https://pubsonline.informs.org/journal/stsy

STOCHASTIC SYSTEMS
Vol. 16, No. 3, September 2026, pp. 307–313
ISSN 1946-5238 (online)

About a Ball Removal Process on Bins

José Correa,[a,b,*] Marcos Kiwi,[b,c] Vasilis Livanos,[b] Eilon Solan,[d] Ron Solan[e]

[a] Departamento de Ingeniería Industrial, Universidad de Chile, Santiago 837-0456, Chile; [b] Centro de Modelamiento Matemático (Centre National de la Recherche Scientifique International Research Laboratory 2807), Universidad de Chile, Santiago 837-0456, Chile; [c] Departamento de Ingeniería Matemática, Universidad de Chile, Santiago 837-0456, Chile; [d] School of Mathematical Sciences, Tel Aviv University, Tel Aviv 6997801, Israel; [e] Broad Institute of Massachusetts Institute of Technology and Harvard, Cambridge, Massachusetts 02142

*Corresponding author

✉ correa@uchile.cl (J. Correa), mk@dim.uchile.cl (M. Kiwi), vas.livanos@gmail.com (V. Livanos), eilonsolan@gmail.com (E. Solan), ron.s.solan@gmail.com (R. Solan)

https://orcid.org/0000-0002-3012-7622 (J. Correa), https://orcid.org/0000-0003-4171-2656 (M. Kiwi), https://orcid.org/0009-0001-1037-8109 (E. Solan),

Received: March 22, 2026
Revised: August 3, 2026
Accepted: August 13, 2026
Published Online in Articles in Advance: September 9, 2026

https://doi.org/10.1287/stsy.2026.0149

Copyright: © 2026 The Author(s)

Open Access Statement: This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. You are free to download this work and share with others, but cannot change in any way or use commercially without permission, and you must attribute this work as "*Stochastic Systems*. Copyright © 2026 The Author(s). https://doi.org/10.1287/stsy.2026.0149, used under a Creative Commons Attribution License: https://creativecommons.org/licenses/by-nc-nd/4.0/."

Abstract. We consider a basic balls-and-bins question in which you have to distribute n balls into k bins. Then, round by round, a ball is removed from a nonempty bin chosen uniformly at random. The process ends when a single nonempty bin remains. The goal is to minimize the expected number of remaining balls. An open problem posed by Will Ma asks whether the initial assignment that minimizes the expected number of remaining balls is one that is as balanced as possible. Using a coupling argument, we answer this conjecture positively, and we discuss the case of nonuniform choice among the nonempty bins.

Funding: Financial support by the Israel Science Foundation [Grant 211/22] and by Agencia Nacional de Investigación y Desarrollo Chile through the Programa de Investigación Asociativa (Project FB210005) and Fondo Nacional de Desarrollo Científico y Tecnológico [Grant 1260036] is gratefully acknowledged.

Keywords: assortment optimization • inventory control • balls and bins

1. Introduction

You have to distribute n balls into k bins. Then, round by round, a ball is removed from a nonempty bin chosen uniformly at random. The process ends when a single nonempty bin remains. The goal is to minimize the expected number of remaining balls. This problem was presented by Will Ma during an open problems session at the Simons Institute for the Theory of Computing. He conjectured that the initial assignments that minimize the expected number of remaining balls are those that distribute the n balls into the k bins as evenly as possible; in other words, the number of balls between any two bins differs by at most one. The purpose of this note is to affirmatively settle this conjecture.

Although this conjecture is very natural and interesting in its own right, one practical motivation comes from assortment optimization. At the beginning of the selling horizon, a seller selects the number of units to stock for each product. Each arriving customer makes a choice among the set of products with remaining inventory. The seller's goal is to pick the stocking quantities to maximize the total expected revenue from the sales net of the stocking cost. Various approximations of the optimal solution under different assumptions on the data have been provided—for example, by Liang et al. (2021), Zhang et al. (2025), and Mouchtaki et al. (2026). The problem we study is a variation of this model, where products are perfect substitutes, customers are indifferent among them, and they are happy and make a purchase whenever they have a choice of products. In this case, they randomly purchase one unit of one product. The above conjecture is equivalent to determining how to initially distribute n items across k product types to maximize the expected number of happy customers. Although unrealistic as a model of customer behavior, it is arguably a particularly simple and mathematically interesting model that is not yet well understood. It is thus a relevant setting in which to develop tools for addressing more general and realistic models.

