

Coupon Collector Problem for Non-Uniform Coupons and Random Quotas

Nathan Shank and Hannah Yang

Moravian College
Department of Mathematics and Computer Science
Moravian College, Bethlehem, PA USA

shank@math.moravian.edu, sthhy01@moravian.edu

Submitted: Aug 24, 2011; Accepted: May 6, 2013; Published: May 16, 2013

Mathematics Subject Classifications: 05D40, 60C05

Abstract

We extend the classical coupon collector problem to find the expected number of selections needed to collect m_i (possible random) copies of coupon i , when the distribution of the coupons is not necessarily equally likely. Upper and lower bounds which provide limiting asymptotics are also obtained for the expected number of selections needed to fulfill a random quota for each coupon.

1 Introduction

The classic coupon collector problem describes a process in which n distinct coupons are placed in an urn. Coupons are randomly selected one at a time (with replacement) until at least one of each type of coupon has been selected. This creates one complete *set* of coupons. The coupon collector problem has a long history (see for example [3, 1]). The coupon collector problem has also been cited as the *single dixie cup problem* [8]. Detailed questions include the *birthday problem*, which considers the number of coupons that must be selected in order to have two copies of any one of the coupons, and the *double dixie cup problem*, which considers the number of coupons that must be selected in order to obtain two complete sets of coupons.

Several other variations of these problems have been explored. Examples include the coupon collector problem with random sample sizes [6], the coupon collector problem with unequally likely coupons [2, 4, 7, 9], and collecting a partial set of coupons [4]. There is also interest in the asymptotics for these problems as the number of possible coupons increases [2, 5, 7, 8, 9].

In the classic paper by Newman and Shepp [8], the authors derived an integral for the expected number of selections needed to form m complete sets of coupons assuming

each of the n coupons is equally likely to be selected. They also derive asymptotics for the expected number of selections as the number of coupons increase and the number of desired sets stays constant. Flajolet et al. [4] derived similar formulas and asymptotics assuming the coupons were not equally likely. Instead, coupon i was assumed to be selected with probability p_i so that $\sum p_i = 1$. In 1998, Papanicolaou et al. [9] expanded these results to allow the p_i to be random variables.

Another interesting variation occurs when we collect a specific amount of each coupon: a quota. Thus we may desire m_i copies of coupon i , where the m_i are not necessarily all equal. If the m_i are constant, $m_i = m$ for all i , then this reduces to collecting m complete sets of coupons. In [7], the author derived an integral formula similar to the one found by Newman and Shepp [8], along with asymptotics for this new *coupon collector problem with quotas*.

To further complicate the problem consider the following application: Assume a biologist would like to track the swim pattern for fish in a particular pond by placing tracking devices on fish. The biologist knows there are 10 different species of fish in the pond which are not all equally distributed. The researcher would like to track one male of each species, however the researcher would be satisfied by tracking two females instead (this restriction could be for environmental reasons or minimizing cost associated to catching the fish). What is the expected number of fish needed to be caught in order to have one male or at least two fish of each species?

We can answer this question by considering the coupon collector problem with *random* quotas. In this paper we consider the expected number of selections needed to acquire m_i copies of coupon i where each coupon type is not necessarily equally likely and the m_i could be random. The generalization to unequally likely coupons and random quotas allows us to find the expected number of fish the researcher needs to catch.

Under certain assumptions of m_i , we also find asymptotics for the expected number of coupon selections required as the number of distinct coupon types increases. We will use the techniques developed in [8] to expand upon the results of May [7] to allow random quotas and consider asymptotics for several variations.

2 Background and definitions

Throughout this manuscript, let $\vec{x}$ denote a possibly infinite vector and let x_i be the i^{th} component of $\vec{x}$. In addition, any bold variable or vector is assumed to be random, for example $\mathbf{X}$ denotes a random variable. For any $m \in \mathbb{N}^+ := \{1, 2, 3, \dots\}$ and any $t \in \mathbb{R}$, let

$$S_m(t) = \sum_{k=0}^{m-1} \frac{t^k}{k!},$$

which is the first m terms of the series expansion of e^t .

We will assume there are n distinct coupon types, however there can be more than one copy of each coupon type. We will refer to an individual coupon as a copy of a coupon type. We shall always assume there are n coupon types, x_i copies of the i^{th} coupon type,

we select a coupon uniformly from all of the X_n coupons where $X_n = \sum_{i=1}^n x_i$, and we will select coupons with replacement.

