Unbalanced allocations

by

Amanda E. Redlich

B.A., University of Chicago (2005)

Submitted to the Department of Mathematics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2010

Author ....
Department of Mathematics
April 2, 2010

Certified by ....
Peter Shor
Morss Professor of Applied Mathematics
Thesis Supervisor

Accepted by ....
Michel X. Goemans
Chairman, Applied Mathematics Committee

Accepted by ....
Bjorn Poonen
Chairman, Department Committee on Graduate Students

Unbalanced allocations

by

Amanda E. Redlich

Submitted to the Department of Mathematics
on April 2, 2010, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

Recently, there has been much research on processes that are mostly random, but also have a small amount of deterministic choice; e.g., Achlioptas processes on graphs. This thesis builds on the balanced allocation algorithm first described by Azar, Broder, Karlin and Upfal. Their algorithm (and its relatives) uses randomness and some choice to distribute balls into bins in a balanced way. Here is a description of the opposite family of algorithms, with an analysis of exactly how unbalanced the distribution can become.

Thesis Supervisor: Peter Shor
Title: Morss Professor of Applied Mathematics

Acknowledgments

Thank you to my advisor, Peter Shor, for mathematical, philological, and financial aid. Thank you to my previous advisors, Igor Pak and Santosh Vempala, for guiding me through my first years of graduate school.

Thank you to John Augustine, Shirley Entzminger, Susan Ferguson, Michelle Gallarelli, Jeff Kahn, Daniel Kleitman, Sachin Lodha, Linda Okun, Tony Pelletier, Jaikumar Radhakrishnan, M. V. Panduranga Rao, Sasanka Roy, Michael Sipser, Cliff Smyth, and Richard Stanley for helpful conversations during my graduate career.

Thank you to those who taught me, especially László Babai, Andrew Blumberg, Diane Herrmann, and Paul Sally at the University of Chicago; María Teresa Holcomb, Teresa Kanoza, William McCall, and David Sykes at Lincoln Land Community College; Gary Gordon, David Kelly, Rebecca Murphy, Marcia Van Pelt, Douglas West, and Kate Wilson; Joel Spencer, who was my math teacher the longest; and James Redlich, who was my math teacher first.

Thank you to the friends and family who supported me: at MIT, my classmates, officemates, and roommates, my Bollywood friends and the thesis writing club; my local family Susan and Elizabeth Redlich and Sachie, Santosh, and Ashwin Karmacharya; Bernie Carver, Sally Millichamp, Michael Houser, and Chris and Marty Reynolds in Springfield; friends of ALUUC, residents of Hawthorne Place and staff of Lincoln Library; my family in Illinois, Missouri, and Wisconsin; my friends in East Setauket, Moorefield, Nashville, New York, St. Paul, and Toronto; my parents Myra Epping and James Redlich; and my sister Margaret Redlich.

Most importantly, I thank my grandparents William and Anna Rose Redlich and Jack and Jo Epping.

Contents

List of Tables

Chapter 1

Introduction

Suppose you have a set of n bins, and you want to distribute m balls into these bins. One way to do so would be to choose a bin uniformly at random for each ball. In the limit, this would produce a Poisson distribution. The expected size of each bin would be m/n. However, bin sizes would range from very small to very large. Recall that under the Poisson distribution the probability of a particular bin having size k is

$$\frac{(\frac{m}{n})^k e^{-m/n}}{k!}.$$

So the expected number of empty bins would be $ne^{-m/n}$. To bound the bin sizes from above, note that, if $m = cn$, the probability of a bin existing of size $\frac{\gamma \log n}{\log \log n}$ is bounded by $n^{1-\gamma+\gamma/\log\log n}$. To bound this probability by $1/n$, it's enough to set $\gamma = 2 + \epsilon$. So we see that almost surely all bins are of size $O(\log n/\log\log n)$. On the other hand, with probability

$$(1/n)(n^{(1/\log\log n)(\log c+1+\log\log\log n)})(\sqrt{\frac{\log\log n}{2\pi\log n}}) \geq (1/n)(\log\log n),$$

there is a bin of size $\log n/\log\log n$.

Suppose you want a more even distribution. For example, if balls are service requests and bins are servers, then you would want to minimize wait time, i.e. bin size. You could try a new method, first described by Azar, Broder, Karlin, and Upfal in [2]: Place the balls sequentially. For the i^{th} ball, select a few (say some constant d) bins uniformly at random, then place the ball in the currently least full of these bins.

This flattens the distribution. It reduces the maximum bin's growth rate from $\log n / \log \log n$ to $\log \log n$. It also decreases the number of empty bins (when $n = m$) from $n/e \approx .37n$ to $.34n$ or less, depending on d.

Furthermore, this is easy to implement. It doesn't require memory of prior ball placements. That is, there is no need to remember previous decisions, as long as bin sizes can be queried. It doesn't require much randomness; there are only dm random choices necessary to place all m balls. It is robust against attacks. For example, when the number of balls is cn, if an adversary were to destroy a bin, there is no strategy that would damage more than c balls in expectation and $O(\log n / \log \log n)$ in the worst case.

Many variations have been explored since this algorithm was first introduced. Initially, it was studied with n bins and $m = cn$ balls (c an arbitrary constant), the number of options d some small constant, and all choices made uniformly at random (i.e. the selection of the d options, and any tie-breaking). Since then, researchers have considered several other cases. To study the case m much greater than n in [3], new proof techniques were developed. In the course of that paper, it was also discovered that asymmetrical tie-breaking can make the overall distribution more level; this notion was studied further in [9].

Often, real-world implementations have further constraints; not every ball (service request) can go in every bin (server). The option sets cannot be uniformly distributed. This case, of d options under different distributions, has been studied in [6] and [5].

Each of these cases has natural real-world applications. In all these applications, the goal is the same: an algorithm using a small amount of randomness and a small amount of bin inspection that produces a flat distribution, with the balls as evenly spaced as possible.

My problem grew out of a different application. Suppose each ball is charged rent for using a bin. Furthermore, suppose the rent is a buy-at-bulk function: placing ten balls in the same bin costs less than placing five balls in two bins. Then it would be beneficial to have many empty bins, and a few bins with many balls in them. This models many real world problems; for example, purchasing software licenses for a large company that needs many applications.

A natural algorithm would be to choose d option bins for each ball, but place the ball in the most (not least) full bin among them. This thesis will focus on analyzing the distribution created by this process. In the next chapter, I analyze what happens when the number of balls and number of bins are roughly equal. The third chapter includes theorems about the distribution that hold for arbitrary numbers of balls and bins. The last chapter considers the case when the number of balls is much greater than the number of bins.

1.1 Definitions

More formally, the distribution analyzed here is

Definition 1. *Let* GREEDY(m, n) *be the algorithm as follows. For time* $t = 1$ *through time* $t = m$*, select uniformly at random from* $[n] = \{1, 2, \ldots n\}$ d *times, for some predetermined* d*. Call the set of at most* d *indices selected at time* t S_t*. Let the bins be labeled* $B_1 \ldots B_n$*. Let the number of balls in* B_i *at time* t *be* $b_i(t)$*. Place the* t^{th} *ball into* B_M*, where* $b_M(t) = \max_{i \in S_t}\{b_i(t)\}$*. If there is a tie, break it uniformly randomly.*

Throughout, we use m for the number of balls and n for the number of bins. We consider the asymptotic behavior as n goes to infinity. Usually $m \geq n$ and d is a small constant (e.g., 2), but many results hold even when this is not the case. I will note which restrictions on m and d are necessary throughout the text.

Often, it is convenient to compare GREEDY with other distributions. For ease of notation, we will call these FAIR (the distribution first described in [2]) and UNIFORM (the uniform distribution).

Definition 2. *Let* FAIR(m, n) *be the algorithm as follows. For time* $t = 1$ *through* *time* $t = m$, *select uniformly at random from* $[n]$ d *times. Let* S_t *be the set of at most* d *indices selected at time* t. *Place the* t^{th} *ball into* B_m, *where* $b_m(t) = \min_{i \in S_t}\{b_i(t)\}$. *If there is a tie, break it uniformly.*

Definition 3. *Let* $\mathrm{UNIFORM}(m, n)$ *be the algorithm as follows. For time* $t = 1$ *through time* $t = m$, *select uniformly at random from* $[n]$ *once. Place the* t^{th} *ball into the bin selected.*

Chapter 2

The case $m = cn$

We first consider the case $m = cn$. Many proof techniques that were used in earlier work on FAIR are also applicable in this case (see e.g. [8], [3], [9]). The differential equations ideas used in [8] are used here for computing the expected number of bins of any constant size. Previous work on UNIFORM is also useful (see e.g. [1], [4]).

2.1 Differential equation techniques

A natural question to ask about GREEDY is how well it achieves the cost minimization goal. As smaller bins are more expensive, we would hope GREEDY has most of its balls in bins of large size, with a few bins of smaller size, and many bins that have no balls at all. We first calculate how many bins of small size there are.

At first glance, this seems easy. Certainly it's easy to analyze how many times a bin is an option. The expected value is d/n, with some lucky bins being options up to $\Omega(\log n / \log\log n)$ times. However, it is possible that a bin could be an option many times without being chosen. If every time the bin is an option, there is another bin in the option set that has more balls, then the bin will end with zero balls. So the size of each bin depends on how many times it is an option and on its relative size each time it is an option.

To analyze this more complicated dependency, we turn to Kurtz's theorem (see Appendix A). The idea of applying Kurtz's Theorem in this context was first developed in [8]. We first use Kurtz's theorem to calculate the expected number of empty bins. We then turn to the more complicated case of bins of arbitrary constant size.

Theorem 2.1.1. *For the case $m = cn$, the expected number of empty bins after all the balls have been distributed is* $(cd - c + 1)^{1/(1-d)}$.

Proof. We will first calculate the expected change in the number of empty bins at each time step, then convert that formula into a differential equation, then solve the differential equation. The solution to the differential equation will also be the desired expected value.

First, let's consider the number of empty bins at an arbitrary time. For convenience we will rescale time so that at time t, tn balls have been distributed. Let $y(t)$ be the proportion of empty bins at time t. Then $y(0)$ is how what fraction of bins are empty before any balls have been placed. Since all bins are empty initially, $y(0) = 1$.

We know that $y(t + 1/n)$ depends on $y(t)$. Recall that our algorithm chooses an option set, then places the ball in the bin in the option set with the most balls in it already. The number of empty bins changes only if a ball is placed in a previously empty bin. In that case, the number decreases.

But a ball would be placed in an empty bin only if an empty bin were the fullest among the options, i.e., if all the bins in the option set were empty. The probability of this happening at time $t + 1/n$ is $y(t)^d$. So the expected change in y from time t to time $t + 1/n$ is $-y(t)^d$.

We now translate this into the language of Kurtz's theorem[1]. As in the appendix, the i^{th} coordinate of X_n will represent the proportion of bins, out of n, that have size i. Since we know that the change in empty bins depends only on the proportion of empty bins, we may truncate $\{X_n\}$ after the first coordinate. So the state space of X_n is contained within $\{n^{-1}k | k \in \mathbb{N}\}$. There is only one possible transition, $\{-1\} = L$. So we need only find $\beta_{-1} = \beta$. By the discussion in the previous paragraph, $\beta(x) = -x^d$.

We now check that the conditions of Kurtz's theorem are satisfied. Since $F(x) = \beta(x)$, the Lipschitz condition is satisfied by M such that $\left|x^d - y^d\right| \leq M \left|x - y\right|$ for all

[1]See Appendix A for a more detailed discussion of Kurtz's theorem. The presentation here will follow the structure of the example there

x and y in the state space. Since x and y must be between 0 and 1, it's enough to find M such that $\left|x^d - y^d\right| \leq M\left|x - y\right|$ for all pairs x and y in $[0,1]$.

Recall that

$$\left|x^d - y^d\right| = |x - y|\left|\sum_{i=0}^{d-1} x^i y^{d-1-i}\right|$$

For x and y in $[0,1]$, $\left|\sum_{i=0}^{d-1} x^i y^{d-1-i}\right| \leq d$. Therefore setting $M = d$ is enough to satisfy the Lipschitz condition.

We now check that $\lim_{n\to\infty} X_n(0) = x_0$ for some constant x_0. Since initially all bins are empty, $X_n(0) = 1$ for all n. Therefore $\lim_{n\to\infty} X_n(0) = 1$, and the second condition is satisfied.

To find a neighborhood K around $\{X(u)|u \leq t\}$, we first need to find X. Recall that $X(t) = x_0 + \int_o^t F(X(u))du$. We know that $F(x) = -x^d$, so we compute X by solving the differential equation $\dot{x} = -x^d$.