Downloaded from informs.org by [186.11.81.201] on 13 October 2026, at 04:18 . For personal use only, all rights reserved.

Downloaded from informs.org by [186.11.81.201] on 13 October 2026, at 04:18 . For personal use only, all rights reserved.

To formalize the problem, we assume that the k bins are labeled by $[k] := \{1,\dots,k\}$ and that an initial assignment of balls to bins is given by $\vec{n} := (n_1,\dots,n_k) \in \mathbb{N}^k$, where $\mathbb{N} = \{0,1,\dots\}$. Instead of the process implicit in the problem description, we consider the following process: at each round, a bin (possibly empty) is chosen according to the uniform distribution on $[k]$, independently of past choices, and if the bin is nonempty, one ball is removed from it. Thus, the sample space is $[k]^\infty$, and the probability distribution $\mathbf{P}$ is the product probability measure under which the coordinates are independent and uniformly distributed on $[k]$. This representation is independent of the initial assignment, because a point in the sample space determines the sequence of selected bins.

For each initial assignment $\vec{n}$, denote by $X_{\vec{n}}$ the (random) number of balls in the last nonempty bin. Thus, if there are two bins and two balls, then $X_{(1,1)}$ and $X_{(2,0)}$ equal one and two, respectively.

We number rounds starting at 0: round 0 refers to the initial configuration, and, for $t \geq 1$, the configuration at the end of round t is the configuration after the first t selections. Let $T(\vec{n})$ denote the first (random) round $t \geq 0$ at which the modified removal process started from $\vec{n}$ has a unique nonempty bin. Note that $X_{\vec{n}}$ is the number of balls remaining at round $T(\vec{n})$ of the modified removal process. Crucially, this random variable has the same distribution as the number of balls remaining when the unmodified removal process terminates. Define

$$f(\vec{n}) := \mathbb{E}[X_{\vec{n}}], \quad \text{for all } \vec{n} \in \mathbb{N}^k.$$

We abuse notation and extend, in the natural way, the domain of f to $\cup_{m=2}^{\infty} \mathbb{N}^m$.

From the definition, we see that $f(a,0) = a$, $f(0,b) = b$, and

$$f(a,b) = \frac{1}{2}f(a-1,b) + \frac{1}{2}f(a,b-1), \quad \text{for all } a,b \in \mathbb{N} \setminus \{0\}. \tag{1}$$

Our main result is:

Theorem 1. *Fix $n,k \geq 2$. The quantity $\min\{f(\vec{n}) : \vec{n} \in \mathbb{N}^k, \sum_{i\in[k]} n_i = n\}$ is attained only by assignments $\vec{n} = (n_1,\dots,n_k) \in \mathbb{N}^k$ satisfying $\sum_{i\in[k]} n_i = n$ and $|n_i - n_j| \leq 1$ for every $i,j \in [k]$.*

The result was previously proved for $k = 2,3$ in Gutiérrez and Subercaseaux (2024). For $k = 3$, the proof argument is a complicated induction (on the initial number of balls n) of six mutually dependent claims. Each claim establishes, under suitable conditions, an inequality concerning the expected number of balls removed throughout the process. The complexity of the induction suggests that any generalization would require significant effort and provide little insight.

Independently of us, Zhou et al. (2026) proved Theorem 1 for any k, by embedding the discrete-time process into a continuous-time analytical framework. Specifically, the embedding recasts the time at which bins empty into independent Erlang random variables. This allows reframing the problem as one concerning order statistics of k independent Erlang variables. This yields a closed-form expression for the marginal contribution of adding one additional ball to any given bin. The closed-form expression is then analyzed and used to establish monotonicity of the expected number of balls removed throughout the process, with respect to the initial number of balls per bin. This directly translates into the optimality of the balanced initial assignment. As observed in Zhou et al. (2026), one key analytical difficulty in analyzing the process is that the objective function—the expected number of balls removed—is not Schur-concave. Thus, typical optimization techniques are not applicable to the problem under consideration. Our proof uses a standard coupling argument and is simpler and significantly shorter.

2. Proof of the Main Result

For completeness, we start by proving Theorem 1 for the case $k = 2$, which involves a coupling argument.

Lemma 1. *Fix $n \in \mathbb{N}$. For every $\vec{n} = (n_1,n_2) \in \mathbb{N}^2$ such that $n_1 + n_2 = n$,*

$$f(\lceil n/2 \rceil, \lfloor n/2 \rfloor) \leq f(n_1,n_2),$$

with strict inequality whenever $|n_1 - n_2| > 1$.

