arXiv:2302.04399v2 [cs.DM] 21 Apr 2023

Balanced Allocations in Batches: The Tower of Two Choices*

Dimitrios Los† and Thomas Sauerwald‡

Department of Computer Science & Technology, University of Cambridge, UK

Abstract

In the balanced allocation framework, the goal is to allocate m balls into n bins, so as to minimize the gap (difference of maximum to average load). The One-Choice process allocates each ball to a bin sampled independently and uniformly at random. The Two-Choice process allocates balls sequentially, and each ball is placed in the least loaded of two sampled bins. Finally, the $(1+\beta)$-process mixes these processes, meaning each ball is allocated using Two-Choice with probability $\beta \in (0,1)$, and using One-Choice otherwise.

Despite Two-Choice being optimal in the sequential setting, it has been observed in practice that it does not perform well in a parallel environment, where load information may be outdated. Following [5], we study such a parallel setting where balls are allocated in batches of size b, and balls within the same batch are allocated with the same strategy and based on the same load information.

For small batch sizes $b \in [n, n \log n]$, it was shown in [22] that Two-Choice achieves an asymptotically optimal gap among all allocation processes with two (or any constant number of) samples.

In this work, we focus on larger batch sizes $b \in [n \log n, n^3]$. It was proved in [20] that Two-Choice leads to a gap of $\Theta(b/n)$. As our main result, we prove that the gap reduces to $\mathcal{O}(\sqrt{(b/n) \cdot \log n})$, if one runs the $(1+\beta)$-process with an appropriately chosen β (in fact this result holds for a larger class of processes). This not only proves the phenomenon that Two-Choice is not the best (leading to the formation of "towers" over previously light bins), but also that mixing two processes (One-Choice and Two-Choice) leads to a process which achieves a gap that is asymptotically smaller than both. We also derive a matching lower bound of $\Omega(\sqrt{(b/n) \cdot \log n})$ for any allocation process, which demonstrates that the above $(1+\beta)$-process is asymptotically optimal.

Our analysis also works in the presence of randomly weighted balls, and also implies exponential tails for the number of bins above a certain load value.

*Full version of a paper appearing in SPAA 2023.

†dimitrios.los@cl.cam.ac.uk

‡thomas.sauerwald@cl.cam.ac.uk

Contents

1 Introduction **3**

2 Notation, Processes and Settings **7**
2.1 Basic Notation 7
2.2 Processes 7
2.3 Conditions on Probability Vectors 8
2.4 Weighted and Batched Settings 10

3 Warm-up: $\mathcal{O}(\sqrt{b/n} \cdot \log n)$ Gap **11**

4 Tight Bound: $\mathcal{O}(\sqrt{(b/n) \cdot \log n})$ Gap **16**
4.1 Step 1: Concentration of the Γ Potential 17
4.1.1 Preliminaries 17
4.1.2 Completing the Proof of Lemma 4.2 23
4.2 Step 2: Completing the Proof of Theorem 4.1 25

5 Lower Bounds on the Gap **27**

6 Experimental Results **29**

7 Conclusions **31**

Bibliography **32**

A Tools **34**
A.1 Auxiliary Probabilistic Claims 34
A.2 Concentration Inequalities 36

1 Introduction

Sequential balanced allocations. In the sequential balanced allocations framework, there are m tasks (balls) to be allocated into n servers (bins). It is well-known that allocating the balls into bins sampled uniformly at random (a.k.a. One-Choice) leads w.h.p.[1] to a maximum load of $\Theta(\log n/\log\log n)$ for $m = n$ and a gap (maximum load minus average load) of $\Theta\big(\sqrt{(m/n)\cdot\log n}\big)$ for $m \geqslant n\log n$.

An improvement over One-Choice is the d-Choice process [3, 6, 15], where each ball is allocated to the least loaded of d bins sampled uniformly at random. For any $m \geqslant n$, this process achieves w.h.p. an $\log_d\log n + \Theta(1)$ gap, i.e., a gap that does not depend on m. For $d = 2$, this great improvement is known as "power-of-two-choices" (see also surveys [28, 38] for more details). Despite the simplistic nature of the balanced allocation framework, the Two-Choice process has had a significant impact on practical applications such as load balancing and distributed storage systems, which was also acknowledged by the "*ACM Paris Kanellakis Theory and Practice Award 2020*" [2] (see also *Applications* below).

Several variants of Two-Choice have been studied. Of particular importance to this work is the $(1+\beta)$-process, where each ball is allocated using Two-Choice with probability $\beta \in (0, 1]$ and One-Choice otherwise. Mitzenmacher [26, Section 4.4.1] introduced this process as a model of Two-Choice with erroneous comparisons. Peres, Talwar and Wieder [33] showed that for $\beta := \beta(n) \ll 1$, it achieves w.h.p. a $\Theta((\log n)/\beta)$ gap (see also [20]), which becomes worse for smaller β, but still remains independent of m. The $(1+\beta)$-process has also been applied to the analysis of Two-Choice in the popular *graphical setting* [4, 16, 33], where bins are organized as vertices in a graph, and each ball is allocated to the lesser loaded of two adjacent vertices of an edge sampled uniformly at random.

Another variant of Two-Choice that has received some attention recently is the family of Two-Thinning processes [12, 13], where the ball is allocated to the second sample only if the first one does not meet a certain criterion, e.g., based on a threshold on its load or a quantile on its rank.

It should be noted that the analyses of all these processes strongly rely on the fact that the load information of each bin is updated after each allocation. In effect this means balls can only be allocated sequentially, which is a downside in distributed and parallel environments.

Outdated information settings. In this work, we demonstrate that in outdated information settings by choosing an appropriately small β, $(1+\beta)$ achieves the asymptotically optimal gap among a large class of processes, including not only Two-Choice (and One-Choice), but even adaptive processes that may allocate with a different scheme after each batch. This confirms earlier empirical observations that the performance of the Two-Choice process deteriorates under outdated information and delays [8, 14, 27, 31, 37].

Berenbrink, Czumaj, Englert, Friedetzky and Nagel [5] introduced the b-Batched setting where balls are allocated in batches of size b. That means, in every batch the b balls are allocated *in parallel*, as the decision where to allocate the ball only depends on the load configuration before that batch of balls arrived. For $b = n$, they proved that Two-Choice achieves w.h.p. an $\mathcal{O}(\log n)$ gap. This bound was recently improved to $\Theta(\log n/\log\log n)$ in [22], and in the same work, it was shown that Two-Choice has a gap that matches the maximum load of One-Choice for b balls, for any batch size $b \in [n\cdot e^{-\log^{\Theta(1)} n}, n\log n]$, and so it is asymptotically optimal. In contrast, for $b \geqslant n\log n$, Two-Choice (and a family of other processes) have w.h.p. a $\Theta(b/n)$ gap [20], a bound which was shown to hold even in the presence of weights and on some graphs. This analysis also demonstrates that increasing d in the d-Choice process, does not always improve the gap, which is in sharp contrast to the sequential setting. In [22], a more powerful setting, τ-Delay was studied for the Two-Choice process, where an adversary

[1] In general, with high probability refers to probability of at least $1 - n^{-c}$ for some constant $c > 0$.

can choose to report for each of the bins any load from the last τ steps. For $b = \tau$, b-Batched is a special instance of τ-Delay and for any $\tau \leqslant n \log n$, the same asymptotic bounds where shown to hold.

Outdated information settings have also been studied in the queuing setting [1, 14, 18, 27, 37]. In particular, Mitzenmacher [27] studied the corresponding version of the b-Batched setting, called the *bulletin board model with periodic updates*, showing that some processes requiring centralized coordination can outperform Two-Choice, but no explicit rigorous bounds were proven. This shortcoming of Two-Choice was characterized as *herd behavior*, meaning that some of the initially lighter bins receive disproportionately many balls, turning them into heavy bins. In another empirically study, Dahlin [8] also observed the herd behavior and suggested similar centralized strategies to improve upon d-Choice. Regarding identifying optimal processes, Whitt [37] remarks:

> *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 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.*

Applications. Recently, several *distributed low-latency schedulers*, including Sparrow [31], Eagle [9], Hawk [10], Peacock [17], Pigeon [36] and Tarcil [11], have used variants of the Two-Choice process. In [31], with regards to the implementation of Sparrow, the authors state:

> *The power of two choices suffers from two remaining performance problems: first, server queue length is a poor indicator of wait time, and second, due to messaging delays, multiple schedulers sampling in parallel may experience race conditions.*

Similar observations have been made in the context of *distributed stream processing* [29, 30] and *load balancers* [24]. These studies support that batch sizes $b = \Omega(n \log n)$ for which Two-Choice is no longer optimal are relevant to real-world applications.

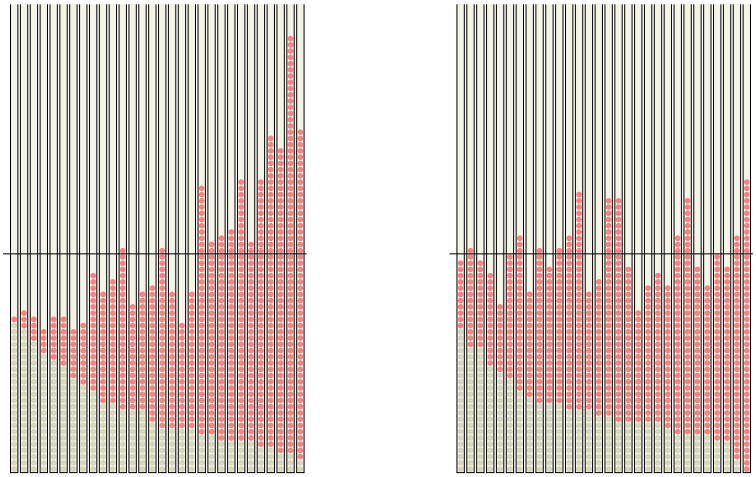
Weighted settings. Several works study balanced allocation processes with *weights* [7, 20, 33, 35]. We will be focusing on weights sampled independently from probability distributions with bounded moment generating functions as in [20] and [33], which includes the geometric, exponential and Poisson distributions.

Our results. In this work, we prove that a family of processes satisfying a mild technical condition achieve the asymptotically optimal gap[2] of $\mathcal{O}\big(\sqrt{(b/n) \cdot \log n}\big)$ in the weighted b-Batched setting for $b \in [2n \log n, n^3]$, leading to roughly a quadratic improvement over the gap of the Two-Choice process. This family of processes includes the $(1+\beta)$-process, which is a process that can be easily implemented in a *decentralized manner*, and demonstrates that by setting $\beta = \sqrt{(n/b) \cdot \log n}$ we attain this asymptotically optimal gap.

We also provide lower bounds establishing the tightness of our upper bounds. Interestingly, the lower bound of $\Omega(\sqrt{(b/n) \cdot \log n})$ applies to a much more powerful class of allocation processes, where the allocation rule is arbitrarily tailored at the beginning of the batch.

The intuition for these optimal processes relates to the herd behavior observed in [27] and [8]. For the d-Choice process, the maximum probability of allocating to a bin is $\max_{i \in [n]} p_i \approx d/n$. This means that, for example, in Two-Choice in a batch of b balls there are some bins that receive $\approx 2b/n$ balls and so a gap of $\approx b/n$ arises. This becomes worse as d grows. To avoid this,

[2]By *optimal* we mean over all processes that choose a probability allocation vector p, where p_i gives the probability to allocate to the i-th heaviest bin, at the beginning of the batch and this vector remains the same throughout the entire batch.



Two-Choice $(1+\beta)$-process

Figure 1.1: The $b = 750$ balls of the latest batch shown in red allocated over the $n = 35$ bins (**left**) for Two-Choice and (**right**) $(1+\beta)$ with $\beta = 1/2$. Observe that Two-Choice allocates more aggressively on the bins that are lightly loaded at the beginning of the batch, while $(1+\beta)$ spreads the allocations more evenly.

we will investigate processes where $\max_{i\in[n]} p_i = (1 + o(1))/n$, which means that in expectation no bin receives too many balls in any particular batch. For example, the $(1 + \beta)$-process has $\max_{i\in[n]} p_i \approx (1+\beta)/n$, which means that this mixing of One-Choice steps with Two-Choice steps circumvents the herd behavior. See Fig. 1.1 for a visualization of how $(1 + \beta)$ achieves a more balanced distribution than Two-Choice over one batch, and Fig. 1.2 for how the gaps of different processes are getting worse with larger $\max_{i\in[n]} p_i$. The asymptotic gap bounds of the One-Choice, Two-Choice and $(1+\beta)$ processes in the b-Batched setting are summarized in Table 1.3. Our results also imply bounds for the shape of the load vector (see Remark 4.3). Our analysis also applies in the presence of randomly weighted balls, and also implies exponential tails for the number of bins above a certain load value.

Our techniques. Our techniques build on and refine those in [20], making use of the hyperbolic cosine potential function [33] and variants. More specifically, a slightly weaker version of our tight upper bound is based on [20, Theorem 3.1] and a refinement of [20, Lemma 4.1]. For our tight gap bound, our approach uses an interplay between two hyperbolic cosine potential functions to prove concentration and then an exponential potential with a larger smoothing parameter to deduce the refined gap. A similar method was used in [20, Section 5], but one crucial novelty here is that we consider allocation processes whose probability allocation vector have a small ℓ_∞ distance from the uniform distribution. We believe that relating and comparing different allocation processes based on their ℓ_∞ distance (or other metrics) could be a promising avenue for future work. This can be also seen as a natural relaxation of the *majorization technique*, which has been the dominant tool to relate different allocation processes [21, 33].

Organization. In Section 2, we introduce the basic notation for balanced allocations, and define the processes and settings that we will be working with. In particular, in Section 2.3 we define general conditions on the probability allocation vector used by the processes, under which our upper bounds on the gap apply. In Section 3, we prove the $\mathcal{O}\big(\sqrt{b/n} \cdot \log n\big)$ bound on the gap for a family of processes in the weighted b-Batched setting. In Section 4, we perform a refined analysis and improve this bound to $\mathcal{O}\big(\sqrt{(b/n) \cdot \log n}\big)$. In Section 5, we show that this achieved gap is asymptotically optimal, and in Section 6, we present some empirical results on the gap of some specific processes. Finally, in Section 7, we summarize the results and conclude

with some open problems.

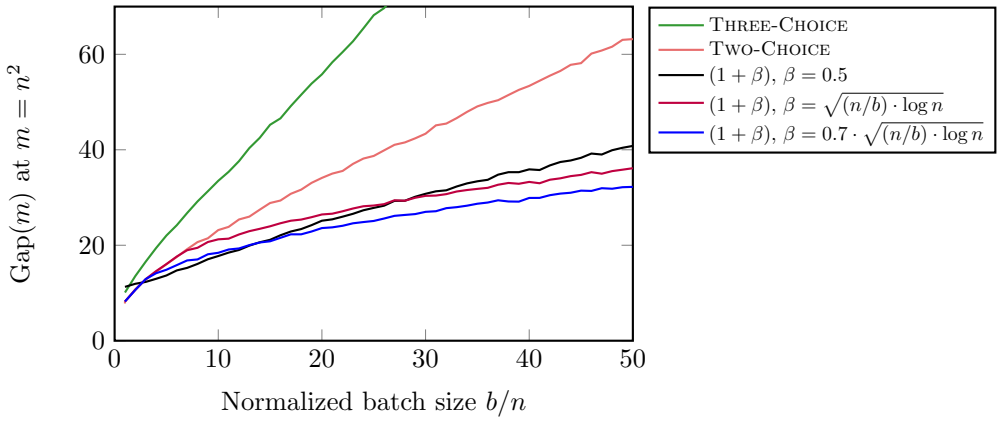


Figure 1.2: In the b-Batched setting for large batch size b, the gaps achieved by the processes are ordered by their maximum entry in the probability allocation vector p: Three-Choice with $\max_{i\in[n]} p_i \approx \frac{3}{n}$, Two-Choice with $\max_{i\in[n]} p_i \approx \frac{2}{n}$, $(1+\beta)$ with $\max_{i\in[n]} p_i \approx \frac{1+\beta}{n}$ for $\beta = 0.5$, $\beta = \sqrt{(n/b) \cdot \log n}$ and $\beta = \sqrt{(n/b) \cdot \log n}$. See Fig. 6.3 for full details of the experiment.