First we find the general solution.

$$\frac{\mathrm{d}x}{x^d} = \mathrm{d}t$$

$$\frac{x^{-(d-1)}}{d-1} + C = t$$

Now we use the initial condition $x(0) = 1$.

$$\frac{1^{-(d-1)}}{d-1} + C = 0$$

$$C = -\frac{1}{d-1}$$

So x is such that

$$-\frac{1}{d-1} + \frac{x^{-(d-1)}}{d-1} = t,$$

and therefore

$$-1 + x^{-(d-1)} = (d-1)t$$

$$x^{-(d-1)} = (d-1)t + 1$$

$$x = ((d-1)t+1)^{-\frac{1}{(d-1)}}.$$

Note that $X = ((d-1)t+1)^{-\frac{1}{(d-1)}}$ is non-negative and decreasing, so for $u \leq t$, $X(u) \leq X(0) = 1$. Therefore we can let $K = [0, 2]$. Then

$$\sum_{l \in L} |l| \sup_{x \in K} \beta_l(x) = \sup_{x \in K} x^d = 2^d,$$

which is finite. So $X_n(c)$ approaches $X(c)$ for any constant c, and the expected proportion of empty bins under GREEDY(cn, n) is $X(c) = ((d-1)c+1)^{-\frac{1}{d-1}}$.

□

Using these results, we can compare statistics for FAIR, UNIFORM, and GREEDY explicitly. For example, the expected number of empty bins after n balls have been distributed is as follows:

Table 2.1: Expected number of empty bins

	GREEDY	UNIFORM	FAIR
$d = 2$	$n/2$	n/e	0.2384
$d = 3$	$n/\sqrt{3}$	n/e	$0.1770n$
d	$nd^{-1/(d-1)}$	n/e	-

Note that, even for $d = 2$, GREEDY is already significantly different from FAIR. More suprisingly, it is different from UNIFORM. The shift from one to two choices increased the number of empty bins significantly. Also note that the proportion of empty bins increases with d. As $d \to \infty$, $d^{-1/(d-1)} \to 1$.

We can use this same technique to compute arbitrary constant bin size statistics. The requirements of Kurtz's theorem are satisfied for any constant bin size and constant time. Since we have scaled time so that at time c, cn balls have been distributed, this means we can compute the expected number of bins of constant size for any number of balls a constant multiple of n.

Theorem 2.1.2. *For the case $m = cn$ and options number d a constant, the expected number of bins of size k under* GREEDY(m, n) *is $y_k(c)$ where y_k satisfies the system* *of differential equations*

$$\{y_i'(t) = 2(y_{i-1}(t) + \ldots + y_0(t))^d - (y_{i-2}(t) + \ldots + y_0(t))^d - (y_i(t) + \ldots + y_0(t))^d\}_{i=2}^k,$$

$$y_1'(t) = 2y_0(t)^d - (y_0(t) + y_1(t))^d,$$

$$y_0'(t) = -y_0(t)^d$$

with initial values $y_0(0) = 1$ *and* $y_i = 0$ *for all* $i > 0$.

Proof. This proof is similar to the previous one. We first informally determine the dependencies of bin size statistics. We then translate this into the language of Kurtz's theorem. We check that the conditions are satisfied, and discover the expected values.

First, compute the dependencies of the bin sizes. Let $y_i(t/n)$ be the proportion of bins of size i at time t. That is, if $b_1(5/4) = 1$, $b_2(5/4) = 1$, $b_3(5/4) = 0$, $b_4(5/4) = 3$, then $y_0(5/4) = 1/4$, $y_1(5/4) = 1/2$, $y_2(5/4) = 0$, $y_3(5/4) = 1/4$. The probability that $y_i(t + 1/n)$ is different from $y_i(t)$ depends on values of $y_j(t)$ for $j \leq i$. The number of bins of size i increases if $S_{t+1/n}$ generates only bins of size at most $i - 1$ and includes at least one bin of size $i - 1$. The number of bins of size i decreases if $S_{t+1/n}$ generates only bins of size at most i, and includes at least one bin of size i.

We know that $L = \{e_i - e_{i-1}\}_{i \in \mathbb{N}}$. Now we need to find β_l. The previous discussion makes it clear that

$$\beta_{e_i - e_{i-1}} = (y_{i-1} + \ldots y_0)^d - (y_{i-2} + \ldots + y_0)^d,$$

because the probability of the option set containing only bins of size at most $i - 1$ is $(y_{i-1} + \ldots y_0)^d$ and the probability of it containing only bins of size less than $i - 1$ is $(y_{i-2} + \ldots + y_0)^d$.

Putting these definitions together, we see that

$$F(y) = \sum_{l \in L} l\beta_l(x) = \sum_{i \in \mathbb{N}} (e_i - e_{i-1})((y_{i-1} + \ldots y_0)^d - (y_{i-2} + \ldots + y_0)^d).$$

The i^{th} coordinate of $F(y)$ is thus

$$2(y_{i-1}(t)+\ldots+y_0(t))^d-(y_{i-2}(t)+\ldots+y_0(t))^d-(y_i(t)+\ldots+y_0(t))^d.$$

As in the previous proof, we truncate X_n to satisfy Kurtz's conditions. Suppose we are interested in calculating the expected number of bins of size k. Then consider only the first $k+1$ coordinates and let $X_n \in \{n^{-1}x | x \in \mathbb{N}^{k+1}\}$. Note that, as before, each coordinate only depends on the coordinates of smaller index. So this truncation is valid.

We now verify that the conditions of Kurtz's theorem are satisfied. As X_n now has $k+1$ coordinates,

$$\begin{aligned}|F(x)-F(y)| = (\sum_{i=1}^{k+1}[2(x_{i-1}(t)+\ldots+x_0(t))^d-(x_{i-2}(t)+\ldots+x_0(t))^d-(x_i(t)+\ldots+x_0(t))^d \\ -2(y_{i-1}(t)+\ldots+y_0(t))^d+(y_{i-2}(t)+\ldots+y_0(t))^d+(y_i(t)+\ldots+y_0(t))^d]^2)^{1/2}.\end{aligned}$$

For ease of notation, let

$$u_i = 2(x_{i-1}+\ldots x_0)^d-(x_{i-2}(t)+\ldots+x_0(t))^d-(x_i(t)+\ldots+x_0(t))^d$$

and

$$v_i = 2(y_{i-1}(t)+\ldots+y_0(t))^d-(y_{i-2}(t)+\ldots+y_0(t))^d-(y_i(t)+\ldots+y_0(t))^d.$$

Then to satisfy the Lipschitz condition we want to find M such that

$$\sum_{i=1}^{k+1}(u_i-v_i)^2 \leq M\sum_{i=0}^{k}(x_i-y_i)^2.$$

We will do this by first fixing j and finding M_j such that

$$(u_j-v_j)^2 \leq M_j\sum_{i=0}^{k}(x_i-y_i)^2.$$

Then letting $M = \sum_{j=1}^{k+1} M_j$ will be sufficient.

Note that

$$u_j - v_j = 2(x_{j-1} + \ldots x_0 - y_{j-1} - \ldots - y_0) \sum_{i=0}^{d-1} (x_{j-1} + \ldots + x_0)^i (y_{j-1} + \ldots + y_0)^{d-1-i}$$
$$-(x_j + \ldots + x_0 - y_j - \ldots - y_0) \sum_{i=0}^{d-1} (x_i + \ldots + x_0)^i (y_i + \ldots + y_0)^{d-1-i}$$
$$-(x_{j-2} + \ldots + x_0 - y_{j-2} - \ldots - y_0) \sum_{i=0}^{d-1} (x_{j-2} + \ldots + x_0)^i (y_{j-2} + \ldots + y_0)^{d-1-i}.$$

Since the x_i and y_i are all between 0 and 1, we know that

$$\sum_{i=0}^{d-1} (x_{j-1} + \ldots + x_0)^i (y_{j-1} + \ldots + y_0)^{d-1-i} \leq dj^{d-1},$$

and similarly

$$\sum_{i=0}^{d-1} (x_j + \ldots + x_0)^i (y_j + \ldots + y_0)^{d-1-i} \leq d(j+1)^{d-1}$$

and

$$\sum_{i=0}^{d-1} (x_{j-2} + \ldots + x_0)^i (y_{j-2} + \ldots + y_0)^{d-1-i} \leq d(j-1)^{d-1}.$$

By the triangle inequality, the overall formula is bounded

$$|u_j - v_j| \leq \left| 2(x_{j-1} + \ldots x_0 - y_{j-1} - \ldots - y_0) \sum_{i=0}^{d-1} (x_{j-1} + \ldots + x_0)^i (y_{j-1} + \ldots + y_0)^{d-1-i} \right|$$
$$+ \left| -(x_j + \ldots + x_0 - y_j - \ldots - y_0) \sum_{i=0}^{d-1} (x_j + \ldots + x_0)^i (y_j + \ldots + y_0)^{d-1-i} \right|$$
$$+ \left| -(x_{j-2} + \ldots + x_0 - y_{j-2} - \ldots - y_0) \sum_{i=0}^{d-1} (x_{j-2} + \ldots + x_0)^i (y_{j-2} + \ldots + y_0)^{d-1-i} \right|$$

So

$$\begin{aligned}|u_j - v_j| \leq 2\left|(x_{j-1} + \ldots + x_0 - y_j - \ldots y_0)\right| dj^{d-1} \\ + \left|(x_j + \ldots x_0 - y_j - \ldots - y_0)\right| d(j+1)^{d-1} \\ + \left|(x_{j-2} + \ldots + x_0 - y_{j-2} - \ldots - y_0)\right| d(j-1)^{d-1}.\end{aligned}$$

Now let m be the index such that $|x_m - y_m| \geq |x_i - y_i|$ for all $i \leq k$. Then by breaking the overall difference $x_i + \ldots + x_0 - y_i - \ldots - y_0$ into pairs $x_r - y_r$, we can use the triangle inequality again and see

$$\begin{aligned}4(k+1)\left|x_m - y_m\right| \geq 2\left|x_{i-1} + \ldots + x_0 - y_{i-1} - \ldots - y_0\right| \\ + \left|x_i + \ldots + x_0 - y_i - \ldots - y_0\right| \\ + \left|x_{i-2} + \ldots + x_0 - y_{i-2} - \ldots - y_0\right|\end{aligned}$$

for all $i \leq k$.

We now combine the previous bounds to see

$$|u_j - v_j| \leq 4(k+1)d(j+1)^{d-1}\left|x_m - y_m\right|.$$

For arbitrary $j \leq k$

$$|u_j - v_j| \leq 4d(k+1)^d\left|x_m - y_m\right|.$$

However, we need to bound $(u_j - v_j)^2$. So square each side to see

$$(u_j - v_j)^2 \leq 16d^2(k+1)^{2d}(x_m - y_m)^2.$$

Note that $\sum_{i=0}^{k}(x_i - y_i)^2 \geq (x_m - y_m)^2$, and we have the bound

$$\sum_{j=1}^{k+1}(u_j - v_j)^2 \leq 16d^2(k+1)^{2d+1}\sum_{i=0}^{k}(x_i - y_i)^2,$$

which means setting $M = 16d^2(k+1)^{2d+1}$ is sufficient for the Lipschitz condition.

We now check the other conditions. For all n, $X_n(0) = e_0$ (the starting position is always that all bins are empty), so $\lim_{n\to\infty} X_n(0) = e_0$. All that remains is to find K such that $\sum_{l\in L} |l| \sup_{x\in K} \beta_l(x)$ is finite. Let $K = [-1,2]^{k+1}$. Then

$$\sup_{x\in k} \beta_{e_i - e_{i-1}}(x) \leq (2+2+\ldots+2)^d + (1+\ldots+1)^d = (2^d+1)i^d.$$

Since all $l \in L$ have $|l| \leq 2$, and $|L| = k+1$, we see

$$\sum_{l\in L} |l| \sup_{x\in K} \beta_l(x) \leq \sum_{i=1}^{k+1} 2(2^d+1)i^d \leq (2^{d+1}+2)(k+1)^{d+1},$$

which is finite

Therefore all the conditions of Kurtz's theorem are satisfied, and so X_n approaches X, where $X = x_0 + \int_0^t F(X(u))du$. That is, each coordinate y_i of X can be determined by solving the system of differential equations

$$\{y_i'(t) = 2(y_{i-1}(t)+\ldots+y_0(t))^d - (y_{i-2}(t)+\ldots+y_0(t))^d - (y_i(t)+\ldots+y_0(t))^d\}_{i=2}^{k},$$

$$y_1'(t) = 2y_0(t)^d - (y_0(t)+y_1(t))^d,$$

$$y_0'(t) = -y_0(t)^d$$

with initial values $y_0(0) = 1$ and $y_i = 0$ for all $i > 0$.