Proof. Without loss of generality, assume $n_1 \geq n_2$. If $|n_1 - n_2| \leq 1$, then $n_1 = \lceil n/2 \rceil$, $n_2 = \lfloor n/2 \rfloor$, and there is nothing to prove. In what follows, we assume $n_1 \geq n_2 + 2$.

It suffices to show that $f(n_1,n_2) > f(n_1 - 1, n_2 + 1)$, because the result then follows by repeatedly applying this inequality. To see that the inequality holds, run the process from the initial assignment (n_1,n_2) and consider a sequence of removals in which, eventually, at some round, bin 1 contains exactly one ball more than bin 2. Denote by t the first round at which this happens and by n_1' the number of balls in bin 2 at the *end* of round t (so

Downloaded from informs.org by [186.11.81.201] on 13 October 2026, at 04:18 . For personal use only, all rights reserved.

that bin 1 contains $n_1' + 1$ at the end of that round). Starting from the initial assignment $(n_1 - 1, n_2 + 1)$, the same sequence of removals, up to round t, leaves bins 1 and 2 with n_1' and $n_1' + 1$ balls, respectively. By symmetry, in any such sequence of removals, the number of balls remaining when the process stops is identically distributed, no matter if the initial assignment is (n_1, n_2) or $(n_1 - 1, n_2 + 1)$.

Consider now the complementary situation in which a sequence of removals starting with the assignment (n_1, n_2) results in bin 1 having, throughout all rounds, at least two more balls than bin 2. Then, bin 1 is the last nonempty bin. Moreover, on the same sequence of removals, but starting with the assignment $(n_1 - 1, n_2 + 1)$, the process takes longer and ends with fewer balls. Finally, note that this complementary situation occurs with positive probability. Indeed, this situation contains, for example, the sequence of chosen bins $(2, 2, \ldots, 2)$ of n_2 consecutive choices of bin 2. □

Remark 1. Lemma 1 in fact yields the stronger statement that $f(n_1, n_2) > f(n_1 - 1, n_2 + 1)$ for all $\vec{n} = (n_1, n_2) \in \mathbb{N}^2$ with $n_1 \geq n_2 + 2$.

A simple property of f is the following:

Lemma 2. $f(r, 1, 1) \leq f(r, 1)$, *for every* $r \geq 1$, *with strict inequality whenever* $r \geq 2$.

Proof. We prove the lemma by induction on r. For $r = 1$, we have $f(1, 1, 1) = f(1, 1) = 1$. Assume that the lemma holds for a certain $r \geq 1$. Then,

$$f(r+1, 1, 1) = \frac{1}{3} f(r, 1, 1) + \frac{2}{3} f(r+1, 1), \tag{2}$$

$$\leq \frac{1}{3} f(r, 1) + \frac{1}{3} f(r, 1) + \frac{1}{3} f(r+1, 0), \tag{3}$$

$$< \frac{1}{2} f(r, 1) + \frac{1}{2} f(r+1, 0), \tag{4}$$

$$= f(r+1, 1), \tag{5}$$

where Equations (2) and (5) hold by the recursion that f satisfies, Equation (3) holds by the induction hypothesis and the recursion, and Equation (4) holds because $f(r, 1) \leq r < r + 1 = f(r+1, 0)$. □

We turn to Theorem 1. Fix an initial assignment $\vec{n}$ such that $|n_i - n_j| \geq 2$ for some i, j. We assume throughout that $n_1 \geq n_j$ for every $j \in [k]$.

Denote by $e_i = (0, \ldots, 0, 1, 0, \ldots, 0)$ the i-th standard unit vector in $\mathbb{R}^k$. By definition, the random variable $X_{\vec{n} - e_1 + e_i}$ indicates the number of balls in the last nonempty bin when the initial assignment is $\vec{n} - e_1 + e_i$. With this notation, for each i such that $n_1 \geq n_i + 2$,

$$\mathbf{E}[X_{\vec{n}}] > \mathbf{E}[X_{\vec{n} - e_1 + e_i}].$$

Note that if $n_1 = n_i + 1$, then by symmetry, $\mathbf{E}[X_{\vec{n}}] = \mathbf{E}[X_{\vec{n} - e_1 + e_i}]$. In what follows, fix i such that $n_1 \geq n_i + 2$.

Given an initial assignment, for every point in our sample space, we can calculate the number of balls in each bin at every round. It will be convenient to assume that if bin j was selected $k_{j,t}$ times up to round t, and if $k_{j,t} > n_j$, then at the end of round t, the bin contains a negative number of balls—that is, $n_j - k_{j,t}$ balls.

