathfrak{F}^s, \mathcal{K}^s \,\right] \leqslant \Phi_1^s \cdot \left(1 - \frac{1}{n}\right) + 2, \tag{6.1}$$

and

$$\mathbf{E}\left[\, \Phi_2^{s+1} \,\middle|\, \mathfrak{F}^s, \mathcal{K}^s \,\right] \leqslant \Phi_2^s \cdot \left(1 - \frac{1}{n}\right) + 2, \tag{6.2}$$

where $\mathcal{K}^s := \mathcal{K}^s_{\phi_1}(q^s)$. Further, let $P \geqslant n^{-4}$. Then, for any steps $t \geqslant 0$ and $\widetilde{t} \in [t, t + n\log^5 n]$, which satisfy

$$\mathbf{Pr}\left[\left\{ \mathrm{Gap}(t - 2n\log^4 n) \leqslant \log^2 n \right\} \cap \bigcap_{s \in [t - 2n\log^4 n, \widetilde{t}]} \mathcal{K}^s \right] \geqslant 1 - P, \tag{6.3}$$

they must also satisfy

$$\mathbf{Pr}\left[\bigcap_{s \in [t, \widetilde{t}]} \left\{ \Phi_2^s \leqslant 8n \right\} \right] \geqslant 1 - (\log^8 n) \cdot P.$$

The statement of this theorem concerns steps in $[t - 2n\log^4 n, \widetilde{t}]$ with $\widetilde{t} \in [t, t + n\log^5 n]$. The interval $[t, \widetilde{t}]$ is the *stabilisation interval*, i.e., the interval where we want to show that $\Phi_2^s \leqslant 8n$ for every $s \in [t, \widetilde{t}]$. The interval $[t - 2n\log^4 n, t]$ is the *recovery* interval where we will show that w.h.p. Φ_2 becomes $\mathcal{O}(n)$ at least *once*, provided we start with a "weak" $\mathcal{O}(\log^2 n)$ gap at step $t - 2n\log^4 n$. For both the recovery and stabilisation intervals we will condition on the event $\mathcal{K}$ holding at every step.

For convenience, we define the event

$$\mathcal{H}^{t - 2n\log^4 n} := \left\{ \mathrm{Gap}(t - 2n\log^4 n) \leqslant \log^2 n \right\}.$$

Several applications of this theorem can be found throughout Chapter 7 and in particular in Sections 7.1 and 7.2.

Dealing with negative indices. To avoid notational clutter in the statements and in the proofs, we will be working with a modified process $\mathcal{P}' = \textsc{Sequential}(\widetilde{q}^t)$ which in negative time steps $t < 0$ performs round robin allocation, i.e., $\widetilde{q}^t = (0, \ldots, 0, 1)$, otherwise follows $\mathcal{P}$. Hence, for any $t < 0$, we have that $|y_i^t| < 1$ and so precondition Eq. (6.3) is trivially satisfied as $z \geqslant 1$. Further, for any $t < 0$, we have that $\Phi_1^t = \Phi_2^t = n$ (since $z \geqslant 1$) and hence preconditions Eq. (6.1) and Eq. (6.2) are trivially satisfied.

6.2 General drop inequality for the expectation

We start by proving a weaker form of the above theorem, where we assume that the probability bound holds in pointwise manner, i.e., using condition $\widetilde{\mathcal{K}}_\phi$.

Lemma 6.3. *Consider any $\mathcal{P} = \textsc{Sequential}(q^s)$ process and any super-exponential potential $\Phi := \Phi(\phi, z)$ with $\phi \in [4, n]$. At any step $s \geqslant 0$ satisfying condition $\widetilde{\mathcal{K}}^s := \widetilde{\mathcal{K}}_\phi^s(q^s)$, we have that*

$$\mathbf{E}\left[\, \Phi^{s+1} \,\middle|\, \mathfrak{F}^s, \widetilde{\mathcal{K}}^s \,\right] \leqslant \Phi^s \cdot \left(1 - \frac{1}{n}\right) + 2.$$

Proof. We consider three cases for the contribution of a bin $i \in [n]$:

Case 1 $[y_i^s < z - 1]$: The contribution of i will remain $\Phi_i^{s+1} = \Phi_i^s = 1$, even if a ball is allocated to bin i. Hence,

$$\mathbf{E}\left[\, \Phi_i^{s+1} \,\middle|\, \mathfrak{F}^s, \widetilde{\mathcal{K}}^s \,\right] = \Phi_i^s = \Phi_i^s \cdot \left(1 - \frac{1}{n}\right) + \frac{1}{n}. \tag{6.4}$$

Case 2 $[y_i^s \in [z-1, z]]$: By the condition $\widetilde{\mathcal{K}}^s$, the probability of allocating a ball to bin i with $y_i^s \geqslant z - 1$ is $q_i^s \leqslant \frac{1}{n} \cdot e^{-\phi}$. Hence, the expected contribution of this bin is at most

$$\begin{aligned}
\mathbf{E}\left[\, \Phi_i^{s+1} \,\middle|\, \mathfrak{F}^s, \widetilde{\mathcal{K}}^s \,\right] &\leqslant \Phi_i^s \cdot e^{\phi \cdot (1 - 1/n)} \cdot q_i^s + \Phi_i^s \cdot (1 - q_i^s) \\
&\leqslant \Phi_i^s \cdot e^{\phi \cdot (1 - 1/n)} \cdot q_i^s + \Phi_i^s \\
&\leqslant \Phi_i^s \cdot e^{\phi \cdot (1 - 1/n)} \cdot \frac{1}{n} \cdot e^{-\phi} + \Phi_i^s \\
&\leqslant \frac{1}{n} + \Phi_i^s = \Phi_i^s \cdot \left(1 - \frac{1}{n}\right) + \frac{2}{n},
\end{aligned} \tag{6.5}$$

using in the last equation that $\Phi_i^s = 1$.

Case 3 $[y_i^s > z]$: Again, by the condition $\widetilde{\mathcal{K}}^s$, the probability of allocating a ball to bin i with $y_i^s > z$ is $q_i^s \leqslant \frac{1}{n} \cdot e^{-\phi}$. Hence,

$$\begin{aligned}
\mathbf{E}\left[\, \Phi_i^{s+1} \,\middle|\, \mathfrak{F}^s, \widetilde{\mathcal{K}}^s \,\right] &\overset{(a)}{=} \Phi_i^s \cdot e^{\phi \cdot (1 - 1/n)} \cdot q_i^s + \Phi_i^s \cdot e^{-\phi/n} \cdot (1 - q_i^s) \\
&\leqslant \Phi_i^s \cdot e^{\phi \cdot (1 - 1/n)} \cdot q_i^s + \Phi_i^s \cdot e^{-\phi/n} \\
&= \Phi_i^s \cdot e^{-\phi/n} \cdot \left(1 + e^{\phi} \cdot q_i^s\right) \\
&\overset{(b)}{\leqslant} \Phi_i^s \cdot \left(1 - \frac{\phi}{2n}\right) \cdot \left(1 + e^{\phi} \cdot \frac{1}{n} \cdot e^{-\phi}\right) \\
&\overset{(c)}{\leqslant} \Phi_i^s \cdot \left(1 - \frac{2}{n}\right) \cdot \left(1 + \frac{1}{n}\right) \\
&\leqslant \Phi_i^s \cdot \left(1 - \frac{1}{n}\right),
\end{aligned} \tag{6.6}$$

using in (a) that $y_i^s > z$ implies that $y_i^s \geqslant z + \frac{1}{n}$ since z is an integer, in (b) that $e^u \leqslant 1 + \frac{1}{2}u$ (for any $-1.5 \leqslant u < 0$) and that $\phi \leqslant n$ and in (c) that $\phi \geqslant 4$.

Aggregating over the three cases, we get the claim:

$$\mathbf{E}\left[\, \Phi^{s+1} \,\middle|\, \mathfrak{F}^s, \widetilde{\mathcal{K}}^s \,\right] = \sum_{i=1}^{n} \mathbf{E}\left[\, \Phi_i^{s+1} \,\middle|\, \mathfrak{F}^s, \widetilde{\mathcal{K}}^s \,\right] \leqslant \sum_{i=1}^{n} \left(\Phi_i^s \cdot \left(1 - \frac{1}{n}\right) + \frac{2}{n} \right) = \Phi^s \cdot \left(1 - \frac{1}{n}\right) + 2. \qquad \square$$

Now, we turn to proving the stronger condition of the theorem, which assumes that the weaker "majorisation" condition $\mathcal{K}^s$ holds.

Lemma 6.2 (Restated, page 126). *Consider any $\mathcal{P} = \text{SEQUENTIAL}(q^s)$ process and any super-exponential potential $\Phi := \Phi(\phi, z)$ with $\phi \in [4, n]$. At any step $s \geqslant 0$, satisfying condition $\mathcal{K}^s := \mathcal{K}_\phi^s(q^s)$, we have that*

$$\mathbf{E}\left[\, \Phi^{s+1} \,\middle|\, \mathfrak{F}^s, \mathcal{K}^s \,\right] \leqslant \Phi^s \cdot \left(1 - \frac{1}{n}\right) + 2.$$

Proof. By combining the upper bounds Eq. (6.4), Eq. (6.5) and Eq. (6.6) from the three cases in the proof of Lemma 6.3, we have that

$$\begin{aligned} \mathbf{E}\left[\, \Phi^{s+1} \middle| \mathfrak{F}^s, \mathcal{K}^s \,\right] \leqslant \sum_{i : y_i^s < z-1} \left(\Phi_i^s \cdot \left(1 - \frac{1}{n}\right) + \frac{1}{n} \right) + \sum_{i : y_i^s \in [z-1, z]} \Phi_i^s \cdot \left(e^{\phi \cdot (1-1/n)} \cdot q_i^s + 1 \right) \\ + \sum_{i : y_i^s > z} \Phi_i^s \cdot \left(e^{\phi \cdot (1-1/n)} \cdot q_i^s + e^{-\phi/n} \right) \end{aligned} \qquad (6.7)$$

Let $\mathcal{B}_{\geqslant z-1}^s := \{ i : y_i^s \geqslant z-1 \}$ be the set of bins with normalised load at least $z-1$. The term that depends on y^s is non-increasing over the loads of the bins in $\mathcal{B}_{\geqslant z-1}^s$, and is given by

$$f(y^s) := \sum_{i : y_i^s \geqslant z-1} e^{\phi (y_i^s - z)^+} \cdot e^{\phi \cdot (1-1/n)} \cdot q_i^s.$$

Therefore, using Lemma B.2 for the vectors $(\widetilde{q}_1^t, \ldots, \widetilde{q}_{|\mathcal{B}_{\geqslant z-1}^t|}^t)$ and $\left(\frac{1}{n} \cdot e^{-\phi}, \ldots, \frac{1}{n} \cdot e^{-\phi}\right)$, and the majorisation in condition $\mathcal{K}^s := \mathcal{K}^s(q^s)$, we have the upper bound of Eq. (6.7) is maximised when all bins $i \in \mathcal{B}_{\geqslant z-1}$ have the same probability

$$q_i^s = \frac{1}{n} \cdot e^{-\phi}.$$

Hence, we can apply Lemma 6.3 to deduce that

$$\mathbf{E}\left[\, \Phi^{s+1} \,\middle|\, \mathfrak{F}^s, \mathcal{K}^s \,\right] \leqslant \Phi^s \cdot \left(1 - \frac{1}{n}\right) + 2. \qquad \square$$

6.3 Concentration

6.3.1 Proof outline of Theorem 6.9

We will now give a summary of the main technical steps in the proof of Theorem 6.9 (an illustration of the key steps is shown in Fig. 6.1). The proof is similar to the proof of Theorem 4.1 for the concentration of the hyperbolic cosine potential, in the sense that we use two instances $\Phi_1 := \Phi_1(\phi_1, z)$ and $\Phi_2 := \Phi_2(\phi_2, z)$ of the super-exponential potential function with $\phi_2 \leqslant \phi_1/84$, such that in steps $s \geqslant 0$ when Φ_1^s is small, then the change of Φ_2^s is very small.

Recovery. By the third precondition Eq. (6.3) of Theorem 6.9, we start with $\operatorname{Gap}(t-2n\log^4 n) \leqslant \log^2 n$, which implies that $\Phi_1^{t-2n\log^4 n} \leqslant e^{\frac{1}{2}\cdot\log^4 n}$ (Claim 6.4). Using the drop inequality for the potential Φ_1 (first precondition Eq. (6.1)), it follows that $\mathbf{E}[\,\Phi_1^s\,] \leqslant 6n$, for any step $s \in [t - n\log^4 n, \widetilde{t}]$ (Lemma 6.7). By using Markov's inequality and a union bound, we can deduce that w.h.p. $\Phi_1^s \leqslant 6n^{12}$ for all steps $s \in [t - n\log^4 n, \widetilde{t}]$. By a simple interplay between two potentials, this implies $\Phi_2^{t-n\log^4 n} \leqslant n^{7/6}$ (Lemma 6.5 (i)). Now using a drop inequality for the potential Φ_2 (second precondition Eq. (6.2)), guarantees that w.h.p. $\Phi_2^{r_0} \leqslant 6n$ for *some single* step $r_0 \in [t - n\log^4 n, t]$ (Lemma 6.8).

Stabilisation. To obtain the stronger statement which holds *for all* steps $s \in [t, \widetilde{t}]$, we will use a concentration inequality. The key point is that for any step r with $\Phi_1^r \leqslant 6n^{12}$ the absolute difference $|\Phi_2^{r+1} - \Phi_2^r|$ is at most $n^{1/3}$, because $\phi_2 \leqslant \frac{\phi_1}{84}$ (by preconditions Eq. (6.1) and Eq. (6.2)). This is crucial for applying the Azuma's inequality for super-martingales (Lemma B.10) to Φ_2 which yields that $\Phi_2^s \leqslant 8n$ for all steps $s \in [t, \widetilde{t}]$ using a smoothing argument (Claim 6.6).

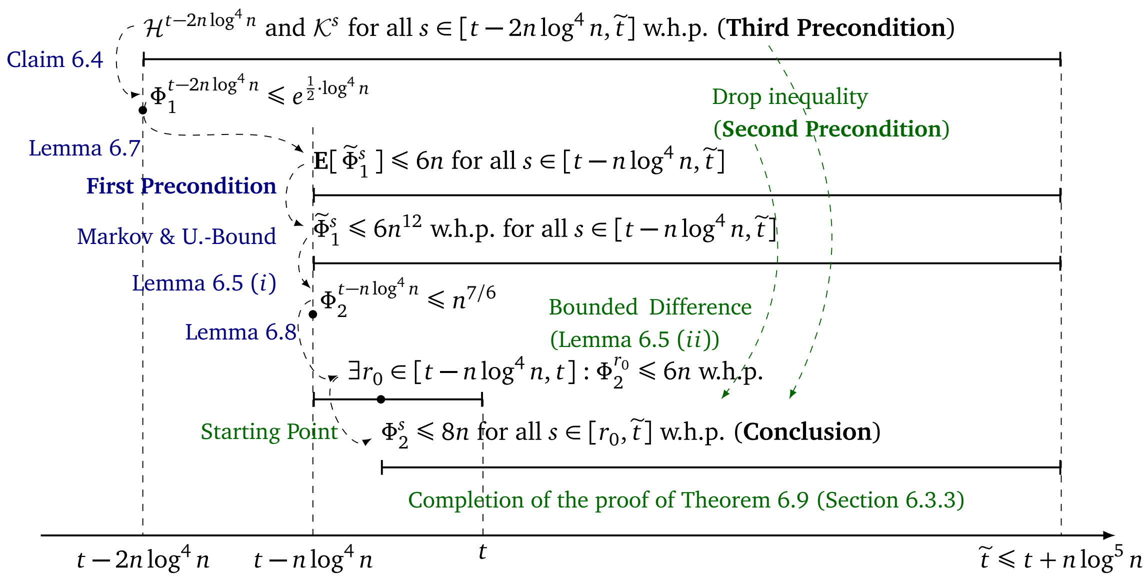


Figure 6.1: Outline for the proof of Theorem 6.9. Results in blue are given in Section 6.3.2, while results in green are used in the completion of the proof in Section 6.3.3.

6.3.2 Deterministic relations between the potential functions

We collect several basic facts about the super-exponential potential functions $\Phi_1 := \Phi_1(\phi_1, z)$ and $\Phi_2 := \Phi_2(\phi_2, z)$ satisfying the preconditions in Theorem 6.9.

We start with a simple upper bound on Φ_1^s using a weak upper bound on the gap at step s.

Claim 6.4. *For any step $s \geqslant 0$, where $\operatorname{Gap}(s) \leqslant \log^2 n$, we have that $\Phi_1^s \leqslant e^{\frac{1}{2}\cdot\log^4 n}$.*

Proof. Since $\phi_1 \leqslant (\log n)/6$, we have that $\Phi_1^s = \sum_{i=1}^n e^{\phi_1 \cdot (y_i^s - z)^+} \leqslant n \cdot e^{\phi_1 \cdot \log^2 n} \leqslant e^{\frac{1}{2}\cdot\log^4 n}$. □

The next lemma is crucial for applying the concentration inequality, since the second statement bounds the maximum additive change of Φ_2^s (assuming Φ_1^s is poly(n)):

Lemma 6.5. *For any step $s \geqslant 0$, if $\Phi_1^s \leqslant 6n^{12}$, then*

$$\begin{aligned} &(i) \quad \Phi_2^s \leqslant n^{7/6}, \\ &(ii) \quad \left|\Phi_2^{s+1} - \Phi_2^s\right| \leqslant n^{1/3}. \end{aligned}$$

Proof. Consider an arbitrary step $s \geqslant 0$ with $\Phi_1^s \leqslant 6n^{12}$. We start by upper bounding the normalised load y_i^s of any bin $i \in [n]$,

$$y_i^s \leqslant z + \frac{\log\left(\Phi_{1,i}^s\right)}{\phi_1} \leqslant z + \frac{\log\left(\Phi_1^s\right)}{\phi_1} \leqslant z + \frac{\log(6n^{12})}{\phi_1} \leqslant z + \frac{14 \log n}{\phi_1}.$$

First statement. Now, we upper bound the contribution of any bin $i \in [n]$ to Φ_2^s,

$$\Phi_{2,i}^s = \exp\!\left(\phi_2 \cdot \left(y_i^s - z\right)^+\right) \leqslant \exp\left(\frac{14 \cdot \phi_2}{\phi_1} \cdot \log n\right) \leqslant n^{1/6}, \tag{6.8}$$

using that $\phi_2 \leqslant \frac{\phi_1}{6 \cdot 14}$. Hence, by aggregating over all bins,

$$\Phi_2^s \leqslant n \cdot n^{1/6} = n^{7/6}. \tag{6.9}$$

Second statement. We will obtain lower and upper bounds for Φ_2^{s+1} in terms of Φ_2^s. For the upper bound, let $i = i^{s+1} \in [n]$ be the bin where the $(s+1)$-th ball is allocated, then

$$\Phi_2^{s+1} \leqslant \Phi_2^s + \Phi_{2,i}^s \cdot e^{\phi_2} \leqslant \Phi_2^s + n^{1/6} \cdot n^{1/6} = \Phi_2^s + n^{1/3},$$

using that Eq. (6.8) and $\phi_2 \leqslant (\log n)/6$. For the lower bound, we pessimistically assume that all bin loads decrease by $1/n$ in step $s+1$, so

$$\Phi_2^{s+1} \geqslant \Phi_2^s \cdot e^{-\phi_2/n} \overset{(a)}{\geqslant} \Phi_2^s \cdot \left(1 - \frac{\phi_2}{n}\right) \overset{(b)}{\geqslant} \Phi_2^s - \frac{n \cdot n^{1/6}}{n} \cdot \log n \geqslant \Phi_2^s - n^{1/3},$$

using in (a) that $e^u \geqslant 1 + u$ (for any u) and in (b) that $\phi_2 \leqslant \log n$ and $\Phi_2^s \leqslant n \cdot n^{1/6}$ by Eq. (6.9). Combining the two bounds we get the second statement. □

The next claim is a simple "smoothness" argument showing that the potential cannot decrease quickly within $\lceil n/\log^2 n\rceil$ steps. The derivation is elementary and relies on the fact that within this time frame the average load changes by at most $2/\log^2 n$.

Claim 6.6. *For any step $s \geqslant 0$ and any step $r \in [s, s + \lceil n/\log^2 n\rceil]$, we have that $\Phi_2^r \geqslant 0.99 \cdot \Phi_2^s$.*

Proof. The normalised load after $r - s$ steps can decrease by at most $\frac{r-s}{n} \leqslant \frac{2}{\log^2 n}$. Hence, for any bin $i \in [n]$,

$$\Phi_{2,i}^r = e^{\phi_2 \cdot (y_i^r - z)^+} \geqslant e^{\phi_2 \cdot (y_i^s - \frac{r-s}{n} - z)^+} \geqslant e^{\phi_2 \cdot (y_i^s - z)^+ - \phi_2 \cdot \frac{2}{\log^2 n}} = \Phi_{2,i}^s \cdot e^{-\frac{2\phi_2}{\log^2 n}} \geqslant \Phi_{2,i}^s \cdot e^{-o(1)} \geqslant 0.99 \cdot \Phi_{2,i}^s,$$

for sufficiently large n, using that $\phi_2 \leqslant (\log n)/6$. By aggregating over all bins, we get the claim. □

Recovery Phase

In this section, we will show for an auxiliary process $\widetilde{\mathcal{P}}$ (to be defined below) that the potential function Φ_2 satisfies $\Phi_2^s \leqslant 6n$ in *at least one step* $s \in [t - n\log^4 n, t]$ w.h.p.

First, we show that for the original process $\mathcal{P}$ in the statement of Theorem 6.9, the potential

$$\widetilde{\Phi}_1^s := \widetilde{\Phi}_1^s(t) := \Phi_1^s \cdot \mathbf{1}_{\cap_{r\in[t-2n\log^4 n, s]}\mathcal{K}^r \cap \mathcal{H}^{t-2n\log^4 n}} \tag{6.10}$$

is small in expectation for *all* steps $s \geqslant t - n\log^4 n$. Note that there is a "recovery time" until the expectation becomes small, of at most $n\log^4 n$ steps after the "weak" bound $\mathrm{Gap}(t-2n\log^4 n) \leqslant \log^2 n$ which follows from the third precondition Eq. (6.3) in Theorem 6.9.

Lemma 6.7. *Consider the potential* $\widetilde{\Phi}_1 := \widetilde{\Phi}_1(t)$ *for any step* $t \geqslant 0$. *Then, for any step* $s \geqslant t - n\log^4 n$,

$$\mathbf{E}\left[\, \widetilde{\Phi}_1^s \,\right] \leqslant 6n.$$

Proof. By the precondition Eq. (6.1) of Theorem 6.9, for any step $s \geqslant t - n\log^4 n$,

$$\mathbf{E}\left[\, \Phi_1^{s+1} \,\middle|\, \mathfrak{F}^s, \mathcal{K}^s \,\right] \leqslant \Phi_1^s \cdot \left(1 - \frac{1}{n}\right) + 2.$$

Next note that whenever $\neg\mathcal{K}^s$ holds, it follows deterministically that $\left\{\widetilde{\Phi}_1^s = \widetilde{\Phi}_1^{s+1} = 0\right\}$, and hence

$$\mathbf{E}\left[\, \widetilde{\Phi}_1^{s+1} \,\middle|\, \mathfrak{F}^s \,\right] \leqslant \widetilde{\Phi}_1^s \cdot \left(1 - \frac{1}{n}\right) + 2. \tag{6.11}$$

We will now upper bound $\mathbf{E}\left[\, \widetilde{\Phi}_1^s \,\middle|\, \mathfrak{F}^{t-2n\log^4 n}, \mathcal{H}^{t-2n\log^4 n} \,\right]$ for any step $s \geqslant t - n\log^4 n$. When $\mathcal{H}^{t-2n\log^4 n}$ holds, by Claim 6.4, it also follows that $\widetilde{\Phi}_1^{t-2n\log^4 n} \leqslant \Phi_1^{t-2n\log^4 n} \leqslant e^{\frac{1}{2}\cdot\log^4 n}$. Hence applying Lemma B.1 (i) (for $a = 1 - \frac{1}{n}$ and $b = 2$) using Eq. (6.11),

$$\begin{aligned}
\mathbf{E}\left[\, \widetilde{\Phi}_1^s \,\middle|\, \mathfrak{F}^{t-2n\log^4 n}, \mathcal{H}^{t-2n\log^4 n} \,\right] &\leqslant \mathbf{E}\left[\, \widetilde{\Phi}_1^s \,\middle|\, \mathfrak{F}^{t-2n\log^4 n}, \widetilde{\Phi}_1^{t-2n\log^4 n} \leqslant e^{\frac{1}{2}\cdot\log^4 n} \,\right] \\
&\leqslant e^{\frac{1}{2}\cdot\log^4 n} \cdot \left(1 - \frac{1}{n}\right)^{s-(t-2n\log^4 n)} + 2n \\
&\overset{(a)}{\leqslant} e^{\frac{1}{2}\cdot\log^4 n} \cdot \left(1 - \frac{1}{n}\right)^{n\log^4 n} + 2n \\
&\overset{(b)}{\leqslant} e^{\frac{1}{2}\cdot\log^4 n} \cdot e^{-\log^4 n} + 2n \leqslant 1 + 2n \leqslant 6n,
\end{aligned}$$

using in (a) that $s \geqslant t - n\log^4 n$ and in (b) that $e^u \geqslant 1 + u$ (for any u). Hence, the claim follows,

$$\mathbf{E}\left[\, \widetilde{\Phi}_1^s \,\right] = \mathbf{E}\left[\, \widetilde{\Phi}_1^s \,\middle|\, \mathcal{H}^{t-2n\log^4 n} \,\right] \cdot \mathbf{Pr}\left[\, \mathcal{H}^{t-2n\log^4 n} \,\right] + 0 \cdot \mathbf{Pr}\left[\, \neg\mathcal{H}^{t-2n\log^4 n} \,\right] \leqslant 6n. \qquad \square$$

The auxiliary process $\widetilde{\mathcal{P}}$. Let $\mathcal{P}$ be the process in the statement of Theorem 6.9, define $t_1 := t - n\log^4 n$ and the stopping time $\sigma := \inf\{s \geqslant t_1 : \{\Phi_1^s > 6n^{11}\} \cup \neg\mathcal{K}^s\}$. Now we define the auxiliary process $\widetilde{\mathcal{P}}_{t_1}$ such that

- in steps $s \in [0, \sigma)$ it follows the allocations of $\mathcal{P}$, and

- in steps $s \in [\sigma, \infty)$ it allocates to the (currently) least loaded bin, which is a bin with normalised load $\leqslant 1 \leqslant z$.

This way the process trivially satisfies $\mathcal{K}^s$ as it never allocates to a bin overloaded by more than 1, and therefore it also satisfies the drop inequalities (Eq. (6.1) and Eq. (6.2)) for any step $s \geqslant t_1$. Furthermore, starting with $\Phi_1^{t_1} \leqslant 6n^{11}$, for any $s \geqslant t_1$ it deterministically satisfies $\Phi_1^s \leqslant 6n^{12}$, since $\Phi_1^{\sigma} \leqslant 6n^{11} \cdot e^{\phi_1} \leqslant 6n^{12}$ (as $\phi_1 \leqslant (\log n)/6$) and for $s \geqslant \sigma$ the potential does not increase.

Further, starting from $\Phi_1^{t_1} \leqslant 6n^{11}$, we will also show that the other potential function Φ_2 w.h.p. becomes linear in *at least one* step in $[t - n\log^4 n, t]$, by using Lemma 6.7.

Lemma 6.8 (Recovery). *For any step $t \geqslant 0$ and the auxiliary process $\widetilde{\mathcal{P}}_{t-n\log^4 n}$, it holds that*

$$\mathbf{Pr}\left[\bigcup_{s\in[t-n\log^4 n,t]} \left\{\Phi_2^s \leqslant 6n\right\} \,\middle|\, \mathfrak{F}^{t-n\log^4 n}, \Phi_1^{t-n\log^4 n} \leqslant 6n^{11} \right] \geqslant 1 - n^{-6}.$$

Proof. Fix any step $t \geqslant 0$ and let $t_1 := t - n\log^4 n$ be the starting point of the analysis. Recall that the auxiliary process $\widetilde{\mathcal{P}}_{t_1}$ satisfies $\mathcal{K}^s$ for any step $s \geqslant t_1$. Further, for any $s \geqslant t_1$, we define the "killed" potential function

$$\widehat{\Phi}_2^s := \Phi_2^s \cdot \mathbf{1}_{\cap_{r\in[t_1,s]}\{\Phi_2^r > 6n\}}.$$

We will show that $\widehat{\Phi}_2$ drops in expectation by a multiplicative factor in every step. We start by showing for Φ_2 that if $\Phi_2^s > 6n$, then it drops in expectation by a multiplicative factor. By the second precondition Eq. (6.2) of Theorem 6.9,

$$\begin{aligned}\mathbf{E}\left[\, \Phi_2^{s+1} \,\middle|\, \Phi_1^{t_1} \leqslant 6n^{11}, \mathfrak{F}^s, \Phi_2^s > 6n \,\right] &= \mathbf{E}\left[\, \Phi_2^{s+1} \,\middle|\, \Phi_1^{t_1} \leqslant 6n^{11}, \mathfrak{F}^s, \mathcal{K}^s, \Phi_2^s > 6n \,\right] \\ &\leqslant \Phi_2^s \cdot \left(1 - \frac{1}{n}\right) + 2 \leqslant \Phi_2^s \cdot \left(1 - \frac{1}{2n}\right) - 6n \cdot \frac{1}{2n} + 2 \leqslant \Phi_2^s \cdot \left(1 - \frac{1}{2n}\right).\end{aligned}$$

Whenever the event $\{\Phi_2^s \leqslant 6n\}$ holds, it follows deterministically that $\left\{\widehat{\Phi}_2^s = \widehat{\Phi}_2^{s+1} = 0\right\}$. So, for the potential $\widehat{\Phi}_2$ we obtain the drop inequality (with one fewer condition),

$$\mathbf{E}\left[\, \widehat{\Phi}_2^{s+1} \,\middle|\, \Phi_1^{t_1} \leqslant 6n^{11}, \mathfrak{F}^s \,\right] \leqslant \widehat{\Phi}_2^s \cdot \left(1 - \frac{1}{2n}\right). \tag{6.12}$$

Using Lemma 6.5 (i), $\Phi_1^{t_1} \leqslant 6n^{11}$ implies that $\widehat{\Phi}_2^{t_1} \leqslant \Phi_2^{t_1} \leqslant n^{7/6}$ and so inductively applying Eq. (6.12) for $n\log^4 n$ steps starting at t_1,we have that

$$\mathbf{E}\left[\, \widehat{\Phi}_2^t \,\middle|\, \mathfrak{F}^{t_1}, \Phi_1^{t_1} \leqslant 6n^{11} \,\right] \leqslant \mathbf{E}\left[\, \widehat{\Phi}_2^t \,\middle|\, \mathfrak{F}^{t_1}, \widehat{\Phi}_2^{t_1} \leqslant n^{7/6} \,\right] \leqslant \widehat{\Phi}_2^{t_1} \cdot \left(1 - \frac{1}{2n}\right)^{n\log^4 n} \leqslant n^{7/6} \cdot e^{-\frac{1}{2}\log^4 n} \leqslant n^{-6},$$

for sufficiently large n, using that $e^u \geqslant 1 + u$ (for any u). Hence, by Markov's inequality,

$$\mathbf{Pr}\left[\, \widehat{\Phi}_2^t \leqslant 1 \,\middle|\, \mathfrak{F}^{t_1}, \Phi_1^{t_1} \leqslant 6n^{11} \,\right] \geqslant 1 - n^{-6}.$$

Since it deterministically holds that $\{\Phi_2^t \geqslant n\}$ for any step $t \geqslant 0$, it follows that if $\{\widehat{\Phi}_2^t \leqslant 1\}$ holds, then also $\{\widehat{\Phi}_2^t = 0\}$. So, we conclude that $\mathbf{1}_{\cap_{r\in[t_1,t]}\{\Phi_2^r > 6n\}} = 0$ holds, i.e.,

$$\mathbf{Pr}\left[\bigcup_{r\in[t_1,t]} \left\{\Phi_2^r \leqslant 6n\right\} \,\middle|\, \mathfrak{F}^{t_1}, \Phi_1^{t_1} \leqslant 6n^{11} \right] \geqslant 1 - n^{-6}. \qquad \square$$

6.3.3 Completing the proof of Theorem 6.9

The proof of Theorem 6.9 shares some of the ideas from the proof of Theorem 4.1. However, there we could more generously take a union bound over the entire time-interval to ensure that the potential is indeed small everywhere with high probability. In some applications (see Section 7.2), we need to apply Theorem 6.9 up to $\omega(1)$ times, which means that we cannot afford to lose a polynomial factor in the error probability. To overcome this, we will partition the time-interval into consecutive intervals of length $\lceil n/\log^2 n \rceil$. Then, we will prove that at the end of each such interval the potential is small w.h.p., and finally use a simple smoothness argument (Claim 6.6) to show that the potential is small w.h.p. in *all* steps.

Theorem 6.9 (Super-exponential potential concentration). *Consider any $\mathcal{P} = \textsc{Sequential}(q^s)$ process for which there exist super-exponential potential functions $\Phi_1 := \Phi_1(\phi_1, z)$ and $\Phi_2 := \Phi_2(\phi_2, z)$ with integer offset $z := z(n) > 0$ and smoothing parameters $\phi_1, \phi_2 \in (0, (\log n)/6]$ with $\phi_2 \leqslant \frac{\phi_1}{84}$, such that they satisfy for any step $s \geqslant 0$,*

$$\mathbf{E}\left[\, \Phi_1^{s+1} \,\middle|\, \mathfrak{F}^s, \mathcal{K}^s \,\right] \leqslant \Phi_1^s \cdot \left(1 - \frac{1}{n}\right) + 2, \tag{6.1}$$

and

$$\mathbf{E}\left[\, \Phi_2^{s+1} \,\middle|\, \mathfrak{F}^s, \mathcal{K}^s \,\right] \leqslant \Phi_2^s \cdot \left(1 - \frac{1}{n}\right) + 2, \tag{6.2}$$

where $\mathcal{K}^s := \mathcal{K}^s_{\phi_1}(q^s)$. Further, let $P \geqslant n^{-4}$. Then, for any steps $t \geqslant 0$ and $\tilde{t} \in [t, t + n\log^5 n]$, which satisfy

$$\mathbf{Pr}\left[\left\{ \mathrm{Gap}(t - 2n\log^4 n) \leqslant \log^2 n \right\} \cap \bigcap_{s \in [t - 2n\log^4 n, \tilde{t}]} \mathcal{K}^s \right] \geqslant 1 - P, \tag{6.3}$$

they must also satisfy

$$\mathbf{Pr}\left[\bigcap_{s \in [t, \tilde{t}]} \left\{ \Phi_2^s \leqslant 8n \right\} \right] \geqslant 1 - (\log^8 n) \cdot P.$$

We will start by proving the following lemma for the auxiliary process $\widetilde{\mathcal{P}}_{t - n\log^4 n}$.

Lemma 6.10. *Consider any step $t \geqslant 0$ and the auxiliary process $\widetilde{\mathcal{P}}_{t-n\log^4 n}$. For $P \geqslant n^{-4}$ as defined in Theorem 6.9, it holds that*

$$\mathbf{Pr}\left[\bigcap_{s \in [t, \tilde{t}]} \left\{ \Phi_2^s \leqslant 8n \right\} \,\middle|\, \Phi_1^{t - n\log^4 n} \leqslant 6n^{11} \right] \geqslant 1 - \frac{1}{4} \cdot (\log^8 n) \cdot P.$$

Proof. Our goal is to apply Azuma's inequality to Φ_2. However, there are two challenges: (i) we cannot afford to take the union bound over all poly(n) steps and (ii) Φ_2 is a super-martingale only when it is sufficiently large. To deal with (i) we will apply Azuma's inequality to sub-intervals of length at most $\lceil n/\log^2 n \rceil$ and then use a smoothness argument (Claim 6.6) to deduce that Φ_2 is small in the steps in between. For (ii) we will define $X^s := X^s(\Phi_2^s)$ in a way to ensure that it is super-martingale at every step and also satisfies the bounded difference inequality.

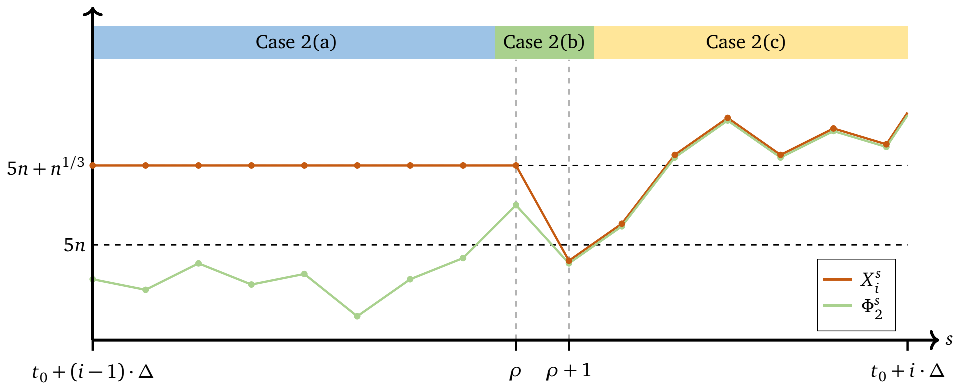


Figure 6.2: Visualisation of three cases in the proof of Claim 6.11 for interval $i \in [\ell]$. Note that as long as $\mathcal{Z}^{t_1}$ holds, then $X_i^s \geqslant \Phi_2^s$.

More specifically, consider an arbitrary step $r \in [t - n\log^4 n, t]$ and partition the interval $(r, \widetilde{t}]$ into

$$\mathcal{I}_1 := (r, r+\Delta],\ \mathcal{I}_2 = (r+\Delta, r+2\Delta],\ \ldots,\ \mathcal{I}_\ell := (r+(\ell-1)\Delta, \widetilde{t}],$$

where $\Delta := \lceil n/\log^2 n \rceil$ and $\ell := \left\lceil \frac{\widetilde{t}-r}{\Delta} \right\rceil \leqslant \left\lceil \frac{t+n\log^5 n-r}{\Delta} \right\rceil \leqslant 2\log^7 n$. In order to prove that Φ_2 is at most $8n$ in every step in $(r, \widetilde{t}]$, we will use our auxiliary lemmas (Section 6.3.2) and Azuma's super-martingale concentration inequality (Lemma B.10) to establish that Φ_2 is at most $7n$ at each of the steps $r+\Delta, r+2\Delta, \ldots, r+(\ell-1)\Delta, \widetilde{t}$. Finally, by using a smoothness argument (Claim 6.6), we will establish that Φ_2 is at most $8n$ at all steps in $(r, \widetilde{t}]$, which is the conclusion of the theorem.

We define the auxiliary random variable $X_i^{r+(i-1)\Delta} := \max\left\{\Phi_2^{r+(i-1)\Delta}, 5n+n^{1/3}\right\}$ for each interval $i \in [\ell]$, and for any $s \in (r+(i-1)\Delta, r+i\Delta]$,

$$X_i^s := \begin{cases} \Phi_2^s & \text{if there exists } u \in [r+(i-1)\Delta, s) \text{ such that } \Phi_2^u \geqslant 5n, \\ 5n+n^{1/3} & \text{otherwise.} \end{cases}$$

Note that whenever the first condition in the definition of X_i^s is satisfied, it remains satisfied until the end of the interval, i.e., until step $r + i\cdot\Delta$. Our next aim is to establish the preconditions of Azuma's inequality (Lemma B.10) for X_i^s.

For convenience, we define $t_1 := t - n\log^4 n$ and the event

$$\mathcal{Z}^{t_1} := \{\Phi_1^{t_1} \leqslant 6n^{11}\}.$$

Claim 6.11. *Fix any interval $\mathcal{I}_i$ for $i \in [\ell]$. Then for the random variables X_i^s, for any step $s \in (r+(i-1)\Delta, r+i\Delta]$, (i) it holds that*

$$\mathbf{E}\left[\, X_i^s \,\middle|\, \mathcal{Z}^{t_1}, \mathfrak{F}^{s-1} \,\right] \leqslant X_i^{s-1},$$

and (ii) it holds that

$$\left(\left|X_i^s - X_i^{s-1}\right| \,\middle|\, \mathcal{Z}^{t_1}, \mathfrak{F}^{s-1}\right) \leqslant 2n^{1/3}.$$

Proof of Claim 6.11. Recall that by the definition of the auxiliary process in Section 6.3.2, conditioning on $\mathcal{Z}^{t_1}$, it satisfies $\{\Phi_1^{s-1} \leqslant 6n^{12}\}$ and $\mathcal{K}^{s-1}$, for any $s \in (r+(i-1)\Delta, r+i\Delta]$, since $r+(i-1)\Delta \geqslant t_1$.

By precondition Eq. (6.2), when $\{\Phi_2^{s-1} \geqslant 4n\}$ also holds, we have that,

$$\mathbf{E}\left[\, \Phi_2^s \,\middle|\, \mathcal{Z}^{t_1}, \mathfrak{F}^{s-1}, \Phi_2^{s-1} \geqslant 4n \,\right] = \mathbf{E}\left[\, \Phi_2^s \,\middle|\, \mathcal{Z}^{t_1}, \mathfrak{F}^{s-1}, \mathcal{K}^{s-1}, \Phi_2^{s-1} \geqslant 4n \,\right]$$

$$\leqslant \Phi_2^{s-1} \cdot \left(1 - \frac{1}{n}\right) + 2 \leqslant \Phi_2^{s-1} - 4n \cdot \frac{1}{n} + 2 \leqslant \Phi_2^{s-1}. \tag{6.13}$$

Further, by Lemma 6.5 (ii), we also have that

$$\left(\left| \Phi_2^s - \Phi_2^{s-1} \right| \;\middle|\; \mathcal{Z}^{t_1}, \mathfrak{F}^{s-1} \right) \leqslant n^{1/3}. \tag{6.14}$$

Case 1 $[\Phi_2^{r+(i-1)\Delta} \geqslant 5n + n^{1/3}]$**:** In this case $X_i^{s-1} = \Phi_2^{s-1}$ for all $s \in (r + (i-1)\Delta, r + i\Delta]$. By Claim 6.6, we also have $\Phi_2^{s-1} \geqslant 0.99 \cdot (5n + n^{1/3}) \geqslant 4n$ (as $\Delta \leqslant \lceil n/\log^2 n \rceil$) and the two statements follow by Eq. (6.13) and Eq. (6.14).

Case 2 $[\Phi_2^{r+(i-1)\Delta} < 5n + n^{1/3}]$**:** Let $\rho := \inf\{u \geqslant r + (i-1)\Delta \colon \Phi_2^u \geqslant 5n\}$. We consider the following three sub-cases (see Fig. 6.2):

- **Case 2(a)** $[s - 1 < \rho]$**:** Here $X_i^s = X_i^{s-1} = 5n + n^{1/3}$, so the two statements hold trivially.
- **Case 2(b)** $[s - 1 = \rho]$**:** We will first establish that

$$5n \leqslant \Phi_2^{s-1} \leqslant 5n + n^{1/3}. \tag{6.15}$$

  The lower bound $\Phi_2^{s-1} \geqslant 5n$ follows by definition of ρ. For the upper bound, we consider the following two cases. If $s - 1 = r + (i-1)\Delta$, then this follows by the assumption for Case 2. Otherwise, we have that $\Phi_2^{s-1} \geqslant 5n$ and $\Phi_2^{s-2} < 5n$. By Eq. (6.14), we obtain that $\Phi_2^{s-1} \leqslant \Phi_2^{s-2} + n^{1/3} < 5n + n^{1/3}$.

  Next, by definition, $X_i^{s-1} = 5n + n^{1/3}$ and $X_i^s = \Phi_2^s$, so by Eq. (6.13),

$$\mathbf{E}\left[X_i^s \;\middle|\; \mathcal{Z}^{t_1}, \mathfrak{F}^{s-1} \right] = \mathbf{E}\left[\Phi_2^s \;\middle|\; \mathcal{Z}^{t_1}, \mathfrak{F}^{s-1} \right] \leqslant \Phi_2^{s-1} < X_i^{s-1},$$

  which establishes the first statement. For the second statement, we have

$$\left| X_i^s - X_i^{s-1} \right| = \left| \Phi_2^s - 5n - n^{1/3} \right| \leqslant \left| \Phi_2^{s-1} - 5n - n^{1/3} \right| + \left| \Phi_2^s - \Phi_2^{s-1} \right| \leqslant 2n^{1/3}.$$

  using in the last inequality that Eq. (6.14) and Eq. (6.15).
- **Case 2(c)** $[s - 1 > \rho]$**:** Here, $X_i^{s-1} = \Phi_2^{s-1}$ and $X_i^s = \Phi_2^s$. Since $\Phi_2^\rho \geqslant 5n$, by Claim 6.6 (as $s - \rho \leqslant \lceil n/\log^2 n \rceil$), we also have that

$$\Phi_2^{s-1} \geqslant 0.99 \cdot \Phi_2^\rho \geqslant 0.99 \cdot 5n \geqslant 4n,$$

  and thus by Eq. (6.13), the first statement follows. The second statement follows by Eq. (6.14). ◊

Now we return to the proof of Lemma 6.10. By Claim 6.11, we have verified that X_i^s satisfies the preconditions of Azuma's inequality for any filtration $\mathfrak{F}^{s-1}$ where $\mathcal{Z}^{t_1}$ holds. So, applying Lemma B.10 for $\lambda = \frac{n}{2\log^7 n}$, $N \leqslant \Delta$ and $D = 2n^{1/3}$, we get for any $i \in [\ell]$,

$$\mathbf{Pr}\left[X_i^{r+i\Delta} \geqslant X_i^{r+(i-1)\Delta} + \lambda \;\middle|\; \mathcal{Z}^{t_1}, \mathfrak{F}^r \right] \leqslant \exp\left(-\frac{n^2/(4\log^{14} n)}{2 \cdot \Delta \cdot (4n^{2/3})} \right) \leqslant P,$$

since $\Delta \leqslant \lceil n/\log^2 n \rceil$ and $P \geqslant n^{-4}$. Taking the union bound over the at most $2\log^7 n$ intervals $i \in [\ell]$, it follows that

$$\mathbf{Pr}\left[\bigcup_{i \in [\ell]} \left\{ X_i^{r+i\Delta} \geqslant X_1^r + i \cdot \frac{n}{2\log^7 n} \right\} \;\middle|\; \mathcal{Z}^{t_1}, \mathfrak{F}^r \right] \leqslant (2\log^7 n) \cdot P \leqslant \frac{1}{8} \cdot (\log^8 n) \cdot P.$$

Next, conditional on $\mathcal{Z}^{t_1}, \mathfrak{F}^r, \{\Phi_2^r \leqslant 6n\}$, we have the following chain of inclusions:

$$\begin{aligned}\bigcap_{i\in[\ell]}\left\{X_i^{r+i\Delta} \leqslant X_1^r + i\cdot\frac{n}{2\log^7 n}\right\} &\overset{(a)}{\subseteq} \bigcap_{i\in[\ell]}\left\{X_i^{r+i\Delta} \leqslant 6n+n\right\} \\ &\overset{(b)}{=} \bigcap_{i\in[\ell]}\left\{\Phi_2^{r+i\Delta} \leqslant 7n\right\} \\ &\overset{(c)}{\subseteq} \bigcap_{s\in[r,\widetilde{t}]}\left\{\Phi_2^s \leqslant \frac{7}{0.99}\cdot n\right\} \\ &\overset{(d)}{\subseteq} \bigcap_{s\in[t,\widetilde{t}]}\left\{\Phi_2^s \leqslant 8n\right\},\end{aligned}$$

where (a) holds since $i \leqslant \ell \leqslant 2\log^7 n$ and $X_1^r \leqslant \max\left\{\Phi_2^r, 5n+n^{1/3}\right\}$, (b) holds since $\Phi_2^{r+i\Delta} \leqslant X_i^{r+i\Delta}$ (which follows from the definition of X_i and Claim 6.11 (ii)), (c) holds by applying the smoothness argument of Claim 6.6 to each interval $i \in [\ell]$, as the length of each interval is at most $\lceil n/\log^2 n\rceil$ and (d) holds since $8 \geqslant \frac{7}{0.99}$ and $r \leqslant t$. This implies that

$$\mathbf{Pr}\left[\bigcap_{s\in[t,\widetilde{t}]}\left\{\Phi_2^s \leqslant 8n\right\} \;\middle|\; \mathcal{Z}^{t_1}, \mathfrak{F}^r, \Phi_2^r \leqslant 6n\right] \geqslant 1 - \frac{1}{8}\cdot(\log^8 n)\cdot P. \tag{6.16}$$

Next, define $\tau := \inf\{r \geqslant t_1 \colon \Phi_2^r \leqslant 6n\}$. By Lemma 6.8,

$$\mathbf{Pr}\left[\,\tau \leqslant t \;\middle|\; \mathcal{Z}^{t_1}\,\right] \geqslant \mathbf{Pr}\left[\,\tau \leqslant t \;\middle|\; \mathfrak{F}^{t_1}, \mathcal{Z}^{t_1}\,\right] \geqslant 1 - n^{-6}. \tag{6.17}$$

We get the conclusion for the $\widetilde{\mathcal{P}}_{t_1}$ process, by combining this with Eq. (6.16) and Eq. (6.17),

$$\begin{aligned}\mathbf{Pr}\left[\bigcap_{s\in[t,\widetilde{t}]}\left\{\Phi_2^s \leqslant 8n\right\} \;\middle|\; \mathcal{Z}^{t_1}\right] &\geqslant \sum_{r=t_1}^{t}\mathbf{Pr}\left[\bigcap_{s\in[t,\widetilde{t}]}\left\{\Phi_2^s \leqslant 8n\right\} \;\middle|\; \mathcal{Z}^{t_1}, \mathfrak{F}^r, \Phi_2^r \leqslant 6n\right]\cdot\mathbf{Pr}\left[\,\tau = r \mid \mathcal{Z}^{t_1}\,\right] \\ &\overset{(6.16)}{\geqslant} \left(1 - \frac{1}{8}\cdot(\log^8 n)\cdot P\right)\cdot\mathbf{Pr}\left[\,\tau \leqslant t \;\middle|\; \mathcal{Z}^{t_1}\,\right] \\ &\overset{(6.17)}{\geqslant} \left(1 - \frac{1}{8}\cdot(\log^8 n)\cdot P\right)\cdot\left(1 - n^{-6}\right) \geqslant 1 - \frac{1}{4}\cdot(\log^8 n)\cdot P. \qquad\square\end{aligned}$$

We now return to the proof of Theorem 6.9 for the original process $\mathcal{P}$.

Proof of Theorem 6.9. Let $t_1 := t - n\log^4 n$. We start by showing that the processes $\mathcal{P}$ and $\widetilde{\mathcal{P}}_{t_1}$ agree with high probability in the interval $[t_1, \widetilde{t}]$ (and so at every step $s \in [t_1, \widetilde{t}]$ we have $\Phi_2^s = \Phi^s_{\widetilde{\mathcal{P}}_{t_1},2}$),

$$\begin{aligned}\mathbf{Pr}\left[\bigcap_{s\in[t_1,\widetilde{t}]}\left\{y^s = y^s_{\widetilde{\mathcal{P}}_{t_1}}\right\}\right] &\geqslant \mathbf{Pr}\left[\bigcap_{s\in[t-2n\log^4 n,\widetilde{t}]}\left\{\Phi_1^s \leqslant 6n^{11}\right\} \cap \bigcap_{s\in[t-2n\log^4 n,\widetilde{t}]}\mathcal{K}^s\right] \\ &\overset{(a)}{\geqslant} \mathbf{Pr}\left[\bigcap_{s\in[t-2n\log^4 n,\widetilde{t}]}\left\{\widetilde{\Phi}_1^s \leqslant 6n^{11}\right\} \cap \mathcal{H}^{t-2n\log^4 n} \cap \bigcap_{s\in[t-2n\log^4 n,\widetilde{t}]}\mathcal{K}^s\right] \\ &\overset{(b)}{\geqslant} 1 - n^2\cdot n^{-11} - P \overset{(c)}{\geqslant} 1 - 2P,\end{aligned} \tag{6.18}$$

using in (a) the definition of $\widetilde{\Phi}_1$ in Eq. (6.10), in (b) Lemma 6.7, Markov's inequality and union bound over $\widetilde{t} - t_1 \leqslant n^2$ steps and precondition Eq. (6.3) and in (c) that $P \geqslant n^{-4}$. Note that this also implies that

$$\mathbf{Pr}\left[\, \Phi_1^{t-n\log^4 n} \leqslant 6n^{11} \,\right] \geqslant 1 - 2P. \tag{6.19}$$

By Lemma 6.10, for the auxiliary process $\widetilde{\mathcal{P}}_{t_1}$, we have that

$$\mathbf{Pr}\left[\, \bigcap_{s\in[t,\widetilde{t}]} \left\{ \Phi^s_{\widetilde{\mathcal{P}}_{t_1},2} \leqslant 8n \right\} \,\middle|\, \Phi_1^{t-n\log^4 n} \leqslant 6n^{11} \,\right] \geqslant 1 - \frac{1}{4} \cdot (\log^8 n) \cdot P.$$

By combining with Eq. (6.19) (since by definition $\mathcal{P}$ and $\widetilde{\mathcal{P}}_{t_1}$ agree in steps $s \leqslant t_1$) we get,

$$\mathbf{Pr}\left[\, \bigcap_{s\in[t,\widetilde{t}]} \left\{ \Phi^s_{\widetilde{\mathcal{P}}_{t_1},2} \leqslant 8n \right\} \,\right] \geqslant \left(1 - \frac{1}{4} \cdot (\log^8 n) \cdot P\right) \cdot (1 - 2P) \geqslant 1 - \frac{3}{4} \cdot (\log^8 n) \cdot P.$$

Finally, using $\mathbf{Pr}[\, A \cap B \,] \geqslant \mathbf{Pr}[\, A \,] - \mathbf{Pr}[\, \neg B \,]$ with Eq. (6.18), we conclude that $\mathcal{P}$ and $\widetilde{\mathcal{P}}_{t_1}$ agree in all steps of the interval $[t_1, \widetilde{t}]$ and so

$$\mathbf{Pr}\left[\, \bigcap_{s\in[t,\widetilde{t}]} \left\{ \Phi_2^s \leqslant 8n \right\} \,\right] \geqslant 1 - \frac{3}{4} \cdot (\log^8 n) \cdot P - 2P \geqslant 1 - (\log^8 n) \cdot P. \qquad \square$$

Applications

In this chapter, we derive gaps for several processes and settings using the tools we developed in the previous chapters.

As a warm-up, in Section 7.1, we show two ways to obtain the $\mathcal{O}(\sqrt{\log n})$ gap for the 2-Dense-Quantile process with probability at least $1-o(1)$ in order to demonstrate the advantage of using the concentration bounds of Chapters 4 and 6. In Section 7.1.1, we obtain the $\mathcal{O}(\sqrt{\log n})$ gap with probability at least $1-\exp(-\sqrt{\log n})$, without using the concentration bounds. In Section 7.1.2, we make use of these concentration bounds to obtain the same $\mathcal{O}(\sqrt{\log n})$ asymptotic bound with probability at least $1-n^{-3}$. Usually, the smaller the gap we want to prove, the larger the difference between the probability guarantee between these two techniques. Therefore, using the concentration bounds is crucial for several results in this chapter.

In Section 7.2, we make repeated use of the super-exponential potential concentration theorem (Theorem 6.9) through a form of *layered induction* to analyse a number of processes including the k-Dense-Quantile, k-Dense-Threshold and Quantile(δ^*) processes. In Section 7.3, we generalise the analysis of k-Dense-Quantile to all processes in the k-Relaxed-Quantile family, which implies bounds for the $(1+\beta)$-process with β close to 1. Then, in Section 7.4, we analyse the g-Adv setting for the Two-Choice process for any $g \leqslant \log n$ and establish upper bounds for a number of other settings including b-Batched, τ-Delay and g-Myopic-Comp for the Two-Choice process. In Section 7.5, we prove tight bounds for a large family of processes in the b-Batched setting. In Section 7.6, we collect results for Two-Choice in the Graphical setting. Finally, in Section 7.7, we outline how layered induction can be used to obtain tight bounds for the Memory process.

7.1 Two analyses for the 2-Dense-Quantile process

In this section, we will show two ways to obtain the $\mathcal{O}(\sqrt{\log n})$ bound on the gap of the 2-Dense-Quantile process with different probability guarantees: (i) in Section 7.1.1 without using concentration bounds and (ii) in Section 7.1.2 with concentration bounds, leading to stronger probability guarantees. Recall that this process is the Quantile(δ_1, δ_2) process where δ_1, δ_2 are the rounded up quantiles of $\widetilde{\delta}_1 := e^{-\frac{1}{4}\cdot\sqrt{\log n}}$ and $\widetilde{\delta}_2 := \frac{1}{2}$ respectively, so that $\widetilde{\delta}_1 \leqslant \delta_1 \leqslant 2\widetilde{\delta}_1$ and $\widetilde{\delta}_2 \leqslant \delta_2$.

7.1.1 Weak probability bounds for 2-Dense-Quantile

In this section, we will present an analysis for the 2-Dense-Quantile process for obtaining the $\mathcal{O}(\sqrt{\log n})$ gap with probability at least $1-e^{-\sqrt{\log n}}$, without using the concentration bounds. In fact, this approach extends to the k-Dense-Quantile process for any constant $k \geqslant 1$ (with weaker probability guarantees as k increases).

We start with the fact that the hyperbolic cosine potential is linear in expectation. This follows because the 2-Dense-Quantile process is majorised by the Median-Quantile process, which by Lemma 3.5 with $\delta = 1/3$, $\epsilon = 1/3$ and $C = 2$ satisfies the preconditions of Theorem 3.2 for any smoothing parameter $\gamma \in \big(0, \frac{1}{8\cdot 3^2\cdot 2^2}\big]$. Therefore, we obtain the following corollary.

Corollary 7.1. *Consider the* 2-DENSE-QUANTILE *process and the hyperbolic cosine potential* $\Gamma := \Gamma(\gamma)$ *for any* $\gamma \in \left(0, \frac{1}{8 \cdot 3^2 \cdot 2^2}\right]$. *Then, there exists a constant* $c > 0$, *such that for any step* $t \geqslant 0$,

$$\mathbf{E}\left[\, \Gamma^t \,\right] \leqslant cn.$$

Next, the main idea is to bound the number of bins with normalised load at least v with probability at least $1 - e^{-\Theta(v)}$, using Markov's inequality. Note that as v gets smaller, the weaker the probability guarantee that we obtain becomes weaker. As we cannot afford to take the union bound over all $n \cdot \mathrm{polylog}(n)$ steps, we apply this bound to $\mathrm{polylog}(n)$ steps at distance n apart and then use the following smoothness claim (where for our case $T = n$).

Claim 7.2 (cf. Lemma 5.28). *Consider the potential* $\Gamma := \Gamma(\gamma)$, *any* $c > 0$ *and any step* $s \geqslant 0$ *for which there exists* $T \geqslant 0$ *such that for some* $s_0 \in [s - T, s]$ *with* $\Gamma^{s_0} \leqslant cn$ *and* $s_1 \in [s, s+T]$ *with* $\Gamma^{s_1} \leqslant cn$. *Then,* $\Gamma^s \leqslant e^{\gamma \frac{T}{n}} \cdot 2cn$.