Under these assumptions, we are interested in finding the expected number of selections until a particular quantity of each of the n coupon type are selected. This expectation depends upon the quantity of each coupon type available, x_i , and the desired quantity, m_i , of each of the n coupon types.

More formally, assume an urn contains x_i copies of coupon i for each $i \in \{1, 2, \dots, n\}$. In addition, assume for each i we need to select m_i copies of coupon i . Let $T_n(\vec{m}, \vec{x})$ denote the expected number of selections until, for each i , m_i coupons of coupon type i have been selected.

In some classical coupon collector problems it is often the case that each coupon type is equally likely to be selected or that we desire to collect multiple complete sets of coupons. In these cases, the variables m_i and x_i may be constants. We will adopt the notation that if a and b are constants, then $T_n(a, \vec{x}) := T_n(\vec{a}, \vec{x})$ where $\vec{a}$ is a vector whose components are all a . Similarly define $T_n(\vec{m}, b) := T_n(\vec{m}, \vec{b})$.

The values of m_i and x_i could also be random. Thus let $\mathbb{E}(T_n(\vec{\mathbf{m}}, \vec{x}))$ denote the expected value of $T_n(\vec{m}, \vec{x})$ over all values of $\vec{m}$, $\mathbb{E}(T_n(\vec{m}, \vec{\mathbf{x}}))$ denote the expected value of $T_n(\vec{m}, \vec{x})$ over all values of $\vec{x}$ and $\mathbb{E}(T_n(\vec{\mathbf{m}}, \vec{\mathbf{x}}))$ denote the expected value of $T_n(\vec{m}, \vec{x})$ over all values of $\vec{m}$ and $\vec{x}$.

3 Main Results

First we will consider a general theorem for $\mathbb{E}(T_n(\vec{\mathbf{m}}, \vec{x}))$, followed by several corollaries which have been individually studied in previous literature.

Theorem 3.1. *For all $i \in \mathbb{N}^+$, let $x_i \in \mathbb{N}^+$ and let $\mathbf{m}_i$ be independent random variables whose support is $\mathbb{N}^+$. Consider an urn such that for each i , the urn contains x_i coupons of type i . Let $X_n = \sum_{i=1}^n x_i$ be the total number of coupons in the urn. Assume each of the X_n coupons are equally likely to be selected. If coupons are randomly selected from the urn with replacement, then the expected number of selections needed until for each $i \in \{1, 2, \dots, n\}$, $\mathbf{m}_i$ coupons of type i have been selected is*

$$\mathbb{E}(T_n(\vec{\mathbf{m}}, \vec{x})) = X_n \int_0^\infty \left(1 - \prod_{i=1}^n (1 - \mathbb{E}(S_{\mathbf{m}_i}(x_i t)) e^{-x_i t}) \right) dt.$$

Consider the application for the biologist trying to catch one male fish or at least two fish of 10 different species of fish. Assume there are 3 rare species which constitute only 5% of the fish population each, 3 dominate species which constitute 15% of the fish population each, and the remaining 4 species make up 10% of the population each. In addition, assume the biologist is equally likely to catch a female or a male fish for each of the 10 species. Evaluating the integral in Theorem 3.1 under these assumptions tells us the expected number of fish the biologist would have to catch is approximately 56.4.

The following corollary provides an explicit formula for the expected number of selections needed to collect one complete set of unevenly distributed coupons.

Corollary 3.2. *Under the assumptions of Theorem 3.1, if $\mathbf{m}_i = 1$ a.s. for all i , then*

$$T_n(1, \vec{x}) = X_n \int_0^\infty \left(1 - \prod_{i=1}^n (1 - e^{-x_i t}) \right) dt.$$

In the classic coupon collector problem, each coupon type is equally likely to be selected. If we consider equally likely coupons, however we collecting a random number of copies of each individual coupon, then we have the following corollary.

Corollary 3.3. *Under the assumptions of Theorem 3.1, if $x_i = 1$ for all i , then*

$$\mathbb{E}(T_n(\vec{\mathbf{m}}, 1)) = n \int_0^\infty \left(1 - \prod_{i=1}^n (1 - \mathbb{E}(S_{\mathbf{m}_i}(t)) e^{-t}) \right) dt.$$

Corollary 3.3 provides an explicit formula for the expected number of selections such that for each i , we collect $\mathbf{m}_i$ copies of coupon i assuming that each of the coupons is equally likely to be selected. This result was considered in [8] for non-random $\mathbf{m}_i$. Note that if x_i is equal to any positive constant for all i , the same result holds.