If $\vec{n}$ has only one nonempty coordinate, transferring one ball from that bin to bin i strictly decreases the objective because $f(r - 1, 1) \leq r - 1 < r = f(r, 0)$. We may therefore assume henceforth that at least two bins are initially nonempty.

Central to our proof is the following stopping time. Given an initial assignment $\vec{n}$, let $S(\vec{n})$ be the smallest $t \geq 0$ such that, at the end of round t, one of the following conditions holds:[1]

a. Bin 1 contains exactly one ball more than bin i.

b. There are exactly two nonempty bins, and one of them contains exactly one ball. Denote by $r \geq 1$ the number of balls in the other bin.

Thus, $S(\vec{n}) = 0$ if (at least) one of the conditions (a) or (b) already holds in the initial configuration.[2] Note also that $S(\vec{n})$ is defined relative to the process started from $\vec{n}$, even when it is subsequently used to couple this process with the one started from $\vec{n} - e_1 + e_i$.

From the definition of $S(\vec{n})$, we have $T(\vec{n}) \geq S(\vec{n})$. We next show that the same lower bound holds for $T(\vec{n} - e_1 + e_i)$. We include this result primarily for clarity, to help discard infeasible events.

Lemma 3. $T(\vec{n} - e_1 + e_i) \geq S(\vec{n})$.

Downloaded from informs.org by [186.11.81.201] on 13 October 2026, at 04:18 . For personal use only, all rights reserved.

Proof. If $S(\vec{n}) = 0$, the result is immediate. Assume henceforth that $S(\vec{n}) \geq 1$.

Take a specific sequence of bins that were selected along the process. If condition (a) is triggered, then at round $S(\vec{n})$, bin 1 has exactly one more ball than bin i. Hence, in the previous round, bin 1 contained exactly two more balls than bin i, and a ball was removed from bin 1 in round $S(\vec{n})$. Because condition (b) has not occurred earlier, there are at least two nonempty bins, and, if there are exactly two nonempty bins, both contain at least two balls. In both cases, when the initial assignment is $\vec{n} - e_1 + e_i$, the removal process does not end at round $S(\vec{n})$, and hence, $T(\vec{n} - e_1 + e_i) > S(\vec{n})$.

If condition (b) is triggered, then $T(\vec{n} - e_1 + e_i)$ may be smaller than $S(\vec{n})$ only if bin 1 contains one ball at round $S(\vec{n})$, so that when the initial assignment is $\vec{n} - e_1 + e_i$, it would contain no balls (if bin i contained more than one ball, then for (b) to be triggered, bin 1 must contain one ball and condition (a) would have triggered before). Because condition (b) was not triggered before, either at round $S(\vec{n})$ the last ball from a *third* bin j has been removed, or at round $S(\vec{n})$, the penultimate ball from bin 1 was removed. In both cases, when the initial assignment is $\vec{n} - e_1 + e_i$, the removal process has not ended before round $S(\vec{n})$, and hence, $T(\vec{n} - e_1 + e_i) \geq S(\vec{n})$. □

We will consider the following events:

- C_a: At round $S(\vec{n})$ condition (a) occurs—that is, bin 1 contains exactly one ball more than bin i.

Outside C_a, when condition (b) first occurs, exactly two nonempty bins remain, and one has one ball. We distinguish whether these bins are $\{1, i\}$, neither of them, or $\{1, j\}$ for some $j \notin \{1, i\}$; in the last case, we further distinguish whether the occupancies are $(1, 1)$, $(r, 1)$, or $(1, r)$. Note that because condition C_a does not occur, bin 1 always contains at least two balls more than bin i, and therefore, it cannot be that i but not 1 is among the last two nonempty bins. We therefore define:

- $F_{1,i}$: Bins 1 and i are the last two nonempty bins.
- $F_\emptyset$: The last two nonempty bins include neither 1 nor i.
- $F_{1,j}^{a,b}$ for $j \notin \{1, i\}$ and $a, b \geq 1$: Bins 1 and j are the last two nonempty bins, and at the first round in which condition (b) is triggered, bins 1 and j contain a and b balls, respectively.