Proof. From step s to s_1, the load of any bin can decrease by at most $\frac{T}{n}$. Hence, we have that for any bin $i \in [n]$,

$$y_i^s \leqslant y_i^{s_1} + \frac{T}{n}.$$

Similarly, from s_0 to s the load of any bin can decrease by at most $\frac{T}{n}$, so

$$y_i^{s_0} \leqslant y_i^s + \frac{T}{n} \quad \Rightarrow \quad -y_i^s \leqslant -y_i^{s_0} + \frac{T}{n}.$$

Combining the two, we get that

$$\Gamma_i^s = e^{\gamma y_i^s} + e^{-\gamma y_i^s} \leqslant e^{\gamma \frac{T}{n}} \cdot \left(e^{-\gamma y_i^{s_0}} + e^{\gamma y_i^{s_1}}\right).$$

By aggregating over all bins $i \in [n]$, we get that

$$\Gamma^s = \sum_{i=1}^n \Gamma_i^s \leqslant \sum_{i=1}^n e^{\gamma \frac{T}{n}} \cdot \left(e^{-\gamma y_i^{s_0}} + e^{\gamma y_i^{s_1}}\right) \leqslant e^{\gamma \frac{T}{n}} \cdot (\Gamma^{s_0} + \Gamma^{s_1}) \leqslant e^{\gamma \frac{T}{n}} \cdot 2cn. \qquad \square$$

Now, we are ready to prove the upper bound on the gap.

Theorem 7.3. *Consider the* 2-DENSE-QUANTILE *process. Then, there exists a constant* $\kappa > 0$, *such that for any step* $m \geqslant 0$,

$$\mathbf{Pr}\left[\, \mathrm{Gap}(m) \leqslant \kappa \cdot \sqrt{\log n} \,\right] \geqslant 1 - e^{-\sqrt{\log n}}.$$

Proof. Consider the potential $\Gamma := \Gamma(\gamma)$ for the smoothing parameter $\gamma := \frac{1}{8 \cdot 3^2 \cdot 2^2}$. By Corollary 7.1 and Markov's inequality, for any step $t \geqslant 0$,

$$\mathbf{Pr}\left[\, \Gamma^t \leqslant cn \cdot e^{2\sqrt{\log n}} \,\right] \geqslant 1 - e^{-2\sqrt{\log n}}.$$

Consider the steps

$$t_0 := m - n \log^2 n, t_1 := m - n(\log^2 n - 1), \ldots, t_{\log^2 n} := m.$$

By the union bound over these at most $\log^2 n + 1 \leqslant 2 \cdot \log^2 n$ steps, we have that

$$\mathbf{Pr}\left[\, \bigcap_{j \in \{0\} \cup [\log^2 n]} \left\{ \Gamma^{t_j} \leqslant cn \cdot e^{2\sqrt{\log n}} \right\} \,\right] \geqslant 1 - 2 \cdot e^{-2\sqrt{\log n}} \cdot \log^2 n.$$

By smoothness, Claim 7.2 with $T = n$, we have that for all steps in between, the potential is at most $2e^{\gamma} cn \cdot e^{2\sqrt{\log n}} \leqslant 4cn \cdot e^{2\sqrt{\log n}}$ (since $\gamma < 1/2$), so

$$\mathbf{Pr}\left[\bigcap_{t \in [m - n\log^2 n, m]} \left\{ \Gamma^t \leqslant 4cn \cdot e^{2\sqrt{\log n}} \right\} \right] \geqslant 1 - 2 \cdot e^{-2\sqrt{\log n}} \cdot \log^2 n. \tag{7.1}$$

Consider the super-exponential potential $\Phi := \Phi(\phi, z)$ with smoothing parameter $\phi := \gamma \cdot \sqrt{\log n}$ and offset $z := \left\lceil \frac{5}{\gamma} \cdot \sqrt{\log n} \right\rceil$. Recall the condition

$$\widetilde{\mathcal{K}}^t = \left\{ \forall i \in [n] \colon \; y_i^t \geqslant z - 1 \;\;\Rightarrow\;\; q_i^t \leqslant \frac{1}{n} \cdot e^{-\phi} \right\},$$

under which, by Lemma 6.2, we have that

$$\mathbf{E}\left[\Phi^{t+1} \,\middle|\, \mathfrak{F}^t, \widetilde{\mathcal{K}}^t \right] \leqslant \Phi^t \cdot \left(1 - \frac{1}{n}\right) + 2.$$

We will now show that when $\left\{\Gamma^t \leqslant 4cn \cdot e^{2\sqrt{\log n}}\right\}$ holds, then the event $\widetilde{\mathcal{K}}^t$ also holds. The number of bins with load at least $z - 1$ is at most

$$4cn \cdot e^{2\sqrt{\log n}} \cdot e^{-\gamma \cdot (z-1)} \leqslant 4cn \cdot e^{2\sqrt{\log n}} \cdot e^{-\gamma \cdot \frac{5}{\gamma} \cdot \sqrt{\log n}} \leqslant n \cdot e^{-\sqrt{\log n}} \leqslant n\widetilde{\delta}_1 \leqslant n\delta_1.$$

Hence, all bins $i \in [n]$ with load at least $y_i^t \geqslant z - 1$ have $\mathrm{Rank}^t(i) \leqslant n\delta_1$ and so

$$q_i^t = \frac{\delta_1}{n} \leqslant \frac{2\widetilde{\delta}_1}{n} = \frac{2}{n} \cdot e^{-\frac{1}{4} \cdot \sqrt{\log n}} \leqslant \frac{1}{n} \cdot e^{-\frac{1}{8} \cdot \sqrt{\log n}} \leqslant \frac{1}{n} \cdot e^{-\phi},$$

using in the last inequality that $\gamma \leqslant 1/8$. Therefore, by Eq. (7.1), we have that

$$\mathbf{Pr}\left[\bigcap_{t \in [m - n\log^2 n, m]} \widetilde{\mathcal{K}}^t \right] \geqslant 1 - 2 \cdot e^{-2\sqrt{\log n}} \cdot \log^2 n. \tag{7.2}$$

Further, when $\left\{\Gamma^{t_0} \leqslant 4cn \cdot e^{2\sqrt{\log n}}\right\}$ holds, we also have that $\mathrm{Gap}(t_0) \leqslant K \cdot \log n$ for some constant $K > 0$. So for sufficiently large n

$$\Phi^{t_0} \leqslant n \cdot e^{\gamma \sqrt{\log n} \cdot K \log n} \leqslant e^{\frac{1}{2} \cdot \log^2 n}.$$

Hence, by Eq. (7.1) and Eq. (7.2),

$$\mathbf{Pr}\left[\{\mathrm{Gap}(t_0) \leqslant K \cdot \log n\} \cap \bigcap_{s \in [t_0, m]} \widetilde{\mathcal{K}}^s \right] \geqslant 1 - 2 \cdot e^{-2\sqrt{\log n}} \cdot \log^2 n. \tag{7.3}$$

By Lemma 6.2, when the event $\widetilde{\mathcal{K}}^t$ holds, we have that

$$\mathbf{E}\left[\Phi^{t+1} \,\middle|\, \mathfrak{F}^t, \widetilde{\mathcal{K}}^t \right] \leqslant \Phi^t \cdot \left(1 - \frac{1}{n}\right) + 2.$$

We define for any step $t \geqslant t_0$,

$$\widetilde{\Phi}^t := \Phi^t \cdot \mathbf{1}_{\{\mathrm{Gap}(t_0) \leqslant K \cdot \log n\} \cap \bigcap_{s \in [t_0, t]} \widetilde{\mathcal{K}}^s}.$$

When $\neg\widetilde{\mathcal{K}}^t$ holds, then $\{\widetilde{\Phi}^t = 0\}$ and $\{\widetilde{\Phi}^{t+1} = 0\}$, and so $\widetilde{\Phi}$ satisfies unconditionally the drop inequality

$$\mathbf{E}\left[\, \widetilde{\Phi}^{t+1} \,\middle|\, \mathfrak{F}^t \,\right] \leqslant \widetilde{\Phi}^t \cdot \left(1 - \frac{1}{n}\right) + 2.$$

Also, we may assume that $\mathrm{Gap}(t_0) \leqslant K \cdot \log n$ and so $\widetilde{\Phi}^{t_0} \leqslant e^{\frac{1}{2}\log^2 n}$, otherwise $\widetilde{\Phi}^t = 0$ for any step $t \geqslant t_0$. Hence, by Lemma B.1 with $a = 1 - \frac{1}{n}$ and $b = 2$, we have that

$$\begin{aligned}
\mathbf{E}\left[\, \widetilde{\Phi}^m \,\right] &= \mathbf{E}\left[\, \widetilde{\Phi}^m \,\middle|\, \mathfrak{F}^{t_0}, \widetilde{\Phi}^{t_0} \leqslant e^{\frac{1}{2}\log^2 n} \,\right] \\
&\leqslant \left(1 - \frac{1}{n}\right)^{m-t_0} \cdot e^{\frac{1}{2}\log^2 n} + 2n \\
&\leqslant \left(1 - \frac{1}{n}\right)^{n\log^2 n} \cdot e^{\frac{1}{2}\log^2 n} + 2n \leqslant e^{-\log^2 n} \cdot e^{\frac{1}{2}\log^2 n} + 2n \leqslant 4n.
\end{aligned}$$

Hence, by Markov's inequality we have that

$$\mathbf{Pr}\left[\, \widetilde{\Phi}^m \leqslant 4n^2 \,\right] \geqslant 1 - n^{-1}.$$

Combining with Eq. (7.3) using the union bound, we get that

$$\begin{aligned}
\mathbf{Pr}\left[\, \Phi^m \leqslant 4n^2 \,\right] &\geqslant \mathbf{Pr}\left[\, \{\widetilde{\Phi}^m \leqslant 4n^2\} \cap \{\mathrm{Gap}(t_0) \leqslant K \cdot \log n\} \cap \bigcap_{s \in [t_0, m]} \widetilde{\mathcal{K}}^s \,\right] \\
&\geqslant 1 - n^{-1} - 2 \cdot e^{-2\sqrt{\log n}} \cdot \log^2 n \geqslant 1 - e^{-\sqrt{\log n}}.
\end{aligned}$$

When $\left\{\Phi^m \leqslant 4n^2\right\}$ holds, we have that

$$\mathrm{Gap}(m) \leqslant z + \frac{\log(4n^2)}{\gamma \cdot \sqrt{\log n}} \leqslant \left\lceil \frac{5}{\gamma} \cdot \sqrt{\log n} \right\rceil + \frac{3}{\gamma} \cdot \sqrt{\log n}.$$

Hence, for the constant $\kappa := 9/\gamma$, we get that

$$\mathbf{Pr}\left[\, \mathrm{Gap}(m) \leqslant \kappa \cdot \sqrt{\log n} \,\right] \geqslant 1 - e^{-\sqrt{\log n}}. \qquad \square$$

7.1.2 Strong probability bounds for 2-Dense-Quantile

Again, we start with the fact that the hyperbolic cosine potential is linear in expectation. This follows because the 2-Dense-Quantile process is majorised by the Median-Quantile process, which by Lemma 3.5 with $\delta = 1/3$, $\epsilon = 1/3$ and $C = 2$ satisfies the preconditions of Theorem 3.2 for any smoothing parameter $\gamma \in \left(0, \frac{1}{8 \cdot 3^2 \cdot 2^2}\right]$.

Corollary 7.4. *Consider the 2-Dense-Quantile process and the hyperbolic cosine potential $\Gamma := \Gamma(\gamma)$ for any $\gamma \in \left(0, \frac{1}{8 \cdot 3^2 \cdot 2^2}\right]$. Then, there exist constants $c_1, c_2 > 0$, such that for any step $t \geqslant 0$,*

$$\mathbf{E}\left[\, \Gamma^{t+1} \,\middle|\, \mathfrak{F}^t \,\right] \leqslant \Gamma^t \cdot \left(1 - \frac{c_1 \gamma}{n}\right) + c_2 \gamma.$$

Then, by Theorem 4.1 with $\gamma_1 := \frac{1}{8 \cdot 3^2 \cdot 2^2}$, $\gamma_2 := \frac{\gamma_1}{84}$ and $\kappa = 6$, we get the following corollary.

Corollary 7.5. *Consider the* 2-DENSE-QUANTILE *process and the hyperbolic cosine potential* $\Gamma_2 := \Gamma_2(\gamma_2)$*. Then, there exists a constant* $c > 0$ *such that for any step* $t \geqslant 0$

$$\mathbf{Pr}\left[\, \Gamma_2^t \leqslant 3cn \,\right] \geqslant 1 - n^{-6}.$$

We now define the super-exponential potential functions $\Phi_1 := \Phi_1(\phi_1, z)$ and $\Phi_2 := \Phi_2(\phi_2, z)$ with $\phi_1 := \gamma_1\sqrt{\log n}$, $\phi_2 := \gamma_2\sqrt{\log n}$ and offset $z := \lceil \frac{3}{\gamma_2} \cdot \sqrt{\log n} \rceil$, so that $z - 1 \geqslant \frac{2}{\gamma_2} \cdot \sqrt{\log n}$. Also, let $\widetilde{\mathcal{K}}^t$ be the event associated with their drop in Theorem 6.9:

$$\widetilde{\mathcal{K}}^t := \widetilde{\mathcal{K}}^t_{\phi_1} := \left\{ \forall i \in [n] \colon\ y_i^t \geqslant z - 1 \ \ \Rightarrow \ \ q_i^t \leqslant \frac{1}{n} \cdot e^{-\gamma_1 \cdot \sqrt{\log n}} \right\} \subseteq \widetilde{\mathcal{K}}^t_{\phi_2}.$$

Lemma 7.6. *For any step* $t \geqslant 0$ *such that* $\Gamma_2^t \leqslant 3cn$*, we have that* $\widetilde{\mathcal{K}}^t$ *holds.*

Proof. Consider an arbitrary step t with $\Gamma_2^t \leqslant 3cn$. Then, in this step the number of bins with normalised load at least $z - 1$ is at most

$$3cn \cdot e^{-\gamma_2 \cdot (z-1)} \leqslant 3cn \cdot e^{-\gamma_2 \cdot \frac{2}{\gamma_2} \cdot \sqrt{\log n}} \overset{(a)}{\leqslant} n \cdot e^{-\sqrt{\log n}} \leqslant n\widetilde{\delta}_1 \leqslant n\delta_1,$$

using in (a) that $\sqrt{\log n} \geqslant \log(3c)$. So any bin $i \in [n]$ with $y_i^t \geqslant z - 1$ has $\mathrm{Rank}^t(i) \leqslant n\delta_1$ and hence

$$q_i^t = \frac{\delta_1}{n} \leqslant \frac{2\widetilde{\delta}_1}{n} = \frac{2}{n} \cdot e^{-\frac{1}{4} \cdot \sqrt{\log n}} \overset{(a)}{\leqslant} \frac{1}{n} \cdot e^{-\frac{1}{8} \cdot \sqrt{\log n}} \overset{(b)}{\leqslant} \frac{1}{n} \cdot e^{-\gamma_1 \cdot \sqrt{\log n}},$$

using in (a) that $\frac{1}{8} \cdot \sqrt{\log n} \geqslant \log 2$ and in (b) that $\gamma_1 \leqslant \frac{1}{8}$. $\square$

Now we are ready to upper bound the gap of 2-DENSE-QUANTILE at an arbitrary step m. Note that compared to Theorem 7.3, the probability bound is stronger.

Theorem 7.7. *Consider the* 2-DENSE-QUANTILE *process. Then, there exists a constant* $\kappa > 0$*, such that for any step* $m \geqslant 0$*,*

$$\mathbf{Pr}\left[\, \mathrm{Gap}(m) \leqslant \kappa \cdot \sqrt{\log n} \,\right] \geqslant 1 - n^{-3}.$$

Proof. By Corollary 7.5, taking the union bound over $2n\log^4 n \leqslant n^2$ steps, we have that

$$\mathbf{Pr}\left[\bigcap_{t \in [m - 2n\log^4 n, m]} \left\{ \Gamma_2^t \leqslant 3cn \right\} \right] \geqslant 1 - n^{-4}.$$

By Lemma 7.6, $\Gamma_2^t \leqslant 3cn$ implies that $\widetilde{\mathcal{K}}^t$ holds and $\mathrm{Gap}(t) \leqslant \log^2 n$, hence

$$\mathbf{Pr}\left[\left\{ \mathrm{Gap}(m - 2n\log^4 n) \leqslant \log^2 n \right\} \cap \bigcap_{t \in [m - 2n\log^4 n, m]} \widetilde{\mathcal{K}}^t \right] \geqslant 1 - n^{-4},$$

and so by Theorem 6.9 for $\widetilde{t} := m$, we get that

$$\mathbf{Pr}\left[\, \Phi_2^m \leqslant 8n \,\right] \geqslant 1 - \frac{(\log n)^8}{n^4} \geqslant 1 - n^{-3}.$$

Finally, when $\left\{ \Phi_2^m \leqslant 8n \right\}$ holds, we have that

$$\mathrm{Gap}(m) \leqslant z + \frac{\log(8n)}{\gamma_2 \cdot \sqrt{\log n}} \leqslant \left\lceil \frac{3}{\gamma_2} \cdot \sqrt{\log n} \right\rceil + \frac{2}{\gamma_2} \cdot \sqrt{\log n},$$

and for $\kappa := 6/\gamma_2$, we conclude that

$$\mathbf{Pr}\left[\, \mathrm{Gap}(m) \leqslant \kappa \cdot \sqrt{\log n} \,\right] \geqslant 1 - n^{-3}. \qquad \square$$

7.2 Layered induction over super-exponential potentials

In this section, we illustrate how to apply the concentration bounds (Theorem 4.1 and Theorem 6.9) in order to obtain $o(\log n)$ gap bounds for k-DENSE-QUANTILE, k-DENSE-THRESHOLD, QUANTILE(δ^*) and g-ADV(TWO-CHOICE).

7.2.1 k-DENSE-QUANTILE process

Recall that the k-DENSE-QUANTILE process is the QUANTILE($\delta_1, \ldots, \delta_k$) process where δ_j is the rounded up (to the next multiple of $1/n$) quantile of $\widetilde{\delta}_j$ defined as:

$$\widetilde{\delta}_j := \begin{cases} \frac{1}{2} & \text{if } j = k, \\ e^{-\frac{1}{4}\cdot(\log n)^{(k-j)/k}} & \text{if } 1 \leqslant j \leqslant k-1. \end{cases}$$

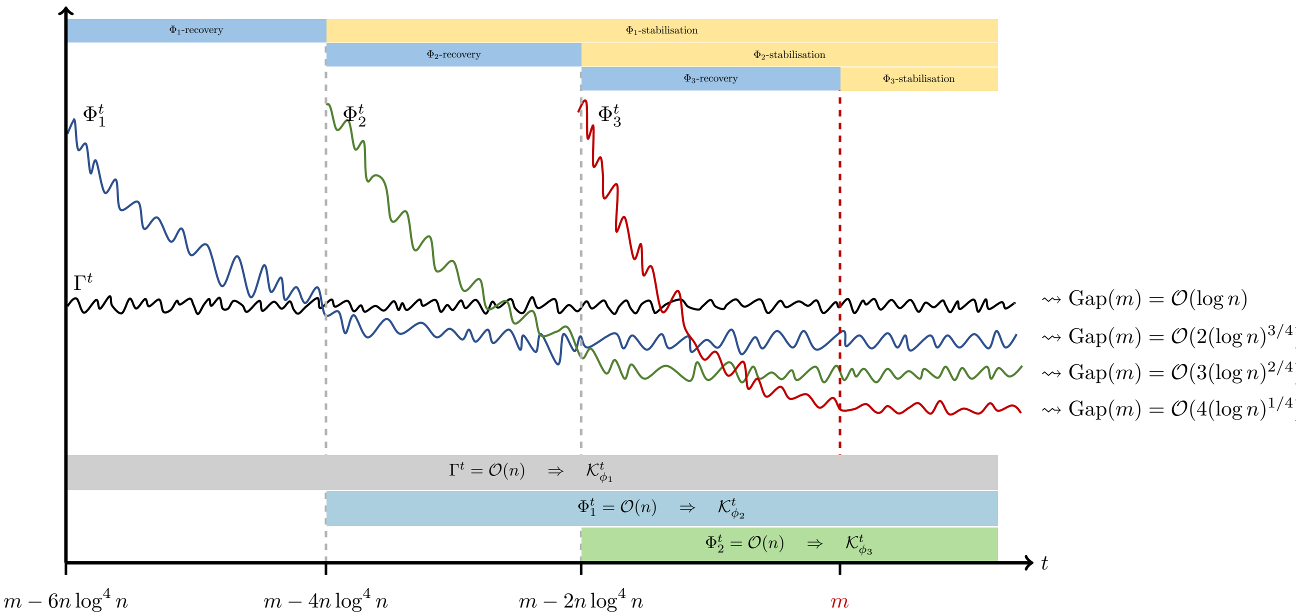


Figure 7.1: *Layered induction* for the k-DENSE-QUANTILE for $k = 4$, showing that when the potential at layer j satisfies $\Phi_j^t = \mathcal{O}(n)$, the the drop condition at layer $j+1$, i.e., $\mathcal{K}_{\phi_{j+1}}^t$ is implied. Then, after a *recovery phase* the potential Φ_{j+1} *stabilises* at $\mathcal{O}(n)$ and implies a tighter bound on the gap.

Similarly, to the 2-DENSE-QUANTILE process, the k-DENSE-QUANTILE process is majorised by the MEDIAN-QUANTILE process, so by Lemma 3.5 it satisfies Theorem 3.2 with $\delta = 1/3$, $\epsilon = 1/3$ and $C = 2$.

Corollary 7.8 (cf. Corollary 7.4). *Consider the k-DENSE-QUANTILE process and the hyperbolic cosine potential $\Gamma := \Gamma(\gamma)$ for any $\gamma \in \left(0, \frac{1}{8\cdot 3^2\cdot 2^2}\right]$. Then, there exist constants $c_1, c_2 > 0$, such that for any step $t \geqslant 0$,*

$$\mathbf{E}\left[\, \Gamma^{t+1} \,\middle|\, \mathfrak{F}^t \,\right] \leqslant \Gamma^t \cdot \left(1 - \frac{c_1 \gamma}{n}\right) + c_2 \cdot \gamma.$$

Then, by Theorem 4.1 with $\gamma_1 := \frac{1}{8\cdot 3^2\cdot 2^2}$, $\gamma_2 := \frac{\gamma_1}{84}$ and $\kappa = 6$, we get the following corollary.

Corollary 7.9. *Consider* k-DENSE-QUANTILE *process and the potential* $\Gamma_2 := \Gamma(\gamma_2)$*. Then, there exists a constant* $c > 0$ *such that for any step* $t \geqslant 0$

$$\mathbf{Pr}\Big[\, \Gamma_2^t \leqslant 3cn \,\Big] \geqslant 1 - n^{-6}.$$

We will now analyse the k-DENSE-QUANTILE process for any $1 \leqslant k \leqslant k_{\max}$, where

$$k_{\max} := \left\lceil \frac{1}{k'} \cdot \log\log n \right\rceil \quad \text{ and } \quad k' := \log\log(\max\{3c, 8, 4^8\}).$$

Note that since $(\log n)^{1/k} \geqslant 4$, $\widetilde{\delta}_j \geqslant \widetilde{\delta}_1 \geqslant e^{-\frac{1}{4}\cdot\log n} = n^{-1/4}$ and so $\widetilde{\delta}_j \leqslant \delta_j \leqslant \frac{1}{n} + \widetilde{\delta}_j \leqslant 2\widetilde{\delta}_j$.

We define the super-exponential potential functions $\Psi_j := \Psi_j(\psi_j, z_j)$ for $1 \leqslant j \leqslant k-1$, with smoothing parameters $\psi_j := \gamma_1 \cdot (\log n)^{j/k}$ and offsets $z_j := \lceil \frac{3}{\gamma_2} \cdot j \cdot (\log n)^{1/k} \rceil$, so that $z_j - 1 \geqslant \frac{2}{\gamma_2} \cdot j \cdot (\log n)^{1/k}$. Similarly we define $\Phi_j := \Phi_j(\phi_j, z_j)$ with a smoothing parameter of $\phi_j := \gamma_2 \cdot (\log n)^{j/k}$ and the same offset z_j. Also, let $\widetilde{\mathcal{K}}_{\phi_j}^t$ be the event associated with their drop in Theorem 6.9. For convenience we also define $\Phi_0 := \Phi_0(\gamma_2, 0)$, and then we have that $\Phi_0^t \leqslant \Gamma_2^t$.

Lemma 7.10. *Let* $C := \max\{3c, 8\}$ *and consider the* k-DENSE-QUANTILE *process for any* $1 \leqslant k \leqslant k_{\max}$*. For any* $1 \leqslant j \leqslant k-1$ *and any step* $t \geqslant 0$ *such that* $\Phi_{j-1}^t \leqslant Cn$*, we have that* $\widetilde{\mathcal{K}}_{\phi_j}^t$ *holds.*

Proof. Consider an arbitrary step t with $\Phi_{j-1}^t \leqslant Cn$. Then, in this step the number of bins with load at least $z_j - 1$ is at most

$$Cne^{-\gamma_2(\log n)^{(j-1)/k}\cdot(z_j - 1 - z_{j-1})} \leqslant Cne^{-\gamma_2(\log n)^{(j-1)/k}\cdot\frac{2}{\gamma_2}\cdot(\log n)^{1/k}} \overset{(a)}{\leqslant} ne^{-(\log n)^{j/k}} \leqslant n\widetilde{\delta}_{k-j} \leqslant n\delta_{k-j},$$

using in (a) that $(\log n)^{1/k} \geqslant \log C$ since $k \leqslant k_{\max}$. So all bins $i \in [n]$ with load at least $z_j - 1$ have $\text{Rank}^t(i) \leqslant n\delta_{k-j}$ and hence (recalling that $\delta_0 = 0$),

$$q_i^t \leqslant \frac{\delta_{k-j-1} + \delta_{k-j}}{n} \leqslant \frac{4\widetilde{\delta}_{k-j}}{n} = \frac{4}{n} \cdot e^{-\frac{1}{4}\cdot(\log n)^{j/k}} \overset{(a)}{\leqslant} \frac{1}{n} \cdot e^{-\frac{1}{8}\cdot(\log n)^{j/k}} \overset{(b)}{\leqslant} \frac{1}{n} \cdot e^{-\gamma_1(\log n)^{j/k}},$$

using in (a) that $(\log n)^{1/k} \geqslant 8\log 4$, since $k \leqslant k_{\max}$ and in (b) that $\gamma_1 \leqslant 1/8$. $\square$

Finally, we can perform a layered induction over these super-exponential potentials to deduce a small gap at an arbitrary step m.

Theorem 7.11. *Consider the* k-DENSE-QUANTILE *process for any* $1 \leqslant k \leqslant k_{\max}$*. Then, there exists a constant* $\kappa > 0$*, such that for any step* $m \geqslant 0$*,*

$$\mathbf{Pr}\Big[\, \mathrm{Gap}(m) \leqslant \kappa \cdot k \cdot (\log n)^{1/k} \,\Big] \geqslant 1 - n^{-3}.$$

Proof. Let $t_j := m - n\log^5 n + 2nj\log^4 n$ for $1 \leqslant j \leqslant k-1$. By Corollary 7.9, taking the union bound over $m - t_0 \leqslant n^2$ steps and using the fact that $\Phi_0^t \leqslant \Gamma_2^t$,

$$\mathbf{Pr}\left[\, \bigcap_{t \in [t_0, m]} \left\{\Phi_0^t \leqslant 3cn\right\} \,\right] \geqslant 1 - n^{-4}.$$

By Lemma 7.10, for each $1 \leqslant j \leqslant k-1$, $\Phi_{j-1}^t \leqslant Cn$ (for $C := \max\{3c, 8\}$) implies that $\widetilde{\mathcal{K}}_{\phi_j}^t$ holds and that $\mathrm{Gap}(t) \leqslant \log^2 n$, and so by inductively applying Theorem 6.9 with starting point $t_j := m - 2n(k-j)\log^4 n$ and $\widetilde{t} := m$, we get that

$$\mathbf{Pr}\left[\bigcap_{t \in [t_{j-1}, m]} \widetilde{\mathcal{K}}^t\right] \geqslant \mathbf{Pr}\left[\bigcap_{t \in [t_{j-1}, m]} \left\{\Phi_{j-1}^t \leqslant 8n\right\}\right] \geqslant 1 - \frac{(\log n)^{8(j-1)}}{n^4},$$

implies that

$$\mathbf{Pr}\left[\bigcap_{t \in [t_j, m]} \left\{\Phi_j^t \leqslant 8n\right\}\right] \geqslant 1 - \frac{(\log n)^{8j}}{n^4}.$$

Hence, by induction we get that for $j = k-1$,

$$\mathbf{Pr}\left[\bigcap_{t \in [t_{k-1}, m]} \left\{\Phi_{k-1}^t \leqslant 8n\right\}\right] \geqslant 1 - \frac{(\log n)^{8(k-1)}}{n^4} \geqslant 1 - n^{-3},$$

using that $k = \mathcal{O}(\log\log n)$. Finally, if $\left\{\Phi_{k-1}^m \leqslant 8n\right\}$ holds, then we have that

$$\mathrm{Gap}(m) \leqslant z_{k-1} + \frac{\log(8n)}{\gamma_2 (\log n)^{(k-1)/k}} \leqslant \left\lceil \frac{3}{\gamma_2} \cdot (k-1) \cdot (\log n)^{1/k} \right\rceil + \frac{2}{\gamma_2} \cdot (\log n)^{1/k},$$

and hence for $\kappa := 5/\gamma_2 > 0$, we get that

$$\mathbf{Pr}\left[\, \mathrm{Gap}(m) \leqslant \kappa \cdot k \cdot (\log n)^{1/k} \,\right] \geqslant 1 - n^{-3}. \qquad \square$$

7.2.2 k-DENSE-THRESHOLD process

The purpose of this section is to show that the layered induction used to analyse the k-DENSE-QUANTILE process in Section 7.2.1, can also be used to analyse the k-DENSE-THRESHOLD process. For space and time considerations, we state the base case as a conjecture (though it is quite likely that the proof follows using the techniques for the strong stabilisation in the g-ADV setting, see Theorem 5.34):

Conjecture 7.12. *For any k-THRESHOLD process whose smallest threshold is $\frac{t}{n}$ (or in fact any process satisfying condition $\mathcal{P}_3$), there exist constants $c > 0$ and $\gamma_2 \in (0, 1/4)$, such that for the hyperbolic cosine potential $\Gamma_2 := \Gamma_2(\gamma_2)$, for any step $t \geqslant 0$,*

$$\mathbf{Pr}\left[\, \Gamma_2^t \leqslant 3cn \,\right] \geqslant 1 - n^{-6}.$$

Now recall that the k-DENSE-THRESHOLD process has thresholds defined by

$$f_j^t := \begin{cases} \frac{t}{n} & \text{if } j = 1, \\ \frac{t}{n} + \left\lceil \frac{3}{\gamma_2} \cdot j \cdot (\log n)^{1/k} \right\rceil - 2 & \text{if } 1 \leqslant j \leqslant k-1. \end{cases}$$

We proceed similarly to Section 7.2.1 defining the super-exponential potential functions Φ_j and Ψ_j, with the same offsets z_j and smoothing parameters as in Section 7.2.1, for any $1 \leqslant j \leqslant k-1$ and $1 \leqslant k \leqslant k_{\max}$. We now prove that when $\Phi_{j-1}^t \leqslant 8n$, we also have that $\widetilde{\mathcal{K}}_j^t$ holds:

Lemma 7.13. *Let $C := \max\{3c, 8\}$ and consider the k-DENSE-THRESHOLD process for any $1 \leqslant k \leqslant k_{\max}$. For any $1 \leqslant j \leqslant k-1$ and any step $t \geqslant 0$ such that $\Phi_{j-1}^t \leqslant Cn$, we have that $\widetilde{\mathcal{K}}_j^t$ holds.*

Proof. Consider an arbitrary step t with $\Phi_{j-1}^t \leqslant Cn$. Then, in this step the number of bins with load at least f_j^t is at most

$$n\delta_{f_j^t}^t \leqslant Cne^{-\gamma_2(\log n)^{(j-1)/k}\cdot(f_j^t - t/n - z_{j-1})} \leqslant Cne^{-\gamma_2(\log n)^{(j-1)/k}\cdot\frac{2}{\gamma_2}\cdot(\log n)^{1/k}} \overset{(a)}{\leqslant} ne^{-(\log n)^{j/k}},$$

using in (a) that $(\log n)^{1/k} \geqslant \log C$ since $k \leqslant k_{\max}$. To allocate to a bin $i \in [n]$ with $y_i^t \geqslant z_j - 1 > f_j^t - t/n$, we need to choose i and one bin $k \in [n]$ with $y_k^t > f_j^t - t/n$, so

$$q_i^t \leqslant \frac{2\delta_{f_j^t}}{n} \leqslant \frac{2}{n} \cdot e^{-(\log n)^{j/k}} \leqslant \frac{1}{n} \cdot e^{-\gamma_1(\log n)^{j/k}},$$

using that $(\log n)^{1/k} \geqslant \log 2$, since $k \leqslant k_{\max}$ and that $\gamma_1 \leqslant 1/2$. □

Theorem 7.14. *Consider the k-DENSE-THRESHOLD process for any $1 \leqslant k \leqslant k_{\max}$. Then, there exists a constant $\kappa > 0$, such that for any step $m \geqslant 0$,*

$$\mathbf{Pr}\left[\, \mathrm{Gap}(m) \leqslant \kappa \cdot k \cdot (\log n)^{1/k} \,\right] \geqslant 1 - n^{-3}.$$

Proof. The proof is the same as Theorem 7.11, as by Conjecture 7.12 and Lemma 7.13 we have the same guarantees on the potentials Γ_2 and Φ_j. □

7.2.3 QUANTILE(δ^*) process

In this section, we will analyse the QUANTILE(δ^*) process with δ^* being the rounded up quantile of $\widetilde{\delta}^* = \frac{(\log\log n)^2}{\log n}$, so $\widetilde{\delta}^* \leqslant \delta^* \leqslant 2\widetilde{\delta}^*$. We start by proving that this process is majorised by TIME-HOMOGENEOUS(p) process with a probability vector p having a bias at a constant δ. This means that we can use Theorem 3.2 and Lemma 3.8 to obtain the following corollary.

Corollary 7.15. *The QUANTILE(δ^*) process is majorised by TIME-HOMOGENEOUS(p) for*

$$p_i = \begin{cases} \frac{1-\epsilon}{n} & \text{if } i \leqslant n/3, \\ \frac{1+\widetilde{\epsilon}}{n} & \text{otherwise,} \end{cases}$$

for any $i \in [n]$ and where $\epsilon = 2\delta^*$ and $\widetilde{\epsilon} = \delta^*$.

By Lemma 3.5 it satisfies the preconditions of Theorem 3.2 for $\delta = 1/3$, $\epsilon = 2\delta^*$ and $C = 2$, so we get the following corollary.

Corollary 7.16. *Consider the QUANTILE(δ^*) process and the potential $\Gamma := \Gamma(\gamma)$ for any $\gamma \in \left(0, \frac{\delta^*}{8\cdot 3\cdot 2^2}\right]$. Then, there exist constants $c_1, c_2 > 0$, such that for any step $t \geqslant 0$,*

$$\mathbf{E}\left[\, \Gamma^{t+1} \,\middle|\, \mathfrak{F}^t \,\right] \leqslant \Gamma^t \cdot \left(1 - \frac{c_1\gamma\delta^*}{n}\right) + c_2\gamma\delta^*.$$

By Theorem 4.1 for $\gamma_1 := \frac{\delta^*}{8\cdot 3\cdot 2^2}$, $\gamma_2 := \frac{\gamma_1}{84}$ and for $\kappa := 6$, we get following lemma.

Lemma 7.17. *Consider the* QUANTILE(δ^*) *process and the potential* $\Gamma_2 := \Gamma_2(\gamma_2)$. *Then, there exists a constant* $c > 0$, *such that for any step* $t \geqslant 0$,

$$\mathbf{Pr}\Big[\, \Gamma_2^t \leqslant 3cn \,\Big] \geqslant 1 - n^{-6}.$$

We now define the super-exponential potential $\Phi_1 := \Phi_1(\phi_1, z)$ with smoothing parameter $\phi_1 := \gamma_1 \cdot \log\log n$ and offset $z := \big\lceil \frac{3}{\gamma_2} \cdot \frac{\log n}{\log\log n} \big\rceil$, so that $z - 1 \geqslant \frac{2}{\gamma_2} \cdot \frac{\log n}{\log\log n}$. Similarly we define $\Phi_2 := \Phi_2(\phi_2, z)$ with smoothing parameter $\phi_2 := \gamma_2 \cdot \log\log n$ and the same offset z. We will now prove that in any step t where $\Gamma_2^t \leqslant 3cn$, we also have that $\widetilde{\mathcal{K}}^t := \widetilde{\mathcal{K}}^t_{\phi_1}$ holds.

Lemma 7.18. *Consider the* QUANTILE(δ^*) *process. For any step* $t \geqslant 0$ *where* $\Gamma_2^t \leqslant 3cn$, *we also have that* $\widetilde{\mathcal{K}}^t$ *holds.*

Proof. Consider an arbitrary step t where $\Gamma_2^t \leqslant 3cn$ holds. Then, the number of bins with normalised load at least $z - 1$ is at most

$$3cn \cdot e^{-\gamma_2 \cdot \delta^* \cdot (z-1)} \leqslant 3cne^{-2\log\log n} \leqslant n \cdot e^{-\log\log n} \leqslant n \cdot e^{2\log\log\log n - \log\log n} = n\widetilde{\delta}^* \leqslant n\delta^*.$$

So any bin i with $y_i^t \geqslant z - 1$ has $\mathrm{Rank}^t(i) \leqslant n\delta^*$ and so

$$q_i^t = \frac{\delta^*}{n} \leqslant \frac{2\widetilde{\delta}^*}{n} \leqslant \frac{2}{n} \cdot e^{2\log\log\log n - \log\log n} \leqslant \frac{1}{n} \cdot e^{-\frac{1}{2}\log\log n} \leqslant \frac{1}{n} \cdot e^{-\gamma_1 \log\log n},$$

since $\gamma_1 = o(1)$. Hence, $\widetilde{\mathcal{K}}^t$ holds. □

Finally, we are ready to prove the bound on the gap for QUANTILE(δ^*).

Theorem 7.19. *Consider the* QUANTILE(δ^*) *process. There exists a constant* $\kappa > 0$ *such that for any step* $m \geqslant 0$,

$$\mathbf{Pr}\left[\, \mathrm{Gap}(m) \leqslant \kappa \cdot \frac{\log n}{\log\log n} \,\right] \geqslant 1 - n^{-3}.$$

Proof. By Lemma 7.17 and using the union bound over the $2n\log^4 n \leqslant n^2$ steps, we have that

$$\mathbf{Pr}\left[\, \bigcap_{t \in [m - 2n\log^4 n, m]} \left\{ \Gamma_2^t \leqslant 3cn \right\} \,\right] \geqslant 1 - n^{-4}.$$

By Lemma 7.18, $\left\{\Gamma_2^t \leqslant 3cn\right\}$ implies that $\widetilde{\mathcal{K}}^t$ holds and that $\mathrm{Gap}(t) \leqslant \log^2 n$, hence

$$\mathbf{Pr}\left[\, \left\{\mathrm{Gap}(m - 2n\log^4 n) \leqslant \log^2 n\right\} \cap \bigcap_{t \in [m - 2n\log^4 n, m]} \widetilde{\mathcal{K}}^t \,\right] \geqslant 1 - n^{-4}.$$

So, by Theorem 6.9, we get that

$$\mathbf{Pr}\Big[\, \Phi_2^m \leqslant 8n \,\Big] \geqslant 1 - n^{-3}.$$

Finally, when $\left\{\Phi_2^m \leqslant 8n\right\}$, we have that

$$\mathrm{Gap}(m) \leqslant z + \frac{\log(8n)}{\gamma_2 \cdot \log\log n} \leqslant \left\lceil \frac{3}{\gamma_2} \cdot \frac{\log n}{\log\log n} \right\rceil + \frac{2}{\gamma_2} \cdot \frac{\log n}{\log\log n},$$

and so for the constant $\kappa := 6/\gamma_2 > 0$,

$$\mathbf{Pr}\left[\, \mathrm{Gap}(m) \leqslant \kappa \cdot \frac{\log n}{\log\log n} \,\right] \geqslant 1 - n^{-3}.$$

□

7.3 k-Relaxed-Quantile condition

We now define the family of k-Relaxed-Quantile$_{\gamma,\epsilon}$ processes for $\gamma \geqslant 1$ and $\epsilon \in (0,1)$, which relaxes the definition of k-Dense-Quantile $=$ Quantile$(\delta_1, \ldots, \delta_k)$, with $1 \leqslant k \leqslant k_{\max}$. More specifically for any step $t \geqslant 0$, the allocation vector q^t satisfies the following: (i) conditions $\mathcal{C}_1$ for $\delta = 1/3$ and constant $\epsilon > 0$, and $\mathcal{C}_2$ for some constant $C > 1$ and (ii) it holds that,

$$(\widetilde{q}_1^t, \ldots, \widetilde{q}_i^t) \preceq \begin{cases} \left(\gamma \cdot \frac{\delta_1}{n}, \ldots, \gamma \cdot \frac{\delta_1}{n}\right) & \text{for any } 1 \leqslant i \leqslant \delta_1 n, \\ \left(\gamma \cdot \frac{\delta_1+\delta_2}{n}, \ldots, \gamma \cdot \frac{\delta_1+\delta_2}{n}\right) & \text{for any } \delta_1 n < i \leqslant \delta_2 n, \\ \vdots & \\ \left(\gamma \cdot \frac{\delta_{k-1}+\delta_k}{n}, \ldots, \gamma \cdot \frac{\delta_{k-1}+\delta_k}{n}\right) & \text{for any } \delta_{k-1} n < i \leqslant \delta_k n. \end{cases}$$

Note that the k-Dense-Quantile process falls into this class for $\gamma = 1$, $\epsilon = 1/3$ and $C > 1$ (cf. Eq. (2.4)).

Theorem 7.20. *Consider any k-Relaxed-Quantile$_{\gamma,\epsilon}$ process for any $1 \leqslant k \leqslant k_{\max}$. Then, there exists a constant $\kappa > 0$, such that for any step $m \geqslant 0$,*

$$\mathbf{Pr}\left[\, \mathrm{Gap}(m) \leqslant \kappa \cdot k \cdot (\log n)^{1/k} \,\right] \geqslant 1 - n^{-3}.$$

The proof of this theorem is omitted as it is similar to the proof of the k-Dense-Quantile with the base case following from the $\mathcal{C}_1$ condition and Corollary 3.6, and the rest of the analysis follows with a smaller (by a constant factor) smoothing parameter.

7.3.1 A d-Thinning process

In this section, we use k-Relaxed-Quantile framework to prove bounds for general d-Thinning, which are close to the lower bound in [77, Proposition 4.1].

Lemma 7.21. *Consider the d-Thinning process for $2 \leqslant d \leqslant k_{\max}$ induced by the quantiles $\delta_{d-1} \ldots, \delta_1$, given by $\delta_j = e^{-\frac{1}{4} \cdot (\log n)^{(d-1-j)/(d-1)}}$ for $1 \leqslant j \leqslant d-2$ and $\delta_{d-1} = 1/3$. Then, there exists a constant $\kappa > 0$ such that for any step $m \geqslant 0$,*

$$\mathbf{Pr}\left[\, \mathrm{Gap}(m) \leqslant \kappa \cdot (d-1) \cdot (\log n)^{1/(d-1)} \,\right] \geqslant 1 - n^{-2}.$$

The definition of this process is such that as we take more samples, we are accepting more bins.

Proof. We will show that this d-Thinning process satisfies the $(d-1)$-Relaxed-Quantile$_{\gamma,\epsilon}$ conditions with $\epsilon = 1/3$ and $\gamma = 1$. Recall the $\delta_1, \ldots, \delta_{d-1}$ in the definition of $(d-1)$-Relaxed-Quantile$_{\gamma,\epsilon}$ and consider two cases for the i-th heaviest bin:

- **Case 1 [$i \leqslant n\delta_1$]:** We allocate to $i \leqslant n\delta_1$ if the first $d-1$ samples are heavy and the last one is equal to i. So,
$$\widetilde{q}_i = \delta_1 \cdot \delta_2 \cdot \ldots \cdot \delta_{d-1} \cdot \frac{1}{n} \leqslant \frac{\delta_1}{n}.$$
- **Case 2 [$n\delta_k < i \leqslant n\delta_{k+1}$, $k < d-1$]:** Let $j_1, \ldots, j_d$ be the sampled bins. The probability of allocating to the i-th heaviest bin is if the first $\ell_1 \geqslant k$ samples were heavy and then we sampled i:
$$\widetilde{q}_i = \frac{1}{n} \cdot \sum_{\ell_1 = j}^{d-1} \prod_{\ell_2 = 1}^{\ell_1} \mathbf{Pr}\left[\, j_{\ell_2} \leqslant n \cdot \delta_{\ell_2} \,\right]$$

$$
\begin{aligned}
&= \frac{1}{n} \cdot \delta_1 \cdot \ldots \cdot \delta_k \cdot (1 + \delta_{k+1} + \delta_{k+1} \cdot \delta_{k+2} + \ldots + \delta_{k+1} \cdot \ldots \cdot \delta_{d-1}) \\
&\leqslant \frac{1}{n} \cdot \delta_1 \cdot \ldots \cdot \delta_k \cdot \Big(1 + \delta_{k+1} + \delta_{k+1} \cdot \frac{1}{3} + \delta_{k+1} \cdot \frac{1}{3^2} + \ldots\Big) \\
&\leqslant \frac{1}{n} \cdot \delta_k \cdot \Big(1 + \frac{3}{2}\delta_{k+1}\Big) \\
&= \frac{1}{n} \cdot \Big(\delta_k + \frac{3}{2} \cdot \delta_k \cdot \delta_{k+1}\Big) \\
&\leqslant \frac{1}{n} \cdot (\delta_k + \delta_{k+1}),
\end{aligned}
$$

using that $\delta_k \leqslant 1/3$ for any k.

This verifies condition (ii) and for condition (i) we have that for any $i \leqslant n\delta_{d-1}$

$$\widetilde{q}_i \leqslant \frac{1}{n} \cdot (\delta_{d-2} + \delta_{d-1}) \leqslant \frac{2}{3n}.$$

Hence, $\mathcal{C}_1$ is also satisfied for $\epsilon = 1/3$. Since all bins i with $\text{Rank}^t(i) > n/3$ have the same q_i^t, it follows that $q_i^t \leqslant \frac{3}{2n}$, so condition $\mathcal{C}_2$ holds for $C = 3/2$. So, by Theorem 7.11, w.h.p. it has an $\mathcal{O}((d-1) \cdot (\log n)^{1/(d-1)})$ gap. $\square$

7.3.2 The $(1+\beta)$-process with β close to 1

We will now prove a bound for the $(1+\beta)$-proces for any $\beta = 1 - e^{-\frac{1}{4}(\log n)^{(k-1)/k}}$ for integer $k = \mathcal{O}(\log\log n)$, by showing that it is a k-RELAXED-QUANTILE$_{3,\beta/3}$ process.

Lemma 7.22. *Consider the $(1+\beta)$-process with $\beta \geqslant 1 - e^{-\frac{1}{4}(\log n)^{(k-1)/k}} = 1 - \widetilde{\delta}_1$ for some integer $1 \leqslant k \leqslant k_{\max}$. Then, it is a k-RELAXED-QUANTILE$_{\gamma,\epsilon}$ process with $\epsilon = \beta/3$ and $\gamma = 3$.*

Proof. Let $\widetilde{q}^t := \widetilde{q}$ be the sorted allocation vector of the $(1+\beta)$-process, where $\beta \geqslant 1 - \widetilde{\delta}_1$ for some integer $k \geqslant 1$.

First, consider any $1 \leqslant j \leqslant k$. Note that as $\widetilde{q}$ is non-decreasing in i and $\beta \geqslant 1 - \delta_1 \geqslant 1 - \delta_j$, and for any $\delta_{j-1}n + 1 \leqslant i \leqslant \delta_j n$,

$$\widetilde{q}_i \overset{(a)}{\leqslant} \widetilde{q}_{\delta_j n} \overset{(b)}{=} (1-\beta) \cdot \frac{1}{n} + \beta \cdot \frac{2n \cdot \delta_j - 1}{n^2} \overset{(c)}{\leqslant} \delta_j \cdot \frac{1}{n} + 1 \cdot 2 \cdot \delta_j \cdot \frac{1}{n} = 3 \cdot \delta_j \cdot \frac{1}{n} \leqslant 3 \cdot (\delta_{j-1} + \delta_j) \cdot \frac{1}{n},$$

where (a) and (b) hold by definition of $(1+\beta)$-process, and inequality (c) uses $\beta \geqslant 1 - \widetilde{\delta}_1 \geqslant 1 - \delta_1 \geqslant 1 - \delta_j$.

Similar to the above, for $\epsilon := \frac{\beta}{3}$, we can upper bound

$$\widetilde{q}_{n/3} = (1-\beta) \cdot \frac{1}{n} + \beta \cdot \frac{2(n/3) - 1}{n^2} \leqslant \frac{1}{n} + \beta \cdot \left(\frac{2}{3n} - \frac{1}{n}\right) = \frac{1 - \frac{1}{3}\beta}{n}. \qquad \square$$

Because $\widetilde{q}$ is non-decreasing, it follows that it satisfies $\mathcal{C}_1$ with $\delta = 1/3$ and $\epsilon = \frac{\beta}{3}$.

Using the above lemma and Theorem 7.20, we get the following upper bound.

Theorem 7.23. *Consider the $(1+\beta)$-process with $\beta \geqslant 1 - e^{-\frac{1}{4}(\log n)^{(k-1)/k}}$ for some integer $1 \leqslant k \leqslant k_{\max}$. Then, there exists a constant $\kappa > 0$ such that for any step $m \geqslant 0$,*

$$\mathbf{Pr}\Big[\,\text{Gap}(m) \leqslant \kappa \cdot k \cdot (\log n)^{1/k}\,\Big] \geqslant 1 - n^{-3}.$$

By inverting the value of k in $\beta = 1 - e^{-\frac{1}{4}\cdot(\log n)^{(k-1)/k}}$, we get that

$$(\log n)^{1/k} = \frac{\log n}{-4\log(1-\beta)}, \quad \text{and} \quad k = \left(\log\left(\frac{\log n}{-4\log(1-\beta)}\right)\right)^{-1} \cdot \log\log n.$$

Hence, if we set $B = \frac{\log n}{-4\log(1-\beta)}$, we get an upper bound of $\mathcal{O}\big(\frac{B}{\log B} \cdot \log\log n\big)$ bound, which matches the lower bound in Lemma C.21 for any $1-\beta = \Omega(e^{-\log^c n})$ with $c = \Omega\left(\frac{1}{\log\log n}\right)$.

7.4 g-Adv setting for Two-Choice with $g \leqslant \log n$

7.4.1 Main setting

In this section, we will complete the proof of the $\mathcal{O}\big(\frac{g}{\log g} \cdot \log\log n\big)$ upper bound for any $g \leqslant \log n$ for Two-Choice in the g-Adv setting. We do this by first proving the key lemma for the drop of the super-exponential potentials (Lemma 7.24) and then complete the layered induction in Theorem 7.25.

Let

$$\alpha_1 := \frac{1}{6\kappa} \leqslant \frac{1}{6\cdot 18}, \tag{7.4}$$

for $\kappa \geqslant \frac{1}{\alpha} = 18 > 0$ the constant in Eq. (5.19) in Lemma 5.25 and

$$\alpha_2 := \frac{\alpha_1}{84} \leqslant \frac{1}{84\cdot 6\cdot 18}. \tag{7.5}$$

We define the function

$$f(k) := (\alpha_1 \log n)^{1/k} = e^{\frac{1}{k}\cdot\log(\alpha_1 \log n)},$$

which is monotone decreasing in $k > 0$, and for $k = 1$, $f(1) = \alpha_1 \log n$. This implies that for every $1 < g < \alpha_1 \log n$, there exists a unique integer $k := k(g) \geqslant 2$ satisfying,

$$(\alpha_1 \log n)^{1/k} \leqslant g < (\alpha_1 \log n)^{1/(k-1)}.$$

This definition implies that $k = \Theta\big(\frac{\log\log n}{\log g}\big)$ and that $k = \mathcal{O}(\log\log n)$, since $g > 1$.

Keeping in mind the previous inequality, we will be making the slightly stronger assumption for $g = \Omega(1)$ (see Claim D.7) that

$$(\alpha_1 \cdot \log n)^{1/k} \leqslant g < \left(\frac{\alpha_2}{4} \cdot \log n\right)^{1/(k-1)}. \tag{7.6}$$

For any g satisfying $\left(\frac{\alpha_2}{4} \cdot \log n\right)^{1/k} \leqslant g < (\alpha_1 \cdot \log n)^{1/k}$, we will obtain the stated $\mathcal{O}\big(\frac{g}{\log g} \cdot \log\log n\big)$ bound by analysing the $\widetilde{g}$-Adv-Comp setting for $\widetilde{g} = (\alpha_1 \cdot (\log n))^{1/k} > g$, since

$$\frac{\widetilde{g}}{\log \widetilde{g}} \leqslant \frac{\widetilde{g}}{\log g} = \frac{\widetilde{g}}{g} \cdot \frac{g}{\log g} \leqslant \left(\frac{4\alpha_1}{\alpha_2}\right)^{1/k} \cdot \frac{g}{\log g} = \mathcal{O}\left(\frac{g}{\log g}\right).$$

We will now define the super-exponential potential functions $\Phi_0, \ldots, \Phi_{k-1}$. The base potential function Φ_0 is just an exponential potential (i.e., has a constant smoothing parameter) defined as

$$\Phi_0^s := \Phi_0^s(\alpha_2, z_0) := \sum_{i=1}^n \Phi_{0,i}^s := \sum_{i=1}^n \exp\left(\alpha_2 \cdot (y_i^s - z_0)^+\right), \tag{7.7}$$

where $\alpha_2 := \frac{\alpha_1}{84}$ and $z_0 := c_5 \cdot g$ for some sufficiently large constant integer $c_5 > 0$ (which was defined in Eq. (5.36) in Lemma 5.33). Further, we define for any integer $1 \leqslant j \leqslant k-1$,

$$\Phi_j^s := \Phi_j^s(\alpha_2 \cdot (\log n) \cdot g^{j-k}, z_j) := \sum_{i=1}^{n} \Phi_{j,i}^s := \sum_{i=1}^{n} \exp\Big(\alpha_2 \cdot (\log n) \cdot g^{j-k} \cdot (y_i^s - z_j)^+\Big), \tag{7.8}$$

where the offsets are given by

$$z_j := c_5 \cdot g + \left\lceil \frac{4}{\alpha_2} \right\rceil \cdot j \cdot g. \tag{7.9}$$

We further define

$$\Psi_0 := \Psi_0(\alpha_1, z_0) := \sum_{i=1}^{n} \Psi_{0,i}^s := \sum_{i=1}^{n} \exp\Big(\alpha_1 \cdot (y_i^s - z_0)^+\Big), \tag{7.10}$$

and for $1 \leqslant j \leqslant k-1$,

$$\Psi_j^s := \Psi_j^s(\alpha_1 \cdot (\log n) \cdot g^{j-k}, z_j) := \sum_{i=1}^{n} \Psi_{j,i}^s := \sum_{i=1}^{n} \exp\Big(\alpha_1 \cdot (\log n) \cdot g^{j-k} \cdot (y_i^s - z_j)^+\Big). \tag{7.11}$$

We are now ready to prove the key lemma for the layered induction.

Lemma 7.24. *Consider the* g-Adv-Comp *setting for any $g \in [\log(2C), \alpha_1 \log n]$, for the constant $C \geqslant 8$ defined in Theorem 5.34 and $\alpha_1 > 0$ defined in Eq. (7.4). Further, let $k := k(g) \geqslant 2$ be the unique integer such that $(\alpha_1 \log n)^{1/k} \leqslant g < (\alpha_1 \log n)^{1/(k-1)}$. Then, for any integer $1 \leqslant j \leqslant k-1$ and any step $s \geqslant 0$, $\Phi_{j-1}^s \leqslant Cn$ implies $\widetilde{\mathcal{K}}_{\psi_j, z_j}^s$.*

Proof. Consider an arbitrary step s with $\Phi_{j-1}^s \leqslant Cn$. Recall the definition of $\widetilde{\mathcal{K}}_{\psi_j, z_j}^s$,

$$\widetilde{\mathcal{K}}_{\psi_j, z_j}^s(q^s) := \left\{ \forall i \in [n]\colon \ y_i^s \geqslant z_j - 1 \ \Rightarrow \ q_i^s \leqslant \frac{1}{n} \cdot e^{-\psi_j} \right\}.$$

Thus, we want to bound the probability to allocate to a bin $i \in [n]$ with load $y_i^s \geqslant z_j - 1$. In order to do this, we will bound the number of bins $\ell \in [n]$ for which the adversary can reverse the comparison of i and ℓ, by bounding the ones with load $y_\ell^s \geqslant z_j - 1 - g$. Recall that $z_j := c_5 \cdot g + \lceil \frac{4}{\alpha_2} \rceil \cdot j \cdot g$. In the analysis below we make use of the following simple bound for $1 \leqslant j \leqslant k-1$,

$$z_j - 1 - g - z_{j-1} = \left\lceil \frac{4}{\alpha_2} \right\rceil \cdot g - 1 - g \geqslant \frac{3}{\alpha_2} \cdot g, \tag{7.12}$$

using that $\alpha_2 \leqslant 1/2$. We consider the cases $j = 1$ and $j > 1$ separately as Φ_0 has a slightly different form than Φ_{j-1} for $j > 1$.

Case 1 [$j = 1$]: The contribution of any bin $\ell \in [n]$ with load $y_\ell^s \geqslant z_1 - 1 - g$ to Φ_0^s is,

$$\Phi_{0,\ell}^s = e^{\alpha_2 \cdot (y_\ell^s - z_0)^+} \geqslant e^{\alpha_2 \cdot (z_1 - 1 - g - z_0)^+} \overset{(7.12)}{\geqslant} e^{3g}.$$

Hence, when $\{\Phi_0^s \leqslant Cn\}$ holds, the number of such bins is at most

$$Cn \cdot e^{-3g} = Cn \cdot e^{-g} \cdot e^{-2g} \leqslant \frac{n}{2} \cdot e^{-2g},$$

using that $g \geqslant \log(2C)$. Hence, the probability of allocating a ball to a bin $i \in [n]$ with $y_i^s \geqslant z_1 - 1$ is at most that of sampling i and a bin $\ell \in [n]$ with $y_\ell^s \geqslant z_1 - 1 - g$, i.e., at most

$$q_i^s \leqslant 2 \cdot \frac{1}{n} \cdot \frac{1}{2} \cdot e^{-2g} \leqslant \frac{1}{n} \cdot e^{-\psi_1},$$

using that $\psi_1 := \alpha_1 \cdot (\log n) \cdot g^{1-k} \leqslant g$, as $g \geqslant (\alpha_1 \log n)^{1/k}$.