Process	Gap in Sequential Setting	Gap in b-Batched Setting	Batch Size
One-Choice	$\Theta\left(\sqrt{(m/n) \cdot \log n}\right)$ [34]	$\Theta\left(\sqrt{(m/n) \cdot \log n}\right)$ [34]	$b \in \mathbb{N}$
Two-Choice	$\log_2 \log n + \mathcal{O}(1)$ [3, 6]	$\Theta(\log \log n)$ [22]	$b \in \Theta(1)$
		$\Theta\left(\frac{\log n}{\log((4n/b)\log n)}\right)$ [22]	$b \in [n \cdot e^{-\log^{\Theta(1)} n}, n \log n]$
		$\Theta\left(b/n\right)$ [20]	$b \in [n \log n, n^3]$
$(1+\beta)$, $\beta = \Theta(1)$	$\Theta(\log n)$ [33]	$\Theta\left(b/n + \log n\right)$ [20]	$b \in [n, n^3]$
$(1+\beta)$-process, $\beta = \sqrt{(n/b) \cdot \log n}$	$\Theta((\log n)/\beta)$ [33]	$\Omega\left(\sqrt{(b/n) \cdot \log n}\right)$ Thm 5.2 $\mathcal{O}\left(\sqrt{(b/n) \cdot \log n}\right)$ Thm 4.1	$b \in [2n \log n, n^3]$
$(1+\beta)$-process, $\beta = \sqrt{n/b}$	$\Theta((\log n)/\beta)$ [33]	$\Omega\left(\sqrt{(b/n) \cdot \log n}\right)$ Thm 5.2 $\mathcal{O}\left(\sqrt{b/n} \cdot \log n\right)$ Cor 3.2	$b \geqslant n \log n$

Table 1.3: Overview of the gap bounds in previous works (rows in Gray) and the gap bounds derived in this work (rows in Green). All gap bounds hold with probability at least $1 - o(1)$. Lower bounds hold for sufficiently large enough m. For the sake of simplicity, we focus on the setting with unit weights and only list results for $(1+\beta)$. Among all these processes, One-Choice produces the worst gap in both settings, even though the gap does not change between the b-Batched and sequential setting. For Two-Choice, the gap becomes b/n in the b-Batched setting with $b = \Omega(n \log n)$, whereas for $(1+\beta)$ the gap is improved to $\sqrt{(b/n) \cdot \log n}$ (for a suitable β).

2 Notation, Processes and Settings

In this section, we introduce notation, processes and settings used throughout this work.

2.1 Basic Notation

We consider the allocation of m balls into n bins, which are labeled $[n] := \{1, 2, \ldots, n\}$. For the moment, the m balls are unweighted (or equivalently, all balls have weight 1). For any step $t \geqslant 0$, x^t is the n-dimensional *load vector*, where x_i^t is the number of balls allocated to bin i in the first t allocations. In particular, $x_i^0 = 0$ for every $i \in [n]$. Finally, the *gap* is defined as

$$\mathrm{Gap}(t) = \max_{i \in [n]} x_i^t - \frac{t}{n}.$$

It will also be convenient to sort the load vector x. To this end, let $\widetilde{x}^t := x^t - \frac{t}{n}$. Then, relabel the bins such that y^t is a permutation of $\widetilde{x}^t$ and $y_1^t \geqslant y_2^t \geqslant \cdots \geqslant y_n^t$. Note that $\sum_{i \in [n]} y_i^t = 0$ and $\mathrm{Gap}(t) = y_1^t$. We call a bin $i \in [n]$ *overloaded*, if $y_i^t \geqslant 0$ and *underloaded* otherwise.

A *probability vector* $p \in \mathbb{R}^n$ is any vector satisfying $\sum_{i=1}^n p_i = 1$ and $p_i \in [0, 1]$ for $i \in [n]$. Following [33], many allocation processes can be described by a time-invariant *probability allocation vector* p^t, which is the probability vector with p_i^t being probability of allocating a ball to the i-th heaviest bin.

By $\mathfrak{F}^t$ we denote the filtration of the process until step t, which in particular reveals the load vector x^t.

2.2 Processes

We start with a formal description of the One-Choice process.

> One-Choice Process:
> Iteration: For each $t \geqslant 0$, sample one bin i, independently and uniformly at random. Then update:
> $$x_i^{t+1} = x_i^t + 1.$$

We continue with a formal description of the Two-Choice process.

> Two-Choice Process:
> Iteration: For each $t \geqslant 0$, sample two bins i_1 and i_2, independently and uniformly at random. Let $i \in \{i_1, i_2\}$ be such that $x_i^t = \min\{x_{i_1}^t, x_{i_2}^t\}$, breaking ties randomly. Then update:
> $$x_i^{t+1} = x_i^t + 1.$$

It is immediate that the probability allocation vector of Two-Choice is

$$p_i = \frac{2i - 1}{n^2}, \qquad \text{for all } i \in [n].$$

Following [33], we recall the definition of the $(1 + \beta)$-process which interpolates between One-Choice and Two-Choice:

$(1+\beta)$ Process:
Parameter: A *mixing factor* $\beta \in (0,1]$.
Iteration: For each $t \geqslant 0$, sample two bins i_1 and i_2, independently and uniformly at random. Let $i \in \{i_1, i_2\}$ be such that $x_i^t = \min\{x_{i_1}^t, x_{i_2}^t\}$, breaking ties randomly. Then update:

$$\begin{cases} x_i^{t+1} = x_i^t + 1 & \text{with probability } \beta, \\ x_{i_1}^{t+1} = x_{i_1}^t + 1 & \text{otherwise.} \end{cases}$$

In other words at each step, the $(1+\beta)$-process allocates the ball following the Two-Choice rule with probability β, and otherwise allocates the ball following the One-Choice rule. Therefore, its probability allocation vector is given by

$$p_i = (1-\beta) \cdot \frac{1}{n} + \beta \cdot \frac{2i-1}{n^2}, \qquad \text{for all } i \in [n].$$

Recall that in [33] (and [20]), it was shown that $\mathrm{Gap}(m) = \mathcal{O}\big(\frac{\log n}{\beta}\big)$ for any $m \geqslant n$ and $\beta \in (0,1]$; so in particular, this gap (bound) does not grow with m.

The next process is another relaxation of Two-Choice.

Quantile(δ) Process:
Parameter: A *quantile* $\delta \in \{1/n, 2/n, \ldots, 1\}$.
Iteration: For each $t \geqslant 0$, sample two bins i_1 and i_2, independently and uniformly at random. Then update:

$$\begin{cases} x_{i_2}^{t+1} = x_{i_2}^t + 1 & \text{if } i_1 \text{ is among the } \delta n \text{ heaviest bins,} \\ x_{i_1}^{t+1} = x_{i_1}^t + 1 & \text{otherwise.} \end{cases}$$

Note that the Quantile(δ) processes can be implemented as a two-phase procedure: First probe the bin i_1 and place the ball there if i_1 is not among the δn heaviest bins. Otherwise, take a second sample i_2 and place the ball there. Since we only need to know whether a bin's rank is above or below a value, the response by a bin can be encoded as a single bit (at the cost of knowing the rank of each bin). The probability allocation vector of Quantile(δ) is given by:

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

2.3 Conditions on Probability Vectors

In [20], the weighted b-Batched setting was analyzed for probability allocation vectors satisfying the following two conditions. The first condition says that the process has a small ε/n bias to place away from overloaded and towards underloaded bins; and the second condition says that no bin has too high probability of being allocated.

- **Condition** $\mathcal{C}_1$: There exist constant quantile[3] $\delta \in (0,1)$ and (not necessarily constant) $\varepsilon \in (0,1)$, such that for any $1 \leqslant k \leqslant \delta n$,

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

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

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

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

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

In the same paper [20, Proposition 7.4] it was shown that any process with $\max_{i\in[n]} p_i \geqslant \frac{1+\varepsilon}{n}$ for $\varepsilon = \Omega(1)$ also has $\mathrm{Gap}(m) = \Omega(b/n)$ for any $b = \Omega(n \log n)$. Therefore, to improve on this asymptotic gao bound, we have to consider processes with $\max_{i\in[n]} p_i = \frac{1+o(1)}{n}$. In our analysis in Sections 3 and 4 we will make use of the following condition based on the ℓ_∞-distance between the probability allocation vector p and the uniform distribution (i.e., One-Choice):

- **Condition** $\mathcal{C}_3$: There exists a $C > 1$, such that

$$\max_{i\in[n]} \left| p_i - \frac{1}{n} \right| \leqslant \frac{C-1}{n}.$$

Note that this condition implies condition $\mathcal{C}_2$ for the same $C > 1$, but unlike $\mathcal{C}_2$ it imposes both an upper and a lower bound on the p_i's. It is easy to verify that $(1+\beta)$-process satisfies all three conditions.

Lemma 2.1. *For any $\beta \in (0, 1]$, the $(1+\beta)$-process satisfies condition $\mathcal{C}_1$ with $\delta = \frac{1}{4}$ and $\varepsilon = \frac{\beta}{2}$, condition $\mathcal{C}_2$ with $C = 1 + \beta$ and condition $\mathcal{C}_3$ with $C = 1 + \beta$.*

Proof. Recall that for the $(1+\beta)$-process, the probability allocation vector satisfies

$$p_i = (1-\beta) \cdot \frac{1}{n} + \beta \cdot \frac{2i-1}{n^2}, \quad \text{for all } i \in [n].$$

We will first show that $\mathcal{C}_1$ holds with $\delta = 1/4$ and $\varepsilon = \beta/2$. For any $1 \leqslant k \leqslant \delta n$, since p is non-decreasing the prefix sums satisfy

$$\sum_{i=1}^{k} p_i \leqslant p_k \cdot k \leqslant p_{\delta n} \cdot k \leqslant ((1-\beta) + \beta \cdot (2\delta)) \cdot \frac{k}{n} = \left(1 - \frac{\beta}{2}\right) \cdot \frac{k}{n},$$

Similarly, for any $\delta n + 1 \leqslant k \leqslant n$, the suffix sums satisfy

$$\begin{aligned}
\sum_{i=k}^{n} p_i &\overset{(a)}{=} \frac{n-k+1}{n} \cdot (1-\beta) + \frac{\beta}{n^2} \cdot (n^2 - (k-1)^2) \\
&= \frac{n-k+1}{n} \cdot (1-\beta) + \frac{\beta}{n^2} \cdot (n-k+1) \cdot (n+k-1) \\
&= \left(1 + \frac{\beta}{n} \cdot (k-1)\right) \cdot \frac{n-k+1}{n} \\
&\overset{(b)}{\geqslant} (1 + \beta \cdot \delta) \cdot \frac{n-k+1}{n} \\
&\overset{(c)}{\geqslant} \left(1 + \varepsilon \cdot \frac{\delta}{1-\delta}\right) \cdot \frac{n-k+1}{n},
\end{aligned}$$

using in (a) that $\sum_{i=1}^{u}(2i-1) = u^2$, in (b) that $k \geqslant \delta n + 1$ and in (c) that $\delta = 1/4$ and $\varepsilon = \beta/2$.

Condition $\mathcal{C}_3$ for $C = 1 + \beta$ (and hence $\mathcal{C}_2$ as well) is verified as follows. As p_i is increasing in $i \in [n]$,

$$\max_{i\in[n]} \left| p_i - \frac{1}{n} \right| = \max\left\{ \frac{1}{n} - p_1, p_n - \frac{1}{n} \right\} = \frac{\beta}{n} - \frac{\beta}{n^2} \leqslant \frac{\beta}{n}. \qquad \square$$

Note that in contrast to Two-Choice which satisfies $\mathcal{C}_3$ for $C = 2 - \frac{1}{n}$, by choosing β small enough we can make the probability allocation vector arbitrarily close to uniform.

We also note that for any process $\mathcal{P}$ satisfying condition $\mathcal{C}_3$ for some $C > 1$, we can define a process $\mathcal{P}'$ satisfying condition $\mathcal{C}_3$ for $C' \in (1, C)$ by mixing the probability allocation vector of $\mathcal{P}$ with that of One-Choice with probability $\eta = \frac{C'-1}{C-1}$.

For instance, the Quantile(1/2) process satisfies condition $\mathcal{C}_3$ for any $C = 1 + 1/2$ (since $\min_{i\in[n]} p_i = \frac{1}{2n}$ and $\max_{i\in[n]} p_i = \frac{3}{2n}$). Therefore, mixing Quantile(1/2) with One-Choice with probability $\eta \in [0, 1]$, gives the following probability allocation vector satisfying condition $\mathcal{C}_3$ for $C = 1 + \eta/2$,

$$p_i = \begin{cases} \frac{1}{n} \cdot (1-\eta) + \frac{1}{2n} \cdot \eta = \frac{1}{n} - \frac{\eta}{2n} & \text{if } i \leqslant \frac{1}{2}n, \\ \frac{1}{n} \cdot (1-\eta) + \frac{3}{2n} \cdot \eta = \frac{1}{n} + \frac{\eta}{2n} & \text{otherwise.} \end{cases}$$

Observation 2.2. *The process obtained by mixing* Quantile(1/2) *with* One-Choice *satisfies condition $\mathcal{C}_1$ with $\delta = 1/2$ and $\varepsilon = \eta/2$, condition $\mathcal{C}_2$ with $C = 1 + \eta/2$ and condition $\mathcal{C}_3$ with $C = 1 + \eta/2$.*

2.4 Weighted and Batched Settings

As in [20], we now extend the definitions of Section 2.1 and Section 2.2 to *weighted balls* and later to the *batched setting*. To this end, let $w^t \geqslant 0$ be the weight of the t-th ball to be allocated for $t \geqslant 1$. By W^t we denote the total weights of all balls allocated after the first $t \geqslant 0$ allocations, so $W^t := \sum_{i=1}^n x_i^t = \sum_{s=1}^t w^s$. The normalized loads are $\widetilde{x}_i^t := x_i^t - \frac{W^t}{n}$, and with y_i^t being again the decreasingly sorted, normalized load vector, we have $\mathrm{Gap}(t) = y_1^t$.

The weight of each ball will be drawn independently from a fixed distribution $\mathcal{W}$ over $[0, \infty)$. Following [33], 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 above conditions (after scaling) are the geometric, exponential, binomial and Poisson distributions.

In the analysis we will be using the following property (see also [33]) and refer to these distributions as Finite-MGF(ζ) (or Finite-MGF(S)):

Lemma 2.3 ([20, Lemma 2.4]). *There exists $S := S(\zeta) \geqslant \max\{1, 1/\zeta\}$, such that for any $\gamma \in (0, \min\{\zeta/2, 1\})$ and any $\kappa \in [-1, 1]$,*

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

We will now describe the allocation of weighted balls into bins using a batch size of $b \geqslant n$. For the sake of concreteness, let us first describe the b-Batched setting if the allocation is done using Two-Choice. For a given batch size consisting of b consecutive balls, each ball of the batch performs the following. First, it samples two bins i_1 and i_2 independently and uniformly at random, and compares the load the two bins had at the beginning of the batch (let us denote the bin which has less load by $i_{\min}$). Secondly, a weight is sampled from the distribution $\mathcal{W}$. Then a weighted ball is added to bin $i_{\min}$. Recall that since the load information is only updated at the beginning of the batch, all allocations of the b balls within the same batch can be performed in parallel.

In the following, we will use a more general framework, where the process of sampling (one or more) bins and then deciding where to allocate the ball to is described by a probability allocation vector p over the n bins (Section 2.1). Also for the analysis, it will be convenient to focus on the normalized and sorted load vector y, which is why the definition below is based on y rather than the actual load vector x.

b-Batched Setting with Weights
Parameters: Batch size $b \geqslant n$, probability allocation vector p, weight distribution $\mathcal{W}$.
Iteration: For each $t = 0 \cdot b, 1 \cdot b, 2 \cdot b, \ldots$:

1. Sample b bins $i_1, i_2, \ldots, i_b$ from $[n]$ following p.
2. Sample b weights $w^{t+1}, w^{t+2}, \ldots, w^{t+b}$ from $\mathcal{W}$.
3. Update for each bin $i \in [n]$,
$$z_i^{t+b} = y_i^t + \sum_{j=1}^{b} w^{t+j} \cdot \mathbf{1}_{i_j=i} - \frac{1}{n} \cdot \sum_{j=1}^{b} w^{t+j}.$$
4. Let y^{t+b} be the vector z^{t+b}, sorted decreasingly.

We also look at the version of the processes that perform random tie-breaking between bins of the same load. For $b = 1$, this makes no observable difference to the process, but for multiple steps, this effectively averages out the probability over (possibly) multiple bins that have the same load. This would, for instance, correspond to Two-Choice, randomly deciding between the two bins if they have the same load. In particular, if p is the original probability allocation vector, then the one with random tie-breaking is $\widetilde{p}(y^t)$ (for t being the beginning of the batch), where

$$\widetilde{p}_i(y^t) := \frac{1}{|\{j \in [n] : y_j^t = y_i^t\}|} \cdot \sum_{j \in [n] : y_j^t = y_i^t} p_j, \quad \text{for all} i \in [n]. \tag{2.1}$$

b-Batched Setting with Weights and Random Tie-Breaking
Parameters: Batch size $b \geqslant n$, probability allocation vector p, weight distribution $\mathcal{W}$.
Iteration: For each $t = 0 \cdot b, 1 \cdot b, 2 \cdot b, \ldots$:

1. Let $\widetilde{p} := \widetilde{p}(y^t)$ be the probability allocation vector accounting for random tie-breaking.
2. Sample b bins $i_1, i_2, \ldots, i_b$ from $[n]$ following $\widetilde{p}$.
3. Sample b weights $w^{t+1}, w^{t+2}, \ldots, w^{t+b}$ from $\mathcal{W}$.
4. Update for each bin $i \in [n]$,
$$z_i^{t+b} = y_i^t + \sum_{j=1}^{b} w^{t+j} \cdot \mathbf{1}_{i_j=i} - \frac{1}{n} \cdot \sum_{j=1}^{b} w^{t+j}.$$
5. Let y^{t+b} be the vector z^{t+b}, sorted decreasingly.