□

So forGREEDY, as for FAIR, Kurtz's theorem can be applied to calculate expected numbers of bins of constant size. For FAIR and similar algorithms, Kurtz's theorem is also effective in estimating numbers of bins of larger sizes. In those cases, the differential equations for higher bin sizes can be bounded in a simple way, which in turn leads to bounds on the probability of very large bins existing. There are no analogous bounds on the differential equations for GREEDY. In the next section, we use other techniques to bound the expected numbers of large bins.

2.2 Maximum bin size

Recall that one of our goals was an algorithm that had many large bins. In particular, the hope was that the largest bin size under GREEDY is much more than that under UNIFORM (and hence FAIR). This seems like a reasonable hope; FAIR decreases the maximum bins size dramatically, so GREEDY ought to increase it.

A natural proof technique would be to use Kurtz's theorem and find a bin size m such that x_m is almost 0, e.g. m such that $x_m < 1/n^2$. Then the probability of a bin of size m existing is almost 0, and we have an upper bound. This idea works well for FAIR, for which

$$F(x_i) = (1 - x_{i-2} - x_{i-3} - \ldots - x_0)^d + (1 - x_i - \ldots x_0) - 2(1 - x_{i-1} - \ldots - x_0)^d.$$

By changing notation and letting $s_i = 1 - x_{i-1} - \ldots - x_0$, the equation simplifies to

$$F(s_i) = s_{i-1}^d - s_i^d.$$

This lends itself to some simple bounds utilizing the facts that $s_{i-1} \geq s_i$ and s_i is between 0 and 1; for example, $\dot{s}_i \leq s_{i-1}^d$. Combining these ideas leads to an upper bound on $s_i - s_{i-1} = x_i$, as desired.

The same technique doesn't work as well for GREEDY. Setting

$$s_i = 1 - x_{i-1} - \ldots - x_0$$

means $F(s_i) = (1 - s_i)^d - (1 - s_{i-1})^d$. The fact that $s_{i-1} \geq s_i$ is no longer useful, as signs are now reversed. Other tricks don't make the equation any easier to estimate.

Instead, we use a coupling between GREEDY(m, n) and UNIFORM(dm, n) to bound the maximum bin size. Unfortunately, this bound is not significantly greater than the upper bound for UNIFORM(m, n); GREEDY does not increase the size of the largest bin.

Theorem 2.2.1. *With probability greater than or equal to $1 - 1/n$, the most full bin* *under* GREEDY(cn, n) *(c an arbitrary constant) has size less than* $\frac{(2+\epsilon)\log n}{\log\log n - \log d - \log c}$ *for all* $\epsilon > 0$.

Proof. Consider UNIFORM(dcn, n) coupled with GREEDY(cn, n) as follows: the $(d(t-1)+i)^{th}$ ball is placed by UNIFORM(dcn, n) in the i^{th} bin chosen for S_t. (If a bin is chosen twice for S_t, then it receives two balls.). This is clearly a valid coupling. Furthermore, if bin B_i receives a ball under GREEDY(cn, n) at time t, then $i \in S_t$, so bin B_i also receives a ball under UNIFORM(dcn, n). Therefore once all cn balls have been distributed under GREEDY(cn, n) and all dcn balls have been distributed under UNIFORM(dcn, n), every bin in the uniform distribution is at least as full as its counterpart in the greedy distribution. Therefore to bound the fullest bin in the greedy distribution, it's enough to bound the fullest bin under the uniform distribution.

Let's formalize this idea. Consider $X' = \sum_{i=1}^{m} X_i'$, where each X_i' is a random variable equal to 1 with probability $1-(1-1/n)^d$ and 0 with probability $(1-1/n)^d$ (so X' has the same distribution as a count of the times some fixed bin B was one of the d options). Let $X = \sum_{i=1}^{n} X_i$ be the sum of random variables, each equal to 1 with probability d/n. Note that $Pr(X \geq m) \geq Pr(X' \geq m)$ for any m. Furthermore, let $Y_i = X_i - d/n$. Note that $E(Y_i) = 0$ and $Y = \sum_{i=1}^{n} Y_i = X - n(d/n) = X - d$.

We now prove a lemma about Y, which will lead to a bound on X, which in turn gives a bound on bin sizes under GREEDY.

Lemma 2.2.2. *For* Y *as defined above, and for arbitrary* β,

$$Pr(|Y| \geq (\beta-1)d) < (e^{\beta-1}\beta^{-\beta})^d.$$

Proof. First recall the standard inequality $Pr(|Y| > a) < 2e^{-2a^2/n}$, where $Y = \sum_{i=1}^{n} Y_i$, $Pr(Y_i = 1 - d/n) = d/n$, and $Pr(Y_i = -d/n) = 1 - d/n$. (See, e.g., Appendix A of [1].) Now consider $E[e^{\lambda Y}]$, for some arbitrary fixed λ.

$$E[e^{\lambda Y}] = \prod_{i=1}^{n} E[e^{\lambda Y_i}] = \prod_{i=1}^{n} [(d/n)(e^{\lambda(1-d/n)} + (1-d/n)(e^{-\lambda d/n})$$
$$= e^{-\lambda d}((d/n)e^{\lambda} + 1 - d/n)^n$$

Now note that $Pr[Y > a] = Pr[e^{\lambda Y} > e^{\lambda a}]$. Apply Markov's inequality to see $Pr[e^{\lambda Y} > e^{\lambda a}] \leq E[e^{\lambda Y}/e^{\lambda a}]$. Combining that with the previous inequality, we see

$$Pr[Y > a] \leq e^{-\lambda(d+a)}((d/n)e^{\lambda} + 1 - d/n)^n$$

for all λ and a. Let $a = (\beta - 1)d$ and $\lambda = \log \beta$ to get $Pr(|Y| \geq (\beta-1)d) < e^{\beta-1}\beta^{-\beta})^d$, as desired. □

We want to upper bound the probability by $1/n^2$. Setting $\beta = \frac{(2+\epsilon)\log n}{d \log\log n}$ is sufficient for arbitrary $\epsilon > \frac{2 \log\log\log n}{\log\log n - \log\log\log 3}$ (for example, any constant ϵ will do):

$$\lim_{n\to\infty} \log(e^{\beta-1}\beta^{-\beta})^d n^2 = \lim_{n\to\infty} d\beta - d - \beta d \log \beta + 2 \log n.$$

Setting $\beta = \frac{(2+\epsilon)\log n}{d \log\log n}$ gives us

$$\lim_{n\to\infty} \frac{(2+\epsilon)\log n}{d \log\log n} - d - \left(\frac{(2+\epsilon)\log n}{\log\log n}\right) \log\left(\frac{(2+\epsilon)\log n}{d \log\log n}\right) + 2\log n$$

$$= \frac{(2+\epsilon)\log n}{\log\log n} - d - \left(\frac{(2+\epsilon)\log n}{\log\log n}\right)(\log(2+\epsilon) + \log\log n - \log d - \log\log\log n) + 2\log n,$$

which is

$$\lim_{n\to\infty} -\epsilon \log n + (\log\log\log n - \log d + \log(2+\epsilon) + 1)\left(\frac{(2+\epsilon)\log n}{\log\log n}\right) - d = -\infty.$$

Therefore the probability is $< 1/n^2$ for n sufficiently large. On the other hand,

$$Pr(|Y| \geq M) \geq Pr(X \geq M + d) \geq Pr(X' \geq M + d).$$

So the probability of a bin having more than $\beta d = \frac{(2+\epsilon)\log n}{\log\log n}$ balls is thus $< 1/n$ for n sufficiently large, and we have an upper bound on the likely size of the maximum bin. $\square$

We can strengthen the above theorem to get an upper bound for non-constant d. Even if d grows with n, as long as it grows slowly, the asymptotic behavior of the maximum is the same as if d were a constant. In fact, if $d(n) = O(\log\log n)$, then the upper bound is the same.

Theorem 2.2.3. *With probability at least* $1-1/n$ *the most full bin under* GREEDY(n,n) *with number of options* $d(n)$ *dependent on* n *has size less than*

$$\frac{(2+\epsilon)\log n}{\log(2+\epsilon)+\log\log n-\log d(n)}$$

for all ϵ *such that*

$$\frac{(2+\epsilon)\log n}{d(n)(\log(2+\epsilon)+\log\log n-\log d(n))} \geq 1.$$

Proof. All that is necessary is a constant ϵ such that $\frac{(2+\epsilon)\log n}{d(n)(\log(2+\epsilon)+\log\log n-\log d(n))} \geq 1$. (Note that this rules out, e.g. $d(n) \geq \log n$.) Let $a = \frac{(2+\epsilon)\log n}{d(n)}$ and $\beta = \frac{a}{\log a} \geq 1$. Then the probability that $X \geq \beta d(n)$ is bounded by $(e^{\beta-1}\beta^{-\beta})^{d(n)} < 1/n^2$ for n sufficiently large. Therefore the maximum bin almost surely contains less than $\beta d(n) = \frac{(2+\epsilon)\log n}{\log(2+\epsilon)+\log\log n-\log d(n)}$ balls. $\square$

So we now have statistics for smaller bins' sizes, and one upper bound on the largest bin size. In the next section, we will turn to different statistics about bin sizes that give another way to compare GREEDY with other distributions.

Chapter 3

Results for arbitrary m

The previous results were limited in several ways. They only held for a number of balls linear in the number of bins. Even with that restriction, we were only able to compute the expected numbers of bins of constant size. Furthermore, the upper bound on the size of the maximum bin didn't take into account the particular structure of GREEDY.

In this chapter, we develop bounds that hold for arbitrary values of m and n. These bounds do take into account the structure of GREEDY, letting us see exactly how the bins are growing and moving. Some of these bounds are on the size of a percentage of the bins, rather than one particular bin, which allows for better estimates. Others look at the overall sequence of bin sizes produced by GREEDY, and compare it with other sequences that are already well-studied.

3.1 Bounds on a fraction of the bins

Our previous results focused on the expected number of bins of an exact size. We now take a different perspective. Rather than calculating the proportion of bins of a certain size, we calculate the size of a certain proportion of bins. That is, we calculate the total number of balls within a certain subset of the bins. We bound the number of balls in the smallest fraction of bins from above and below. These bounds are valid for arbitrary numbers of balls and bins, unlike previous bounds. They are tight; the gap between the upper and lower bounds is independent of m, n, and the fraction of bins considered.

Theorem 3.1.1. *Under* GREEDY(m, n), *the expected number of balls in the last* xn *bins, once the bins have been arranged in decreasing order, is at most* $x^d m$ *for all values of* x, m, n, *and* d. *The probability of the last* xn *bins having at least* k *balls is upper bounded by* $1 - \sum_{i=0}^{k-1} \binom{m}{i}(1-x)^{m-i}x^i$.

Proof. We first relabel the bins and redefine GREEDY to make it easier to analyze. Let $i(t)$ be the index of the i^{th} largest bin at time t. Break ties in this labeling randomly; for example, if $b_1(t) = 3$, $b_2(t) = 3$, and $b_3(t) = 2$, then $1(t)$ is equally likely to be 1 or 2. Now think of the option sets in terms of this relabeling. Choosing $(i_1, \ldots, i_d)$ from $[n]^d$ at time t gives as options bins $B_{i_1(t)}, \ldots, B_{i_d(t)}$. The bin that gets the ball is the one with the smallest index under this labeling. For ease of notation, let $i_m = \min\{i_1, \ldots, i_d\}$.

Note that this redefined version is indistinguishable from the original. Option sets are still chosen uniformly at random, and the largest bin in the option set is the one which gets a ball. Ties are broken uniformly randomly, as before.

This redefinition both confuses and clarifies the algorithm. It is helpful in determining which bin gets the ball. In the original labeling, it was equally likely that $b_1(t) > b_2(t)$ or $b_2(t) > b_1(t)$. Now, it is always true that $b_{1(t)}(t) \geq b_{2(t)}(t)$. On the other hand, it is no longer the case that giving the t^{th} ball to bin i implies bin i is larger at time $t+1$. For example, suppose the bin sizes are $2, 2, 2, 1, 1, 0$ and $S_t = \{3, 4, 6\}$ (so $i_1 = 3$, $i_2 = 4$, and $i_3 = 6$). Then the ball goes into bin $B_{3(t)}$. At time $t+1$, the configuration is $3, 2, 2, 1, 1, 0$. Although bin $B_{3(t)}$ was given the ball, $b_{3(t)} = b_{3(t+1)}$. The increase is for $b_{1(t)}$: $b_{1(t+1)}(t+1) = b_{1(t)}(t) + 1$.