Case 2 [$j > 1$]**:** The contribution of any bin $\ell \in [n]$ with load $y_\ell^s \geqslant z_j - 1 - g$ to Φ_{j-1}^s is,

$$\Phi_{j-1,\ell}^s = e^{\alpha_2 \cdot (\log n) \cdot g^{j-1-k} \cdot (y_\ell^s - z_{j-1})} \overset{(7.12)}{\geqslant} e^{\alpha_2 \cdot (\log n) \cdot g^{j-1-k} \cdot (\frac{3}{\alpha_2} \cdot g)} = e^{3 \cdot (\log n) \cdot g^{j-k}}.$$

Hence, when $\{\Phi_{j-1}^s \leqslant Cn\}$ holds, the number of such bins is at most

$$Cn \cdot e^{-3 \cdot (\log n) \cdot g^{j-k}} \leqslant \frac{n}{2} \cdot e^{-2 \cdot (\log n) \cdot g^{j-k}},$$

using that $e^{-(\log n) \cdot g^{j-k}} \leqslant e^{-(\log n) \cdot g^{2-k}} \leqslant e^{-(\log n) \cdot g \cdot \frac{1}{\alpha_1 \log n}} \leqslant e^{-\log(2C)} = \frac{1}{2C}$, since $j > 1$, $g \leqslant (\alpha_1 \log n)^{1/(k-1)}$, $g \geqslant \log(2C)$ and $\alpha_1 \leqslant 1$.

Hence, the probability of allocating a ball to a bin $i \in [n]$ with $y_i^s \geqslant z_j - 1$ is at most that of sampling i and a bin $\ell \in [n]$ with $y_\ell^s \geqslant z_j - 1 - g$, i.e., at most

$$q_i^s \leqslant 2 \cdot \frac{1}{n} \cdot \frac{1}{2} \cdot e^{-2 \cdot (\log n) \cdot g^{j-k}} = \frac{1}{n} \cdot e^{-2 \cdot (\log n) \cdot g^{j-k}} \leqslant \frac{1}{n} \cdot e^{-\psi_j},$$

recalling that $\psi_j := \alpha_1 \cdot (\log n) \cdot g^{j-k}$ for $\alpha_1 \leqslant 1$.

Combining the two cases, we conclude that the event $\widetilde{\mathcal{K}}_{\psi_j, z_j}^s$ holds at step s. $\square$

Theorem 7.25. *Consider the g-*Adv-Comp *setting for any $g \in (1, \log n]$. Then, there exists a constant $\widetilde{\kappa} > 0$ such that for any step $m \geqslant 0$,*

$$\mathbf{Pr}\left[\operatorname{Gap}(m) \leqslant \widetilde{\kappa} \cdot \frac{g}{\log g} \cdot \log \log n\right] \geqslant 1 - n^{-3}.$$

Proof. Let $g_{\min} := \max\left\{\log(2C), \frac{\alpha_2}{4\sqrt{\alpha_1}}\right\}$. We consider three cases depending on the value of $1 \leqslant g \leqslant \log n$:

Case 1 [$\min\left\{\frac{\alpha_2}{4}, c_6\right\} \cdot \log n \leqslant g \leqslant \log n$]**:** (for $c_6 > 0$ as defined in Eq. (5.22)) In this case, the $\mathcal{O}(\log n)$ upper bound follows by the $\mathcal{O}(g + \log n)$ upper bound of Theorem 5.26.

Case 2 [$1 < g < g_{\min}$]**:** For $1 < g \leqslant g_{\min}$, g is constant and the $\mathcal{O}(\log \log n)$ upper bound on the gap will follow by considering the $\widetilde{g}$-Adv-Comp setting with $\widetilde{g} = \lceil g_{\min} \rceil$. This setting encompasses g-Adv-Comp as $\widetilde{g} \geqslant g$ and is analysed in Case 3.

Case 3 [$g_{\min} \leqslant g < \min\left\{\frac{\alpha_2}{4}, c_6\right\} \cdot \log n$]**:** Recall that for any $g < \alpha_1 \log n$, we defined the unique integer $k := k(g) \geqslant 2$ satisfying,

$$(\alpha_1 \log n)^{1/k} \leqslant g < (\alpha_1 \log n)^{1/(k-1)},$$

and as explained in Eq. (7.6), since $g < \frac{\alpha_2}{4} \log n$, we may assume that the following stronger condition holds

$$(\alpha_1 \log n)^{1/k} \leqslant g < \left(\frac{\alpha_2}{4} \log n\right)^{1/(k-1)},$$

where the inequalities are valid using Claim D.7 and that $g \geqslant \frac{\alpha_2}{4\sqrt{\alpha_1}}$.

Let $t_j := m - 2n(k-j)\log^4 n$ for any integer $0 \leqslant j \leqslant k-1$. We will proceed by induction on the potential functions Φ_j. The base case follows by applying Theorem 5.34 (using $g \leqslant c_6 \log n$ and $t_0 \geqslant m - n\log^5 n$),

$$\mathbf{Pr}\left[\bigcap_{t\in[t_0,m]} \{\Phi_0^t \leqslant Cn\} \right] \geqslant 1 - n^{-4}. \tag{7.13}$$

We will now prove the induction step.

Lemma 7.26 (Induction step). *Consider the* g-Adv-Comp *setting for any* $g \geqslant \max\left\{\log(2C), \frac{\alpha_2}{4\sqrt{\alpha_1}}\right\}$ *satisfying* $(\alpha_1 \log n)^{1/k} \leqslant g < (\frac{\alpha_2}{4}\log n)^{1/(k-1)}$ *for some integer* $k \geqslant 2$*, where* $C > 0$ *is the constant defined in Theorem 5.34 and* $\alpha_1, \alpha_2 > 0$ *are defined in Eq. (7.4) and Eq. (7.5). Then, for any integer* $1 \leqslant j \leqslant k-1$ *and any step* $m \geqslant 0$*, if it holds that*

$$\mathbf{Pr}\left[\bigcap_{t\in[t_{j-1},m]} \{\Phi_{j-1}^t \leqslant Cn\} \right] \geqslant 1 - \frac{(\log n)^{8(j-1)}}{n^4},$$

then it also follows that

$$\mathbf{Pr}\left[\bigcap_{t\in[t_j,m]} \{\Phi_j^t \leqslant Cn\} \right] \geqslant 1 - \frac{(\log n)^{8j}}{n^4}.$$

Proof of Lemma 7.26. Consider an arbitrary integer j with $1 \leqslant j \leqslant k-1$ and assume that

$$\mathbf{Pr}\left[\bigcap_{t\in[t_{j-1},m]} \{\Phi_{j-1}^t \leqslant Cn\} \right] \geqslant 1 - \frac{(\log n)^{8(j-1)}}{n^4}.$$

By Lemma 7.24, we have that $\{\Phi_{j-1}^t \leqslant Cn\}$ implies $\widetilde{\mathcal{K}}_{\psi_j,z_j}^t$. Furthermore, $\{\Phi_{j-1}^{t_{j-1}} \leqslant Cn\}$ also implies $\{\mathrm{Gap}(t_{j-1}) \leqslant \log^2 n\}$. Hence, it also holds that

$$\mathbf{Pr}\left[\left\{\mathrm{Gap}(t_{j-1}) \leqslant \log^2 n\right\} \cap \bigcap_{t\in[t_{j-1},m]} \widetilde{\mathcal{K}}_{\psi_j,z_j}^t \right] \geqslant 1 - \frac{(\log n)^{8(j-1)}}{n^4}. \tag{7.14}$$

Applying Lemma 6.2 for the potentials Φ_j and Ψ_j, since $\psi_j \geqslant \phi_j = \alpha_2 \cdot (\log n) \cdot g^{j-k} \geqslant \alpha_2 \cdot (\log n) \cdot g^{1-k} \geqslant 4$ (as $g < (\frac{\alpha_2}{4}\log n)^{1/(k-1)}$), for any step $t \geqslant 0$ it holds that

$$\mathbf{E}\left[\Phi_j^{t+1} \,\middle|\, \mathfrak{F}^t, \widetilde{\mathcal{K}}_{\psi_j,z_j}^t \right] \leqslant \Phi_j^t \cdot \left(1 - \frac{1}{n}\right) + 2, \tag{7.15}$$

and

$$\mathbf{E}\left[\Psi_j^{t+1} \,\middle|\, \mathfrak{F}^t, \widetilde{\mathcal{K}}_{\psi_j,z_j}^t \right] \leqslant \Psi_j^t \cdot \left(1 - \frac{1}{n}\right) + 2. \tag{7.16}$$

Hence, by Eq. (7.15), Eq. (7.16) and Eq. (7.14), the preconditions of Theorem 6.9 are satisfied for starting step $t_j := m - 2n(k-j) \cdot \log^4 n$, $P := (\log n)^{8(j-1)}/n^4$ and terminating step at $\widetilde{t} := m$, and so we conclude (since $C \geqslant 8$) that

$$\mathbf{Pr}\left[\bigcap_{t\in[t_j,m]} \left\{\Phi_j^t \leqslant Cn\right\} \right] \geqslant 1 - \frac{(\log n)^{8j}}{n^4}. \qquad \square$$

Returning to the proof of Theorem 7.25, inductively applying Lemma 7.26 for $k-1$ times and using Eq. (7.13) as a base case, we get that

$$\mathbf{Pr}\left[\bigcap_{t\in[t_{k-1},m]}\left\{\Phi_{k-1}^t \leqslant Cn\right\}\right] \geqslant 1-\frac{(\log n)^{8(k-1)}}{n^4} \geqslant 1-n^{-3},$$

using in the last step that $k=\mathcal{O}(\log\log n)$. When $\{\Phi_{k-1}^m \leqslant Cn\}$ occurs, the gap at step m cannot be more than $z_k := c_5 g + \lceil \frac{4}{\alpha_2} \rceil \cdot k \cdot g$, since otherwise we would get a contradiction

$$\begin{aligned} Cn \geqslant \Phi_{k-1}^m &\geqslant \exp\Big(\alpha_2 \cdot (\log n) \cdot g^{(k-1)-k} \cdot (z_k - z_{k-1})\Big) \\ &= \exp\Big(\alpha_2 \cdot (\log n) \cdot g^{-1} \cdot \Big(\Big\lceil \frac{4}{\alpha_2} \Big\rceil \cdot g\Big)\Big) \geqslant \exp(4\cdot \log n) = n^4. \end{aligned}$$

Hence, $\text{Gap}(m) \leqslant z_k = c_5 g + \lceil \frac{4}{\alpha_2} \rceil \cdot k \cdot g$.

By the assumption on g, we have

$$g < \left(\frac{\alpha_2}{4}\log n\right)^{1/(k-1)} \;\Rightarrow\; \log g < \frac{1}{k-1}\cdot \log\left(\frac{\alpha_2}{4}\log n\right) \;\Rightarrow\; k < 1 + \frac{\log\left(\frac{\alpha_2}{4}\log n\right)}{\log g},$$

using that $g > 1$. Since $\alpha_2 > 0$ and $c_5 > 0$ are constants, we conclude that there exists a constant $\widetilde{\kappa} > 0$ such that

$$\mathbf{Pr}\left[\text{Gap}(m) \leqslant \widetilde{\kappa} \cdot \frac{g}{\log g} \cdot \log\log n\right] \geqslant 1-n^{-3}. \qquad \square$$

In the above, we actually proved the following slightly stronger corollary, which we will use in Section 7.4.3. This is based on the insight that, in order to prove the above gap bound at step m, the only assumption on the steps $[0, t_0)$, with $t_0 := m - n\log^5 n - \Delta_r$, is that the coarse bound of $\mathcal{O}(g\log(ng))$ on the difference between maximum and minimum load must hold at step t_0 (see, e.g., Lemma 5.23 and Theorem 5.34). While during the interval $[t_0, m]$ the process is required to be an instance of g-Adv-Comp, in the interval $[0, t_0)$ the process can be arbitrary as long as the coarse gap bound holds at step t_0.

Corollary 7.27. *Consider any $g \in (1, \log n]$, $m \geqslant 0$, $t_0 := m - n\log^5 n - \Delta_r$ for $\Delta_r := \Delta_r(g) > 0$ as defined in Lemma 5.23 and $c_3 > 0$ the constant in Theorem 3.21. Further, consider a process which, in steps $[t_0, m]$, is an instance of g-Adv-Comp setting. Then, there exists a constant $\widetilde{\kappa} > 0$, such that*

$$\mathbf{Pr}\left[\text{Gap}(m) \leqslant \widetilde{\kappa} \cdot \frac{g}{\log g} \cdot \log\log n \;\middle|\; \mathfrak{F}^{t_0}, \max_{i\in[n]} \left|y_i^{t_0}\right| \leqslant c_3 g \log(ng)\right] \geqslant 1-n^{-3}.$$

7.4.2 Relaxed g-Adv setting

In order to analyse the τ-Delay, b-Batched and probabilistic noisy settings, we study the relaxed g_1-Adv-Comp setting, where in most steps the adversary uses a $g_2 \ll g_1$. To make this more specific, consider an arbitrary instance of g_1-Adv-Comp and define D^t as the set of pairs of bins (of unequal load) whose comparison is reversed with non-zero probability by the adversary G^t to allocate ball $t+1$,

$$D^t(\mathfrak{F}^t) := \left\{(i,j) \in [n]\times[n] : y_i^t > y_j^t \wedge \Big(G^t(\mathfrak{F}^t, i, j, 1) > 0 \vee G^t(\mathfrak{F}^t, j, i, 2) > 0\Big)\right\},$$

and then define the largest load difference that could be reversed by the process at step $t+1$,

$$g^t(\mathfrak{F}^t) := \max_{(i,j)\in D^t(\mathfrak{F}^t)} \left| y_i^t - y_j^t \right|.$$

This can be seen as the "effective g-bound" of the process in step $t+1$. Note that $g^t \leqslant g_1$ holds deterministically.

Lemma 7.28. *Consider the g_1-ADV-COMP setting for any $g_1 \in [1, n\log n]$, and consider any $g_2 \in [1, \log^2 n]$. If for every step $t \geqslant 0$, we have*

$$\mathbf{Pr}\left[\, g^t \leqslant g_2 \,\right] \geqslant 1 - n^{-6},$$

then there exists a constant $\widetilde{\kappa} > 0$ such that for any step $m \geqslant 0$,

$$\mathbf{Pr}\left[\, \mathrm{Gap}(m) \leqslant \widetilde{\kappa} \cdot \frac{g_2}{\log g_2} \cdot \log\log n \,\right] \geqslant 1 - n^{-2}.$$

Proof of Lemma 7.28. Let $\mathcal{P}$ be a process satisfying the preconditions in the statement and we will define the auxiliary process $\widetilde{\mathcal{P}}_{t_0}$ for some step $t_0 \geqslant 0$ (to be specified below). Consider the stopping time $\sigma := \inf\{s \geqslant t_0 : g^s > g_2\}$. Then, the auxiliary process $\widetilde{\mathcal{P}}_{t_0}$ is defined so that

- in steps $s \in [0, \sigma)$, it makes the same allocations as $\mathcal{P}$, and
- in steps $s \in [\sigma, \infty)$, it makes the same allocations as the g_2-BOUNDED process.

This way $\widetilde{\mathcal{P}}_{t_0}$ is a g_2-ADV-COMP process for all steps $s \geqslant t_0$. Let $\widetilde{y}$ be the normalised load vector for $\widetilde{\mathcal{P}}_{t_0}$, then it follows by the precondition that w.h.p. the two processes agree for any interval $[t_0, m]$ with $m - t_0 \leqslant n^3$, i.e.,

$$\mathbf{Pr}\left[\bigcap_{s\in[t_0,m]} \{y^s = \widetilde{y}^s\} \right] \geqslant \mathbf{Pr}\left[\bigcap_{s\in[t_0,m]} \{g^s \leqslant g_2\} \right] \geqslant 1 - n^{-6}\cdot n^3 = 1 - n^{-3}. \qquad (7.17)$$

For $m \leqslant n^3$, the upper bound follows directly by Theorem 7.25 for $\widetilde{\mathcal{P}}_0$ and taking the union bound with Eq. (7.17), i.e., that $\widetilde{\mathcal{P}}_0$ agrees with $\mathcal{P}$.

For $m > n^3$, the analysis is slightly more challenging. We need to show that the process recovers from the weak upper bound obtained from the g_1-ADV-COMP setting. Let $\Gamma := \Gamma(\gamma)$ be as defined in Eq. (3.1) with $\gamma := -\log(1 - \frac{1}{17\cdot 32})/g_2$ for the $\widetilde{\mathcal{P}}_{t_0}$ process (i.e., the $\widetilde{y}$ load vector). Also, let $t_0 := m - n^3$ and $t_1 := m - n\log^5 n - \Delta_r$ where $\Delta_r := \Delta_r(g_2) = \Theta(ng_2(\log(ng_2))^2)$ is the recovery time defined in Lemma 5.23. In this analysis, we consider the following three phases (see Fig. 7.2):

- $[0, t_0]$: The process $\mathcal{P}$ is an instance of the g_1-ADV-COMP setting with $g_1 = n\log n$. Hence, by Theorem 3.21, it follows that at step t_0 w.h.p. $\mathrm{Gap}(t_0) = \mathcal{O}(n\log^2 n)$.
- $(t_0, t_1]$: The process $\mathcal{P}$ w.h.p. agrees with $\widetilde{\mathcal{P}}_{t_0}$ which is a g_2-ADV-COMP process. We will use an analysis similar to that in Section 3.3 to prove that w.h.p. $\mathrm{Gap}(t_1) = \mathcal{O}(g_2\log(ng_2))$.
- $(t_1, m]$: The process $\mathcal{P}$ w.h.p. continues to agree with $\widetilde{\mathcal{P}}_{t_0}$ which is a g_2-ADV-COMP process and so by Corollary 7.27 this implies that w.h.p. $\mathrm{Gap}(m) = \mathcal{O}\big(\frac{g_2}{\log g_2}\cdot\log\log n\big)$.

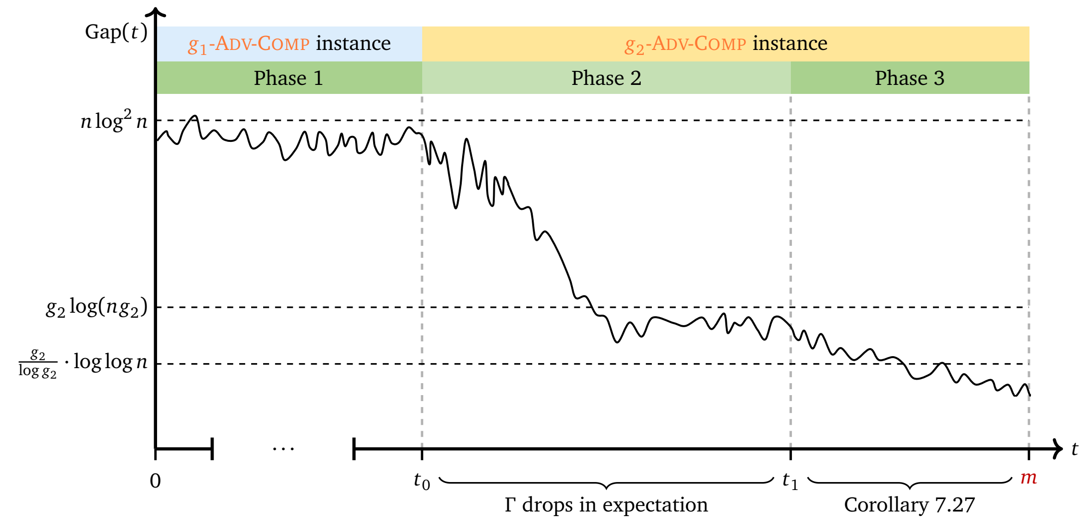


Figure 7.2: The three phases in the proof of Lemma 7.28.

Phase 1 $[0, t_0]$**:** Using Theorem 3.21 (iii) (for $g := n \log n$) and for $c_3 > 0$ being the constant defined in Eq. (3.32), we have that,

$$\mathbf{Pr}\left[\max_{i \in [n]} \left|\tilde{y}_i^{t_0}\right| \leqslant 3c_3 n \log^2 n\right] \geqslant 1 - n^{-14}, \tag{7.18}$$

using $c_3 g \log(ng) \leqslant c_3 n \cdot \log n \cdot \log(n^2 \cdot \log n) \leqslant 3c_3 n \log^2 n$.

Phase 2 $(t_0, t_1]$**:** Let us now turn our attention to the interval $(t_0, t_1]$, where $\widetilde{\mathcal{P}}_{t_0}$ is a g_2-Adv-Comp process. So, by Theorem 3.21 (i) (for $g := g_2$), there exists a constant $c_1 \geqslant 1$, such that for any step $t \geqslant t_0$

$$\mathbf{E}\left[\, \Gamma^{t+1} \,\middle|\, \mathfrak{F}^t \,\right] \leqslant \Gamma^t \cdot \left(1 - \frac{\gamma}{64n}\right) + c_1.$$

At step t_0, when $\{\max_{i \in [n]} |\tilde{y}_i^{t_0}| \leqslant 3c_3 n \log^2 n\}$ holds, we also have that $\Gamma^{t_0} \leqslant 2n \cdot e^{3\gamma c_3 n \log^2 n}$. Hence, applying Lemma B.1 (i) (with $a = 1 - \frac{\gamma}{64n}$ and $b = c_1$), for step t_1 we have

$$\begin{aligned}
\mathbf{E}\left[\, \Gamma^{t_1} \,\middle|\, \mathfrak{F}^{t_0}, \max_{i \in [n]} \left|\tilde{y}_i^{t_0}\right| \leqslant 3c_3 n \log^2 n \,\right] &\leqslant \mathbf{E}\left[\, \Gamma^{t_1} \,\middle|\, \mathfrak{F}^{t_0}, \Gamma^{t_0} \leqslant 2n \cdot e^{3\gamma c_3 n \log^2 n} \,\right] \\
&\leqslant \Gamma^{t_0} \cdot \left(1 - \frac{\gamma}{64n}\right)^{t_1 - t_0} + \frac{64c_1}{\gamma} \cdot n \\
&\overset{(a)}{\leqslant} 2n \cdot e^{3\gamma c_3 n \log^2 n} \cdot e^{-\frac{\gamma}{64n} \cdot \frac{1}{2} n^3} + \frac{64c_1}{\gamma} \cdot n \\
&\leqslant \frac{100c_1}{\gamma} \cdot n.
\end{aligned}$$

using in (a) that $e^u \geqslant 1 + u$ and $t_1 - t_0 \geqslant \frac{1}{2} n^3$. By Markov's inequality, we have that,

$$\mathbf{Pr}\left[\, \Gamma^{t_1} \leqslant \frac{100c_1}{\gamma} \cdot n^4 \,\middle|\, \mathfrak{F}^{t_0}, \max_{i \in [n]} \left|\tilde{y}_i^{t_0}\right| \leqslant 3c_3 n \log^2 n \,\right] \geqslant 1 - n^{-3}.$$

When the event $\left\{\Gamma^{t_1} \leqslant \frac{100c_1}{\gamma} \cdot n^4\right\}$ holds, it implies that

$$\mathrm{Gap}(t_1) \leqslant \frac{1}{\gamma} \cdot \left(\log\left(\frac{100c_1}{\gamma}\right) + 4\log n\right) \overset{(a)}{=} \frac{c_3 g_2}{16} \cdot \left(\mathcal{O}(1) + \log\left(\frac{c_3 g_2}{16}\right) + 4\log n\right) \leqslant c_3 g_2 \log(ng_2),$$

using in (a) that $c_3 := \frac{16}{\gamma g_2}$ (defined in Eq. (3.32)). Therefore,

$$\mathbf{Pr}\left[\max_{i\in[n]}\left|\widetilde{y}_i^{t_1}\right| \leqslant c_3 g_2 \log(n g_2) \;\middle|\; \mathfrak{F}^{t_0}, \max_{i\in[n]}\left|\widetilde{y}_i^{t_0}\right| \leqslant 3 c_3 n \log^2 n\right] \geqslant 1 - n^{-3}. \tag{7.19}$$

Phase 3 $(t_1, m]$**:** Now, we turn our attention to the steps in $(t_1, m]$, where $\widetilde{\mathcal{P}}_{t_0}$ is again a g_2-ADV-COMP process. Therefore, applying Corollary 7.27 (for $t_0 := t_1 = m - n\log^5 n - \Delta_r$ and $g := g_2$), there exists a constant $\widetilde{\kappa} > 0$ such that

$$\mathbf{Pr}\left[\mathrm{Gap}_{\widetilde{\mathcal{P}}_{t_0}}(m) \leqslant \widetilde{\kappa} \cdot \frac{g_2}{\log g_2} \cdot \log\log n \;\middle|\; \mathfrak{F}^{t_1}, \max_{i\in[n]}\left|\widetilde{y}_i^{t_1}\right| \leqslant c_3 g_2 \log(n g_2)\right] \geqslant 1 - n^{-3}. \tag{7.20}$$

By combining Eq. (7.18), Eq. (7.19) and Eq. (7.20), we have that

$$\mathbf{Pr}\left[\mathrm{Gap}_{\widetilde{\mathcal{P}}_{t_0}}(m) \leqslant \widetilde{\kappa} \cdot \frac{g_2}{\log g_2} \cdot \log\log n\right] \geqslant \left(1 - n^{-14}\right) \cdot \left(1 - n^{-3}\right) \cdot \left(1 - n^{-3}\right) \geqslant 1 - 3n^{-3}.$$

Finally, by Eq. (7.17) we have that w.h.p. $\mathcal{P}$ and $\widetilde{\mathcal{P}}_{t_0}$ agree in every step in $[t_0, m]$, so by taking the union bound we conclude

$$\mathbf{Pr}\left[\mathrm{Gap}_{\mathcal{P}}(m) \leqslant \widetilde{\kappa} \cdot \frac{g_2}{\log g_2} \cdot \log\log n\right] \geqslant 1 - 3n^{-3} - n^{-3} \geqslant 1 - n^{-2}. \qquad \square$$

7.4.3 Applications

An upper bound for the probabilistic noise setting

Proposition 7.29. *Consider the ρ-NOISY-COMP setting with $\rho(\delta)$ being any non-decreasing function in δ with $\lim_{\delta\to\infty}\rho(\delta) = 1$. For any $n \in \mathbb{N}$, define $\delta^* := \delta^*(n) = \min\{\delta \geqslant 1 \colon \rho(\delta) \geqslant 1 - n^{-4}\}$[1]. Then, there exists a constant $\kappa > 0$, such that for any step $m \geqslant 0$,*

$$\mathbf{Pr}\left[\max_{i\in[n]}\left|y_i^m\right| \leqslant \kappa \cdot \delta^* \log(n\delta^*)\right] \geqslant 1 - n^{-3}.$$

Note that for the σ-NOISY-LOAD process where $\rho(\delta)$ has Gaussian tails (see Eq. (2.3)), we have $\delta^* = \mathcal{O}(\sigma \cdot \sqrt{\log n})$. The choice of δ^* in Proposition 7.29 ensures that in most steps, all possible comparisons among bins with load difference greater than δ^* will be correct, implying that the process satisfies the condition of g-ADV-COMP with $g = \delta^*$.

Proof. We will analyse the hyperbolic cosine potential $\Gamma := \Gamma(\gamma)$ as defined in Eq. (3.1), with $\gamma := -\log(1 - \frac{1}{17\cdot 32})/\delta^*$. We first state a trivial upper bound on $\mathbf{E}\left[\,\Gamma^{t+1} \,\middle|\, \mathfrak{F}^t\,\right]$ in terms of Γ^t, which holds deterministically for all steps $t \geqslant 0$ (cf. Lemma 5.28 (i)),

$$\Gamma^{t+1} = \sum_{i=1}^{n} \Gamma_i^{t+1} \leqslant \sum_{i=1}^{n} e^{\gamma} \cdot \Gamma_i^t = e^{\gamma} \cdot \Gamma^t \leqslant (1 + 2\gamma) \cdot \Gamma^t,$$

using that $e^{\gamma} \leqslant 1 + 2\gamma$ for $0 < \gamma \leqslant 1$.

We will now provide a better upper bound, exploiting that with high probability all possible comparisons between bins that differ by at least δ^* will be correct. Again, consider any step $t \geqslant 0$. Let us

[1] Not to be confused with the optimal quantile in QUANTILE(δ^*).

assume that in step t, we first determine the outcome of the load comparisons among all n^2 possible bin pairs. Only then we sample two bins, and allocate the ball following the pre-determined outcome of the load comparison. For any two bins $i_1, i_2 \in [n]$ with $|x_{i_1}^t - x_{i_2}^t| \geqslant \delta^*$, we have

$$\rho(|x_{i_1}^t - x_{i_2}^t|) \geqslant \rho(\delta^*) \geqslant 1 - n^{-4}.$$

Hence by the union bound over all n^2 pairs, we can conclude that with probability at least $1-n^{-2}$, all comparisons among bin pairs with load difference at least δ^* are correct. Let us denote this event by $\mathcal{G}^t$, so $\mathbf{Pr}\left[\, \mathcal{G}^t \,\right] \geqslant 1 - n^{-2}$ for every step $t \geqslant 0$. Conditional on $\mathcal{G}^t$, the process in step t is an instance of g-Adv-Comp with $g = \delta^*$. Therefore, by Theorem 3.21 (i), there exists a constant $c_1 \geqslant 1$, so that the hyperbolic cosine potential satisfies

$$\mathbf{E}\left[\, \Gamma^{t+1} \,\middle|\, \mathfrak{F}^t, \mathcal{G}^t \,\right] \leqslant \Gamma^t \cdot \left(1 - \frac{\gamma}{64n}\right) + c_1.$$

Now combining our two upper bounds on $\mathbf{E}\left[\, \Gamma^{t+1} \,\middle|\, \mathfrak{F}^t \,\right]$, we conclude

$$\begin{aligned}
\mathbf{E}\left[\, \Gamma^{t+1} \,\middle|\, \mathfrak{F}^t \,\right] &\leqslant \mathbf{E}\left[\, \Gamma^{t+1} \,\middle|\, \mathfrak{F}^t, \mathcal{G}^t \,\right] \cdot \mathbf{Pr}\left[\, \mathcal{G}^t \,\right] + \mathbf{E}\left[\, \Gamma^{t+1} \,\middle|\, \mathfrak{F}^t \,\right] \cdot \left(1 - \mathbf{Pr}\left[\, \mathcal{G}^t \,\right]\right) \\
&\overset{(a)}{\leqslant} \mathbf{E}\left[\, \Gamma^{t+1} \,\middle|\, \mathfrak{F}^t, \mathcal{G}^t \,\right] \cdot \left(1 - n^{-2}\right) + \mathbf{E}\left[\, \Gamma^{t+1} \,\middle|\, \mathfrak{F}^t \,\right] \cdot n^{-2} \\
&\leqslant \left(\Gamma^t \cdot \left(1 - \frac{\gamma}{64n}\right) + c_1\right) \cdot \left(1 - n^{-2}\right) + (1 + 2\gamma) \cdot \Gamma^t \cdot n^{-2} \\
&\leqslant \Gamma^t \cdot \left(1 - \frac{\gamma}{64n}\right) + c_1 - \Gamma^t \cdot n^{-2} + \Gamma^t \cdot \frac{\gamma}{64n} \cdot n^{-2} + \Gamma^t \cdot n^{-2} + \Gamma^t \cdot 2\gamma \cdot n^{-2} \\
&\leqslant \Gamma^t \cdot \left(1 - \frac{\gamma}{96n}\right) + c_1,
\end{aligned}$$

using in (a) that $\mathbf{Pr}\left[\, \mathcal{G}^t \,\right] \geqslant 1 - n^{-2}$. So, using Lemma B.1 (ii) (with $a = 1 - \frac{\gamma}{96n}$ and $b = c_1$) since $\Gamma^0 = 2n \leqslant \frac{100c_1}{\gamma} \cdot n$, we get for any $t \geqslant 0$,

$$\mathbf{E}\left[\, \Gamma^t \,\right] \leqslant \frac{100c_1}{\gamma} \cdot n.$$

Using Markov's inequality yields, $\mathbf{Pr}\left[\, \Gamma^t > \frac{100c_1}{\gamma} \cdot n^4 \,\right] \leqslant n^{-3}$. Now the claim follows, since the event $\left\{\Gamma^t \leqslant \frac{100c_1}{\gamma} \cdot n^4\right\}$ for $\gamma = \Theta\left(\frac{1}{\delta^*}\right)$, implies that

$$\max_{i \in [n]} \left|y_i^t\right| \leqslant \frac{1}{\gamma} \log\left(\frac{100c_1}{\gamma} \cdot n^4\right) = \mathcal{O}(\delta^* \cdot \log(n\delta^*)). \qquad \square$$

Upper bounds for delay settings

In this section, we will prove tight upper bounds for the Two-Choice process in the τ-Delay and the b-Batched settings for a range of values for the delay parameter τ and batch size b. In particular for $\tau = n$, we show:

Theorem 7.30. *Consider the τ-Delay setting with $\tau = n$. Then, there exists a constant $\kappa > 0$ such that for any step $m \geqslant 0$,*

$$\mathbf{Pr}\left[\, \mathrm{Gap}(m) \leqslant \kappa \cdot \frac{\log n}{\log\log n} \,\right] \geqslant 1 - n^{-2}.$$

This implies the same upper bound for the b-BATCHED setting for $b = n$, since it is an instance of the τ-DELAY setting with $\tau = n$. This bound improves the $\mathcal{O}(\log n)$ bound in [28, Theorem 1] and can be easily seen to be asymptotically tight due to the $\Omega\big(\frac{\log n}{\log\log n}\big)$ lower bound for ONE-CHOICE with n balls (Observation C.5).

We will analyse τ-DELAY using a more general approach which also works for other choices of the delay parameter $\tau \leqslant n \log n$. First note that τ-DELAY is an instance of g_1-ADV-COMP with $g_1 := \tau - 1$. This follows since for all steps, a bin could sampled (and be allocated to) at most $\tau - 1$ times during the last $\tau - 1$ steps. However, for a typical execution we expect each bin to be incremented much less frequently during the last $\tau - 1$ steps, and thus the process is with high probability an instance of g_2-ADV-COMP for some $g_2 \ll g_1$.

We will now use Lemma 7.28 to prove Theorem 7.30.

Proof of Theorem 7.30. Any bin can be allocated to at most $n-1$ times during an interval of $\tau-1 = n-1$ steps, so an adversary with $g_1 = n-1$ who is also aware of the entire history of the process, can simulate τ-DELAY at any step t, by using the allocation information in steps $[t-n, t]$.

To obtain the bound for g_2, note that in TWO-CHOICE, at each step two bins are sampled for each ball. So in $n-1$ steps of TWO-CHOICE, there are $2(n-1)$ bins sampled using ONE-CHOICE. By the properties of ONE-CHOICE (Corollary B.21), we have that for any consecutive $n-1$ allocations, with probability at least $1-n^{-6}$ we sample (and allocate to) no bin more than $11 \log n / \log\log n$ times. In such a sequence of bin samples, we can simulate τ-DELAY using g_2-ADV-COMP with $g_2 = 11 \log n / \log\log n$. Hence, for any step $t \geqslant 0$,

$$\mathbf{Pr}\big[\, g^t \leqslant g_2 \,\big] \geqslant 1 - n^{-6}.$$

Since the precondition of Lemma 7.28 holds for $g_1 = n - 1$ and $g_2 = 11 \log n / \log\log n$, we get that there exists a constant $\kappa > 0$, such that

$$\mathbf{Pr}\left[\, \mathrm{Gap}(m) \leqslant \kappa \cdot \frac{\log n}{\log\log n} \,\right] \geqslant 1 - n^{-2}. \qquad \square$$

The same argument also applies for any $\tau \in [n \cdot e^{-\log^c n}, n \log n]$. For ONE-CHOICE with 2τ balls the gap is w.h.p. $\mathrm{polylog}(n)$ (e.g., see Lemma B.20) and so we can apply Lemma 7.28 with $g_2 = \mathrm{polylog}(n)$, to obtain the gap bound of

$$\frac{g_2}{\log g_2} \cdot \log\log n = \Theta(g_2).$$

Corollary 7.31. *There exists a constant $\kappa > 0$, such that the τ-DELAY setting with any $\tau \in [n \cdot e^{-\log^c n}, n \log n]$, where $c > 0$ is any constant, for any step $m \geqslant 0$, it holds that*

$$\mathbf{Pr}\left[\, \mathrm{Gap}(m) \leqslant \kappa \cdot \frac{\log n}{\log\big(\frac{4n}{\tau} \cdot \log n\big)} \,\right] \geqslant 1 - n^{-2}.$$

Remark 7.32. *A matching lower bound holds for the b-BATCHED setting for any batch size $b \in [n \cdot e^{-\log^c n}, n \log n]$. This follows by the lower bound for ONE-CHOICE with b balls (e.g., see Lemma B.24) which matches the gap of b-BATCHED in the first batch (Observation C.5).*

Therefore, Corollary 7.31 and Remark 7.32 establish that w.h.p. $\mathrm{Gap}(m) = \Theta\big(\frac{\log n}{\log((4n/b) \cdot \log n)}\big)$ for the b-BATCHED setting for any $b \in [n \cdot e^{-\log^c n}, n \log n]$. However, the following remark (which also applies to the τ-DELAY setting), establishes that there are regions where the b-BATCHED setting has an asymptotically worse gap than ONE-CHOICE with b balls.

Remark 7.33. *For any τ (or b) being $n^{1-\epsilon}$ for any constant $\epsilon \in (0,1)$, the* ONE-CHOICE *process has* $\mathrm{Gap}(b) = \mathcal{O}(1)$ *w.h.p. (see Corollary B.22). Hence, by Lemma 7.28 with $g_2 = \mathcal{O}(1)$,* τ-DELAY *has for any step $m \geqslant 0$,* $\mathrm{Gap}(m) = \mathcal{O}(\log\log n)$ *w.h.p., which is asymptotically tight by Observation C.7 for $m = n$.*

7.5 b-BATCHED setting: Tighter bounds

In Section 3.2.2, we proved an $\mathcal{O}\big(\frac{b}{n} \cdot \log n\big)$ bound for processes satisfying conditions $\mathcal{C}_1$ and $\mathcal{C}_2$, and in Section 3.2.3, we proved an $\mathcal{O}\big(\sqrt{\frac{b}{n}} \cdot \log n\big)$ bound for processes satisfying conditions $\mathcal{C}_1$ and $\mathcal{C}_3$. In this section, using an interplay between two potentials, we improve these bounds to $\mathcal{O}\big(\frac{b}{n} + \log n\big)$ and $\mathcal{O}\big(\sqrt{\frac{b}{n} \cdot \log n}\big)$ respectively, which are asymptotically tight as we show in Appendix C.1. More specifically, we prove the following two theorems with their direct corollaries for concrete processes.

Theorem 7.39 (Restated, page 162). *Consider any* SEQUENTIAL(q^t) *process with q^t satisfying condition $\mathcal{C}_1$ for constant $\delta \in (0,1)$ and constant $\epsilon \in (0,1)$ as well as condition $\mathcal{C}_2$ for some constant $C > 1$, at every step $t \geqslant 0$. Further, consider the* WEIGHTED b-BATCHED *setting with any $n \leqslant b \leqslant n^3$ and weights from a* FINITE-MGF(S) *distribution with constant $S \geqslant 1$. Then, there exists a constant $\kappa := \kappa(\delta, \epsilon, C, S) > 0$, such that for any step $m \geqslant 0$ being a multiple of b,*

$$\mathbf{Pr}\left[\max_{i \in [n]} y_i^m \leqslant \kappa \cdot \left(\frac{b}{n} + \log n\right)\right] \geqslant 1 - n^{-2}.$$

Corollary 7.34. *Consider the* WEIGHTED b-BATCHED *setting with any $b \in [2n \log n, n^3]$ and weights from a* FINITE-MGF(S) *distribution with constant $S \geqslant 1$. Then, for the* TWO-CHOICE, *the* $(1+\beta)$*-process and the* QUANTILE(δ) *process with constant $\delta > 0, \beta \in (0,1)$, we have that there exists a constant $\kappa := \kappa(S) > 0$ such that for any step $m \geqslant 0$ being a multiple of b,*

$$\mathbf{Pr}\left[\max_{i \in [n]} y_i^m \leqslant \kappa \cdot \left(\frac{b}{n} + \log n\right)\right] \geqslant 1 - n^{-2}.$$

Theorem 7.35. *Consider the* WEIGHTED b-BATCHED *setting with any $b \in [2n \log n, n^3]$ and weights from a* FINITE-MGF(S) *distribution with constant $S \geqslant 1$. Further let $\epsilon = \sqrt{(n/b) \log n}$. Consider any* SEQUENTIAL(q^t) *process with probability allocation vector q^t satisfying condition $\mathcal{C}_1$ for constant $\delta \in (0,1)$ and ϵ as well as condition $\mathcal{C}_3$ for $C = 1 + \epsilon$, at every step $t \geqslant 0$. Then, there exists a constant $\kappa := \kappa(\delta, S) > 0$, such that for any step $m \geqslant 0$ being a multiple of b,*

$$\mathbf{Pr}\left[\max_{i \in [n]} y_i^m \leqslant \kappa \cdot \sqrt{\frac{b}{n} \cdot \log n}\right] \geqslant 1 - n^{-2}.$$

The following corollary gives concrete instances of processes that give asymptotically better bounds for the b-BATCHED setting than TWO-CHOICE.

Corollary 7.36. *Consider the* WEIGHTED b-BATCHED *setting with any $b \in [2n \log n, n^3]$ and weights from a* FINITE-MGF(S) *distribution with constant $S \geqslant 1$. Then, for the* $(1+\beta)$ *and the* η-MIXED(QUANTILE$(1/2)$, ONE-CHOICE) *processes with $\beta = \eta = \sqrt{(n/b) \cdot \log n}$, we have that there exists a constant $\kappa := \kappa(S) > 0$ such that for any step $m \geqslant 0$ being a multiple of b,*

$$\mathbf{Pr}\left[\max_{i \in [n]} y_i^m \leqslant \kappa \cdot \sqrt{\frac{b}{n} \cdot \log n}\right] \geqslant 1 - n^{-2}.$$

The proofs of these two theorems (Theorems 7.35 and 7.39) are quite similar, so we only present the details for Theorem 7.39 and refer the reader to [114] for the proof of Theorem 7.35.

7.5.1 Proof outline of Theorem 7.39

There are two key steps in the proof of Theorem 7.39:

Step 1: Similarly to the analysis in Chapter 4 and Chapter 6, we will use two instances of the hyperbolic cosine potential, in order to show that it is concentrated at $\mathcal{O}(n)$. More specifically, we will be using $\Gamma_1 := \Gamma_1(\gamma_1)$ (defined in Eq. (3.1)) with the smoothing parameter $\gamma_1 := \frac{\epsilon\delta}{40 \cdot C^2 \cdot S^2} \cdot \min\left\{\frac{1}{\log n}, \frac{n}{b}\right\}$ and $\Gamma_2 := \Gamma_2(\gamma_2)$ with $\gamma_2 := \frac{\gamma_1}{8 \cdot 30}$, i.e., with a smoothing parameter which is a large constant factor smaller than γ_1. So, in particular $\Gamma_2^t \leqslant \Gamma_1^t$ at any step $t \geqslant 0$. Also, note that by varying $b \in [n, n \log n]$, both smoothing factors do not change, but this will not affect the upper bound, as we shall see below.

In the following lemma, proven in Appendix D.5, we show that w.h.p. $\Gamma_2 = \mathcal{O}(n)$ for $\log^3 n$ consecutive batches.

Lemma 7.37 (Restatement of Lemma D.8). *Consider any process satisfying the conditions in Theorem 7.39. Let $\widetilde{c} := 2 \cdot \frac{8c}{\delta}$ where $c := c(\delta) > 0$ is the constant from Theorem 3.2. Then, for any step $t \geqslant 0$ being a multiple of b,*

$$\mathbf{Pr}\left[\bigcap_{j \in [0, \log^3 n]} \left\{\Gamma_2^{t+j\cdot b} \leqslant \widetilde{c} \cdot n\right\}\right] \geqslant 1 - n^{-3}.$$

The proof follows the usual interplay between the two hyperbolic cosine potentials, in that conditioning on $\Gamma_1^t = \operatorname{poly}(n)$ implies that $\Delta\Gamma_2^{t+1} = \mathcal{O}(\frac{n}{b} \cdot n^{1/4})$ (Lemma D.10 (ii)). This in turn allows us to apply a bounded difference inequality to prove concentration for Γ_2. In contrast to Chapter 4 and Section 6.3, here we need a slightly different concentration inequality Theorem B.12, as in a single batch the load of a bin may change by a large amount (with small probability). The complete proof is given in Appendix D.5.

Step 2: Consider an arbitrary step $s = t + j \cdot b$ where $\{\Gamma_2^s \leqslant \widetilde{c} \cdot n\}$ holds. Then, the number of bins i with load y_i^s at least $z := \frac{1}{\gamma_2} \cdot \log(\widetilde{c}/\delta) = \Theta(\max\{b/n, \log n\})$ is at most $\widetilde{c}n \cdot e^{-\gamma_2 z} = \delta n$. With this in mind, we define the following potential function for any step $t \geqslant 0$, which only takes into account bins that are overloaded by at least z balls:

$$\Lambda^t := \Lambda^t(\lambda, z) := \sum_{i : y_i^t \geqslant z} \Lambda_i^t := \sum_{i : y_i^t \geqslant z} e^{\lambda \cdot (y_i^t - z)},$$

where $\lambda := \min\left\{\frac{\epsilon}{4CS}, \frac{n \log n}{b}\right\}$. This means that when $\{\Gamma_2^s \leqslant \widetilde{c} \cdot n\}$ holds, the probability of allocating to one of these bins is $q_i^s \leqslant \frac{1-\epsilon}{n}$, because of the condition $\mathcal{C}_1$. Hence, the potential drops in expectation over one batch (Lemma 7.38) and this means that w.h.p. $\Lambda^m = \operatorname{poly}(n)$, which implies that $\operatorname{Gap}(m) = \mathcal{O}(z + \lambda^{-1} \cdot \log n) = \mathcal{O}(b/n + \log n)$ gap.

7.5.2 Completing the proof of Theorem 7.39

We will now show that when $\Gamma_2^t = \mathcal{O}(n)$, the stronger potential function Λ^t drops in expectation. This will allow us to prove that $\Lambda^m = \operatorname{poly}(n)$ and deduce that w.h.p. $\operatorname{Gap}(m) = \mathcal{O}(b/n + \log n)$.

Lemma 7.38. *Consider any process satisfying the conditions in Theorem 7.39. Let $\widetilde{c} := 2 \cdot \frac{8c}{\delta}$ where $c := c(\delta) > 0$ is the constant from Theorem 3.2. For any step $t \geqslant 0$ being a multiple of b,*

$$\mathbf{E}\left[\Lambda^{t+b} \,\middle|\, \mathfrak{F}^t, \Gamma_2^t \leqslant \widetilde{c} \cdot n\right] \leqslant \Lambda^t \cdot e^{-\frac{\lambda\epsilon}{2n} \cdot b} + n \cdot e^{\frac{C\lambda}{n} \cdot b}.$$

Proof. Consider an arbitrary step $t \geqslant 0$ being a multiple of b and consider a sorted labelling of the bins. Assuming that $\{\Gamma_2^t \leqslant \widetilde{c} \cdot n\}$ holds, the number of bins with load $y_i^t \geqslant z$ is at most

$$\widetilde{c} \cdot n \cdot e^{-\gamma_2 \cdot z} = \widetilde{c} \cdot n \cdot e^{-\log(\widetilde{c}/\delta)} = \delta \cdot n.$$

For any bin $i \in [n]$ with $y_i^t \geqslant z$, we get as in Eq. (3.18) (using that $\lambda \leqslant 1$),

$$\mathbf{E}\left[\, \Lambda_i^{t+b} \,\middle|\, \mathfrak{F}^t \,\right] \leqslant \Lambda_i^t \cdot \left(1 + \left(q_i^t - \frac{1}{n}\right) \cdot \lambda + 2 \cdot q_i^t \cdot S\lambda^2\right)^b.$$

The upper bound on $\mathbf{E}\left[\, \Lambda^{t+b} \,\middle|\, \mathfrak{F}^t, \Gamma_2^t \leqslant \widetilde{c} \cdot n \,\right]$ is maximised when $q_i^t = \frac{1-\epsilon}{n}$, using Lemma B.2 since there are at most δn such bins (i.e., $i \leqslant \delta n$). So,

$$\begin{aligned}\sum_{i : y_i^t \geqslant z} \mathbf{E}\left[\, \Lambda_i^{t+b} \,\middle|\, \mathfrak{F}^t, \Gamma_2^t \leqslant \widetilde{c} \cdot n \,\right] &\overset{(a)}{\leqslant} \sum_{i : y_i^t \geqslant z} \Lambda_i^t \cdot \left(1 - \frac{\lambda\epsilon}{n} + 2 \cdot C \cdot S \cdot \frac{\lambda^2}{n}\right)^b \\ &\overset{(b)}{\leqslant} \sum_{i : y_i^t \geqslant z} \Lambda_i^t \cdot \left(1 - \frac{\lambda\epsilon}{2n}\right)^b \overset{(c)}{\leqslant} \sum_{i : y_i^t \geqslant z} \Lambda_i^t \cdot e^{-\frac{\lambda\epsilon}{2n} \cdot b},\end{aligned}$$

using in (a) that $q_i^t \leqslant \frac{C}{n}$, in (b) that $\lambda \leqslant \frac{\epsilon}{4CS}$ and in (c) that $1 + v \leqslant e^v$ for any v. For the rest of the bins with $i > \delta n$,

$$\begin{aligned}\sum_{i : y_i^t < z} \mathbf{E}\left[\, \Lambda_i^{t+b} \,\middle|\, \mathfrak{F}^t \,\right] &\leqslant \sum_{i : y_i^t < z} \Lambda_i^t \cdot \left(1 + \left(q_i^t - \frac{1}{n}\right) \cdot \lambda + 2 \cdot q_i^t \cdot S\lambda^2\right)^b \\ &\overset{(a)}{\leqslant} \sum_{i : y_i^t < z} \Lambda_i^t \cdot \left(1 + \frac{C}{n} \cdot \lambda - \frac{1}{n} \cdot \lambda + 2 \cdot \frac{C}{n} \cdot S\lambda^2\right)^b \\ &\overset{(b)}{\leqslant} \sum_{i : y_i^t < z} \Lambda_i^t \cdot \left(1 + \frac{C\lambda}{n}\right)^b \overset{(c)}{\leqslant} \sum_{i : y_i^t < z} \left(1 + \frac{C\lambda}{n}\right)^b \overset{(d)}{\leqslant} \sum_{i : y_i^t < z} e^{\frac{C\lambda}{n} \cdot b},\end{aligned}$$

using in (a) that $q_i^t \leqslant \frac{C}{n}$, in (b) that $\lambda \leqslant \frac{\epsilon}{4CS}$, in (c) that $\Lambda_i^t \leqslant 1$ (as there are at most δn bins with normalised load at least z) and in (d) that $1 + v \leqslant e^v$ for any v.

Aggregating the contributions over all bins,

$$\mathbf{E}\left[\, \Lambda^{t+b} \,\middle|\, \mathfrak{F}^t, \Gamma_2^t \leqslant \widetilde{c} \cdot n \,\right] \leqslant \sum_{i : y_i^t \geqslant z} \Lambda_i^t \cdot e^{-\frac{\lambda\epsilon}{2n} \cdot b} + \sum_{i : y_i^t < z} e^{\frac{C\lambda}{n} \cdot b} \leqslant \Lambda^t \cdot e^{-\frac{\lambda\epsilon}{2n} \cdot b} + n \cdot e^{\frac{C\lambda}{n} \cdot b}. \qquad \square$$

Theorem 7.39. *Consider any* SEQUENTIAL(q^t) *process with q^t satisfying condition $\mathcal{C}_1$ for constant $\delta \in (0,1)$ and constant $\epsilon \in (0,1)$ as well as condition $\mathcal{C}_2$ for some constant $C > 1$, at every step $t \geqslant 0$. Further, consider the* WEIGHTED b-BATCHED *setting with any $n \leqslant b \leqslant n^3$ and weights from a* FINITE-MGF(S) *distribution with constant $S \geqslant 1$. Then, there exists a constant $\kappa := \kappa(\delta, \epsilon, C, S) > 0$, such that for any step $m \geqslant 0$ being a multiple of b,*

$$\mathbf{Pr}\left[\, \max_{i \in [n]} y_i^m \leqslant \kappa \cdot \left(\frac{b}{n} + \log n\right) \right] \geqslant 1 - n^{-2}.$$

Proof. Consider first the case when $m \geqslant b \cdot \log^3 n$. Let $t_0 = m - b \cdot \log^3 n$. Let $\mathcal{E}^t := \{\Gamma_2^t \leqslant \widetilde{c} \cdot n\}$. Then using Lemma D.8,

$$\mathbf{Pr}\left[\, \bigcap_{j \in [0, \log^3 n]} \mathcal{E}^{t_0 + j \cdot b} \right] \geqslant 1 - n^{-3}. \tag{7.21}$$

We define the killed potential $\widetilde{\Lambda}$, with $\widetilde{\Lambda}^{t_0} := \Lambda^{t_0}$ and for $j > 0$,

$$\widetilde{\Lambda}^{t_0+j\cdot b} := \Lambda^{t_0+j\cdot b} \cdot \mathbf{1}_{\cap_{s\in[0,j]}\mathcal{E}^{t_0+s\cdot b}}.$$

By Lemma 7.38 for $t = t_0 + j \cdot b$, we have that

$$\mathbf{E}\Big[\, \widetilde{\Lambda}^{t_0+(j+1)\cdot b} \;\Big|\; \mathfrak{F}^{t_0+j\cdot b}, \mathcal{E}^{t_0+s\cdot b} \,\Big] \leqslant \widetilde{\Lambda}^{t_0+j\cdot b} \cdot e^{-\frac{\lambda\epsilon}{2n}\cdot b} + n \cdot e^{\frac{C\lambda}{n}\cdot b}.$$

When $\mathcal{E}^{t_0+j\cdot b}$ does not hold, it follows deterministically that $\widetilde{\Lambda}^{t_0+(j+1)\cdot b} = \widetilde{\Lambda}^{t_0+j\cdot b} = 0$. Hence, we have the following unconditional drop inequality

$$\mathbf{E}\Big[\, \widetilde{\Lambda}^{t_0+(j+1)\cdot b} \;\Big|\; \mathfrak{F}^{t_0+j\cdot b} \,\Big] \leqslant \widetilde{\Lambda}^{t_0+j\cdot b} \cdot e^{-\frac{\lambda\epsilon}{2n}\cdot b} + n \cdot e^{\frac{C\lambda}{n}\cdot b}. \tag{7.22}$$

Assuming $\mathcal{E}^{t_0}$ holds, we have

$$\max_{i\in[n]} y_i^{t_0} \leqslant \frac{1}{\gamma_2} \cdot (\log \widetilde{c} + \log n) \leqslant \frac{2}{\gamma_2} \cdot \log n,$$

for sufficiently large n. Recalling that $\gamma_2 = \Theta(\lambda \cdot \log n)$, there exists a constant $\kappa_1 > 0$ such that

$$\widetilde{\Lambda}^{t_0} \leqslant n \cdot e^{\lambda \cdot y_1^{t_0}} \leqslant e^{\kappa_1 \log^2 n}.$$

Applying Lemma B.1 to Eq. (7.22) with $a := e^{-\frac{\lambda\epsilon}{2n}\cdot b}$ and $b := n \cdot e^{\frac{C\lambda}{n}\cdot b}$ for $\log^3 n$ steps,

$$\mathbf{E}\left[\, \widetilde{\Lambda}^{m} \;\middle|\; \mathfrak{F}^{t_0}, \widetilde{\Lambda}^{t_0} \leqslant e^{\kappa_1 \log^2 n} \,\right] \leqslant e^{\kappa_1 \log^2 n} \cdot a^{\log^3 n} + \frac{n \cdot e^{\frac{C\lambda}{n}\cdot b}}{1-a} \overset{(a)}{\leqslant} 1 + 1.5 \cdot n \cdot e^{\frac{C\lambda}{n}\cdot b} \leqslant 2 \cdot n \cdot e^{\frac{C\lambda}{n}\cdot b} \overset{(b)}{\leqslant} 2 \cdot n^{1+\kappa_2}. \tag{7.23}$$

using in (a) that $\frac{\lambda\epsilon}{2n} \cdot b = \Omega(1)$ and a is a constant < 1 and in (b) that $\frac{C\lambda}{n} \cdot b \leqslant \kappa_2 \cdot \log n$ for some constant $\kappa_2 > 0$, since $\lambda = \min\left\{\frac{\epsilon}{4CS}, \frac{n \log n}{b}\right\}$.

By Markov's inequality, we have

$$\mathbf{Pr}\left[\, \widetilde{\Lambda}^{m} \leqslant 2 \cdot n^{4+\kappa_2} \;\middle|\; \mathfrak{F}^{t_0}, \widetilde{\Lambda}^{t_0} \leqslant e^{\kappa_1 \log^2 n} \,\right] \geqslant 1 - n^{-3}.$$

Hence, by Eq. (7.21),

$$\mathbf{Pr}\Big[\, \widetilde{\Lambda}^{m} \leqslant 2 \cdot n^{4+\kappa_2} \,\Big] = \mathbf{Pr}\Big[\, \widetilde{\Lambda}^{m} \leqslant 2 \cdot n^{4+\kappa_2} \;\Big|\; \mathcal{E}^{t_0} \,\Big] \cdot \mathbf{Pr}\Big[\, \mathcal{E}^{t_0} \,\Big] \geqslant \left(1 - n^{-3}\right) \cdot \left(1 - n^{-3}\right) \geqslant 1 - 2n^{-3}. \tag{7.24}$$

Combining Eq. (7.21) and Eq. (7.24), we have

$$\mathbf{Pr}\Big[\, \Lambda^{m} \leqslant 2 \cdot n^{4+\kappa_2} \,\Big] \geqslant \mathbf{Pr}\left[\, \left\{\widetilde{\Lambda}^{m} \leqslant 2 \cdot n^{4+\kappa_2}\right\} \cap \bigcap_{j\in[0,\log^3 n]} \mathcal{E}^{t_0+j\cdot b} \,\right] \geqslant 1 - 2n^{-3} - n^{-3} \geqslant 1 - n^{-2}.$$

Finally, $\{\Lambda^m \leqslant 2 \cdot n^{4+\kappa_2}\}$ implies that

$$\max_{i\in[n]} y_i^{m} \leqslant z + \frac{\log 2}{\lambda} + \frac{1}{\lambda} \cdot (4 + \kappa_2) \cdot \log n = \mathcal{O}(b/n + \log n),$$

since $\lambda = \min\left\{\frac{\epsilon}{4CS}, \frac{n \log n}{b}\right\} = \Theta(\max\{b/n, \log n\})$, so the claim follows.

For the case when $m < b \cdot \log^3 n$, note that $\widetilde{\Lambda}^{t_0} \leqslant n$ deterministically, which is a stronger starting point in Eq. (7.23) to prove that $\mathbf{E}[\, \Lambda^m \,] \leqslant 2 \cdot n^{1+\kappa_2}$, which in turn implies the gap bound. $\square$

7.6 Graphical setting for Two-Choice

7.6.1 Graphical in the Weighted b-Batched setting

In [152], the authors proved bounds on the gap for the $(1+\beta)$-process (in the setting without batches) where balls are sampled from a Finite-MGF(ζ) distribution with constant $\zeta > 0$. Then, they used a majorisation argument to deduce gap bounds for the Graphical setting for the Two-Choice process (from here onwards referred to just as Graphical). However, due to the involved majorisation argument not working for weights, all results for graphical allocation in [152] assume balls are unweighted. This lack of results for weighted graphical allocations is summarised as [152, Open Question 1]. By leveraging the results in previous sections, we are able to fill this "gap".

For a d-regular (and connected) graph G, let us define the **conductance** as:

$$\phi(G) := \min_{S \subseteq V \,:\, 1 \leqslant |S| \leqslant n/2} \frac{|E(S, V \setminus S)|}{|S| \cdot d},$$

where $|E(S, V \setminus S)|$ counts (once) the edges between the sets S and $V \setminus S$. We will call a family of graphs an *expander*, if ϕ is at least a constant bounded below from 0 (as $n \to \infty$).