3 Warm-up: $\mathcal{O}(\sqrt{b/n} \cdot \log n)$ Gap

In this section, we will refine the analysis of [20, Section 4] to prove an $\mathcal{O}(\sqrt{b/n} \cdot \log n)$ bound on the gap for a family of processes. This will also be used as a starting point for the analysis in Section 4 to obtain the tighter bound. The main theorem that we prove is the following.

Theorem 3.1. *Consider any allocation process with probability allocation vector p^t satisfying conditions $\mathcal{C}_1$ for constant $\delta \in (0, 1)$ and (not necessarily constant) $\varepsilon \in (0, 1)$ as well as 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* $S \geqslant 1$ *and a batch size* $b \geqslant \frac{2CS}{(C-1)^2} \cdot n$. *Then, there exists a constant* $k := k(\delta) > 0$, *such that for any step* $m \geqslant 0$ *being a multiple of* b,

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

Recall that by Lemma 2.1, the $(1+\beta)$-process satisfies condition $\mathcal{C}_1$ with $\varepsilon = \frac{\beta}{2}$ and $\delta = \frac{1}{4}$, and conditions $\mathcal{C}_2$ and $\mathcal{C}_3$ with $C = 1 + \beta$.

In particular, by choosing $\beta = \Theta\big(\sqrt{n/b}\big)$ we get a process that is asymptotically better than TWO-CHOICE and which is within just a $\sqrt{\log n}$ multiplicative factor from the optimal gap bound proven for unit weights in Section 5.

Corollary 3.2. *Let* $b \geqslant n \log n$ *and consider the weighted* b-BATCHED *setting with weights from a* FINITE-MGF(S) *distribution with* $S \in [1, b/4n]$. *Then, there exists a constant* $k > 0$ *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,

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

The analysis is based on the *hyperbolic cosine potential* which is defined for smoothing parameter $\gamma > 0$ as

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

We also decompose Γ^t by defining

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

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 $\Delta\Gamma_i^{t+1} := \Gamma_i^{t+1} - \Gamma_i^t$.

We will make use of the following drift theorem shown in [20]. Note that in statement of the theorem, *rounds* could consist of multiple single-step allocations and in that case p^t is not necessarily the probability allocation vector, but it could be a probability vector giving an estimate for the "average number of balls" allocated to a bin.

Theorem 3.3 (cf. [20, Theorem 3.1]). *Consider any allocation process* $\mathcal{P}$ *and a probability vector* p^t *satisfying condition* $\mathcal{C}_1$ *for some constant* $\delta \in (0,1)$ *and some* $\varepsilon \in (0,1)$ *at every round* $t \geqslant 0$. *Further assume that there exist* $K > 0$, $\gamma \in \big(0, \min\big\{1, \frac{\varepsilon\delta}{8K}\big\}\big]$ *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,*

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

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* $t \geqslant 0$,

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

and

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

Now we will show that any process satisfying condition $\mathcal{C}_3$, also satisfies the preconditions of Theorem 3.3 for the expected change of the potential functions Φ and Ψ over one batch.

Lemma 3.4. *Consider any allocation process with probability allocation vector p^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 $\gamma \in (0, \frac{n}{2(C-1)\cdot b}]$ and any step $t \geqslant 0$ being a multiple of b,*

$$\mathbf{E}\left[\, \Phi^{t+b} \,\middle|\, \mathfrak{F}^t \,\right] \leqslant \sum_{i=1}^{n} \Phi_i^t \cdot \left(1 + \left(p_i^t - \frac{1}{n}\right) \cdot b \cdot \gamma + \frac{5(C-1)^2 b}{n} \cdot b \cdot \frac{\gamma^2}{n}\right), \tag{3.2}$$

and

$$\mathbf{E}\left[\, \Psi^{t+b} \,\middle|\, \mathfrak{F}^t \,\right] \leqslant \sum_{i=1}^{n} \Psi_i^t \cdot \left(1 + \left(\frac{1}{n} - p_i^t\right) \cdot b \cdot \gamma + \frac{5(C-1)^2 b}{n} \cdot b \cdot \frac{\gamma^2}{n}\right). \tag{3.3}$$

The proof proceeds in a similar manner to [20, Lemma 4.1], but we bound the terms in (3.6) and (3.10) more tightly using the new condition $\mathcal{C}_3$. Compared to the statement of [20, Lemma 4.1], the coefficients of the term $\frac{\gamma^2}{n}$ change from $5C^2S^2\frac{b^2}{n}$ to $5(C-1)^2\frac{b^2}{n}$. Note that C is replaced by $C-1$, which makes a difference when $C = 1 + o(1)$, and that S does not appear as we have assumed that $b \geqslant \frac{2CS}{(C-1)^2} \cdot n$.

Proof. Consider an arbitrary step $t \geqslant 0$ being a multiple of b and for convenience let $p = p^t$. 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

$$\gamma \leqslant \frac{n}{2(C-1)\cdot b} \leqslant \frac{C-1}{4CS}. \tag{3.4}$$

Consider an arbitrary bin $i \in [n]$. Define the binary vector $Z \in \{0,1\}^b$, where Z_j indicates whether the j-th ball was allocated to bin i. The expected change for the overload potential Φ_i^t of the bin is given by,

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

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

$$\begin{aligned}
&\sum_{z\in\{0,1\}^b} \mathbf{Pr}\left[\, Z = z \,\right] \cdot \mathbf{E}\left[\, e^{\gamma \sum_{j=1}^{b}\left(z_j w^{t+j} - \frac{w^{t+j}}{n}\right)} \,\middle|\, \mathfrak{F}^t, Z = z \,\right] \\
&\quad \overset{(a)}{=} \sum_{z\in\{0,1\}^b} \prod_{j=1}^{b} (p_i)^{z_j} (1-p_i)^{1-z_j} (\mathbf{E}[\, e^{\gamma W(1-\frac{1}{n})} \,])^{z_j} (\mathbf{E}[\, e^{-\gamma W/n} \,])^{1-z_j} \\
&\quad \overset{(b)}{\leqslant} \sum_{z\in\{0,1\}^b} \prod_{j=1}^{b} \left(p_i \cdot \left(1 + \gamma \cdot \left(1 - \frac{1}{n}\right) + S\gamma^2\right)\right)^{z_j} \cdot \left((1-p_i) \cdot \left(1 - \frac{\gamma}{n} + \frac{S\gamma^2}{n^2}\right)\right)^{1-z_j} \\
&\quad \overset{(c)}{=} \left(p_i \cdot \left(1 + \gamma \cdot \left(1 - \frac{1}{n}\right) + S\gamma^2\right) + (1-p_i) \cdot \left(1 - \frac{\gamma}{n} + \frac{S\gamma^2}{n^2}\right)\right)^b \\
&\quad = \left(1 + \gamma \cdot \left(p_i - \frac{1}{n}\right) + p_i \cdot S\gamma^2 + (1-p_i) \cdot \frac{S\gamma^2}{n^2}\right)^b \\
&\quad \overset{(d)}{\leqslant} \left(1 + \gamma \cdot \left(p_i - \frac{1}{n}\right) + 2 \cdot p_i \cdot S\gamma^2\right)^b,
\end{aligned} \tag{3.5}$$

using in (a) that the weights are independent given $\mathfrak{F}^t$, in (b) Lemma 2.3 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 $p_i \geqslant \frac{1}{n^2}$ by condition $\mathcal{C}_3$ for $C \in (1, 1.9)$. Let us define

$$u_i := \left(p_i - \frac{1}{n}\right) \cdot \gamma + 2 \cdot p_i \cdot S\gamma^2. \tag{3.6}$$

We will now show that $|u_i \cdot b| \leqslant 2(C-1) \cdot b \cdot \frac{\gamma}{n} \leqslant 1$, which holds indeed since

$$\begin{aligned}
|u_i \cdot b| &= \left|\left(p_i - \frac{1}{n}\right) \cdot b \cdot \gamma + 2 \cdot p_i \cdot b \cdot S\gamma^2\right| \\
&\leqslant \left|\left(p_i - \frac{1}{n}\right) \cdot b \cdot \gamma\right| + 2 \cdot p_i \cdot b \cdot S\gamma^2 \\
&\overset{(a)}{\leqslant} \frac{C-1}{n} \cdot b \cdot \gamma + 2 \cdot \frac{C}{n} \cdot b \cdot S\gamma^2 \\
&= (C - 1 + 2CS\gamma) \cdot b \cdot \frac{\gamma}{n} \\
&\overset{(b)}{\leqslant} 2(C-1) \cdot b \cdot \frac{\gamma}{n} \qquad (3.7) \\
&\overset{(c)}{\leqslant} 1, \qquad (3.8)
\end{aligned}$$

using in (a) that $\left|p_i - \frac{1}{n}\right| \leqslant \frac{C-1}{n}$ by condition $\mathcal{C}_3$, in (b) that $\gamma \leqslant \frac{C-1}{2CS}$ by (3.4) and in (c) that $\gamma \leqslant \frac{n}{2(C-1)\cdot b}$.

Then,

$$\begin{aligned}
\mathbf{E}\left[\, \Phi_i^{t+b} \,\middle|\, \mathfrak{F}^t \,\right] &\overset{(a)}{\leqslant} \Phi_i^t \cdot e^{u_i \cdot b} \\
&\overset{(b)}{\leqslant} \Phi_i^t \cdot \left(1 + u_i \cdot b + (u_i \cdot b)^2\right) \\
&\overset{(3.6)}{=} \Phi_i^t \cdot \left(1 + \left(p_i - \frac{1}{n}\right) \cdot b \cdot \gamma + 2 \cdot p_i \cdot b \cdot S\gamma^2 + (u_i \cdot b)^2\right) \\
&\overset{(3.7)}{\leqslant} \Phi_i^t \cdot \left(1 + \left(p_i - \frac{1}{n}\right) \cdot b \cdot \gamma + 2 \cdot p_i \cdot b \cdot S\gamma^2 + \left(2(C-1) \cdot b \cdot \frac{\gamma}{n}\right)^2\right) \\
&\overset{(c)}{\leqslant} \Phi_i^t \cdot \left(1 + \left(p_i - \frac{1}{n}\right) \cdot b \cdot \gamma + \frac{5(C-1)^2 b}{n} \cdot b \cdot \frac{\gamma^2}{n}\right),
\end{aligned}$$

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 (3.8), and in (c) that $\frac{(C-1)^2 b}{n} \cdot b \cdot \frac{\gamma^2}{n} \geqslant 2 \cdot \frac{C}{n} \cdot b \cdot S\gamma^2 \geqslant 2 \cdot p_i \cdot b \cdot S\gamma^2$, since $b \geqslant \frac{2CS}{(C-1)^2} \cdot n$.

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:

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

$$\overset{(b)}{\leqslant} \sum_{z\in\{0,1\}^b}\prod_{j=1}^{b}\left(p_i\cdot\left(1-\gamma\cdot\left(1-\frac{1}{n}\right)+S\gamma^2\right)\right)^{z_j}\cdot\left((1-p_i)\cdot\left(1+\frac{\gamma}{n}+\frac{S\gamma^2}{n^2}\right)\right)^{1-z_j}$$

$$\overset{(c)}{=} \left(p_i\cdot\left(1-\gamma\cdot\left(1-\frac{1}{n}\right)+S\gamma^2\right)+(1-p_i)\cdot\left(1+\frac{\gamma}{n}+\frac{S\gamma^2}{n^2}\right)\right)^b$$

$$= \left(1+\left(\frac{1}{n}-p_i\right)\cdot\gamma+p_i\cdot S\gamma^2+(1-p_i)\cdot\frac{S\gamma^2}{n^2}\right)^b$$

$$\overset{(d)}{\leqslant} \left(1+\left(\frac{1}{n}-p_i\right)\cdot\gamma+2\cdot p_i\cdot S\gamma^2\right)^b, \tag{3.9}$$

using in (a) that the weights are independent given $\mathfrak{F}^t$, in (b) Lemma 2.3 twice with $\kappa=-\left(1-\frac{1}{n}\right)$ and with $\kappa=\frac{1}{n}$ respectively, in (c) the binomial theorem and in (d) that $p_i\geqslant\frac{1}{n^2}$ by condition $\mathcal{C}_3$ for $C\in(1,1.9)$. Let us define

$$\widetilde{u}_i := \left(\frac{1}{n}-p_i\right)\cdot\gamma+2\cdot p_i\cdot S\gamma^2. \tag{3.10}$$

Similarly, to (3.8), we get that

$$|\widetilde{u}_i\cdot b| \leqslant \left|\left(\frac{1}{n}-p_i\right)\cdot b\cdot\gamma\right|+2\cdot p_i\cdot b\cdot S\gamma^2 \leqslant 2(C-1)\cdot b\cdot\frac{\gamma}{n} \tag{3.11}$$

$$\leqslant 1. \tag{3.12}$$

So,

$$\begin{aligned}
\mathbf{E}\left[\,\Psi_i^{t+b}\,\middle|\,\mathfrak{F}^t\,\right] &\overset{(a)}{\leqslant} \Psi_i^t\cdot e^{\widetilde{u}_i\cdot b}\\
&\overset{(b)}{\leqslant} \Psi_i^t\cdot\left(1+\widetilde{u}_i\cdot b+(\widetilde{u}_i\cdot b)^2\right)\\
&\overset{(3.10)}{=} \Psi_i^t\cdot\left(1+\left(\frac{1}{n}-p_i\right)\cdot b\cdot\gamma+2\cdot p_i\cdot S\gamma^2\cdot b+(\widetilde{u}_i\cdot b)^2\right)\\
&\overset{(3.11)}{\leqslant} \Psi_i^t\cdot\left(1+\left(\frac{1}{n}-p_i\right)\cdot b\cdot\gamma+2\cdot p_i\cdot b\cdot\gamma^2+\left(2(C-1)\cdot b\cdot\frac{\gamma}{n}\right)^2\right)\\
&\overset{(c)}{\leqslant} \Psi_i^t\cdot\left(1+\left(\frac{1}{n}-p_i\right)\cdot b\cdot\gamma+\frac{5(C-1)^2b}{n}\cdot b\cdot\frac{\gamma^2}{n}\right),
\end{aligned}$$

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 (3.12), and in (c) that $\frac{(C-1)^2b}{n}\cdot b\cdot\frac{\gamma^2}{n}\geqslant 2\cdot\frac{C}{n}\cdot b\cdot S\gamma^2\geqslant 2\cdot p_i\cdot b\cdot S\gamma^2$, since $b\geqslant\frac{2CS}{(C-1)^2}\cdot n$. □

Having verified the preconditions for Theorem 3.3, we are now ready to prove the bound on the gap for this family of processes.

Remark 3.5. *The same upper bound in Theorem 3.1 also holds for processes with random tie breaking. The reason for this is that (i) averaging probabilities in* (2.1) *can only reduce the maximum entry (and increase the minimum) in the allocation vector $\widetilde{p}^t$, i.e. $\max_{i\in[n]}\widetilde{p}_i^t(x^t)\leqslant\max_{i\in[n]}p_i$, so it still satisfies condition $\mathcal{C}_3$ 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* (3.2) *and* (3.3) *in Lemma 3.4 remain the same.*

Proof of Theorem 3.1. Consider the b-Batched setting at steps that are a multiple of b and rounds consisting of b consecutive allocations. By Lemma 3.4, the preconditions of Theorem 3.3

are satisfied for $K := 5 \cdot (C-1)^2 \cdot \frac{b}{n}$, $R := b$ and $\gamma := \frac{\varepsilon\delta}{8K} = \frac{\varepsilon\delta}{40\cdot(C-1)^2\cdot\frac{b}{n}} \leqslant \frac{n}{2(C-1)\cdot b}$, since $\varepsilon \leqslant C-1$ and also $\gamma \leqslant 1$ since $b \geqslant \frac{2CS}{(C-1)^2} \cdot n$, $C > 1$ and $S \geqslant 1$ (as in (3.4)),

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

Hence, by Theorem 3.3, there exists a constant $c := c(\delta) > 0$ such that for any step $m \geqslant 0$ which is a multiple of b,

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

Therefore, by Markov's inequality

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

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

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

4 Tight Bound: $\mathcal{O}(\sqrt{(b/n)\cdot\log n})$ Gap

In this section, we will prove the stronger $\mathcal{O}\big(\sqrt{(b/n)\cdot\log n}\big)$ bound on the gap for a family of processes in the weighted b-Batched setting (with $b \in [2n\log n, n^3]$). More specifically, these processes are a subset of the ones analyzed in Section 3 and include the $(1+\beta)$-process with $\beta = \sqrt{(n/b)\log n}$, as well as Quantile(1/2) mixed with One-Choice. As we will show in Section 5, these processes achieve the asymptotically optimal bound.

Theorem 4.1. *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 $\varepsilon = \sqrt{(n/b)\cdot\log n}$. Consider any process with probability allocation vector p^t satisfying at every step $t \geqslant 0$, condition $\mathcal{C}_1$ for constant $\delta \in (0,1)$ and ε, as well as condition $\mathcal{C}_3$ for $C = 1+\varepsilon$. Then, there exists a constant $\kappa := \kappa(\delta, S) > 0$, such that for any step $m \geqslant 0$ being a multiple of b,*

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

There are two key steps in the proof:

Step 1: Similarly to the analysis in [21], we will use two instances of the hyperbolic cosine potential (defined in (3.1)), in order to show that it is concentrated at $\mathcal{O}(n)$. More specifically, we will be using $\Gamma_1 := \Gamma_1(\gamma_1)$ with the smoothing parameter $\gamma_1 := \frac{\delta}{40S} \cdot \sqrt{n/(b\log n)}$ 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 $t \geqslant 0$.