So putting the ball into bin B_{i_m} doesn't guarantee that $b_{i_m(t+1)}(t+1) = b_{i_m(t)}(t)+1$, but it does guarantee that the increase must be in a bin of index at most i_m. The increased bin may move to the left in the reordering, but never to the right. Therefore we do know that

$$\sum_{j=1}^{i_m} b_{j(t)}(t) + 1 = \sum_{j=1}^{i_m} b_{j(t+1)}(t+1).$$

We can use this observation to get bounds on ball placement. It is clear from the above equation that $\sum_{j>(1-x)n} b_{j(t)}$ can increase only if $i_m > (1-x)n$. In other words, the last x of the bins increase in size only if the list of options is contained within them. This happens at each time step with probability x^d. Therefore the expected number of times this has happened, once all m balls have been distributed, is mx^d. So once all balls have been distributed, the number of balls in the union of the last x bins has expectation at most mx^d.

Furthermore, the last x bins contain at least k balls only if the option set has been contained within the last x bins at least k times. Therefore the probability of the last x bins containing at least k balls is at most

$$1-\sum_{i=0}^{k-1}\binom{m}{i}Pr(S_t \text{ is in last } x)^i Pr(S_t \text{ is not in last } x)^{m-i} = 1-\sum_{i=0}^{k-1}\binom{m}{i}x^{di}(1-x^d)^{m-i}.$$

□

This gives us another statistic by which to compare GREEDY and UNIFORM. As the expected number of bins of size k under UNIFORM(m,n) is known to be

$$\frac{(m/n)^k e^{-m/n}}{k!},$$

we can compute the expected fraction of bins that contain a particular fraction of balls. For example, when $m = n$, the expected number of bins of size 0 is n/e, of size 1 is n/e, and of size 2 is $n/2e$. So if we take just the emptiest bins until we have half the balls, the expected number of bins would be

$$n/e + n/e + (n/2 - n/e)(1/2) \approx 0.8n.$$

We can use the same type of calculations to get Table 3.1.

Note that, as desired, the majority of the balls under GREEDY are concentrated in the largest few bins. This effect becomes more pronounced as d grows. For example, when $d \geq 4$, half the balls take up at least the lower $2^{-1/4} \simeq 0.84$. Therefore there are

Table 3.1: Expected lower fraction of bins x containing fraction of balls y, $m = n$

	UNIFORM	$d = 2$	$d = 3$	d
$y = 1/3$	$x = 0.7$	$x \geq 1/\sqrt{3}$	$x \geq 1/3^{1/3}$	$x \geq (1/3)^{1/d}$
$y = 1/2$	$x = 0.8$	$x \geq 1/\sqrt{2}$	$x \geq 1/2^{1/3}$	$x \geq (1/2)^{1/d}$
$y = 2/3$	$x = 0.88$	$x \geq \sqrt{(2/3)}$	$x \geq (2/3)^{1/3}$	$x \geq (2/3)^{1/d}$

at least half the balls within the top .16 bins, unlike under UNIFORM, where they take up the top 0.2 bins.

Also note that this argument holds regardless of m, n, and d. It is equally valid for $n = m$, m linear in n, and m exponential in n. It is valid for d constant or a function of n or m. It is also valid for arbitrary x, as long as $xn \geq 1$. For example, we can use this to see the expected number of balls in the last $1/\sqrt{n}$ fraction of the bins is at most $n^{-d/2}m$. Furthermore, the last bin (i.e. the last $1/n$ fraction of bins) remains empty until $m = \Theta(n^d)$; $mx^d = m(1/n)^d = \frac{m}{n^d}$.

Theorem 3.1.2. *The expected number of balls under* GREEDY(m,n)*in the last xn bins, once the bins have been arranged in decreasing order, is at least*

$$\left(\frac{2d+1}{2d+2}\right)^d \left(\frac{1}{2d+2+x-2dx}\right) x^d m \geq e^{-1/2} \left(\frac{1}{2d+2}\right) x^d m$$

for all values of m, n, d, and x.

Proof. This proof relies on a case-by-case analysis of the behavior of GREEDY. We first prove a weaker bound, to demonstrate a simple version of the technique used. We then alter parameters to prove the stronger bound.

Note that, if a bin within the last xn bins receives a ball, the only case in which the total number of balls in the last xn bins doesn't increase is if there were a string of bins of the same size that overlapped the dividing line. For example, suppose $x = 1/2$, the bins are 322221, and the option set is $\{4, 5, 6\}$. This option set is contained within the last half, so under our estimate we would expect the last half to increase. The ball goes in the 4^{th} biggest bin, but the new bin size sequence is 332221. The last half has the same size as it did before, because the increase "leapt up" the string of 2s.

When the $(xn+j)^{th}$ bin is given a ball, but the $(xn-k)^{th}$ bin is the one that increases, then the bins between the $(xn-k)^{th}$ and the $(xn+j)^{th}$ must have all had the same size. More formally, if $i_m(t) = xn + j$ and

$$b_{(xn+j)(t)}(t) = b_{(xn+j)(t+1)}(t+1)$$

and

$$b_{(xn-k)(t)}(t) + 1 = b_{(xn-k)(t+1)}(t+1),$$

then

$$b_{(xn-k)(t)}(t) = b_{(xn-k+1)(t)}(t) = \ldots = b_{(xn+j)(t)}(t).$$

In order to find a lower bound on the size of the last x bins, we will bound how often there is a string of equally-sized bins stretching from the left to the right of $B_{xn}(t)$. We now give a sketch of the argument's structure, using simple parameters. We then optimize the parameters to find a tighter bound.

Suppose at a particular time step all the option bins are within the last xn. Call this type of time step "good". There are two possibilities for each good round: the increased bin will be within the last $2xn$ bins, or it won't. Let g be the number of good time steps. Then the first or the second case will happen at least $g/2$ times. If the first case happens at least $g/2$ times, then the last $3xn$ bins will have at least $g/2$ balls.

The second case is more tricky. Suppose it happens at least $g/2$ times. Consider the $(1-2x)n-1^{st}$ largest bin (i.e., the smallest bin outside of the least $2x$). Every time a case-two step occurs, that bin must be in a string of equally-sized bins that stretches from some index greater than $(1-x)n$ to some index less than $(1-2x)n$. Furthermore, that string's length decreases every time a case-two step occurs. Therefore, after at most $n-2xn$ case-two steps, the string no longer contains the $(1-2x)n-1^{st}$ bin. This means that boundary bin must have increased in size. If there are $g/2$ case-two steps, then the $(1-2x)n - 1^{st}$ bin must contain at least

$$\frac{\text{number of case-two steps}}{n-2xn} = \frac{g}{2(1-2x)n}$$

balls. There are xn bins at least as full as that one in the last $3xn$, so the last $3xn$ bins must contain at least

$$(xn)\left(\frac{g}{2(1-2x)n}\right) = \frac{gx}{2(1-2x)}$$

balls. Therefore, the last $3xn$ bins contain at least

$$(g/2)(\min\{1, x/(1-2x)\})$$

balls overall. The expected value of g is $x^d m$, so the last $3xn$ bins are expected to contain at least

$$(x^d m/2)(\min\{1, x/(1-2x)\})$$

balls.

This isn't the tightest possible bound, of course. All the parameters used can be altered to give a better bound. For example, we can alter the proportion of cases one and two. For any $p+q=1$, case one happens at least p of the time or case two happens at least q. This gives bounds of

$$pmx^d$$

and

$$\left(\frac{qmx^d}{n(1-2x)}\right)nx.$$

Setting them equal to find the best p and q gives $q = \frac{1-2x}{1-x}$ and $p = \frac{x}{1-x}$, which gives the lower bound

$$\frac{mx^{d+1}}{1-x}$$

in either case.

The other parameter to tighten is the split, i.e. the boundaries for case one and case two. Let δn be the overall section whose balls are bounded (corresponding to $3xn$ in the original proof). Break this large δ section into three smaller sections, of length γn (the emptiest bins, corresponding to the smallest xn bins in the original proof), βn (next emptiest, the middle xn), and αn (most full within the last δn, the largest xn). Case one will now mean that the option set was contained within the last γ bins, and the increase occurred within the last $\beta + \gamma$. Case two is when the option set was contained within the last γ bins, but the last $\beta + \gamma$ did not increase.

We know that case one happens p times or case two happens q times. If case one happens p times, then the expected increase of the last δ bins is

$$pm\gamma^d.$$

If case two happens q times, then as before we need to divide the expected number of case two occurrences by the maximum length of a string of equal bins that spans the boundary between the larger bins and the last $\beta + \gamma$, then multiply it by the number of bins within δ but above $\beta + \gamma$. This gives an expected increase of

$$\left(\frac{qm\gamma^d}{n(1-\beta-\gamma)}\right)\alpha n.$$

So overall we have the lower bound

$$\min\{pm\gamma^d, \left(\frac{qm\gamma^d}{n(1-\beta-\gamma)}\right)\alpha n\}$$

for the number of balls in the last δn bins.

Note that this bound is largest when $\beta = 0$. Setting the two bounds to be equal to find the best p and q, we see

$$q = \frac{1-\delta+\alpha}{1-\delta+2\alpha}$$

and

$$p = \frac{\gamma^\alpha}{1-\delta+2\alpha}.$$

Now plug in $\alpha + \gamma = \delta$ and the above value for p to the generic lower bound

$$pm\gamma^d = \frac{m\alpha(\delta-\alpha)^d}{1+2\alpha-\delta}$$

and solve the derivative for zero to see the maximum is at

$$\alpha = \frac{\sqrt{(1-\delta)^2(1+d)^2 + 2d(\delta)(1-\delta)} - (1-\delta)(1+d)}{2d}.$$

This is pretty unwieldy. Approximate the maximum by setting $\alpha = \frac{\delta}{2d+2}$. Then

$$\gamma = \delta - \alpha = \delta\left(\frac{2d+1}{d+2}\right)$$

and

$$pm\gamma^d = \frac{\left(\frac{\delta}{2+2d}\right)\left(\frac{2d+1}{2d+2}\right)^d \delta^d}{1-\frac{\delta}{1+d}-\delta} m = \left(\frac{2d+1}{2d+2}\right)^d \left(\frac{1}{2d+2+\delta-2d\delta}\right)\delta^d m.$$

Recall that $(1-\epsilon)^{1/\epsilon} \geq e^{-1}$ for $0 < \epsilon < 1$, so

$$\left(\frac{2d+1}{2d+2}\right)^d = \left(\left(1-\frac{1}{2d+2}\right)^{2d+2}\right)^{\frac{d}{2d+2}} \geq e^{-\frac{d}{2d+2}} \geq e^{-1/2}.$$

Also note that $1 < 2d$, so

$$\frac{1}{2d+2+\delta-2d\delta} \geq \frac{1}{2d+2}.$$

So we have the simpler lower bound

$$E(\text{number of balls in the last } \delta n \text{ bins}) \geq e^{-1/2}\left(\frac{1}{2d+2}\right)\delta^d m$$

$\square$

Note that the upper and lower bounds differ from each other by only a constant factor, independent of x, m, and n. We know therefore that as the fraction of bins x increases, the fraction of balls increases as x^d. The only unknown is the correct constant.

3.2 Majorization results

The previous results give us tight (up to constant) estimates on the number of balls in subsets of bins. They don't estimate the size of specific bins. We now bound specific bins by using the reordering according to size. Here, we use coupling to compare the bin size sequence produced by GREEDY with that produced by UNIFORM.

Theorem 3.2.1. *For all values of m and n, the expected sequence of bin sizes under* GREEDY*, arranged in decreasing order, majorizes the expected sequence of bin sizes under* UNIFORM(m, n).

Theorem 3.2.2. *Given a sequence $\mathbf{x}$, let $M_{\mathbf{x}}$ be the set of sequences with the same length and sum that majorize x. Then for any particular bin size sequence $\mathbf{a} = a_1, \ldots a_n$ with probability of p_1 under* UNIFORM(m, n)*, the probability of $M_{\mathbf{a}}$ under* GREEDY(m, n) *has probability $p_2 \geq p_1$.*

Theorem 3.2.3. *For any S a set of sequences, the probability of $\cup_{\mathbf{x} \in S} M_{\mathbf{x}}$ under* GREEDY(m, n) *is at least as great as the probability of $\cup_{\mathbf{x} \in S} \mathbf{x}$ under* UNIFORM(m, n).