Lemma 7.40. *Consider Graphical on a d-regular graph with conductance ϕ. Then, in any step $t \geqslant 0$, the sorted allocation vector $\widetilde{q}^t$ satisfies for all $1 \leqslant k \leqslant n/2$,*

$$\sum_{i=1}^{k} \widetilde{q}_i^t \leqslant (1-\phi) \cdot \frac{k}{n},$$

and similarly, for any $n/2 + 1 \leqslant k \leqslant n$,

$$\sum_{i=k}^{n} \widetilde{q}_i^t \geqslant (1+\phi) \cdot \frac{n-k+1}{n}.$$

Further, $\max_{i \in [n]} \widetilde{q}_i^t \leqslant \frac{2}{n}$. Thus, the vector $\widetilde{q}^t$ satisfies condition $\mathcal{C}_1$ with $\delta = 1/2$, $\epsilon = \phi$ and condition $\mathcal{C}_2$ with $C = 2$.

The proof of this lemma closely follows [152, Proof of Theorem 3.2].

Proof. Fix any load vector x^t in step t. Consider any $1 \leqslant k \leqslant n/2$. Let S_k be the k bins with the largest load. Hence in order to allocate a ball into S_k, both endpoints of the sampled edge must be in S_k, and so

$$\sum_{i=1}^{k} q_i^t = \frac{2 \cdot |E(S_k, S_k)|}{2 \cdot |E|} \leqslant \frac{d \cdot k - \phi \cdot d \cdot k}{n \cdot d} = (1-\phi) \cdot \frac{k}{n},$$

where the inequality used that $d \cdot |S_k| = |E(S_k, V \setminus S_k)| + 2 \cdot |E(S_k, S_k)|$ and the definition of conductance ϕ. Now, we will consider the suffix sums for $n/2 + 1 \leqslant k \leqslant n$. We start by upper bounding the prefix sum up to $k-1$,

$$\sum_{i=1}^{k-1} \widetilde{q}_i^t \leqslant \frac{|S_{k-1}| \cdot d - |V \setminus S_{k-1}| \cdot \phi \cdot d}{nd} \leqslant \frac{(k-1) \cdot d - (n-k+1) \cdot \phi \cdot d}{nd} = \frac{(k-1) - (n-k+1) \cdot \phi}{n},$$

where the inequality used our assumption that G has conductance ϕ. Therefore, we lower bound the suffix sum by

$$\sum_{i=k}^{n} \widetilde{q}_i^t = 1 - \sum_{i=1}^{k-1} \widetilde{q}_i^t \geqslant 1 - \frac{(k-1)-(n-k+1)\cdot\phi}{n} = (1+\phi)\cdot\frac{n-k+1}{n}.$$

Hence, $\widetilde{q}^t$ satisfies condition $\mathcal{C}_1$ with $\epsilon = \phi$.

Finally, we also know that $q_i^t \leqslant \frac{d}{d\cdot n/2} = \frac{2}{n}$, for any bin $i \in [n]$, since in the worst-case we allocate a ball to bin i whenever one of its d incident edges are chosen. $\square$

The next result is for the non-batched GRAPHICAL setting.

Theorem 7.41. *Consider* GRAPHICAL *on a d-regular graph with conductance $\phi > 0$. Further, assume that balls are sampled from a* FINITE-MGF(S) *distribution with $S > 1$. Then, there exists a constant $\kappa > 0$ such that for any step $m \geqslant 0$,*

$$\mathbf{Pr}\left[\max_{i\in[n]} |y_i^m| \leqslant \kappa \cdot S \cdot \frac{\log n}{\phi}\right] \geqslant 1 - n^{-2}.$$

Proof. By Lemma 3.5, we have that for the potentials $\Phi := \Phi(\gamma)$ and $\Psi := \Psi(\gamma)$ with $\gamma := \frac{\phi}{32S}$,

$$\mathbf{E}\left[\, \Delta\Phi_i^{t+1} \,\middle|\, x^t \,\right] \leqslant \Phi_i^t \cdot \left(\left(q_i^t - \frac{1}{n}\right)\cdot\gamma + 4S\cdot\frac{\gamma^2}{n}\right),$$

and

$$\mathbf{E}\left[\, \Delta\Psi_i^{t+1} \,\middle|\, x^t \,\right] \leqslant \Psi_i^t \cdot \left(\left(\frac{1}{n} - q_i^t\right)\cdot\gamma + 4S\cdot\frac{\gamma^2}{n}\right).$$

Hence, applying Theorem 3.2 for $\epsilon := \phi$, $\delta := 1/2$, we get for the potential $\Gamma := \Gamma(\gamma)$ and any step $m \geqslant 0$,

$$\mathbf{E}\left[\, \Gamma^m \,\right] \leqslant \frac{8c}{\delta}\cdot n,$$

for some constant $c := c(\delta) > 0$. Hence, by Markov's inequality

$$\mathbf{Pr}\left[\Gamma^m \leqslant \frac{8c}{\delta}\cdot n^3\right] \geqslant 1 - n^{-2}.$$

The event $\{\Gamma^m \leqslant \frac{8c}{\delta}\cdot n^3\}$ implies that

$$\max_{i\in[n]} |y_i^m| \leqslant \log\left(\frac{8c}{\delta}\right) + 3\cdot\frac{32S}{\phi}\cdot\log n = \mathcal{O}\left(S\cdot\frac{\log n}{\phi}\right). \qquad \square$$

The next result is a Theorem 7.41 which applies for the b-BATCHED GRAPHICAL setting.

Theorem 7.42. *Consider* GRAPHICAL *on a d-regular graph with conductance $\phi > 0$. Further, consider the batched setting with $b \geqslant n$ and assume that balls are sampled from a* FINITE-MGF(ζ) *distribution with constant $\zeta > 0$. Then, there exists a constant $\kappa := \kappa(\zeta) > 0$ such that it holds for any step $m \geqslant 0$ being a multiple of b,*

$$\mathbf{Pr}\left[\max_{i\in[n]} |y_i^m| \leqslant \kappa\cdot\frac{b}{n}\cdot\frac{\log n}{\phi}\right] \geqslant 1 - n^{-2}.$$

Further, if the conductance ϕ is lower bounded by a constant > 0 (i.e., G is an expander), and $n \leqslant b \leqslant n^3$, then there exists a constant $\kappa := \kappa(\zeta) > 0$ such that for any $m \geqslant 0$ being a multiple of b,

$$\mathbf{Pr}\left[\, y_1^m \leqslant \kappa\cdot\left(\frac{b}{n} + \log n\right)\right] \geqslant 1 - n^{-2}.$$

Note that our first gap bound generalises [152, Theorem 3.2], which is a gap bound of $\mathcal{O}(\frac{\log n}{\phi})$ in the setting without batches and weights. Similarly, our second result extends the $\mathcal{O}(\log n)$ bound from [152] for expanders, and proves that the same gap bound applies in the WEIGHTED b-BATCHED setting with any $b = \mathcal{O}(n \log n)$.

Proof. The first result follows directly from Lemma 7.40 and Theorem 3.14. For the second result, $\epsilon = \phi$ is a constant > 0, and we can apply the refined gap bound from Theorem 7.39. □

7.6.2 GRAPHICAL on dense expanders

We now analyse the GRAPHICAL setting on dense expander graphs. To this end, we first recall some basic notation of spectral graph theory and expansion. For an undirected graph G, the normalised Laplacian Matrix of G is an $n \times n$-matrix defined by

$$\mathbf{L} = \mathbf{I} - \mathbf{D}^{-1/2} \cdot \mathbf{A} \cdot \mathbf{D}^{1/2},$$

where $\mathbf{I}$ is the identity matrix, $\mathbf{A}$ is the adjacency matrix and $\mathbf{D}$ is the diagonal matrix where $D_{u,u} = \deg(u)$ for any vertex $u \in V$. Further, let $\lambda_1 \leqslant \lambda_2 \leqslant \cdots \leqslant \lambda_n$ be the n eigenvalues of $\mathbf{L}$, and let $\lambda := \max_{i \in [2,n]} |1 - \lambda_i|$ be the spectral expansion of G. Further, for any set $U \subseteq V$ define $\operatorname{vol}(U) := \sum_{v \in U} \deg(v)$. Note that for a d-regular graph, we have $\operatorname{vol}(U) = d \cdot |U|$ and $\operatorname{vol}(V) = dn$.

We now recall the following (stronger) version of the Expander Mixing Lemma (cf. [52]):

Lemma 7.43 (Expander Mixing Lemma). *For any subsets $X, Y \subseteq V$,*

$$\left| |2E(X,Y)| - \frac{\operatorname{vol}(X) \cdot \operatorname{vol}(Y)}{\operatorname{vol}(V)} \right| \leqslant \lambda \cdot \frac{\sqrt{\operatorname{vol}(X) \cdot \operatorname{vol}(\overline{X}) \cdot \operatorname{vol}(Y) \cdot \operatorname{vol}(\overline{Y})}}{\operatorname{vol}(V)},$$

where $\operatorname{vol}(\overline{X}) = \operatorname{vol}(V \setminus X)$.

In the following, we consider G to be a d-regular graph.

Proposition 7.44. *Consider GRAPHICAL on a d-regular graph G with spectral expansion λ. Then, at any step $t \geqslant 0$, the sorted allocation vector $\widetilde{q}^t$ satisfies the following three inequalities.*

(i) For any $1 \leqslant k \leqslant \lambda \cdot n$,

$$\sum_{i=1}^{k} \widetilde{q}_i^t \leqslant 2\lambda \cdot \frac{k}{n}.$$

(ii) For any $\lambda \cdot n \leqslant k$,

$$\sum_{i=1}^{k} \widetilde{q}_i^t \leqslant 2 \cdot \left(\frac{k}{n}\right)^2.$$

(iii) For any $1 \leqslant k \leqslant n$,

$$\sum_{i=1}^{k} \widetilde{q}_i^t \leqslant \frac{k}{n} \cdot \left(1 - (1-\lambda) \cdot \frac{n-k}{n}\right).$$

Proof. First statement. Fix $1 \leqslant k \leqslant \lambda \cdot n$. and let S_k be the k bins with the largest load. Using Lemma 7.43 for $X = Y = S_k$:

$$2|E(S_k, S_k)| \leqslant d \cdot \frac{|S_k| \cdot |S_k|}{n} + \lambda \cdot \frac{d|S_k|(n - |S_k|)}{n} \leqslant d \cdot \frac{|S_k| \cdot |S_k|}{n} + \lambda d \cdot |S_k|.$$

Since $|S_k| \leqslant \lambda \cdot n$, we conclude that

$$2|E(S_k, S_k)| \leqslant 2\lambda d \cdot |S_k|.$$

Note that in GRAPHICAL, we allocate a ball to one of the k bins with the largest load if and only if we sample an edge in $E(S_k, S_k)$. Using this and the upper bound on $|E(S_k, S_k)|$ from above, it follows that

$$\sum_{i=1}^{|S_k|} \widetilde{q}_i^t = \frac{2|E(S_k, S_k)|}{2|E|} \leqslant \frac{2\lambda d |S_k|}{nd} = 2\lambda \cdot \frac{|S_k|}{n}.$$

Second statement. Consider now the case where $\lambda n \leqslant |S_k|$. Then,

$$2|E(S_k, S_k)| \leqslant d \cdot \frac{|S_k| \cdot |S_k|}{n} + \lambda d \cdot \frac{|S_k| \cdot (n - |S_k|)}{n} \leqslant d \cdot \frac{|S_k| \cdot |S_k|}{n} + d \cdot \frac{|S_k| \cdot |S_k|}{n} \cdot \frac{n - |S_k|}{n} \leqslant 2d \cdot \frac{|S_k| \cdot |S_k|}{n}.$$

and therefore,

$$\sum_{i=1}^{|S_k|} \widetilde{q}_i^t = \frac{2|E(S_k, S_k)|}{2|E|} \leqslant \frac{2d \cdot \frac{|S_k| \cdot |S_k|}{n}}{nd} = 2 \cdot \left(\frac{|S_k|}{n} \right)^2.$$

Third statement. Finally, consider the general case where $1 \leqslant |S_k| \leqslant n$. Then using Lemma 7.43,

$$2|E(S_k, S_k)| \leqslant d \cdot \frac{|S_k| \cdot |S_k|}{n} + \lambda d \cdot \frac{|S_k|(n - |S_k|)}{n} = d \cdot |S_k| \cdot \left(1 - (1 - \lambda) \cdot \frac{n - |S_k|}{n} \right).$$

and therefore,

$$\sum_{i=1}^{|S_k|} \widetilde{q}_i^t = \frac{2|E(S_k, S_k)|}{2|E|} \leqslant \frac{d|S_k| \cdot \left(1 - (1-\lambda) \cdot \frac{n - |S_k|}{n}\right)}{nd} = \frac{|S_k|}{n} \cdot \left(1 - (1 - \lambda) \cdot \frac{n - |S_k|}{n} \right). \qquad \square$$

Lemma 7.45. *Consider* GRAPHICAL *on a d-regular graph G with spectral expansion $\lambda \leqslant e^{-\frac{1}{4}(\log n)^{(k-1)/k}}$ for some integer $k \geqslant 2$. Then, for any step $t \geqslant 0$, the process is a k-*RELAXED-QUANTILE$_{\gamma,\epsilon}$ *process with $\gamma = 2$ and $\epsilon = 1/12$.*

Proof. Consider the GRAPHICAL setting on the graph G defined in the statement. We will verify that it is a k-RELAXED-QUANTILE$_{\gamma,\epsilon}$ process.

We start by verifying that $\widetilde{q}^t$ satisfies the first condition of k-RELAXED-QUANTILE$_{\gamma,\epsilon}$, i.e., condition $\mathcal{C}_1$ with $\delta = 1/3$ and $\epsilon = 1/12$. First, consider any prefix sum of the sorted allocation vector $\widetilde{q}^t$ for any $j \in [1, n/3]$. Then by Proposition 7.44 (iii),

$$\sum_{i=1}^{j} \widetilde{q}_i^t \leqslant \frac{j}{n} \cdot \left(1 - (1 - \lambda) \cdot \frac{n - j}{n} \right) \leqslant \frac{j}{n} \cdot \left(1 - (1 - \lambda) \cdot \frac{2}{3} \right).$$

Since by assumption $\lambda \leqslant 1/2$, we have $1 - (1 - \lambda) \cdot \frac{2}{3} \leqslant 1 - 1/3 = 1 - 4\epsilon$ for $\epsilon = 1/12$.

Similarly, for any $j \in [(2/3)n, n]$, then by Proposition 7.44 (iii),

$$\sum_{i=j}^{n} \widetilde{q}_i^t = \left(1 - \frac{j}{n} \right) \cdot \left(1 + \frac{j}{n} \cdot (1 - \lambda) \right) \geqslant \left(1 - \frac{j}{n} \right) \cdot \left(1 + \frac{2}{3} \cdot (1 - \lambda) \right),$$

and thus again, since $\lambda \leqslant 1/2$, we have $1 + \frac{2}{3} \cdot (1 - \lambda) \geqslant 1 + 4\epsilon$.

We will now verify the second condition of k-RELAXED-QUANTILE$_{\gamma,\epsilon}$. Let $\delta_1, \ldots, \delta_k$ be the k quantiles in the condition. then, for any $1 \leqslant j \leqslant \delta_1 n$, it follows by Proposition 7.44 (i),

$$\sum_{i=1}^{j} \widetilde{q}_i^t \leqslant 2\lambda \cdot \frac{j}{n} \cdot \leqslant 2\lambda \cdot \frac{\delta_1}{n} \leqslant \gamma \cdot \frac{\delta_1}{n},$$

since $\gamma = 2$ and $\lambda \leqslant 1/2$. Therefore, for any $1 \leqslant j \leqslant \delta_1 n$,

$$(\widetilde{q}_1^t, \ldots, \widetilde{q}_j^t) \preceq \left(\gamma \cdot \frac{\delta_1}{n}, \ldots, \gamma \cdot \frac{\delta_1}{n} \right).$$

Consider now any prefix sum over $\widetilde{q}^t$, where $j \leqslant \delta_{\ell+1} n$ for any $1 \leqslant \ell < k$. Then. by Proposition 7.44 (ii), we have that

$$\sum_{i=1}^{j} \widetilde{q}_i^t \leqslant 2 \cdot \left(\frac{j}{n} \right)^2 \leqslant 2 \cdot \frac{\delta_{\ell+1}}{n} \cdot j \leqslant 2 \cdot \frac{\delta_\ell + \delta_{\ell+1}}{n} \cdot j = \gamma \cdot \frac{\delta_\ell + \delta_{\ell+1}}{n} \cdot j.$$

Therefore, for any $j \leqslant \delta_{\ell+1} n$

$$(\widetilde{q}_1^t, \ldots, \widetilde{q}_j^t) \preceq \left(\gamma \cdot \frac{\delta_\ell + \delta_{\ell+1}}{n}, \ldots, \gamma \cdot \frac{\delta_\ell + \delta_{\ell+1}}{n} \right).$$

This concludes the proof showing that $\widetilde{q}^t$ satisfies the k-RELAXED-QUANTILE$_{\gamma,\epsilon}$ condition, for any step $t \geqslant 0$. □

Theorem 7.46. *Consider* GRAPHICAL *on a d-regular graph G with spectral expansion $\lambda \leqslant 1/2$. Further, let k be the largest integer with $1 \leqslant k \leqslant k_{\max}$ such that $e^{-\frac{1}{4}(\log n)^{(k-1)/k}} \geqslant \widetilde{\lambda} := \max\{\lambda, e^{-\frac{1}{4}(\log n)^{(k_{\max}-1)/k_{\max}}}\}$. Then, there exists a constant $\kappa > 0$ such that for any step $m \geqslant 0$,*

$$\mathbf{Pr}\left[\operatorname{Gap}(m) \leqslant \kappa \cdot k \cdot \left(\frac{\log n}{\log(1/\widetilde{\lambda})} \right)^{(k+1)/k} \right] \geqslant 1 - n^{-3}.$$

We will postpone the proof for the moment, and first state the following bound that follows immediately from the theorem above when λ decays polynomially in n:

Corollary 7.47 (Special case of Theorem 7.46). *Consider* GRAPHICAL *on d-regular graph on G with spectral expansion $\lambda \leqslant n^{-c}$ for some constant $c > 0$. Then there is a constant $\kappa = \kappa(c) > 0$ such that for any $m \geqslant 1$,*

$$\mathbf{Pr}\left[\operatorname{Gap}(m) \leqslant \kappa \cdot \log\log n \right] \geqslant 1 - n^{-3}.$$

Note that $\lambda \leqslant n^{-c_1}$ captures a *relaxed, multiplicative* approximation of Ramanujan graphs (it is in fact more relaxed than the existing notion "weakly Ramanujan"). Recently, [170] proved that for any $\operatorname{poly}(n) \leqslant d \leqslant n/2$, a random d-regular graph satisfies the constraint on λ with probability at least $1 - n^{-1}$.

Further, we remark that the above result in some sense extends one of the main results of [100] which states that for any graph with degree $n^{1/\log\log n}$, graphical balanced allocation achieves a gap of at most $\Theta(\log\log n)$ in the lightly loaded case ($m = n$). Our result above also refines a previous result of [152] which states that for any expander graph, a gap bound of $\mathcal{O}(\log n)$ holds (even in the heavily-loaded case $m \geqslant n$). In conclusion, we see that the gap bound of $\mathcal{O}(\log\log n)$ extends from the complete graph (which is the "original" TWO-CHOICE process) to other graphs, provided we have a strong expansion and high density.

Proof of Theorem 7.46. By Lemma 7.45 the process is an instance of k-RELAXED-QUANTILE$_{\gamma,\epsilon}$ for $\gamma = 2$ and $\epsilon = 1/12$. Therefore, by Theorem 7.20, it follows that there exists a constant $\kappa > 0$ such that

$$\mathbf{Pr}\left[\, \mathrm{Gap}(m) \leqslant \kappa \cdot k \cdot (\log n)^{1/k} \,\right] \geqslant 1 - n^{-3}.$$

By assumption on k, we have

$$-\frac{1}{4} \cdot (\log n)^{(k-1)/k} \geqslant \log(\widetilde{\lambda}), \tag{7.25}$$

but also, as k is chosen as large as possible,

$$-\frac{1}{4} \cdot (\log n)^{k/(k+1)} \leqslant \log(\widetilde{\lambda}).$$

Thus

$$(\log n)^{1/k} \leqslant \left(\frac{\log n}{\log(1/\widetilde{\lambda})} \right)^{(k+1)/k}. \tag{7.26}$$

Applying the gap bound of $\kappa \cdot k \cdot (\log n)^{1/k}$ from Theorem 7.20, and using 7.26 we get the conclusion. $\square$

7.7 MEMORY process

In this section, we state the main results for the MEMORY process. The proofs and omitted details can be found in [118]. Our main result is an upper bound for the HETEROGENEOUS setting, demonstrating that it performs better than TWO-CHOICE, for which the gap diverges even when $a = 10$ and $b = 10$ [176].

Theorem 7.48 ([118, Theorem 1.1]). *Consider the HETEROGENEOUS(MEMORY, $\mathcal{S}$) process where $\mathcal{S}$ is an (a, b)-biased sampling distribution, for arbitrary constants $a, b \geqslant 1$. Then, there exists a constant $\kappa := \kappa(a, b) > 0$ such that for every step $m \geqslant 0$,*

$$\mathbf{Pr}\left[\, \mathrm{Gap}(m) \leqslant \kappa \cdot \log\log n \,\right] \geqslant 1 - n^{-3}.$$

For the case of the uniform sampling distribution, we get a matching lower bound by Theorem C.26.

For the upper bound, the base case follows by the analysis of the RESET-MEMORY process. Then, for the layered induction we proceed as we have done so far for the k-DENSE-QUANTILE and k-DENSE-THRESHOLD process, but several modifications are needed. We outline these modifications below, but for space reasons, we will not include all lemmas.

As a corollary of the analysis, we also get the following upper bound for the $(1, 1, d)$-RESET-MEMORY process.

Theorem 7.49 ([118, Theorem 1.5]). *Consider the $(1, 1, d)$-RESET-MEMORY process for any constant $d \geqslant 1$. Then, there exists a constant $\kappa > 0$, such that for any step $m \geqslant 0$, we have that*

$$\mathbf{Pr}\left[\, \mathrm{Gap}(m) \leqslant \kappa \cdot \log n \,\right] \geqslant 1 - n^{-3}.$$

Full Potentials. We will be using layered induction over super-exponential potential functions, similar to the one used in Section 7.2, but with some differences (see discussion below). We now define the super-exponential potential functions for $1 \leqslant j \leqslant j_{\max} - 1$,

$$\Psi_j^t := \sum_{i=1}^n \Psi_{j,i}^t := \sum_{i=1}^n e^{\alpha_1 \cdot \nu^j \cdot (y_i^t - z_j)^+}, \quad \text{and} \quad \Phi_j^t := \sum_{i=1}^n \Phi_{j,i}^t := \sum_{i=1}^n e^{\alpha_2 \cdot \nu^j \cdot (y_i^t - z_j)^+},$$

where we set

$$z_j := \frac{5\nu}{\alpha_2} \cdot j, \quad \nu := \max\{\log(2Cb), 36b\}, \quad C := \max\{6c, 6\}, \quad j_{\max} = \log_\nu\left(\frac{\alpha_2}{2\nu}\log n\right), \tag{7.27}$$

and $\alpha_1, \alpha_2, c > 0$ are constants with $\alpha_1 = 6 \cdot 14 \cdot \alpha_2$. Our aim is to prove that $\Phi_{j_{\max}-1}^m = \mathcal{O}(n)$, which implies that $\mathrm{Gap}(m) = \mathcal{O}(\log\log n)$.

The folded process. In the j-th layer of the layered induction (for $1 \leqslant j \leqslant j_{\max} - 1$), we analyse the following folded process of which MEMORY is an instance. For this, we group the steps into consecutive rounds (of varying lengths), and refer to the s-th step within the round as substep s. Further, we let $y_i^{r,s}$ be the normalised load of bin i after substep s of round r. Then, we define the *folded process* as follows:

- For each round $r \geqslant 0$, sample bin $i := i(r) \in [n]$ according to the sampling distribution $\mathcal{S}$:
  - **Case A:** If $y_i^{r,0} \geqslant z_{j-1} + \frac{2\nu}{\alpha_2}$, then allocate one ball to an arbitrary bin ℓ with $y_\ell^{r,0} \leqslant y_i^{r,0}$, and proceed to the next round.
  - **Case B:** Otherwise, start a sequence of consecutive phases each consisting of $\frac{\nu}{\alpha_2}$ substeps (that is, each phase $k \geqslant 1$ consists of substeps $s \in [(k-1) \cdot \frac{\nu}{\alpha_2}, k \cdot \frac{\nu}{\alpha_2})$ within the current round r). In each substep s, we sample one bin $i = i(r,s)$ according to $\mathcal{S}$ and allocate one ball to an arbitrary bin ℓ with $y_\ell^{r,s} \leqslant z_{j-1} + \frac{4\nu}{\alpha_2}$. At the end of each phase, we also complete the round if either of the following two conditions hold:
    - ⋆ **Condition 1:** In none of the substeps s of the current phase did we sample a bin ℓ with $y_\ell^{r,s} < z_{j-1} + \frac{2\nu}{\alpha_2}$ at the corresponding substep s.
    - ⋆ **Condition 2:** We have completed $k_j := e^{\nu^{j+1}} \cdot \log^3 n \leqslant n^{1/7}$ phases.

Partial potentials. For the *recovery phase*, i.e., showing that at some *step* s in an interval of n·polylog(n) length we have $\Phi^s \leqslant Cn$, we will need larger drop rates for the potentials, so we will be using the following potentials defined only over the heavy bins

$$\dot{\Psi}_j^t := \sum_{i=1}^n \dot{\Psi}_{j,i}^t := \sum_{i:y_i^t \geqslant z_j} e^{\alpha_1 \cdot \nu^j \cdot (y_i^t - z_j)}, \quad \text{and} \quad \dot{\Phi}_j^t := \sum_{i=1}^n \dot{\Phi}_{j,i}^t := \sum_{i:y_i^t \geqslant z_j} e^{\alpha_2 \cdot \nu^j \cdot (y_i^t - z_j)},$$

for $\alpha_1, \alpha_2, z_j > 0$ defined as above. In contrast to Φ_j (and Ψ_j) which are always $\geqslant n$ (since each bin contributes at least 1), $\dot{\Phi}_j$ (and $\dot{\Psi}_j$) could be as small as 0. Also, $\dot{\Phi}_j^t \leqslant \Phi_j^t \leqslant \dot{\Phi}_j^t + n$ (and $\dot{\Psi}_j^t \leqslant \Psi_j^t \leqslant \dot{\Psi}_j^t + n$).

Potentials over rounds. We also define versions of the Φ_j and Ψ_j potentials indexed by a round $r \geqslant 0$ (note the starting step of the first round may not be equal to 0):

$$\overline{\Psi}_j^r := \sum_{i=1}^n e^{\alpha_1 \cdot \nu^j \cdot (y_i^{r,0} - z_j)^+}, \quad \text{and} \quad \overline{\Phi}_j^r := \sum_{i=1}^n e^{\alpha_2 \cdot \nu^j \cdot (y_i^{r,0} - z_j)^+}.$$

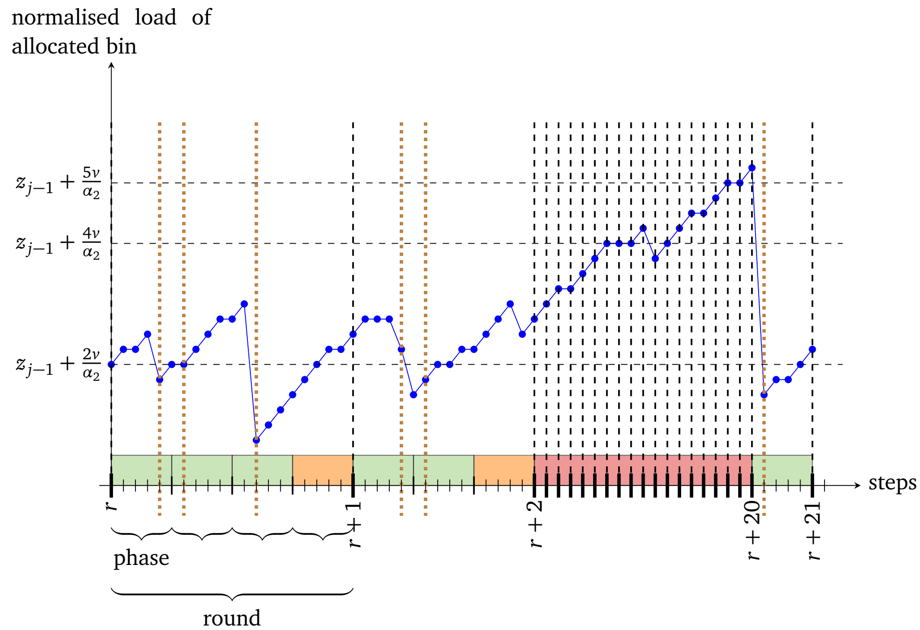


Figure 7.3: The phases and rounds of the folded process. Brown lines indicate the first substep within a phase in which a light bin was sampled (as can be seen in the second phase of round $r+1$, this does not necessarily mean that this bin is going to be used for allocation or as a cache). As shown, it is only possible to allocate to a bin with normalised load above $z_{j-1} + \frac{5\nu}{\alpha_2}$ after a long sequence of red rounds.

Similarly, we define the partial potential functions over rounds

$$\ddot{\Psi}_j^r := \sum_{i : y_i^t \geqslant z_j} e^{\alpha_1 \cdot \nu^j \cdot (y_i^{r,0} - z_j)}, \quad \text{and} \quad \ddot{\Phi}_j^r := \sum_{i : y_i^t \geqslant z_j} e^{\alpha_2 \cdot \nu^j \cdot (y_i^{r,0} - z_j)}.$$

Differences to previous applications: These potentials are similar in form to the ones used in Section 7.2 for $k = \Theta(\log\log n)$. However, the analysis is different as the potentials drop in expectation only when considering a sufficiently long interval (e.g., the folded version of the process). For example, starting from a state where the cache has load at least $z_j + 1$, the potential Φ_j will increase in expectation over one step. Considering rounds consisting of several balls, introduces several challenges:

- **Issue 1:** In each round we could allocate as many as k_j balls, which could be $\Omega(n^\epsilon)$. This would mean, that starting from a round r_0 with $\Phi^{r_0} = \mathcal{O}(e^{0.5 \log^3 n})$ and having a drop inequality similar to that in Lemma 6.2, e.g.,

$$\mathbf{E}\left[\, \overline{\Phi}_j^{r+1} \,\middle|\, \mathfrak{F}^r, \overline{\Phi}_{j-1}^r \leqslant Cn \,\right] \leqslant \overline{\Phi}_j^r \cdot \left(1 - \frac{1}{n}\right) + 2,$$

we may need $\Omega(n \log^3 n)$ rounds to prove that the potential becomes $\mathcal{O}(n)$ in expectation. In these rounds, we could allocate $\Omega(n^{1+\epsilon} \cdot \log^3 n)$ balls (k_j in each round) and so the length of the interval of the entire analysis would need to be $\omega(n \cdot \mathrm{polylog}(n))$. However, it would not be possible to tolerate a $\mathrm{poly}(n)$ probability decrease in each layer, as we have $j_{\max} = \Theta(\log\log n)$ layers.

Solution: Define the potential function $\ddot{\Phi}_j$ over just the bins with normalised load at least z_j. For this potential function, we can show that:

$$\mathbf{E}\left[\, \ddot{\Phi}_j^{r+1} \,\middle|\, \mathfrak{F}^r, \overline{\Phi}_{j-1}^r \leqslant 2Cn \,\right] \leqslant \ddot{\Phi}_j^r \cdot \left(1 - \frac{e^{\nu^{j+1}}}{n}\right) + e^{-\nu^j}.$$

This means that starting from a round r_0 with $\Phi^{r_0} = \mathcal{O}(e^{0.5 \log^3 n})$, we need to wait only for $n \cdot e^{-\nu^{j+1}} \cdot \log^3 n$ rounds, so at most $n \cdot \log^6 n$ steps, for the potential to become $\mathcal{O}(n)$.

- **Issue 2:** Unfortunately, for stabilisation, i.e., showing that $\Phi_j^s = \mathcal{O}(n)$ for $n \cdot \text{polylog}(n)$ steps, we cannot use the partial potential function $\ddot{\Phi}_j$, as it could change by $\Omega(n^{1/3})$ in a single round.

  Consider the case where there are $n \cdot e^{-\nu^j}$ bins (for $j = 1$), whose load is $z_j + \frac{k_j}{n}$. Then in a single round, we could allocate k_j balls only in light bins, so that the potential becomes 0. Hence, the potential decreases by $n \cdot e^{-\nu^j} \cdot e^{\alpha_2 k_j / n} = \Omega(n)$, for $j = 1$. This means that we can no longer apply the concentration inequality, as the bounded difference condition is not strong enough.

  Solution: For this part of the analysis, we use the full potential $\overline{\Phi}_j$ and a stopping time to guarantee that the number of balls allocated at every application of the concentration inequality is at most $n / \log^2 n$. This allows us to apply the smoothness argument to argue that the potential is $\mathcal{O}(n)$ in every step in the interval.

Chapter $7^{1/2}$: DP for balanced allocations

In this chapter, we show a simple application of dynamic programming to balanced allocations processes and some of the insights that can be obtained for small values of n and m.

Exact probabilities. For any Markovian (meaning that $p^t(\mathfrak{F}^t) = p(x)$) and index-independent (meaning that $p(x) = p(\texttt{sorted}(x))$) process we can use the following forward dynamic programming equations to compute the exact probability $Q^t(x) = \mathbf{Pr}[\, x^t = x \,]$ for each sorted load vector x,

$$Q^{t+1}(\texttt{sorted}(x + \mathbf{e}_i)) \mathrel{+}= Q^t(x) \cdot p_i(x), \quad \text{for all } i \in [n],$$

and using that $Q^0(x) = \mathbf{1}_{x=\mathbf{0}}$. Note that when computing Q^{t+1} we only need the values from Q^t, so we can use memoisation.

Having access to the exact probabilities means that we can plot the distribution of any function $f : \mathbb{N}^n \to \mathbb{R}$ of the load vector, such as the gap, minimum load, the ℓ_2-norm of the load vector and so on (see Fig. 7.5 and Table 7.4). In addition, we can compute for instance the expectation of these quantities or other properties of these random variables.

Process/Parameters	$m = n = 10$	$m = n = 20$	$m = n = 30$
One-Choice	2.748	3.231	3.492
$(1 + \beta)$ for $\beta = 1/2$	2.369	2.699	2.913
Median-Quantile	2.260	2.542	2.743
Mean-Thinning	2.242	2.550	2.763
Quantile(δ^*)	2.168	2.348	2.497
Two-Choice	2.061	2.152	2.224
Memory	2.209	2.427	2.586

Table 7.4: Exact computation (to 4 significant figures) of $\mathbf{E}[\, \mathrm{Gap}(m) \,]$ for $m = n$ and $n \in \{10, 20, 30\}$ using dynamic programming.

This confirms the empirical observation in [136, Section 5] that for small values of n, Two-Choice has a smaller gap than Memory. We also confirm that is still the case for some values of $m \geqslant n$ in Table 7.7.

Computing optimal parameters. Using a similar dynamic programming computation, we compute the exact expected gap for the Memory* process, an optimal version of Memory which can look at the entire load vector and also decide whether to allocate to a cache (and which one). In particular, the first two columns in Table 7.8 show that it is not always optimal to allocate to the cache, even if the process has exactly one cache. The third and fourth column shows the improvement for two caches and raises the question whether the improvement is just in the lower order terms or whether it leads to a gap that is $o(\log \log n)$.

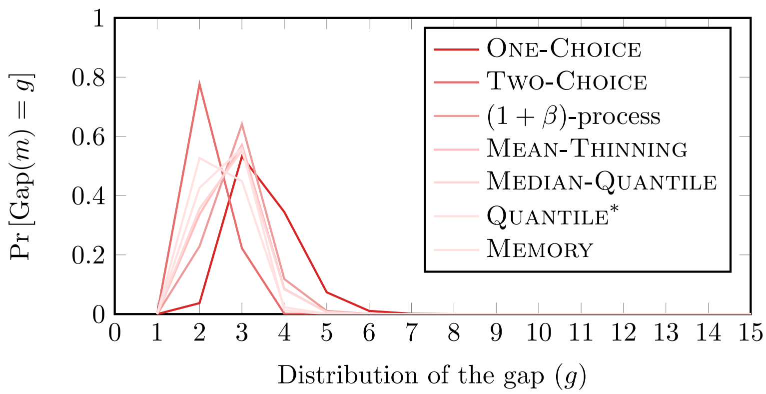


Figure 7.5: Exact probability distribution for Gap(m) for various processes for $m = n = 30$.

	TWO-CHOICE	(1, 1)-MEMORY	(1, 2)-MEMORY	(1, 3)-MEMORY
10	2.061	2.209	2.053	1.996
20	2.152	2.427	2.168	2.066
30	2.224	2.586	2.266	2.119

Table 7.6: Exact computation of $\mathbf{E}[\,\text{Gap}(m)\,]$ for $m = n$ and $n \in \{10, 20, 30\}$ using dynamic programming. The table shows that for small values of n, TWO-CHOICE is better than MEMORY with $d = 1$, sometimes better than MEMORY with $d = 2$, but worse than MEMORY with $d = 3$.

	$m = n$	$m = 2n$	$m = 3n$	$m = 4n$
TWO-CHOICE ($n = 10$)	2.061	2.132	2.168	2.186
MEMORY ($n = 10$)	2.209	2.335	2.372	2.385
TWO-CHOICE ($n = 15$)	2.113	2.205	2.250	2.273
MEMORY ($n = 15$)	2.326	2.462	2.500	2.512

Table 7.7: Exact computation of $\mathbf{E}[\,\text{Gap}(m)\,]$ for $n \in \{10, 15\}$ and $m \in \{n, 2n, 3n, 4n\}$ using dynamic programming. As in Table 7.6, it confirms that TWO-CHOICE has a smaller expected gap than MEMORY.

n	(1, 1)-MEMORY	(1, 1)-MEMORY*	(1, 2)-MEMORY*
10	2.209	2.099	2.015
20	2.427	2.201	2.043
30	2.586	2.294	2.070

Table 7.8: Exact computation of $\mathbf{E}[\,\text{Gap}(m)\,]$ for $m = n$ and $n \in \{5, 10, 15, 20, 25, 30, 35\}$ using dynamic programming. The table shows that for small values of n, TWO-CHOICE is better than MEMORY* with one cache, but worse than MEMORY* with two caches.

Conclusions

In summary, we introduced a set of techniques for analysing balanced allocation processes in the heavily-loaded case. The techniques are summarised as follows: (i) a refined analysis for the expectation of the hyperbolic cosine potential (Chapter 3) and its concentration (Chapter 4), (ii) an interplay between *absolute value*, *quadratic* and *exponential* potentials (Chapter 5), (iii) an analysis of *super-exponential* potentials in expectation and their concentration (Chapter 6) and (iv) a layered induction over super-exponential potentials (Chapter 7). These techniques allowed us to obtain asymptotically tight bounds for large families of processes and in a broad variety of settings, including outdated, noisy and incomplete information (see Tables A.1 to A.4). These reveal several insights into the balanced allocations problem, such as

- The characterisation of *adversarial noise* in the Two-Choice process (Section 7.4.1). As a corollary of this, we obtained bounds for Two-Choice with *random* and *delay noise* (Section 7.4.3).
- A phase transition in the b-Batched setting, where for batch size $b \leqslant n \log n$, Two-Choice is asymptotically optimal among processes making a constant number of samples (Section 7.4.3); and for $b \gg n \log n$ where instances of the $(1+\beta)$-process and Quantile(δ) processes have a gap that improves that of Two-Choice roughly quadratically (Section 7.5).
- Exploring various novel ways for obtaining balanced allocations, such as the "power of filling" underloaded bins using processes like Twinning and Packing (Section 5.2.2), which also achieve an $\mathcal{O}(\log n)$ gap in sparse regular graphs, like cycles (Remark 2.20).
- Deriving gap bounds for the weighted Graphical setting making progress on [152, Open Problem 1], by extending their $\mathcal{O}(\log n/\phi)$ bound on the gap for regular expanders with conductance ϕ to hold in the presence of weights and even for the b-Batched setting (Section 7.6.1). Further, we extend the results of [100] to prove sub-logarithmic gap bounds for expanders in the heavily-loaded case (Section 7.6.2).
- For Two-Thinning processes, we proved a lower bound that refutes [75, Problem 1.3] (see Appendix C.3). Following [76], we analysed more easily-realisable processes, namely Quantile(δ^*), which asymptotically achieves this optimal gap and Relative-Threshold processes which are within polyloglog(n) factors of the optimal. Then, we established a "power of two queries" phenomenon for processes that use two samples and make more than one queries per sample (Section 7.2.1). The generality of the analysis also gave us a near-tight bound on d-Thinning processes with $d \geqslant 2$ (Section 7.3.1).
- Investigating the "power of Memory" when sampling with heterogeneous distributions of arbitrary constant imbalance, and also demonstrating its robustness to resets (Section 7.7).
- In [116], we also used some of the lower bound techniques and the interplay between the quadratic potential and number of empty bins, to resolve two conjectures in the Repeated Balls-into-Bins setting [23].

There are several open questions remaining in the area of balanced allocations and various interesting processes and settings to be explored. We think that a large number of processes and settings can be analysed using these techniques and extensions thereof. A small set of these are the following:

- Establishing tight lower and upper bounds for d-Thinning, k-Quantile and k-Threshold. The result for the Quantile(δ^*) process hints that the bound for k-Dense-Quantile process need not be tight. As shown in [76] for $k = 1$, the optimal bounds may differ if the number of balls m is known in advance. This raises the questions: (i) are there easily-realisable processes that achieve the optimal bounds when m is known in advance and (ii) does this also extend for $k \geqslant 2$?

- Establishing tight bounds for Two-Choice in the Graphical setting for arbitrary graphs, like cycles. This problem has also been stated as [152, Open Problem 2] and in related settings in [23]. Another interesting direction is to design simpler processes that improve upon Two-Choice and are more easily-realisable than the ones in [20]. It may be of interest to explore how some of the processes like Mean-Thinning perform in the Graphical setting.

- Analysing other real-world processes in the outdated and noisy settings, such as the (k,d)-Choice process [146]. It would also be interesting to explore more closely the connection between the outdated setting and that of having multiple allocators, which is one of the main motivations in real-world systems for studying settings with delayed information.

- There are several interesting questions regarding the Memory process. One possible direction is to quantify the advantage of the optimal Memory strategy when $M \geqslant 1$ (it is not optimal to allocate in the least loaded of the sampled and the cached bin) and whether this gives any advantage over Two-Choice in the presence of weights. Another could be to obtain tight bounds for (a,b)-biased distributions for not necessarily constant $a, b > 1$.

- Analysing variants of the d-Thinning process which penalise the number of samples rejected (and thus improve the sample efficiency), e.g., on the k-th sample you allocate k balls (or a number of balls that is given by a non-decreasing function $f(k)$).

- Devising techniques for deriving bounds that are tight up to lower order terms for the various settings considered.

- Analysing noisy settings where the noise parameters are dependent on the bins (i.e., are heterogeneous) and unknown, so they have to be learnt.

Bibliography

[1] Anders Aamand, Mathias Bæk Tejs Knudsen, and Mikkel Thorup. Power of d Choices with Simple Tabulation. In *45th International Colloquium on Automata, Languages, and Programming (ICALP'18)*, Vol. 107. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 5:1–5:14. doi 24

[2] Dimitris Achlioptas, Raissa M. D'Souza, and Joel Spencer. Explosive percolation in random networks. *Science* 323, 5920 (2009), 1453–1455. doi 25

[3] Micah Adler, Soumen Chakrabarti, Michael Mitzenmacher, and Lars Rasmussen. Parallel randomized load balancing. *Random Structures & Algorithms* 13, 2 (1998), 159–188. doi 24, 200, 202, 203, 204

[4] Susanne Albers and Michael Mitzenmacher. Average-case analyses of first fit and random fit bin packing. In *19th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'98)*. ACM, New York, 290–299. 25

[5] Dan Alistarh, James Aspnes, and Rati Gelashvili. Space-Optimal Majority in Population Protocols. In *29th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'18)*. SIAM, 2221–2239. doi 25

[6] Dan Alistarh, Alkida Balliu, Dimitrios Los, and Sean Ovens. A Brief Summary of PODC 2022. *SIGACT News* 54, 1 (mar 2023), 106–112. doi 18

[7] Dan Alistarh, Trevor Brown, Justin Kopinsky, Jerry Zheng Li, and Giorgi Nadiradze. Distributionally Linearizable Data Structures. In *30th Annual ACM Symposium on Parallel Algorithms and Architectures (SPAA'18)*. ACM, 133–142. doi 14, 16, 25, 43

[8] Dan Alistarh, Rati Gelashvili, and Joel Rybicki. Fast Graphical Population Protocols. In *25th International Conference on Principles of Distributed Systems (OPODIS'21)*, Vol. 217. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 14:1–14:18. doi 24, 25

[9] Dan Alistarh, Justin Kopinsky, Jerry Li, and Giorgi Nadiradze. The Power of Choice in Priority Scheduling. In *36th Annual ACM-SIGOPT Principles of Distributed Computing (PODC'17)*. ACM, 283–292. doi 25

[10] Noga Alon, Itai Benjamini, Eyal Lubetzky, and Sasha Sodin. Non-backtracking random walks mix faster. *Commun. Contemp. Math.* 9, 4 (2007), 585–603. doi 25

[11] Noga Alon, Eyal Lubetzky, and Ori Gurel-Gurevich. Choice-memory tradeoff in allocations. In *50th Annual IEEE Symposium on Foundations of Computer Science (FOCS'09)*. IEEE Computer Soc., Los Alamitos, CA, 230–238. doi 24

[12] Eitan Altman and Philippe Nain. Closed-Loop Control with Delayed Information. In *ACM SIGMETRICS Joint International Conference on Measurement and Modeling of Computer Systems (PERFORMANCE'92)*. ACM, 193–204. doi 13

[13] Juhan Aru, Bhargav Narayanan, Alex Scott, and Ramarathnam Venkatesan. Balancing sums of random vectors. *Discrete Anal.* 4 (2018), 17 pages. doi 24

[14] Gilad Asharov, Moni Naor, Gil Segev, and Ido Shahaf. Searchable symmetric encryption: optimal locality in linear space via two-dimensional balanced allocations. In *48th Annual ACM Symposium on Theory of Computing (STOC'16)*. ACM, 1101–1114. doi 25

[15] John Augustine, William K. Moses Jr., Amanda Redlich, and Eli Upfal. Balanced Allocation: Patience is not a Virtue. In *27th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'16)*. SIAM, 655–671. doi 24, 25

[16] Chen Avin and Bhaskar Krishnamachari. The Power of Choice in Random Walks: An Empirical Study. In *9th ACM Symposium on Modeling Analysis and Simulation of Wireless and Mobile Systems (MSWiM'06)*. ACM, 219–228. doi 25

[17] Yossi Azar, Andrei Z. Broder, Anna R. Karlin, Michael Mitzenmacher, and Eli Upfal. 2020. The ACM Paris Kanellakis Theory and Practice Award. https://www.acm.org/media-center/2021/may/technical-awards-2020. 3, 12

[18] Yossi Azar, Andrei Z. Broder, Anna R. Karlin, and Eli Upfal. Balanced allocations. *SIAM J. Comput.* 29, 1 (1999), 180–200. doi 3, 11, 24, 190, 205, 209, 211, 224

[19] Yossi Azar, Ilan Reuven Cohen, and Debmalya Panigrahi. Randomized algorithms for online vector load balancing. In *29th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'18)*. SIAM, Philadelphia, PA, 980–991. doi 24

[20] Nikhil Bansal and Ohad N. Feldheim. The power of two choices in graphical allocation. In *54th Annual ACM Symposium on Theory of Computing (STOC'22)*. ACM, 52–63. doi 14, 24, 176

[21] Rakesh D. Barve, Edward F. Grove, and Jeffrey Scott Vitter. Simple randomized mergesort on parallel disks. *Parallel Comput.* 23, 4-5 (1997), 601–631. doi 24

[22] Tuğkan Batu, Petra Berenbrink, and Colin Cooper. Chains-into-bins processes. *J. Discrete Algorithms* 14 (2012), 21–28. doi 24

[23] Luca Becchetti, Andrea E. F. Clementi, Emanuele Natale, Francesco Pasquale, and Gustavo Posta. Self-stabilizing repeated balls-into-bins. *Distributed Comput.* 32, 1 (2019), 59–68. doi 13, 175, 176

[24] Omri Ben-Eliezer, Lior Gishboliner, Dan Hefetz, and Michael Krivelevich. Very fast construction of bounded-degree spanning graphs via the semi-random graph process. *Random Struct. Algorithms* 57, 4 (2020), 892–919. doi 25

[25] Omri Ben-Eliezer, Dan Hefetz, Gal Kronenberg, Olaf Parczyk, Clara Shikhelman, and Milos Stojakovic. Semi-random graph process. *Random Struct. Algorithms* 56, 3 (2020), 648–675. doi 25

[26] Itai Benjamini and Yury Makarychev. Balanced allocation: memory performance tradeoffs. *Ann. Appl. Probab.* 22, 4 (2012), 1642–1649. doi 24

[27] Petra Berenbrink, André Brinkmann, Tom Friedetzky, and Lars Nagel. Balls into non-uniform bins. *J. Parallel Distributed Comput.* 74, 2 (2014), 2065–2076. doi 14

[28] Petra Berenbrink, Artur Czumaj, Matthias Englert, Tom Friedetzky, and Lars Nagel. Multiple-Choice Balanced Allocation in (Almost) Parallel. In *16th International Workshop on Randomization and Computation (RANDOM'12)*. Springer-Verlag, 411–422. doi 3, 13, 15, 41, 159, 192

[29] Petra Berenbrink, Artur Czumaj, Angelika Steger, and Berthold Vöcking. Balanced allocations: the heavily loaded case. *SIAM J. Comput.* 35, 6 (2006), 1350–1385. doi 3, 11, 25, 190

[30] Petra Berenbrink, Tom Friedetzky, Christopher Hahn, Lukas Hintze, Dominik Kaaser, Peter Kling, and Lars Nagel. Infinite Balanced Allocation via Finite Capacities. In *41st IEEE International Conference on Distributed Computing Systems, (ICDCS'21)*. IEEE, 965–975. doi 14

[31] Petra Berenbrink, Tom Friedetzky, Peter Kling, Frederik Mallmann-Trenn, Lars Nagel, and Chris Wastell. Self-Stabilizing Balls and Bins in Batches - The Power of Leaky Bins. *Algorithmica* 80, 12 (2018), 3673–3703. doi 13

[32] Petra Berenbrink, Tom Friedetzky, Peter Kling, Frederik Mallmann-Trenn, and Chris Wastell. Plurality Consensus in Arbitrary Graphs: Lessons Learned from Load Balancing. In *24th Annual European Symposium on Algorithms (ESA'16)*, Vol. 57. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 10:1–10:18. doi 14

[33] Petra Berenbrink, Kamyar Khodamoradi, Thomas Sauerwald, and Alexandre Stauffer. Balls-into-bins with nearly optimal load distribution. In *25th Annual ACM Symposium on Parallel Algorithms and Architectures (SPAA'13)*. ACM, 326–335. doi 24, 25

[34] Philippe Besnard, Sylvie Doutre, and Andreas Herzig. Encoding argument graphs in logic. In *Information processing and management of uncertainty in knowledge-based systems. Part II (Commun. Comput. Inf. Sci., Vol. 443)*. Springer, [Cham], 345–354. 24

[35] Meena Boppana, Rani Hod, Michael Mitzenmacher, and Tom Morgan. Voronoi choice games. In *43rd International Colloquium on Automata, Languages, and Programming (ICALP'16)*, Vol. 55. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, Art. No. 23, 13. 24

[36] Prosenjit Bose, Karim Douïeb, and Stefan Langerman. Dynamic optimality for skip lists and B-trees. In *19th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'08)*. ACM, New York, 1106–1114. 25

[37] Karl Bringmann, Thomas Sauerwald, Alexandre Stauffer, and He Sun. Balls into bins via local search: cover time and maximum load. In *31st International Symposium on Theoretical Aspects of Computer Science (STACS'14)*, Vol. 25. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 187–198. 24

[38] Karl Bringmann, Thomas Sauerwald, Alexandre Stauffer, and He Sun. Balls into bins via local search: cover time and maximum load. *Random Structures & Algorithms* 48, 4 (2016), 681–702. doi 24

[39] Andrei Broder and Michael Mitzenmacher. Using multiple hash functions to improve IP lookups. In *20th IEEE Conference on Computer Communications (INFOCOM'01)*, Vol. 3. IEEE, 1454–1463. doi 25

[40] Andrei Broder and Michael Mitzenmacher. Multidimensional balanced allocations. In *16th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'05)*. ACM, 195–196. 24

[41] Andrei Z. Broder and Ravi Kumar. 2020. A Note on Double Pooling Tests. arXiv:2004.01684 25

[42] Amarjit Budhiraja, Debankur Mukherjee, and Ruoyu Wu. Supermarket model on graphs. *Ann. Appl. Probab.* 29, 3 (2019), 1740–1777. doi 24

[43] John W. Byers, Jeffrey Considine, and Michael Mitzenmacher. Geometric Generalizations of the Power of Two Choices. In *16th Annual ACM Symposium on Parallel Algorithms and Architectures (SPAA'04)*. ACM, 54–63. doi 24

[44] Julie Anne Cain, Peter Sanders, and Nick Wormald. The random graph threshold for k-orientability and a fast algorithm for optimal multiple-choice allocation. In *18th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'07)*. ACM, 469–476. 25

[45] Nicoletta Cancrini and Gustavo Posta. Propagation of chaos for a balls into bins model. *Electronic Communications in Probability* 24 (2019), 1 – 9. doi 13

[46] Nicoletta Cancrini and Gustavo Posta. Mixing time for the Repeated Balls into Bins dynamics. *Electronic Communications in Probability* 25 (2020), 1 – 14. doi

[47] Nicoletta Cancrini and Gustavo Posta. Propagation of chaos for a general balls into bins dynamics. *Electronic Journal of Probability* 26 (2021), 1 – 20. doi 13

[48] L. Elisa Celis, Omer Reingold, Gil Segev, and Udi Wieder. Balls and Bins: Smaller Hash Families and Faster Evaluation. *SIAM J. Comput.* 42, 3 (2013), 1030–1050. doi 24

[49] May-Ru Chen and Ching-Zong Wei. A new urn model. *J. Appl. Probab.* 42, 4 (2005), 964–976. doi 24

[50] Xue Chen. Derandomized Balanced Allocation. In *30th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'19)*. SIAM, 2513–2526. doi 24
[51] Fan Chung and Linyuan Lu. Concentration inequalities and martingale inequalities: a survey. *Internet Math.* 3, 1 (2006), 79–127. 196
[52] Fan R. K. Chung. 1997. *Spectral Graph Theory*. American Mathematical Society, USA. 166
[53] Ilan Cohen, Sungjin Im, and Debmalya Panigrahi. Online two-dimensional load balancing. In *47th International Colloquium on Automata, Languages, and Programming (ICALP'20)*, Vol. 168. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, Art. No. 34, 21. 24
[54] Richard Cole, Alan Frieze, Bruce M. Maggs, Michael Mitzenmacher, Andréa W. Richa, Ramesh Sitaraman, and Eli Upfal. On balls and bins with deletions. In *2nd International Workshop on Randomization and Computation (RANDOM'98)*, Vol. 1518. Springer, Berlin, 145–158. doi 24, 25
[55] Artur Czumaj, Friedhelm Meyer auf der Heide, and Volker Stemann. Contention resolution in hashing based shared memory simulations. *SIAM J. Comput.* 29, 5 (2000), 1703–1739. doi 25
[56] Artur Czumaj and Volker Stemann. Randomized allocation processes. *Random Structures & Algorithms* 18, 4 (2001), 297–331. doi 24
[57] Søren Dahlgaard, Mathias Bæk Tejs Knudsen, Eva Rotenberg, and Mikkel Thorup. The Power of Two Choices with Simple Tabulation. In *27th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'16)*. SIAM, 1631–1642. doi 24
[58] Michael Dahlin. Interpreting Stale Load Information. *IEEE Trans. Parallel Distributed Syst.* 11, 10 (2000), 1033–1047. doi 4, 13, 15
[59] Luc Devroye. 1986. *Lecture notes on bucket algorithms*. Progress in Computer Science, Vol. 6. Birkhäuser Boston, Inc., Boston, MA, New York. viii+146 pages. doi 24
[60] Luc Devroye, Gabor Lugosi, Gahyun Park, and Wojciech Szpankowski. Multiple choice tries and distributed hash tables. In *18th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'07)*. ACM, New York, 891–899. 25
[61] Luc Devroye and Pat Morin. Cuckoo hashing: further analysis. *Inform. Process. Lett.* 86, 4 (2003), 215–219. doi 24
[62] Martin Dietzfelbinger and Friedhelm Meyer auf der Heide. Simple, Efficient Shared Memory Simulations. In *5th Annual ACM Symposium on Parallel Algorithms and Architectures (SPAA'93)*. ACM, 110–119. doi 25
[63] Martin Dietzfelbinger and Christoph Weidling. Balanced allocation and dictionaries with tightly packed constant size bins. *Theoret. Comput. Sci.* 380, 1-2 (2007), 47–68. doi 24
[64] Martin Dietzfelbinger and Philipp Woelfel. Almost random graphs with simple hash functions. In *35th Annual ACM Symposium on Theory of Computing (STOC'03)*. ACM, 629–638. doi 24
[65] Benjamin Doerr, Leslie Ann Goldberg, Lorenz Minder, Thomas Sauerwald, and Christian Scheideler. Brief Announcement: Stabilizing Consensus with the Power of Two Choices. In *Distributed Computing*. Springer Berlin Heidelberg, 528–530. 25
[66] Eleni Drinea, Alan Frieze, and Michael Mitzenmacher. Balls and Bins Models with Feedback. In *13th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'02)*. SIAM, 308–315. 24
[67] Devdatt P. Dubhashi and Alessandro Panconesi. 2009. *Concentration of measure for the analysis of randomized algorithms*. Cambridge University Press, Cambridge. xvi+196 pages. doi 24, 196
[68] Derek L. Eager, Edward D. Lazowska, and John Zahorjan. Adaptive load sharing in homogeneous distributed systems. *IEEE Transactions on Software Engineering* SE-12, 5 (1986), 662–675. doi 11
[69] Jeff Edmonds and Kirk Pruhs. Balanced Allocations of Cake. In *47th Annual IEEE Symposium on Foundations of Computer Science (FOCS'06)*. IEEE Computer Society, 623–634. doi 25