In the following lemma, proven in Section 4.1, we show that w.h.p. $\Gamma_2 = \mathcal{O}(n)$ for any $\log^3 n$ consecutive batches.

Lemma 4.2. *Consider any process satisfying the conditions in Theorem 4.1. Let $\widetilde{c} := 2 \cdot \frac{8c}{\delta}$ where $c := c(\delta) > 0$ is the constant from Theorem 3.3. Then, for any step $t \geqslant 0$ being a multiple of b,*

$$\mathbf{Pr}\left[\,\bigcap_{j\in[0,\log^3 n]} \left\{\Gamma_2^{t+j\cdot b} \leqslant \widetilde{c}n\right\}\,\right] \geqslant 1 - n^{-3}.$$

The proof follows the interplay between the two hyperbolic cosine potentials, in that conditioning on $\Gamma_1^t = \text{poly}(n)$ (which follows w.h.p. by the analysis in Section 3) implies that $\left|\Delta\Gamma_2^{t+1}\right| \leqslant n^{1/4} \cdot \sqrt{(n/b) \cdot \log n}$ (Lemma 4.5 (ii)). This in turn allows us to apply a bounded difference inequality to prove concentration for Γ_2. In contrast to [21] and [22], here we need a slightly different concentration inequality Theorem A.6 (also used in [20]), 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 Section 4.1.

Step 2: Consider an arbitrary step $s = t + j \cdot b$ where $\{\Gamma_2^s \leqslant \widetilde{c}n\}$ holds. Then, the number of bins i with load y_i^s at least $z := \frac{1}{\gamma_2} \cdot \log(\widetilde{c}/\delta) = \Theta(\sqrt{(b/n) \cdot \log n})$ 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 $\lambda := \frac{\varepsilon}{4CS} = \Theta(\sqrt{(n/b) \cdot \log n})$ and we define $\Lambda_i^t = 0$ for the rest of the bins i. This means that when $\{\Gamma_2^s \leqslant \widetilde{c}n\}$ holds, the probability of allocating to one of these bins is $p_i^s \leqslant \frac{1-\varepsilon}{n}$, because of the condition $\mathcal{C}_1$. Hence, the potential drops in expectation over one batch (Lemma 4.9) and this means that w.h.p. $\Lambda^m = \text{poly}(n)$, which implies that $\text{Gap}(m) = \mathcal{O}(z + \lambda^{-1} \cdot \log n) = \mathcal{O}(\sqrt{(b/n) \cdot \log n})$ gap.

4.1 Step 1: Concentration of the Γ Potential

Recall that in Theorem 4.1, we considered the weighted b-Batched setting with any $b \in [2n \log n, n^3]$ and weights sampled independently from a Finite-MGF(S) distribution with constant $S \geqslant 1$, for any allocation process with probability allocation vector p^t satisfying condition $\mathcal{C}_1$ for constant $\delta \in (0, 1)$ and $\varepsilon \in (0, 1)$ as well as condition $\mathcal{C}_3$ for some $C > 1$, at every step $t \geqslant 0$.

Lemma 4.2 (Restated, page 16). *Consider any process satisfying the conditions in Theorem 4.1. Let $\widetilde{c} := 2 \cdot \frac{8c}{\delta}$ where $c := c(\delta) > 0$ is the constant from Theorem 3.3. Then, for any step $t \geqslant 0$ being a multiple of b,*

$$\mathbf{Pr}\left[\bigcap_{j \in [0, \log^3 n]} \left\{\Gamma_2^{t+j \cdot b} \leqslant \widetilde{c}n\right\}\right] \geqslant 1 - n^{-3}.$$

Remark 4.3. *The number of bins $i \in [n]$ with normalized load value y_i^t at least z is w.h.p. at most $\widetilde{c}n \cdot e^{-\gamma_2 z}$, where $\gamma_2 = \Theta(\sqrt{(n/b) \cdot \log n})$.*

The proof of this lemma is similar to the proofs in [20, Section 5] and [21, Section 5], 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 $\gamma_1 := \frac{\delta}{40S} \cdot \sqrt{n/(b \log n)}$ and $\gamma_2 := \frac{\gamma_1}{8 \cdot 30}$.

The rest of this section is organized as follows. In Section 4.1.1, we establish some basic properties for the potentials Γ_1 and Γ_2 and in Section 4.1.2 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 4.2. Then, in Section 4.2, we complete the proof of Theorem 4.1.

4.1.1 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). By a simple Chernoff bound and a union bound, we can deduce that this holds for a poly(n)-long interval.

Lemma 4.4 (cf. [20, Lemma 5.4]). *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*

$$\mathbf{Pr}\left[\bigcap_{s\in[t_0,t_1]} \mathcal{H}^s\right] \geqslant 1 - n^{-10}$$

We will now show that when $\Gamma_1^t = \text{poly}(n)$ and $\mathcal{H}^t$ holds, then $\Delta\Gamma_2^{t+1}$ is small.

Lemma 4.5. *Consider any process satisfying the conditions in Lemma 4.2 and any step $t \geqslant 0$, such that $\Gamma_1^t \leqslant 2\widetilde{c}n^{26}$ and $\mathcal{H}^t$ holds. Then, we have that*

$$(i) \qquad \Gamma_2^t \leqslant n^{5/4},$$
$$(ii) \qquad |\Gamma_2^{t+1} - \Gamma_2^t| \leqslant n^{1/4} \cdot \sqrt{\frac{n}{b} \cdot \log n}.$$

Further, let $\widehat{x}^t$ be the load vector obtained by moving any 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. Recall that $\gamma_1 := \frac{\delta}{40S} \cdot \sqrt{n/(b \log n)}$ and $\gamma_2 := \frac{\gamma_1}{8\cdot 30}$. Consider any step $t \geqslant 0$, such that $\Gamma_1^t \leqslant 2\widetilde{c}n^{26}$ and $\mathcal{H}^t$ holds. We start by bounding the load of any bin $i \in [n]$,

$$\Gamma_1^t \leqslant 2\widetilde{c}n^{26} \Rightarrow e^{\gamma_1 \cdot y_i^t} + e^{-\gamma_1 \cdot y_i^t} \leqslant \widetilde{c}n^{26} \Rightarrow y_i^t \leqslant \frac{27}{\gamma_1} \log n \wedge -y_i^t \leqslant \frac{27}{\gamma_1} \log n, \tag{4.1}$$

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 (4.1), we bound the contribution of any bin $i \in [n]$ to Γ_2^t as follows,

$$\Gamma_{2i}^t = e^{\gamma_2 y_i^t} + e^{-\gamma_2 y_i^t} \leqslant 2 \cdot e^{\gamma_2 \cdot \frac{27}{\gamma_1} \log n} \leqslant 2n^{1/8}, \tag{4.2}$$

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. Let bin $j \in [n]$ be the bin where the j-th ball was allocated. We consider the following cases for the contribution of a bin i to Γ_{2i}^t:

Case 1 [$i = j$ and $y_j^t \geqslant 0$]: Since $j \in [n]$ is overloaded, we have that

$$\begin{aligned}\left|\Delta\Gamma_{2j}^{t+1}\right| &\leqslant \Gamma_{2j}^t \cdot e^{\gamma_2 \cdot \frac{15}{\zeta} \cdot \log n} - \Gamma_{2j}^t \overset{(a)}{\leqslant} \Gamma_{2j}^t \cdot \left(1 + \gamma_2 \cdot \frac{30}{\zeta} \cdot \log n\right) - \Gamma_{2j}^t \\ &= \Gamma_{2j}^t \cdot \gamma_2 \cdot \frac{30}{\zeta} \cdot \log n \overset{(b)}{\leqslant} n^{1/8} \cdot \sqrt{\frac{n}{b} \cdot \log n},\end{aligned}$$

using in (a) the Taylor estimate $e^v \leqslant 1 + 2v$ (for $v \leqslant 1$) and that $\gamma_2 \cdot \frac{15}{\zeta} \cdot \log n \leqslant 1$, since $\gamma_2 \leqslant \frac{\delta}{60S} \cdot \sqrt{\frac{n}{b \log n}}$ and $S \geqslant \frac{1}{\zeta}$ and in (b) that (4.2), $\gamma_2 \leqslant \frac{\delta}{60S} \cdot \sqrt{\frac{n}{b \log n}}$ and $S \geqslant \frac{1}{\zeta}$.

Case 2 [$i = j$ and $y_j^t < 0$]: Similarly, if j is underloaded, we have that

$$\begin{aligned}\left|\Delta\Gamma_{2j}^{t+1}\right| &\leqslant \Gamma_{2j}^t - \Gamma_{2j}^t \cdot e^{-\gamma_2 \cdot \frac{15}{\zeta} \cdot \log n} \leqslant \Gamma_{2j}^t - \Gamma_{2j}^t \cdot \left(1 - \gamma_2 \cdot \frac{30}{\zeta} \cdot \log n\right) \\ &= \Gamma_{2j}^t \cdot \gamma_2 \cdot \frac{30}{\zeta} \cdot \log n \leqslant n^{1/8} \cdot \sqrt{\frac{n}{b} \cdot \log n}.\end{aligned}$$

Case 3 [$i \neq j$ and $y_i^t \geqslant 0$]: The contribution of the rest of the bins is due to the change in the average load. More specifically, for any overloaded bin $i \in [n] \setminus \{j\}$,

$$\begin{aligned}\left|\Delta\Gamma_{2i}^{t+1}\right| &\leqslant \Gamma_{2i}^t \cdot e^{\gamma_2 \cdot \frac{15}{\zeta} \cdot \frac{\log n}{n}} - \Gamma_{2i}^t \leqslant \Gamma_{2i}^t \cdot \left(1 + 2 \cdot \gamma_2 \cdot \frac{15}{\zeta} \cdot \frac{\log n}{n}\right) - \Gamma_{2i}^t \\ &= \Gamma_{2i}^t \cdot \gamma_2 \cdot \frac{30}{\zeta} \cdot \frac{\log n}{n} \leqslant \sqrt{\frac{\log n}{bn}} \cdot n^{1/8}.\end{aligned}$$

Case 4 [$i \neq j$ and $y_i^t < 0$]: Similarly, for any underloaded bin $i \in [n] \setminus \{j\}$,

$$\begin{aligned}\left|\Delta\Gamma_{2i}^{t+1}\right| &\leqslant \Gamma_{2i}^t - \Gamma_{2i}^t \cdot e^{-\gamma_2 \cdot \frac{15}{\zeta} \cdot \frac{\log n}{n}} \\ &\leqslant \Gamma_{2i}^t - \Gamma_{2i}^t \cdot \left(1 - 2 \cdot \gamma_2 \cdot \frac{15}{\zeta} \cdot \frac{\log n}{n}\right) \\ &= \Gamma_{2i}^t \cdot \gamma_2 \cdot \frac{30}{\zeta} \cdot \frac{\log n}{n} \leqslant \sqrt{\frac{\log n}{bn}} \cdot n^{1/8}.\end{aligned}$$

Hence, aggregating over all bins

$$\left|\Delta\Gamma_2^{t+1}\right| \leqslant \left|\Delta\Gamma_{2j}^{t+1}\right| + \sum_{i \in [n] \setminus \{j\}} \left|\Delta\Gamma_{2i}^{t+1}\right| \leqslant n^{1/8} \cdot \sqrt{\frac{n}{b} \cdot \log n} + n \cdot \sqrt{\frac{\log n}{bn}} \cdot n^{1/8} \leqslant n^{1/4} \cdot \sqrt{\frac{n}{b} \cdot \log n},$$

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, it follows that $w^t \leqslant \frac{15}{\zeta} \cdot \log n$, so for bin i,

$$\Gamma_{1i}^t(\widehat{x}^t) \leqslant e^{\gamma_1 w^t} \cdot \Gamma_{1i}^t(x^t) \leqslant 2 \cdot \Gamma_{1i}^t(x^t),$$

since $\gamma_1 \leqslant \frac{1}{40 \cdot S \cdot \log n}$ and $S > 1/\zeta$. Similarly, for bin j.

$$\Gamma_{1j}^t(\widehat{x}^t) \leqslant e^{\gamma_1 w^t} \cdot \Gamma_{1j}^t(x^t) \leqslant 2 \cdot \Gamma_{1j}^t(x^t),$$

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 4.6. *Consider any process satisfying the conditions in Lemma 4.2. Then, there exists a constant $\widetilde{c} := \widetilde{c}(\delta)$ such that for any step $t \geqslant 0$ being a multiple of b,*

$$(i) \quad \mathbf{E}[\,\Gamma_1^t\,] \leqslant \frac{\widetilde{c}}{2} \cdot n, \quad \text{and} \quad (ii) \quad \mathbf{E}[\,\Gamma_2^t\,] \leqslant \frac{\widetilde{c}}{2} \cdot n.$$

Further,

$$(iii) \quad \mathbf{E}\left[\,\Gamma_2^{t+b} \;\middle|\; \mathfrak{F}^t, \Gamma_2^t \geqslant \widetilde{c}n\,\right] \leqslant \Gamma_2^t \cdot \left(1 - \frac{1}{\log n}\right),$$

and

$$(iv) \quad \mathbf{E}\left[\,\Gamma_2^{t+b} \;\middle|\; \mathfrak{F}^t, \Gamma_2^t \leqslant \widetilde{c}n\,\right] \leqslant \widetilde{c}n - \frac{n}{\log^2 n}.$$

Proof. First/Second statement. Recall that $C = 1 + \varepsilon$ and $\varepsilon = \sqrt{\frac{n}{b} \cdot \log n}$. By Lemma 3.4 with $K := 5 \cdot (C-1)^2 \cdot \frac{b}{n}$, $R := b$ and $\gamma_1 := \frac{\delta}{40S} \cdot \sqrt{\frac{n}{b \log n}}$, the preconditions for $\Phi := \Phi(\gamma_1)$ and $\Psi := \Psi(\gamma_1)$ in Theorem 3.3 are satisfied. To apply Theorem 3.3 we just need to verify (i) that $\gamma_1 \leqslant \frac{\varepsilon\delta}{8K} \leqslant 1$, which holds since

$$\frac{\delta}{40S} \cdot \sqrt{\frac{n}{b \log n}} \leqslant \frac{\delta}{40} \cdot \sqrt{\frac{n}{b \log n}} = \frac{\varepsilon\delta}{8K} \leqslant 1$$

and (ii) that $b \geqslant \frac{2CS}{(C-1)^2} \cdot n$, which follows since

$$\frac{2CS}{(C-1)^2} \cdot n = 2CS \cdot \frac{b}{\log n} \leqslant b.$$

Hence, by Theorem 3.3 we get the conclusion by setting $\widetilde{c} := 16c/\delta$, for some constant $c := c(\delta) > 0$.

Similarly for the potential $\Gamma_2 := \Gamma_2(\gamma_2)$ since $\gamma_2 \leqslant \gamma_1$.