Recall the example of the biologist interesting in catching one male or at least two fish of 10 different species of fish. If each species is equally likely to be caught and the gender of the fish caught is also equally likely to be female or male then the integral in Corollary 3.3 evaluates to approximately 39.995. Thus the biologist can expect to catch approximately 40 fish.

Finally, we can combine the previous two corollaries and consider collecting $\mathbf{m}$ complete sets of coupons assuming that each coupon type is equally likely to be selected. Thus we only have one random variable $\mathbf{m}$ and not a sequence of random variables $\mathbf{m}_i$.

Corollary 3.4. *Under the assumptions of Theorem 3.1, if for all i , $x_i = 1$ and $\mathbf{m}_i = \mathbf{m}$, then*

$$\mathbb{E}(T_n(\mathbf{m}, 1)) = n \int_0^\infty (1 - (1 - \mathbb{E}(S_{\mathbf{m}}(t)) e^{-t})^n) dt.$$

Corollary 3.4 was developed in [8] by Newman and Shepp for a non-random $\mathbf{m}$.

For small values of n and x_i , and simple distributions of $\mathbf{m}_i$ (including constants), the previous explicit formulas can be easily computed. Asymptotics of these expectations have also been considered in previous literature. For example, Newman and Shepp [8] considered collecting m sets of n equally likely coupons. They showed that for a fixed m , and for n large, the expected number of selections is $n(\log n + (m-1)\log \log n + c_m + o(1))$, where c_m is a constant depending on m . In [9] the authors provide a technique for finding asymptotics for unequally likely coupons. In addition, [9] provides explicit examples of distributions which have different asymptotic behavior than those found in [8]. We expand upon these asymptotic results.

Let $\mathbf{m}_{n-} = \min_{i \leq n} \mathbf{m}_i$ and $x_n^- = \min_{i \leq n} x_i$. Let m^- denote the minimum of the support of $\mathbf{m}_i$ and x^- denote the minimum support of $\mathbf{x}_i$. Similarly define $\mathbf{m}_{n+}$, m^+ , x^+ , and x_n^+ for the maximum. In order to have a valid coupon collector problem, there must be at least one copy of each coupon type in the urn. We will assume we are collecting at least one copy of each type of coupon. Thus we can assume $\mathbf{m}_{n-}, m^-, x^-, x_n^- \geq 1$

Theorem 3.5. *Under the assumptions of Theorem 3.1,*

$$\frac{x_n^-}{x_n^+} \mathbb{E}(T_n(\mathbf{m}_{n-}, 1)) \leq \mathbb{E}(T_n(\vec{\mathbf{m}}, \vec{x})) \leq \frac{x_n^+}{x_n^-} \mathbb{E}(T_n(\mathbf{m}_{n+}, 1)).$$

Theorem 3.5 provides an upper and lower bound for $\mathbb{E}(T_n(\vec{\mathbf{m}}, \vec{x}))$. Notice that the bounds depend on $x_n^-, x_n^+, \mathbf{m}_{n-}$, and $\mathbf{m}_{n+}$.

In [8] the authors showed that $T_n(m, 1) = n[\log n + (m - 1) \log \log n + c_m + o(1)]$ for fixed m and $n \rightarrow \infty$. Under some general assumptions on the behavior of x_n^-, x_n^+, m^- , and m^+ we can use this result to obtain the following corollary for bounds on the asymptotics for $\mathbb{E}(T_n(\vec{\mathbf{m}}, \vec{x}))$.

Corollary 3.6. *Under the assumptions of Theorem 3.1, if $\lim_{n \rightarrow \infty} \frac{x_n^-}{x_n^+} = \alpha$ where $0 < \alpha < \infty$, $m^- > 0$, and $m^+ < \infty$, then*

$$\begin{aligned} \alpha n(\log n + (m^- - 1) \log \log n + c^- + o(1)) &\leq \mathbb{E}(T_n(\vec{\mathbf{m}}, \vec{x})) \\ &\leq \frac{1}{\alpha} n(\log n + (m^+ - 1) \log \log n + c^+ + o(1)), \end{aligned}$$

as $n \rightarrow \infty$, where c^- and c^+ are constants depending on m^- and m^+ respectively.