[70] Jeff Edmonds, Kirk Pruhs, and Jaisingh Solanki. Confidently Cutting a Cake into Approximately Fair Pieces. In *4th International Conference on Algorithmic Aspects in Information and Management (AAIM'08)*. Springer-Verlag, 155–164. doi 25
[71] Carl G. Esseen. A moment inequality with an application to the central limit theorem. *Skand. Aktuarietidskr.* 39 (1956), 160–170. doi 195
[72] Guy Even and Moti Medina. Revisiting randomized parallel load balancing algorithms. *Theoret. Comput. Sci.* 444 (2012), 87–99. doi 24
[73] Bin Fan, Dave G. Andersen, Michael Kaminsky, and Michael D. Mitzenmacher. Cuckoo Filter: Practically Better Than Bloom. In *10th ACM Conference on Emerging Networking Experiments and Technologies (CoNEXT'14)*. ACM, 75–88. doi 24
[74] Sándor Fekete, Ekkehard Köhler, and Jürgen Teich. Optimal FPGA Module Placement with Temporal Precedence Constraints. In *Proceedings of the Conference on Design, Automation and Test in Europe (DATE '01)*. IEEE Press, 658–667. 24
[75] Ohad N. Feldheim and Ori Gurel-Gurevich. The power of thinning in balanced allocation. *Electron. Commun. Probab.* 26 (2021), Paper No. 34, 8. doi 3, 12, 25, 175, 215
[76] Ohad N. Feldheim, Ori Gurel-Gurevich, and Jiange Li. 2021. Long-term balanced allocation via thinning. arXiv:2110.05009 [math.PR] arXiv:2110.05009. 16, 25, 27, 175, 176, 190
[77] Ohad N. Feldheim and Jiange Li. Load balancing under d-thinning. *Electron. Commun. Probab.* 25 (2020), Paper No. 1, 13. doi 148, 190, 215
[78] Jon Feldman, Aranyak Mehta, Vahab Mirrokni, and S. Muthukrishnan. Online stochastic matching: beating $1-\frac{1}{e}$. In *50th Annual IEEE Symposium on Foundations of Computer Science (FOCS'09)*. IEEE Computer Soc., Los Alamitos, CA, 117–126. doi 25
[79] William Feller. 1971. *An introduction to probability theory and its applications. Vol. II*. John Wiley & Sons, Inc., New York-London-Sydney. xxiv+669 pages. 201, 203
[80] Daniel Fernholz and Vijaya Ramachandran. The k-orientability thresholds for $G_{n,p}$. In *18th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'07)*. ACM, New York, 459–468. 25
[81] Nikolaos Fountoulakis, Megha Khosla, and Konstantinos Panagiotou. The multiple-orientability thresholds for random hypergraphs. *Combin. Probab. Comput.* 25, 6 (2016), 870–908. doi 25
[82] Armando Fox, Steven D. Gribble, Yatin Chawathe, Eric A. Brewer, and Paul Gauthier. Cluster-Based Scalable Network Services. In *16th ACM Symposium on Operating Systems Principles (SOSP'97)*. Association for Computing Machinery, 78–91. doi 13
[83] Michael J. Freedman, Carmit Hazay, Kobbi Nissim, and Benny Pinkas. Efficient set intersection with simulation-based security. *J. Cryptology* 29, 1 (2016), 115–155. doi 25
[84] Alan Frieze, Páll Melsted, and Michael Mitzenmacher. An analysis of random-walk cuckoo hashing. *SIAM J. Comput.* 40, 2 (2011), 291–308. doi 24
[85] Francis Galton. 1889. *Natural inheritance*. Macmillan, London. 282 pages. https://www.biodiversitylibrary.org/item/76404 78
[86] Pu Gao and Nicholas Wormald. Orientability thresholds for random hypergraphs. *Combin. Probab. Comput.* 24, 5 (2015), 774–824. doi 25
[87] Agelos Georgakopoulos, John Haslegrave, Thomas Sauerwald, and John Sylvester. Choice and bias in random walks. In *11th Innovations in Theoretical Computer Science Conference (ITCS'20)*, Vol. 151. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, Art. No. 76, 19. 25
[88] George Giakkoupis and Vassos Hadzilacos. A scheme for load balancing in heterogenous distributed hash tables. In *24th Annual ACM-SIGOPT Principles of Distributed Computing (PODC'05)*. ACM, 302–311. doi 25
[89] Gaston H. Gonnet. Expected length of the longest probe sequence in hash code searching. *J. Assoc. Comput. Mach.* 28, 2 (1981), 289–304. doi 11

[90] Catherine Greenhill, Bernard Mans, and Ali Pourmiri. Balanced Allocation on Dynamic Hypergraphs. In *24th International Workshop on Randomization and Computation (RANDOM'20)*, Vol. 176. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 11:1–11:22. doi 14, 24

[91] Anupam Gupta, Ravishankar Krishnaswamy, Amit Kumar, and Sahil Singla. Online Carpooling Using Expander Decompositions. In *40th IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science (FSTTCS'20)*, Vol. 182. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 23:1–23:14. doi 25

[92] Michel Hanna, Socrates Demetriades, Sangyeun Cho, and Rami Melhem. CHAP: Enabling Efficient Hardware-Based Multiple Hash Schemes for IP Lookup. In *8th International IFIP-TC 6 Networking Conference (NETWORKING'09)*. Springer-Verlag, 756–769. doi 25

[93] Carmit Hazay and Kobbi Nissim. Efficient set operations in the presence of malicious adversaries. *J. Cryptology* 25, 3 (2012), 383–433. doi 25

[94] Susan Hohenberger and Stephen A. Weis. Honest-Verifier Private Disjointness Testing Without Random Oracles. In *Privacy Enhancing Technologies*. Springer Berlin Heidelberg, 277–294. 25

[95] Zhiyi Huang, Xinkai Shu, and Shuyi Yan. The power of multiple choices in online stochastic matching. In *54th Annual ACM Symposium on Theory of Computing (STOC'22)*. ACM, 91–103. doi 25

[96] Kiyoshi Itô and Henry P. McKean, Jr. 1965. *Diffusion processes and their sample paths*. Academic Press, Inc., Publishers, New York; Springer-Verlag, Berlin-New York, Berlin. xvii+321 pages. doi 44

[97] Kazuo Iwama and Akinori Kawachi. Approximated Two Choices in Randomized Load Balancing. In *Algorithms and Computation*. Springer Berlin Heidelberg, 545–557. 16

[98] Norman L. Johnson and Samuel Kotz. 1977. *Urn models and their application: An approach to modern discrete probability theory*. John Wiley & Sons, New York-London-Sydney. xiii+402 pages. 24

[99] Richard M. Karp, Michael Luby, and Friedhelm Meyer auf der Heide. Efficient PRAM simulation on a distributed memory machine. *Algorithmica* 16, 4-5 (1996), 517–542. doi 3, 11, 25

[100] Krishnaram Kenthapadi and Rina Panigrahy. Balanced allocation on graphs. In *17th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'06)*. ACM, New York, 434–443. doi 4, 14, 16, 168, 175

[101] Adam Kirsch, Michael Mitzenmacher, and Udi Wieder. More robust hashing: cuckoo hashing with a stash. *SIAM J. Comput.* 39, 4 (2009), 1543–1561. doi 24

[102] Valentin F. Kolchin, Boris A. Sevastyanov, and Vladimir P. Chistyakov. 1978. *Random allocations*. John Wiley & Sons, New York-London-Sydney. xi+262 pages. 24, 25

[103] Samuel Kotz and Narayanaswamy Balakrishnan. 1997. Advances in urn models during the past two decades. In *Advances in combinatorial methods and applications to probability and statistics*. Birkhäuser Boston, Boston, MA, 203–257. 24

[104] Elias Koutsoupias, Marios Mavronicolas, and Paul G. Spirakis. Approximate Equilibria and Ball Fusion. In *9th International Colloquium on Structural Information and Communication Complexity (SIROCCO'02)*, Vol. 13. Carleton Scientific, 223–235. 25

[105] Łukasz Krzywiecki and Mirosław Kutyłowski. Coalition Resistant Anonymous Broadcast Encryption Scheme Based on PUF. In *Trust and Trustworthy Computing*. Springer Berlin Heidelberg, 48–62. 25

[106] Joy Kuri and Anurag Kumar. Optimal control of arrivals to queues with delayed queue length information. *IEEE Trans. Automat. Control* 40, 8 (1995), 1444–1450. doi 13

[107] Samuel Kutin. 2002. *Extensions to McDiarmid's inequality when differences are bounded with high probability*. Technical Report. University of Chicago. 196, 234

[108] Marc Lelarge. A new approach to the orientation of random hypergraphs. In *23rd Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'12)*. ACM, New York, 251–264. 25

[109] Christoph Lenzen, Merav Parter, and Eylon Yogev. Parallel Balanced Allocations: The Heavily Loaded Case. In *31st Annual ACM Symposium on Parallel Algorithms and Architectures (SPAA'19)*. ACM, 313–322. `doi` 24

[110] Christoph Lenzen and Roger Wattenhofer. Tight bounds for parallel randomized load balancing [extended abstract]. In *43rd Annual ACM Symposium on Theory of Computing (STOC'11)*. ACM, New York, 11–20. `doi` 24

[111] Dimitrios Los and Thomas Sauerwald. Balanced Allocations in Batches: Simplified and Generalized. In *34th Annual ACM Symposium on Parallel Algorithms and Architectures (SPAA'22)*. ACM, 389–399. `doi`

[112] Dimitrios Los and Thomas Sauerwald. Balanced Allocations with Incomplete Information: The Power of Two Queries. In *13th Innovations in Theoretical Computer Science Conference (ITCS'22)*, Vol. 215. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 103:1–103:23. `doi`

[113] Dimitrios Los and Thomas Sauerwald. Balanced Allocations with the Choice of Noise. In *41st Annual ACM-SIGOPT Principles of Distributed Computing (PODC'22)*. ACM, 164–175. `doi` 18

[114] Dimitrios Los and Thomas Sauerwald. Balanced Allocations in Batches: The Tower of Two Choices. In *35th Annual ACM Symposium on Parallel Algorithms and Architectures (SPAA'23)*. ACM, 51–61. `doi` 160, 207

[115] Dimitrios Los and Thomas Sauerwald. 2023. An Improved Drift Theorem for Balanced Allocations. arXiv:2308.11083 [math.PR] 53

[116] Dimitrios Los and Thomas Sauerwald. Tight Bounds for Repeated Balls-Into-Bins. In *40th International Symposium on Theoretical Aspects of Computer Science (STACS'23)*, Vol. 254. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 45:1–45:22. `doi` 21, 175

[117] Dimitrios Los, Thomas Sauerwald, and John Sylvester. Balanced Allocations: Caching and Packing, Twinning and Thinning. In *33rd Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'22)*. SIAM, 1847–1874. `doi` 17, 50, 90, 91, 93, 94, 95, 190

[118] Dimitrios Los, Thomas Sauerwald, and John Sylvester. Balanced Allocations with Heterogeneous Bins: The Power of Memory. In *34th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'23)*. SIAM, 4448–4477. `doi` 17, 22, 169, 224

[119] Dimitrios Los, Thomas Sauerwald, and John Sylvester. 2023. Mean-Biased Processes for Balanced Allocations. arXiv:2308.05087 [math.PR] 90

[120] Malwina J. Luczak and Colin McDiarmid. On the maximum queue length in the supermarket model. *Ann. Probab.* 34, 2 (2006), 493–527. `doi` 24

[121] Malwina J. Luczak and James Norris. Strong approximation for the supermarket model. *Ann. Appl. Probab.* 15, 3 (2005), 2038–2061. `doi` 24

[122] Malwina J. Luczak and James R. Norris. Averaging over fast variables in the fluid limit for Markov chains: Application to the supermarket model with memory. *The Annals of Applied Probability* 23, 3 (2013), 957 – 986. 25

[123] Malwina J. Luczak and Eli Upfal. Reducing network congestion and blocking probability through balanced allocation. In *40th Annual IEEE Symposium on Foundations of Computer Science (FOCS'99)*. IEEE Computer Soc., Los Alamitos, CA, 587–595. `doi` 25

[124] Hosam M. Mahmoud. 2009. *Pólya urn models*. CRC Press, Boca Raton, FL. xii+290 pages. 24

[125] Yishay Mansour, Aviad Rubinstein, Shai Vardi, and Ning Xie. Converting Online Algorithms to Local Computation Algorithms. In *39th International Colloquium on Automata, Languages, and Programming (ICALP'12)*. Springer-Verlag, 653–664. `doi` 25

[126] Albert W. Marshall and Ingram Olkin. 1983. Inequalities via majorization—an introduction. In *General inequalities, 3 (Oberwolfach, 1981)*. Internat. Schriftenreihe Numer. Math., Vol. 64. Birkhäuser, Basel, 165–187. 35

[127] Marios Mavronicolas and Paul Spirakis. The price of selfish routing. In *33rd Annual ACM Symposium on Theory of Computing (STOC'01)*. ACM, 510–519. doi 25

[128] Friedhelm Meyer auf der Heide, Christian Scheideler, and Volker Stemann. Exploiting storage redundancy to speed up randomized shared memory simulations. *Theoret. Comput. Sci.* 162, 2 (1996), 245–281. doi 24, 25

[129] Vahab S. Mirrokni, Mikkel Thorup, and Morteza Zadimoghaddam. Consistent Hashing with Bounded Loads. In *29th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'18)*. SIAM, 587–604. doi 14

[130] Michael Mitzenmacher. Load Balancing and Density Dependent Jump Markov Processes. In *37th Annual IEEE Symposium on Foundations of Computer Science (FOCS'96)*. IEEE Computer Society, 213–222. 25

[131] Michael Mitzenmacher. 1996. *The Power of Two Choices in Randomized Load Balancing*. Ph. D. Dissertation. University of California at Berkeley. 12

[132] Michael Mitzenmacher. On the analysis of randomized load balancing schemes. *Theory Comput. Syst.* 32, 3 (1999), 361–386. doi 34

[133] Michael Mitzenmacher. Studying balanced allocations with differential equations. *Combin. Probab. Comput.* 8, 5 (1999), 473–482. doi 25

[134] Michael Mitzenmacher. How useful is old information? *IEEE Transactions on Parallel and Distributed Systems* 11, 1 (2000), 6–20. doi 4, 13, 15, 207

[135] Michael Mitzenmacher. More Analysis of Double Hashing for Balanced Allocations. In *13th Workshop on Analytic Algorithmics and Combinatorics (ANALCO'16)*. SIAM, 1–9. doi 24

[136] Michael Mitzenmacher, Balaji Prabhakar, and Devavrat Shah. Load Balancing with Memory. In *43rd Annual IEEE Symposium on Foundations of Computer Science (FOCS'02)*. IEEE, 799–808. doi 4, 13, 33, 173

[137] Michael Mitzenmacher, Andréa W. Richa, and Ramesh Sitaraman. 2001. The power of two random choices: a survey of techniques and results. In *Handbook of randomized computing, Vol. I, II*. Comb. Optim., Vol. 9. Kluwer Acad. Publ., Dordrecht, 255–312. doi 11, 24, 194

[138] Michael Mitzenmacher and Eli Upfal. 2017. *Probability and computing* (second ed.). Cambridge University Press, Cambridge. xx+467 pages. Randomization and probabilistic techniques in algorithms and data analysis. 24, 25, 195, 199, 200, 202, 208, 212, 213

[139] Yves Mocquard, Bruno Sericola, and Emmanuelle Anceaume. Balanced Allocations and Global Clock in Population Protocols: An Accurate Analysis. In *25th International Colloquium on Structural Information and Communication Complexity (SIROCCO'18)*, Vol. 11085. Springer, 296–311. doi 25

[140] Rajeev Motwani, Rina Panigrahy, and Ying Xu. Fractional Matching Via Balls-and-Bins. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*. Springer Berlin Heidelberg, 487–498. 25

[141] Rajeev Motwani and Prabhakar Raghavan. 1995. *Randomized algorithms*. Cambridge University Press, Cambridge. xiv+476 pages. doi 24

[142] Giorgi Nadiradze. 2021. *On Achieving Scalability through Relaxation*. Ph. D. Dissertation. IST Austria. doi 73, 191

[143] Ankur Narang, Sourav Dutta, and Souvik Bhattacherjee. 2011. Multidimensional Balanced Allocation for Multiple Choice & $(1+\beta)$ Processes. arXiv:1111.0762 24

[144] Kay Ousterhout, Patrick Wendell, Matei Zaharia, and Ion Stoica. Sparrow: distributed, low latency scheduling. In *ACM SIGOPS 24th Symposium on Operating Systems Principles, SOSP '13, Farmington, PA, USA, November 3-6, 2013*. ACM, 69–84. doi 24
[145] Rasmus Pagh and Flemming Friche Rodler. Cuckoo hashing. *J. Algorithms* 51, 2 (2004), 122–144. doi 24
[146] Gahyun Park. A generalization of multiple choice balls-into-bins: tight bounds. *Algorithmica* 77, 4 (2017), 1159–1193. doi 24, 176
[147] Gahyun Park, Minseok Kwon, Ramprasad Tamilselvan, and Seungjoon Lee. Two-Way Trees: A Distributed Algorithm for Efficient Replica Search and Placement. In *1st Symposium on Algorithmic Principles of Computer Systems (APOCS'20)*. SIAM, 133–146. doi 25
[148] Sarvar Patel, Giuseppe Persiano, and Kevin Yeo. What Storage Access Privacy is Achievable with Small Overhead?. In *38th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems, (PODS'19)*. ACM, 182–199. doi 25
[149] Mihai Patrascu and Mikkel Thorup. The Power of Simple Tabulation Hashing. *J. ACM* 59, 3 (2012), 14:1–14:50. doi 24
[150] David A. Patterson, Garth Gibson, and Randy H. Katz. A Case for Redundant Arrays of Inexpensive Disks (RAID). In *14th ACM International Conference on Management of Data (SIGMOD'88)*. ACM, 109–116. doi 24
[151] Robin Pemantle. A survey of random processes with reinforcement. *Probab. Surv.* 4 (2007), 1–79. doi 24
[152] Yuval Peres, Kunal Talwar, and Udi Wieder. Graphical balanced allocations and the $(1+\beta)$-choice process. *Random Structures & Algorithms* 47, 4 (2015), 760–775. doi 3, 4, 12, 14, 17, 19, 22, 25, 34, 35, 37, 40, 52, 53, 54, 65, 164, 166, 168, 175, 176, 190, 191, 199, 227
[153] Benny Pinkas, Thomas Schneider, and Michael Zohner. Scalable Private Set Intersection Based on OT Extension. *ACM Trans. Priv. Secur.* 21, 2 (2018), 7:1–7:35. doi 25
[154] Pattarawit Polpinit and Montira Thipchote. Upper bounds for the convergence rate of a randomized local search in a load-balancing game with variable-capacity resources. In *The 8th Electrical Engineering/ Electronics, Computer, Telecommunications and Information Technology (ECTI'11)*. IEEE, 520–523. doi 25
[155] Anastasiia Postnikova, Nikita Koval, Giorgi Nadiradze, and Dan Alistarh. Multi-queues can be state-of-the-art priority schedulers. In *27th ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming (PPoPP'22)*. ACM, 353–367. doi 25
[156] Ali Pourmiri. Balanced allocation on graphs: A random walk approach. *Random Struct. Algorithms* 55, 4 (2019), 980–1009. doi 25
[157] Martin Raab and Angelika Steger. "Balls into bins"—a simple and tight analysis. In *2nd International Workshop on Randomization and Computation (RANDOM'98)*, Vol. 1518. Springer, 159–170. doi 11, 192, 199, 200
[158] Amanda Redlich. A Power-of-Two-Choices Unbalanced Allocation Process. *SIAM Journal on Discrete Mathematics* 31, 1 (2017), 477–488. 24, 25
[159] Omer Reingold and Shai Vardi. New techniques and tighter bounds for local computation algorithms. *J. Comput. System Sci.* 82, 7 (2016), 1180–1200. doi 25
[160] Adones Rukundo, Aras Atalar, and Philippas Tsigas. Brief Announcement: 2D-Stack – A Scalable Lock-Free Stack Design That Continuously Relaxes Semantics for Better Performance. In *38th Annual ACM-SIGOPT Principles of Distributed Computing (PODC'18)*. ACM, 407–409. doi 25
[161] Luís M. S. Russo. Range minimum queries in minimal space. *Theor. Comput. Sci.* 909 (2022), 19–38. doi 25
[162] Peter Sanders and Berthold Vöcking. Tail bounds and expectations for random arc allocation and applications. *Combin. Probab. Comput.* 12, 3 (2003), 301–318. doi 24

[163] Thomas Sauerwald and He Sun. Tight Bounds for Randomized Load Balancing on Arbitrary Network Topologies. In *53rd Annual IEEE Symposium on Foundations of Computer Science (FOCS'12)*. IEEE Computer Society, 341–350. doi 125
[164] Ketan Savla, Emilio Frazzoli, and Francesco Bullo. Traveling salesperson problems for the Dubins vehicle. *IEEE Trans. Automat. Control* 53, 6 (2008), 1378–1391. doi 25
[165] Thomas Schickinger and Angelika Steger. Simplified Witness Tree Arguments. In *27th Conference on Current Trends in Theory and Practice of Informatics (SOFSEM'00)*, Vol. 1963. Springer, 71–87. doi 24
[166] Devravat Shah and Balaji Prabhakar. The use of memory in randomized load balancing. In *IEEE International Symposium on Information Theory (ISIT'02)*. IEEE, 125. doi 13
[167] Zhiwei Shang, Simon Oya, Andreas Peter, and Florian Kerschbaum. Obfuscated Access and Search Patterns in Searchable Encryption. In *28th Annual Network and Distributed System Security Symposium (NDSS'21)*. The Internet Society, 1–18. 25
[168] Amritpal Singh and Shalini Batra. Fingerprint based duplicate detection in streamed data. *Comput. Inform.* 37, 6 (2018), 1313–1338. 25
[169] Kunal Talwar and Udi Wieder. Balanced allocations: the weighted case. In *39th Annual ACM Symposium on Theory of Computing (STOC'07)*. ACM, 256–265. doi 14, 25
[170] Konstantin Tikhomirov and Pierre Youssef. The spectral gap of dense random regular graphs. *Ann. Probab.* 47, 1 (2019), 362–419. doi 168
[171] Benny Van Houdt. A mean field model for a class of garbage collection algorithms in flash-based solid state drives. *Queueing Syst.* 77, 2 (2014), 149–176. doi 25
[172] Berthold Vöcking. How Asymmetry Helps Load Balancing. In *40th Annual IEEE Symposium on Foundations of Computer Science (FOCS'99)*. IEEE, 131–141. doi 24
[173] Nikita D. Vvedenskaya, Roland L. Dobrushin, and Fridrikh I. Karpelevich. A queueing system with a choice of the shorter of two queues—an asymptotic approach. *Problemy Peredachi Informatsii* 32, 1 (1996), 20–34. 24
[174] Lutz Warnke. On the method of typical bounded differences. *Combin. Probab. Comput.* 25, 2 (2016), 269–299. doi 25
[175] Ward Whitt. Deciding Which Queue to Join: Some Counterexamples. *Oper. Res.* 34, 1 (1986), 55–62. doi 13
[176] Udi Wieder. Balanced allocations with heterogenous bins. In *19th Annual ACM Symposium on Parallel Algorithms and Architectures (SPAA'07)*. ACM, 188–193. doi 4, 14, 25, 31, 41, 169
[177] Udi Wieder. Hashing, Load Balancing and Multiple Choice. *Found. Trends Theor. Comput. Sci.* 12, 3-4 (2017), 275–379. doi 11, 24
[178] Nicholas Charles Wormald. 1999. *The differential equation method for random graph processes and greedy algorithms*. Wydawnictwo Naukowe Pwn, 73–155. 25
[179] Qiaomin Xie, Xiaobo Dong, Yi Lu, and Rayadurgam Srikant. Power of d Choices for Large-Scale Bin Packing: A Loss Model. *SIGMETRICS Perform. Eval. Rev.* 43, 1 (jun 2015), 321–334. doi 25
[180] Songnian Zhou. A trace-driven simulation study of dynamic load balancing. *IEEE Transactions on Software Engineering* 14, 9 (1988), 1327–1341. doi 12

Index

Summary of explicit bounds

Process	Range	Lower bound	Reference	Upper bound	Reference
Two-Choice	–	$\Omega(\log\log n)$	[18, Thm 1.1]	$\mathcal{O}(\log\log n)$	[29, Thm 2]
$(1+\beta)$	$\beta \leqslant 1-\Omega(1)$	$\Omega\big(\frac{\log n}{\beta}\big)$	[152, Sec 4]	$\mathcal{O}\big(\frac{\log n}{\beta} + \frac{\log(1/\beta)}{\beta}\big)$	[152, Cor 2.12]
$(1+\beta)$	$\beta \leqslant 1-\Omega(1)$	–	–	$\mathcal{O}\big(\frac{\log n}{\beta}\big)$	Thm 3.7
$(1+\beta)$	$\beta = 1-e^{-\frac{1}{4}(\log n)^{1-1/k}}$ for $k \in [1, \log\log n]$	–	–	$\mathcal{O}\big(k \cdot (\log n)^{1/k}\big)$	Thm 7.23
$(1+\beta)$	$\beta = 1-\Omega(e^{-\frac{1}{4}\log^{1-1/k} n})$ for $1-\frac{1}{k} \in [\Omega(1), 1)$	$\Omega\big(k \cdot (\log n)^{1/k}\big)$	Lem C.21	–	–
$(1+\beta)$	$\beta = 1-\mathrm{poly}(n^{-1})$	$\Omega(\log\log n)$	Obs C.7	$\mathcal{O}(\log\log n)$	Thm 7.23
Memory	–	$\Omega(\log\log n)$	Thm C.26	$\mathcal{O}(\log\log n)$	Thm 7.48
$(1,1,2)$-Reset-Memory	–	$\Omega(\log n)$	Lem C.27	$\mathcal{O}(\log n)$	Thm 7.23
Packing	–	$\Omega\big(\frac{\log n}{\log\log n}\big)$	[117, Thm 10.2]	$\mathcal{O}(\log n)$	[117, Thm 4.3]
Twinning	–	$\Omega(\log n)$	Cor C.24	$\mathcal{O}(\log n)$	Cor 5.12
Mean-Thinning	–	$\Omega(\log n)$	Cor C.24	$\mathcal{O}(\log n)$	Cor 5.12
Relative-Threshold$(f(n))$	$f(n) \geqslant 0$	–	–	$f(n)+\mathcal{O}(\log n)$	Cor 5.13
Relative-Threshold$(f(n))$	$f(n) \geqslant \log n$	$\Omega(f(n))$	Lem C.25	–	–
Two-Thinning(f^t)	–	$\Omega\big(\frac{\log n}{\log\log n}\big)$	Thm C.15	–	–
Two-Thinning(f^t)	–	–	–	$\mathcal{O}\big(\frac{\log n}{\log\log n}\big)$	[76, Thm 1]
Quantile(δ)	–	–	–	$\mathcal{O}\big(\frac{\log n}{\delta}\big)$	Cor 3.9
Quantile(δ^*)	$\delta^* = \frac{(\log\log n)^2}{\log n}$	–	–	$\mathcal{O}\big(\frac{\log n}{\log\log n}\big)$	Thm 7.19
k-Dense-Quantile	$k \in [1, \log\log n]$	–	–	$\mathcal{O}\big(k \cdot (\log n)^{1/k}\big)$	Thm 7.11
k-Dense-Threshold	$k \in [1, \log\log n]$	–	–	$\mathcal{O}\big(k \cdot (\log n)^{1/k}\big)$	Thm 7.14
d-Thinning	–	–	–	$\mathcal{O}\big((d-1)\cdot(\log n)^{1/(d-1)}\big)$	Lem 7.21
d-Thinning	–	–	–	$\Omega\big(d \cdot \big(\frac{d \log n}{\log\log n}\big)^{1/d}\big)$	[77, Prop 4.1]

Table A.1: Overview of the lower and upper bounds for different processes derived in previous works (rows in Gray) and in this work (rows in Green). All upper bounds hold for all values of $m \geqslant n$ w.h.p., while lower bounds may only hold for a suitable value of m w.h.p.. In all the above.

Process	Graphical	Batch size	Weights	Gap bound	Reference
Two-Choice	d-reg., conduct. ϕ	–	–	$\mathcal{O}\big(\frac{\log n}{\phi}\big)$	[152, Thm 3.2]
Two-Choice	d-reg., conduct. ϕ	–	random	$\mathcal{O}\big(\frac{\log n}{\phi}\big)$	Thm 7.41
Two-Choice	d-reg., conduct. ϕ	$b \geqslant n$	random	$\mathcal{O}\big(\frac{b}{n} \cdot \frac{\log n}{\phi}\big)$	Thm 7.42
Two-Choice	expander, $d = \mathcal{O}(1)$	$b \in [n, n^3]$	random	$\mathcal{O}\big(\frac{b}{n} + \log n\big)$	Thm 7.42
Two-Choice	expander, spectral gap λ $e^{-\frac{1}{4}(\log n)^{1-1/k}} \geqslant \lambda$ for $k \in [1, \log\log n]$	–	–	$\mathcal{O}\big(k \cdot (\log n)^{1/k}\big)$	Thm 7.46
Two-Choice	expander, spectral gap λ $\lambda = \mathrm{poly}(n^{-1})$	–	–	$\mathcal{O}(\log\log n)$	Cor 7.47

Table A.3: Overview of the gap bounds in previous works (rows in Gray) and the gap bounds derived in this work (rows in Green) for graphical allocations. The upper bounds on the gap hold for all values of m, while some of the lower bounds may only hold for certain m.

Setting	Range	Lower bound	Reference	Upper bound	Reference
g-Bounded	$1 \leqslant g$	–	–	$\mathcal{O}(g \cdot \log(ng))$	[142]
g-Adv-Comp	$1 \leqslant g$	–	–	$\mathcal{O}(g \cdot \log(ng))$	Thm 3.21
g-Adv-Comp	$1 \leqslant g$	–	–	$\mathcal{O}(g + \log n)$	Thm 5.26
g-Adv-Comp	$1 < g \leqslant \log n$	–	–	$\mathcal{O}\big(\frac{g}{\log g} \cdot \log\log n\big)$	Thm 7.25
g-Myopic-Comp	$\frac{\log n}{\log\log n} \leqslant g$	$\Omega(g)$	Pro C.8	–	–
g-Myopic-Comp	$1 < g \leqslant \frac{\log n}{\log\log n}$	$\Omega\big(\frac{g}{\log g} \cdot \log\log n\big)$	Obs C.7 Thm C.9	–	–
σ-Noisy-Load	$1 \leqslant \sigma$	–	–	$\mathcal{O}\big(\sigma \sqrt{\log n} \cdot \log(n\sigma)\big)$	Pro 7.29
σ-Noisy-Load	$2 \cdot (\log n)^{-1/3} \leqslant \sigma$	$\Omega\big(\min\{1, \sigma\} \cdot (\log n)^{1/3}\big)$	Pro C.11	–	–
σ-Noisy-Load	$4 \leqslant \sigma$	$\Omega\big(\min\{\sigma^{4/5}, \sigma^{2/5} \cdot \sqrt{\log n}\}\big)$	Pro C.11	–	–

Table A.2: Overview of the lower and upper bounds for different noise settings for Two-Choice derived in previous works (rows in Gray) and in this work (rows in Green). All upper bounds hold for all values of $m \geqslant n$ w.h.p., while lower bounds may only hold for a suitable value of m w.h.p.. Recall that g-Bounded and g-Myopic-Comp can be simulated by g-Adv-Comp and $(2g)$-Adv-Comp, respectively.

Process	Range	Lower bound	Reference	Upper bound	Reference
TWO-CHOICE	$b = n$	–	–	$\mathcal{O}(\log n)$	[28, Thm 1]
b-BATCHED $\mathcal{C}_1, \mathcal{C}_2$	$b \geqslant n$	–	–	$\mathcal{O}\big(\frac{b}{n} \cdot \log n\big)$	Thm 3.14
b-BATCHED $\mathcal{C}_1, \mathcal{C}_2$	$b \in [n, n^3]$	–	–	$\mathcal{O}\big(\frac{b}{n} + \log n\big)$	Thm 7.39
b-BATCHED $\mathcal{C}_1, \mathcal{C}_3$	$b \geqslant n \log n$	–	–	$\mathcal{O}\big(\sqrt{\frac{b}{n}} \cdot \log n\big)$	Cor 3.18
b-BATCHED $\mathcal{C}_1, \mathcal{C}_3$	$b \geqslant n \log n$	–	–	$\mathcal{O}\big(\sqrt{\frac{b}{n}} \cdot \log n\big)$	Cor 7.36
$(1+\beta)$, $\beta \leqslant 1 - \Omega(1)$	$b \geqslant 1$	$\Omega(\log n)$	–	–	Prop C.2
TWO-CHOICE, $(1+\beta)$, $\beta = \Omega(1)$	$b \geqslant n \log n$	$\Omega(\frac{b}{n})$	–	–	Prop C.3
τ-DELAY(TWO-CHOICE)	$\tau = n$	–	–	$\mathcal{O}\big(\frac{\log n}{\log \log n}\big)$	Thm 7.30
τ-DELAY(TWO-CHOICE)	$\tau \in [n \cdot e^{-(\log n)^c}, n \log n]$	–	–	$\mathcal{O}\Big(\frac{\log n}{\log((4n/\tau) \log n)}\Big)$	Rem 7.31
τ-DELAY(TWO-CHOICE)	$\tau = n^{1-\epsilon}$	–	–	$\mathcal{O}(\log \log n)$	Rem 7.33
b-BATCHED(TWO-SAMPLE)	$b \geqslant n \log n$	$\Omega(\sqrt{\frac{b}{n}} \cdot \log n)$	Obs C.5	–	–
b-BATCHED(TWO-CHOICE)	$b = n$	$\Omega(\frac{\log n}{\log \log n})$	Obs C.5	–	–
b-BATCHED(TWO-CHOICE)	$b \in [n \cdot e^{-(\log n)^c}, n \log n]$	$\Omega\Big(\frac{\log n}{\log((4n/b) \log n)}\Big)$	[157, Thm 1]	–	–
b-BATCHED(TWO-CHOICE)	$b = n^{1-\epsilon}$	$\Omega(\log \log n)$	Obs C.7	–	–

Table A.4: Overview of the gap bounds in previous works (rows in Gray) and the gap bounds derived in this work (rows in Green). The upper bounds on the gap hold for all values of m, while some of the lower bounds may only hold for certain m. The parameters c, ϵ can be any constant in $(0, 1)$.

TOOLS

B.1 Auxiliary lemmas

We begin with a simple lemma for a sequence of random variables whose expectation satisfies a recurrence inequality.

Lemma B.1. *Consider any sequence of random variables* $(X^i)_{i\in\mathbb{N}}$ *for which there exist* $0 < a < 1$ *and* $b > 0$*, such that every* $i \geqslant 1$*,*

$$\mathbf{E}\left[X^i \mid X^{i-1} \right] \leqslant X^{i-1} \cdot a + b.$$

Then, (i) *for every* $i \geqslant 0$*,*

$$\mathbf{E}\left[X^i \mid X^0 \right] \leqslant X^0 \cdot a^i + \frac{b}{1-a}.$$

Further, (ii) *if* $X^0 \leqslant \frac{b}{1-a}$ *holds, then for every* $i \geqslant 0$*,*

$$\mathbf{E}\left[X^i \right] \leqslant \frac{b}{1-a}.$$

Proof. First statement. We will prove by induction that for every $i \in \mathbb{N}$,

$$\mathbf{E}\left[X^i \mid X^0 \right] \leqslant X^0 \cdot a^i + b \cdot \sum_{j=0}^{i-1} a^j.$$

For $i = 0$, $\mathbf{E}\left[X^0 \mid X^0 \right] \leqslant X^0$. Assuming the induction hypothesis holds for some $i \geqslant 0$, then since $a > 0$,

$$\begin{aligned}
\mathbf{E}\left[X^{i+1} \mid X^0 \right] &= \mathbf{E}\left[\mathbf{E}\left[X^{i+1} \mid X^i \right] \mid X^0 \right] \leqslant \mathbf{E}\left[X^i \mid X^0 \right] \cdot a + b \\
&\leqslant \left(X^0 \cdot a^i + b \cdot \sum_{j=0}^{i-1} a^j \right) \cdot a + b \\
&= X^0 \cdot a^{i+1} + b \cdot \sum_{j=0}^{i} a^j.
\end{aligned}$$

The claims follows using that $\sum_{j=0}^{i} a^j \leqslant \sum_{j=0}^{\infty} a^j = \frac{1}{1-a}$, for any $a \in (0,1)$.

Second statement. We will prove this claim by induction. Then, assuming that $\mathbf{E}\left[X^i \right] \leqslant \frac{b}{1-a}$ holds for $i \geqslant 0$, we have for $i+1$

$$\mathbf{E}\left[X^{i+1} \right] = \mathbf{E}\left[\mathbf{E}\left[X^{i+1} \mid X^i \right] \right] \leqslant \mathbf{E}\left[X^i \right] \cdot a + b \leqslant \frac{b}{1-a} \cdot a + b = \frac{b}{1-a}. \qquad \square$$

Next, we proceed with a standard fact, whose proof we give for completeness.

Lemma B.2. *Let* $p, q \in \mathbb{R}^n$ *be two probability vectors such that* $p \succeq q$ *and* $c \in \mathbb{R}^n$ *be non-negative and non-increasing. Then,*

$$\langle p, c \rangle \geqslant \langle q, c \rangle.$$

Proof. We will consider a sequence of moves between p and q, which gradually moves probability mass from lower to higher coordinates. Specifically, we define the following sequence:

$$\begin{aligned}
r^1 &= (p_1, p_2, p_3, p_4, \ldots, p_n) = p \\
r^2 &= (q_1, p_2 + (p_1 - q_1), p_3, p_4, \ldots, p_n) \\
r^3 &= (q_1, q_2, p_3 + (p_1 + p_2 - q_1 - q_2), p_4, \ldots, p_n) \\
r^4 &= (q_1, q_2, q_3, p_4 + (p_1 + p_2 + p_3 - q_1 - q_2 - q_3), \ldots, p_n) \\
&\vdots \\
r^n &= \left(q_1, q_2, q_3, \ldots, q_{n-1}, p_n + \sum_{i=1}^{n-1}(p_i - q_i)\right) = q,
\end{aligned}$$

where in the last equation we used $p_n + \sum_{i=1}^{n-1}(p_i - q_i) = p_n - p_n + q_n = q_n$.

For any $1 \leqslant k < n$, since r^k and r^{k+1} differ only in the k-th and $(k+1)$-st coordinate, and $\sum_{i=1}^{k}(p_i - q_i) \geqslant 0$, we conclude it follows that

$$\begin{aligned}
\langle r^k, c\rangle - \langle r^{k+1}, c\rangle &\geqslant r^k_k c_k + r^k_{k+1} c_{k+1} - r^{k+1}_k c_k + r^{k+1}_{k+1} c_{k+1} \\
&= c_k \cdot \left(\left(p_k + \sum_{i=1}^{k-1}(p_i - q_i)\right) - q_k\right) + c_{k+1} \cdot \left(p_{k+1} - \left(p_{k+1} + \sum_{i=1}^{k}(p_i - q_i)\right)\right) \\
&= (c_k - c_{k+1}) \cdot \sum_{i=1}^{k}(p_i - q_i) \\
&\geqslant 0.
\end{aligned}$$

Hence $\langle p, c\rangle = \langle r^1, c\rangle \geqslant \langle r^2, c\rangle \geqslant \cdots \geqslant \langle r^n, c\rangle = \langle q, c\rangle$.

□

Lemma B.3. *The function $f(z) = z \cdot e^{k/z}$ for $k > 0$, is decreasing for $z \in (0, k]$.*

Proof. By differentiating,

$$f'(z) = e^{k/z} - z \cdot e^{k/z} \cdot \frac{k}{z^2} = e^{k/z} \cdot \left(1 - \frac{k}{z}\right).$$

For $z \in (0, k)$, $f'(z) < 0$, so f is decreasing. □

B.2 Concentration inequalities

In this section, we state several well-known concentration inequalities.

Lemma B.4 (Multiplicative factor Chernoff Binomial Bound [137]). *Let $X^1, \ldots, X^n$ be independent binary random variables with $\mathbf{Pr}\left[X^i = 1\right] = p$. Then,*

$$\mathbf{Pr}\left[\sum_{i=1}^{n} X^i \geqslant npe\right] \leqslant e^{-np},$$

and

$$\mathbf{Pr}\left[\sum_{i=1}^{n} X^i \leqslant \frac{np}{e}\right] \leqslant e^{\left(\frac{2}{e}-1\right)np}.$$

Next we state a Chernoff bound for Poisson random variables.

Lemma B.5 (Theorem 5.4 from [138]). *Let $X \sim \text{Pois}(\lambda)$, then for any $0 < \delta < 1$,*

$$\mathbf{Pr}[\, X \leqslant (1-\delta)\cdot\lambda \,] \leqslant e^{-\lambda\delta^2/2},$$

and

$$\mathbf{Pr}[\, X \geqslant (1+\delta)\cdot\lambda \,] \leqslant e^{-\lambda\delta^2/3}.$$

Theorem B.6 (Berry-Esseen [71]). *Let $X^1,\ldots,X^n$ be a sequence of i.i.d. random variables with mean μ, variance σ^2 and central moment $\rho = \mathbf{E}[\,|X^i-\mu|^3\,]$. Then, there exists a constant $C>0$ such that for $\alpha\in\mathbb{R}$*

$$\left|\mathbf{Pr}\left[\frac{\frac{1}{n}\sum_{i=1}^n X^i-\mu}{\frac{\sigma}{\sqrt{n}}}\leqslant\alpha\right]-\widetilde{\Phi}(\alpha)\right|\leqslant C\cdot\frac{\rho}{\sigma^3\sqrt{n}},$$

where $\widetilde{\Phi}$ is the cumulative distribution of the standard normal distribution.

Lemma B.7 (Berry-Esseen for Poisson r.vs). *Let $X\sim\text{Po}(m)$, where $m\in\mathbb{N}$, then for any $\alpha\in\mathbb{R}$*

$$\left|\mathbf{Pr}\left[X\leqslant m+\alpha\sqrt{m}\right]-\widetilde{\Phi}(\alpha)\right|\leqslant C\cdot\frac{\rho}{\sigma^3\sqrt{m}}.$$

Proof. The sum of n independent Poisson r.vs. with parameters $(k_i)_{i=1}^n$ is a Poisson r.v. with parameter $\sum_{i=1}^n k_i$ (e.g. [138, Lemma 5.2]). Hence, we can write X as the sum of m r.vs. $X^i\sim\text{Po}(1)$. Then, applying Theorem B.6 gives,

$$\left|\mathbf{Pr}\left[\frac{\frac{\sum_{i=1}^n X^i}{m}-\mu}{\frac{\sigma}{\sqrt{m}}}\leqslant\alpha\right]-\widetilde{\Phi}(\alpha)\right|=\left|\mathbf{Pr}\left[X\leqslant m+\alpha\sqrt{m}\right]-\widetilde{\Phi}(\alpha)\right|\leqslant C\cdot\frac{\rho}{\sigma^3\sqrt{m}}.$$

□

The next lemma is a standard Chernoff bound for sum of independent random variables whose moment generating function is bounded.

Lemma B.8. *Assume $X^1,X^2,\ldots,X^k$ are independent samples from a distribution $\mathcal{W}$, for which there is a constant $\lambda>0$ such that $\mathbf{E}[\,\mathcal{W}\,]=1$ and $\mathbf{E}[\,e^{\lambda\mathcal{W}}\,]\leqslant S$. Then for $X:=\sum_{i=1}^k X^i$, it holds for that*

$$\mathbf{Pr}[\, X\geqslant 2\log(S)/\lambda\cdot k\,]\leqslant\exp(-\log(S)\cdot k).$$

Furthermore, for the special case $k=1$, we have for any $c>0$,

$$\mathbf{Pr}\left[X^1\geqslant 1/\lambda\cdot(c\cdot\log(n)+\log(S))\right]\leqslant n^{-c}.$$

Proof. Let $t\in(0,\lambda]$ to be specified later. Then,

$$\begin{aligned}
\mathbf{Pr}[\, X\geqslant 2\log(S)/\lambda\cdot k\,] &= \mathbf{Pr}\left[e^{tX}\geqslant e^{t\cdot 2\log(S)/\lambda\cdot k}\right]\\
&\leqslant \mathbf{E}\left[e^{tX}\right]\cdot\exp(-t\cdot 2\log(S)/\lambda\cdot k)\\
&= \left(\mathbf{E}\left[e^{tX^1}\right]\right)^k\cdot\exp(-t\cdot 2\log(S)/\lambda\cdot k)\\
&\leqslant \left(\mathbf{E}[\,e^{\lambda X^1}\,]\right)^{k\cdot t/\lambda}\cdot\exp(-t\cdot 2\log(S)/\lambda\cdot k)
\end{aligned}$$

$$\leqslant S^{k\cdot t/\lambda} \cdot \exp\left(-t \cdot 2\log(S)/\lambda \cdot k\right)$$
$$= \exp\left(k \cdot (\log(S) \cdot t/\lambda - t \cdot 2\log(S)/\lambda)\right),$$

where the second inequality is due to Jensen's inequality. Choosing $t = \lambda$ yields the claim.

For the second statement, for any $c > 0$,

$$\begin{aligned}\mathbf{Pr}\left[\, X^1 \geqslant 1/\lambda \cdot (c \cdot \log(n) + \log(S)) \,\right] &\leqslant \mathbf{Pr}\left[\, e^{\lambda X^1} \geqslant e^{-c\cdot\log(n) - \log(S)} \,\right] \\ &\leqslant \mathbf{E}\left[\, e^{\lambda W} \,\right] \cdot e^{-c\cdot\log(n)-\log(S)} \\ &\leqslant S \cdot n^{-c} \cdot \frac{1}{S} = n^{-c}. \end{aligned}$$
□

B.3 Concentration inequalities for martingales

In this section, we state several well-known inequalities for martingales.

Lemma B.9 ([51, Theorems 6.1 & 6.5]). *Consider a martingale $X^0, \ldots, X^N$ with filtration $\mathfrak{F}^0, \ldots, \mathfrak{F}^N$ satisfying $|X^i - X^{i-1}| \leqslant M$ and $\mathbf{Var}\left[\, X^i \mid \mathfrak{F}^{i-1} \,\right] \leqslant \sigma_i^2$ for any $i \in [N]$, then for any $\lambda > 0$,*

$$\mathbf{Pr}\left[\, \left|X^N - \mathbf{E}\left[\, X^N \,\right]\right| \geqslant \lambda \,\right] \leqslant 2 \cdot \exp\left(-\frac{\lambda^2}{2 \cdot (\sum_{i=1}^N \sigma_i^2 + M\lambda/3)}\right).$$

Lemma B.10 (Azuma's Inequality for Super-Martingales [67, Problem 6.5]). *Let $X^0, \ldots, X^n$ be a super-martingale satisfying $|X^i - X^{i-1}| \leqslant c_i$ for any $i \in [n]$, then for any $\lambda > 0$,*

$$\mathbf{Pr}\left[\, X^n \geqslant X^0 + \lambda \,\right] \leqslant \exp\left(-\frac{\lambda^2}{2 \cdot \sum_{i=1}^n c_i^2}\right).$$

A concentration inequality with a bad event. Following [107], we will now give the definition for *strongly difference-bounded* and then give the statement for a bounded differences inequality with bad events.

Definition B.11 (Strongly difference-bounded – Definition 1.6 in [107]). *Let $\Omega_1, \ldots, \Omega_N$ be probability spaces. Let $\Omega = \prod_{k=1}^N \Omega_k$ and let X be a random variable on Ω. We say that X is strongly difference-bounded by (η_1, η_2, ξ) if the following holds: there is a "bad" subset $\mathcal{B} \subseteq \Omega$, where $\xi = \mathbf{Pr}[\, \omega \in \mathcal{B} \,]$. If $\omega, \omega' \in \Omega$ differ only in the k-th coordinate, and $\omega \notin B$, then*

$$|X(\omega) - X(\omega')| \leqslant \eta_2.$$

Furthermore, for any ω and ω' differing only in the k-th coordinate,

$$|X(\omega) - X(\omega')| \leqslant \eta_1.$$

Theorem B.12 (Theorem 3.3 in [107]). *Let $\Omega_1, \ldots, \Omega_N$ be probability spaces. Let $\Omega = \prod_{k=1}^N \Omega_k$, and let X be a random variable on Ω which is strongly difference-bounded by (η_1, η_2, ξ). Let $\mu = \mathbf{E}[\, X \,]$. Then for any $\lambda > 0$ and any $\gamma_1, \ldots, \gamma_N > 0$,*

$$\mathbf{Pr}[\, X \geqslant \mu + \lambda \,] \leqslant \exp\left(-\frac{\lambda^2}{2 \cdot \sum_{k\in[N]} (\eta_2 + \eta_1 \gamma_k)^2}\right) + \xi \cdot \sum_{k\in[N]} \frac{1}{\gamma_k}.$$

B.4 Probabilistic inequalities

We start with showing that a random variable with bounded MGF also has bounded fourth moment.

Lemma B.13. *Consider a random variable W with $\mathbf{E}\left[\, e^{\lambda W} \,\right] < \infty$ for some $\lambda > 0$. then*

$$\mathbf{E}\left[\, W^4 \,\right] < \left(\left(\frac{8}{\lambda}\right) \cdot \log\left(\frac{8}{\lambda}\right)\right)^4 + \mathbf{E}\left[\, e^{\lambda W} \,\right].$$

Proof. Let $\kappa := (8/\lambda) \cdot \log(8/\lambda)$. Consider $x \geqslant \max(0, \kappa) =: \kappa^*$. Then

$$e^{\lambda x/4} = e^{\lambda x/8} \cdot e^{\lambda x/8} \geqslant e^{\log(8/\lambda)} \cdot e^{\lambda x/8} \geqslant \frac{8}{\lambda} \cdot \frac{\lambda x}{8} = x,$$

using that $e^z \geqslant z$ for any z. Hence,

$$e^{\lambda x} = (e^{\lambda x/4})^4 \geqslant x^4.$$

Hence, if p_x is the pdf of W, then

$$\begin{aligned}
\mathbf{E}\left[\, W^4 \,\right] &= \int_{x=0}^{\infty} x^4 \cdot p_x dx = \int_{x=0}^{\kappa^*} x^4 \cdot p_x dx + \int_{x=\kappa^*}^{\infty} x^4 \cdot p_x dx \\
&\leqslant \int_{x=0}^{\kappa^*} \kappa^4 \cdot p_x dx + \int_{x=\kappa^*}^{\infty} e^{\lambda x} \cdot p_x dx \\
&\leqslant \kappa^4 \cdot \int_{x=0}^{\infty} p_x dx + \int_{x=0}^{\infty} e^{\lambda x} \cdot p_x dx = \kappa^4 + \mathbf{E}\left[\, e^{\lambda W} \,\right].
\end{aligned}$$

□

Lemma B.14. *Consider any $n \geqslant 2$ and $\lambda \geqslant 16 \cdot \log n$. Let $X^1, \ldots, X^n$ be independent Poisson random variables with $X^i \sim \mathrm{Pois}(\lambda)$, and denote by for $Y_{(n)}, Y_{(n-1)}$ the smallest and second smallest of the X^i's. Then there exist constants $\kappa_1, \kappa_2 > 0$ such that,*

$$\mathbf{Pr}\left[\, Y_{(n-1)} - Y_{(n)} \geqslant \kappa_1 \cdot \sqrt{\lambda / \log n} \,\right] \geqslant \kappa_2.$$

Proof. Let $X \sim \mathrm{Pois}(\lambda)$, where $\lambda := m/n \geqslant 16 \cdot \log n$. Let $k \geqslant 0$ be the minimal integer such that

$$\mathbf{Pr}\left[\, \mathrm{Pois}(\lambda) \leqslant k \,\right] \geqslant n^{-1}.$$

By Lemma B.5 for $\delta := \sqrt{4 \cdot \lambda^{-1} \cdot \log n}$, we have

$$\mathbf{Pr}\left[\, X \leqslant \lambda - \sqrt{4 \cdot \lambda \cdot \log n} \,\right] \leqslant e^{-\lambda \cdot \delta^2/2} = e^{-2 \log n} = n^{-2}.$$

Hence it follows that $k \geqslant \lambda - 2 \cdot \sqrt{\lambda \cdot \log n}$. Next note that

$$\frac{\mathbf{Pr}\left[\, \mathrm{Pois}(\lambda) = k+1 \,\right]}{\mathbf{Pr}\left[\, \mathrm{Pois}(\lambda) = k \,\right]} = \frac{\lambda}{k+1}, \tag{B.1}$$

which, since $k \geqslant \frac{1}{2}\lambda$ (as $\lambda \geqslant 16 \log n$), also implies that

$$\mathbf{Pr}\left[\, \mathrm{Pois}(\lambda) \leqslant k \,\right] \leqslant 2 \cdot n^{-1}.$$

Our next claim is that

$$\mathbf{Pr}\left[\, \mathrm{Pois}(\lambda) = k \,\right] \leqslant 2 \cdot n^{-1} \cdot 1 / \sqrt{\lambda / \log n}.$$

We will now derive this claim. We have

$$
\begin{aligned}
2n^{-1} &\geqslant \mathbf{Pr}[\,\mathrm{Pois}(\lambda) \leqslant k\,] \\
&= \sum_{j=0}^{k} \mathbf{Pr}[\,\mathrm{Pois}(\lambda) = j\,] \\
&\overset{(a)}{=} \sum_{j=0}^{k} \mathbf{Pr}[\,\mathrm{Pois}(\lambda) = k\,] \cdot \prod_{i=j}^{k-1} \frac{i}{\lambda} \\
&\geqslant \sqrt{\lambda/\log n} \cdot \mathbf{Pr}[\,\mathrm{Pois}(\lambda) = k\,] \cdot \left(\frac{k-\sqrt{\lambda}}{\lambda}\right)^{\sqrt{\lambda/\log n}} \\
&\overset{(b)}{\geqslant} \sqrt{\lambda/\log n} \cdot \mathbf{Pr}[\,\mathrm{Pois}(\lambda) = k\,] \cdot \left(\frac{\lambda - 3\sqrt{\lambda \log n}}{\lambda}\right)^{\sqrt{\lambda/\log n}} \\
&= \sqrt{\lambda/\log n} \cdot \mathbf{Pr}[\,\mathrm{Pois}(\lambda) = k\,] \cdot \left(1 - \frac{3}{\sqrt{\lambda/\log n}}\right)^{\sqrt{\lambda/\log n}} \\
&\overset{(c)}{\geqslant} \sqrt{\lambda/\log n} \cdot \mathbf{Pr}[\,\mathrm{Pois}(\lambda) = k\,] \cdot c_1,
\end{aligned}
$$

for some constant $c_1 > 0$, where in (a) we used Eq. (B.1) and in (b) we used that $k \geqslant \lambda - 2\sqrt{\lambda \log n}$, and in (c) that $\lambda \geqslant 16 \log n$.

Next we wish to upper bound

$$
\mathbf{Pr}\Big[\,\mathrm{Pois}(\lambda) \leqslant k + c_2 \cdot \sqrt{\lambda/\log n}\,\Big],
$$

for some constant $c_2 > 0$. Note that

$$
\begin{aligned}
&\mathbf{Pr}\Big[\,\mathrm{Pois}(\lambda) \leqslant k + c_2 \cdot \sqrt{\lambda/\log n}\,\Big] \\
&\overset{(a)}{\leqslant} \mathbf{Pr}[\,\mathrm{Pois}(\lambda) \leqslant k\,] + \sum_{i=1}^{c_2\sqrt{\lambda/\log n}} \frac{\lambda^i}{k \cdot (k+1) \cdot \ldots \cdot (k+i-1)} \cdot \mathbf{Pr}[\,\mathrm{Pois}(\lambda) = k\,] \\
&\leqslant 2n^{-1} + c_2 \cdot \sqrt{\lambda/\log n} \cdot \frac{\lambda^{c_2\sqrt{\lambda/\log n}}}{k^{c_2\sqrt{\lambda/\log n}}} \cdot 2 \cdot n^{-1} \cdot 1/\sqrt{\lambda/\log n} \\
&= 2n^{-1} + 2c_2 \cdot \left(1 - \frac{1}{c\sqrt{\lambda \cdot \log n}}\right)^{-c_2\sqrt{\lambda/\log n}} \cdot n^{-1} \\
&\leqslant c_3 \cdot n^{-1},
\end{aligned}
\tag{B.2}
$$

for another constant $c_3 > 0$, where (a) is due to Eq. (B.1).

We now use the principle of deferred decisions when exposing the n independent Poisson variables with mean λ denoted by $X^1, X^2, \ldots, X^n$ one by one. Let $\tau := \min\{j : X^j \leqslant k\}$. With probability $1-(1-1/n)^n \geqslant 1 - 1/e$, we have $\tau < n$. Conditional on that, $X^{\tau+1}, \ldots, X^n$ are still $n - \tau$ independent Poisson variables with mean λ. Due to Eq. (B.2), the probability that all of the following Poisson random variables are larger than $k + c_2 \cdot \sqrt{\lambda/\log n}$ is at least

$$
\left(1 - c_3 \cdot n^{-1}\right)^{\tau} \geqslant \left(1 - c_3 \cdot n^{-1}\right)^{n} \geqslant c_4,
$$

where $c_4 > 0$ is another constant.

Hence with probability at least $(1-1/e)\cdot c_4$, we have a gap of at least $c_2 \cdot \sqrt{\lambda/\log n}$ between $Y_{(n-1)}$ and $Y_{(n)}$. $\square$

B.5 Facts about the One-Choice process

In this section, we collect several facts about the One-Choice process. We first restate the so-called Poisson approximation method.

Lemma B.15 ([138, Corollary 5.11]). *Let $(x^T)_{i\in[n]}$ be the load vector after T steps of One-Choice. Further, let $(\widetilde{x}^T)_{i\in[n]}$ be n independent Poisson random variables with parameter $\lambda = T/n$ each. Further, let $\mathcal{E}$ be any event which is determined by x^T, and further assume that $\mathbf{Pr}[\,\mathcal{E}\,]$ is either monotonically increasing in T or monotonically decreasing in T. Further, let $\widetilde{\mathcal{E}}$ be the corresponding event determined by $\widetilde{x}^T$. Then,*

$$\mathbf{Pr}[\,\mathcal{E}\,] \leqslant 2 \cdot \mathbf{Pr}\Big[\,\widetilde{\mathcal{E}}\,\Big].$$

B.5.1 Maximum load: The $m = cn\log n$ balls case

The next standard result was also used in [152, Section 4] and is based on [157]. For convenience of the reader, we give the self-contained proof for obtaining high probability bounds.