Consequently, the following collection of events partitions the sample space.

$$\mathcal{F} = \{C_a, F_{1,i} \setminus C_a, F_\emptyset \setminus C_a, (F_{1,j}^{1,1} \setminus C_a)_{j \notin \{1,i\}}, (F_{1,j}^{r,1} \setminus C_a)_{j \notin \{1,i\}; r \geq 2}, (F_{1,j}^{1,r} \setminus C_a)_{j \notin \{1,i\}; r \geq 2}\}.$$

Note that $\mathcal{F}$ is defined relative to the stopping time $S(\vec{n})$, which assumes that the initial assignment is $\vec{n}$. Because these events cover the sample space, by the law of iterated expectations, we have

$$\begin{aligned}\mathbf{E}[X_{\vec{n}}] &= \mathbf{P}(C_a) \cdot \mathbf{E}[X_{\vec{n}} | C_a] + \mathbf{P}(F_{1,i} \setminus C_a) \cdot \mathbf{E}[X_{\vec{n}} | F_{1,i} \setminus C_a] \\ &\quad + \mathbf{P}(F_\emptyset \setminus C_a) \cdot \mathbf{E}[X_{\vec{n}} | F_\emptyset \setminus C_a] + \sum_{j \notin \{1,i\}} \mathbf{P}(F_{1,j}^{1,1} \setminus C_a) \cdot \mathbf{E}[X_{\vec{n}} | F_{1,j}^{1,1} \setminus C_a] \\ &\quad + \sum_{j \notin \{1,i\}} \sum_{r \geq 2} (\mathbf{P}(F_{1,j}^{r,1} \setminus C_a) \cdot \mathbf{E}[X_{\vec{n}} | F_{1,j}^{r,1} \setminus C_a] + \mathbf{P}(F_{1,j}^{1,r} \setminus C_a) \cdot \mathbf{E}[X_{\vec{n}} | F_{1,j}^{1,r} \setminus C_a]). \end{aligned} \tag{6}$$

An analogous equality holds for $\mathbf{E}[X_{\vec{n}-e_1+e_i}]$. Note that although the random variable $X_{\vec{n}-e_1+e_i}$ is defined relative to the initial assignment $\vec{n} - e_1 + e_i$, the partition $\mathcal{F}$ which is used in the analogous equality for $\mathbf{E}[X_{\vec{n}-e_1+e_i}]$ is defined relative to the initial assignment $\vec{n}$. This coupling constitutes the core of our proof strategy.

Next, we show that the first four summands in (6) are at least as large as the corresponding summands in the analogous expression for $\mathbf{E}[X_{\vec{n}-e_1+e_i}]$.

Lemma 4. *The following hold*:

i. $\mathbf{E}[X_{\vec{n}} | C_a] = \mathbf{E}[X_{\vec{n}-e_1+e_i} | C_a]$.

ii. $\mathbf{E}[X_{\vec{n}} | F_{1,i} \setminus C_a] > \mathbf{E}[X_{\vec{n}-e_1+e_i} | F_{1,i} \setminus C_a]$.

iii. $X_{\vec{n}} = X_{\vec{n}-e_1+e_i}$ *on* $F_\emptyset \setminus C_a$.

iv. $X_{\vec{n}} = X_{\vec{n}-e_1+e_i} = 1$ *on* $F_{1,j}^{1,1} \setminus C_a$, *for every* $j \notin \{1, i\}$.

Proof. On C_a, bin 1 contains one more ball than bin i at round $S(\vec{n})$. Hence, when the initial assignment is $\vec{n} - e_1 + e_i$, bin 1 contains one ball fewer, whereas bin i contains one additional ball at round $S(\vec{n})$. By symmetry, $X_{\vec{n}}$ conditioned on C_a and $X_{\vec{n}-e_1+e_i}$ conditioned on C_a are identically distributed, and item (i) follows.

On $F_{1,i} \setminus C_a$ the last two nonempty bins are 1 and i, and condition (b) triggers at round $S(\vec{n})$. Hence, bin 1 contains at least two more balls than bin i, which, in turn, contains one ball. By Remark 1, we have $f(m, 1) > f(m-1, 2)$ for every $m > 2$, and therefore, item (ii) follows.

Downloaded from informs.org by [186.11.81.201] on 13 October 2026, at 04:18 . For personal use only, all rights reserved.

On $F_\emptyset \setminus C_a$, bin 1 is empty, and hence, bin i contains a negative number of balls. Even if we transferred in the initial assignment a ball from bin 1 to bin i, it would not affect the identity of the last two bins. This establishes item (iii).