Third statement. Furthermore, by Lemma 3.4 and Theorem 3.3, we also get that for any $t \geqslant 0$,

$$\mathbf{E}\left[\, \Gamma_2^{t+b} \;\middle|\; \mathfrak{F}^t \,\right] \leqslant \Gamma_2^t \cdot \left(1 - b \cdot \frac{\varepsilon\delta}{8n} \cdot \gamma_2\right) + b \cdot c\gamma_2\varepsilon. \tag{4.3}$$

We define the constant

$$\begin{aligned}\widetilde{c}_1 &:= \frac{1}{2} \cdot b \cdot \frac{\varepsilon\delta}{8n} \cdot \gamma_2 = b \cdot \frac{\delta^2}{16 \cdot 8 \cdot 30 \cdot 40S \cdot n} \cdot \sqrt{\frac{n}{b \log n}} \cdot \sqrt{\frac{n \log n}{b}} \\ &= \frac{\delta^2}{16 \cdot 8 \cdot 30 \cdot 40 \cdot S}.\end{aligned}$$

When $\left\{\Gamma_2^t \geqslant \widetilde{c}n\right\}$ holds, then (4.3) yields,

$$\begin{aligned}\mathbf{E}\left[\, \Gamma_2^{t+b} \,\middle|\, \mathfrak{F}^t, \Gamma_2^t \geqslant \widetilde{c}n \,\right] &\leqslant \Gamma_2^t \cdot \left(1 - 2 \cdot \widetilde{c}_1\right) + b \cdot c\gamma_2\varepsilon \\ &= \Gamma_2^t - \widetilde{c}_1 \cdot \Gamma_2^t + \left(b \cdot c\gamma_2\varepsilon - \widetilde{c}_1 \cdot \Gamma_2^t\right) \\ &\leqslant \Gamma_2^t - \widetilde{c}_1 \cdot \Gamma_2^t + \left(b \cdot c\gamma_2\varepsilon - \frac{1}{2} \cdot b \cdot \frac{\varepsilon\delta}{8n} \cdot \gamma_2 \cdot \frac{16c}{\delta} \cdot n\right) \\ &\leqslant \Gamma_2^t \cdot \left(1 - \frac{1}{\log n}\right).\end{aligned}$$

Fourth statement. Similarly, when $\Gamma_1^t < \widetilde{c}n$, (4.3) yields,

$$\begin{aligned}\mathbf{E}\left[\, \Gamma_2^{t+b} \,\middle|\, \mathfrak{F}^t, \Gamma_2^t < \widetilde{c}n \,\right] &\leqslant \widetilde{c}n \cdot \left(1 - 2 \cdot \widetilde{c}_1\right) + b \cdot c\gamma_2\varepsilon \\ &= \widetilde{c}n - \widetilde{c} \cdot \widetilde{c}_1 \cdot n + \left(b \cdot c\gamma_2\varepsilon - \widetilde{c} \cdot \widetilde{c}_1 \cdot n\right) \\ &= \widetilde{c}n - \frac{\widetilde{c}}{\log n} \cdot n \leqslant \widetilde{c}n - \frac{n}{\log^2 n}.\end{aligned}$$

□

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 4.7. *Let $\widetilde{c} := 2 \cdot \frac{8c}{\delta}$ be the constant defined in Lemma 4.6. For any $b \in [2n \log n, n^3]$ and for any step $t \geqslant 0$ being a multiple of b,*

$$\mathbf{Pr}\left[\bigcap_{s \in [t, t+2b\log^3 n]} \left\{\Gamma_1^s \leqslant \widetilde{c}n^{26}\right\}\right] \geqslant 1 - n^{-10}.$$

Proof. We will start by bounding Γ_1^s at steps s being a multiple of b. Using Lemma 4.6 (i), Markov's inequality and the union bound over $2\log^3 n + 1$ steps, we have for any $t \geqslant 0$,

$$\mathbf{Pr}\left[\bigcap_{s\in[0,2\log^3 n]} \left\{\Gamma_1^{t+s\cdot b} \leqslant \widetilde{c}n^{12}\right\}\right] \geqslant 1 - (4\log^3 n)\cdot n^{-11}. \tag{4.4}$$

Now, assuming that $\Gamma_1^{t+s\cdot b} \leqslant \widetilde{c}n^{12}$, we will upper bound Γ_1 for the steps in between, i.e., for $\Gamma_1^{t+s\cdot b+r}$ 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 (3.5) in Lemma 3.4 (since $\gamma_1 \leqslant 1$ and p satisfies condition $\mathcal{C}_3$),

$$\begin{aligned}
\mathbf{E}\left[\,\Phi_{1i}^{t+s\cdot b+r} \,\middle|\, \mathfrak{F}^{t+s\cdot b}, \Phi_{1i}^{t+s\cdot b}\right] &\leqslant \Phi_{1i}^{t+s\cdot b}\cdot\left(1 + \left(p_i - \frac{1}{n}\right)\cdot\gamma_1 + 2\cdot p_i\cdot S\gamma_1^2\right)^r \\
&\overset{(a)}{\leqslant} \Phi_{1i}^{t+s\cdot b}\cdot\left(1 + \frac{C-1}{n}\cdot\gamma_1 + 2\cdot\frac{C}{n}\cdot S\gamma_1^2\right)^r \\
&\overset{(b)}{\leqslant} \Phi_{1i}^{t+s\cdot b}\cdot\left(1 + 2\cdot\frac{C-1}{n}\cdot\gamma_1\right)^r \\
&\leqslant \Phi_{1i}^{t+s\cdot b}\cdot e^{2\gamma_1(C-1)\cdot\frac{r}{n}} \\
&\overset{(c)}{\leqslant} 2\cdot\Phi_{1i}^{t+s\cdot b},
\end{aligned}$$

using in (a) that $p_i - \frac{1}{n} \leqslant \frac{C-1}{n}$ by condition $\mathcal{C}_3$, (b) that $\gamma_1 \leqslant \frac{C-1}{2CS}$ (as in (3.4)) and in (c) that $\gamma_1 \leqslant \frac{\delta}{40}\cdot\sqrt{\frac{n}{b\log n}}$ and $C - 1 = \sqrt{\frac{n\log n}{b}}$. Similarly, using (3.9) in Lemma 3.4,

$$\begin{aligned}
\mathbf{E}\left[\,\Psi_{1i}^{t+s\cdot b+r} \,\middle|\, \mathfrak{F}^{t+s\cdot b}, \Psi_{1i}^{t+s\cdot b}\right] &\leqslant \Psi_{1i}^{t+s\cdot b}\cdot\left(1 + \left(\frac{1}{n} - p_i\right)\cdot\gamma_1 + 2\cdot p_i\cdot S\gamma_1^2\right)^r \\
&\overset{(a)}{\leqslant} \Psi_{1i}^{t+s\cdot b}\cdot\left(1 + \frac{C-1}{n}\cdot\gamma_1 + 2\cdot\frac{C}{\gamma_1}\cdot S\gamma_1^2\right)^r \\
&\overset{(b)}{\leqslant} \Psi_{1i}^{t+s\cdot b}\cdot\left(1 + 2\cdot\frac{C-1}{n}\cdot\gamma_1\right)^r \\
&\leqslant \Psi_{1i}^{t+s\cdot b}\cdot e^{2\gamma(C-1)\cdot\frac{r}{n}} \\
&\overset{(c)}{\leqslant} 2\cdot\Psi_{1i}^{t+s\cdot b},
\end{aligned}$$

using in (a) that $\frac{1}{n} - p_i \leqslant \frac{C-1}{n}$ by condition $\mathcal{C}_3$, (b) that $\gamma_1 \leqslant \frac{C-1}{2CS}$ and in (c) that $\gamma_1 \leqslant \frac{\delta}{40}\cdot\sqrt{\frac{n}{b\log n}}$ and $C - 1 = \sqrt{\frac{n\log n}{b}}$. Hence, combining and 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 $r \in [0, b)$,

$$\mathbf{Pr}\left[\Gamma_1^{t+s\cdot b+r} \leqslant n^{14}\cdot\Gamma_1^{t+s\cdot b}\right] \geqslant 1 - 2n^{-14}.$$

Hence, by a union bound over the $2b\log^3 n \leqslant 2n^3\log^3 n$ possible steps (since $b \leqslant n^3$) for $s \in [0, 2\log^3 n]$ and $r \in [0, b)$,

$$\mathbf{Pr}\left[\bigcap_{r\in[0,b]}\bigcap_{s\in[0,2\log^3 n]} \left\{\Gamma_1^{t+s\cdot b+r} \leqslant n^{14}\cdot\Gamma_1^{t+s\cdot b}\right\}\right] \geqslant 1 - 2n^{-14}\cdot 2b\log^3 n \geqslant 1 - \frac{1}{2}n^{-10}. \tag{4.5}$$

Finally, taking the union bound of (4.4) and (4.5), we conclude

$$\begin{aligned}\mathbf{Pr}\left[\bigcap_{s\in[t,t+2b\log^3 n]}\left\{\Gamma_1^s \leqslant \widetilde{c}n^{26}\right\}\right] &\geqslant \Pr\left[\bigcap_{r\in[0,b]}\bigcap_{s\in[0,2\log^3 n]}\left\{\Gamma_1^{t+s\cdot b+r} \leqslant n^{14}\cdot\Gamma_1^{t+s\cdot b}\right\}\right.\\ &\qquad\left.\cap\bigcap_{s\in[0,2\log^3 n]}\left\{\Gamma_1^{t+s\cdot b}\leqslant\widetilde{c}n^{12}\right\}\right]\\ &\geqslant 1-\frac{1}{2}n^{-10}-(4\log^3 n)\cdot n^{-11}\geqslant 1-n^{-10}.\end{aligned}$$

□

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 4.8 (Recovery). *Consider any process satisfying the conditions in Lemma 4.2. Let $\widetilde{c} := 2\cdot\frac{8c}{\delta}$ be the constant defined in Lemma 4.6. For any step $t\geqslant 0$ being a multiple of b,*

$$\mathbf{Pr}\left[\bigcup_{s\in[0,\log^3 n]}\left\{\Gamma_2^{t+s\cdot b}\leqslant\widetilde{c}n\right\}\right]\geqslant 1-2n^{-8}.$$

Proof. By Lemma 4.6 (ii), using Markov's inequality at step t being a multiple of b, we have

$$\mathbf{Pr}\left[\,\Gamma_2^t\leqslant\widetilde{c}n^9\,\right]\geqslant 1-n^{-8}. \tag{4.6}$$

We will be assuming $\Gamma_2^t\leqslant\widetilde{c}n^9$. By Lemma 4.6 (iii), for any step $r\geqslant 0$, then

$$\mathbf{E}\left[\,\Gamma_2^{r+1}\,\middle|\,\mathfrak{F}^r,\Gamma_2^r>\widetilde{c}n\,\right]\leqslant\Gamma_2^r\cdot\Big(1-\frac{1}{\log n}\Big).$$

In order to prove that $\Gamma_2^{t+s\cdot b}$ is small for some $s\in[0,b\log^3 n]$, we define the "killed" potential function for any $r\in[0,\log^3 n]$,

$$\widehat{\Gamma}_2^{t+r\cdot b}:=\Gamma_2^{t+r\cdot b}\cdot\mathbf{1}_{\bigcap_{s\in[0,r]}\{\Gamma_2^{t+s\cdot b}>\widetilde{c}n\}}.$$

Note that $\widehat{\Gamma}_2^{t+r\cdot b}\leqslant\Gamma_2^{t+r\cdot b}$ and that $\left\{\widehat{\Gamma}_2^{t+r\cdot b}=0\right\}$ implies that $\left\{\widehat{\Gamma}_2^{t+r\cdot b+1}=0\right\}$. Hence, the $\widehat{\Gamma}$ potential satisfies unconditionally the drop inequality of Lemma 4.6 (iii), that is,

$$\mathbf{E}\left[\,\widehat{\Gamma}_2^{t+(r+1)\cdot b}\,\middle|\,\mathfrak{F}^{t+r\cdot b},\widehat{\Gamma}_2^{t+r\cdot b}\,\right]\leqslant\widehat{\Gamma}_2^{t+r\cdot b}\cdot\Big(1-\frac{1}{\log n}\Big).$$

Inductively applying this for $\log^3 n$ batches and using that $\widehat{\Gamma}_2^t\leqslant\Gamma_2^t\leqslant\widetilde{c}n^9$,

$$\mathbf{E}\left[\,\widehat{\Gamma}_2^{t+(\log^3 n)\cdot b}\,\middle|\,\mathfrak{F}^t,\Gamma_2^t\leqslant\widetilde{c}n^9\,\right]\leqslant\Gamma_2^t\cdot\Big(1-\frac{1}{\log n}\Big)^{\log^3 n}\leqslant\widetilde{c}n^9\cdot e^{-\log^2 n}<n^{-7}.$$

So by Markov's inequality,

$$\mathbf{Pr}\left[\,\widehat{\Gamma}_2^{t+(\log^3 n)\cdot b}<n\,\middle|\,\Gamma_2^t\leqslant\widetilde{c}n^9\,\right]\geqslant 1-n^{-8}$$

By combining with (4.6),

$$\mathbf{Pr}\left[\,\widehat{\Gamma}_2^{t+(\log^3 n)\cdot b}<n\,\right]\geqslant\big(1-n^{-8}\big)\cdot\big(1-n^{-8}\big)\geqslant 1-2n^{-8}.$$

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 $1-2n^{-8}$, we have that $\widehat{\Gamma}_2^{t+(\log^3 n)\cdot b}=0$ or equivalently the event

$$\neg\bigcap_{s\in[0,\log^3 n]}\left\{\Gamma_2^{t+s\cdot b}>\widetilde{c}n\right\},$$

holds, which implies the conclusion. □

4.1.2 Completing the Proof of Lemma 4.2

We are now ready to prove Lemma 4.2, using a method of bounded differences with a bad event Theorem A.6 ([19, Theorem 3.3]).

Proof of Lemma 4.2. Our starting point is to apply Lemma 4.8, 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,

$$\mathbf{Pr}\left[\bigcup_{\rho\in[-\log^3 n,0]}\left\{\Gamma_2^{t+\rho\cdot b}\leqslant \widetilde{c}n\right\}\right]\geqslant 1-2n^{-8}. \tag{4.7}$$

Note that if $t < b\log^3 n$, then deterministically $\Gamma_2^0 = 2n \leqslant \widetilde{c}n$ (which corresponds to $\rho = -t/b$).

We are now going to apply the concentration inequality Theorem A.6 to each of the batches starting at $t+\rho\cdot b, \ldots, t+(\log^3 n)\cdot b$ and show that the potential remains $\leqslant \widetilde{c}n$ at the last step of each batch. More specifically, we will show that for any $\widetilde{r} \in [\rho, \log^3 n]$, for $r = t + b\cdot\widetilde{r}$,

$$\mathbf{Pr}\left[\,\Gamma_2^{r+b} > \widetilde{c}n \,\middle|\, \mathfrak{F}^r, \Gamma_2^r \leqslant \widetilde{c}n\,\right] \leqslant 3n^{-4}.$$

We will show this by applying Theorem A.6 for all steps of the batch $[r, r+b]$. We define the good event

$$\mathcal{G}_r := \mathcal{G}_r^{r+b} := \bigcap_{s\in[r,r+b]}\left(\left\{\Gamma_1^s \leqslant \widetilde{c}n^{26}\right\}\cap\mathcal{H}^s\right),$$

and the bad event $\mathcal{B}_r := (\mathcal{G}_r)^c$. Using a union bound over Lemma 4.4 and Lemma 4.7,

$$\mathbf{Pr}\left[\bigcap_{s\in[t-b\log^3 n, t+b\log^3 n]}\left(\left\{\Gamma_1^s \leqslant \widetilde{c}n^{26}\right\}\cap\mathcal{H}^s\right)\right]\geqslant 1-2n^{-10}. \tag{4.8}$$

Consider any $u \in [r, r+b]$. Further, we define the slightly weaker good event, $\widetilde{\mathcal{G}}_r^u := \bigcap_{s\in[r,u]}\left(\left\{\Gamma_1^s \leqslant 2\widetilde{c}n^{26}\right\}\cap\mathcal{H}^s\right)$ 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 $(n^{5/4}, n^{1/4}\cdot\sqrt{(n/b)\cdot\log n}, 2n^{-10})$ (Definition A.5).

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,

$$\begin{aligned}\left|\widehat{\Gamma}_r^{r+b}(\omega) - \widehat{\Gamma}_r^{r+b}(\omega')\right| &\leqslant \max_{\widetilde{\omega}}\widehat{\Gamma}_r^{r+b}(\widetilde{\omega}) - \min_{\widetilde{\omega}}\widehat{\Gamma}_r^{r+b}(\widetilde{\omega})\\ &\leqslant \max_{\widetilde{\omega}\in\widetilde{\mathcal{G}}_r^{r+b}}\Gamma_2^{r+b}(\widetilde{\omega}) - 0\\ &\leqslant n^{5/4},\end{aligned}$$

where in the last inequality we used Lemma 4.5 (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 $u \in [r, r+b]$,

$$\Gamma_1^u(\omega') \leqslant 2\cdot\Gamma_1^u(\omega) \leqslant 2\widetilde{c}n^{26},$$

where the first inequality is by Lemma 4.5 (iii). Hence $\omega' \in \widetilde{\mathcal{G}}_r^{r+b}$, so $\mathbf{1}_{\widetilde{\mathcal{G}}_r^{r+b}(\omega')} = 1$ and $\widehat{\Gamma}_r^{r+b}(\omega') = \Gamma_2^{r+b}(\omega')$. Similarly, for $\omega \in \mathcal{G}_r \subseteq \widetilde{\mathcal{G}}_r^{r+b}$, we have $\widehat{\Gamma}_r^{r+b}(\omega) = \Gamma_2^{r+b}(\omega)$ and by Lemma 4.5 (ii),

$$\left|\widehat{\Gamma}_r^{r+b}(\omega) - \widehat{\Gamma}_r^{r+b}(\omega')\right| = \left|\Gamma_2^{r+b}(\omega) - \Gamma_2^{r+b}(\omega')\right| \leqslant n^{1/4}\cdot\sqrt{\frac{n}{b}\cdot\log n}.$$

Within a single batch all allocations are independent, so we apply Theorem A.6, choosing $\gamma_k := \frac{1}{b}$ and $N := b$, which states that for any $T > 0$ and $\mu := \mathbf{E}\left[\, \widehat{\Gamma}_r^{r+b} > \mu + T \,\middle|\, \mathfrak{F}^r, \Gamma_2^r \leqslant \widetilde{c}n \,\right]$,

$$\begin{aligned}
&\mathbf{Pr}\left[\, \widehat{\Gamma}_r^{r+b} > \mu + T \,\middle|\, \mathfrak{F}^r, \Gamma_2^r \leqslant \widetilde{c}n \,\right] \\
&\quad\leqslant \exp\left(-\frac{T^2}{2 \cdot \sum_{k=1}^{b} (n^{1/4} \cdot \sqrt{\frac{n}{b} \cdot \log n} + n^{5/4} \cdot \frac{1}{b})^2}\right) + 2n^{-10} \cdot \sum_{k=1}^{b} b.
\end{aligned}$$