Corollary 3.6 follows easily from Theorem 3.5, Theorem 2 of [8], and the observation that if $0 < a \leq b$ then $T_n(a, 1) \leq T_n(b, 1)$, i.e. the expected amount of time to collect a complete sets of coupons is less than the expected amount of time to collect b complete sets of coupons.

We can also consider the number of copies of each coupon in the urn to be random, thus the $\mathbf{x}_i$ are random in addition to the $\mathbf{m}_i$. In doing so, it is more difficult to find a simple expression for the expected number of selections because of the dependency. However we are able to find an upper bound on the asymptotics based on some simple assumptions. Similar to [5], the authors also studied asymptotics of the random coupon collector problem with random $\mathbf{x}_i$. In sections 3 and 4 of [5] the author provided several examples of distributions of $\mathbf{x}_i$ which produced asymptotics which were not bounded by constant multiples of those in Corollary 3.6 and [8].

The following theorem provides bounds for random $\mathbf{x}_i$. In Corollary 3.8 we find restrictions on the distribution of the $\mathbf{x}_i$ which produce asymptotics which are bounded by constant multiples of the limits found in Corollary 3.6 and [8].

Theorem 3.7. *In addition to the assumptions of Theorem 3.1 assume $\mathbf{x}_i$ are independent random variables such that $1 \leq x^-$ and $x^+ < \infty$. Then*

$$\frac{\mu_{\mathbf{x}_n}}{nx^+} \mathbb{E}(T_n(\vec{\mathbf{m}}, 1)) \leq \mathbb{E}(T_n(\vec{\mathbf{m}}, \vec{\mathbf{x}})) \leq \frac{\mu_{\mathbf{x}_n}}{nx^-} \mathbb{E}(T_n(\vec{\mathbf{m}}, 1)),$$

where $\mu_{\mathbf{x}_n} = \mathbb{E}(\mathbf{X}_n)$.

Corollary 3.8. *Under the assumptions of Theorem 3.7, if $\lim_{n \rightarrow \infty} \frac{\mu_{\mathbf{x}_n}}{n} = \mu$, $0 < m^-$, and $m^+ < \infty$, then*

$$\begin{aligned} \frac{\mu}{x^+} n(\log n + (m^- - 1) \log \log n + c^- + o(1)) \\ \leq \mathbb{E}(T_n(\vec{\mathbf{m}}, \vec{\mathbf{x}})) \\ \leq \frac{\mu}{x^-} n(\log n + (m^+ - 1) \log \log n + c^+ + o(1)), \end{aligned}$$

as $n \rightarrow \infty$, where c^- and c^+ are constants depending on $\mathbf{m}^-$ and $\mathbf{m}^+$ respectively.

Again, Corollary 3.8 follows easily from Theorem 3.7, Theorem 2 of [8], and the observation that if $0 < a \leq b$ then $T_n(a, 1) \leq T_n(b, 1)$.

4 Proofs of main results

Proof of Theorem 3.1

Proof. Let $\mathbf{q}_i$ be the probability of not obtaining $\mathbf{m}_k$ copies of coupon k for all $1 \leq k \leq n$ in the first i selections. Thus $\mathbf{q}_i$ represents the probability of failure at selection i . Let $\mathbf{N}_i$ denote the number of ways of selecting i coupons from the urn and failing to obtain $\mathbf{m}_k$ copies of coupon k for some $1 \leq k \leq n$. Thus

$$\mathbf{q}_i = \frac{\mathbf{N}_i}{(\sum_{j=1}^n x_j)^i} = \frac{\mathbf{N}_i}{(X_n)^i}. \quad (1)$$

Let $\{(x_1 + x_2 + \cdots + x_n)^i / \vec{\mathbf{m}}\}$ denote the expansion of the polynomial $(y_1 + y_2 + \cdots + y_n)^i$, with all terms whose power for y_k is at least $\mathbf{m}_k$ for all $1 \leq k \leq n$ removed, evaluated at $(x_1, x_2, \dots, x_n)$.

Notice that

$$(X_n)^i = (x_1 + x_2 + \cdots + x_n)^i \text{ and } \mathbf{N}_i = \{(x_1 + x_2 + \cdots + x_n)^i / \vec{\mathbf{m}}\}. \quad (2)$$

For any positive integer valued random variable $X \geq 0$ we know $\mathbb{E}(X) = \sum_{i=0}^{\infty} \mathbb{P}(X > i)$.