Lemma B.16. *Consider the One-Choice process for $m = cn\log n$ balls where $c \geqslant 1/\log n$. Then,*

$$\mathbf{Pr}\left[\,\mathrm{Gap}(m) \geqslant \frac{\sqrt{c}}{10}\cdot\log n\,\right] \geqslant 1 - n^{-2}.$$

Proof. In order to use the Poisson Approximation [138, Chapter 5], let $\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_n$ be n independent Poisson random variables with parameter $\lambda = \frac{m}{n} = c\log n$. Then,

$$\mathbf{Pr}\left[\,\hat{x}_i \geqslant \lambda + \frac{\sqrt{c}}{10}\cdot\log n\,\right] \geqslant \mathbf{Pr}\left[\,\hat{x}_i = \lambda + \frac{\sqrt{c}}{10}\cdot\log n\,\right] = e^{-\lambda}\cdot\frac{\lambda^{\lambda+\frac{\sqrt{c}}{10}\cdot\log n}}{(\lambda+\frac{\sqrt{c}}{10}\cdot\log n)!}.$$

Using that $z! \leqslant \sqrt{2\pi z}\left(\frac{z}{e}\right)^z e^{\frac{1}{12z}}$ for any integer $z \geqslant 1$,

$$\begin{aligned}
\mathbf{Pr}\left[\,\hat{x}_i = \lambda + \frac{\sqrt{c}}{10}\cdot\log n\,\right] &\geqslant \frac{1}{4\cdot\sqrt{2\pi\lambda}}\cdot e^{-\lambda}\cdot\left(\frac{e\lambda}{\lambda+\frac{\sqrt{c}}{10}\cdot\log n}\right)^{\lambda+\frac{\sqrt{c}}{10}\cdot\log n} \\
&\geqslant \frac{1}{4\cdot\sqrt{2\pi\lambda}}\cdot e^{\frac{\sqrt{c}}{10}\log n}\cdot\left(1+\frac{1}{10\sqrt{c}}\right)^{-\lambda-\frac{\sqrt{c}}{10}\cdot\log n} \\
&\geqslant \frac{1}{4\cdot\sqrt{2\pi\lambda}}\cdot e^{\frac{\sqrt{c}}{10}\log n}\cdot e^{-\frac{1}{10\sqrt{c}}\cdot(\lambda+\frac{\sqrt{c}}{10}\cdot\log n)} \\
&\geqslant \frac{1}{4\cdot\sqrt{2\pi\lambda}}\cdot e^{\frac{\sqrt{c}}{10}\log n-\frac{1}{10\sqrt{c}}\lambda-\frac{1}{100}\log n} \\
&\geqslant \frac{1}{4\cdot\sqrt{2\pi\lambda}}\cdot e^{-\frac{1}{100}\log n}
\end{aligned}$$

Since for any $k \geqslant 0$,

$$\frac{\mathbf{Pr}[\,\hat{x}_i = k+1\,]}{\mathbf{Pr}[\,\hat{x}_i = k\,]} = \frac{\lambda}{k+1},$$

we conclude that

$$\mathbf{Pr}\left[\,\hat{x}_i \geqslant \lambda + \frac{\sqrt{c}}{10}\cdot\log n\,\right] \geqslant \sum_{k=0}^{\sqrt{\lambda}-1}\mathbf{Pr}\left[\,\hat{x}_i = \lambda + \frac{\sqrt{c}}{10}\cdot\log n + k\,\right]$$

$$
\begin{aligned}
&\geqslant \sqrt{\lambda}\cdot \mathbf{Pr}\left[\, \hat{x}_i = \lambda + \frac{\sqrt{c}}{10}\cdot \log n + \sqrt{\lambda}\,\right] \\
&\geqslant \sqrt{\lambda}\cdot \mathbf{Pr}\left[\, \hat{x}_i = \lambda + \frac{\sqrt{c}}{10}\cdot \log n\,\right]\cdot \prod_{k=1}^{\sqrt{\lambda}}\left(\frac{\lambda}{\lambda + \frac{\sqrt{c}}{10}\cdot \log n + k}\right) \\
&\geqslant \sqrt{\lambda}\cdot \frac{1}{4\cdot\sqrt{2\pi\lambda}}\cdot e^{-\frac{1}{100}\log n}\cdot \left(\frac{\lambda}{\lambda + \frac{\sqrt{c}}{10}\cdot \log n + \sqrt{\lambda}}\right)^{\sqrt{\lambda}} \\
&\geqslant \sqrt{\lambda}\cdot \frac{1}{4\cdot\sqrt{2\pi\lambda}}\cdot e^{-\frac{1}{100}\log n}\cdot \left(1 + \frac{1}{5\sqrt{c}}\right)^{-\sqrt{\lambda}} \\
&\geqslant \sqrt{\lambda}\cdot \frac{1}{4\cdot\sqrt{2\pi\lambda}}\cdot e^{-\frac{1}{100}\log n}\cdot e^{-\frac{1}{5}\sqrt{\log n}} \\
&\geqslant e^{-\frac{1}{99}\log n} = n^{-1/99},
\end{aligned}
$$

where the last inequality holds for sufficiently large n. Hence,

$$
\mathbf{Pr}\left[\bigcup_{i=1}^{n}\left\{\hat{x}_i \geqslant \lambda + \frac{\sqrt{c}}{10}\cdot \log n\right\}\right] \geqslant 1 - \left(1 - n^{-1/99}\right)^n \geqslant 1 - n^{-3}.
$$

Hence for $\widetilde{\mathcal{E}} := \left\{\max_{i\in[n]} \hat{x}_i \geqslant \lambda + \frac{\sqrt{c}}{10}\cdot \log n\right\}$, we have $\mathbf{Pr}\left[\,\neg\widetilde{\mathcal{E}}\,\right] \leqslant n^{-3}$. Note that $\widetilde{\mathcal{E}}$ is a monotone event under adding balls, and thus with $\mathcal{E} := \left\{\mathrm{Gap}(m) \geqslant \lambda + \frac{\sqrt{c}}{10}\cdot \log n\right\}$, we have by [138, Corollary 5.11])

$$
\mathbf{Pr}\left[\,\neg\mathcal{E}\,\right] \leqslant 2\cdot \mathbf{Pr}\left[\,\neg\widetilde{\mathcal{E}}\,\right] \leqslant 2\cdot n^{-3} \leqslant n^{-2}. \qquad \square
$$

B.5.2 Maximum load: The very lightly-loaded case

The following facts about the (very) lightly-loaded region of ONE-CHOICE, follow from the concentration inequalities stated before. The results by Raab and Steger [157] do not cover the region $m \ll n/\operatorname{polylog}(n)$, do not provide an estimate for the number of balls with height at least k and also the bounds are not derived for at least $1 - n^{-c}$ probability.

Lemma B.17. *Consider the* ONE-CHOICE *process with $m = \frac{n}{\log^c n}$ balls into n bins, where $c > 0$ is an arbitrary constant. Then, for any constant $\alpha > 0$ and for sufficiently large n,*

$$
\mathbf{Pr}\left[\,\mathrm{Gap}(m) > \frac{1}{c+1}\cdot \frac{\log n}{\log\log n}\,\right] \geqslant 1 - \frac{2}{n^{\alpha}}.
$$

Proof. We will bound the probability of event $\mathcal{E}$, that the maximum load is less than $M = \frac{1}{c+1}\cdot \log n/\log\log n$. The maximum load is a function that is increasing with the number of balls.

The technique of Poissonisation [3, Theorem 12] states that for ONE-CHOICE, the probability of a monotonically increasing event (in this case $\mathcal{E}$) is bounded by twice the probability that the event holds for independent Poisson r.vs. in place of the load r.vs.

We define $\mathcal{E}'$ to be the event that the maximum load is less than M, for n Poisson r.vs. Thus, $\mathbf{Pr}\left[\,\mathcal{E}\,\right] \leqslant 2\cdot \mathbf{Pr}\left[\,\mathcal{E}'\,\right]$. We bound $\mathbf{Pr}\left[\,\mathcal{E}'\,\right]$ by bounding the probability that no bin has load exactly M. We want

$$
\mathbf{Pr}\left[\,\mathcal{E}'\,\right] \leqslant \left(1 - \frac{e^{-\frac{1}{\log^c n}}\left(\frac{1}{\log^c n}\right)^M}{M!}\right)^n \leqslant \exp\left(-n\frac{e^{-\frac{1}{\log^c n}}\left(\frac{1}{\log^c n}\right)^M}{M!}\right) \leqslant \frac{1}{n^{\alpha}}.
$$

This is equivalent to showing that

$$-n\frac{e^{-\frac{1}{\log^c n}}\left(\frac{1}{\log^c n}\right)^M}{M!} < -\alpha \log n \iff \log n - \frac{1}{\log^c n} - Mc\log\log n - \log M! > \log(\alpha \log n)$$
$$\iff \log n - \log(\alpha\log n) - \frac{1}{\log^c n} > Mc\log\log n + \log M!.$$

Using Stirling's upper bound [79, Equation 9.1],

$$\begin{aligned}
Mc\log\log n + \log M! &< Mc\log\log n + M(\log M - 1) + \log M \\
&= M(c\log\log n - \log(c+1) + \log\log n - \log\log\log n - 1) + \log M \\
&= M(c+1)\log\log n - CM - M\log\log\log n + \log M \\
&= \log n - CM - M\log\log\log n + \log M \\
&< \log n - \log(\alpha\log n) - \frac{1}{\log^c n},
\end{aligned}$$

for sufficiently large n, since $\log(\alpha\log n) + \frac{1}{\log^c n} = o(M\log\log\log n - M)$ for any constant $\alpha > 0$. Hence, we get the desired lower bound. $\square$

We now extend Lemma B.17 to a case with fewer balls.

Lemma B.18. *(cf. Lemma B.17) Consider the* ONE-CHOICE *process with $m = \frac{n}{e^{u\log^c n}}$ (for constants $0 < c < 1$ and $u > 0$) balls into n bins. Then, for any constant $k > 0$ with $u \cdot k < 1$, for any constant $\alpha > 0$ and for sufficiently large n,*

$$\mathbf{Pr}\left[\,\mathrm{Gap}(m) \geqslant k\cdot(\log n)^{1-c}\,\right] \geqslant 1 - \frac{2}{n^\alpha}.$$

Proof. We define $\mathcal{E}$ and $\mathcal{E}'$ as in Lemma B.17. We bound $\mathbf{Pr}[\,\mathcal{E}'\,]$ by bounding the probability that no bin has load exactly $M = k\cdot(\log n)^{1-c}$. We claim

$$\mathbf{Pr}\left[\,\mathcal{E}'\,\right] \leqslant \left(1 - \frac{e^{-e^{-u\log^c n}}\left(e^{-u\log^c n}\right)^M}{M!}\right)^n \leqslant \exp\left(-n\frac{e^{-e^{-u\log^c n}}\left(e^{-u\log^c n}\right)^M}{M!}\right) < \frac{1}{n^\alpha},$$

which is equivalent to showing that

$$-n\frac{e^{-e^{-u\log^c n}}\left(e^{-u\log^c n}\right)^M}{M!} < -\alpha\log n \Leftrightarrow \log n - e^{-u\log^c n} - Mu\log^c n - \log M! > \log(\alpha\log n)$$
$$\Leftrightarrow \log n - \log(\alpha\log n) - e^{-u\log^c n} > Mu\log^c n + \log M!.$$

Using Stirling's upper bound [79, Equation 9.1],

$$\begin{aligned}
Mu\log^c n + \log M! &< Mu\log^c n + M(\log M - 1) + \log M \\
&= M(u\log^c n + \log k + (1-c)\log\log n - 1) + \log M \\
&= ku\cdot\log n + M(\log k + (1-c)\log\log n - 1) + \log M \\
&< \log n - \log(\alpha\log n) - e^{-u\log^c n},
\end{aligned}$$

for sufficiently large n, since $\log(\alpha\log n) + e^{-u\log^c n} + M(\log k + (1-c)\log\log n - 1) + \log M = o((1-u\cdot k)\log n)$ for any constant $\alpha > 0$ and $u\cdot k < 1$. Hence, we get the desired lower bound. $\square$

We use the following well known results for ONE-CHOICE.

Lemma B.19. *Consider the* ONE-CHOICE *process, for any $\alpha > 0$, any bin $i \in [n]$, and any step $m \geqslant 0$,*

$$\mathbf{E}[\, e^{\alpha x_i^m} \,] \leqslant e^{\frac{m}{n} \cdot (e^{\alpha} - 1)}.$$

Proof. We will proceed inductively to show that $\mathbf{E}[\, e^{\alpha x_i^m} \,] \leqslant e^{\frac{m}{n} \cdot (e^{\alpha} - 1)}$. The base case follows since $\mathbf{E}[\, e^{\alpha x_i^0} \,] = 1 \leqslant 1$. For $m \geqslant 1$, let Z_i^m indicate whether the m-th ball was allocated to bin $i \in [n]$ and assume that $\mathbf{E}[\, e^{\alpha x_i^{m-1}} \,] \leqslant e^{\frac{m-1}{n} \cdot (e^{\alpha} - 1)}$ holds, then

$$\begin{aligned}
\mathbf{E}\Big[\, e^{\alpha x_i^m} \,\Big] &= \mathbf{E}\Big[\, e^{\alpha (Z_i^m + x_i^{m-1})} \,\Big] = \mathbf{E}\Big[\, e^{\alpha Z_i^m} \,\Big] \cdot \mathbf{E}\Big[\, e^{\alpha x_i^{m-1}} \,\Big] \\
&= \Big(\frac{1}{n} \cdot e^{\alpha} + \Big(1 - \frac{1}{n}\Big) \cdot e^0\Big) \cdot \mathbf{E}\Big[\, e^{\alpha x_i^{m-1}} \,\Big] = \Big(1 + \frac{1}{n} \cdot (e^{\alpha} - 1)\Big) \cdot \mathbf{E}\Big[\, e^{\alpha x_i^{m-1}} \,\Big] \\
&\overset{(a)}{\leqslant} e^{\frac{1}{n} \cdot (e^{\alpha} - 1)} \cdot \mathbf{E}\Big[\, e^{\alpha x_i^{m-1}} \,\Big] \\
&\leqslant e^{\frac{1}{n} \cdot (e^{\alpha} - 1)} \cdot e^{\frac{m-1}{n} \cdot (e^{\alpha} - 1)} = e^{\frac{m}{n} \cdot (e^{\alpha} - 1)},
\end{aligned}$$

using in (a) that $1 + u \leqslant e^u$ (for any u). □

We now proceed to obtain an upper bound for the maximum load of ONE-CHOICE for any $m \leqslant 2n \log n$. As we will show in Lemma B.23, this bound is asymptotically tight.

Lemma B.20 (cf. [3, Lemma 14]). *Consider the* ONE-CHOICE *process for any $m \leqslant 2n \log n$. Then,*

$$\mathbf{Pr}\left[\, \max_{i \in [n]} x_i^m \leqslant 11 \cdot \frac{\log n}{\log(\frac{4n}{m} \cdot \log n)} \,\right] \geqslant 1 - n^{-6}.$$

Proof. Using Lemma B.19, for the given m and $\alpha = \log(\frac{4n}{m} \cdot \log n) > 0$, we have that for any bin $i \in [n]$,

$$\mathbf{E}[\, e^{\alpha x_i^m} \,] \leqslant e^{\frac{m}{n} \cdot e^{\log(\frac{4n}{m} \cdot \log n)}} = n^4.$$

Hence, by Markov's inequality,

$$\mathbf{Pr}\Big[\, e^{\alpha x_i^m} \leqslant n^{11} \,\Big] \geqslant 1 - n^{-7}.$$

When this event holds, we have

$$x_i^m \leqslant \frac{1}{\alpha} \cdot \log(n^{11}) \leqslant 11 \cdot \frac{\log n}{\log(\frac{4n}{m} \cdot \log n)}.$$

By taking the union bound over all bins $i \in [n]$, we get the claim. □

For $m = \mathcal{O}(n)$, this recovers the well-known $\mathcal{O}\big(\frac{\log n}{\log \log n}\big)$ bound.

Corollary B.21 (cf. [138, Chapter 5]). *Consider the* ONE-CHOICE *process for $m = 2n$. Then,*

$$\mathbf{Pr}\left[\, \max_{i \in [n]} x_i^m \leqslant 11 \cdot \frac{\log n}{\log \log n} \,\right] \geqslant 1 - n^{-6}.$$

For $m = \mathcal{O}(n^{1-\epsilon})$ for some constant $\epsilon \in (0,1)$, this shows that $\mathrm{Gap}(m) = \mathcal{O}(1)$.

Corollary B.22. *Consider the* ONE-CHOICE *process with $m = 2n^{1-\epsilon}$ for any constant $\epsilon \in (0,1)$. Then,*

$$\mathbf{Pr}\left[\, \max_{i \in [n]} x_i^m \leqslant \frac{11}{\epsilon} \,\right] \geqslant 1 - n^{-6}.$$

In the following lemma, we prove that the ONE-CHOICE bound obtained in Lemma B.20 is asymptotically tight.

Lemma B.23 (cf. Lemma 14 in [3]). *Consider the* ONE-CHOICE *process with $m \leqslant n \log n$. Then, there exists a constant $\kappa > 0$, such that*

$$\mathbf{Pr}\left[\max_{i \in [n]} x_i^m \geqslant \frac{1}{4} \cdot \frac{\log n}{\log(\frac{4n}{m} \cdot \log n)} \right] \geqslant 1 - n^{-1}.$$

Proof. We will bound the probability of event $\mathcal{E}$, that the maximum load is less than $M = \frac{1}{4} \cdot \frac{\log n}{\log(\frac{4n}{m} \cdot \log n)}$. Clearly, the probability of the event $\mathcal{E}$ is monotonically increasing in the number of balls (while keeping M fixed).

Following Lemma B.15, it suffices to bound the probability of the event $\widetilde{\mathcal{E}}$ which is that the maximum value of n independent Poisson random variables with parameter $\lambda = \frac{m}{n}$ is less than M. We want to show that

$$\mathbf{Pr}\left[\, \widetilde{\mathcal{E}} \,\right] \leqslant \left(1 - \frac{e^{-\frac{m}{n}} \left(\frac{m}{n}\right)^M}{M!} \right)^n \leqslant \exp\left(-n \cdot \frac{e^{-m/n} \left(\frac{m}{n}\right)^M}{M!} \right) < n^{-1}.$$

This is equivalent to showing that

$$\begin{aligned} -n \cdot \frac{e^{-m/n} \left(\frac{m}{n}\right)^M}{M!} < -\log n &\iff \log n - \frac{m}{n} + M \cdot \log\left(\frac{m}{n}\right) - \log M! > \log \log n \\ &\iff \log n + M \cdot \log\left(\frac{m}{n}\right) > \frac{m}{n} + \log M! + \log \log n. \end{aligned}$$

Using $\log M! \leqslant M \cdot (\log M - 1) + \log M$ (e.g., in [79, Equation 9.1]), we deduce that

$$\begin{aligned} &\frac{m}{n} + M \cdot \left(\log M - 1 - \log\left(\frac{m}{n}\right) \right) + \log M + \log \log n \\ &\quad < \frac{3}{4} \cdot \log n + \frac{\log n}{4 \cdot \log\left(\frac{4n \log n}{m}\right)} \cdot \left(\log \log n - \log\left(4 \cdot \log\left(\frac{4n \log n}{m} \right) \right) - \log\left(\frac{m}{n}\right) \right) \\ &\quad = \frac{3}{4} \cdot \log n + \frac{\log n}{4 \cdot \log\left(\frac{4n \log n}{m}\right)} \cdot \left(\log\left(\frac{4n \log n}{m} \right) - \log\left(16 \cdot \log\left(\frac{4n \log n}{m} \right) \right) \right) \\ &\quad \leqslant \frac{3}{4} \cdot \log n + \frac{\log n}{4 \cdot \log\left(\frac{4n \log n}{m}\right)} \cdot \log\left(\frac{4n \log n}{m} \right) \\ &\quad = \frac{3}{4} \cdot \log n + \frac{1}{4} \cdot \log n = \log n, \end{aligned}$$

for sufficiently large n and using that $\log\left(\frac{4n \log n}{m}\right) > 0$ as $m \leqslant n \log n$. Hence, we get the desired lower bound. $\square$

Combining with Lemma B.16, we also get the asymptotically tight bound on the gap

Lemma B.24 (cf. [3, Lemma 14]). *Consider the* ONE-CHOICE *process with $m \leqslant n \log n$. Then, there exists a constant $\kappa > 0$, such that*

$$\mathbf{Pr}\left[\mathrm{Gap}(m) \geqslant \kappa \cdot \frac{\log n}{\log\left(\frac{4n}{m} \cdot \log n\right)} \right] \geqslant 1 - n^{-1}.$$

Proof. For sufficiently small constant $C \in (0,1)$, for any $m \leqslant Cn\log n$ we have that

$$\frac{1}{4} \cdot \frac{\log n}{\log\left(\frac{4n}{m} \cdot \log n\right)} \geqslant 2 \cdot \frac{m}{n},$$

and hence the conclusion follows by Lemma B.23 for $\kappa = \frac{1}{8}$. For $m > Cn\log n$, the stated bound follows from Lemma B.16. □

B.5.3 Number of bins above a certain load

Lemma B.25. *Consider the* ONE-CHOICE *process for $m = n\log^2 n$. With probability at least $1 - o(n^{-2})$, there are at least $cn\log n$ balls with at least $\frac{m}{n} + \frac{a}{2}\log n$ height for $a = 0.4$ and $c = 0.25$.*

Proof. Consider the event $\mathcal{E}$ that the number of balls with load above $\frac{a}{2}\log n$ is at most $\frac{1}{5}\log n$. Since $\mathcal{E}$ is monotonically increasing in the number of balls, its probability is bounded by twice the probability of the event occurring for independent Poisson random variables [3, Theorem 12].

By Berry-Esseen inequality for Poisson random variables (Lemma B.7), for sufficiently large n and since $\epsilon = (\log n)^{-4}$,

$$\left|\mathbf{Pr}\left[\, Y \geqslant a \,\right] - \widetilde{\Phi}(a)\right| \leqslant \epsilon \Rightarrow \widetilde{\Phi}(a) - \epsilon \leqslant \mathbf{Pr}\left[\, X \geqslant \log^2 n + a\log n \,\right] \leqslant \widetilde{\Phi}(a) + \epsilon.$$

For $a = 0.4$, we get $\widetilde{\Phi}(a) \leqslant 0.35$. Let $X_i := \mathbf{1}(Y_i \geqslant \log^2 n + a\log n)$ and let $X := \sum_{i=1}^n X_i$, then X is a Binomial distribution with $p \leqslant 0.35$. Using the lower tail Chernoff bound for the Binomial distribution (Lemma B.4),

$$\mathbf{Pr}\left[\sum_{i=1}^n X_i \leqslant \frac{np}{e}\right] \leqslant e^{-\Omega(n)}.$$

For sufficiently large n, the RHS can be made $o(1/n^2)$, hence there are at least np/e bins with load at least $\frac{m}{n} + a\log n$ w. p. $1 - o(1/n^2)$. This means that w.h.p. at least $np/e \cdot a\log n = \frac{npa}{e}\log n \leqslant 0.26 \cdot n\log n$ balls have height $\frac{m}{n} + \frac{a}{2}\log n = \frac{m}{n} + 0.4\log n$. □

Lemma B.26. *(cf. Lemma B.25) Consider the* ONE-CHOICE *process with $m = Kn\sqrt{\log n} - \mathcal{O}(Kn\sqrt{\log n} \cdot e^{-\sqrt{\log n}})$ balls, for $K = 1/10$. Then, with probability at least $1 - n^{-4}$, there are at least $\frac{1}{20} \cdot e^{-0.21\sqrt{\log n}} n\sqrt{\log n}$ balls with height at least $\frac{3}{20} \cdot \sqrt{\log n}$.*

Proof. Let $C := \frac{1}{20}$ and note that $m = K(1 - o(1))n\sqrt{\log n}$. Using Poissonisation [3, Theorem 12], the probability that the statement of the lemma does not hold is upper bounded by twice the probability for the corresponding event with n independent Poisson random variables $X_1, X_2, \ldots, X_n$ with parameter $\lambda = \frac{m}{n} = K(1 - o(1))\sqrt{\log n}$. For a single Poisson random variable X, we lower bound the probability that $X \geqslant u$ for $u = (K + 2 \cdot C)\sqrt{\log n}$,

$$\begin{aligned}
\mathbf{Pr}\left[\, X \geqslant u \,\right] \geqslant \mathbf{Pr}\left[\, X = u \,\right] = \frac{e^{-\lambda}\lambda^u}{u!} &\geqslant \frac{e^{-\lambda}\lambda^u}{eu(u/e)^u} = e^{-\lambda + u - 1 - \log u}\left(\frac{\lambda}{u}\right)^u \\
&\geqslant \exp\left((K + 2 \cdot C)\sqrt{\log n} \cdot \log\left(\frac{K(1 - o(1))}{K + 2 \cdot C}\right)\right) \\
&\geqslant \exp\left(-0.8(K + 2 \cdot C)\sqrt{\log n}\right) > \exp\left(-0.2\sqrt{\log n}\right),
\end{aligned}$$

where the penultimate inequality used $\log\left(\frac{K(1-o(1))}{(K+2\cdot C)}\right) > -0.8$. Using Lemma B.4, this implies that w.p. $1 - n^{-4}$ at least $ne^{-0.20\sqrt{\log n} - 1} \geqslant ne^{-0.21\sqrt{\log n}}$ bins have load at least $(K + 2 \cdot C)\sqrt{\log n}$, so at least $e^{-0.21\sqrt{\log n}} \cdot Cn\sqrt{\log n}$ balls have height at least $(K + C)\sqrt{\log n}$. □

Lower bounds

In this chapter, we prove lower bounds for various of the processes and settings analysed in the previous chapters. More specifically, in Appendix C.1, we prove lower bounds for settings with outdated information. In Appendix C.2, we prove lower bounds for settings with adversarial and random noise. Next, in Appendix C.3, we prove lower bounds for Two-Thinning processes. Finally, in Appendix C.5, we prove lower bounds for Mean-Biased processes.

We employ the following set of techniques:

- Use of *coupling* and *majorisation* with a One-Choice process (e.g., Observation C.7, Theorem C.15 and Lemma C.27). To analyse the One-Choice process, we often employ *Poissonisation* (see Appendix B.5).
- Use of upper bounds on (exponential) *potential functions* to bound the number of bins with a normalised load in a given range (e.g., Theorem C.26).
- Use of *layered induction* (e.g., Theorem C.9 and Theorem C.26), in a spirit similar to [18], but with significant modifications.

C.1 Lower bounds for settings with outdated information

For the lower bounds we always assume that balls are unweighted (or equivalently, have unit weight).

We will now establish an interesting behaviour for a family of processes (including $(1+\beta)$ with constant $\beta \in (0,1)$ and Quantile(δ) with constant $\delta \in (0,1)$), that in the b-Batched setting for any $b \in [n, n\log n]$, the gap is $\Theta(\log n)$. We do this by proving a matching lower bound for Theorem 7.39.

We recall the following result which assumes no batching, i.e., balls are allocated sequentially using perfect knowledge about the bin loads.

Lemma C.1. *Consider any* Sequential(q^t) *process (in the unweighted setting) with allocation vector q^t satisfying $\min_{i\in[n]} q_i^t \geqslant \frac{C}{n}$ for some constant $C>0$ at every step $t \geqslant 0$. Then, for $m = \frac{C}{16} n \log n$,*

$$\mathbf{Pr}\left[\, \mathrm{Gap}(m) \geqslant \frac{C}{16}\cdot \log n \,\right] \geqslant 1 - 2n^{-2}.$$

Proof. The proof follows via a coupling with One-Choice. Since $\min_{i\in[n]} q_i^t \geqslant \frac{C}{n}$ at every step $t \geqslant 0$, by a Chernoff bound, with probability at least $1 - n^{-\omega(1)}$ we have that at least $\frac{C^2}{4\cdot 16}\cdot n\log n$ balls are allocated using One-Choice. Hence, by Lemma B.16, we get that

$$\mathbf{Pr}\left[\, \max_{i\in[n]} y_i^m \geqslant \left(\frac{C^2}{4\cdot 16} + \sqrt{\frac{C^2}{4\cdot 16}} \right)\cdot \log n \,\right] \geqslant 1 - 2n^{-2}.$$

Since $m = \frac{C}{16} n\log n$, we conclude that that

$$\mathbf{Pr}\left[\, \mathrm{Gap}(m) \geqslant \frac{C}{16}\cdot \log n \,\right] \geqslant 1 - 2n^{-2}. \qquad \square$$

By majorisation, it follows that any TIME-HOMOGENEOUS(p) process with a non-decreasing probability vector p has a worse gap in the b-BATCHED setting with $b > 1$ than for $b = 1$.

Proposition C.2. *Consider any* TIME-HOMOGENEOUS(p) *process with non-decreasing probability vector p satisfying $\min_{i\in[n]} p_i \geqslant \frac{C}{n}$ for some constant $C > 0$, in the (unweighted)* b-BATCHED *setting for any $b \geqslant 1$. Then, for $m = \frac{C}{16} \cdot n \log n$,*

$$\mathbf{Pr}\left[\, \mathrm{Gap}(m) \geqslant \frac{C}{16} \cdot \log n \,\right] \geqslant 1 - 2n^{-2}.$$

Note that this statement applies to the $(1+\beta)$-process for constant $\beta \in (0,1)$ and QUANTILE(δ) for constant $\delta \in (0,1)$, but it does not apply to TWO-CHOICE.

Proof. Consider the process $\mathcal{P} =$ TIME-HOMOGENEOUS(p). Recall that $\mathcal{Q} =$ b-BATCHED($\mathcal{P}$) is by definition a SEQUENTIAL(q^t) process with $q^t = q^t(\mathfrak{F}^t)$ being a permutation of p. Since p is non-decreasing, it follows that $\widetilde{q}^t \succeq p$.

By applying majorisation (Theorem 2.5) for processes $\mathcal{P}$ and $\mathcal{Q}$, for the first m steps, we get that there is a coupling such that the sorted load vector $\widetilde{y}^m_{\mathcal{Q}}$ majorizes $\widetilde{y}^m_{\mathcal{P}}$, and in particular,

$$(\widetilde{y}^m_{\mathcal{Q}})_1 \geqslant (\widetilde{y}^m_{\mathcal{P}})_1 \quad \Rightarrow \quad \mathrm{Gap}_{\mathcal{Q}}(m) \geqslant \mathrm{Gap}_{\mathcal{P}}(m).$$

Hence the statement of the lemma follows by Lemma C.1. $\square$

Now, we turn our attention to proving lower bounds for the same family of processes when the batch size is $b \geqslant n \log n$. This lower bound follows just by looking at the load of the bin i with probability $q_i^0 \geqslant \frac{C}{n}$ in the first batch.

Proposition C.3. *Consider any* TIME-HOMOGENEOUS(p) *process with p satisfying $\max_{i\in[n]} p_i \geqslant \frac{C}{n}$ for some $C > 1$, in the (unweighted)* b-BATCHED *setting with $b \geqslant n \log n$. Then, for $\gamma := \min(C-1, 0.5)$, any bin $j = \mathrm{argmax}_{i\in[n]} p_i$ satisfies*

$$\mathbf{Pr}\left[\, y_j^b \geqslant \frac{\gamma}{4} \cdot \frac{b}{n} \,\right] \geqslant 1 - n^{-\gamma^2/8}.$$

Proof. For convenience, let us define $\gamma := \min(C-1, 0.5)$, so $\gamma \in (0, 1/2)$. Note that during the first batch consisting of $b \geqslant n \log n$ balls, the load vector is never updated and all balls are allocated using the same probability vector p. Hence each ball will be allocated to some bin i with probability at least $\frac{C}{n}$, independently. Let $X := x_i^b = \sum_{j=1}^{b} X_j$, where the X_j's are independent Bernoulli random variables with $\mathbf{E}[\, X_j \,] \geqslant \frac{1+\gamma}{n}$. Hence $\mathbf{E}[\, X \,] \geqslant b \cdot \frac{1+\gamma}{n}$. Using the following Chernoff bound, which states that for any $\lambda > 0$,

$$\mathbf{Pr}[\, X \leqslant (1-\lambda) \cdot \mathbf{E}[\, X \,]\,] \leqslant \exp\bigl(-\lambda^2/2 \cdot \mathbf{E}[\, X \,]\bigr).$$

Picking $\lambda = \gamma/2$ implies

$$\mathbf{Pr}\left[\, x_i^b \leqslant (1-\gamma/2) \cdot (1+\gamma) \cdot \frac{b}{n} \,\right] \leqslant \exp\left(-\frac{\gamma^2}{8} \cdot \frac{b}{n}\right) \leqslant n^{-\gamma^2/8},$$

where the last inequality used our assumption that $b \geqslant n \log n$. If $x_i^b \geqslant (1-\gamma/2) \cdot (1+\gamma) \cdot \frac{b}{n}$, then this implies for the normalised load,

$$y_j^b = x_i^b - \frac{b}{n} \geqslant \frac{\gamma}{2} \cdot \frac{b}{n} - \frac{\gamma^2}{2} \cdot \frac{b}{n} \geqslant \frac{\gamma}{4} \cdot \frac{b}{n},$$

where the last inequality used $\gamma \leqslant 1/2$. $\square$

Unlike Proposition C.2, Proposition C.3 can be applied to TWO-CHOICE and any d-CHOICE process with $d = \mathcal{O}(1)$.

Herd phenomenon. For $b \geqslant n \log n$, this $\Omega\big(\frac{b}{n}\big)$ lower bound establishes a sharp contrast between TWO-CHOICE and the $(1+\beta)$ process, as the later for $\beta = \sqrt{\frac{n}{b} \cdot \log n}$ can achieve the asymptotically optimal $\mathcal{O}\big(\sqrt{\frac{b}{n} \cdot \log n}\big)$ gap as shown in Section 7.5. It also demonstrates that increasing d in d-CHOICE does not necessarily help, a phenomenon which had been observed by Mitzenmacher [134] but not rigorously proven (to the best of our knowledge).

Observation C.4. *Consider any d-SAMPLE process with $d = \mathcal{O}(1)$ which uses random tie breaking, in the b-BATCHED setting with any $b \geqslant n \log n$, and assume all balls have unit weight. Then, we have that*

$$\mathbf{Pr}\left[\mathrm{Gap}(b) \geqslant \frac{1}{10} \cdot \sqrt{\frac{b}{n} \cdot \log n} \right] \geqslant 1 - n^{-2}.$$

Proof. In the first batch, any such process behaves exactly like ONE-CHOICE. Hence the result follows immediately from a known lower bound for ONE-CHOICE for b balls into n bins by Lemma B.16. $\square$

The same $\Omega(\sqrt{\frac{b}{n} \cdot \log n})$ lower bound was shown in [114, Theorem 5.2] to also hold for any process and at any batch (not just in the first one).

Small batch sizes. For small batch sizes $n \cdot e^{-\log^c n} \leqslant b \leqslant n \log n$ with $c = \mathcal{O}(1)$, we have that TWO-CHOICE is optimal among all TWO-SAMPLE (or in fact any d-SAMPLE process with $d = \mathcal{O}(1)$). This follows from the upper bound in Corollary 7.31 and the lower bound that follows from ONE-CHOICE in the first batch (e.g., Lemma B.17).

Observation C.5. *Consider any d-SAMPLE process with $d = \mathcal{O}(1)$ and random tie breaking, in the b-BATCHED setting with any $b \in [n \cdot e^{\log^c n}, n \log n]$ with any constant $c > 0$, and assume all balls have unit weight. Then, we have that*

$$\mathbf{Pr}\left[\mathrm{Gap}(b) \geqslant \kappa \cdot \frac{\log n}{\log\big(\frac{4n}{b} \cdot \log n\big)} \right] \geqslant 1 - n^{-1}.$$

Proof. During the allocation of the first batch consisting of b balls, the load information of all bins is never updated, i.e., their load "estimate" equals zero. Hence in these first n steps, as ties are broken randomly, b-BATCHED behaves exactly like ONE-CHOICE. Hence, by Lemma B.24 there exists a constant $\kappa > 0$,

$$\mathbf{Pr}\left[\mathrm{Gap}(b) \geqslant \kappa \cdot \frac{\log n}{\log\big(\frac{4n}{b} \cdot \log n\big)} \right] \geqslant 1 - n^{-1}. \qquad \square$$

Random tie-breaking. Let us remark that in the proof of Proposition C.3, we assumed that the allocation process uses the same probability vector and bin labelling in all steps of the same batch. In particular, this analysis does not apply to TWO-CHOICE with random tie-breaking. However, TWO-CHOICE with random tie-breaking will allocate all balls in the first batch following ONE-CHOICE. Exploiting this, we can then prove that by the end of the batch, there is a unique bin which attains the minimum load if $b = \Omega(n \log n)$, which means for the second batch we can apply Proposition C.3, and conclude that a lower bound of $\Omega\big(\frac{b}{n}\big)$ holds with constant probability > 0.

We will make use of the following property of n independent Poisson random variables, which is proven in Appendix B.4:

Lemma B.14 (Restated, page 197). *Consider any $n \geqslant 2$ and $\lambda \geqslant 16 \cdot \log n$. Let $X^1, \ldots, X^n$ be independent Poisson random variables with $X^i \sim \mathrm{Pois}(\lambda)$, and denote by for $Y_{(n)}, Y_{(n-1)}$ the smallest and second smallest of the X^i's. Then there exist constants $\kappa_1, \kappa_2 > 0$ such that,*

$$\mathbf{Pr}\left[\, Y_{(n-1)} - Y_{(n)} \geqslant \kappa_1 \cdot \sqrt{\lambda / \log n} \,\right] \geqslant \kappa_2.$$

Using this we can now derive the lower bound for allocation processes with random tie-breaking.

Lemma C.6. *Consider a* TIME-HOMOGENEOUS(p) *process with probability vector p and random tie-breaking, such that $p_n \geqslant \frac{C}{n}$ for some constant $C \in (1, 1.5]$ in the (unweighted) b-*BATCHED *setting with $b \geqslant \frac{384}{(C-1)^2} n \log n$. Then, there exists a constant $\kappa := \kappa(C) > 0$, such that*

$$\mathbf{Pr}\left[\, \mathrm{Gap}(2b) \geqslant \frac{C-1}{8} \cdot \frac{b}{n} \,\right] \geqslant \kappa.$$

Proof. Initially, all bins have load 0, so the first b balls will be allocated using ONE-CHOICE. In order to use the Poisson Approximation Method [138, Theorem 5.6], let $\widetilde{X}_1, \widetilde{X}_2, \ldots, \widetilde{X}_n$ be n independent Poisson distributed random variables with rate $\lambda = (b - 4 \cdot \sqrt{b})/n$. By Lemma B.8, the sum $S_n := \sum_{i=1}^n \widetilde{X}_i$ is in the range $[b - 8\sqrt{b}, b]$, with probability at least $1 - o(1)$. By Lemma B.14, we have that with at least constant $\kappa_2 > 0$ probability, the difference between the smallest and second smallest bin is at least $\kappa_1 \cdot \sqrt{\lambda / \log n}$, for some constant $\kappa_1 > 0$.

Consider now the allocation of the remaining $b - S_n \leqslant 8\sqrt{b}$ balls. The average load of a bin through these balls is $8\sqrt{b}/n$. Using Markov's inequality, the smallest bin does not receive more than $16\sqrt{b}/n$ additional balls with probability at least $1/2$.

Since $\kappa_1 \cdot \sqrt{\lambda / \log n} \geqslant \kappa_1 \cdot \sqrt{0.5 \cdot b/n \cdot 1/\log n} \geqslant 16\sqrt{b}/n$ we can conclude that there is still a unique minimally loaded bin after the allocation of all b balls. Further, by using a Chernoff bound for ONE-CHOICE, it follows that

$$\mathbf{Pr}\left[\, y_n^b \leqslant b/n - \sqrt{6 \cdot b/n \log n} \,\right] \geqslant 1 - n^{-2}.$$

Taking the union bound, we conclude that at the end of the first batch, the following holds:

$$\mathbf{Pr}\left[\, y_n^b \in \left[-\sqrt{6 \cdot b/n \log n}, y_{n-1}^b - 1 \right] \right] \geqslant \kappa_1 \cdot \frac{1}{2} - o(1) - n^{-2}. \tag{C.1}$$

Conditioning on $y_n^b \leqslant y_{n-1}^b - 1$, we have $\widetilde{p}_n(x^b) \geqslant p_n \geqslant \frac{C}{n}$. For simplicity, let us fix label n to be the index of the bin with smallest load at time b. Applying Proposition C.3 to the allocations made in the second batch to bin n, we conclude that there is a constant $\gamma > 0$ such that

$$\mathbf{Pr}\left[\, x_n^{2b} - x_n^b \geqslant \left(1 + \frac{\gamma}{4}\right) \cdot \frac{b}{n} \;\middle|\; y_n^b \in \left[-\sqrt{6 \cdot b/n \log n}, y_{n-1}^b - 1 \right] \right] \geqslant 1 - n^{-\gamma^2/8}. \tag{C.2}$$

Both events in Eq. (C.1) and Eq. (C.2) hold with probability at least $\kappa_1 \cdot \frac{1}{3}$, and in this case,

$$\begin{aligned}
x_n^{2b} &= x_n^b + x_n^{2b} - x_n^b \\
&\geqslant \frac{b}{n} - \sqrt{6 \cdot b/n \log n} + \left(1 + \frac{C-1}{4}\right) \cdot \frac{b}{n} \\
&\geqslant \frac{2b}{n} - \sqrt{6 \cdot b/n \log n} + \frac{C-1}{4} \cdot \frac{b}{n} \\
&\overset{(a)}{\geqslant} \frac{2b}{n} + \frac{C-1}{8} \cdot \frac{b}{n},
\end{aligned}$$

where we have used in (a) that if $b \geqslant \frac{384}{(C-1)^2} n \log n$ then,

$$\frac{C-1}{4} \cdot \frac{b}{n} \geqslant 2 \cdot \sqrt{6 \cdot b/n \log n} \quad \Leftrightarrow \quad b \geqslant \frac{384}{(C-1)^2} \cdot n \log n.$$

Hence $\text{Gap}(2b) \geqslant \frac{C-1}{8} \cdot \frac{b}{n}$. □

C.2 Lower bounds in noisy settings

In this section, we will state and prove lower bounds on the gap of Two-Choice in various noisy settings. Recall that the g-Bounded and g-Myopic-Comp processes are specific instances of the g-Adv-Comp setting. Our main result is for g-Myopic-Comp, where we will prove a lower bound of $\Omega(g + \frac{g}{\log g} \cdot \log \log n)$, which matches the upper bounds in Sections 5.3 and 7.4.1 for all $g \geqslant 0$ (Corollary C.10). Table C.1 gives a summary of the results.

Process	Range	Lower Bound	Reference
Any g-Adv-Comp	$0 \leqslant g$	$\log_2 \log n + \Omega(1)$	Obs C.7
g-Myopic-Comp	$2 \leqslant g$	$\Omega(g)$	Prop C.8
g-Myopic-Comp	$10 \leqslant g \leqslant \frac{1}{8} \cdot \frac{\log n}{\log \log n}$	$\Omega\big(\frac{g}{\log g} \cdot \log \log n\big)$	Thm C.9
σ-Noisy-Load	$2 \cdot (\log n)^{-1/3} \leqslant \sigma$	$\Omega(\min\{1, \sigma\} \cdot (\log n)^{1/3})$	Prop C.11
σ-Noisy-Load	$32 \leqslant \sigma$	$\Omega(\min\{\sigma^{4/5}, \sigma^{2/5} \cdot \sqrt{\log n}\})$	Prop C.11

Table C.1: Overview of the lower bounds for different noise settings. All of these hold for a particular value of m with high probability.

We start with a simple lower bound which follows immediately by majorisation with the Two-Choice process without noise. This lower bound holds regardless of which strategy the adversary uses.

Observation C.7. *There exists a constant $\kappa > 0$ such that for any $g \geqslant 0$ and any instance of the g-Adv-Comp setting,*

$$\mathbf{Pr}\big[\, \text{Gap}(n) \geqslant \log_2 \log n - \kappa \,\big] \geqslant 1 - n^{-1}.$$

Proof. For the Two-Choice process without noise, it was shown in [18, Theorem 3.3] that there exists a constant $\kappa > 0$ such that $\mathbf{Pr}\big[\, \text{Gap}(n) \geqslant \log_2 \log n - \kappa \,\big] \geqslant 1 - n^{-1}$.

At any step $t \geqslant 0$, the probability allocation vector p of Two-Choice without noise is majorised the probability allocation vector q^t of any instance of Two-Choice in the g-Adv-Comp setting, as q^t is formed by p and possibly reallocating some probability mass from light to heavy bins. Hence, the lower bound follows by majorisation (see Theorem 2.5). □

We proceed by analysing the g-Myopic-Comp process by coupling its allocations with that of a One-Choice process.

Proposition C.8. *Consider the* g-MYOPIC-COMP *process. Then,* (i) *for any* $g \in [2, 6\log n]$ *and for* $m = \frac{1}{2} \cdot ng$*, it holds that*

$$\mathbf{Pr}\left[\operatorname{Gap}(m) \geqslant \frac{1}{35} \cdot g \right] \geqslant 1 - n^{-2}.$$

Further, (ii) *for any* $g \geqslant 6\log n$ *and for* $m = ng^2/(32\log n)$*,*

$$\mathbf{Pr}\left[\operatorname{Gap}(m) \geqslant \frac{1}{60} \cdot g \right] \geqslant 1 - 2n^{-2}.$$

Recall that for $g = \Omega(\operatorname{polylog}(n))$, our upper bound on the gap is

$$\mathcal{O}\left(g + \frac{g}{\log g} \cdot \log\log n \right) = \mathcal{O}(g).$$

This means that the lower bound in Proposition C.8 is matching for those g. For smaller values of g, a stronger lower bound will be presented in Theorem C.9.

Proof. First statement. Consider g-MYOPIC-COMP with $m = \frac{1}{2} \cdot ng$ balls and define the stopping time $\tau := \inf\{t \geqslant 0 \colon \max_{i \in [n]} x_i^t \geqslant g\}$. Note that $\tau \leqslant m$ implies there is a bin j with $x_j^{\tau} \geqslant g$, and hence

$$\operatorname{Gap}(m) \geqslant x_j^m - \frac{m}{n} \geqslant x_j^{\tau} - \frac{1}{2} g \geqslant g - \frac{1}{2} g > \frac{1}{35} g.$$

Let us now assume $\tau > m$. In that case, all bins have an absolute load in $[0, g]$ and are therefore indistinguishable. Hence during steps $1, 2, \ldots, m$, the g-MYOPIC-COMP process behaves exactly like ONE-CHOICE.

By Lemma B.16 (for $c := \frac{1}{2} \cdot \frac{g}{\log n} \geqslant \frac{1}{\log n}$ since $g \geqslant 2$) it holds for ONE-CHOICE that,

$$\mathbf{Pr}\left[\operatorname{Gap}(m) \geqslant \frac{1}{10\sqrt{2}} \cdot \sqrt{g \cdot \log n} \right] \geqslant 1 - n^{-2},$$

which, as $g \cdot \log n \geqslant \frac{1}{6} g^2$, implies that

$$\mathbf{Pr}\left[\operatorname{Gap}(m) \geqslant \frac{1}{35} \cdot g \right] \geqslant 1 - n^{-2}.$$

Second statement. We will first prove for the g-MYOPIC-COMP process that w. p. $1 - o(n^{-1})$ for any step $1 \leqslant t \leqslant m$ and any bin $i \in [n]$, it holds that $|y_i^t| \leqslant \frac{g}{2}$. To this end, fix any bin $i \in [n]$ and define for any $1 \leqslant t \leqslant m$, $Z^t = Z^t(i) := \sum_{j=1}^{t} (Y_j - \frac{1}{n})$, where the Y_j's are independent Bernoulli variables with parameter $1/n$ each. Clearly, Z^t forms a martingale, and $\mathbf{E}[\, Z^t \,] = 0$ for all $1 \leqslant t \leqslant m$. Further, let $\tau := \inf\{t \geqslant 1 \colon |Z^t| > \frac{g}{2}\}$. Then $Z^{t \wedge \tau}$ is also martingale. Further, we have

$$\mathbf{Var}\left[\, Z^{t+1} \mid Z^t \,\right] = \frac{1}{n} \cdot \left(1 - \frac{1}{n}\right) \leqslant \frac{1}{n} =: \sigma^2.$$

Also $|Z^{t+1} - Z^t| \leqslant 1 =: M$. Hence by a martingale inequality Lemma B.9,

$$\mathbf{Pr}\left[\, \left|Z^{m \wedge \tau}\right| \geqslant \lambda \,\right] \leqslant 2 \cdot \exp\left(- \frac{\lambda^2}{2(\sum_{i=1}^{m} \sigma^2 + M\lambda/3)} \right),$$

with $\lambda = \frac{g}{2}$,

$$\mathbf{Pr}\left[\, |Z^{m \wedge \tau}| \geqslant \frac{g}{2} \,\right] \leqslant 2 \cdot \exp\left(-\frac{g^2}{8(m \cdot \frac{1}{n} + \frac{g}{6})} \right) = 2 \cdot \exp\left(-\frac{g^2}{8 \cdot (\frac{g^2}{32 \log n} + \frac{g}{6})} \right) \leqslant 2n^{-3},$$

where the last inequality holds since $8 \cdot \left(\frac{g^2}{32 \log n} + \frac{g}{6} \right) \leqslant 8 \cdot \left(\frac{g^2}{32 \log n} + \frac{g^2}{32 \log n} \right) = \frac{g^2}{2 \log n}$, using that $g \geqslant 6 \log n$.

If the event $|Z^{m \wedge \tau}| \leqslant \frac{g}{2}$ (or equivalently $\tau > m$) occurs, then this implies that bin i deviates from the average load by at most $\frac{g}{2}$ in all steps $1, 2, \ldots, m$. By the union bound, this holds with probability $1 - 2n^{-2}$ for all n bins. Consequently, the g-Myopic-Comp process behaves exactly like One-Choice until time m with probability $1 - 2n^{-2}$.

By Lemma B.16 (for $c := \frac{g^2}{32 \log^2 n}$) it holds for One-Choice that,

$$\mathbf{Pr}\left[\, \mathrm{Gap}(m) \geqslant \frac{g}{10 \cdot \sqrt{32}} \,\right] \geqslant 1 - n^{-2}.$$

The claim follows by taking the union bound and using that $\frac{g}{10 \cdot \sqrt{32}} \geqslant \frac{g}{60}$. □

Theorem C.9. *Consider the g-Myopic-Comp process for any $g \in [10, \frac{1}{8} \cdot \frac{\log n}{\log \log n}]$. Then, there exists $\ell := \ell(g, n)$ (defined in Eq. (C.3)), such that for $m = n \cdot \ell$, it holds that*

$$\mathbf{Pr}\left[\, \mathrm{Gap}(m) \geqslant \frac{1}{8} \cdot \frac{g}{\log g} \cdot \log \log n \,\right] \geqslant 1 - n^{-\omega(1)}.$$

The proof of this lower bound is similar to the method used by Azar, Broder, Karlin and Upfal [18] to prove the $\Omega(\log \log n)$ lower bound for Two-Choice without noise, in the sense that it follows a layered induction approach, but some additional care is needed. For example, the induction step size in the load is g and not 1, and the outcome of a load comparison depends on the load difference.

Proof. In the proof we will divide the allocation of $m = n \cdot \ell$ balls into ℓ consecutive phases, each of which lasts for n steps, where

$$\ell := \left\lfloor \frac{\log(\frac{1}{8} \log n / \log g)}{\log g} \right\rfloor. \tag{C.3}$$

First, we verify that $\ell \geqslant 1$,

$$\ell = \left\lfloor \frac{\log(\frac{1}{8} \log n / \log g)}{\log g} \right\rfloor \overset{(a)}{\geqslant} \left\lfloor \frac{\log(\frac{1}{8} \log n / \log g)}{\log(\frac{1}{8} \log n / \log n)} \right\rfloor \overset{(b)}{\geqslant} \left\lfloor \frac{\log(\frac{1}{8} \log n / \log n)}{\log(\frac{1}{8} \log n / \log n)} \right\rfloor = 1,$$

where (a) used that $g \leqslant \frac{1}{8} \log n / \log \log n$ and (b) that $g \leqslant \log n$.

Next we define for any $k = 1, 2, \ldots, \ell$ the following event:

$$\mathcal{E}_k := \left\{ \left| \left\{ i \in [n] \colon x_i^{k \cdot n} \geqslant k \cdot g \right\} \right| \geqslant n \cdot g^{-\sum_{j=1}^{k} g^j} \right\}.$$

The main goal of this proof is to show that $\mathcal{E}_\ell$ occurs with high probability. Assuming for the moment that $\mathcal{E}_\ell$ indeed occurs, let us verify that the lower bound on the gap follows. First, note that $\mathcal{E}_\ell$ implies the existence of a bin $i \in [n]$ with $x_i^m \geqslant \ell \cdot g$ and thus $\mathrm{Gap}(m) \geqslant x_i^m - \frac{m}{n} \geqslant \ell \cdot g - \ell = \ell \cdot (g-1)$, since by the choice of $\ell = \left\lfloor \frac{\log(\frac{1}{8}\log n/\log g)}{\log g} \right\rfloor$,

$$g^{-\sum_{j=1}^{k} g^j} \overset{(a)}{\geqslant} g^{-4g^\ell} \geqslant g^{-\frac{1}{2}\log n/\log g} = n^{-1/2}, \tag{C.4}$$

where in (a) we used that for any $g \geqslant 2$,

$$\sum_{j=1}^{k} g^j \leqslant 2\sum_{j=0}^{\ell} g^j = 2 \cdot \frac{g^{\ell+1}-1}{g-1} \leqslant 2 \cdot \frac{g^{\ell+1}}{\frac{1}{2}g} = 4g^\ell. \tag{C.5}$$

Thus $n \cdot g^{-\sum_{j=1}^{k} g^j} \geqslant 1$. Secondly, we verify that $\ell \cdot (g-1) = \Omega\big(\frac{g}{\log g} \cdot \log\log n\big)$,

$$\begin{aligned}
\ell \cdot (g-1) &\overset{(a)}{\geqslant} \frac{1}{4} \cdot \frac{\log(\frac{1}{8}\log n/\log g)}{\log g} \cdot g \\
&= \frac{1}{4} \cdot \frac{\log(1/8) + \log\log n - \log\log g}{\log g} \cdot g \\
&\overset{(b)}{\geqslant} \frac{1}{8} \cdot \frac{\log\log n}{\log g} \cdot g,
\end{aligned} \tag{C.6}$$

where (a) holds since $\lfloor u \rfloor \geqslant u/2$ for $u \geqslant 1$ (as $\ell \geqslant 1$) and $g-1 \geqslant g/2$, since $g \geqslant 2$, and (b) holds since for sufficiently large n, $(1/4) \cdot \log\log n \geqslant -\log(1/8)$ and $(1/4) \cdot \log\log n \geqslant \log\log g$ as $g \leqslant \log n$.

In order to establish that $\mathcal{E}_\ell$ occurs w. p. $1 - o(n^{-1})$, we will proceed by induction and prove that for any $k \geqslant 1$, with $\epsilon_k := n^{-\omega(1)}$ and $\mathcal{E}_0 := \Omega$,

$$\mathbf{Pr}\left[\, \mathcal{E}_k \mid \mathcal{E}_{k-1} \,\right] \geqslant 1 - \epsilon_k.$$

Induction Base ($k = 1$). Here we consider the allocation of the first n balls into n bins. We are interested in the number of bins which reach load level g during that phase. Note that as long as the loads of both sampled bins are smaller than g, the allocation follows that of One-Choice; if one of the bins has a load which is already larger than g, then the load difference may force the process to place a ball in the lighter of the two bins. In the following, we will (pessimistically) assume that the allocation of *all* n balls follows One-Choice.

Instead of the original One-Choice process which produces a load vector $(x_i^n)_{i\in[n]}$, we consider the Poisson Approximation (Lemma B.15) and analyse the load vector $(\widetilde{x}_i^n)_{i\in[n]}$, where the $\widetilde{x}_i^n$, $i \in [n]$ are independent Poisson random variables with mean $n/n = 1$. Clearly, for any $i \in [n]$,

$$\mathbf{Pr}\left[\, \widetilde{x}_i^n \geqslant g \,\right] \geqslant \mathbf{Pr}\left[\, \widetilde{x}_i^n = g \,\right] = e^{-1} \cdot \frac{1^g}{g!} \geqslant 2 \cdot g^{-g},$$

using in the last inequality Stirling's approximation (e.g. [138, Lemma 5.8]) and $g \geqslant 5$,

$$g! \leqslant e \cdot g \cdot \left(\frac{g}{e}\right)^g = \frac{e^{-1}}{2} \cdot g^g \cdot \left(\frac{2e^2 g}{e^g}\right) \leqslant \frac{e^{-1}}{2} \cdot g^g.$$

Let $Y := \left|\left\{ i \in [n] \colon \widetilde{x}_i^n \geqslant g \right\}\right|$. Then $\mathbf{E}\left[\, Y \,\right] \geqslant 2n \cdot g^{-g}$. By a standard Chernoff Bound,

$$\mathbf{Pr}\left[\, Y \leqslant n \cdot g^{-g} \,\right] \leqslant \mathbf{Pr}\left[\, Y \leqslant \frac{1}{2} \cdot \mathbf{E}\left[\, Y \,\right] \right] \leqslant \exp\left(-\frac{1}{8} \cdot \mathbf{E}\left[\, Y \,\right]\right) \leqslant n^{-\omega(1)},$$

where we have used $g^{-g} \geqslant g^{-4g^{\ell}} \geqslant n^{-1/2}$, due to Eq. (C.4). Hence by the Poisson Approximation (Lemma B.15), $\mathbf{Pr}[\, \mathcal{E}_1 \,] \geqslant 1 - 2 \cdot n^{-\omega(1)} = 1 - n^{-\omega(1)}$.

Induction Step ($k-1 \longrightarrow k$, $k \geqslant 2$). For the induction step, we analyse phase $k = 2, \ldots, \ell$ and we will lower bound $\mathbf{Pr}[\, \mathcal{E}_k \mid \mathcal{E}_{k-1} \,]$. Assuming $\mathcal{E}_{k-1}$ occurs, there are at least $n \cdot g^{-\sum_{j=1}^{k-1} g^j}$ bins whose load is at least $(k-1) \cdot g$ at the beginning of phase k. Let us call such a set of bins $\mathcal{B}_{k-1}$, which can be assumed to satisfy with equality:

$$|\mathcal{B}_{k-1}| = n \cdot g^{-\sum_{j=1}^{k-1} g^j}. \tag{C.7}$$

Next note that whenever we sample two bins i_1, i_2 from $\mathcal{B}_{k-1}$, we allocate the ball to a random bin among $\{i_1, i_2\}$ if both bins have not reached load level $k \cdot g$. Therefore, in order to lower bound the number of bins in $\mathcal{B}_{k-1}$ which reach load level $k \cdot g$ by the end of phase k, we may pessimistically assume that if two bins in $\mathcal{B}_{k-1}$ are sampled, the ball will be *always* placed in a randomly sampled bin. Note that the probability that a ball will be allocated into the set $\mathcal{B}_{k-1}$ is lower bounded by $\left(\frac{|\mathcal{B}_{k-1}|}{n}\right)^2$. Let Z denote the number of balls allocated to $\mathcal{B}_{k-1}$ in phase k. Then $\mathbf{E}[\, Z \,] \geqslant n \cdot \left(\frac{|\mathcal{B}_{k-1}|}{n}\right)^2$. Using a Chernoff Bound for $\widetilde{m} := \frac{2}{3} \cdot n \cdot \left(\frac{|\mathcal{B}_{k-1}|}{n}\right)^2$,

$$\mathbf{Pr}[\, Z \leqslant \widetilde{m} \,] \leqslant \mathbf{Pr}\left[\, Z \leqslant \frac{2}{3} \cdot \mathbf{E}[\, Z \,] \,\right] \leqslant \exp\left(-\frac{1}{18} \cdot \mathbf{E}[\, Z \,]\right) \leqslant n^{-\omega(1)},$$

where we used that

$$\mathbf{E}[\, Z \,] \geqslant n \cdot g^{-2\sum_{j=1}^{k-1} g^j} \overset{(a)}{\geqslant} n \cdot g^{-2g^k} \overset{(b)}{\geqslant} n \cdot n^{-1/2},$$

using in (a) that for any integer $g \geqslant 2$, $\sum_{j=1}^{k-1} g^j \leqslant \frac{g^k - 1}{g-1} \leqslant g^k$, and in (b) that $k \leqslant \ell$ and the property of ℓ in Eq. (C.4).