By Lemma 4.6 (iv), we have $\mu \leqslant \mathbf{E}[\, \widehat{\Gamma}_r^{r+b} \mid \Gamma_2^r < \widetilde{c}n \,] \leqslant \mathbf{E}[\, \Gamma_2^{r+b} \mid \Gamma_2^r < \widetilde{c}n \,] \leqslant \widetilde{c}n - n/\log^2 n$. Hence, for $T := n/\log^2 n$, since $2n \log n \leqslant b \leqslant n^3$, we have

$$\begin{aligned}
\mathbf{Pr}\left[\, \widehat{\Gamma}_r^{r+b} > \widetilde{c}n \,\middle|\, \mathfrak{F}^r, \Gamma_2^r \leqslant \widetilde{c}n \,\right] &\leqslant \exp\left(-\frac{n^2/\log^4 n}{2 \cdot b \cdot (2 \cdot n^{1/4} \cdot \sqrt{\frac{n}{b} \cdot \log n})^2}\right) + 2n^{-10} \cdot b^2 \\
&\leqslant \exp\left(-\frac{n^{1/2}}{8 \cdot \log^5 n}\right) + 2n^{-10} \cdot n^6 \leqslant 3n^{-4}.
\end{aligned}$$

Let $\mathcal{K}_\rho^{\widetilde{r}} := \mathcal{G}_\rho^{t+\widetilde{r} \cdot b} \cap \{\Gamma_2^{t+\rho \cdot b} \leqslant \widetilde{c}n\}$ 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

$$\mathbf{Pr}\left[\, \widehat{\Gamma}_r^{t+(\widetilde{r}+1) \cdot b} \cdot \mathbf{1}_{\mathcal{K}_\rho^{\widetilde{r}+1}} > \widetilde{c}n \,\middle|\, \mathfrak{F}^r, \widehat{\Gamma}_r^{t+\widetilde{r} \cdot b} \cdot \mathbf{1}_{\mathcal{K}_\rho^{\widetilde{r}}} \leqslant \widetilde{c}n \,\right] \leqslant 3n^{-4}. \tag{4.9}$$

By union bound of (4.7) and (4.8),

$$\begin{aligned}
\mathbf{Pr}\left[\, \bigcup_{\rho \in [-\log^3 n]} \mathcal{K}_\rho^{\log^3 n} \,\right] &\geqslant \mathbf{Pr}\left[\, \mathcal{G}_{-\log^3 n}^{\log^3 n} \cap \bigcup_{\rho \in [-\log^3 n, 0]} \left\{ \Gamma_2^{t+\rho \cdot b} \leqslant \widetilde{c}n \right\} \,\right] \\
&\geqslant 1 - 2n^{-8} - 2n^{-10} \geqslant 1 - 3n^{-8}.
\end{aligned} \tag{4.10}$$

Let

$$\mathcal{A} := \bigcap_{\widetilde{r} \in [0, \log^3 n]} \left\{ \Gamma_2^{t+\widetilde{r} \cdot b} \leqslant \widetilde{c}n \right\},$$

and

$$\mathcal{A}_\rho := \bigcap_{\widetilde{r} \in [\rho, \log^3 n]} \left\{ \widehat{\Gamma}_r^{t+\widetilde{r} \cdot b} \cdot \mathbf{1}_{\mathcal{K}_\rho^{\widetilde{r}}} \leqslant \widetilde{c}n \right\}.$$

Then,

$$\begin{aligned}
\mathbf{Pr}\left[\, \mathcal{A}_\rho \,\middle|\, \Gamma_2^{t+\rho \cdot b} \leqslant \widetilde{c}n \,\right] &\geqslant \prod_{\widetilde{r} \in [\rho, \log^3 n - 1]} \mathbf{Pr}\left[\bigcap_{\widetilde{s} \in [\rho+1, \widetilde{r}+1]} \left\{ \widehat{\Gamma}_r^{t+\widetilde{s} \cdot b} \cdot \mathbf{1}_{\mathcal{K}_\rho^{\widetilde{s}}} \leqslant \widetilde{c}n \right\} \right. \\
&\qquad \left. \,\middle|\, \bigcap_{\widetilde{s} \in [\rho+1, \widetilde{r}-1]} \left\{ \widehat{\Gamma}_r^{t+\widetilde{s} \cdot b} \cdot \mathbf{1}_{\mathcal{K}_\rho^{\widetilde{s}}} \leqslant \widetilde{c}n \right\}, \widehat{\Gamma}_r^{t+\widetilde{r} \cdot b} \cdot \mathbf{1}_{\mathcal{K}_\rho^{\widetilde{s}}} \leqslant \widetilde{c}n \right] \\
&\geqslant \prod_{\widetilde{r} \in [\rho, \log^3 n - 1]} \mathbf{Pr}\left[\, \widehat{\Gamma}_r^{\widetilde{r}+b} \cdot \mathbf{1}_{\mathcal{K}_\rho^{\widetilde{r}+1}} > \widetilde{c}n \,\middle|\, \mathfrak{F}^{t+\widetilde{r} \cdot b}, \widehat{\Gamma}_r^{t+\widetilde{r} \cdot b} \cdot \mathbf{1}_{\mathcal{K}_\rho^{\widetilde{r}}} \leqslant \widetilde{c}n \,\right] \\
&\geqslant (1 - 3n^{-4})^{2 \log^3 n} \geqslant 1 - 6n^{-4} \cdot \log^3 n,
\end{aligned}$$

where in the last inequality we have used (4.9) and the fact $\rho \geqslant -\log^3 n$. So,

$$\begin{aligned}
\mathbf{Pr}\left[\, \mathcal{A}_\rho \,\right] &= \mathbf{Pr}\left[\, \mathcal{A}_\rho \,\middle|\, \Gamma_2^{t+\rho \cdot b} \leqslant \widetilde{c}n \,\right] \cdot \mathbf{Pr}\left[\, \Gamma_2^{t+\rho \cdot b} \leqslant \widetilde{c}n \,\right] + 1 \cdot \mathbf{Pr}\left[\, \neg \left\{ \Gamma_2^{t+\rho \cdot b} \leqslant \widetilde{c}n \right\} \,\right] \\
&\geqslant 1 - 6n^{-4} \cdot \log^3 n.
\end{aligned} \tag{4.11}$$

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 (4.10) and (4.11), that

$$\mathbf{Pr}\left[\mathcal{A}\right] \geqslant \mathbf{Pr}\left[\bigcup_{\rho\in[-\log^3 n,0]} \mathcal{K}_\rho^{\log^3 n} \cap \bigcap_{\rho\in[-\log^3 n,0]} \mathcal{A}_\rho\right] \geqslant 1 - 3n^{-8} - 6n^{-4}\cdot\log^6 n \geqslant 1 - n^{-3}.\square$$

4.2 Step 2: Completing the Proof of Theorem 4.1

We will now show that when $\Gamma_2^t = \mathcal{O}(n)$, the stronger potential function Λ^t drops in expectation over the next batch. This will allow us to prove that $\Lambda^m = \mathrm{poly}(n)$ and deduce that w.h.p. $\mathrm{Gap}(m) = \mathcal{O}\big(\sqrt{(b/n)\cdot\log n}\big)$.

Lemma 4.9. *Consider any process satisfying the conditions in Theorem 4.1. Let $\widetilde{c} := 2\cdot\frac{8c}{\delta}$ where $c := c(\delta) > 0$ is the constant from Theorem 3.3. For any step $t \geqslant 0$ being a multiple of b,*

$$\mathbf{E}\left[\,\Lambda^{t+b}\,\middle|\,\mathfrak{F}^t, \Gamma_2^t \leqslant \widetilde{c}n\,\right] \leqslant \Lambda^t\cdot e^{-\frac{\lambda\varepsilon}{2n}\cdot b} + n^2.$$

Proof. Consider an arbitrary step $t \geqslant 0$ being a multiple of b and consider a labeling of the bins so that they are sorted by load. Assuming that $\{\Gamma_2^t \leqslant \widetilde{c}n\}$ holds, the number of bins with load $y_i^t \geqslant z$ is at most

$$\widetilde{c}n\cdot e^{-\gamma_2\cdot z} = \widetilde{c}n\cdot e^{-\log(\widetilde{c}/\delta)} = \delta n.$$

For any bin $i \in [n]$ with $y_i^t \geqslant z$, we get as in (3.5) (using that $\lambda \leqslant 1$ and that p satisfies $\mathcal{C}_3$ for $C \in (1, 1.9)$),

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

Since there are at most δn such bins (i.e., $i \leqslant \delta n$), p satisfies condition $\mathcal{C}_1$ and the normalized vector y^t is sorted, by Lemma A.2 the upper bound on $\mathbf{E}\left[\,\Lambda^{t+b} \mid \mathfrak{F}^t, \Gamma_2^t \leqslant \widetilde{c}n\,\right]$ is maximized when $p_i^t = \frac{1-\varepsilon}{n}$, so

$$\begin{aligned}\sum_{i:y_i^t\geqslant z}\mathbf{E}\left[\,\Lambda_i^{t+b}\,\middle|\,\mathfrak{F}^t, \Gamma_2^t \leqslant \widetilde{c}n\,\right] &\overset{(a)}{\leqslant} \sum_{i:y_i^t\geqslant z}\Lambda_i^t\cdot\left(1 - \frac{\lambda\varepsilon}{n} + 2CS\cdot\frac{\lambda^2}{n}\right)^b \\ &\overset{(b)}{\leqslant} \sum_{i:y_i^t\geqslant z}\Lambda_i^t\cdot\left(1 - \frac{\lambda\varepsilon}{2n}\right)^b \overset{(c)}{\leqslant} \sum_{i:y_i^t\geqslant z}\Lambda_i^t\cdot e^{-\frac{\lambda\varepsilon}{2n}\cdot b},\end{aligned}$$

using in (a) that $p_i^t \leqslant \frac{C}{n}$, in (b) that $\lambda = \frac{\varepsilon}{4CS}$ and in (c) that $1 + v \leqslant e^v$ for any v. For the rest of the bins with $i > \delta n$,

$$\begin{aligned}\mathbf{E}\left[\,\Lambda_i^{t+b}\,\middle|\,\mathfrak{F}^t\,\right] &\leqslant \Lambda_i^t\cdot\left(1 + \left(p_i^t - \frac{1}{n}\right)\cdot\lambda + 2\cdot p_i^t\cdot S\lambda^2\right)^b \\ &\overset{(a)}{\leqslant} \Lambda_i^t\cdot\left(1 + \frac{C-1}{n}\cdot\lambda + 2CS\cdot\frac{\lambda^2}{n}\right)^b \\ &\overset{(b)}{\leqslant} \Lambda_i^t\cdot\left(1 + 2\cdot\frac{C-1}{n}\cdot\lambda\right)^b \\ &\overset{(c)}{\leqslant} \left(1 + 2\cdot\frac{C-1}{n}\cdot\lambda\right)^b \\ &\overset{(d)}{\leqslant} e^{2\cdot\frac{C-1}{n}\cdot\lambda b} \overset{(e)}{\leqslant} n,\end{aligned}$$

using in (a) that $p_i^t \leqslant \frac{C}{n}$, in (b) that $2CS \cdot \frac{\lambda^2}{n} = \frac{\varepsilon}{2n} \cdot \lambda \leqslant \frac{C-1}{2n} \cdot \lambda$ since $\lambda = \frac{\varepsilon}{4CS}$ and $\varepsilon = C-1$, in (c) that $\Lambda_i^t \leqslant 1$, in (d) that $1 + v \leqslant e^v$ for any v and in (e) that $2 \cdot \frac{C-1}{n} \cdot \lambda b \leqslant \frac{\varepsilon^2}{2CS} \cdot \frac{b}{n} \leqslant \log n$ (since $C > 1$ and $S \geqslant 1$).

Aggregating the contributions over all bins,

$$\mathbf{E}\left[\ \Lambda^{t+b} \ \middle|\ \mathfrak{F}^t, \Gamma_2^t \leqslant \widetilde{c}n\ \right] \leqslant \sum_{i:y_i^t \geqslant z} \Lambda_i^t \cdot e^{-\frac{\lambda\varepsilon}{2n}\cdot b} + \sum_{i:y_i^t < z} n \leqslant \Lambda^t \cdot e^{-\frac{\lambda\varepsilon}{2n}\cdot b} + n^2. \qquad \square$$

Now we are ready to complete the proof of Theorem 4.1.

Proof of Theorem 4.1. First consider the case when $m \geqslant b \cdot \log^3 n$. Let $t_0 = m - b \cdot \log^3 n$. Let $\mathcal{E}^t := \left\{\Gamma_2^t \leqslant \widetilde{c}n\right\}$. Then using Lemma 4.2,

$$\mathbf{Pr}\left[\ \bigcap_{j \in [0, \log^3 n]} \mathcal{E}^{t_0 + j\cdot b}\ \right] \geqslant 1 - n^{-3}. \tag{4.12}$$

We define the killed potential $\widetilde{\Lambda}$, with $\widetilde{\Lambda}^{t_0} := \Lambda^{t_0}$ and for $j > 0$,

$$\widetilde{\Lambda}^{t_0 + j\cdot b} := \Lambda^{t_0 + j\cdot b} \cdot \mathbf{1}_{\cap_{s\in[0,j]} \mathcal{E}^{t_0 + s\cdot b}}.$$

Since $\widetilde{\Lambda}^t \leqslant \Lambda^t$, we have that by Lemma 4.9 for $t = t_0 + j \cdot b$, we have that

$$\mathbf{E}\left[\ \widetilde{\Lambda}^{t_0+(j+1)\cdot b} \ \middle|\ \mathfrak{F}^{t_0+j\cdot b}, \Gamma_2^{t_0+j\cdot b} \leqslant \widetilde{c}n\ \right] \leqslant \widetilde{\Lambda}^{t_0+j\cdot b} \cdot e^{-\frac{\lambda\varepsilon}{2n}\cdot b} + n^2.$$

When $\mathcal{E}^{t_0+j\cdot b}$ does not hold, then deterministically $\widetilde{\Lambda}^{t_0+(j+1)\cdot b} = \widetilde{\Lambda}^{t_0+j\cdot b} = 0$. Hence, we have the following unconditional drop inequality

$$\mathbf{E}\left[\ \widetilde{\Lambda}^{t_0+(j+1)\cdot b} \ \middle|\ \mathfrak{F}^{t_0+j\cdot b}\ \right] \leqslant \widetilde{\Lambda}^{t_0+j\cdot b} \cdot e^{-\frac{\lambda\varepsilon}{2n}\cdot b} + n^2. \tag{4.13}$$

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 A.1 to (4.13) with $a := e^{-\frac{\lambda\varepsilon}{2n}\cdot b}$ and $b := n^2$ for $\log^3 n$ steps,

$$\mathbf{E}\left[\ \widetilde{\Lambda}^m \ \middle|\ \mathfrak{F}^{t_0}, \widetilde{\Lambda}^{t_0} \leqslant e^{\kappa_1 \log^2 n}\ \right] \leqslant e^{\kappa_1 \log^2 n} \cdot a^{\log^3 n} + \frac{b}{1-a} \overset{(a)}{\leqslant} 1 + 1.5 \cdot b \leqslant 2n^2. \tag{4.14}$$

using in (a) that $\frac{\lambda\varepsilon}{2n} \cdot b = \Omega(\log n)$, since $\lambda = \frac{\varepsilon}{4CS}$ and $\varepsilon = \sqrt{(n/b) \cdot \log n}$.

By Markov's inequality, we have

$$\mathbf{Pr}\left[\ \widetilde{\Lambda}^m \leqslant 2n^5 \ \middle|\ \mathfrak{F}^{t_0}, \widetilde{\Lambda}^{t_0} \leqslant e^{\kappa_1 \log^2 n}\ \right] \geqslant 1 - n^{-3}.$$

Hence, by (4.12),

$$\mathbf{Pr}\left[\ \widetilde{\Lambda}^m \leqslant 2n^5\ \right] = \mathbf{Pr}\left[\ \widetilde{\Lambda}^m \leqslant 2n^5 \ \middle|\ \mathcal{E}^{t_0}\ \right] \cdot \mathbf{Pr}\left[\ \mathcal{E}^{t_0}\ \right] \geqslant \left(1 - n^{-3}\right) \cdot \left(1 - n^{-3}\right) \geqslant 1 - 2n^{-3}.$$

Combining with (4.12), we have

$$\mathbf{Pr}\left[\,\Lambda^m \leqslant 2n^5\,\right] \geqslant \mathbf{Pr}\left[\left\{\widetilde{\Lambda}^m \leqslant 2n^5\right\} \cap \bigcap_{j\in[0,\log^3 n]} \mathcal{E}^{t_0+j\cdot b}\right] \geqslant 1-2n^{-3}-n^{-3} \geqslant 1-n^{-2}.$$

Finally, $\{\Lambda^m \leqslant 2n^5\}$ implies that

$$\max_{i\in[n]} y_i^m \leqslant z + \frac{\log 2}{\lambda} + \frac{5\log n}{\lambda} = \mathcal{O}\left(\sqrt{(b/n)\cdot\log n}\right),$$

since $\lambda = \frac{\varepsilon}{4CS} = \Theta(\sqrt{(n\log n)/b})$, so the claim follows.