On $F^{1,1}_{1,j} \setminus C_a$, we have $X_{\vec{n}} = 1$. Immediately before round $S(\vec{n})$, for some $\ell \notin \{1,j\}$, bins $1, j, \ell$ must each have contained one ball. Indeed, if bins 1 and j had contained two and one balls, respectively, or one and two balls, respectively, then condition (b) would have occurred earlier. Moreover, the additional nonempty bin ℓ cannot be i, because condition (a) was not triggered in an earlier round. Hence, in this case $X_{\vec{n}-e_1+e_i} = 1$. As a result, item (iv) holds. □

Recall that $k_{\ell,t}$ denotes the number of times bin ℓ is selected in the first t rounds and that $f(a,b) = f(b,a)$ by symmetry. Observe that for $r \geq 2$:

$$\begin{aligned}
\mathbf{E}[X_{\vec{n}} | F^{r,1}_{1,j} \setminus C_a] &= \mathbf{E}[X_{\vec{n}} | F^{1,r}_{1,j} \setminus C_a] = f(r,1),\\
\mathbf{E}[X_{\vec{n}-e_1+e_i} | F^{r,1}_{1,j} \setminus C_a, k_{i,S(\vec{n})} > n_i] &= f(r-1,1),\\
\mathbf{E}[X_{\vec{n}-e_1+e_i} | F^{r,1}_{1,j} \setminus C_a, k_{i,S(\vec{n})} = n_i] &= f(r-1,1,1),\\
\mathbf{E}[X_{\vec{n}-e_1+e_i} | F^{1,r}_{1,j} \setminus C_a] &= f(r,0).
\end{aligned} \tag{7}$$

Regarding the last equality, on $F^{1,r}_{1,j} \setminus C_a$ bin 1 contains one ball at round $S(\vec{n})$, and because C_a does not hold, bin i contains a negative number of balls. Moreover, because condition (b) is triggered, there are two possibilities for what happened at round $S(\vec{n})$: either a ball was removed from bin 1 or a ball was removed from some bin $\ell \notin \{1,j,i\}$. In both cases, when the initial assignment is $\vec{n} - e_1 + e_i$, at the beginning of round $S(\vec{n})$, there are r balls in bin j and one ball in another bin (1 or ℓ), and at the end of that round, that other bin is empty.

Because $\mathbf{P}(F_{1,i} \setminus C_a) > 0$, Lemma 4 implies that, to prove $\mathbf{E}[X_{\vec{n}}] > \mathbf{E}[X_{\vec{n}-e_1+e_i}]$, it suffices to show that for every $j \notin \{1,i\}$ and every $r \geq 2$,

$$\begin{aligned}
&\mathbf{P}(F^{r,1}_{1,j} \setminus C_a) \cdot \mathbf{E}[X_{\vec{n}} | F^{r,1}_{1,j} \setminus C_a] + \mathbf{P}(F^{1,r}_{1,j} \setminus C_a) \cdot \mathbf{E}[X_{\vec{n}} | F^{1,r}_{1,j} \setminus C_a]\\
&\quad \geq \mathbf{P}(F^{r,1}_{1,j} \setminus C_a) \cdot \mathbf{E}[X_{\vec{n}-e_1+e_i} | F^{r,1}_{1,j} \setminus C_a] + \mathbf{P}(F^{1,r}_{1,j} \setminus C_a) \cdot \mathbf{E}[X_{\vec{n}-e_1+e_i} | F^{1,r}_{1,j} \setminus C_a].
\end{aligned} \tag{8}$$

Fix then $j \notin \{1,i\}$. By Lemma 2, for every $r \geq 2$, we have $f(r-1,1,1) \leq f(r-1,1)$, and hence, by (7), Equation (8) is implied by

$$(\mathbf{P}(F^{r,1}_{1,j} \setminus C_a) + \mathbf{P}(F^{1,r}_{1,j} \setminus C_a)) \cdot f(r,1) \geq \mathbf{P}(F^{r,1}_{1,j} \setminus C_a) \cdot f(r-1,1) + \mathbf{P}(F^{1,r}_{1,j} \setminus C_a) \cdot f(r,0).$$

By (1) and because $f(r-1,1) < r = f(r,0)$ for every $r \geq 2$, the inequality above holds, provided $\mathbf{P}(F^{r,1}_{1,j} \setminus C_a) \geq \mathbf{P}(F^{1,r}_{1,j} \setminus C_a)$. Thus, all that remains to establish Theorem 1 is to show this last inequality. We do so in what follows.