Conditioning on this event, in phase k we have a One-Choice process with $\widetilde{m}$ balls into $\widetilde{n} := |\mathcal{B}_{k-1}|$ bins, which w.l.o.g. will be labelled $1, 2, \ldots, \widetilde{n}$. Again, we apply the Poisson approximation and define $(\widetilde{x}^{\widetilde{m}})_{i \in [\widetilde{n}]}$ as $\widetilde{n}$ independent Poisson random variables with mean λ given by

$$\lambda := \frac{\widetilde{m}}{\widetilde{n}} = \frac{\frac{2}{3} \cdot n \cdot \left(\frac{|\mathcal{B}_{k-1}|}{n}\right)^2}{|\mathcal{B}_{k-1}|} = \frac{2}{3} \cdot \frac{|\mathcal{B}_{k-1}|}{n} = \frac{2}{3} \cdot g^{-\sum_{j=1}^{k-1} g^j},$$

using Eq. (C.7). With that, it follows for any bin $i \in \mathcal{B}_{k-1}$,

$$\mathbf{Pr}[\, \widetilde{x}_i \geqslant g \,] \geqslant \mathbf{Pr}[\, \widetilde{x}_i = g \,] = e^{-\lambda} \cdot \frac{\lambda^g}{g!} \overset{(a)}{\geqslant} e^{-1} \cdot \frac{\left(\frac{2}{3} \cdot g^{-\sum_{j=1}^{k-1} g^j}\right)^g}{g!} \overset{(b)}{\geqslant} 2 \cdot g^{-\sum_{j=1}^{k} g^j},$$

using in (a) that $\lambda \leqslant 1$ and in (b) Stirling's approximation (e.g. [138, Lemma 5.8]) and that $g \geqslant 10$,

$$g! \leqslant e \cdot g \cdot \left(\frac{g}{e}\right)^g = \frac{e^{-1}}{2} \cdot \left(\frac{2}{3}\right)^g \cdot g^g \cdot \left(2e^2 g \cdot \left(\frac{3}{2e}\right)^g\right) \leqslant \frac{e^{-1}}{2} \cdot \left(\frac{2}{3}\right)^g \cdot g^g.$$

Let $Y := |\{i \in [\widetilde{n}] \colon \widetilde{x}_i \geqslant g\}|$. Then $\mathbf{E}[\, Y \,] \geqslant 2\widetilde{n} \cdot g^{-\sum_{j=1}^{k} g^j}$. Thus by a Chernoff Bound,

$$\mathbf{Pr}\left[\, Y \leqslant \widetilde{n} \cdot g^{-\sum_{j=1}^{k} g^j} \,\right] \leqslant \mathbf{Pr}\left[\, Y \leqslant \frac{1}{2} \cdot \mathbf{E}[\, Y \,] \,\right] \leqslant \exp\left(-\frac{1}{8} \cdot \mathbf{E}[\, Y \,]\right) \leqslant n^{-\omega(1)},$$

where we used that

$$\mathbf{E}[\, Y \,] \geqslant 2\widetilde{n} \cdot g^{-\sum_{j=1}^{k} g^j} = 2n \cdot g^{-\sum_{j=1}^{k-1} g^j} \cdot g^{-\sum_{j=1}^{k} g^j} \geqslant 2n \cdot g^{-2\sum_{j=1}^{k} g^j} \overset{\text{(C.5)}}{\geqslant} 2n \cdot g^{-4g^k} \overset{(a)}{\geqslant} 2n^{1/2},$$

where in (a) we used $k \leqslant \ell$ and Eq. (C.4). Thus by the union bound and Poisson Approximation, $\mathbf{Pr}[\, \mathcal{E}_k \mid \mathcal{E}_{k-1} \,] \geqslant 1 - n^{-\omega(1)} - 2 \cdot n^{-\omega(1)} = 1 - n^{-\omega(1)}$, which completes the induction.

Finally, by a simple union bound,

$$\mathbf{Pr}[\, \mathcal{E}_\ell \,] \geqslant 1 - \sum_{k=1}^{\ell} \mathbf{Pr}[\, \neg \mathcal{E}_k \mid \mathcal{E}_{k-1} \,] \geqslant 1 - \ell \cdot n^{-\omega(1)} \geqslant 1 - n^{-\omega(1)}.$$

As verified in Eq. (C.6), $\mathcal{E}_\ell$ implies $\mathrm{Gap}(m) \geqslant \frac{1}{8} \cdot \frac{g}{\log g} \cdot \log \log n$, and therefore the proof is complete. $\square$

Combining Observation C.7, Proposition C.8 and Theorem C.9, we get:

Corollary C.10. *Consider the* g-MYOPIC-COMP *process for any* $g \geqslant 1$*. Then, there exists an* $m := m(g) \geqslant 0$*, such that*

$$\mathbf{Pr}\left[\, \mathrm{Gap}(m) = \Omega\left(g + \frac{g}{\log g} \cdot \log \log n \right) \right] \geqslant 1 - n^{-1}.$$

We proceed with two lower bounds for the σ-NOISY-LOAD process.

Proposition C.11. *Consider the* σ-NOISY-LOAD *process with* $\rho(\delta) := 1 - \frac{1}{2} \cdot \exp\big(-(\delta/\sigma)^2\big)$ *for some (not necessarily constant)* $\sigma > 0$*. Then,* (i) *for any* $\sigma \geqslant 2 \cdot (\log n)^{-1/3}$*,*

$$\mathbf{Pr}\left[\, \mathrm{Gap}(n) \geqslant \min\left\{ \frac{1}{8} \cdot (\log n)^{1/3}, \frac{1}{2}\sigma \cdot (\log n)^{1/3} \right\} \right] \geqslant 1 - 2n^{-1}.$$

Further, (ii) *for any* $\sigma \geqslant 32$*, it holds for* $m = \frac{1}{2}\sigma^{4/5} \cdot n$*,*

$$\mathbf{Pr}\left[\, \mathrm{Gap}(m) \geqslant \min\left\{ \frac{1}{2}\sigma^{4/5}, \frac{1}{30}\sigma^{2/5} \cdot \sqrt{\log n} \right\} \right] \geqslant 1 - 2n^{-2}.$$

Proof. First statement. Let $\tau := \inf\{t \geqslant 1 \colon \max_{i \in [n]} x_i^t \geqslant \sigma \cdot (\log n)^{1/3}\}$. If $\tau \leqslant n$, then

$$\mathrm{Gap}(n) \geqslant x_i^n - \frac{n}{n} \geqslant x_i^\tau - 1 \geqslant \sigma \cdot (\log n)^{1/3} - 1 \geqslant \frac{1}{2}\sigma \cdot (\log n)^{1/3}.$$

Consider now the case $\tau > n$. For any step $t \leqslant \tau$, we will perform a "sub-sampling" of the correct comparison in (possibly) two stages as follows. Let $i_1 = i_1^t$ and $i_2 = i_2^t$ be the two sampled bins, and δ be their load difference. Let $Z_1 \sim \mathrm{Ber}(1 - \exp(-(\delta/\sigma)^2))$ and $Z_2 \sim \mathrm{Ber}(1/2)$ be two independent random variables, which can be thought as the outcome of two biased coin flips that will be used to determine whether the load comparison is correct. If $Z_1 = 1$, then the comparison is correct (regardless of what Z_2 is). However, if $Z_1 = 0$, then the comparison is correct if and only if $Z_2 = 1$. Overall, the probability of a correct comparison (if $\delta > 0$) is equal to

$$\begin{aligned} \mathbf{Pr}[\, Z_1 = 1 \,] + \mathbf{Pr}[\, Z_1 = 0 \,] \cdot \mathbf{Pr}[\, Z_2 = 1 \mid Z_1 = 0 \,] &= 1 - \exp\big(-(\delta/\sigma)^2\big) + \exp\big(-(\delta/\sigma)^2\big) \cdot \frac{1}{2} \\ &= 1 - \frac{1}{2} \exp\big(-(\delta/\sigma)^2\big). \end{aligned}$$

Further, conditional on $Z_1 = 0$, the ball will be placed in a random bin among $\{i_1, i_2\}$. Hence as long as $t \leqslant \tau$, we can couple the allocation of each ball by g-MYOPIC-COMP to an allocation by ONE-CHOICE with probability at least $\exp(-(\delta/\sigma)^2) \geqslant \exp(-(\log n)^{2/3})$. Using a Chernoff bound, with probability $1 - n^{-\omega(1)}$, we we can couple the allocation of at least $n/2 \cdot \exp(-(\log n)^{2/3})$ balls out of the first n balls with that of ONE-CHOICE. Consequently, using Lemma B.23 the maximum load (and gap) is at least $\frac{1}{8}(\log n)^{1/3}$ with probability at least $1 - n^{-1}$. Combining the two cases we get the claim.

Second Statement. Consider any $\sigma \geqslant 32$ and define the stopping time $\tau := \inf\{t \geqslant 1 \colon \max_{i\in[n]} x_i^t \geqslant \sigma^{4/5}\}$. Let $m := \frac{1}{2}\sigma^{4/5} \cdot n$. If $\tau \leqslant m$, then there is a bin $i \in [n]$ with $x_i^\tau \geqslant \sigma^{4/5}$, and

$$\mathrm{Gap}(m) \geqslant x_i^m - \frac{m}{n} \geqslant x_i^\tau - \frac{1}{2}\sigma^{4/5} \geqslant \frac{1}{2}\sigma^{4/5}.$$

Otherwise, in each step until m, the load difference between any two sampled bins is at most $\sigma^{4/5}$, and therefore each ball is coupled with the allocation of a ONE-CHOICE process with probability at least $\exp\big(-(\sigma^{4/5}/\sigma)^2\big) \geqslant 1 - \frac{1}{\sigma^{2/5}}$.

Let X be the number of ONE-CHOICE allocations in the first m steps. Using the following standard Chernoff bound,

$$\mathbf{Pr}[\, X \geqslant (1-\delta) \cdot \mathbf{E}[\, X\,]\,] \geqslant 1 - e^{-\frac{1}{2}\delta^2 \cdot \mathbf{E}[\, X\,]},$$

with $\delta = \frac{1}{\sigma^{2/5}} \leqslant 1$ (since $\sigma \geqslant 1$) and $\mathbf{E}[\, X\,] \geqslant \big(1 - \frac{1}{\sigma^{2/5}}\big) \cdot m$, we get that

$$\mathbf{Pr}\left[\, X \geqslant \frac{1}{2}\left(1 - 2 \cdot \frac{1}{\sigma^{2/5}}\right) \cdot \sigma^{4/5} \cdot n \,\right] \geqslant 1 - e^{-\frac{1}{2}\delta^2 \cdot \mathbf{E}[\, X\,]} \geqslant 1 - n^{-\omega(1)}.$$

Therefore, using Lemma B.16 (for $c := \frac{1}{2} \cdot \big(1 - 2 \cdot \frac{1}{\sigma^{2/5}}\big) \cdot \sigma^{4/5} \cdot \frac{1}{\log n} \geqslant \frac{1}{\log n}$), we get that

$$\begin{aligned}
\mathrm{Gap}(m) &\geqslant \frac{1}{10} \cdot \sqrt{\frac{1}{2}\left(1 - \frac{2}{\sigma^{2/5}}\right) \cdot \sigma^{4/5} \cdot \log n} + \frac{1}{2}\left(1 - \frac{1}{\sigma^{2/5}}\right) \cdot \sigma^{4/5} - \frac{m}{n} \\
&= \frac{1}{10} \cdot \sqrt{\frac{1}{2}\left(1 - \frac{2}{\sigma^{2/5}}\right) \cdot \sigma^{4/5} \cdot \log n} + \frac{1}{2}\left(1 - \frac{1}{\sigma^{2/5}}\right) \cdot \sigma^{4/5} - \frac{1}{2}\sigma^{4/5} \\
&= \frac{1}{10} \cdot \sqrt{\frac{1}{2}\left(1 - \frac{2}{\sigma^{2/5}}\right) \cdot \sigma^{4/5} \cdot \log n} - \frac{1}{2}\sigma^{2/5} \\
&\overset{(a)}{\geqslant} \frac{1}{20}\sigma^{2/5} \cdot \sqrt{\log n} - \frac{1}{2} \cdot \sigma^{2/5} \\
&\geqslant \frac{1}{30}\sigma^{2/5} \cdot \sqrt{\log n},
\end{aligned}$$

using in (a) that $\sigma \geqslant 32$. So, combining the two cases we get the claim. □

C.3 Lower bounds for TWO-THINNING processes

Now, we turn our attention to proving lower bounds for TWO-THINNING. In the lightly-loaded case (i.e., $m = n$), [75] proved an upper bound of $(2 + o(1)) \cdot (\sqrt{2 \log n / \log\log n})$ on the maximum load for a uniform THRESHOLD (f)-process with $f = \sqrt{2 \log n / \log\log n}$ ([77] extended this to $d > 2$). They also proved that this strategy is asymptotically optimal. In [75, Problem 1.3], the authors suggest that the $\mathcal{O}(\sqrt{\log n / \log\log n})$ bound on the gap extends to the heavily-loaded case. Here we will disprove this, establishing a slightly larger lower bound of $\Omega(\sqrt{\log n})$ (Theorem C.20). We also derive additional lower bounds (Theorem C.15 and Corollary C.16) that demonstrate that any QUANTILE or THRESHOLD process will "frequently" attain a gap which is even as large as $\Omega(\log n / \log\log n)$. These demonstrate that the QUANTILE(δ^*) process is asymptotically optimal.

Let us describe the intuition behind this bound in case of uniform quantiles, neglecting some technicalities. Consider QUANTILE(δ) and the equivalent TWO-THINNING instance where a ball is placed in the first bin if its load is among the $(1 - \delta) \cdot n$ lightest bins, and otherwise it is placed in a new (second) bin chosen uniformly at random (Lemma 2.19). We have two cases:

Case 1: We choose most times a "large" δ. Then we allocate approximately $m \cdot \delta$ balls to their second bin choice which is uniform over all n bins. This will lead to a behaviour close to ONE-CHOICE (Lemma C.12).

Case 2: We choose most times a "small" δ. Then we allocate approximately $m \cdot (1-\delta)$ balls with the first bin choice, which is a ONE-CHOICE process over the $n \cdot (1-\delta)$ lightest bins. As we establish in Lemma C.13, for small δ there are simply "too many" light bins that will reach a high load level, so the process is again close to ONE-CHOICE.

Lemma C.12 (Restated, page 218). *For any* QUANTILE(δ^t) *(or* THRESHOLD(f^t)*) process,*

$$\mathbf{Pr}\left[\max_{t \in [1, n \log^2 n]} \mathrm{Gap}(t) \geqslant \frac{1}{8} \cdot \frac{\log n}{\log \log n} \right] \geqslant 1 - o(n^{-2}).$$

In fact, as shown in Corollary C.16, this lower bound holds for a significant proportion of time-steps. We also show a lower bound for fixed m, which is derived in a similar way as Theorem C.15, but with a different parameterisation of "large" and "small" quantiles:

Lemma C.12 (Restated, page 221). *For any adaptive* QUANTILE(δ^t) *(or* THRESHOLD(f^t)*) process, with* $m = \frac{1}{10} \cdot n\sqrt{\log n}$ *balls, it holds that*

$$\mathbf{Pr}\left[\mathrm{Gap}(m) \geqslant \frac{1}{20} \sqrt{\log n} \right] \geqslant 1 - o(n^{-2}).$$

C.3.1 Preliminaries

Let us first formalise the intuition of the lower bound. Recall that we will analyse the adaptive case, which means that the quantiles at each step t may depend on the full history of the process $\mathfrak{F}^t$. We also remind the reader that any adaptive THRESHOLD(f^t) process can be simulated by QUANTILE(δ^t) (Lemma D.1), which is why we will do the analysis below for QUANTILE(δ^t) only.

The next lemma proves that if within n consecutive allocations a large quantile is used too often, then QUANTILE(δ) restricted to the heavy loaded bins generates a high maximum load, similar to ONE-CHOICE.

Lemma C.12. *Consider any* QUANTILE(δ^t) *process during the time-interval* $[t, t+n)$*. If* QUANTILE(δ^t) *allocates at least* $n/(\log n)^2$ *balls with a quantile larger than* $(\log n)^{-2}$ *in* $[t, t+n)$*, then*

$$\mathbf{Pr}\left[\mathrm{Gap}(t+n) \geqslant \frac{1}{8} \frac{\log n}{\log \log n} \right] \geqslant 1 - o(n^{-4}).$$

Proof. Assume there are at least $n/(\log n)^2$ allocations with quantile larger than $(\log n)^{-2}$. Then, using Lemma B.4, w. p. at least $1 - o(n^{-4})$, at least $\frac{1}{e} \frac{n}{\log^2 n} \cdot \frac{1}{\log^2 n} \geqslant \frac{n}{\log^5 n}$ balls are thrown using ONE-CHOICE.

Consider now the load configuration before the *batch*, i.e. the next n balls are allocated. If $\mathrm{Gap}(t) \geqslant \log n$, then $\mathrm{Gap}(t+n) \geqslant \frac{1}{8} \log n / \log \log n$, as a load can decrease by at most 1 in n steps. So we can assume $\mathrm{Gap}(t) < \log n$. Let B be the set of bins whose load is at least the average load at time t, then $|B| \geqslant n/\log n$. Using Lemma B.4, w. p. at least $1 - o(n^{-4})$ the batch will allocate at least $n/(\log n)^6$ balls to the bins of B. Hence, by Lemma B.17, at least one bin in B will increase its load by an additive factor of $\frac{1}{7} \log n / \log \log n$ w. p. at least $1 - o(n^{-4})$. Since the average load only increases by one during the batch, there will be a gap of $\frac{1}{8} \log n / \log \log n$ w.h.p., and our claim is established. $\square$

The next lemma implies that if for most allocations the largest quantile is too small, then the allocations on the lightest bins follows that of ONE-CHOICE, and we end up with a high maximum load.

Lemma C.13. *Consider any* QUANTILE(δ^t) *process with* $m = n\log^2 n$ *balls that allocates at most* n *balls with a quantile larger than* $(\log n)^{-2}$. *Then,*

$$\mathbf{Pr}\left[\,\mathrm{Gap}(m) \geqslant 0.2\log n\,\right] \geqslant 1 - o(n^{-2}).$$

The proof of this lemma is similar to Lemma C.12, but a bit more complex. We define a coupling between the QUANTILE(δ^t) process and the ONE-CHOICE process. We couple the allocation of balls whose first sample is among the $(1-\delta^t)\cdot n$-lightest bins with a ONE-CHOICE process. The balls whose first sample is among the $\delta^t \cdot n$-heaviest bins are allocated differently, and cause our process to diverge from an original ONE-CHOICE process. However, we prove that the number of different allocations is too small to change the order of the gap.

Proof. We will use the following coupling between the allocations of QUANTILE(δ^t) and ONE-CHOICE. At each step $t \in [1, n\log^2 n]$, we first sample a bin index $j \in [n]$ uniformly at random. In the ONE-CHOICE process, we place the ball in the j-th most loaded bin. In the QUANTILE process:

1. If $j > \delta^t \cdot n$, we place the ball in the j-th most loaded bin (of QUANTILE), and we say that the processes agree.
2. If $j \leqslant \delta^t \cdot n$, we sample another bin index $\widetilde{j} \in [n]$ uniformly at random and place the ball in the $\widetilde{j}$-th most loaded bin (of QUANTILE).

Let y^s and z^s be the sorted load vectors of ONE-CHOICE and the QUANTILE process respectively at step $s \geqslant 0$. Further, let $L(s) := d_{\ell_1}(y^s, z^s)$ be the ℓ_1-distance between these vectors. Note that $L(0) = 0$. If in a step both processes place a ball in the j-th most loaded bin, using a simple coupling argument (see Lemma C.14 below for details) it follows that

$$L(t+1) \leqslant L(t).$$

Otherwise, if in a step the processes place a ball in a different bin, since only two positions in the load vectors can increase by one, then

$$L(t+1) \leqslant L(t) + 2.$$

Hence by induction over s, if k is the number of steps for which the processes disagree, then

$$L(n\log^2 n) \leqslant 2\cdot k.$$

We will next show an upper bound on k, which in turn implies an upper bound on $L(n\log^2 n)$. First, for each of the at most n steps $t \in [1, n\log^2 n]$ for which $\delta^t \geqslant (\log n)^{-2}$, we (pessimistically) assume that the two processes always disagree. Secondly, for the at most $n\log^2 n$ steps $t \in [1, n\log^2 n]$ with $\delta^t \leqslant (\log n)^{-2}$, using a Chernoff bound (Lemma B.4), we have w. p. $1 - o(n^{-2})$ in at most $(n\log^2 n)\cdot(\log n)^{-2}\cdot e = ne$ of these steps s, the case that $j \leqslant \delta^s \cdot n$, i.e., the two processes disagree. Now if this event occurs,

$$k \leqslant n\cdot 1 + n\cdot e \leqslant 2n\cdot e \qquad \Rightarrow \qquad L(n\log^2 n) \leqslant 4n\cdot e.$$

By Lemma B.25, there are constants $a = 0.4, c = 0.25$ such that with probability $1 - o(n^{-2})$, the ONE-CHOICE load vector $y^{n\log^2 n}$ has at least $0.25n\log n$ balls with height at least $0.2\log n$. However, any load vector which has no balls at height $0.2\log n$ must have a ℓ_1-distance of at least $0.25n\log n$ to $y^{n\log^2 n}$, and thus we conclude by the union bound that $\mathrm{Gap}(n\log^2 n) \geqslant \frac{a}{2}\log n$ holds with probability $1 - 2o(n^{-2})$. □

Lemma C.14. *Let $\widetilde{y}_1$ and $\widetilde{y}_2$ be two non-increasing sorted load vectors. Consider the sorted vectors $\widetilde{y}_1 + \mathbf{e}_i$ and $\widetilde{y}_2 + \mathbf{e}_i$ after incrementing the value at index i. Then, $d_{\ell_1}(\widetilde{y}_1, \widetilde{y}_2) \geqslant d_{\ell_1}(\widetilde{y}_1 + \mathbf{e}_i, \widetilde{y}_2 + \mathbf{e}_i)$.*

Proof. If the items being updated end up both in the same indices (after sorting), then their ℓ_1 distance remains unchanged.

Let $u := \widetilde{y}_{1i}$ and $v := \widetilde{y}_{2i}$ for the updated index i in the (old) sorted load vector. To obtain the new sorted load vector, we have to search in both $\widetilde{y}_1$ and $\widetilde{y}_2$ from right to left for the leftmost entry being equal to u and being equal to v, respectively, and then increment these values. Then, there are the following three cases to consider (in bold is the value to be updated):

Case 1 $u < v$: Let $v < w_1 \leqslant \ldots \leqslant w_k$, where w_k is the matching value for $u+1$ in z, then $w_k > v \Rightarrow w_k \geqslant u+2$

$$
\begin{array}{c|ccccccc}
\widetilde{y}_1 & \ldots & u & \ldots & u & u & \ldots & \mathbf{u} & \ldots \\
\hline
\widetilde{y}_2 & \ldots & w_k & \ldots & w_1 & v & \ldots & \mathbf{v} & \ldots
\end{array}
\rightarrow
\begin{array}{c|cccccccc}
\widetilde{y}_1 + \mathbf{e}_i & \ldots & u+1 & \ldots & u & u & \ldots & u & \ldots \\
\hline
\widetilde{y}_2 + \mathbf{e}_i & \ldots & w_k & \ldots & w_1 & v+1 & \ldots & v & \ldots \\
 & & \underbrace{}_{-1} & & & \underbrace{}_{+1} & & &
\end{array}
$$

Case 2 $u < v$: Let $u < w_1 \leqslant \ldots \leqslant w_k$, where w_k is the matching value for $v+1$ in y

$$
\begin{array}{c|ccccccc}
\widetilde{y}_1 & \ldots & w_k & \ldots & w_1 & u & \ldots & \mathbf{u} & \ldots \\
\hline
\widetilde{y}_2 & \ldots & v & \ldots & v & v & \ldots & \mathbf{v} & \ldots
\end{array}
\rightarrow
\begin{array}{c|cccccccc}
\widetilde{y}_1 + \mathbf{e}_i & \ldots & w_k & \ldots & w_1 & u+1 & \ldots & u & \ldots \\
\hline
\widetilde{y}_2 + \mathbf{e}_i & \ldots & v+1 & \ldots & v & v & \ldots & v & \ldots \\
 & & \underbrace{}_{\leqslant 1} & & & \underbrace{}_{-1} & & &
\end{array}
$$

Case 3 $u = v$: Let $u < w_1 \leqslant \ldots \leqslant w_k$, where w_k is the matching value for $u+1$ in z

$$
\begin{array}{c|ccccccc}
\widetilde{y}_1 & \ldots & w_k & \ldots & w_1 & u & \ldots & \mathbf{u} & \ldots \\
\hline
\widetilde{y}_2 & \ldots & u & \ldots & u & u & \ldots & \mathbf{u} & \ldots
\end{array}
\rightarrow
\begin{array}{c|cccccccc}
\widetilde{y}_1 + \mathbf{e}_i & \ldots & w_k & \ldots & w_1 & u+1 & \ldots & u & \ldots \\
\hline
\widetilde{y}_2 + \mathbf{e}_i & \ldots & u+1 & \ldots & u & u & \ldots & u & \ldots \\
 & & \underbrace{}_{-1} & & & \underbrace{}_{+1} & & &
\end{array}
$$

$\square$

C.3.2 Lower bound for a range of values

With Lemmas C.12 and C.13 proven in the previous subsection, we can now derive a lower bound for any QUANTILE(δ^t) (or THRESHOLD(f^t)) process, establishing Theorem C.15. After the proof, we also state two simple consequences that follow immediately from this result.

Theorem C.15. *For any* QUANTILE(δ^t) *(or* THRESHOLD(f^t)*) process,*

$$
\mathbf{Pr}\left[\max_{t \in [1, n\log^2 n]} \mathrm{Gap}(t) \geqslant \frac{1}{8} \cdot \frac{\log n}{\log\log n} \right] \geqslant 1 - o(n^{-2}).
$$

Proof. Since any adaptive THRESHOLD(f^t) can be simulated by an adaptive QUANTILE(δ^t) process (see Lemma D.1), it suffices to prove the claim for adaptive QUANTILE(δ^t) processes. We will allow the adversary to run two processes, and then choose one that achieves a gap of $< \frac{1}{8} \log n / \log\log n$ (if such exists):

- **Process $\mathcal{P}_1$.** The adversary has to allocate $m = n\log^2 n$ balls into n bins. The adversary wins if for all steps $t \in [m]$, $\mathrm{Gap}(t) < \frac{1}{8}\log n/\log\log n$, and, Condition $\mathcal{K}_1$, at least n out of the m quantiles are larger than $(\log n)^{-2}$.
- **Process $\mathcal{P}_2$.** The adversary has to allocate $m = n\log^2 n$ balls into n bins. The adversary wins if $\mathrm{Gap}(m) < \frac{1}{8}\log n/\log\log n$ and, Condition $\mathcal{K}_2$, at least $m-n = n\log^2 n - n$ out of the m quantiles are at most $(\log n)^{-2}$.

Note that the conditions $\mathcal{K}_1$ and $\mathcal{K}_2$ form a disjoint partition. We will prove that the adversary cannot win any of the two games with probability greater than n^{-2}. Now recall the original process, the one we would like to analyse:

- **Process $\mathcal{P}_3$ (adaptive** QUANTILE(δ^t)**).** The adversary has to allocate $m = n\log^2 n$ balls into bins at each step. The adversary wins if $\mathrm{Gap}(t) < \frac{1}{8}\log n/\log\log n$ for all $t \in [m]$.

We will show below that $\mathbf{Pr}[\,\text{adversary wins } \mathcal{P}_1\,] = o(n^{-2})$ and $\mathbf{Pr}[\,\text{adversary wins } \mathcal{P}_2\,] = o(n^{-2})$, and these bounds hold for the best possible strategies an adversary can use in each game, respectively. Assuming that these bounds hold and by noticing that exactly one of $\mathcal{K}_1$ and $\mathcal{K}_2$ must hold for $\mathcal{P}_3$,

$$\mathbf{Pr}[\,\mathcal{P}_3 \text{ wins}\,] = \mathbf{Pr}[\,\mathcal{P}_3 \text{ wins}, \mathcal{K}_1\,] + \mathbf{Pr}[\,\mathcal{P}_3 \text{ wins}, \mathcal{K}_2\,] \leqslant \mathbf{Pr}[\,\mathcal{P}_1 \text{ wins}\,] + \mathbf{Pr}[\,\mathcal{P}_2 \text{ wins}\,] \leqslant o(n^{-2}).$$

Analysis of Process 1: Let $\mathcal{E}_t$ be the event that (i) QUANTILE allocates at least $n/(\log n)^2$ balls with a quantile larger than $(\log n)^{-2}$ in the interval $[t, t+n)$, and (ii) $\mathrm{Gap}(t+n) < \frac{1}{8}\log n/\log\log n$. Note that this is the negation of Lemma C.12, so by union bound over $1 \leqslant t \leqslant m-n$,

$$\mathbf{Pr}\left[\,\bigcup_{t=1}^{m-n} \mathcal{E}_t\,\right] \leqslant n\log^2 n \cdot o(n^{-4}) = o(n^{-2}).$$

Note that if none of the $\mathcal{E}_t$ for $1 \leqslant t \leqslant m-n$ occur, then the adversary allocates at most $n/(\log n)^2 \cdot (\log n)^2 \geqslant n$ out of the m balls with a quantile at least $(\log n)^{-2}$. Therefore,

$$\mathbf{Pr}[\,\text{adversary wins } \mathcal{P}_1\,] \leqslant o(n^{-2}).$$

Analysis of Process 2: The analysis of $\mathcal{P}_2$ follows directly by Lemma C.13. $\square$

Let us also observe a slightly stronger statement which follows directly from Theorem C.15:

Corollary C.16. *For any* QUANTILE(δ^t) *process, it holds that,*

$$\mathbf{Pr}\left[\,\bigcup_{t\in[1,n\log^2 n]} \left\{ \min_{s\in\left[t,t+\frac{1}{16}n\frac{\log n}{\log\log n}\right)} \mathrm{Gap}(s) \geqslant \frac{1}{16}\cdot\frac{\log n}{\log\log n} \right\}\,\right] \geqslant 1 - n^{-2}.$$

Proof of Corollary C.16. If there is a step t for which $\mathrm{Gap}(t) \geqslant \frac{1}{8}\cdot\log n/\log\log n$, then for any s with $t \leqslant s \leqslant t + \frac{1}{16}\cdot\log n/\log\log n$, $\mathrm{Gap}(s) \geqslant \mathrm{Gap}(t) - (s-t)/n \geqslant \frac{1}{16}\cdot\log n/\log\log n$. Hence the statement follows from Theorem C.15. $\square$

In other words, the corollary states that for $\Omega(n\log n/\log\log n)$ (consecutive) steps in $[1, \Theta(n\log^2 n)]$, the gap is $\Omega(\log n/\log\log n)$. This is in contrast to the behaviour of the process QUANTILE(δ_1, δ_2), for which our result in Section 7.1.2 implies that with high probability the gap is *always* below $\mathcal{O}(\sqrt{\log n})$ during any time-interval of the same length.

For uniform QUANTILE(δ), we are always running either process $\mathcal{P}_1$ or $\mathcal{P}_2$, so the following strengthened version of Theorem C.15 holds:

Corollary C.17. *For any uniform* QUANTILE(δ) *process for $m = n\log^2 n$ balls,*

$$\mathbf{Pr}\left[\,\mathrm{Gap}(m) \geqslant \frac{1}{8}\cdot\frac{\log n}{\log\log n}\,\right] \geqslant 1 - o(n^{-2}).$$

Proof. Since δ is fixed, in the proof of Theorem C.15, we are always running either process $\mathcal{P}_1$ or $\mathcal{P}_2$. For process $\mathcal{P}_1$, $\mathcal{E}_{m-n}$ holds w.p. $1-o(n^{-4})$, so there is an $\Omega(\log n/\log\log n)$ gap at m. For process $\mathcal{P}_2$, there is an $\Omega(\log n/\log\log n)$ gap at m w.p. $1-o(n^{-2})$. Hence, in both cases the gap at step m is $\Omega(\log n/\log\log n)$ w.p. $1-o(n^{-2})$. $\square$

C.3.3 Lower bound for fixed $m = \Theta(n\sqrt{\log n})$

We now prove a version of Theorem C.15 that establishes a lower bound of $\Omega(\sqrt{\log n})$ on the gap for a *fixed* value m. It follows the same proof as Theorem C.15 except that the parameters are different: (i) $m = \Theta(n\sqrt{\log n})$ and (ii) Condition $\mathcal{K}_1$ is defined as having at least $m\cdot e^{-\sqrt{\log n}}$ out of the m quantiles being at least $e^{-\sqrt{\log n}}$. Lemma C.18 is the modified Lemma C.12 and Lemma C.19 is the modified Lemma C.13.

Lemma C.18. *Consider any adaptive* QUANTILE(δ^t) *process during the time-interval $[t, t+n)$. If the process allocates at least $n/e^{\sqrt{\log n}}$ balls with a quantile larger than $e^{-\sqrt{\log n}}$ in $[t, t+n)$, then*

$$\mathbf{Pr}\left[\,\mathrm{Gap}(t+n) \geqslant \frac{1}{5}\sqrt{\log n}\,\right] \geqslant 1 - o(n^{-4}).$$

Proof. Assume there are at least $n/e^{\sqrt{\log n}}$ allocations with quantile larger than $e^{-\sqrt{\log n}}$. Then, using Lemma B.4, w. p. at least $1-o(n^{-4})$, at least $\frac{1}{e}\frac{n}{e^{\sqrt{\log n}}}\cdot\frac{1}{e^{\sqrt{\log n}}} \geqslant \frac{n}{e^{3\sqrt{\log n}}}$ balls are thrown using ONE-CHOICE.

Consider now the load configuration before the batch is allocated. If $\mathrm{Gap}(t) \geqslant \frac{1}{4}\sqrt{\log n}$, then $\mathrm{Gap}(t+n) \geqslant \frac{1}{5}\sqrt{\log n}$, as a load can decrease by at most 1 in n steps. So we can assume $\mathrm{Gap}(t) < \frac{1}{4}\sqrt{\log n}$. Let B be the set of bins whose load is at least the average load at time t, then $|B| \geqslant n/(\frac{1}{4}\sqrt{\log n})$. Using Lemma B.4, w. p. at least $1-o(n^{-4})$ the batch will allocate at least $n/(e\cdot e^{3\sqrt{\log n}}\cdot(\frac{1}{4}\sqrt{\log n})) \geqslant n/e^{4\sqrt{\log n}}$ balls to the bins of B. Hence, using Lemma B.18 with $c = 1/2$, $u = 4$ and $k = \frac{2}{9}$ at least one bin in B will increase its load by an additive factor of $\frac{2}{9}\sqrt{\log n}$ w. p. at least $1-o(n^{-4})$. Since the average load only increases by one during the batch, we have created a gap of $\frac{2}{9}\sqrt{\log n} - 1 > \frac{1}{5}\sqrt{\log n}$, and our claim is established. $\square$

Lemma C.19. *Consider any adaptive* QUANTILE(δ^t) *process with $m = \frac{1}{10}n\sqrt{\log n}$ balls that allocates at most n balls with a quantile larger than $e^{-\sqrt{\log n}}$, then*

$$\mathbf{Pr}\left[\,\mathrm{Gap}(m) \geqslant \frac{1}{20}\sqrt{\log n}\,\right] \geqslant 1 - o(n^{-2}).$$

Proof. Let $C = 1/20$ and $K = 1/10$. We will use the same coupling as in the proof of Lemma C.13. We now obtain an upper bound on the number of steps k where the two processes disagree, which in turn implies an upper bound on $L(m)$. First, for each of the at most n steps $t \in [1, m]$ for which $\delta^t \geqslant e^{-\sqrt{\log n}}$, we (pessimistically) assume that the two processes always disagree. Secondly, for the at most m steps $t \in [1, m]$ with $\delta^t \leqslant e^{-\sqrt{\log n}}$, using a Chernoff bound (Lemma B.4), we have w. p. $1-o(n^{-2})$ in at most $e\cdot(Kn\sqrt{\log n})\cdot e^{-\sqrt{\log n}}$ of these steps s the case that $j \leqslant \delta^s\cdot n$, i.e., the two processes disagree. Now if this event occurs,

$$k \leqslant n\cdot 1 + n\cdot e \leqslant 2n\cdot e \qquad \Rightarrow \qquad L(Kn\sqrt{\log n}) \leqslant 4\cdot e\cdot(Kn\sqrt{\log n})\cdot e^{-\sqrt{\log n}}.$$

By Lemma B.26, with probability $1 - o(n^{-2})$, the ONE-CHOICE load vector $y^{Kn\sqrt{\log n}}$ has at least $e^{-0.21\sqrt{\log n}} \cdot Cn\sqrt{\log n}$ balls with at least $(K+C)\cdot\sqrt{\log n}$ height. However, any load vector which has no balls at height $(K+C)\cdot\sqrt{\log n}$ must have a ℓ_1-distance of at least $e^{-0.21\sqrt{\log n}} \cdot Cn\sqrt{\log n} \cdot K \cdot \sqrt{\log n} > L(m)$ to $y^{Kn\sqrt{\log n}}$, and thus we conclude by the union bound that $\mathrm{Gap}(Kn\sqrt{\log n}) \geqslant C\sqrt{\log n}$ holds with probability $1 - 2o(n^{-2})$. □

Theorem C.20. *For any adaptive* QUANTILE(δ^t) *(or* THRESHOLD(f^t)*) process, with* $m = \frac{1}{10} \cdot n\sqrt{\log n}$ *balls, it holds that*

$$\mathbf{Pr}\left[\,\mathrm{Gap}(m) \geqslant \frac{1}{20}\sqrt{\log n}\,\right] \geqslant 1 - o(n^{-2}).$$

Proof. Let $K = 1/10$. Since any adaptive THRESHOLD(f^t) can be simulated by an adaptive QUANTILE(δ^t) process (see Lemma D.1), it suffices to prove the claim for adaptive QUANTILE(δ^t) processes. We will allow the adversary to run two processes, and then choose one that achieves a gap smaller than $C\sqrt{\log n}$ (if such exists):

- **Process $\mathcal{P}_1$.** The adversary has to allocate $m = Kn\sqrt{\log n}$ balls into n bins. The adversary wins if for step m, $\mathrm{Gap}(m) < C\sqrt{\log n}$, and, Condition $\mathcal{K}_1$, at least $(Kn\sqrt{\log n}) \cdot e^{-\sqrt{\log n}}$ out of the m quantiles are larger than $e^{-\sqrt{\log n}}$.
- **Process $\mathcal{P}_2$.** The adversary has to allocate $m = Kn\sqrt{\log n}$ balls into n bins. The adversary wins if for step m, $\mathrm{Gap}(m) < C\sqrt{\log n}$ and, Condition $\mathcal{K}_2$, at least $m - (Kn\sqrt{\log n}) \cdot e^{-\sqrt{\log n}}$ out of the m quantiles are at most $e^{-\sqrt{\log n}}$.

We will prove that the adversary cannot win any of the two games with probability greater than n^{-2}. Now recall the original process, the one we would like to analyse:

- **Process $\mathcal{P}_3$ (adaptive QUANTILE(δ^t)).** The adversary has to allocate $m = Kn\sqrt{\log n}$ balls into bins using one adaptive query at each step. The adversary wins if $\mathrm{Gap}(m) < C\sqrt{\log n}$.

Again, we will show below that $\mathbf{Pr}[\,\text{adversary wins } \mathcal{P}_1\,] = o(n^{-2})$ and $\mathbf{Pr}[\,\text{adversary wins } \mathcal{P}_2\,] = o(n^{-2})$, and these bounds imply that $\mathbf{Pr}[\,\mathcal{P}_3 \text{ wins}\,] = o(n^{-2})$. We now turn to the analysis of $\mathcal{P}_1$ and $\mathcal{P}_2$:

Analysis of Process 1: Let $\mathcal{E}_t$ be the event that (i) QUANTILE allocates at least $n \cdot e^{-\sqrt{\log n}}$ balls with a quantile at least $e^{-\sqrt{\log n}}$ in the interval $[t, t+n]$, and (ii) $\mathrm{Gap}(t+n) \leqslant \frac{1}{5}\sqrt{\log n}$. Note that this is the negation of Lemma C.18, so by union bound over $1 \leqslant t \leqslant m - n$,

$$\mathbf{Pr}\left[\,\bigcup_{t=1}^{m-n} \mathcal{E}_t\,\right] \leqslant K \cdot n\sqrt{\log n} \cdot o(n^{-4}) = o(n^{-2}).$$

Note that if none of the $\mathcal{E}_t$ for $1 \leqslant t \leqslant m - n$ occur, then we either have $\mathrm{Gap}(t) \geqslant \frac{1}{5}\sqrt{\log n}$ at some time $t \leqslant m$ (implying $\mathrm{Gap}(m) \geqslant (\frac{1}{5} - \frac{1}{10})\sqrt{\log n} \geqslant \frac{1}{20}\sqrt{\log n}$), or the adversary allocates less than $\frac{n}{e^{\sqrt{\log n}}} \cdot K\sqrt{\log n}$ out of the m balls with a quantile at least $e^{-\sqrt{\log n}}$. Therefore,

$$\mathbf{Pr}[\,\text{adversary wins } \mathcal{P}_1\,] = o(n^{-2}).$$

Analysis of Process 2: The analysis of $\mathcal{P}_2$ follows directly by Lemma C.19. □

C.4 Lower bound for $(1+\beta)$-process with large β

In this section, we prove a simple bound for the $(1+\beta)$-process which matches our upper bound in Theorem 7.23 up to multiplicative constants for any $1-\beta = \Omega(e^{-\log^c n})$.

Lemma C.21. *Consider the $(1+\beta)$-process with $1-\beta = e^{-\frac{1}{4}\cdot\log^c n}$ for any $c \in \left[\frac{1}{\log\log n}, 1\right)$. Then, there exists $m := m(\beta)$ and a constant $\kappa > 0$ such that*

$$\mathbf{Pr}\left[\, \mathrm{Gap}(m) \geqslant \kappa \cdot \frac{\log n}{-\log(1-\beta)} \,\right] \geqslant 1 - o(1).$$

Proof. Consider $m := \frac{1}{10}\cdot n\log^{1-c} n$ balls. Then, by a Chernoff bound with probability at least $1-o(1)$ we have that at least $\frac{1}{10}(1-\beta)\cdot m$ balls are allocated using ONE-CHOICE. Hence, by Lemma B.23,

$$\mathbf{Pr}\left[\, \max_{i\in[n]} x_i^m \geqslant \frac{1}{4}\cdot \frac{\log n}{\log\left(\frac{4\log^c n}{\frac{1}{10^2}\cdot(1-\beta)}\right)} \,\right] \geqslant 1-o(1).$$

The expression for the maximum load is given by

$$\frac{1}{4}\cdot\frac{\log n}{\log\left(\frac{4\log^c n}{\frac{1}{10^2}\cdot(1-\beta)}\right)} = \frac{1}{4}\cdot\frac{\log n}{\log\left(400\cdot(\log n)\cdot e^{\frac{1}{4}\log^c n}\right)} \geqslant \frac{1}{4}\cdot\frac{\log n}{2\log^c n} = \frac{1}{8}\cdot\log^{1-c} n,$$

using that $e^{\frac{1}{4}\log^c n} \geqslant 400\log^c n$ for any $c = \Omega\left(\frac{1}{\log\log n}\right)$. Hence, subtracting the average, we get the conclusion

$$\frac{1}{8}\cdot\log^{1-c} n - \frac{1}{10}\cdot\log^{1-c} n \geqslant \frac{1}{40}\cdot\log^{1-c} n = \Omega\left(\frac{\log n}{-\log(1-\beta)}\right). \qquad \square$$

C.5 Lower bounds for MEAN-BIASED processes

We shall now define a new condition for allocation processes which is satisfied for many natural processes, including TWINNING, MEAN-THINNING, and $(1+\beta)$ with constant β.

- **Condition $\mathcal{P}_4$**: for any $\epsilon > 0$ there exists a constant $0 < k_3 \leqslant 1$ such that for all steps $t \geqslant 0$ with $\delta^t \in (\epsilon, 1-\epsilon)$ and all bins $i \in [n]$ we have
$$q_i^t \geqslant \frac{k_3}{n}.$$

So essentially this condition implies that in any step t where the mean quantile δ^t is in $(\epsilon, 1-\epsilon)$, there is at least a $\Omega(1/n)$-probability of allocating to each bin.

We shall now observe that

- For the MEAN-THINNING process, the probability of allocating to an overloaded bin $i \in [n]$ is $q_i^t = \frac{\delta^t}{n}$. Thus condition $\mathcal{P}_4$ is satisfied with $k_3 := \epsilon$.
- For the TWINNING process, we have $q_i^t = \frac{1}{n}$. So $\mathcal{P}_4$ is satisfied with $k_3 := 1$.
- The $(1+\beta)$-process has $q_i^t = \frac{1-\beta}{n} + \frac{\beta(2i-1)}{n^2} > \frac{1-\beta}{n}$, satisfying $\mathcal{P}_4$ with $k_3 := 1-\beta$.

The next claim shows that for many steps the mean quantile is not at the extremes.

Claim C.22. *Consider any* MEAN-BIASED *process. For any $m = \Theta(n \log n)$ and $\epsilon > 0$, let $G_1^m := G_1^m(\epsilon)$ be the number of steps $s \in [1, m]$ with $\delta^s \in (\epsilon, 1-\epsilon)$. Then, there exists some $\epsilon > 0$ and a constant $\kappa := \kappa(\epsilon) > 0$ such that*

$$\mathbf{Pr}\left[\, G_1^m \geqslant \kappa \cdot m \,\right] \geqslant 1 - 2 \cdot n^{-12}.$$

Proof. By applying Lemma 5.7 for $t_0 = 0$, (where $\Lambda^0 = n$), we get that there exists a constant $C > 0$ such that $\Delta^t \leqslant C \cdot n$ for a constant fraction of the rounds t in the range $[1, m]$ with probability at least $1 - n^{-12}$. The claim then follows by applying Lemma 5.5 at each of these rounds. □

Lemma C.23. *Consider any* MEAN-BIASED *process satisfying condition $\mathcal{P}_4$. Let $k := k_3\kappa/1000$ where k_3 is specified by $\mathcal{P}_4$ and κ is given by Claim C.22. Then*

$$\mathbf{Pr}\left[\, \mathrm{Gap}(k \cdot n \log n) \geqslant w\_k \cdot \log n \,\right] \geqslant 1 - n^{-1}.$$

Proof. Let $m = k \cdot n \log n$ and let q^t be the allocation vector of the process at step $t \geqslant 0$. By Claim C.22 w.h.p. there exists some constants $\epsilon > 0$ and $\kappa > 0$ such that there are at least κm rounds $s \in [0, m]$ with $\delta^s \in (\epsilon, 1-\epsilon)$. Denote this set of rounds by S and observe that by condition $\mathcal{P}_4$, for any $s \in S$ and $i \in [n]$ we have $q_i^s \geqslant k_3/n$.

Observe that we can couple the location of the balls allocated as rounds in S to locations under a ONE-CHOICE process as follows: before each round $s \in S$ we sample an independent Bernoulli random variable $X_s \sim \mathsf{Ber}(k_3)$ with success probability k_3. If $X_s = 1$ then we allocate the ball(s) to a uniformly random bin. Otherwise, if $X_s = 0$ we allocate the ball(s) to the i-th loaded bin with probability $(q_i^s - k_3/n)/(1-k_3)$. If we let $X = \sum_{s \in S} X_s$ then it follows that, conditional on $|S| \geqslant \kappa m$ we have $X > k_3\kappa m/2$ w.h.p. by the Chernoff bound. It follows that w.h.p. at least $(k_3\kappa k/2) \cdot n \log n$ balls are allocated according to the ONE-CHOICE protocol.

By Lemma B.16, when $cn \log n$ balls are allocated using ONE-CHOICE, for any constant $c > 0$, then with probability at least $1 - n^{-2}$, the max load is at least $(c + \sqrt{c}/10) \log n$. Thus if we choose $k = k_3\kappa/(800 w\_^2)$, since each ball has weight at least 1 and at most $w\_$, we have

$$\frac{\mathrm{Gap}(m)}{\log n} \geqslant \frac{k_3\kappa k}{2} + \frac{1}{10}\sqrt{\frac{k_3\kappa k}{2}} - w\_k \geqslant \frac{1}{10}\sqrt{\frac{k_3\kappa k}{2}} - w\_k = w\_k,$$

with probability $1 - n^{-1}$ by taking the union bound of these three events. □

Hence, we can deduce from the lemma above by recalling that MEAN-THINNING, TWINNING and $(1+\beta)$ all satisfy $\mathcal{P}_4$ and either the conditions $\mathcal{P}_2$ and $\mathcal{W}_3$, or, $\mathcal{W}_2$ and $\mathcal{P}_3$:

Corollary C.24. *For either the* MEAN-THINNING, TWINNING *or* $(1+\beta)$*-processes, there exist a constant $k > 0$ (different for each process) such that*

$$\mathbf{Pr}\left[\, \mathrm{Gap}(k \cdot n \log n) \geqslant k \cdot \log n \,\right] \geqslant 1 - n^{-1}.$$

Finally, we prove a tight lower bound for any RELATIVE-THRESHOLD$(f(n))$ process with $f(n) \geqslant \log n$.

Lemma C.25. *Consider any* RELATIVE-THRESHOLD$(f(n))$ *process with $f(n) \geqslant \log n$, then for $m := \frac{100n \cdot (f(n))^2}{\log n}$*

$$\mathbf{Pr}\left[\, \bigcup_{t \in [0,m]} \{\mathrm{Gap}(t) \geqslant f(n)\} \,\right] \geqslant 1 - n^{-1}.$$

Proof. Since $f(n) \geqslant \log n$, it follows that $m \geqslant n \log n$. We also define the stopping time

$$\tau := \inf\{t \geqslant 0 : \text{Gap}(t) \geqslant f(n)\}.$$

Clearly, if $\tau \leqslant m$, then there is a step t in $[0, m]$ with $\text{Gap}(t) \geqslant f(n)$. Otherwise, in these m steps we are always allocating to the first sample and so we can couple the allocations to that of a One-Choice process (call it $\mathcal{P}$). By Lemma B.16, we have that

$$\mathbf{Pr}\left[\,\text{Gap}_{\mathcal{P}}(m) \geqslant \frac{1}{10} \cdot \sqrt{\frac{m}{n} \cdot \log n} = f(n)\,\right] \geqslant 1 - n^{-1}.$$

Hence,

$$\mathbf{Pr}\left[\,\bigcup_{t \in [0,m]} \{\text{Gap}(t) \geqslant f(n)\}\,\right] \geqslant \mathbf{Pr}\left[\,\text{Gap}_{\mathcal{P}}(m) \geqslant \frac{1}{10} \cdot \sqrt{\frac{m}{n} \cdot \log n} = f(n)\,\right] \geqslant 1 - n^{-1}. \qquad \square$$

C.6 Lower bounds for Memory process

In this section, we state the lower bounds for the Memory processes. The omitted details and complete proofs can be found in [118].

Theorem C.26 ([118, Theorem 1.2]). *Consider the Memory process (with the uniform sampling distribution). Then, there is a constant $\kappa > 0$ such that for every step $m \geqslant n$,*

$$\mathbf{Pr}\left[\,\text{Gap}(m) \geqslant \kappa \cdot \log \log n\,\right] \geqslant 1 - n^{-1}.$$

On a high level, the proof of this theorem follows the layered induction argument used by [18] to lower bound the gap of Two-Choice in the lightly-loaded case. However, for the Memory process in the heavily loaded case, we require some additional arguments to bootstrap the induction and also deal with the correlations between the bins.

We also state a simple lower bound for the $(1, 1, d)$-Reset-Memory process.

Lemma C.27. *For the $(1, 1, d)$-Reset-Memory process with constant $d > 0$ and $m = \frac{1}{400d} n \log n$, we have that*

$$\mathbf{Pr}\left[\,\text{Gap}(m) \geqslant \frac{1}{400d} \cdot \log n\,\right] \geqslant 1 - n^{-2}.$$

Proof. In $(1, 1, d)$-Reset-Memory, every d steps the cache is reset and so the ball is allocated using One-Choice. Hence, in m steps, there are m/d balls allocated using One-Choice. By Lemma B.16, we get

$$\mathbf{Pr}\left[\,\max_{i \in [n]} y_i^m \geqslant \left(\frac{1}{400d^2} + \frac{1}{200d}\right) \cdot \log n\,\right] \geqslant 1 - n^{-2}.$$

Therefore, by definition of $m = \frac{1}{400d} n \log n$,

$$\mathbf{Pr}\left[\,\text{Gap}(m) \geqslant \frac{1}{400d} \cdot \log n\,\right] \geqslant 1 - n^{-2}. \qquad \square$$

Omitted proofs

D.1 Relation between Threshold and Quantile processes

Lemma D.1. *Any* Threshold$(f_1^t, \ldots, f_k^t)$ *process can be simulated by a* Quantile$(\delta_1^t, \ldots, \delta_k^t)$ *process.*

Proof. Consider an arbitrary time step $t \geqslant 0$. Since the process is adaptive, we are allowed to determine the value of δ_j^t by looking at the load distribution x^t. We want to choose δ_j^t, such that comparing the rank $i \leqslant \delta_j^t n$ gives the same answer as $f_j^t \leqslant x_i^t$ for every $i \in [n]$. This can be achieved by choosing δ_j^t to be the largest possible quantile such that $y^t_{\delta_j^t \cdot n} \leqslant f_j^t(t)$. This way any $i \leqslant \delta_j^t n$ will have $x_i^t \leqslant \delta_j^t$ and these will be the only such i's by construction. Hence, at each time step the probability vectors of Quantile$(\delta_1^t, \ldots, \delta_k^t)$ and Threshold$(f_1^t, \ldots, f_k^t)$ will be the same. $\square$

Lemma D.2. *For any* Quantile$(\delta_1^t, \ldots, \delta_k^t)$ *process, there exist thresholds* $f_1^t, \ldots, f_k^t$ *and probability vector* $(\beta_1^t, \ldots, \beta_k^t)$ *such that* $(\beta_1^t, \ldots, \beta_k^t)$–Mixed(Threshold$(f_1^t), \ldots,$ Threshold(f_k^t)).

In other words, there is a reduction from Quantile to adaptive Threshold, but the Threshold process must have the ability to randomise between different instances of Threshold.

Proof. Let us first prove the claim for $k = 1$, that is, Quantile(δ) can be simulated by an adaptive randomised threshold process with one threshold. Since we only analyse one time-step t, we will for simplicity omit this dependency and write $\delta = \delta^t$.

Let δ^1 be the quantile where the values equal to $y_{n \cdot \delta}$ start and δ^2, where they end (so $\delta^1 \leqslant \delta \leqslant \delta^2$). Sampling between a threshold of $y_{n \cdot \delta}$ and $y_{n \cdot \delta} + 1$ with probability $\alpha \in [0,1]$ interpolates between the Quantile(δ^1) and Quantile(δ^2). Let p^1 and p^2 be the probability vectors for Quantile(δ^1) and Quantile(δ^2), then the probability vector q for this adaptive randomised threshold process is given by,

$$q_i = \alpha \cdot p_i^1 + (1-\alpha) \cdot p_i^2.$$

At $i \leqslant n \cdot \delta^1 \leqslant n \cdot \delta^2$, we have,

$$q_i = \alpha \cdot \frac{\delta^1}{n} + (1-\alpha) \cdot \frac{\delta^2}{n}.$$

We pick $\alpha = \frac{\delta^2 - \delta}{\delta^2 - \delta^1} \in [0,1]$ so that $q_i = \frac{\delta}{n}$ for $i \leqslant n \cdot \delta^1$. Then for $i \geqslant n \cdot \delta^2 \geqslant n \cdot \delta^1$, we get

$$q_i = \alpha \cdot \frac{1+\delta^1}{n} + (1-\alpha) \cdot \frac{1+\delta^2}{n} = \frac{\alpha + (1-\alpha)}{n} + \frac{\alpha \cdot \delta^1 + (1-\alpha) \cdot \delta^2}{n} = \frac{1+\delta}{n},$$

by the choice of α. So at the indices $i \in [n] \setminus (\delta^1 n, \delta^2 n]$ agree with Quantile(δ).

At the indices $n \cdot \delta^1 < i \leqslant n \cdot \delta^2$, the probability is shared between bins with the same load, so the effect is indistinguishable (see Fig. D.1), in terms of the resulting load vectors.

We will extend this idea to $k > 1$ quantiles, by replacing each quantile δ_j with a mixture of two thresholds $y_{n \cdot \delta_j}$ and $y_{n \cdot \delta_j} + 1$ with probability α_j. For this, we define δ_j^1 and δ_j^2 with $\delta_j^2 \geqslant \delta_j \geqslant \delta_j^1$ to be the left and right quantiles for the values of $y_{n \cdot \delta_j}$.

To argue that there exist coefficients α_j such that the two processes are equivalent, we start with the probability vector q of the Quantile$(\delta_1, \ldots, \delta_k)$ process. For each $j \in [k]$, construct the probability

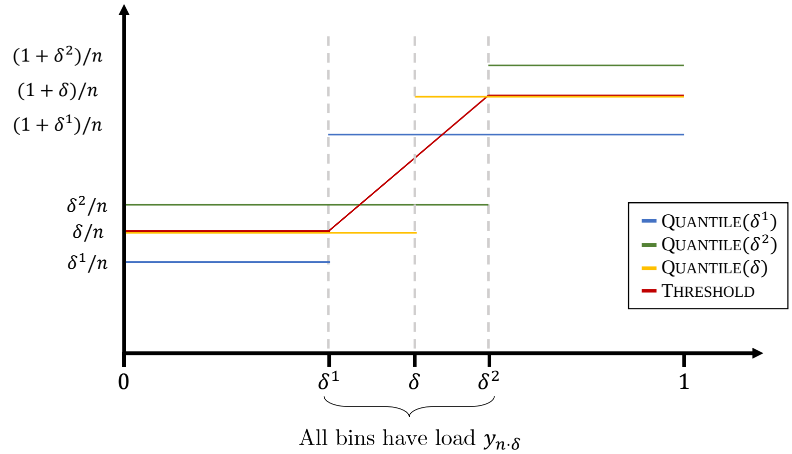


Figure D.1: The ONE-THRESHOLD process which uses a threshold of $y_{n\cdot\delta}$ probability α and $y_{n\cdot\delta}+1$ with probability $1-\alpha$, corresponds to mixing the probability vectors of QUANTILE(δ^1) and QUANTILE(δ^2). The resulting probability vector differs from QUANTILE(δ) only in the region $(n\cdot\delta_1, n\cdot\delta_2]$, where by design all bins have load $y_{n\cdot\delta}$. Hence, the effect of the two processes is indistinguishable.

vector q^j which agrees with q at all $i \leqslant n\cdot\delta_j$, except possibly for values equal to $y_{n\cdot\delta_j}$. For these values at $i \leqslant \delta_j\cdot n$, we will ensure that the processes have the same aggregate probability, so the effect on these bins will be indistinguishable.