Proof. These three theorems follow from a simple coupling. Consider running two algorithms in parallel. Each process has its own set of n bins ($\{A_1, \ldots, A_n\}$ and $\{B_1, \ldots, B_n\}$, respectively) and m balls which will be distributed sequentially into the A_i (respectively B_i). We will assume the process runs from time 1 to time m, with the i^{th} ball being placed at time i. Consider the bin size sequence. As before, let bin $A_{i(t)}$ be the i^{th} largest bin at time t (breaking ties uniformly randomly), and define $B_{i(t)}$ analogously. Let $a_{i(t)}$ be the number of balls in bin $A_{i(t)}$ at time t, similarly define $b_{i(t)}$. At time t, choose option set $i_1, \ldots, i_d \in [n]$ uniformly randomly. For the greedy algorithm, place the t^{th} ball in bin $A_{i_m(t)}$, where $m = \min\{i_1, \ldots i_d\}$. For the uniform algorithm, place the t^{th} ball in bin $B_{i_1(t)}$.

At any time t, the sequence $a_{1(t)}, \dots a_{n(t)}$ majorizes the sequence $b_{1(t)}, \dots, b_{n(t)}$. This can be proved by induction on time, although the proof is less trivial then it might look at first glance. Suppose i balls have been distributed, and $a_{1(i)}, \dots a_{n(i)}$ majorizes the sequence $b_{1(i)}, \dots, b_{n(i)}$. Look at what happens when the $i + 1^{st}$ ball is distributed.

By the definitions of A and B, the bin in which ball $i + 1$ is placed is the same as, or to the left of, the bin for B at time i. However, that does not necessarily mean that the position in the ordering of $A_{m(i)}$ at time $i + 1$ is to the left of the position in the ordering of $B_{i_1(i)}$ at time $i + 1$. For example, suppose $A = 7632$ and $B = 6552$ at time i, and the chosen indices are 3 and 4. In that case, for A, 3 becomes 4, and for B, 5 becomes 6. At time $i + 1$, $A = 7642$ and $B = 6652$. The bin whose size increased for B was to the left of the bin whose size increased for A.

So we have to be more subtle. Suppose it does happen that the bin whose size increased for B is, at time $i+1$, to the left of the bin whose size increased for A. For convenience, let its index at time $i+1$ be a for A and b for B. Let the indices at time i be α and β, respectively. Then for this to have happened, all bins in A with indices between α and a must have equal size at time i, and similarly for all bins in B with indices between β and b.

Now, suppose for contradiction that A at time $i + 1$ no longer majorizes B. In particular, there is some index j such that

$$\sum_{j=1}^{k} a_{j(i+1)} < \sum_{j=1}^{k} b_{j(i+1)}.$$

For this to be true, $\beta \leq k < \alpha$ and

$$\sum_{j=1}^{k} a_{j(i)} = \sum_{j=1}^{k} b_{j(i)}.$$

Suppose $a_{k(i)} < b_{k(i)}$. By the above discussion, $b_{k(i)} = b_{k+1(i)}$ and, by definition,

$a_{k(i)} \geq a_{k+1(i)}$. Therefore
$$\sum_{j=1}^{k+1} a_{j(i)} < \sum_{j=1}^{k+1} b_{j(i)},$$
which contradicts the induction hypothesis.

Suppose $a_{k(i)} > b_{k(i)}$. Then it's even simpler: $\sum_{j=1}^{k-1} a_{j(i)} < \sum_{j=1}^{k-1} b_{j(i)}$.

Suppose $a_{k(i)} = b_{k(i)}$. Since A majorizes B at time i, for

$$\mu = \max_{j<k} a_{j(i)} \neq b_{j(i)},$$

$a_{\mu(i)} > b_{\mu(i)}$. Therefore, if μ exists and is greater than 1,

$$\sum_{j=1}^{\mu-1} a_{j(i)} < \sum_{j=1}^{\mu-1} b_{j(i)}.$$

Furthermore, such a μ must exist. Suppose it did not. Then A and B would be equal up through the k^{th} bin at time i. But we also know that all bins between β and b are equal for B and not for A. Therefore the bin size for A must decrease at some index l, for $\beta \leq l \leq b$. But then A no longer majorizes B, as $\sum_{j=1}^{l} a_{j(i)} < \sum_{j=1}^{l} b_{j(i)}$.

Suppose $\mu = 1$. Then $\sum_{j=1}^{k} a_{j(i)} > \sum_{j=1}^{k} b_{j(i)}$, contradicting our original assumption.

So we have now proved that $a_{1(t)}, \ldots a_{n(t)}$ majorizes $b_{1(t)}, \ldots b_{n(t)}$. From this fact, all three theorems follow:

To prove the first theorem, recall that the weighted averages of pairs of sequences, for each of which one majorizes the other, must majorize each other. That is, if there exists a bijection $f : S \to T$ such that $f(\mathbf{x})$ majorizes $\mathbf{x}$, then for any $\{\lambda_{\mathbf{x}}\}_{\mathbf{x}\in S}$ such that $\sum_{\mathbf{x}\in S} \lambda_{\mathbf{x}} = 1$, $\sum_{\mathbf{x}\in S} \lambda_{\mathbf{x}}\mathbf{x}$ is majorized by $\sum_{\mathbf{x}\in S} \lambda_{\mathbf{x}} f(\mathbf{x})$. Expected value fits this definition, so this coupling proves that the expected bin size sequence under GREEDY majorizes the expected bin size sequence under UNIFORM.

To prove the second theorem, note that for any sequence of option sets that generated $\mathbf{a}$ under UNIFORM, the same sequence will lead to some $\mathbf{b}$ under GREEDY, where $\mathbf{b}$ majorizes $\mathbf{a}$. The probability of $\mathbf{a}$ is the probability of any of those bin size sequences occurring. If any of those sequences occurs, then $\mathbf{b}$ will majorize $\mathbf{a}$. So $Pr(M_{\mathbf{a}})$ under GREEDY is at least the probability of any of those sequences occuring, which is $Pr(\mathbf{a})$ under UNIFORM. So $p_2 \geq p_1$, as desired.

To prove the third theorem, we generalize the previous paragraphs. Let $S'_{\mathbf{x}}$ be the set of option set sequences $\mathbf{s}$ that lead to a bin size sequence in $M_{\mathbf{x}}$ under GREEDY. Let $S''_{\mathbf{x}}$ be the set of option set sequences $\mathbf{s}$ that lead to the bin size sequence $\mathbf{x}$ under UNIFORM. Then the coupling tells us that $S''_{\mathbf{x}} \subseteq S'_{\mathbf{x}}$, so the probability of $S'_{\mathbf{x}}$ is at least the probability of $S''_{\mathbf{x}}$. By the definitions of S' and S'' we know that $S''_{\mathbf{x}}$ is disjoint from $S''_{\mathbf{y}}$ for all $\mathbf{x} \neq \mathbf{y}$. Thus $\sum_{\mathbf{x} \in S} |S''_{\mathbf{x}}| \leq \sum_{\mathbf{x} \in S} |S'_{\mathbf{x}}|$, and so the probability of $\cup_{\mathbf{x} \in S} M_{\mathbf{x}}$ under GREEDY is at least the probability of $\cup_{\mathbf{x} \in S} \mathbf{x}$ under UNIFORM.

□

We can also bound the bin size sequence of GREEDY from above by coupling with UNIFORM. The key is to consider UNIFORM with d times as many balls. Then the same arguments used to prove the lower bound also prove an upper bound.

Theorem 3.2.4. *For all values of m and n, under* GREEDY(m, n) *the expected sequence of bin sizes, arranged in decreasing order, is majorized by the expected sequence of bin sizes under* UNIFORM(dm, n).

Theorem 3.2.5. *Given a sequence $\mathbf{x}$, let $M_{\mathbf{x},y}$ be the set of sequences of the same length as $\mathbf{x}$ and with sum y (y is not necessarily the same as the sum of $\mathbf{x} = \sum_{i=1}^{n} x_i$) that majorize $\mathbf{x}$. Then for any particular bin sequence $\mathbf{b} = b_1, \ldots b_n$ with probability p_2 under* GREEDY(m, n), *the probability p_1 under* UNIFORM(dm, n) *of $M_{\mathbf{b},dm}$ is greater than p_2.*

Theorem 3.2.6. *For all S sets of sequences, the probability of $\cup_{\mathbf{x} \in S} M_{\mathbf{x},dm}$ under* UNIFORM(dm, n) *is at least as great as the probability of $\cup_{\mathbf{x} \in S} x$ under* GREEDY(m, n).

Proof. A similar argument works for this second group of theorems. As before, A will be the greedy process and B the uniformly random process (but now with d times as many balls as A). That is, A is GREEDY(m, n) and B is UNIFORM(dm, n). The same induction shows that the sequence generated by GREEDY(m, n) is majorized by that of UNIFORM(dm, n). Modify the coupling to allow a step to be (A doesn't get a ball and B does) or (A and B both get balls). The two processes are coupled by choosing $i_1, \ldots i_d$ uniformly at random from $[n]$ at time steps of the form xd for x a natural number. At time steps of the form $xd + i_m(xd)$, bins $A_{i_m(xd)}$ and $B_{i_m(xd)}$ get a ball. At time steps of the form $xd + j$, $j \neq i_m(xd)$, bins in A do not get a ball, but bin B_{i_j} does. At time dm, then, all m balls have been distributed in A and all dm have been distributed in B.

We can show by an argument similar to the previous one that $a_1(dm) \ldots a_n(dm)$ is majorized by $b_1(dm) \ldots b_n(dm)$. Suppose $a_1(i) \ldots a_n(i) = \mathbf{a}(i)$ is majorized by $b_1(i) \ldots b_n(i) = \mathbf{b}(i)$. We want to prove that $\mathbf{a}(i+1)$ is majorized by $\mathbf{b}(i+1)$. If $i+1$ is not of the form $xd + i_m(xd)$, then $\mathbf{a}(i) = \mathbf{a}(i+1)$. Since $\mathbf{b}(i+1)$ majorizes $\mathbf{b}(i)$, we know that $\mathbf{b}(i+1)$ majorizes $\mathbf{a}(i+1)$. On the other hand, if $i+1 = xd + i_m(xd)$, then the same bin gets a ball under A and B. We are now in the situation described in the previous proof. We have two bin size sequences at time i, one majorizing the other, and the bins receiving balls under A and B are both indexed by $xd + m(xd)$. So the same analysis works here as did in the previous case. Thus we know that the sequence generated by A is majorized by the sequence generated by B.

As before, we use this argument to prove all three theorems. The weighted average of majorizing sequences is a majorizing sequence, so the expected sequence of bin sizes under GREEDY(m, n) is majorized by the expected sequence of bin sizes under UNIFORM(dm, n). The option sets argument also proves that $p_1 \geq p_2$, as before. And a coupling of option set sequences again proves the third theorem.

$\square$

These theorems are powerful tools in analyzing the growth of GREEDY(m, n). For example, they show that the largest bin under GREEDY(m, n) is at least as large as the largest bin under UNIFORM(m, n). On the other hand, it is less than UNIFORM(dm, n). As a corollary to these theorems, then, we can generate Theorem 2.2.1 and Theorem 2.2.3.

Chapter 4

Results for m much larger than n

In this final chapter, we consider the case $m >> n$. Once the number of balls gets much larger than the number of bins, UNIFORM and FAIR both approach an even distribution of m/n balls per bin. On the other hand, GREEDY becomes more lopsided. How lopsided GREEDY becomes, and how quickly, will be discussed below.

4.1 Relative position of bins

The difficulty of estimating the size of a particular bin in the previous chapter was due to the bins and balls changing position. If bin B is currently part of a string of equally-sized bins, then giving it a ball means B leaps to the head of the string after reordering. Note that the probability of any particular bin getting a ball depends only on its position in the size ordering, so this is especially frustrating. We now bound the probability of two bins changing position.