Note that $n_\ell - k_{\ell,t}$ is the number of balls remaining in bin ℓ at the end of round t (possibly becoming negative). Let S_b denote the first round at which condition (b) holds—that is, the first round at which exactly two bins are nonempty and one of them contains exactly one ball. With the above notation,

$$C_a^c = \{n_1 - k_{1,t} \geq n_i - k_{i,t} + 2 \text{ for every } 0 \leq t \leq S_b\}. \tag{9}$$

On C_a^c, we have $S(\vec{n}) = S_b$.

We now define the reflection map $\Psi : [k]^\infty \to [k]^\infty$. Informally, this map swaps the selections of bins 1 and j after the last round in which they contain the same number of balls, thereby sending a path in $F^{1,r}_{1,j}$ to a path in $F^{r,1}_{1,j}$ while preserving the event C_a^c. Formally, for $\omega \in [k]^\infty$ with $S_b(\omega) < \infty$, let

$$\tau(\omega) := \max\{t : 0 \leq t \leq S_b(\omega), n_1 - k_{1,t}(\omega) = n_j - k_{j,t}(\omega)\},$$

whenever this set is nonempty, and set $\Psi(\omega) := \omega$ otherwise. When τ is defined, let $\Psi(\omega)$ be the point obtained from ω by interchanging the symbols 1 and j in every coordinate ω_t with $t > \tau(\omega)$, all other coordinates (in particular, all selections of bin i) being left unchanged.

Lemma 5. *The map Ψ is a measure-preserving involution of $([k]^\infty, \mathbf{P})$.*

Proof. The map Ψ is clearly an involution and, in particular, a bijection. Moreover, Ψ is measure preserving. Indeed, for each $s \geq 0$, on the set $\{\tau = s\}$, the function Ψ interchanges the symbols 1 and j after time s and, hence, is measure preserving; and on the set where τ is undefined, Ψ is the identity. □

Downloaded from informs.org by [186.11.81.201] on 13 October 2026, at 04:18 . For personal use only, all rights reserved.

Lemma 6. *For all $j \notin \{1, i\}$ and all $r \geq 2$, we have $\mathbf{P}(F_{1,j}^{r,1} \setminus C_a) \geq \mathbf{P}(F_{1,j}^{1,r} \setminus C_a)$.*

Proof. Note that τ is defined on $F_{1,j}^{1,r}$. Indeed, $n_1 \geq n_j$, and on this event in round S_b, bin 1 contains less balls than bin j, so there must be a round in which the two bins contain the same number of balls. We next assert that $\Psi(F_{1,j}^{1,r} \setminus C_a) \subseteq C_a^c$. Indeed, let $\omega \in F_{1,j}^{1,r} \cap C_a^c$. Observe that:

- Because ω and $\Psi(\omega)$ coincide up to time τ and condition (a) is not triggered along ω up to that time (because $\omega \in C_a^c$), it is not triggered along $\Psi(\omega)$ up to time τ as well.
- Because $\omega \in F_{1,j}^{1,r}$, in each round between $\tau + 1$ and S_b, bin 1 contains more balls along $\Psi(\omega)$ than along ω. This is because τ is the last round in which the two bins contain the same number of balls, and along ω, at time S_b bin j contains more balls than bin 1. Hence, if condition (a) is not triggered along ω between those rounds, it is not triggered also along $\Psi(\omega)$.

These two items imply that $\Psi(\omega) \in C_a^c$, thus establishing the assertion.

Next, we assert that $\Psi(F_{1,j}^{1,r}) \subseteq F_{1,j}^{r,1}$. Indeed, fix $\omega \in F_{1,j}^{1,r}$. At time $\tau(\omega)$, bins 1 and j contain the same number of balls—say, z. Because $\omega \in F_{1,j}^{1,r}$, along ω, between rounds $\tau(\omega) + 1$ and $S_b(\omega)$, the number of rounds in which bins 1 and j have been selected is $z - 1$ and $z - r$, respectively. Because up to time $\tau(\omega)$ both ω and $\Psi(\omega)$ coincide, and after that time, the symbols 1 and j are swapped, $\Psi(\omega) \in F_{1,j}^{r,1}$.