In each step we create probability vectors p^{1j} and p^{2j}, by adding quantiles δ_j^1 and δ_j^2 respectively to q^{j-1}. These affect only the values of the entries in $(n\cdot\delta_{j-1}, n\cdot\delta_{j+1}]$. As in the one query case, we choose $\alpha_j := \frac{\delta_j^2-\delta_j}{\delta_j^2-\delta_j^1}$ such that, for $i \in (n\cdot\delta_{j-1}, n\cdot\delta_j^1]$

$$q_i^j = \alpha_j\cdot\left(\frac{\delta_{j-1}+\delta_j^1}{n}\right)+(1-\alpha_j)\cdot\left(\frac{\delta_{j-1}+\delta_j^2}{n}\right) = \frac{\delta_{j-1}}{n}+\alpha_j\cdot\frac{\delta_j^1}{n}+(1-\alpha_j)\cdot\frac{\delta_j^2}{n} = \frac{\delta_{j-1}+\delta_j}{n} = q_i,$$

and for $i \in (n\cdot\delta_j^2, n\cdot\delta_{j+1}]$,

$$q_i^j = \alpha_j\cdot\left(\frac{\delta_j^1+\delta_{j+1}}{n}\right)+(1-\alpha_j)\cdot\left(\frac{\delta_j^2+\delta_{j+1}}{n}\right) = \frac{\delta_{j+1}}{n}+\frac{\alpha_j\delta_j^1+(1-\alpha_j)\delta_j^2}{n} = \frac{\delta_{j-1}+\delta_j}{n} = q_i.$$

The linear weighting preserves the following property: Let B be a set of bins, then if $\sum_{b\in B} p_b^{1j} = \sum_{b\in B} p_b^{2j}$ then $\sum_{b\in B} q_b^j = \sum_{b\in B} p_b^{1j} = \sum_{b\in B} p_b^{2j}$. This implies that:

1. If $p_i^{1j} = p_i^{2j}$, then $q_i^j = p_i^{1j} = p_i^{2j}$.
2. Let B_x be the set of bins in $[1, \delta_{j-1}\cdot n]$ with equal load x. By the inductive argument, in q^j the probability of allocating a ball to x will be the same as in that of q.

Hence, this ensures that each step extends the agreement of probability vector q^j and q to each bin $i \in [1, \delta_{j+1}\cdot n]$. The only possible exceptions are bins with equal load, where the probability mass is just rearranged among them. Hence, q^k will be equivalent to q for the given load vector. □

Lemma D.3. *Any* QUANTILE$(\delta_1,\ldots,\delta_k)$ *process can be simulated by an adaptive (and randomised)* $(2k)$-THINNING *process.*

Proof. We may assume that QUANTILE $(\delta_1,\ldots,\delta_k)$ will process $2k$ queries one by one, and alternate between the two bins. First, send the largest quantile to bin i_1, then send the largest to bin i_2, then send the second largest to bin i_1, etc. and stop as soon as you receive a negative answer. Therefore, for ease of notation, let us set $\gamma_i := \delta_{k-i}$ for $i \in [k]$.

Further, let i_1 and i_2 be two chosen bins, and $\widetilde{i}$ be the bin where the ball is finally placed. Note that

$$\mathbf{Pr}\left[\, \mathrm{Rank}^t(\widetilde{i}) \leqslant n \cdot \gamma_j \,\right] = \gamma_j \cdot \gamma_j.$$

since $\widetilde{i} \in \{i_1, i_2\}$ will be of rank at least $n \cdot \gamma_j$ if and only if both bins i_1 and i_2 satisfy $\mathrm{Rank}^t(i_1) \leqslant n \cdot \gamma_j$ and $\mathrm{Rank}^t(i_2) \leqslant n \cdot \gamma_j$; and those bins are chosen independently.

On the other hand, consider now an adaptive $(2k)$-THINNING process with increasing load thresholds $f_1 \leqslant f_2 \leqslant \ldots \leqslant f_{2k}$ and $2k$ bin choices $i_1, i_2, \ldots, i_{2k}$, which are chosen uniformly and independently at random. Each load threshold f_j applied to bin i_j will be randomised so that it simulates a QUANTILE $(\gamma_{\lfloor (j-1)/2 \rfloor})$ see (Lemma D.2). Further, let $\bar{i}$ be the final bin of this allocation process.

First, the bin i_ℓ in iteration ℓ will not be accepted with probability

$$\mathbf{Pr}\left[\, \mathrm{Rank}^t(i_\ell) \leqslant n \cdot \gamma_{1+\lfloor \ell/2 \rfloor} \,\right] = \gamma_{1+\lfloor \ell/2 \rfloor},$$

and using the independence of the first $2j$ sampled different bins, we obtain

$$\begin{aligned}
\mathbf{Pr}\left[\, \mathrm{Rank}^t(\bar{i}) \leqslant n \cdot \gamma_j \,\right] &= \prod_{\ell=1}^{2j} \mathbf{Pr}\left[\, \mathrm{Rank}^t(i_\ell) \leqslant n \cdot \gamma_{1+\lfloor \ell/2 \rfloor} \,\right] \\
&= \gamma_1 \cdot \gamma_1 \cdot \gamma_2 \cdot \gamma_2 \cdot \ldots \cdot \gamma_j \cdot \gamma_j \leqslant \gamma_j \cdot \gamma_j \leqslant \mathbf{Pr}\left[\, \mathrm{Rank}^t(\widetilde{i}) \leqslant n \cdot \gamma_j \,\right].
\end{aligned}$$

□

D.2 WEIGHTED setting

Lemma D.4. *There exists $S := S(\zeta) \geqslant \max\{1, 1/\zeta\}$, such that for any $\alpha \in (0, \min\{\zeta/2, 1\})$ and any $\kappa \in [-1, 1]$,*

$$\mathbf{E}\left[\, e^{\alpha \cdot \kappa \cdot \mathcal{W}} \,\right] \leqslant 1 + \alpha \cdot \kappa + S\alpha^2 \cdot \kappa^2.$$

Proof. This proof closely follows the argument in [152, Lemma 2.1]. Let $M(z) = \mathbf{E}\left[\, e^{z\mathcal{W}} \,\right]$, then using Taylor's Theorem (mean value form remainder), for any $z \in [-\alpha, \alpha]$ there exists $\xi \in [-\alpha, \alpha]$ such that

$$M(z) = M(0) + M'(0) \cdot z + M''(\xi) \cdot \frac{1}{2} \cdot z^2 = 1 + z + M''(\xi) \cdot \frac{1}{2} \cdot z^2.$$

By the assumptions on α and ζ,

$$\begin{aligned}
M''(\xi) = \mathbf{E}[\, \mathcal{W}^2 e^{\xi \mathcal{W}} \,] &\overset{(a)}{\leqslant} \sqrt{\mathbf{E}[\, \mathcal{W}^4 \,] \cdot \mathbf{E}[\, e^{2\xi \mathcal{W}} \,]} \overset{(b)}{\leqslant} \frac{1}{2} \cdot \left(\mathbf{E}[\, \mathcal{W}^4 \,] + \mathbf{E}[\, e^{2\xi \mathcal{W}} \,] \right) \\
&\overset{(c)}{\leqslant} \frac{1}{2} \cdot \left(\left(\frac{8}{\zeta} \cdot \log\left(\frac{8}{\zeta} \right) \right)^4 + \mathbf{E}[\, e^{\zeta \mathcal{W}} \,] + \mathbf{E}[\, e^{\zeta \mathcal{W}} \,] \right).
\end{aligned}$$

where (a) uses the Cauchy-Schwartz inequality $|\mathbf{E}[\, X \cdot Y \,]| \leqslant \sqrt{\mathbf{E}[\, X^2 \,] \mathbf{E}[\, Y^2 \,]}$ for random variables X and Y, (b) uses a mean inequality, and (c) uses Lemma B.13. Now defining

$$S := 2 \cdot \max\left\{ \left(\frac{8}{\zeta} \cdot \log\left(\frac{8}{\zeta} \right) \right)^4, 2 \cdot \mathbf{E}[\, e^{\zeta \mathcal{W}} \,], 1/2 \right\},$$

and choosing $z := \kappa \cdot \alpha$, the lemma follows. □

D.3 MEAN-BIASED processes

Lemma D.5. *For any constant $\eta > 0$, the $(1+\eta)$-process is a* MEAN-BIASED *process.*

Proof. Let q^t be the allocation vector of $(1+\eta)$ with parameter $\eta \in (0,1)$; that is, with probability η it performs MEAN-THINNING, otherwise ONE-CHOICE. Then for any $i \in B_+^t$,

$$q_i^t = (1-\eta)\cdot\frac{1}{n} + \eta\cdot\frac{\delta^t}{n} = \frac{1}{n} - \frac{\eta\cdot(1-\delta^t)}{n}.$$

Similarly, for any $i \in B_-^t$,

$$q_i^t = (1-\eta)\cdot\frac{1}{n} + \eta\cdot\frac{1+\delta^t}{n} = \frac{1}{n} + \frac{\eta\cdot\delta^t}{n}.$$

Thus for $k_1 = k_2 = \eta \in (0,1)$, the $(1+\eta)$ process satisfies $\mathcal{P}_3$ and $\mathcal{W}_2$. $\square$

Lemma D.6. *For any constant $\beta \in (0,1]$, the $(1+\eta)$-process with $\eta = \beta$ majorizes $(1+\beta)$-process at each step.*

Proof. We will show that for any step $t \geqslant 0$ and for any load vector, the $(1+\eta)$ for $\eta = \beta$ allocation vector majorizes the allocation vector of $(1+\beta)$-process. So, by Theorem 2.5, the claim will follow.

Recall that the $(1+\beta)$ allocation vector p^t is given by,

$$p_i^t = p_i = (1-\beta)\cdot\frac{1}{n} + \frac{\beta(2i-1)}{n^2}.$$

The sorted allocation vector $\widetilde{q}^t$ for the $(1+\eta)$ process is non-decreasing and uniform over B_-^t and B_+^t, so majorization follows immediately once we prove that

$$\sum_{i=1}^{|B_+^t|} p_i \leqslant \sum_{i=1}^{|B_+^t|} \widetilde{q}_i^t.$$

For the allocation vector p we have,

$$\sum_{i=1}^{\delta^t n} p_i = \sum_{i=1}^{\delta^t n}(1-\beta)\cdot\frac{1}{n} + \sum_{i=1}^{\delta^t n}\frac{\beta(2i-1)}{n^2} = (1-\beta)\cdot\delta^t + \frac{\beta(\delta^t\cdot n)^2}{n^2} = \delta^t - \beta\cdot(\delta^t - (\delta^t)^2).$$

Similarly, for the allocation vector $\widetilde{q}^t$ we have,

$$\sum_{i=1}^{\delta^t n}\widetilde{q}_i^t = \sum_{i=1}^{\delta^t n}\frac{1}{n} - \sum_{i=1}^{\delta^t n}\frac{\beta(1-\delta^t)}{n} = \delta^t - \beta\cdot(\delta^t - (\delta^t)^2).$$ $\square$

D.4 g-ADV setting

Claim D.7. *Consider $\alpha_1, \alpha_2 > 0$ as defined in Eq. (7.4) and Eq. (7.5) respectively. Then, for any $g \geqslant \frac{\alpha_2}{4\sqrt{\alpha_1}}$ and $k := k(g) \geqslant 2$ being the unique integer such that $(\alpha_1 \log n)^{1/k} < g \leqslant (\alpha_1 \log n)^{1/(k-1)}$, it holds that*

$$(\alpha_1\cdot(\log n))^{1/k} \leqslant \left(\frac{\alpha_2}{4}\cdot(\log n)\right)^{1/(k-1)}.$$

Proof. Let $R := \frac{\alpha_2}{4\sqrt{\alpha_1}} = \frac{\sqrt{\alpha_1}}{4\cdot 84} \leqslant 1$ (using that $\alpha_2 = \frac{\alpha_1}{84}$ and $\alpha_1 \leqslant 1$). By rearranging the target inequality,

$$\begin{aligned}
\left(\frac{\alpha_2}{4}\cdot(\log n)\right)^{1/(k-1)} \cdot (\alpha_1 \cdot (\log n))^{-1/k} &= \exp\left(\frac{1}{k-1}\cdot \log\left(\frac{\alpha_2}{4}\log n\right) - \frac{1}{k}\cdot\log(\alpha_1 \log n)\right) \\
&\overset{(a)}{\geqslant} \exp\left(\frac{1}{k-1}\cdot\left(\log R + \frac{1}{k}\cdot \log\log n\right)\right) \\
&\overset{(b)}{\geqslant} \exp\left(\frac{1}{k-1}\cdot(\log R - \log R)\right) = 1.
\end{aligned}$$

using in (a) that $-\frac{1}{k}\log\alpha_1 \geqslant -\frac{1}{2(k-1)}\log\alpha_1$ and in (b) that $k \leqslant \frac{1}{-\log R}\cdot\log\log n$ since $g \geqslant R^{-1}$. □

D.5 Concentration for the b-BATCHED setting

In this chapter, we prove that the hyperbolic cosine potential concentrates at $\mathcal{O}(n)$ for a large family of processes in the b-BATCHED setting for $n \leqslant b \leqslant n^3$. In particular, we complete the proof of the following lemma stated in Section 7.5, which is essential for getting the $\mathcal{O}\big(\frac{b}{n} + \log n\big)$ bound on the gap (Theorem 7.39).

Recall that in Theorem 7.39, we considered the WEIGHTED b-BATCHED setting with any $n \leqslant b \leqslant n^3$ and weights from a FINITE-MGF(ζ, S) distribution with constant $\zeta, S \geqslant 1$, for any SEQUENTIAL(q^t) process with q^t satisfying condition $\mathcal{C}_1$ for constant $\delta \in (0,1)$ and constant $\epsilon \in (0,1)$ as well as condition $\mathcal{C}_2$ for some constant $C > 1$, at every step $t \geqslant 0$.

Lemma D.8. *Consider any process satisfying the conditions in Theorem 7.39. Let $\tilde{c} := 2\cdot\frac{8c}{\delta}$ where $c := c(\delta) > 0$ is the constant from Theorem 3.2. Then, for any step $t \geqslant 0$ being a multiple of b,*

$$\mathbf{Pr}\left[\bigcap_{j\in[0,\log^3 n]}\left\{\Gamma_2^{t+j\cdot b} \leqslant \tilde{c}\cdot n\right\}\right] \geqslant 1 - n^{-3}.$$

The proof of this lemma is similar to the proof of Theorem 4.1 in Chapter 4, in that we use the interplay between two instances of the hyperbolic cosine potential $\Gamma_1 := \Gamma_1(\gamma_1)$ and $\Gamma_2 := \Gamma_2(\gamma_2)$ with smoothing parameter γ_2 being a large constant factor smaller than γ_1. More specifically, we will be working with $\gamma_1 := \frac{\epsilon\delta}{40\cdot C^2\cdot S^2}\cdot\min\left\{\frac{1}{\log n}, \frac{n}{b}\right\}$ and $\gamma_2 := \frac{\gamma_1}{8\cdot 30}$.

The rest of this chapter is organised as follows. In Appendix D.5.1, we establish some basic properties for the potentials Γ_1 and Γ_2 and in Appendix D.5.2 we use these to show that w.h.p. $\Gamma_2^t = \mathcal{O}(n)$ for at least $\log^3 n$ batches, and complete the proof of Lemma D.8.

D.5.1 Preliminaries

We define the following event, for any step $t \geqslant 0$

$$\mathcal{H}^t := \left\{w^t \leqslant \frac{15}{\zeta}\cdot\log n\right\},$$

which means that the weight of the ball sampled in step t is $\mathcal{O}(\log n)$ (since by assumption $\zeta > 0$ is constant).

Lemma D.9. *Consider any* FINITE-MGF(ζ) *distribution* $\mathcal{W}$ *with constant* $\zeta > 0$. *Then, for any steps* $t_0 \geqslant 0$ *and* $t_1 \in [t_0, t_0 + n^3 \log^3 n]$, *we have that*

$$\mathbf{Pr}\left[\bigcap_{s \in [t_0, t_1]} \mathcal{H}^s \right] \geqslant 1 - n^{-10}$$

Proof. Consider an arbitrary step $s \in [t_0, t_1]$. Since w^s is sampled according to W with $\mathbf{E}[\, e^{\zeta W} \,] < \infty$, by Lemma B.8, we have that

$$\mathbf{Pr}\left[\, w^s > \frac{15}{\zeta} \cdot \log n \,\right] \leqslant n^{-14}.$$

By taking the union bound over the interval $[t_0, t_1]$ and since $t_1 - t_0 \leqslant n^4$, we get the conclusion. $\square$

We will now show that when $\Gamma_1^t = \mathrm{poly}(n)$ and $\mathcal{H}^t$ holds, then $\Delta\Gamma_2^{t+1}$ is small.

Lemma D.10. *Consider any process satisfying the conditions in Lemma D.8 and any step* $t \geqslant 0$, *such that* $\Gamma_1^t \leqslant 2\widetilde{c} \cdot n^{26}$ *and* $\mathcal{H}^t$ *holds. Then, we have that*

$$\begin{aligned} &(i) && \Gamma_2^t \leqslant n^{5/4}, \\ &(ii) && \left|\Gamma_2^{t+1} - \Gamma_2^t\right| \leqslant \frac{n}{b} \cdot n^{1/4}. \end{aligned}$$

Further, let $\widehat{x}^t$ *be the load vector obtained by moving the* t*-th ball of the load vector* x^t *to some other bin, then*

$$(iii) \qquad \Gamma_1^t(\widehat{x}^t) \leqslant 2 \cdot \Gamma_1^t(x^t).$$

Proof. Consider any step $t \geqslant 0$, such that $\Gamma_1^t \leqslant 2\widetilde{c} \cdot n^{26}$ and $\mathcal{H}^t$ holds. We start by bounding the load of any bin. For any bin $i \in [n]$,

$$\Gamma_1^t \leqslant 2\widetilde{c} \cdot n^{26} \Rightarrow e^{\gamma_1 \cdot y_i^t} + e^{-\gamma_1 \cdot y_i^t} \leqslant \widetilde{c} \cdot n^{26} \Rightarrow y_i^t \leqslant \frac{27}{\gamma_1} \log n \wedge -y_i^t \leqslant \frac{27}{\gamma_1} \log n, \tag{D.1}$$

where in the second implication we used $\log(2\widetilde{c}) + \frac{26}{\gamma_1} \log n \leqslant \frac{27}{\gamma_1} \log n$, for sufficiently large n.

First statement. Using Eq. (D.1), we can bound the contribution of any bin $i \in [n]$ to Γ_2^t as follows,

$$\Gamma_{2i}^t \leqslant e^{\gamma_2 y_i^t} + e^{-\gamma_2 y_i^t} \leqslant 2 \cdot e^{\gamma_2 \cdot \frac{27}{\gamma_1} \log n} \leqslant 2 \cdot n^{1/8}, \tag{D.2}$$

using that $\gamma_2 := \frac{\gamma_1}{8 \cdot 30}$. By aggregating, we get the first claim $\Gamma_1^t = \sum_{i=1}^n \Gamma_{1i}^t \leqslant 2 \cdot n \cdot n^{1/8} \leqslant n^{5/4}$.

Second statement. Consider the change for the bin $j \in [n]$ where the ball was allocated. Since $\gamma_2 < \frac{1}{40 \cdot S \cdot \log n}$ and $S > \frac{1}{\zeta}$, we have $\gamma_2 \cdot \frac{15}{\zeta} \cdot \log n \leqslant 1$ and so by a Taylor estimate, $e^{\gamma_2 \cdot \frac{15}{\zeta} \cdot \log n} \leqslant 1 + 2 \cdot \gamma_2 \cdot \frac{15}{\zeta} \cdot \log n$. If $j \in [n]$ is an overloaded bin ($y_j^t \geqslant 0$), then

$$\left|\Delta\Gamma_{2j}^{t+1}\right| \leqslant \Gamma_{2j}^t \cdot e^{\gamma_2 \cdot \frac{15}{\zeta} \cdot \log n} - \Gamma_{2j}^t \leqslant \Gamma_{2j}^t \cdot \left(1 + \gamma_2 \cdot \frac{30}{\zeta} \cdot \log n\right) - \Gamma_{2j}^t = \Gamma_{2j}^t \cdot \gamma_2 \cdot \frac{30}{\zeta} \cdot \log n \leqslant \frac{n}{b} \cdot n^{1/8} \cdot \log n,$$

using Eq. (D.2) and $\gamma_2 \leqslant \frac{\epsilon\delta}{40 \cdot C^2 \cdot S^2} \cdot \frac{n}{b}$. Similarly, if j is underloaded ($y_j^t < 0$), then

$$\left|\Delta\Gamma_{2j}^{t+1}\right| \leqslant \Gamma_{2j}^t - \Gamma_{2j}^t \cdot e^{-\gamma_2 \cdot \frac{15}{\zeta} \cdot \log n} \leqslant \Gamma_{2j}^t - \Gamma_{2j}^t \cdot \left(1 - \gamma_2 \cdot \frac{30}{\zeta} \cdot \log n\right) = \Gamma_{2j}^t \cdot \gamma_2 \cdot \frac{30}{\zeta} \cdot \log n \leqslant \frac{n}{b} \cdot n^{1/8} \cdot \log n.$$

The contribution of the rest of the bins is due to the change in the average load. In particular, for any overloaded bin $i \in [n] \setminus \{j\}$,

$$\left|\Delta\Gamma_{2i}^{t+1}\right| \leqslant \Gamma_{2i}^{t} \cdot e^{\gamma_2 \cdot \frac{15}{\zeta} \cdot \frac{\log n}{n}} - \Gamma_{2i}^{t} \leqslant \Gamma_{2i}^{t} \cdot \left(1 + 2 \cdot \gamma_2 \cdot \frac{15}{\zeta} \cdot \frac{\log n}{n}\right) - \Gamma_{2i}^{t} = \Gamma_{2i}^{t} \cdot \gamma_2 \cdot \frac{30}{\zeta} \cdot \frac{\log n}{n} \leqslant \frac{1}{b} \cdot \log n \cdot n^{1/8}.$$

Similarly, for an underloaded bin $i \in [n] \setminus \{j\}$,

$$\left|\Delta\Gamma_{2i}^{t+1}\right| \leqslant \Gamma_{2i}^{t} - \Gamma_{2i}^{t} \cdot e^{-\gamma_2 \cdot \frac{15}{\zeta} \cdot \frac{\log n}{n}} \leqslant \Gamma_{2i}^{t} - \Gamma_{2i}^{t} \cdot \left(1 - 2 \cdot \gamma_2 \cdot \frac{15}{\zeta} \cdot \frac{\log n}{n}\right) = \Gamma_{2i}^{t} \cdot \gamma_2 \cdot \frac{30}{\zeta} \cdot \frac{\log n}{n} \leqslant \frac{1}{b} \cdot \log n \cdot n^{1/8}.$$

Hence, aggregating over all bins

$$\left|\Delta\Gamma_{2}^{t+1}\right| \leqslant \left|\Delta\Gamma_{2j}^{t+1}\right| + \sum_{i \in [n]\setminus\{j\}} \left|\Delta\Gamma_{2i}^{t+1}\right| \leqslant 2 \cdot \frac{n}{b} \cdot n^{1/8} \cdot \log n + n \cdot \frac{1}{b} \cdot \log n \cdot n^{1/8} \leqslant \frac{n}{b} \cdot n^{1/4},$$

for sufficiently large n.

Third statement. Let $i, j \in [n]$ be the differing bins between x^t and $\widehat{x}^t$. Then since $\mathcal{H}^t$ holds, $w^t \leqslant \frac{15}{\zeta} \cdot \log n$, so for any bin $i \in [n]$

$$\Gamma_{1i}(\widehat{x}^t) \leqslant e^{\gamma w^t} \cdot \Gamma_{1i}^{t}(x^t) \leqslant 2 \cdot \Gamma_{1i}^{t}(x^t),$$

since $\gamma < \frac{1}{40 \cdot S \cdot \log n}$ and $S > 1/\zeta$. Similarly, for bin j,

$$\Gamma_{1j}(\widehat{x}^t) \leqslant e^{\gamma w^t} \cdot \Gamma_{1j}^{t}(x^t) \leqslant 2 \cdot \Gamma_{1j}^{t}(x^t).$$

Hence,

$$\Gamma_{1}^{t}(\widehat{x}^t) = \sum_{k=1}^{n} \Gamma_{1k}^{t}(\widehat{x}^t) \leqslant \sum_{k=1}^{n} 2 \cdot \Gamma_{1k}^{t}(x^t) = 2 \cdot \Gamma_{1}^{t}(x^t). \qquad \square$$

Next, we will show that $\mathbf{E}[\, \Gamma_2 \,] = \mathcal{O}(n)$ and that when Γ_2 is sufficiently large, it drops in expectation over the next batch.

Lemma D.11. *Consider any process satisfying the conditions in Lemma D.8. Then, for any step $t \geqslant 0$ being a multiple of b,*

$$(i) \quad \mathbf{E}[\, \Gamma_1^t \,] \leqslant \frac{\widetilde{c}}{2} \cdot n, \quad \textit{and} \quad (ii) \quad \mathbf{E}[\, \Gamma_2^t \,] \leqslant \frac{\widetilde{c}}{2} \cdot n.$$

Further, there exists a constant $\widetilde{c}_1 := \widetilde{c}_1(\epsilon, \delta) > 0$ such that

$$(iii) \quad \mathbf{E}\left[\, \Gamma_2^{t+b} \,\middle|\, \mathfrak{F}^t, \Gamma_2^t \geqslant \widetilde{c} \cdot n \,\right] \leqslant \Gamma_2^t \cdot \left(1 - \frac{\widetilde{c}_1}{\log n}\right),$$

and

$$(iv) \quad \mathbf{E}\left[\, \Gamma_2^{t+b} \,\middle|\, \mathfrak{F}^t, \Gamma_2^t \leqslant \widetilde{c} \cdot n \,\right] \leqslant \widetilde{c} \cdot n - \frac{n}{\log^2 n}.$$

Proof. First/Second statement. The statements follow immediately by Lemma 3.13 and Theorem 3.2, by setting $\widetilde{c} := 16c/\delta$, since $c := c(\delta) > 0$.

Also, using Lemma 3.13 and Theorem 3.2 for smoothing parameter γ_2, we get that for any $t \geqslant 0$,

$$\mathbf{E}\left[\, \Gamma_2^{t+b} \,\middle|\, \mathfrak{F}^t \,\right] \leqslant \Gamma_2^t \cdot \left(1 - b \cdot \frac{\epsilon\delta}{8n} \cdot \gamma_2\right) + b \cdot c\gamma_2\epsilon. \tag{D.3}$$

Third statement. Let $\widetilde{c}_3 := \frac{1}{2} \cdot b \cdot \frac{\epsilon\delta}{8n} \cdot \gamma_2 \geqslant \widetilde{c}_1 / \log n$, for some constant $\widetilde{c}_1 > 0$ since $\gamma_2 = \Theta(\min\{n/b, 1/\log n\})$ and $\epsilon > 0$ is constant. When $\Gamma_2^t \geqslant \widetilde{c} \cdot n$, then Eq. (D.3) yields,

$$\begin{aligned}
\mathbf{E}\left[\Gamma_2^{t+b} \,\middle|\, \mathfrak{F}^t, \Gamma_2^t \geqslant \widetilde{c} \cdot n \right] &\leqslant \Gamma_2^t \cdot \left(1 - 2 \cdot \widetilde{c}_3\right) + b \cdot c\gamma_2\epsilon \\
&\leqslant \Gamma_2^t - \widetilde{c}_3 \cdot \Gamma_2^t + \left(b \cdot c\gamma_2\epsilon - \widetilde{c}_3 \cdot \Gamma_2^t\right) \\
&\leqslant \Gamma_2^t - \widetilde{c}_3 \cdot \Gamma_2^t + \left(b \cdot c\gamma_2\epsilon - \frac{1}{2} \cdot b \cdot \frac{\epsilon\delta}{8n} \cdot \gamma_2 \cdot \frac{16c}{\delta} \cdot n\right) \\
&\leqslant \left(1 - \frac{\widetilde{c}_1}{\log n}\right) \cdot \Gamma_2^t.
\end{aligned}$$

Fourth statement. Similarly, when $\Gamma_1^t < \widetilde{c} \cdot n$, Eq. (D.3) yields,

$$\begin{aligned}
\mathbf{E}\left[\Gamma_2^{t+b} \,\middle|\, \mathfrak{F}^t, \Gamma_2^t < \widetilde{c} \cdot n \right] &\leqslant \widetilde{c} \cdot n \cdot \left(1 - 2 \cdot \widetilde{c}_3\right) + b \cdot c\gamma_2\epsilon \\
&= \widetilde{c} \cdot n - \widetilde{c} \cdot \widetilde{c}_3 \cdot n + \left(b \cdot c\gamma_2\epsilon - \widetilde{c} \cdot \widetilde{c}_3 \cdot n\right) \\
&\leqslant \widetilde{c} \cdot n - \frac{\widetilde{c} \cdot \widetilde{c}_1}{\log n} \cdot n \leqslant \widetilde{c} \cdot n - \frac{n}{\log^2 n}.
\end{aligned}$$

□

In the next lemma, we show that w.h.p. Γ_1 is poly(n) for every step in an interval of length $2b \log^3 n$.

Lemma D.12. *Let $\widetilde{c} := 2 \cdot \frac{8c}{\delta}$ be the constant defined in Lemma D.11. For any $n \leqslant b \leqslant n^3$ and for any step $t \geqslant 0$ being a multiple of b,*

$$\mathbf{Pr}\left[\bigcap_{s \in [t, t+2\log^3 n]} \left\{ \Gamma_1^s \leqslant \widetilde{c} \cdot n^{26} \right\} \right] \geqslant 1 - n^{-10}.$$

Proof. We will start by bounding Γ_1^s at steps s being a multiple of b. Using Lemma D.11 (i), Markov's inequality and the union bound, we have for any $t \geqslant 0$,

$$\mathbf{Pr}\left[\bigcap_{s \in [0, 2\log^3 n]} \left\{ \Gamma_1^{t+s\cdot b} \leqslant \widetilde{c} \cdot n^{12} \right\} \right] \geqslant 1 - \frac{2\log^3 n}{n^{11}}. \tag{D.4}$$

Given that $\Gamma_1^{t+s\cdot b} \leqslant \widetilde{c} \cdot n^{12}$, we will upper bound $\Gamma_1^{t+s\cdot b+r}$ for any $r \in [0, b)$. To this end, recalling that $\Gamma_{1i}^{t+s\cdot b+r} := \Phi_{1i}^{t+s\cdot b+r} + \Psi_{1i}^{t+s\cdot b+r}$, we will upper bound for each bin $i \in [n]$ the terms $\Phi_{1i}^{t+s\cdot b+r}$ and $\Psi_{1i}^{t+s\cdot b+r}$ separately. Proceeding using Eq. (3.18) in Lemma 3.13 (since $\gamma_1 \leqslant 1$ and p satisfies $\mathcal{C}_2$),

$$\begin{aligned}
\mathbf{E}\left[\Phi_{1i}^{t+s\cdot b+r} \,\middle|\, \mathfrak{F}^{t+s\cdot b} \right] &\leqslant \Phi_{1i}^{t+s\cdot b} \cdot \left(1 + \frac{C\gamma_1}{n} + 2 \cdot \frac{C}{n} \cdot S\gamma_1^2\right)^r \\
&\overset{(a)}{\leqslant} \Phi_{1i}^{t+s\cdot b} \cdot \left(1 + \frac{2C\gamma_1}{n}\right)^r \\
&\leqslant \Phi_{1i}^{t+s\cdot b} \cdot e^{2\gamma_1 C \cdot \frac{r}{n}} \leqslant \Phi_{1i}^{t+s\cdot b} \cdot e^{2\gamma_1 C \cdot \frac{b}{n}} \overset{(b)}{\leqslant} 2 \cdot \Phi_{1i}^{t+s\cdot b},
\end{aligned}$$

using in (a) that $\gamma_1 \leqslant \frac{\epsilon\delta}{40 \cdot C^2 \cdot S^2} \leqslant \frac{2}{S}$ and in (b) that $\gamma_1 \leqslant \frac{\epsilon\delta}{40 \cdot C^2 \cdot S^2} \cdot \frac{n}{b} \leqslant \frac{1}{4C} \cdot \frac{n}{b}$. Similarly, using Eq. (3.20) in Lemma 3.13,

$$\mathbf{E}\left[\Psi_{1i}^{t+s\cdot b+r} \,\middle|\, \mathfrak{F}^{t+s\cdot b} \right] \leqslant \Psi_{1i}^{t+s\cdot b} \cdot \left(1 + \frac{C\gamma_1}{n} + 2 \cdot \frac{C}{n} \cdot S\gamma_1^2\right)^r$$

$$\overset{(a)}{\leqslant} \Psi_{1i}^{t+s\cdot b} \cdot \left(1 + \frac{2C\gamma_1}{n}\right)^r$$

$$\leqslant \Psi_{1i}^{t+s\cdot b} \cdot e^{2\gamma C \cdot \frac{r}{n}} \leqslant \Psi_{1i}^{t+s\cdot b} \cdot e^{2\gamma C \cdot \frac{b}{n}} \overset{(b)}{\leqslant} 2 \cdot \Psi_{1i}^{t+s\cdot b},$$

using in (a) that $\gamma_1 \leqslant \frac{\epsilon\delta}{40\cdot C^2\cdot S^2} \leqslant \frac{2}{S}$ and in (b) that $\gamma_1 \leqslant \frac{\epsilon\delta}{40\cdot C^2\cdot S^2} \cdot \frac{n}{b} \leqslant \frac{1}{4C} \cdot \frac{n}{b}$. Hence, aggregating over the bins,

$$\mathbf{E}\left[\, \Gamma_1^{t+s\cdot b+r} \,\middle|\, \mathfrak{F}^{t+s\cdot b}, \Gamma_1^{t+s\cdot b} \,\right] \leqslant 2 \cdot \Gamma_1^{t+s\cdot b}.$$

Applying Markov's inequality, for any $r \in [0, b)$,

$$\mathbf{Pr}\left[\, \Gamma_1^{t+s\cdot b+r} \leqslant n^{14} \cdot \Gamma_1^{t+s\cdot b} \,\right] \geqslant 1 - 2 \cdot n^{-14}.$$

Hence, by a union bound over the $2b \cdot \log^3 n \leqslant 2 \cdot n^3 \cdot \log^3 n$ possible rounds for $s \in [0, 2\log^3 n]$ and $r \in [0, b)$,

$$\mathbf{Pr}\left[\bigcap_{r\in[0,b]} \bigcap_{s\in[0,2\log^3 n]} \left\{\Gamma_1^{t+s\cdot b+r} \leqslant n^{14} \cdot \Gamma_1^{t+s\cdot b}\right\} \right] \geqslant 1 - 2 \cdot n^{-14} \cdot 2b\log^3 n \geqslant 1 - \frac{1}{2} \cdot n^{-10}. \tag{D.5}$$

Finally, taking the union bound of Eq. (D.4) and Eq. (D.5), we conclude

$$\begin{aligned}
&\mathbf{Pr}\left[\bigcap_{s\in[t,t+2b\log^3 n]} \left\{\Gamma_1^s \leqslant \widetilde{c} \cdot n^{26}\right\} \right] \\
&\geqslant \mathbf{Pr}\left[\bigcap_{r\in[0,b]} \bigcap_{s\in[0,2\log^3 n]} \left\{\Gamma_1^{t+s\cdot b+r} \leqslant n^{14} \cdot \Gamma_1^{t+s\cdot b}\right\} \cap \bigcap_{s\in[0,2\log^3 n]} \left\{\Gamma_1^{t+s\cdot b} \leqslant \widetilde{c} \cdot n^{12}\right\} \right] \\
&\geqslant 1 - \frac{1}{2} \cdot n^{-10} - \frac{2\log^3 n}{n^{11}} \geqslant 1 - n^{-10}.
\end{aligned}$$

□

We will now show that w.h.p. there is a step every $b\log^3 n$ steps, such that the exponential potential Γ_2 becomes $\mathcal{O}(n)$. We call this the *recovery* phase.

Lemma D.13 (Recovery). *Let $\widetilde{c} := 2 \cdot \frac{8c}{\delta}$ be the constant defined in Lemma D.11. For any step $t \geqslant 0$ being a multiple of b,*

$$\mathbf{Pr}\left[\bigcup_{s\in[0,\log^3 n]} \left\{\Gamma_2^{t+s\cdot b} \leqslant \widetilde{c} \cdot n\right\} \right] \geqslant 1 - 2 \cdot n^{-8}.$$

Proof. By Lemma D.11 (ii), using Markov's inequality at step t being a multiple of b, we have

$$\mathbf{Pr}\left[\, \Gamma_2^t \leqslant \widetilde{c} \cdot n^9 \,\right] \geqslant 1 - n^{-8}. \tag{D.6}$$

We will be assuming $\Gamma_2^t \leqslant \widetilde{c} \cdot n^9$. By Lemma D.11 (iii), there exists a constant $\widetilde{c}_1 > 0$ such that for any step $r \geqslant 0$, then

$$\mathbf{E}\left[\, \Gamma_2^{r+1} \,\middle|\, \mathfrak{F}^r, \Gamma_2^r > \widetilde{c} \cdot n \,\right] \leqslant \Gamma_2^r \cdot \left(1 - \frac{\widetilde{c}_1}{\log n}\right).$$

In order to prove that Γ_2 is small in some $s \in [0, b\log^3 n]$, we define the "killed" potential function for any $r \in [0, \log^3 n]$,

$$\widehat{\Gamma}_2^{t+r\cdot b} := \Gamma_2^{t+r\cdot b} \cdot \mathbf{1}_{\bigcap_{s\in[0,r]}\{\Gamma_2^{t+s\cdot b} > \widetilde{c}\cdot n\}}.$$

Note that $\widehat{\Gamma}_2^{t+r\cdot b} \leqslant \Gamma_2^{t+r\cdot b}$ and that $\{\widehat{\Gamma}_2^{t+r\cdot b} = 0\}$ implies that $\{\widehat{\Gamma}_2^{t+r\cdot b+1} = 0\}$. Hence, the $\widehat{\Gamma}$ potential satisfies unconditionally the drop inequality of Lemma D.11 (iii), that is,

$$\mathbf{E}\Big[\ \widehat{\Gamma}_2^{t+(r+1)\cdot b}\ \Big|\ \mathfrak{F}^{t+r\cdot b}, \widehat{\Gamma}_2^{t+r\cdot b}\ \Big] \leqslant \widehat{\Gamma}_2^{t+r\cdot b} \cdot \Big(1 - \frac{\widetilde{c}_1}{\log n}\Big).$$

Inductively applying this for $\log^3 n$ batches, and since $\widetilde{c}_1 := \widetilde{c}_1(\epsilon, \delta) > 0$ is a constant,

$$\mathbf{E}\Big[\ \widehat{\Gamma}_2^{t+(\log^3 n)\cdot b}\ \Big|\ \mathfrak{F}^t, \Gamma_2^t\ \Big] \leqslant \Gamma_2^t \cdot \Big(1 - \frac{\widetilde{c}_1}{\log n}\Big)^{\log^3 n} \leqslant e^{-\widetilde{c}_1 \cdot \log^2 n} \cdot \widetilde{c} \cdot n^9 < n^{-7}.$$

So by Markov's inequality,

$$\mathbf{Pr}\Big[\ \widehat{\Gamma}_2^{t+(\log^3 n)\cdot b} < n\ \Big|\ \Gamma_2^t \leqslant \widetilde{c} \cdot n^9\ \Big] \geqslant 1 - n^{-8}$$

By union bound with Eq. (D.6),

$$\begin{aligned}\mathbf{Pr}\Big[\ \widehat{\Gamma}_2^{t+(\log^3 n)\cdot b} \geqslant n\ \Big] &= \mathbf{Pr}\Big[\ \widehat{\Gamma}_2^{t+(\log^3 n)\cdot b} \geqslant n\ \Big|\ \Gamma_2^t \leqslant \widetilde{c} \cdot n^9\ \Big] \cdot \mathbf{Pr}\big[\ \Gamma_2^t \leqslant \widetilde{c} \cdot n^9\ \big] + \mathbf{Pr}\big[\ \Gamma_2^t > \widetilde{c} \cdot n^9\ \big] \\ &< n^{-8} + n^{-8} = 1 - 2 \cdot n^{-8}.\end{aligned}$$

Due to the definition of Γ_2, at any step $t \geqslant 0$, deterministically $\Gamma_2^t \geqslant 2n$. So, we conclude that w.p. at least $1 - 2 \cdot n^{-8}$, we have that $\widehat{\Gamma}_2^{t+(\log^3 n)\cdot b} = 0$ or equivalently the event

$$\neg \bigcap_{s \in [0, \log^3 n]} \big\{\Gamma_2^{t+s\cdot b} > \widetilde{c} \cdot n\big\},$$

holds, which implies the conclusion. $\square$

D.5.2 Completing the proof of Lemma D.8

We are now ready to prove Lemma D.8, using a method of bounded differences with a bad event Theorem B.12 ([107, Theorem 3.3]).

Consider any process satisfying the conditions in Theorem 7.39. Let $\widetilde{c} := 2 \cdot \frac{8c}{\delta}$ where $c := c(\delta) > 0$ is the constant from Theorem 3.2. Then, for any step $t \geqslant 0$ being a multiple of b,

$$\mathbf{Pr}\left[\ \bigcap_{j \in [0, \log^3 n]} \big\{\Gamma_2^{t+j\cdot b} \leqslant \widetilde{c} \cdot n\big\}\ \right] \geqslant 1 - n^{-3}.$$

Proof. Our starting point is to apply Lemma D.13, which proves that there is at least one step $t + \rho \cdot b \in [t - b\log^3 n, t]$ with $\rho \in [-\log^3 n, 0]$ such that the potential Γ_2 is small,

$$\mathbf{Pr}\left[\ \bigcup_{\rho \in [-\log^3 n, 0]} \big\{\Gamma_2^{t+\rho\cdot b} \leqslant \widetilde{c} \cdot n\big\}\ \right] \geqslant 1 - 2 \cdot n^{-8}. \tag{D.7}$$

Note that if $t < b \cdot \log^3 n$, then deterministically $\Gamma_2^0 = 2n \leqslant \widetilde{c} \cdot n$ (which corresponds to $\rho = -t/b$).

We are now going to apply the concentration inequality Theorem B.12 to each of the batches starting at $t + \rho \cdot b, \ldots, t + (\log^3 n) \cdot b$ and show that the potential remains $\leqslant \widetilde{c} \cdot n$ at the last step of each batch. In particular, we will show that for any $\widetilde{r} \in [\rho, \log^3 n]$, for $r = t + b \cdot \widetilde{r}$,

$$\mathbf{Pr}\big[\ \Gamma_2^{r+b} > \widetilde{c} \cdot n\,\big|\ \mathfrak{F}^r, \Gamma_2^r \leqslant \widetilde{c} \cdot n\ \big] \leqslant 3 \cdot n^{-4}.$$

We will show this by applying Theorem B.12 for all steps of the batch $[r, r+b]$. We define the good event

$$\mathcal{G}_r := \mathcal{G}_r^{r+b} := \bigcap_{s \in [r,r+b]} \left(\left\{ \Gamma_1^s \leqslant \widetilde{c} \cdot n^{26} \right\} \cap \mathcal{H}^s \right),$$

and $\mathcal{B}_r := (\mathcal{G}_r)^c$ the bad event. Using a union bound over Lemma D.9 and Lemma D.12,

$$\mathbf{Pr}\left[\bigcap_{s \in [t-b\log^3 n, t+b\log^3 n]} \left(\left\{ \Gamma_1^s \leqslant \widetilde{c} \cdot n^{26} \right\} \cap \mathcal{H}^s \right) \right] \geqslant 1 - 2n^{-10}. \tag{D.8}$$

Consider any $u \in [r, r+b]$. Further, we define the slightly weaker good event,

$$\widetilde{\mathcal{G}}_r^u := \bigcap_{s \in [r,u]} \left(\left\{ \Gamma_1^s \leqslant 2\widetilde{c} \cdot n^{26} \right\} \cap \mathcal{H}^s \right)$$

and the "killed" potential,

$$\widehat{\Gamma}_r^u := \Gamma_2^u \cdot \mathbf{1}_{\widetilde{\mathcal{G}}_r^u}.$$

We will show that the sequence $\widehat{\Gamma}_r^r, \ldots, \widehat{\Gamma}_r^{r+b}$ is strongly difference-bounded by $\left(n^{5/4}, \frac{n}{b} \cdot n^{1/4}, 2 \cdot n^{-10}\right)$ (Definition B.11).

Let $\omega \in [n]^b$ be an allocation vector encoding the allocations made in $[r, r+b]$. Let ω' be an allocating vector resulting from ω by changing one arbitrary allocation. It follows that,

$$\left| \widehat{\Gamma}_r^{r+b}(\omega) - \widehat{\Gamma}_r^{r+b}(\omega') \right| \leqslant \max_{\widetilde{\omega}} \widehat{\Gamma}_r^{r+b}(\widetilde{\omega}) - \min_{\widetilde{\omega}} \widehat{\Gamma}_r^{r+b}(\widetilde{\omega}) \leqslant \max_{\widetilde{\omega} \in \widetilde{\mathcal{G}}_r^{r+b}} \Gamma_2^{r+b}(\widetilde{\omega}) - 0 \leqslant n^{5/4},$$

where in the last inequality we used Lemma D.10 (i) that for any $\widetilde{\omega} \in \widetilde{\mathcal{G}}_r^{r+b}$, we have $\widehat{\Gamma}_r^{r+b}(\widetilde{\omega}) \leqslant \Gamma_2^{r+b}(\widetilde{\omega}) \leqslant n^{5/4}$.

We will now derive a refined bound by additionally assuming that $\omega \in \mathcal{G}_r$. Then, for any $u \in [r, r+b]$,

$$\Gamma_1^{r+u}(\omega') \leqslant 2 \cdot \Gamma_1^{r+u}(\omega) \leqslant 2\widetilde{c} \cdot n^{26},$$

where the first inequality is by Lemma D.10 (iii). Hence $\omega' \in \widetilde{\mathcal{G}}_r^{r+b}$, so $\mathbf{1}_{\widetilde{\mathcal{G}}_r^{r+b}(\omega')} = 1$ and $\widehat{\Gamma}_r^{r+b}(\omega') = \Gamma_{2r}^{r+b}(\omega')$. Similarly, for $\omega \in \mathcal{G}_r \subseteq \widetilde{\mathcal{G}}_r^{r+b}$, we have $\widehat{\Gamma}_r^{r+b}(\omega) = \Gamma_{2r}^{r+b}(\omega)$ and by Lemma D.10 (ii),

$$\left| \widehat{\Gamma}_r^{r+b}(\omega) - \widehat{\Gamma}_r^{r+b}(\omega') \right| = \left| \Gamma_2^{r+b}(\omega) - \Gamma_2^{r+b}(\omega') \right| \leqslant \frac{n}{b} \cdot n^{1/4}.$$

Within a single batch all allocations are independent, so we apply Theorem B.12, choosing $\gamma_k := \frac{1}{b}$ and $N := b$, which states that for any $\lambda > 0$ and $\mu := \mathbf{E}\left[\widehat{\Gamma}_r^{r+b} > \mu + \lambda \mid \mathfrak{F}^r, \Gamma_2^r \leqslant \widetilde{c} \cdot n \right]$,

$$\mathbf{Pr}\left[\widehat{\Gamma}_r^{r+b} > \mu + \lambda \,\middle|\, \mathfrak{F}^r, \Gamma_2^r \leqslant \widetilde{c} \cdot n \right] \leqslant \exp\left(-\frac{\lambda^2}{2 \cdot \sum_{k=1}^{b} (\frac{n}{b} \cdot n^{1/4} + n^{5/4} \cdot \frac{1}{b})^2} \right) + 2 \cdot n^{-10} \cdot \sum_{k=1}^{b} b.$$

By Lemma D.11 (iv), we have $\mu \leqslant \mathbf{E}[\, \widehat{\Gamma}_r^{r+b} \mid \Gamma_2^r < \widetilde{c} \cdot n \,] \leqslant \mathbf{E}[\, \Gamma_2^{r+b} \mid \Gamma_2^r < \widetilde{c} \cdot n \,] \leqslant \widetilde{c} \cdot n - n/\log^2 n$. Hence, for $\lambda := n/\log^2 n$, since $n \leqslant b \leqslant n^3$, we have

$$\begin{aligned} \mathbf{Pr}\left[\widehat{\Gamma}_r^{r+b} > \widetilde{c} \cdot n \,\middle|\, \mathfrak{F}^r, \Gamma_2^r \leqslant \widetilde{c} \cdot n \right] &\leqslant \exp\left(-\frac{n^2/\log^4 n}{2 \cdot b \cdot (2 \cdot \frac{n}{b} \cdot n^{1/4})^2} \right) + 2n^{-10} \cdot b^2 \\ &\leqslant \exp\left(-\frac{b}{8 \cdot \log^4 n \cdot n^{1/2}} \right) + 2n^{-10} \cdot n^6 \leqslant 3 \cdot n^{-4}. \end{aligned}$$

Let $\mathcal{K}_\rho^{\widetilde{r}} := \mathcal{G}_\rho^{t+\widetilde{r}\cdot b} \cap \{\Gamma_2^{t+\rho\cdot b} \leqslant \widetilde{c}\cdot n\}$ for $\widetilde{r} \in [\rho, \log^3 n]$. For any $\widetilde{r} \geqslant \rho$, since $\mathcal{K}_\rho^{\widetilde{r}+1} \subseteq \mathcal{K}_\rho^{\widetilde{r}}$, we have

$$\mathbf{Pr}\left[\, \widehat{\Gamma}_r^{t+(\widetilde{r}+1)\cdot b} \cdot \mathbf{1}_{\mathcal{K}_\rho^{\widetilde{r}+1}} > \widetilde{c}\cdot n \,\middle|\, \mathfrak{F}^r, \widehat{\Gamma}_r^{t+\widetilde{r}\cdot b} \cdot \mathbf{1}_{\mathcal{K}_\rho^{\widetilde{r}}} \leqslant \widetilde{c}\cdot n \,\right] \leqslant 3\cdot n^{-4}. \tag{D.9}$$

By union bound of Eq. (D.7) and Eq. (D.8),

$$\begin{aligned}\mathbf{Pr}\left[\bigcup_{\rho\in[-\log^3 n]} \mathcal{K}_\rho^{\log^3 n}\right] &\geqslant \mathbf{Pr}\left[\mathcal{G}_{-\log^3 n}^{\log^3 n} \cap \bigcup_{\rho\in[-\log^3 n,0]} \left\{\Gamma_2^{t+\rho\cdot b} \leqslant \widetilde{c}\cdot n\right\}\right] \\ &\geqslant 1-2\cdot n^{-8} - 2\cdot n^{-10} \geqslant 1-3\cdot n^{-8}.\end{aligned} \tag{D.10}$$

Let $\mathcal{A} := \bigcap_{\widetilde{r}\in[0,\log^3 n]} \left\{\Gamma_2^{t+\widetilde{r}\cdot b} \leqslant \widetilde{c}\cdot n\right\}$ and $\mathcal{A}_\rho := \bigcap_{\widetilde{r}\in[\rho,\log^3 n]} \left\{\widehat{\Gamma}_r^{t+\widetilde{r}\cdot b}\cdot \mathbf{1}_{\mathcal{K}_\rho^{\widetilde{r}}} \leqslant \widetilde{c}\cdot n\right\}$. Then,

$$\begin{aligned}\mathbf{Pr}\left[\,\mathcal{A}_\rho \,\middle|\, \Gamma_2^{t+\rho\cdot b} \leqslant \widetilde{c}\cdot n\,\right] &\geqslant \prod_{\widetilde{r}\in[\rho,\log^3 n-1]} \mathbf{Pr}\Bigg[\bigcap_{\widetilde{s}\in[\rho+1,\widetilde{r}+1]} \left\{\widehat{\Gamma}_r^{t+\widetilde{s}\cdot b}\cdot \mathbf{1}_{\mathcal{K}_\rho^{\widetilde{s}}} \leqslant \widetilde{c}\cdot n\right\} \\ &\qquad\qquad \Bigg|\, \bigcap_{\widetilde{s}\in[\rho+1,\widetilde{r}-1]} \left\{\widehat{\Gamma}_r^{t+\widetilde{s}\cdot b}\cdot \mathbf{1}_{\mathcal{K}_\rho^{\widetilde{s}}} \leqslant \widetilde{c}\cdot n\right\}, \widehat{\Gamma}_r^{t+\widetilde{r}\cdot b}\cdot \mathbf{1}_{\mathcal{K}_\rho^{\widetilde{s}}} \leqslant \widetilde{c}\cdot n\Bigg] \\ &\geqslant \prod_{\widetilde{r}\in[\rho,\log^3 n-1]} \mathbf{Pr}\left[\,\widehat{\Gamma}_r^{\widetilde{r}+b}\cdot \mathbf{1}_{\mathcal{K}_\rho^{\widetilde{r}+1}} > \widetilde{c}\cdot n \,\middle|\, \mathfrak{F}^{t+\widetilde{r}\cdot b}, \widehat{\Gamma}_r^{t+\widetilde{r}\cdot b}\cdot \mathbf{1}_{\mathcal{K}_\rho^{\widetilde{r}}} \leqslant \widetilde{c}\cdot n\,\right] \\ &\geqslant (1-3n^{-4})^{2\log^3 n} \geqslant 1-6\cdot n^{-4}\cdot \log^3 n,\end{aligned}$$

where in the last inequality we have used Eq. (D.9) and the fact $\rho \geqslant -\log^3 n$. So,

$$\begin{aligned}\mathbf{Pr}\left[\,\mathcal{A}_\rho\,\right] &= \mathbf{Pr}\left[\,\mathcal{A}_\rho \,\middle|\, \Gamma_2^{t+\rho\cdot b} \leqslant \widetilde{c}\cdot n\,\right]\cdot \mathbf{Pr}\left[\,\Gamma_2^{t+\rho\cdot b} \leqslant \widetilde{c}\cdot n\,\right] + 1\cdot \mathbf{Pr}\left[\,\neg\left\{\Gamma_2^{t+\rho\cdot b} \leqslant \widetilde{c}\cdot n\right\}\,\right] \\ &\geqslant 1-6\cdot n^{-4}\cdot \log^3 n.\end{aligned} \tag{D.11}$$

Note that for any $\rho \in [-\log^3 n, 0]$, we have that $\mathcal{A}_\rho \cap \mathcal{K}_\rho^{\log^3 n} \subseteq \mathcal{A}$. Hence we conclude by the union bound of Eq. (D.10) and Eq. (D.11), that

$$\mathbf{Pr}\left[\,\mathcal{A}\,\right] \geqslant \mathbf{Pr}\left[\bigcup_{\rho\in[-\log^3 n,0]} \mathcal{K}_\rho^{\log^3 n} \cap \bigcap_{\rho\in[-\log^3 n,0]} \mathcal{A}_\rho\right] \geqslant 1-3\cdot n^{-8} - 6\cdot n^{-4}\cdot \log^6 n \geqslant 1-n^{-3}.$$

□

EXPERIMENTS

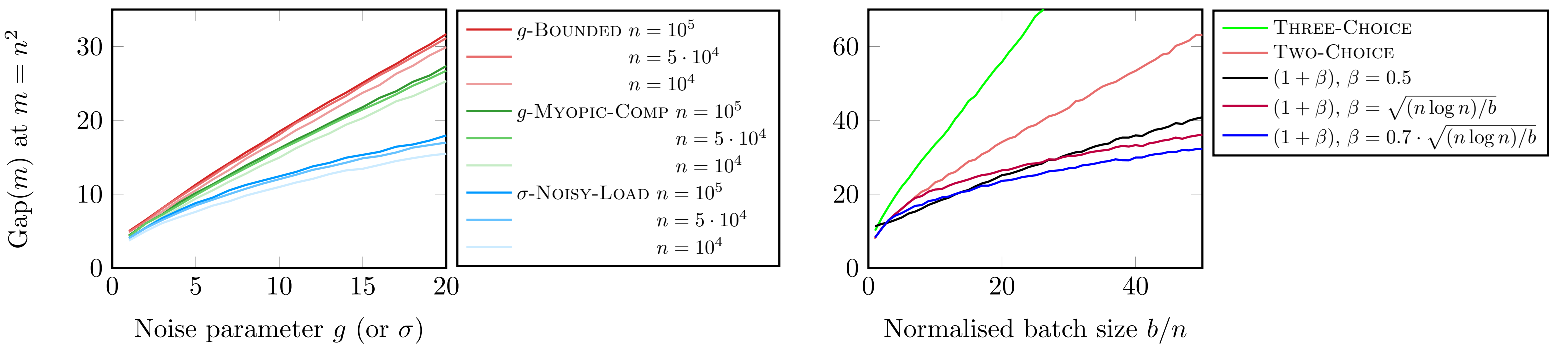


Figure E.1: The Gap(m) for $n \in \{10^4, 5 \cdot 10^4, 10^5\}$ and $m = n^2$ for noisy settings of TWO-CHOICE and for various processes in the b-BATCHED setting (25 repetitions).

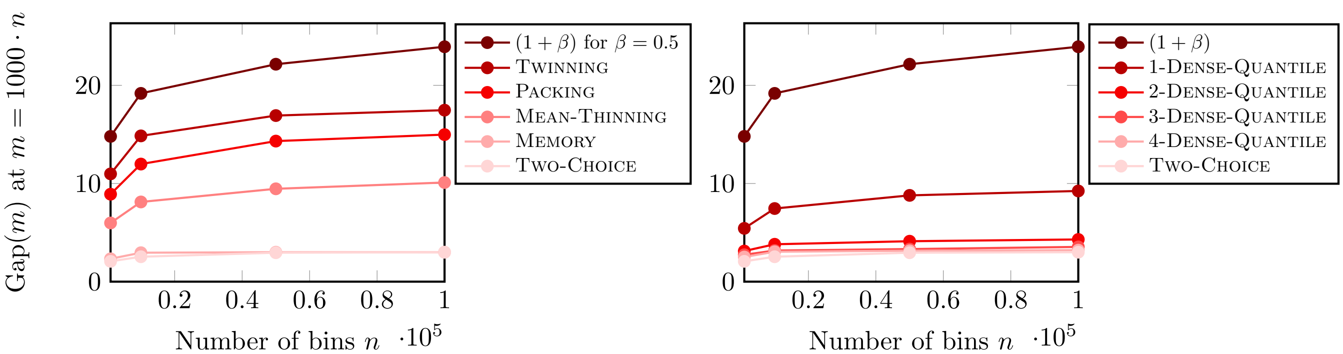


Figure E.2: The Gap(m) for $n \in \{5 \cdot 10^3, 10^4, 5 \cdot 10^4, 10^5\}$ and $m = 1000 \cdot n$ for various processes (25 repetitions).

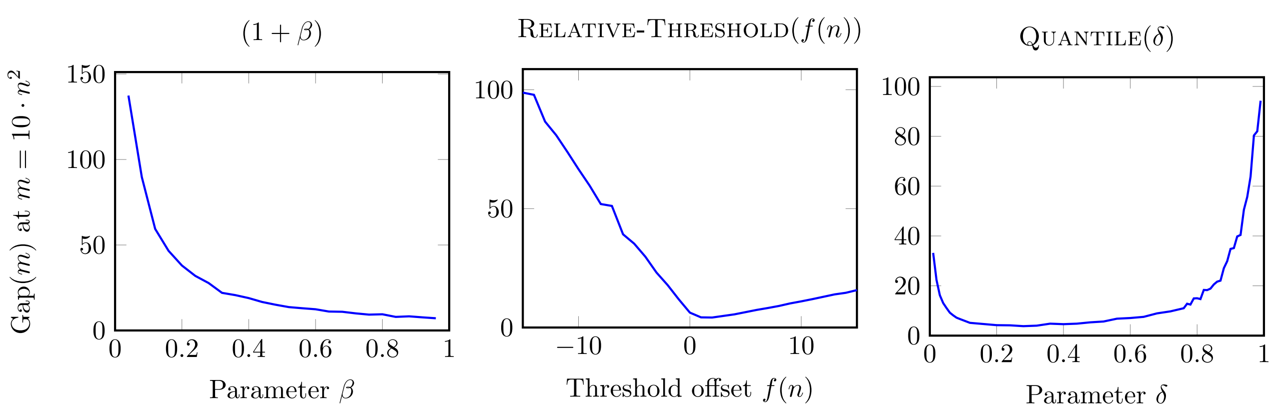


Figure E.3: The Gap(m) for $n = 10^3$ and $m = 10^7$ for different parameters of the $(1 + \beta)$, RELATIVE-THRESHOLD($f(n)$) and QUANTILE(δ) processes (15 repetitions).