Thus by conditioning on $\vec{\mathbf{m}}$ then substituting in (1) and (2) we have

$$\begin{aligned}\mathbb{E}(T_n(\vec{\mathbf{m}}, \vec{x})) &= \mathbb{E}\left(\sum_{i=0}^{\infty} \mathbf{q}_i\right) \\ &= \mathbb{E}\left(\sum_{i=0}^{\infty} \frac{\mathbf{N}_i}{(X_n)^i}\right) \\ &= \mathbb{E}\left(\sum_{i=0}^{\infty} \frac{\{(x_1 + x_2 + \cdots + x_n)^i / \vec{\mathbf{m}}\}}{(X_n)^i}\right).\end{aligned}$$

A simple integration by parts shows $n \int_0^{\infty} \frac{t^i}{i!} e^{-nt} dt = \frac{1}{n^i}$. Thus

$$\mathbb{E}(T_n(\vec{\mathbf{m}}, \vec{x})) = X_n \mathbb{E}\left(\sum_{i=0}^{\infty} \int_0^{\infty} \frac{\{(x_1 t + x_2 t + \cdots + x_n t)^i / \vec{\mathbf{m}}\}}{i!} e^{-X_n t} dt\right).$$

Tonelli's theorem justifies changing the order of integration, summation, and expectation. Hence

$$\mathbb{E}(T_n(\vec{\mathbf{m}}, \vec{x})) = X_n \int_0^{\infty} \mathbb{E}\left(\left(\sum_{i=0}^{\infty} \frac{\{(x_1 t + x_2 t + \cdots + x_n t)^i / \vec{\mathbf{m}}\}}{i!}\right) e^{-X_n t}\right) dt. \quad (3)$$

The series expansion for the exponential function shows

$$\begin{aligned}e^{X_n t} - \prod_{k=1}^n (e^{x_k t} - S_{\mathbf{m}_k}(x_k t)) &= e^{x_1 t + x_2 t + \cdots + x_n t} - \prod_{k=1}^n (e^{x_k t} - S_{\mathbf{m}_k}(x_k t)) \\ &= \sum_{i=0}^{\infty} \frac{(x_1 t + x_2 t + \cdots + x_n t)^i}{i!} \\ &\quad - \left(\sum_{i=m_1}^{\infty} \frac{(x_1 t)^i}{i!}\right) \left(\sum_{i=m_2}^{\infty} \frac{(x_2 t)^i}{i!}\right) \cdots \left(\sum_{i=m_n}^{\infty} \frac{(x_n t)^i}{i!}\right) \\ &= \sum_{i=0}^{\infty} \frac{\{(x_1 t + x_2 t + \cdots + x_n t)^i / \vec{\mathbf{m}}\}}{i!}.\end{aligned}$$

Substituting the previous identity into equation (3) reveals

$$\begin{aligned}\mathbb{E}(T_n(\vec{\mathbf{m}}, \vec{x})) &= X_n \int_0^{\infty} \mathbb{E}\left(\left(e^{X_n t} - \prod_{i=1}^n (e^{x_i t} - S_{\mathbf{m}_i}(x_i t))\right) e^{-X_n t}\right) dt \\ &= X_n \int_0^{\infty} \mathbb{E}\left(1 - \prod_{i=1}^n (1 - S_{\mathbf{m}_i}(x_i t) e^{-x_i t})\right) dt.\end{aligned}$$

Independence of $\mathbf{m}_i$ gives

$$\mathbb{E}(T_n(\vec{\mathbf{m}}, \vec{x})) = X_n \int_0^\infty \left(1 - \prod_{i=1}^n (1 - \mathbb{E}(S_{\mathbf{m}_i}(x_i t)) e^{-x_i t}) \right) dt,$$

which completes the proof. $\square$

In order to prove Theorem 3.5, first notice that for any $x, t, m > 0$, the function $f(x) = S_m(xt)e^{-xt}$ is positive. Differentiating with respect to x shows

$$\begin{aligned} f'(x) &= -te^{-xt}S_m(xt) + te^{-xt}S_{m-1}(xt) \\ &= -te^{-xt}(S_m(xt) - S_{m-1}(xt)) \\ &= -te^{-xt} \frac{(xt)^{m-1}}{(m-1)!}, \end{aligned}$$

which is clearly negative for $x, t, m > 0$. Thus $S_m(xt)e^{-xt}$ is non-increasing for $x, t, m > 0$.