Theorem 4.1.1. *For any starting configuration of bins and balls, if bin B_i is currently $\delta n/(d-1)$ balls more full than bin B_j, then the probability of bin B_i becoming smaller than bin B_j at any time in the future (i.e. for any value of m) under* GREEDY *is at most $e^{-\delta}$.*

Proof. Let's view the changing ball statistics of the two bins as a random walk. Fix two bins, without loss of generality say B_1 and B_2, and consider $|b_1 - b_2|$ at each time step. For most steps, this statistic doesn't change; usually the ball doesn't get placed in B_1 or B_2. Let's condition on one of those two bins getting a ball. If the larger bin gets a ball, the distance between them increases, so the statistic increases by 1. If the smaller bin gets a ball, it decreases. This is a random walk with a reflecting barrier at 0. Since a larger bin is more likely to get a ball than a smaller bin, it is biased in favor of $+1$

How much more likely is the larger bin to get a ball? Suppose B_1 and B_2 are currently ranked the i^{th} and j^{th} bins, with $i > j$, where the 1^{st} bin is the smallest and the n^{th} bin is the biggest. Then the probability, given that one of the two bins gets a ball, of the bigger bin getting it is

$$\frac{i^d - (i-1)^d}{i^d - (i-1)^d + j^d - (j-1)^d}.$$

That is because the i^{th} bin gets a ball with probability $\left(\frac{i}{n}\right)^d - \left(\frac{i-1}{n}\right)^d$ overall, and the j^{th} with probability $\left(\frac{j}{n}\right)^d - \left(\frac{j-1}{n}\right)^d$. So the probability that, given that either the i^{th} or j^{th} gets a ball, of the i^{th} getting it is

$$\frac{\left(\frac{i}{n}\right)^d - \left(\frac{i-1}{n}\right)^d}{\left(\frac{i}{n}\right)^d - \left(\frac{i-1}{n}\right)^d + \left(\frac{j}{n}\right)^d - \left(\frac{j-1}{n}\right)^d} = \frac{i^d - (i-1)^d}{i^d - (i-1)^d + j^d - (j-1)^d}.$$

We now minimize this probability. First, note the minimum must be such that $j + 1 = i$. That is clear from considering the model; the closer together two bins are, the closer together their respective probabilities of getting a ball are. If the two bins are neighbors, then the only option sets for which bin i would get a ball but bin $i-1$ would not are those which include both bins.

So we need to minimize

$$\frac{i^d - (i-1)^d}{i^d - (i-1)^d + (i-1)^d - (i-2)^d} = \frac{i^d - (i-1)^d}{i^d - (i-2)^d}.$$

We first take the derivative.

$$\frac{d}{dx}\left[\frac{x^d-(x-1)^d}{x^d-(x-2)^d}\right] =$$
$$\frac{(x^d-(x-2)^d)(dx^{d-1}-d(x-1)^{d-1})-(dx^{d-1}-d(x-2)^{d-1})(x^d-(x-1)^d)}{(x^d-(x-2)^d)^2}$$

Note that the denominator is positive, so we can determine the behavior of the function by looking at the sign of the numerator.

$$\begin{aligned}
&(x^d-(x-2)^d)(x^{d-1}-(x-1)^{d-1})-(x^{d-1}-(x-2)^{d-1})(x^d-(x-1)^d)\\
&= (x-(x-2))\left(\sum_{j=0}^{d-1} x^j(x-2)^{d-1-j}\right)(x-(x-1))\left(\sum_{j=0}^{d-2} x^j(x-1)^{d-2-j}\right)\\
&\quad -(x-(x-2))\left(\sum_{j=0}^{d-2} x^j(x-2)^{d-2-j})(x-(x-1)\right)\left(\sum_{j=0}^{d-1} x^j(x-1)^{d-1-j}\right)\\
&= 2\sum_{j=0}^{d-1} x^j(x-2)^{d-1-j}\sum_{j=0}^{d-2} x^j(x-1)^{d-2-j} - 2\sum_{j=0}^{d-2} x^j(x-2)^{d-2-j}\sum_{j=0}^{d-1} x^j(x-1)^{d-1-j}
\end{aligned}$$

Let $a = \sum_{j=0}^{d-2} x^j(x-2)^{d-2-j}$ and $b = \sum_{j=0}^{d-2} x^j(x-1)^{d-2-j}$. Then the difference becomes

$$\begin{aligned}
&2((x-2)\sum_{j=0}^{d-2} x^j(x-2)^{d-2-j}+x^{d-1})\left(\sum_{j=0}^{d-2} x^j(x-1)^{d-2-j}\right)\\
&\quad -2\left(\sum_{j=0}^{d-2} x^j(x-2)^{d-2-j}\right)((x-1)\sum_{j=0}^{d-2} x^j(x-1)^{d-2-j})\\
&= 2[(a(x-2)+x^{d-1})(b)-a((x-1)b+x^{d-1})]\\
&= 2[abx-2ab+bx^{d-1}-abx+ab-ax^{d-1}]\\
&= 2[(b-a)x^{d-1}-ab]
\end{aligned}$$

We want this to be negative: $(b-a)x^{d-1}-ab<0$. Therefore we need to show

$$\frac{bx^{d-1}}{b+x^{d-1}} < a.$$

Expand b and a to get the equivalent inequality

$$\frac{x^{d-1}\left(\frac{x^{d-1}-(x-1)^{d-1}}{x-(x-1)}\right)}{\left(\frac{x^{d-1}-(x-1)^{d-1}}{x-(x-1)}\right)+x^{d-1}} < \frac{x^{d-1}-(x-2)^{d-1}}{x-(x-2)}.$$

Simplifying further gives

$$\frac{x^{d-1}(x^{d-1}-(x-1)^{d-1})}{2x^{d-1}-(x-1)^{d-1}} < (1/2)(x^{d-1}-(x-2)^{d-1}).$$

Dividing the left hand numerator and denominator by x^{d-1}gives

$$\frac{x^{d-1}-(x-1)^{d-1}}{2x^{d-1}-(\frac{x-1}{x})^{d-1}} < (1/2)(x^{d-1}-(x-2)^{d-1}).$$

Note that $x^{d-1}-(x-1)^{d-1} < x^{d-1}-(x-2)^{d-1}$ and $2x^{d-1}-(\frac{x-1}{x})^{d-1} > 2$, so the inequality is true.

Therefore the original derivative is negative, and the original function is minimized at $i = n$. That is, the probability of the larger of two bins getting a ball is minimized when they are the largest and second largest bins. In that case the larger bin gets a ball with probability

$$\frac{n^d-(n-1)^d}{n^d-(n-2)^d}.$$

We will now formalize the coupling of the changing gap size with a random walk. For ease of notation, let i_j be the j^{th} time step at which bin B_1 or B_2 gets a ball. Let X_i be the position of a random walk with bias ϵ at time i, where

$$\frac{1+\epsilon}{2} = \frac{n^d-(n-1)^d}{n^d-(n-2)^d}.$$

That is, the random walk has the same bias as that between the largest and second largest bins.

Couple the sequence $|b_1(i_1)-b_2(i_1)|, |b_1(i_2)-b_2(i_2)|, \ldots$ with $X_1, X_2, \ldots$ as follows. If at time j the larger bin has probability γ of being chosen over the smaller bin, then $X_j = X_{j+1} + 1$ with probability

$$\gamma - \frac{1+\epsilon}{2}$$

if

$$|b_1(i_j) - b_2(i_j)| > |b_1(i_{j-1}) - b_2(i_{j-1})|$$

and $X_j = X_{j-1} - 1$ with probability 1 if

$$|b_1(i_j) - b_2(i_j)| < |b_1(i_{j-1}) - b_2(i_{j-1})|.$$

That is, the random walk takes a -1 step every time the bin sizes step closer together, and additionally may take a -1 step even if the bins don't step closer, in such a way that the probability of a +1 step in the random walk is always $\frac{1+\epsilon}{2}$.

If B_1 and B_2 switch relative positions, then there exists a time step t at which $b_1(t) = b_2(t)$. Therefore they switch positions only if there exists t such that $|b_1(t) - b_2(t)| = 0$. The coupling above shows that if the bin size gap reaches 0 at time t, the random walk must have also reached 0 at time t or earlier. We now bound the probability that the random walk reaches 0.

This bound uses a gambler's ruin argument (see Appendix B). The probability of ruin starting from position x is

$$\left(\frac{q}{p}\right)^x = \left(\frac{(1-\epsilon)/2}{(1+\epsilon)/2}\right)^x = \left(\frac{1-\epsilon}{1+\epsilon}\right)^x.$$

Therefore to bound this probability by $e^{-\delta}$, we need

$$x \geq \frac{\delta}{\log((1+\epsilon)/(1-\epsilon))}.$$

Since $1/2 + \epsilon/2 = \frac{n^d-(n-1)^d}{n^d-(n-2)^d}$, we know that $\frac{1+\epsilon}{1-\epsilon} = \frac{n^d-(n-1)^d}{(n-1)^d-(n-2)^d}$. Recall that n is approaching infinity, so we can estimate using leading terms

$$\frac{n^d-(n-1)^d}{(n-1)^d-(n-2)^d} \sim \frac{n^d-n^d+dn^{d-1}-\binom{d}{2}n^{d-2}}{n^d-dn^{d-1}+\binom{d}{2}n^{d-2}-n^d+2dn^{d-1}-4\binom{d}{2}n^{d-2}}$$
$$= \frac{dn-\binom{d}{2}}{dn-3\binom{d}{2}} = 1+\frac{2\binom{d}{2}}{dn-3\binom{d}{2}} = 1+\frac{d-1}{n-(3/2)(d-1)}$$

For small α, $\log(1+\alpha)\sim\alpha$. Therefore $\log(1+\frac{d-1}{n-(3/2)(d-1)})\sim\frac{d-1}{n-(3/2)(d-1)}$. So

$$\frac{\delta n}{(d-1)} > \frac{\delta(n-(3/2)(d-1))}{(d-1)} \sim \frac{\delta}{\log\left(1+\frac{(d-1)}{n-(3/2)(d-1)}\right)}$$
$$\sim \frac{\delta}{\log\frac{n^d-(n-1)^d}{(n-1)^d-(n-2)^d}}.$$

If $x > \frac{\delta n}{2(d-1)}$, for n sufficiently large, then a random walk starting at x with bias ϵ will reach 0 with probability less than $e^{-\delta}$. Therefore the probability of two bins switching position under GREEDY(m,n), if they start at least $\frac{\delta n}{2(d-1)}$ apart, is less than $e^{-\delta}$ □

Note that the gap needed is linear in n. So for $m = cn$, for example, this theorem is not useful; we know the largest bin has size on the order of $\log n/\log\log n$. Therefore there are no two bins that are far enough apart from each other for this theorem to apply.

However, once m is larger, this gap does appear. For example, if $m = \delta n^2/2(d-1)$, then the expected bin size under UNIFORM(m,n) is $\delta n/2(d-1)$. We know that the bin sizes under GREEDY(m,n) majorize UNIFORM(m,n), so that means there is expected to be a bin under GREEDY(m,n) of size at least $\delta n/2(d-1)$. Therefore the largest bin has size at least $\delta n/2(d-1)$.

We also know from the upper bounds on the number of balls in the last few bins, that that last $\left(\frac{2c(d-1)}{\delta}\right)^{1/d} n^{1-2/d}$ bins have at most c balls total. Therefore the last $\left(\frac{2c(d-1)}{\delta}\right)^{1/d} n^{1-2/d} - c$ bins are empty. So for constant c the gap between the largest bin and the smallest $\left(\frac{2c(d-1)}{\delta}\right)^{1/d} n^{1-2/d} - c$ bins is at least $\delta n/2(d-1)$. So we can apply this theorem to see that once $\delta n^2/2(d-1)$ balls have been distributed, none of the smallest $\left(\frac{2c(d-1)}{\delta}\right)^{1/d} n^{1-2/d} - c$ will become the largest bin at any time in the future.

4.2 Bins of equal size

If we knew putting a ball in the currently i^{th} largest bin at time t would make the i^{th} largest bin one ball bigger at time $t+1$, it would be easy to predict how big each bin gets. Unfortunately, that's not true. If all the bins between i and j have the same size, then putting a ball in any bin between i and j will mean the i^{th} bin is the one that increases. So it is useful to study how strings of equally-sized bins can occur.

In this section, we first prove a warm-up theorem about how putting a ball in one bin instead of another can affect the final distribution. We then bound the probability that any two bins have a constant size gap. In particular, this bounds the probability that any two bins have the same size. This gives us a bound on how long any string of equally-sized bins is likely to be.

Theorem 4.2.1. *For any distribution of balls* $\mathbf{b} = (b_1, b_2, \ldots b_n)$*, and any* $i \neq j$*, consider the effect of* GREEDY(m, n) *on initial configurations of* $\mathbf{b} + e_i$ *and* $\mathbf{b} + e_j$*. For* $m = O(n \log n)$*, the final size of* B_i *starting from* $\mathbf{b} + e_i$ *will be greater than the final size of* B_i *starting from* $\mathbf{b} + e_j$ *with high probability.*

Proof. One would think that giving a ball to one bin instead of another would have only a positive effect on the winning bin's final size. However, it could happen that the bin that initially receives the ball would have been better off if the ball had gone to the other bin. There are some sequences of option sets that can lead to B_i's final size being greater starting from $\mathbf{b} + e_j$ than $\mathbf{b} + e_i$.