For the case when $m < b\cdot\log^3 n$, it deterministically holds that $\widetilde{\Lambda}^{t_0} \leqslant n$, which is a stronger starting point in (4.14) to prove that $\mathbf{E}[\,\Lambda^m\,] \leqslant 2n^5$, which in turn implies the gap bound. □

5 Lower Bounds on the Gap

In this section, we prove two lower bounds of $\Omega(\sqrt{(b/n)\cdot\log n})$ on the gap. Both lower bounds hold even in the unit weights case.

Observation 5.1. *Consider the b-*Batched *setting with any $b \geqslant n\log n$, and assume all balls have unit weights. Then, for any process which uses the same probability allocation vector within each batch with random tie breaking,*

$$\mathbf{Pr}\left[\,\mathrm{Gap}(b) \geqslant \frac{1}{10}\cdot\sqrt{(b/n)\cdot\log n}\,\right] \geqslant 1-n^{-2}.$$

Proof. Any such process behaves exactly like One-Choice in the first batch and so the lower bound follows from that of One-Choice for b balls into n bins (cf. [34] and [23, Lemma A.2]). □

The next lower bound is more involved. This bound also applies to processes which are allowed to adjust the probability allocation vector from one batch to another arbitrarily; e.g., the probability for a heavily underloaded bin might be set close to (or even equal to) 1, and similarly, the probability for a heavily overloaded bin might be set close to (or equal to) 0. Additionally, the lower bound below applies to any two consecutive batches, and not only to the end of the first batch as in Observation 5.1.

Theorem 5.2. *Consider the b-*Batched *setting with any $b = \Omega(n\log n)$ in the unit weights case. Furthermore, consider an allocation process which may adaptively change the probability allocation vector for each batch. Then there is a constant $\kappa > 0$ such that for any allocation process (which may adaptively change the probability for each batch) it holds that for every $t \geqslant 0$ being a multiple of b,*

$$\mathbf{Pr}\left[\,\max\{\mathrm{Gap}(t),\mathrm{Gap}(t+b)\} \geqslant \kappa\cdot\sqrt{(b/n)\cdot\log n}\,\right] \geqslant 1/2.$$

Proof. In the proof, we shall prove a slightly stronger statement:

$$\mathbf{Pr}\left[\,\max\{\mathrm{Gap}(t),\mathrm{Gap}(t+b)\} \geqslant \kappa\cdot\sqrt{(b/n)\cdot\log n} \;\middle|\; \mathfrak{F}^t\,\right] \geqslant 1/2.$$

That is, there is no load configuration and no probability allocation vector (depending on $\mathfrak{F}^t$) such that the gap is small, both before and at the end of an arbitrary batch.

For notational convenience, we will prove this statement by assuming that $t = 0$, and x^0 is an arbitrary load vector satisfying $\sum_{i\in[n]} x_i^0 = 0$ (in other words, we shift time backwards

by t steps) and $p = p^0$ is the probability allocation vector used by the process. Consider one arbitrary bin $j \in [n]$. Then,

$$\mathbf{E}\left[x_j^b - \frac{b}{n} + z_j\right] = b \cdot p_j - \frac{b}{n} + z_j =: \varphi_j$$

For a sufficiently large constant $C > 0$, let us now assume $\max_{j\in[n]} z_j \leqslant C/2 \cdot \sqrt{(b/n)\cdot \log n}$; clearly, if this is not the case, we already have a large gap already before the next batch.

Next consider a bin $j \in [n]$ with

$$p_j \leqslant \frac{1}{n} + \frac{1}{b} \cdot \left(-10C \cdot \sqrt{(b/n)\cdot \log n}\right).$$

We will now apply a Chernoff bound (Lemma A.4) for $x_j^b \sim \mathsf{Bin}(b, p_j)$, with $\delta := C\cdot\sqrt{(n/b)\cdot \log n}$, $\mu := b \cdot p_j$ and $\mu_H := \frac{b}{n} - 10C \cdot \sqrt{(b/n)\cdot \log n} \geqslant \mu$ to get that

$$\mathbf{Pr}\left[x_j^b \geqslant \frac{b}{n} - C \cdot \sqrt{(b/n)\cdot \log n}\right] \leqslant \mathbf{Pr}\left[x_j^b \geqslant \mu_H \cdot (1+\delta)\right] \leqslant e^{-\delta^2\mu_H/3} = e^{-\frac{C^2}{3}\cdot \log n} \leqslant n^{-4},$$

using that $C \geqslant 4$. If $\left\{x_j^b \leqslant \frac{b}{n} - C \cdot \sqrt{(b/n)\cdot \log n}\right\}$ occurs, then

$$x_j^b - \frac{b}{n} + z_j \leqslant -C \cdot \sqrt{(b/n)\cdot \log n} + z_j \leqslant -C/2 \cdot \sqrt{(b/n)\cdot \log n} \leqslant 0,$$

and thus bin j will not contribute to the gap at step b.

Hence in the remainder of the proof, we would like to assume that for all bins $j \in [n]$,

$$p_j \geqslant \frac{1}{n} + \frac{1}{b} \cdot \left(-10C \cdot \sqrt{(b/n)\cdot \log n}\right) =: p_{\text{low}}$$

Note $p_{\text{low}} \leqslant 1/n$. Consider now a transformation of the probability vector $(p_i)_{i\in[n]}$ into $(\widetilde{p}_i)_{i\in[n]}$, where $\widetilde{p}$ satisfies for all $j \in [n]$,$p_{\text{low}} \leqslant \widetilde{p}_j \leqslant \max\{p_{\text{low}}, p_j\}$. In other words, in $\widetilde{p}$ we only increase probabilities of bins $j \in [n]$, for which $p_j < p_{\text{low}}$. Let us define $\mathcal{J} := \{j \in [n] \colon p_j < p_{\text{low}}\}$. For $b \geqslant (20C^2)n \log n$, this implies $\widetilde{p}_j \geqslant \frac{1}{2n}$ for all $j \in [n]$.

Further, let $(x_i^b)_{i\in[n]}$ be a load vector where the locations of the next b balls are sampled according to p, and $(\widetilde{x}_i^b)_{i\in[n]}$ be a load vector where these locations are sampled according to $\widetilde{p}$. Clearly, there is a coupling so that for every $j \in [n] \setminus \mathcal{J}$, $x_j^b \geqslant \widetilde{x}_j^b$ (since $p_j^b \geqslant \widetilde{p}_j^b$). Further, for any $j \in \mathcal{J}$, by a union bound,

$$\mathbf{Pr}\left[\max_{j\in\mathcal{J}} x_j^b > 0\right] \leqslant |\mathcal{J}| \cdot n^{-4} \leqslant n^{-3}.$$

Hence it follows that, for any threshold $T > 0$,

$$\mathbf{Pr}\left[\max_{j\in[n]} x_j^b \geqslant T\right] \geqslant \mathbf{Pr}\left[\max_{j\in[n]} \widetilde{x}_j^b \geqslant T\right] - \mathbf{Pr}\left[\max_{j\in\mathcal{J}} x_j^b \geqslant T\right] \geqslant \mathbf{Pr}\left[\max_{j\in[n]} \widetilde{x}_j^b \geqslant T\right] - n^{-3}.$$

Therefore, in the remainder of the proof, we will lower bound $\mathbf{Pr}\left[\max_{j\in[n]\setminus\mathcal{J}} \widetilde{x}_j^b \geqslant T\right]$ for a suitable value of $T = \Omega(\sqrt{b/n \cdot \log n})$. We will also use the definition

$$\widetilde{\varphi}_j := b \cdot \widetilde{p}_j - \frac{b}{n} + z_j.$$

Finally, we define $\xi = 0.1$ as a (sufficiently) small constant.

Case 1: We have at least $n - n^\xi$ bins for which $\widetilde{\varphi}_j \leqslant -C\sqrt{b/n}$. Since $\sum_{i\in[n]} \widetilde{\varphi}_i = 0$, this implies that there must be at least one bin with $j \in [n]$ with $\widetilde{\varphi}_j \geqslant \frac{n-n^\xi}{n^\xi} \cdot C \cdot \sqrt{b/n} \geqslant 1/2 \cdot n^{1-\xi} \cdot \sqrt{b/n}$. Further, using that the median of a $\mathsf{Bin}(N, q)$ r.v. is either $\lfloor Nq \rfloor$ or $\lceil Nq \rceil$, then

$$\mathbf{Pr}\left[x_j^b \geqslant \mathbf{E}\left[x_j^b\right] - \frac{1}{4}n^{1-\xi}\right] \geqslant \mathbf{Pr}\left[x_j^b \geqslant \left\lfloor \mathbf{E}\left[x_j^b\right]\right\rfloor\right] \geqslant 1/2,$$

it follows that with probability at least $1/2$ we will have a large gap.

Case 2: We have at least n^ξ bins with $\widetilde{\varphi}_j \geqslant -C\sqrt{b/n}$; call this set $\mathcal{B}$. We further know that, due to the definition of $\widetilde{p}$, we have for all bins $j \in [n]$ that $\widetilde{p}_j \geqslant \frac{1}{2n}$. Hence, we set $T := b\cdot\widetilde{p}_j + \kappa\cdot\sqrt{b\cdot\widetilde{p}_j\cdot\log n}$, and applying Lemma A.3 yields for any bin $j \in \mathcal{B}$, $\mathbf{Pr}\left[\widetilde{x}_j^b \geqslant T\right] \geqslant n^{-\xi/2}$. Since $\widetilde{p}_j \geqslant \frac{1}{2n}$, $|\mathcal{B}| \geqslant n^\xi$, the claim follows. □

6 Experimental Results

In this section, we complement our theoretical analysis with some experimental results for the b-Batched setting. In Fig. 6.1, we plot the gap of the $(1+\beta)$-process for various batch sizes and different values of $\beta \in (0, 1]$ (Two-Choice corresponding to $\beta = 1$). The plot strongly suggests the existence of an optimal β, which seems to increase as the batch size b grows.

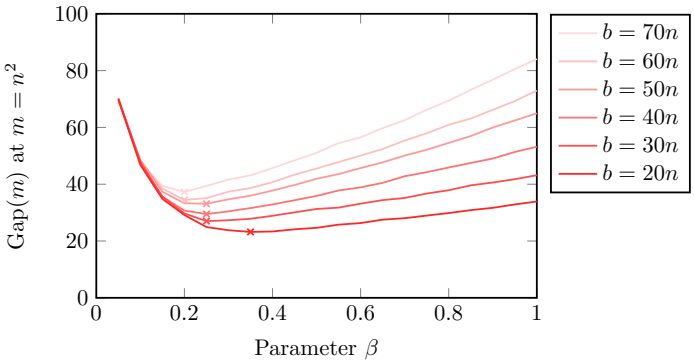


Figure 6.1: Average gap for the $(1+\beta)$-process in the b-Batched setting with unit weights for $n = 1.000$ bins, $m = n^2$ balls, for various batch sizes and parameter values $\beta \in (0, 1]$ (averaged over 25 runs).

In Fig. 6.2, we present the corresponding empirical results of Fig. 6.1 for the Quantile process (mixed with One-Choice). As with the $(1+\beta)$-process, the optimal mixing factor η tends to increase as the batch size grows. The Quantile with the optimized mixing factor seems to perform slightly worse than the optimized $(1+\beta)$-process.

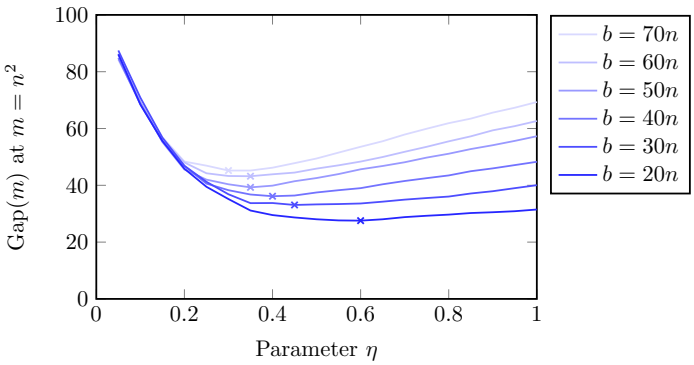


Figure 6.2: Average gap for the Quantile process (mixed with One-Choice with probability $\eta \in (0, 1]$) in the b-Batched setting with unit weights for $n = 1.000$ bins, $m = n^2$ balls, for various batch sizes and parameter values $\eta \in [0, 1]$ (averaged over 25 runs).

In Fig. 6.3, we plot the gap of Two-Choice, Three-Choice and $(1+\beta)$ versus the batch size. For small values of b, the gap of Two-Choice and Three-Choice is small, but soon grows rapidly, diverging from the asymptotically optimal $(1+\beta)$-processes as predicted by the theoretical analysis. Similar, results are observed for weights sampled from an exponential distribution Fig. 6.4.

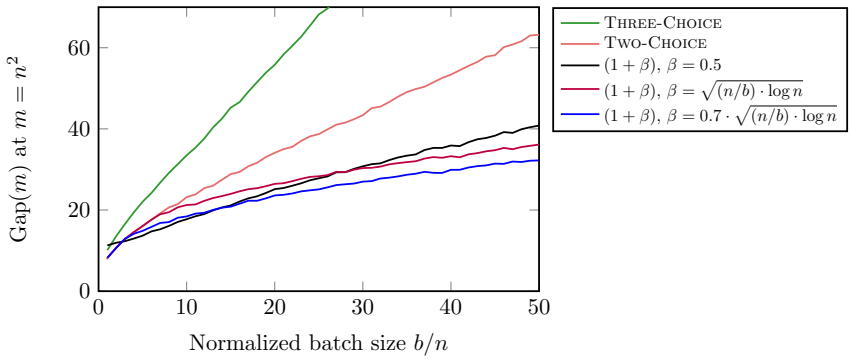


Figure 6.3: Average gap for Three-Choice, Two-Choice and $(1+\beta)$ with $\beta = 0.5$, $\beta = \sqrt{(n/b) \cdot \log n}$ and $\beta = 0.7 \cdot \sqrt{(n/b) \cdot \log n}$ in the b-Batched setting with unit weights for $n = 1.000$ bins, $m = n^2$ balls vs batch size $b \in \{n, \ldots, 50n\}$ (averaged over 50 runs).

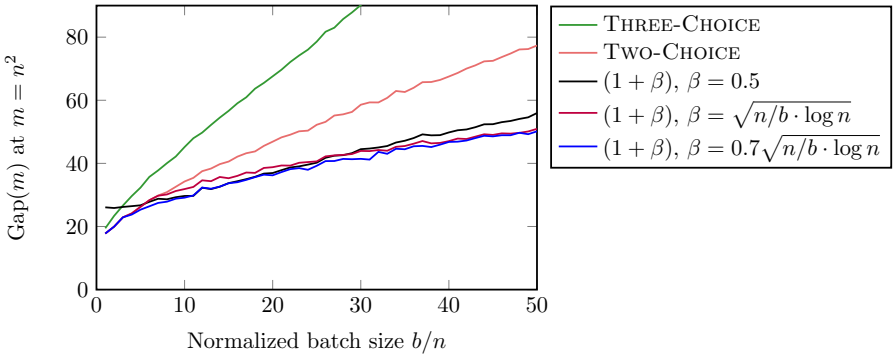


Figure 6.4: Average gap for Three-Choice, Two-Choice and $(1+\beta)$ with $\beta = 0.5$, $\beta = \sqrt{(n/b) \cdot \log n}$ and $\beta = 0.7 \cdot \sqrt{(n/b) \cdot \log n}$ in the b-Batched setting with weights from an $\mathsf{Exp}(1)$ distribution for $n = 1.000$ bins, $m = n^2$ balls vs batch size $b \in \{n, \ldots, 50n\}$ (averaged over 50 runs).

Finally, in Table 6.5, we show the gap of the $(1+\beta)$ and Quantile compared to Two-Choice and One-Choice with b balls (which is the theoretically optimal attainable value), for slightly larger values of $n \in \{10^4, 10^5\}$. The for large b, the $(1+\beta)$ has roughly half the gap of Two-Choice and is close to the theoretically optimal value of One-Choice for $m = b$ balls.

		Two-Choice	Quantile	$(1+\beta)$	One-Choice* for $m = b$
$n = 10^4$	$b = 20n$	36.45	30.15	**26.60**	19.00
	$b = 50n$	70.10	45.75	**39.00**	29.75
	$b = 80n$	100.85	55.65	**46.80**	35.80
$n = 10^5$	$b = 20n$	39.90	34.1	**29.95**	22.40
	$b = 50n$	75.55	50.3	**44.20**	34.30
	$b = 80n$	111.10	64.9	**55.20**	41.95

Table 6.5: Average gap for Two-Choice, Quantile (with $\eta = \sqrt{(n/b) \cdot \log n}$) and $(1+\beta)$ (with $\beta = 0.7\sqrt{(n/b) \cdot \log n}$) in the b-Batched setting with $b \in \{20n, 50n, 80n\}$ and $n \in \{10^4, 10^5\}$ (averaged over 20 runs). The last column gives the average gap for One-Choice with $m = b$ balls which is the theoretically optimal attainable value.