The two established claims imply $\Psi(F_{1,j}^{1,r} \setminus C_a) \subseteq F_{1,j}^{r,1} \setminus C_a$. To conclude, observe that

$$\mathbf{P}(F_{1,j}^{1,r} \setminus C_a) = \mathbf{P}(\Psi(F_{1,j}^{1,r} \cap C_a^c)) \leq \mathbf{P}(F_{1,j}^{r,1} \setminus C_a),$$

where the first equality is because Ψ is measure preserving and injective (Lemma 5) and the inequality is by monotonicity of $\mathbf{P}$. □

3. Nonuniform Selection of Bins

In the case where the bins from which balls are removed are chosen nonuniformly, it is natural to ask whether the optimal initial assignment of balls to bins is proportional to the probability of picking each bin. The goal of this section is to show that this is not the case, even for the two-bin scenario.

Consider the case where $k = 2$ and $n = 6$. We claim that the initial assignment $(5, 1)$ is not optimal when, at each round, bins 1 and 2 are chosen with probabilities $5/6$ and $1/6$, respectively. Note that the assignment $(5, 1)$ is perfectly proportional to the selection probabilities.

Let $f_p(a, b)$ denote the expected number of balls remaining when the process stops, starting from the initial assignment (a, b), when bin 1 is selected with probability p and bin 2 with probability $1 - p$ at each round. To achieve this section's stated goal, it suffices to show that $f_{5/6}(4, 2) < f_{5/6}(5, 1)$. That is, the perfectly proportional initial assignment is not optimal: moving one ball from the heavier bin to the lighter bin reduces the expected number of balls in the last nonempty bin.

First, note that $f_p(a, b)$ satisfies the following recursion:

$$f_p(a,b) = \begin{cases} a, & \text{if } b = 0, \\ b, & \text{if } a = 0, \\ pf_p(a-1,b) + (1-p)f_p(a,b-1), & \text{otherwise.} \end{cases}$$

Using this relation, one can verify that the following expression holds for $b = 1$:

$$f_p(a,1) = a - \frac{p - p^a}{1-p} \quad \text{for } a > 0.$$

Similarly, one can verify that for $b = 2$:

$$f_p(a,2) = a + 2 - \frac{2}{1-p} + p^a\left(a + 2 + \frac{2p}{1-p}\right) \quad \text{for } a \in \mathbb{N}.$$

Simple calculations show that

$$f_{5/6}(5,1) = \frac{3125}{1296} \approx 2.41, \;\; f_{5/6}(4,2) = \frac{139}{81} \approx 1.72.$$

In particular, the proportional initial assignment $(5, 1)$ is not optimal.

Although the optimal allocation is not balanced, we conjecture that for every number of bins, the L_∞ distance between the optimal allocation and the balanced allocation is $O(\sqrt{n})$.

Acknowledgments

The authors thank Bernardo Subercaseaux for calling to their attention the problem addressed in this manuscript and for several insightful conversations. The authors also thank Will Ma and Elchanan Mossel for useful discussions and pointers to the literature. The authors acknowledge the use of Claude (Anthropic, July 2026) in identifying an error in an earlier proof of Lemma 6 and in developing the argument given here. Part of this work was done while the third author was a postdoc and the fourth author was a visitor at the Centro de Modelamiento Matemático at Universidad de Chile.

Endnotes

[1] Although $S(\vec{n})$ depends on i, we opt not to make the dependence explicit in the notation.

[2] Both conditions are triggered simultaneously if at round $S(\vec{n})$, bin i contains zero balls; some bin $k \neq 1, i$ contains more than one ball; bin 1 contains two balls at the beginning of the round and is selected at that round; and all other bins are empty.

References

Gutiérrez F, Subercaseaux B (2024) Assortment optimization for conference goodies with indifferent attendees. Preprint, submitted March 5, https://arxiv.org/abs/2403.03330.

Liang AJ, Jasin S, Uichanco J (2021) Assortment and inventory planning under dynamic substitution with MNL model: An LP approach and an asymptotically optimal policy. Technical report, University of Michigan, Ann Arbor, MI.

Mouchtaki O, El Housni O, Gallego G, Goyal V, Humair S, Kim S, Sadighian A, Wu J (2026) Joint assortment and inventory planning under the Markov chain choice model. *Management Sci.*, ePub ahead of print February 4, https://doi.org/10.1287/mnsc.2023.01322.

Zhang J, Ma W, Topaloglu H (2025) Technical note—Leveraging the degree of dynamic substitution in assortment and inventory planning. *Oper. Res.* 73(3):1248–1259.

Zhou Z, Wang T, Zhang J (2026) Optimal inventory allocation for indifferent goods under dynamic substitution. *Oper. Res.* 74(4):2001–2012.

Downloaded from informs.org by [186.11.81.201] on 13 October 2026, at 04:18 . For personal use only, all rights reserved.