For example, suppose $\mathbf{b} = (1, 1, 1)$, $i = 1$, and $j = 2$ (so we are comparing GREEDY on $\mathbf{b} + e_1 = (2, 1, 1)$ and $\mathbf{b} + e_2 = (1, 2, 1)$). Further suppose that the option sets are $S_1 = S_2 = \{1, 3\}$ and $S_3 = S_4 = \{2, 3\}$. The following table shows possible outcomes.

Note that a tie is broken at time $t = 3$ for initial configuration $(2, 1, 1)$ and at time

Table 4.1: Bin sizes under different configurations and tie breaks

	$B_2 > B_3$	$B_3 > B_2$	$B_1 > B_3$	$B_3 > B_1$
$t = 0$	211	211	121	121
$t = 1$	311	311	221	122
$t = 2$	411	411	321	123
$t = 3$	421	412	331	124
$t = 4$	431	413	341	125

$t = 1$ for initial configuration $(1, 2, 1)$. Recall that GREEDY breaks ties uniformly at random. In the above table, $B_i > B_j$ indicates that the tie between bins B_i and B_j is broken in favor of bin B_i. Therefore we see that, if the tie is randomly broken in favor of B_2 at time $t = 3$ and in favor of B_3 at time $t = 1$, $b_2(4) = 3$ starting from $(2, 1, 1)$ and $b_2(4) = 2$ starting from $(1, 2, 1)$. The effect of bin B_2 being larger initially is to make B_2 smaller after more balls have been placed.

Note that the difficulty in the above case was the result of option sets intersecting. Ball placement at time 1 influenced placement at time 2, for example, because their option sets were the same. In general, if the option sets containing B_i and the option sets containing B_j have no intersection or chain of intersections, it is impossible for an extra ball in B_i to cause B_j to increase. This condition is necessary for our "paradoxical" event to occur.

We analyze the probability of options sets intersecting now. For ease of notation, assume the starting configurations are $\mathbf{b} + e_1$ and $\mathbf{b} + e_2$. We call the chain of intersections generated by a sequence of option sets an "influence set". We define T_i to be the influence set at time i. Initially, $T_0 = \{B_1, B_2\}$. Given a sequence of option sets $\{S_i\}$, define T_i recursively.

$$T_i = T_{i-1} \cup \{x | x \in S_i \text{ and } S_i \cap T_{i-1} \neq \emptyset\}$$

Note that the only case in which there could be a paradox is if a subsequent option set S_t contains B_1 or B_2 and some other bin which was already influenced by the initial choice of e_1 or e_2. So it is enough to bound the probability of $|S_t \cap T_{t-1}| \geq 2$.

Let's do so. If S_t does contain both B_1 or B_2 and some bin that is in T_{t-1}, then there must be a subsequence of option sets $S_t, S_{t_{r-1}}, S_{t_{r-2}}, \ldots S_{t_1}$ such that S_{t_i} intersects $S_{t_{i-1}}$ and S_{t_1} contains B_1 or B_2. There are $\binom{t}{r}$ choices for indices of a subsequence of length r. The probability that any particular length-r subsequence is intersecting is bounded by $(d/n)^r$. There are two choices for S_{t_1} and S_t, to contain B_1 or B_2, and the probability of each is less than d/n. So the overall probability of an intersecting subsequence of length r is bounded by $4\binom{t}{r}(d/n)^{r+1}$. Then for the existence of any sequence of any length, we have the bound

$$4\sum_{r=1}^{t}\binom{t}{r}(d/n)^r \leq (4d/n)(1+d/n)^r \to (4d/n)e^{td/n}.$$

Note that when $t = cn\log n$ for c any constant, this is $O(n^{c-1})$. In particular, if $c < 1$, this is $o(1)$. □

We now turn to a more complex theorem.

Theorem 4.2.2. *For each j, for any constant δ, for all $m = \Omega(n^2 \log n)$ any pair of bins outside a set of size δn are with high probability at least j balls different from each other.*

Proof. This theorem uses the same sorts of ideas in its proof as the previous one did. In order to show two bins are likely to have different sizes, we analyze the option sets and determine which types of option sets have an effect. We then count the number of effective option sets. Finally, we bound how likely it is that a sequence of option sets will have the wrong effect.

We first bound the probability of any two bins having the same size, once the process has run a sufficiently long time. Let the number of balls to be distributed be $m = m' + t$, where appropriate values for m' and t will be discussed later. We fix two bins, A and B, and consider the probability that they are of the same size after all m balls have been distributed. We distribute the balls in two phases. In the first phase, distribute m' balls. We will then analyze the remaining t steps based on the bins' positions at time m'.

We first find a set of bins that are much smaller than bin A or bin B at time m'. If we assume that A and B are within the largest $(1-\delta)n$ bins at time m', we can fix g, and find ϵ such that the last ϵ bins at time m each have size at most $\min\{a(m), b(m)\} - g$. (Appropriate values for δ, g and ϵ will be discussed later.)

Now look at the final t rounds. We'll reveal the option sets $\{S_i\}_{i=1}^{t}$ in several phases. In the first phase, we'll reveal all the sets that don't contain A or B. Then, reveal all the sets that contain A or B and at least one bin outside of the least ϵ at time m'.

Call the remaining sets "important". These sets are useful, as (since the bins competing with A or B are all much smaller), they guarantee A or B gets a ball. To assure this guarantee, we also need to be careful that the other bins in the important option sets haven't gotten too much bigger, i.e. they are still smaller than A or B. We choose the gap size x so that cannot happen. The exact value of x is discussed below.

Now reveal all the non-A or B elements of the q important option sets. Each of them contains exactly one of A or B. So there are 2^q possibilities, $\{A, B\}^q$, once the other elements are revealed. Let's partially order the vectors by setting $A < B$ (so, e.g., $ABBAB < BBBAB$). We will show that this partial ordering corresponds to an ordering in terms of bin size: let $\mathbf{v} < \mathbf{u}$ if, for any fixed sequence of sets with A or B hidden in the q important sets, the $\mathbf{v}$ sequence of As and Bs would result in fewer balls in bin B and more in bin A than the $\mathbf{u}$ sequence with high probability.

This is just a repetition of the theorem above. We already know that, as t was chosen to be small, there is unlikely to be a chain of intersection within the option sets. So the placement of a ball into A or B at any of the important stages is unlikely to increase the size of the other bin. If $\mathbf{v} < \mathbf{u}$ under the $A < B$ vector ordering, then the $\mathbf{v}$ sequence will generate a smaller B than the $\mathbf{u}$ sequence, unless there is a chain of intersection within the option sets. The previous theorem guarantees with high probability that this won't happen.

This tells us that, with high probability, the set of revealed sequences such that A and B have the same size is an anti-chain. Therefore, with high probability, it has size at most $\binom{q}{q/2}$ by Sperner's Lemma. So, given that there are exactly q important option sets, and that the gap between the last ϵn bins and A and B is as expected, and that the sequence of important sets is non-intersecting, the probability of A and B having the same number of balls after all m balls have been distributed is at most $\frac{\binom{q}{q/2}}{2^q}$. We can bound q so that this probability is small, to complete the proof.

We now calculate the actual probabilities. First consider δ. Note that, if A and B are in the upper $1-\delta$ proportion of bins at time m', then A and B have at least as many balls as the δn^{th} bin at time m'. That bin has at least as many balls as the average of the last δn bins' sizes. Recall that we know the number of balls in the last δ bins is at least (number of times option set is within last δ)/$(e^{1/2}(2d-2))$. So if the option set is within the last δ x times, then the number of balls in the last δ bins is at least $x/e^{1/2}(2d-2)$.

Note that the option sets are distributed uniformly, so we can use the standard Chernoff bound

$$Pr(x < \delta^d m' - \alpha) < e^{-2\alpha^2/m'}$$

to see that the number of balls in the last δ bins is at least $\delta^d m' - \alpha$ with probability at least $1 - e^{-2\alpha/m'}$. Because the δn^{th} bin has at least as many balls as the average, if A and B are in the upper $(1-\delta)n$ bins, they will each have at least

$$\frac{\delta^d m' - \alpha}{\delta n}$$

balls with probability at least

$$1 - e^{-2\alpha^2/m'}.$$

Now turn to ϵ. Suppose the number of balls in the least $2\epsilon n$ bins at time m' is x. Then the ϵn^{th} smallest bin would contain at most $x/\epsilon n$ balls: because there are x balls total, and the next ϵn bins each have at least as many balls as the ϵn^{th} smallest, $\epsilon n y \leq x$ where y is the number of balls in the ϵn^{th} bin.

Furthermore, we can bound x using previous results. Recall that x is at most the number of times the option set is within the last $2\epsilon n$. Again, the Chernoff bound tells us that

$$Pr(x > (2\epsilon)^d m' + \beta) < e^{-2\beta^2/m'}.$$

So the least ϵn bins will each have size at most

$$\frac{2^d \epsilon^d m' + \beta}{\epsilon n}$$

with probability at least

$$1 - e^{-2\beta^2/m'}.$$

We can combine these two results to see that, with probability at least

$$(1 - e^{-2\alpha^2/m'})(1 - e^{-2\beta^2/m'}),$$

the gap between A or B and any bin within the least ϵn will have size at least

$$\frac{\delta^d m' - \alpha}{\delta n} - \frac{2^d \epsilon^d m' + \beta}{\epsilon n}.$$

To make the probability of error exponentially small in n, it is thus sufficient to let $\alpha = \sqrt{\alpha' n m'}$ and $\beta = \sqrt{\beta' m' n}$ for any constants α' and β'. Since $\sqrt{nm'} < m'$ for $m' = \Omega(n^2 \log n)$, by choosing suitable ϵ given δ, it is possible to guarantee a gap of size $g = c_0(m'/n)$ for any arbitrary constant c_0.

Recall that we assumed the gaps between A or B and the least ϵn were so large at time m' that A and B would still be larger than the least ϵn at time $m' + t$. We now determine exactly how large a gap is necessary to guarantee it will not be closed after t steps. Each bin in the least ϵn may increase only if it is a member of an option set. So it is enough to bound the number of times any of the least bins appear in an option set during the last t time steps. Again, note that the number of times x that one of the least ϵn bins is an option can be Chernoff bounded: $Pr(x > \epsilon dt + \lambda) \leq e^{-2\lambda^2/t}$. If $\lambda = \sqrt{\lambda' t n}$ for some constant λ', this bound is again exponentially small in n.

We now turn to t and q. First note that we can again use Chernoff bounds to guarantee a value of q within a certain range. The probability of an option set being important is $2((\epsilon+1/n)^d-\epsilon^d)$. Therefore the number of important rounds in the last t steps is at least $2((\epsilon+1/n)^d-\epsilon^d)t-\gamma$ with probability at least $1-e^{-2\gamma^2/t}$. Again, setting $\gamma=\sqrt{\gamma' nt}$ is sufficient to give an exponentially small error probability.

Given that the important sets are non-intersecting, the probability of the important set sequence making A and B have equal size is at most

$$\frac{\binom{q}{q/2}}{2^q}\sim\frac{\sqrt{2}}{\sqrt{\pi q}}\leq\frac{\sqrt{2}}{\sqrt{\pi(2((\epsilon+1/n)^d-\epsilon^d)t-\gamma)}}.$$

We finally choose t. To guarantee the smaller bins don't get too big, we need t such that

$$g\geq c_0(m'/n)>\epsilon dt+\lambda'\sqrt{tn}.$$

To guarantee the important set sequence doesn't make A and B equal, it is enough to find t such that

$$\frac{\sqrt{2}}{\sqrt{\pi(2((\epsilon+1/n)^d-\epsilon^d)t-\gamma)}}=O(\frac{1}{\sqrt{\log n}}).$$

Setting $t\geq c_1 n\log n$ and $m'\geq(c_1/c_0)\epsilon dn^2\log n$ for any constant c_1 is sufficient for both conditions. So overall we need

$$m=m'+t'\geq(c_1)((\epsilon d/c_0)n+1)n\log n$$

for arbitrary constants ϵ, c_0, c_1. Therefore $m=\Omega(n^2\log n)$ is enough to guarantee A and B are not equal with probability

$$\frac{(1-\delta)^2(1-e^{-2\alpha' n})(1-e^{-2\beta' n})(1-e^{-2\lambda' n})(1-e^{-2\gamma' n})}{c\sqrt{\log n}}.$$

Note that the same argument could apply for any arbitrary difference j between A and B. Important sets are the same, and the sequences that guarantee B is exactly j balls greater than A still form an antichain. So B and A are almost surely at least j balls apart for $m=\Omega(n^2\log n)$.