7 Conclusions

In this work, we revisited the outdated information setting of [5], where balls are allocated to bins in batches of size b, using the load information available at the beginning of the batch. We established that by defining the mixing factor β carefully as a function of the batch size b, $(1+\beta)$ achieves the asymptotically optimal gap for any $b \geqslant n \log n$. That is, by having β chosen appropriately small, $(1+\beta)$ circumvents the "herd behavior" (as called in [27]), where some of the previously underloaded bins are chosen too frequently, turning them into heavily overloaded bins in the next batch. Similarly, β should also not be too small, as otherwise the process would be too close to One-Choice.

There are several directions for future work. First, recall that our lower bounds apply to a large class of processes which allocate all balls within the *same batch* independently. However, there are processes which allocate multiple balls in a coordinated way. For example, the process of Park [32] draws d samples, and then places into each of the k least loaded bins one ball. It would be interesting to explore the gap of this type of processes in the b-Batched setting. A second avenue is to analyze Two-Thinning processes (and in particular processes that use a fixed load threshold relative to the average) in outdated information settings. An experimental study of threshold processes with outdated information was already conducted in 1989 [25, Figure 8], but no rigorous bounds were proven. A third possibility is to investigate whether the $(1+\beta)$ and related processes are superior to Two-Choice in other settings, like the τ-Delay or random noise settings studied in [22]. Finally, one could study settings where the load information of bins is updated at different rates, depending on the specific bin. In such a setting, when deciding between sampled bins, both their reported load estimates and update rates should be taken into account.

Bibliography

[1] Eitan Altman and Philippe Nain. 1992. 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

[2] 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] Yossi Azar, Andrei Z. Broder, Anna R. Karlin, and Eli Upfal. 1999. Balanced allocations. *SIAM J. Comput.* 29, 1 (1999), 180–200. doi

[4] Nikhil Bansal and Ohad N. Feldheim. 2022. The power of two choices in graphical allocation. In *54th Annual ACM Symposium on Theory of Computing (STOC'22)*. ACM, 52–63. doi

[5] Petra Berenbrink, Artur Czumaj, Matthias Englert, Tom Friedetzky, and Lars Nagel. 2012. Multiple-Choice Balanced Allocation in (Almost) Parallel. In *16th International Workshop on Randomization and Computation (RANDOM'12)*. Springer-Verlag, 411–422. doi

[6] Petra Berenbrink, Artur Czumaj, Angelika Steger, and Berthold Vöcking. 2006. Balanced allocations: the heavily loaded case. *SIAM J. Comput.* 35, 6 (2006), 1350–1385. doi

[7] Petra Berenbrink, Tom Friedetzky, Zengjian Hu, and Russell Martin. 2008. On weighted balls-into-bins games. *Theoret. Comput. Sci.* 409, 3 (2008), 511–520. doi

[8] Michael Dahlin. 2000. Interpreting Stale Load Information. *IEEE Trans. Parallel Distributed Syst.* 11, 10 (2000), 1033–1047. doi

[9] Pamela Delgado, Diego Didona, Florin Dinu, and Willy Zwaenepoel. 2016. Job-aware Scheduling in Eagle: Divide and Stick to Your Probes. In *7th ACM Symposium on Cloud Computing (SoCC'16)*. ACM, 497–509. doi

[10] Pamela Delgado, Florin Dinu, Anne-Marie Kermarrec, and Willy Zwaenepoel. 2015. Hawk: Hybrid Datacenter Scheduling. In *2015 USENIX Annual Technical Conference (USENIX'15)*. USENIX, 499–510.

[11] Christina Delimitrou, Daniel Sánchez, and Christos Kozyrakis. 2015. Tarcil: reconciling scheduling speed and quality in large shared clusters. In *6th ACM Symposium on Cloud Computing (SoCC'15)*. ACM, 97–110. doi

[12] Ohad N. Feldheim and Ori Gurel-Gurevich. 2021. The power of thinning in balanced allocation. *Electron. Commun. Probab.* 26 (2021), Paper No. 34, 8. doi

[13] Ohad N. Feldheim and Jiange Li. 2020. Load balancing under d-thinning. *Electronic Communications in Probability* 25 (2020), Paper No. 1, 13. doi

[14] Armando Fox, Steven D. Gribble, Yatin Chawathe, Eric A. Brewer, and Paul Gauthier. 1997. Cluster-Based Scalable Network Services. In *16th ACM Symposium on Operating Systems Principles (SOSP'97)*. ACM, 78–91. doi

[15] Richard M. Karp, Michael Luby, and Friedhelm Meyer auf der Heide. 1996. Efficient PRAM simulation on a distributed memory machine. *Algorithmica* 16, 4-5 (1996), 517–542. doi

[16] Krishnaram Kenthapadi and Rina Panigrahy. 2006. Balanced allocation on graphs. In *17th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'06)*. SIAM, 434–443. doi

[17] Mansour Khelghatdoust and Vincent Gramoli. 2018. Peacock: Probe-Based Scheduling of Jobs by Rotating Between Elastic Queues. In *24th International Conference on Parallel and Distributed Computing (Euro-Par'18)*, Vol. 11014. Springer, 178–191. doi

[18] Joy Kuri and Anurag Kumar. 1995. Optimal control of arrivals to queues with delayed queue length information. *IEEE Trans. Automat. Control* 40, 8 (1995), 1444–1450. doi

[19] Samuel Kutin. 2002. *Extensions to McDiarmid's inequality when differences are bounded with high probability.* Technical Report. University of Chicago.

[20] 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, 389–399. doi

[21] 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)*, Vol. 215. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 103:1–103:23. doi

[22] 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, 164–175. doi

[23] 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, 45:1–45:22. doi

[24] Yi Lu, Qiaomin Xie, Gabriel Kliot, Alan Geller, James R. Larus, and Albert G. Greenberg. 2011. Join-Idle-Queue: A novel load balancing algorithm for dynamically scalable web services. *Perform. Evaluation* 68, 11 (2011), 1056–1071. doi

[25] Ravi Mirchandaney, Don Towsley, and John A. Stankovic. 1989. Analysis of the Effects of Delays on Load Sharing. *IEEE Trans. Comput.* 38, 11 (nov 1989), 1513–1525. doi

[26] Michael Mitzenmacher. 1999. On the analysis of randomized load balancing schemes. *Theory Comput. Syst.* 32, 3 (1999), 361–386. doi

[27] Michael Mitzenmacher. 2000. How Useful Is Old Information? *IEEE Trans. Parallel Distributed Syst.* 11, 1 (2000), 6–20. doi

[28] 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., Netherlands, 255–312. doi

[29] Muhammad Anis Uddin Nasir, Gianmarco De Francisci Morales, David García-Soriano, Nicolas Kourtellis, and Marco Serafini. 2015. The power of both choices: Practical load balancing for distributed stream processing engines. In *31st IEEE International Conference on Data Engineering (ICDE'15)*. IEEE, 137–148. doi

[30] Muhammad Anis Uddin Nasir, Gianmarco De Francisci Morales, Nicolas Kourtellis, and Marco Serafini. 2016. When two choices are not enough: Balancing at scale in Distributed Stream Processing. In *32nd IEEE International Conference on Data Engineering (ICDE'16)*. IEEE, 589–600. doi

[31] Kay Ousterhout, Patrick Wendell, Matei Zaharia, and Ion Stoica. 2013. Sparrow: distributed, low latency scheduling. In *24th ACM SIGOPS Symposium on Operating Systems Principles (SOSP'13)*. ACM, 69–84. doi

[32] Gahyun Park. 2011. A generalization of multiple choice balls-into-bins. In *30th Annual ACM-SIGOPT Principles of Distributed Computing (PODC'11)*. ACM, 297–298. doi

[33] Yuval Peres, Kunal Talwar, and Udi Wieder. 2015. Graphical balanced allocations and the $(1+\beta)$-choice process. *Random Structures & Algorithms* 47, 4 (2015), 760–775. doi

[34] Martin Raab and Angelika Steger. 1998. "Balls into bins"—a simple and tight analysis. In *2nd International Workshop on Randomization and Computation (RANDOM'98)*, Vol. 1518. Springer, 159–170. doi

[35] Kunal Talwar and Udi Wieder. 2007. Balanced allocations: the weighted case. In *39th Annual ACM Symposium on Theory of Computing (STOC'07)*. ACM, 256–265. doi

[36] Zhijun Wang, Huiyang Li, Zhongwei Li, Xiaocui Sun, Jia Rao, Hao Che, and Hong Jiang. 2019. Pigeon: an Effective Distributed, Hierarchical Datacenter Job Scheduler. In *10th ACM Symposium on Cloud Computing (SoCC'19)*. ACM, 246–258. doi

[37] Ward Whitt. 1986. Deciding Which Queue to Join: Some Counterexamples. *Oper. Res.* 34, 1 (1986), 55–62. doi

[38] Udi Wieder. 2017. Hashing, Load Balancing and Multiple Choice. *Found. Trends Theor. Comput. Sci.* 12, 3-4 (2017), 275–379. doi

A Tools

A.1 Auxiliary Probabilistic Claims

For convenience, we add the following well-known inequality for a sequence of random variables, whose expectations are related through a recurrence inequality.

Lemma A.1. *Consider a sequence of random variables $(X^i)_{i\in\mathbb{N}}$ such that there exist $a \in (0,1)$ 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, for every $i \geqslant 1$,

$$\mathbf{E}\left[\, X^i \mid X^0 \,\right] \leqslant X^0 \cdot a^i + \frac{b}{1-a}.$$

Proof. 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 $i = 0$, it trivially holds that $\mathbf{E}\left[\, X^0 \mid X^0 \,\right] \leqslant X^0$. 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 for $a \in (0,1)$, $\sum_{j=0}^{\infty} a^j = \frac{1}{1-a}$. □

For the next lemma, we define for two n-dimensional vectors x, y, $\langle x, y \rangle := \sum_{i=1}^{n} x_i \cdot y_i$.

Lemma A.2 ([20, Lemma A.7]). *Let $(p_k)_{k=1}^n, (q_k)_{k=1}^n$ be two probability vectors and $(c_k)_{k=1}^n$ be non-negative and non-increasing. Then if p majorizes q, i.e., for all $1 \leqslant k \leqslant n$, $\sum_{i=1}^{k} p_i \geqslant \sum_{i=1}^{k} q_i$ holds, then*

$$\langle p, c \rangle \geqslant \langle q, c \rangle.$$

We continue with an "anti-concentration" result, i.e., a lower bound on the probability that a binomial random variable is significantly larger than its expectation.

Lemma A.3. *Let m, n be integers such that $m \geqslant n \log n$. Further, let p be a probability satisfying $p \in [1/(2n), 1/2]$ and let $X \sim \mathsf{Bin}(m, p)$. Then for any constant $\xi \in (0,1)$, there exists a constant $\kappa \geqslant 0$, such that*

$$\mathbf{Pr}\left[\, X \geqslant m \cdot p + \kappa \cdot \sqrt{mp \cdot \log n} \,\right] \geqslant n^{-\xi/2}.$$

Proof. Since $X \sim \mathsf{Bin}(m, p)$, we know that

$$\mathbf{Pr}\left[\, X = m \cdot p \,\right] = \binom{m}{m \cdot p} \cdot p^{m \cdot p} \cdot (1-p)^{m - m \cdot p}.$$

Let $\mu := m \cdot p$ and $f(z) := \mathbf{Pr}\left[\, X = z \,\right]$ for any $z \in [0, m]$. Then, for any integer $k \geqslant 1$,

$$\begin{aligned}
\frac{f(\mu+k)}{f(\mu+(k-1))} &= \frac{\frac{m!}{(\mu+k)!(m-\mu-k)!}}{\frac{m!}{(\mu+k-1)!(m-\mu-k+1)!}} \cdot \frac{p}{1-p} \\
&= \frac{m-\mu-k+1}{\mu+k} \cdot \frac{p}{1-p} \\
&= \frac{m \cdot (1 - \frac{\mu+k-1}{m})}{(1+\frac{k}{\mu})m \cdot p} \cdot \frac{p}{1-p} \\
&= \frac{1 - \frac{\mu+k-1}{m}}{1+\frac{k}{\mu}} \cdot \frac{1}{1-p} \\
&= \frac{1 - p \cdot \frac{\mu+k-1}{\mu}}{1+\frac{k}{\mu}} \cdot \frac{1}{1-p}.
\end{aligned}$$

The first factor is decreasing in k. Hence, for any $K \geqslant 1$,

$$\begin{aligned}
f(\mu+K) &\geqslant f(\mu) \cdot \left(\frac{\mu \cdot (1-p) - (K-1)p}{\mu+K}\right)^K \cdot \left(\frac{1}{1-p}\right)^K \\
&= f(\mu) \cdot \left(\frac{(\mu+K)\cdot(1-p) - K(1+p)}{(\mu+K)\cdot(1-p)}\right)^K \\
&= f(\mu) \cdot \left(1 - \frac{K(1+p)}{(\mu+K)\cdot(1-p)}\right)^{(\frac{(\mu+K)\cdot(1-p)}{K(1+p)}-1)\cdot\frac{K}{\frac{(\mu+K)\cdot(1-p)}{K(1+p)}-1}}
\end{aligned}$$

We will now make use of the fact that $(1-\frac{1}{N})^{N-1} \geqslant e^{-1}$, where $N := \frac{(\mu+K)\cdot(1-p)}{K\cdot(1+p)}$. With this we have,

$$\begin{aligned}
f(\mu+K) &= f(\mu) \cdot \left(1 - \frac{1}{N}\right)^{(N-1)\cdot\frac{K}{N-1}} \\
&\geqslant f(\mu) \cdot e^{-\frac{K}{\frac{(\mu+K)\cdot(1-p)}{K(1+p)}-1}} \\
&\geqslant f(\mu) \cdot e^{-\frac{2K}{\frac{(\mu+K)\cdot(1-p)}{K(1+p)}}} \\
&\geqslant f(\mu) \cdot e^{-\frac{8K^2}{(\mu+K)}} \\
&\geqslant f(\mu) \cdot e^{-\frac{16K^2}{\mu}},
\end{aligned}$$

having used that $p \leqslant 1/2$ and $K \leqslant 4 \cdot \mu$. Further, by a concentration bound, e.g., by Chebyshev's inequality,

$$\mathbf{Pr}\left[\, |X-\mu| \geqslant 2\sqrt{\mathbf{Var}\left[\, X \,\right]} \,\right] \leqslant \frac{1}{4}.$$

This implies that at least $\frac{3}{4}$ of the probability of X is on the interval $[\mu - 2\sqrt{\mathbf{Var}\left[\, X \,\right]}, \mu + 2\sqrt{\mathbf{Var}\left[\, X \,\right]}]$. Since the mode, i.e., largest probability is $\lfloor (m+1)p \rfloor$ or $\lceil (m+1)p \rceil - 1$, it follows that

$$f(\mu) \geqslant \frac{3}{8 \cdot \mathbf{Var}\left[\, X \,\right]} = \frac{3}{8 \cdot \sqrt{m \cdot p(1-p)}} \geqslant \frac{3}{8\sqrt{\mu}}.$$

Then, it follows that

$$\mathbf{Pr}\left[\, \widetilde{x}_j \geqslant \mu + \xi/2 \cdot \sqrt{\mu \log n} \,\right] \geqslant \sum_{K=\xi/100\cdot\sqrt{\mu\log n}}^{m} f(\mu+K)$$

$$
\begin{aligned}
&\geqslant \sum_{K=\xi/100\cdot\sqrt{\mu\log n}}^{\xi/50\cdot\sqrt{\mu\log n}} f(\mu+K) \\
&\geqslant f(\mu)\cdot \sum_{K=\xi/100\cdot\sqrt{\mu\log n}}^{\xi/50\cdot\sqrt{\mu\log n}} e^{-\frac{16K^2}{\mu}} \\
&\geqslant \frac{3}{8\sqrt{\mu}}\cdot \xi/100\cdot\sqrt{\mu\log n}\cdot e^{-16\xi^2/2500\cdot\log n} \\
&\geqslant n^{-32/2500\xi^2} \geqslant n^{-\xi/2}.
\end{aligned}
$$

□

A.2 Concentration Inequalities

We now proceed by stating a standard Chernoff bound.

Lemma A.4 (Chernoff Bound). *Let $X^1,\dots,X^n$ be independent random variables taking values in $\{0,1\}$. Let $X=\sum_{i=1}^n X^i$, $\mu=\mathbf{E}[X]$ and $\delta\geqslant 0$ be arbitrary. Then, for any $\mu_H\geqslant\mu$,*

$$\mathbf{Pr}[X\geqslant(1+\delta)\cdot\mu_H]\leqslant e^{-\delta^2\mu_H/3}.$$

Following [19], we will now give the definition for *strongly difference-bounded* and then give the statement for a bounded differences inequality with bad events.

Definition A.5 (Strongly difference-bounded – [19, Definition 1.6]). *Let $\Omega_1,\dots,\Omega_N$ 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 A.6 ([19, Theorem 3.3]). *Let $\Omega_1,\dots,\Omega_N$ 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 $\gamma_1,\dots,\gamma_N>0$,*

$$\mathbf{Pr}[X\geqslant\mu+\lambda]\leqslant\exp\left(-\frac{\lambda^2}{2\cdot\sum_{k=1}^N(\eta_2+\eta_1\gamma_k)^2}\right)+\xi\cdot\sum_{k=1}^N\frac{1}{\gamma_k}.$$