□

Furthermore, we can use this result to bound the total number of equally sized bins.

Corollary 4.2.3. *With high probability, for* $m = \Omega(n^2 \log n)$*, the greatest number of bins of the same size is less than* y*, for any* y *such that* $\frac{n}{(\log n)^{1/4}} = o(y)$*. (For example, with high probability there are no more than* $\frac{n}{(\log n)^{1/5}}$ *bins of any one size.)*

Proof. First, let x_i be the number of bins of the same size as B_i. Then

$$E\left(\sum_{i=1}^{n} x_i) = O(n^2/\sqrt{\log n}\right).$$

We can use Markov's Inequality to see that

$$Pr\left(\sum_{i=1}^{n} x_i > a\right) = O(n^2/a\sqrt{\log n}).$$

Now, suppose there were y bins of the same size. Then the sum would be at least $y(y-1)/2$ (as there would be at least $\binom{y}{2}$ pairs of the same size). The probability of that happening is

$$O(n^2/y^2\sqrt{\log n}).$$

If y is such that

$$\frac{n}{(\log n)^{1/4}} = o(y),$$

this probability goes to zero. $\square$

Appendix A

Kurtz's theorem

We now justify the use of differential equations in calculating expected values. This appendix will define necessary terms, state Kurtz's theorem, and demonstrate the technique by applying it to a simple example. The source of this theorem is [7]; it was originally used in this context in [8].

Definition 4. *A density dependent family of Markov chains is a sequence of jump Markov processes* $\{X_n\}$ *that all have the same transition functions, once renormalized. More formally, it satisfies the following requirements: For each* $n \in \mathbb{N}$*, the state space of* X_n *is contained within* $\{n^{-1}k | k \in \mathbb{Z}^*\}$*, where* $\mathbb{Z}^*$ *is* $\mathbb{Z}^d$ *for some constant* d *or* $\mathbb{Z}^{\mathbb{N}}$*. There exists a set of non-negative functions* β_l*, where* l *ranges over the set of transitions* L*, that generates the transition probability for any pair of states in any chain* X_n*. That is, the probability of going from* x *to* y *in* X_m *is* $q_{x,y}^{(n)} = \beta_{n(x-y)}(x)$*. For convenience, let* $F(x) = \sum_{l \in L} l\beta_l(x)$*, i.e. the expected change in* x*.*

Note that GREEDY, FAIR and UNIFORM all generate density dependent families of Markov chains. For example, consider UNIFORM(m, n). In this setting, $\mathbb{Z}^* = \mathbb{Z}^{\mathbb{N}}$. The i^{th} coordinate represents the proportion of bins of size i. The possible transitions are that, for arbitrary i, the number of bins of size i increases by one and the number of size $i-1$ decreases by one. Therefore $L = \{e_i - e_{i-1}\} \subset \mathbb{Z}^{\mathbb{N}}$. The probability of the transition $(e_i - e_{i-1})/n$ is just the proportion of bins of size $i-1$. Therefore $\beta_{e_i - e_{i-1}}(x) = e_{i-1} \cdot x$. Note that β doesn't change with n. For all X_n, i.e. for all numbers of bins, the probability of a bin statistic increasing or decreasing is dependent on the same coordinate in the same way.

Theorem A.0.4 (Kurtz). *Suppose $\{X_n\}$ is a density dependent family of Markov chains such that $|F(x) - F(y)| \leq M\,|x - y|$ for some constant M, $\lim_{n\to\infty} X_n(0) = x_0$. Let $X(t) = x_0 + \int_0^t F(X(u))du$. If there exists a neighborhood K around $\{X(u)|u \leq t\}$ such that $\sum_{l\in L} |l| \sup_{x\in K} \beta_l(x)$ is finite, then $\lim_{n\to\infty} \sup_{u\leq t} |X_n(u) - X(u)| = 0$. In other words, as n tends to infinity, X_n tends to X.*

Note that X can be thought of as a solution to the differential equation $\dot{X} = F(X)$. In most applications of Kurtz's theorem, this differential equation is solved, then used to predict the limit of the Markov process. Let's see how this works by doing a simple example, analyzing UNIFORM.

First, UNIFORM(m, n) is a density dependent family of Markov chains, so the first condition is satisfied. Second,

$$F(x) = \sum_{l\in L} l\beta_l(x) = \sum_{i=1}^{m} (e_i - e_{i-1})(e_{i-1} \cdot x) = \sigma x - x,$$

where σ is the operation that translates each coordinate to its right neighbor, and leaves the first coordinate 0, i.e. $\sigma(x_0, \ldots x_n) = 0, x_1, \ldots x_{n-1}$. Therefore

$$|F(x) - F(y)| \leq |x - y| + |\sigma x - \sigma y| \leq 2\,|x - y|$$

and the second condition is satisfied. Third, $X_n(0) = e_0$ for all n, so in particular $\lim_{n\to\infty} X_n(0) = e_0$ and $x_0 = e_0$.

To check the next condition, we must first find X. Note that $X(t) = \sum_{i=0}^{m+1} \frac{t^i e^{-t}}{i!}$ satisfies $\dot{X} = F(X)$. (We normalize so that nt balls have been placed at time t.) Because we have the Lipschitz condition $|F(x) - F(y)| \leq M\,|x - y|$ and the initial values, there is exactly one X that satisfies $\dot{X} = F(X)$. So we know that is the X we seek.

Now we find K a neighborhood of $\{X(u)|u \leq t\}$ that makes the sum finite. This is difficult; for each coordinate $l \leq nt$, $\beta_l(x)$ can be as large as

$$\frac{l^l e^{-l}}{l!} \sim \frac{1}{\sqrt{2\pi l}}.$$

Therefore

$$\sum_{l \in L} |l| \sup_{x \in \{X(u) | u \leq t\}} \beta_l(x) \geq \sum_{i=0}^{nt} 2\left(1/\sqrt{2\pi l}\right) \geq (1/\sqrt{2\pi}) \log nt - 1,$$

which diverges as n goes to infinity.

However, for specific applications, we can find a converging sum. Suppose we just want to know the number of bins of some constant size, say c. Then we could look at a new X_n, where we truncate each vector at the c^{th} coordinate. The i^{th} coordinate is still the proportion of bins of size i for $i \leq c$. Note that the $\beta_l(x_i)$ depends only on x_j for $j \leq i$, so this truncated version will still accurately simulate the smaller bin statistics. Furthermore, we can now let K be $[-1, 2]^{c+1}$ and bound $\sup_{x \in K} \beta_l(x)$ by 2, to get

$$\sum_{l \in L} |l| \sup_{x \in k} \beta_l(x) \leq 6c$$

for all t and n. So the final condition is satisfied. Therefore as n tends to infinity, at fixed time t the expected number of bins of size i for any constant approaches $\frac{t^i e^{-t}}{i!}$. For example, if we consider UNIFORM(n, n), then the expected proportion of bins of size 1 is e^{-1}. The same technique can be used for any constant bin size at any constant time, i.e. for any number of balls that is a constant multiple of n.

Appendix B

Gambler's Ruin

We now discuss the random walk results necessary in bounding bin position switches. We start with gambler's ruin in the finite case, and then turn to the infinite case. Much of this discussion comes from [4].

Definition 5. *A random walk will here refer to a random walk on* $\mathbb{N}$*, which makes a +1 step with probability* p *and a -1 step with probability* q*. More formally, let*

$$X_{k,c} = c + \sum_{i=1}^{k} Y_i,$$

where the Y_i *are independent random variables, each with probability* p *of being 1 and* q *of being -1.* $X_{k,c}$ *thus represents the position of a random walk on* $\mathbb{Z}$ *at time* k *starting at* c*.*

We will be considering the relationship between the starting position c and the probability of the walk reaching 0. We first consider the finite case, calculating the probability of reaching 0 before some constant a. We then use this result to calculate the probability of ever reaching 0, for an arbitrarily long walk.

Theorem B.0.5. *The probability of a random walk as defined above reaching* 0 *before it reaches* a *is* $\frac{(q/p)^a - (q/p)^c}{(q/p)^a - 1}$.

Proof. We first define q_z to be the probability of reaching 0 before a from starting position z. Then for $z = 1, 2, \ldots a-1$ we have

$$q_z = pq_{z+1} + qq_{z-1}.$$

That's because from z the next step is to $z+1$ with probability p and to $z-1$ with probability q. We can rearrange that equation to see

$$p(q_{z+1} - q_z) = q(q_z - q_{z-1}).$$

Therefore

$$q_{z+1} - q_z = (q/p)(q_z - q_{z-1}).$$

We also know $q_a = 0$ and $q_0 = 1$. Since we have difference equations and boundary conditions, we can now solve for q_c and get

$$\frac{(q/p)^a - (q/p)^c}{(q/p)^a - 1},$$

as desired. □

We now consider the infinite case.

Theorem B.0.6. *The probability of a random walk as defined above reaching* 0 *at any time is* $\min\{(q/p)^c, 1\}$

Proof. We simply take the limit of the previous formula as $a \to \infty$. When $q < p$, $\lim_{a\to\infty}(q/p)^a = 0$, and

$$\begin{aligned}
\lim_{a\to\infty} \frac{(q/p)^a - (q/p)^c}{(q/p)^a - 1} &= \lim_{a\to\infty} \frac{(q/p)^a - 1}{(q/p)^a - 1} + \frac{1 - (q/p)^c}{(q/p)^a - 1} \\
&= 1 - \frac{1 - (q/p)^c}{1 - (q/p)^a} \\
&= \left(\frac{q}{p}\right)^c.
\end{aligned}$$

When $q \geq p$, $\lim_{a\to\infty}(q/p)^a = \infty$, and

$$\begin{aligned}
\lim_{a\to\infty} \frac{(q/p)^a - (q/p)^c}{(q/p)^a - 1} &= \lim_{a\to\infty} \frac{(q/p)^a - 1}{(q/p)^a - 1} + \frac{1 - (q/p)^c}{(q/p)^a - 1} \\
&= 1 - \frac{1 - (q/p)^c}{1 - (q/p)^a} = 1.
\end{aligned}$$

□

This theorem can be used to find the necessary bias for a random walk to reach 0 with a certain low probability, which is what we need for our applications.

Definition 6. *Say a random walk has bias ϵ if the probability of a +1 step is $(1+\epsilon)/2$ and a -1 step is $(1-\epsilon)/2$.*

Then we have the following corollary.

Corollary B.0.7. *To guarantee probability of at most $e^{-\delta}$ of reaching 0 for a random walk with bias ϵ, it is sufficient to start at*

$$c = \frac{\delta}{\log((1+\epsilon)/(1-\epsilon))}.$$

This follows from the theorem by some simple manipulation. When the bias is ϵ, $p = (1+\epsilon)/2$ and $q = (1-\epsilon)/2$. So $(q/p)^c$ becomes $\left(\frac{1-\epsilon}{1+\epsilon}\right)^c$. Then

$$\left(\frac{1-\epsilon}{1+\epsilon}\right)^c < e^{-\delta}$$

when

$$c(\log(1-\epsilon) - \log(1-\epsilon)) < -\delta,$$

and we see

$$c > \frac{\delta}{\log(\frac{1+\epsilon}{1-\epsilon})}.$$

Bibliography

[1] Noga Alon and Joel Spencer. *The probabilistic method.* John Wiley and Sons, Inc., Hoboken, New Jersey, 2008.

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

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

[4] William Feller. *An introduction to probability theory and its applications.* John Wiley and Sons, Inc., New York, 1968.

[5] P. Brighten Godfrey. Balls and bins with structure: balanced allocations on hypergraphs. In *Proceedings of the Nineteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 511–517, New York, 2008. ACM.

[6] Krishnaram Kenthapadi and Rina Panigrahy. Balanced allocation on graphs. In *Proceedings of the Seventeenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 434–443, New York, 2006. ACM.

[7] Thomas G. Kurtz. *Approximation of population processes*, volume 36 of *CBMS-NSF Regional Conference Series in Applied Mathematics.* Society for Industrial and Applied Mathematics (SIAM), Philadelphia, Pa., 1981.

[8] Michael Mitzenmacher. Studying balanced allocations with differential equations. *Combin. Probab. Comput.*, 8(5):473–482, 1999.

[9] Berthold Vöcking. How asymmetry helps load balancing. *J. ACM*, 50(4):568–589 (electronic), 2003.