Proof of Theorem 3.5

Proof. By Theorem 3.1 and by the definition of x_n^+ , we have

$$\begin{aligned} \mathbb{E}(T_n(\vec{\mathbf{m}}, \vec{x})) &= X_n \int_0^\infty \left(1 - \prod_{i=1}^n (1 - \mathbb{E}(S_{\mathbf{m}_i}(x_i t)) e^{-x_i t}) \right) dt \\ &\leq x_n^+ n \int_0^\infty \left(1 - \prod_{i=1}^n (1 - \mathbb{E}(S_{\mathbf{m}_i}(x_i t)) e^{-x_i t}) \right) dt. \end{aligned} \quad (4)$$

Since $S_m(xt)e^{-xt}$ is non-increasing for $x, t, m > 0$ and by the definition of x_n^- and $\mathbf{m}_{n+}$, (4) is bounded above by

$$\mathbb{E}(T_n(\vec{\mathbf{m}}, \vec{x})) \leq x_n^+ n \int_0^\infty \left(1 - \prod_{i=1}^n (1 - \mathbb{E}(S_{\mathbf{m}_{n+}}(x_n^- t)) e^{-x_n^- t}) \right) dt.$$

Let $u = x_n^- t$. Thus by the definition of $\mathbb{E}(T_n(\mathbf{m}_{n+}, 1))$ we have

$$\begin{aligned} \mathbb{E}(T_n(\vec{\mathbf{m}}, \vec{x})) &\leq \frac{x_n^+}{x_n^-} \left[n \int_0^\infty (1 - (1 - \mathbb{E}(S_{\mathbf{m}_{n+}}(u)) e^{-u})^n) du \right] \\ &= \frac{x_n^+}{x_n^-} \mathbb{E}(T_n(\mathbf{m}_{n+}, 1)) \end{aligned}$$

Similarly, we derive that $\frac{x_n^-}{x_n^+} \mathbb{E}(T_n(\mathbf{m}_{n-}, 1)) \leq \mathbb{E}(T_n(\vec{\mathbf{m}}, \vec{x}))$ which completes the proof. $\square$

Proof of Theorem 3.7

Proof. Conditioning on $\vec{x}$ we have

$$\mathbb{E}(T_n(\vec{m}, \vec{x})) = \mathbb{E}(\mathbb{E}(T_n(\vec{m}, \vec{x}) | \vec{x})).$$

Applying Theorem 3.1 we see

$$\mathbb{E}(T_n(\vec{m}, \vec{x})) = \mathbb{E}\left(\mathbf{X}_n \int_0^\infty \left(1 - \prod_{i=1}^n (1 - \mathbb{E}(S_{\mathbf{m}_i}(\mathbf{x}_i t)) e^{-\mathbf{x}_i t})\right) dt | \vec{x}\right).$$

Since $S_m(xt)e^{-xt}$ is non-increasing for $x, t, m > 0$ we obtain the following upper bound

$$\mathbb{E}(T_n(\vec{m}, \vec{x})) \leq \mathbb{E}\left(\mathbf{X}_n \int_0^\infty \left(1 - \prod_{i=1}^n (1 - \mathbb{E}(S_{\mathbf{m}_i}(x^- t)) e^{-x^- t})\right) dt | \vec{x}\right).$$

By Theorem 3.1 the previous becomes

$$\mathbb{E}(T_n(\vec{m}, \vec{x})) \leq \frac{1}{nx^-} \mathbb{E}(\mathbf{X}_n \mathbb{E}(T_n(\vec{m}, x^-)) | \vec{x}).$$

Since $\mathbb{E}(T_n(\vec{m}, x^-))$ does not depend on $\vec{x}$, we have

$$\mathbb{E}(T_n(\vec{m}, \vec{x})) \leq \frac{\mu_{\mathbf{X}_n}}{nx^-} \mathbb{E}(T_n(\vec{m}, x^-))$$

where $\mu_{\mathbf{X}_n} = \mathbb{E}(\mathbf{X}_n)$.

Similarly, we can derive that

$$\frac{\mu_{\mathbf{X}_n}}{nx^+} \mathbb{E}(T_n(\vec{m}, 1)) \leq \mathbb{E}(T_n(\vec{m}, \vec{x}))$$

which completes the proof. $\square$

References

- [1] G. Bloom, D. Sandell, and S. Holst. *Problems and Snapshots from the World of Probability*. Springer, New York, 1994.
- [2] M. Brown, E. A. Pekoz, and S.M. Ross. Coupon collecting. *Probability in the Engineering and Informational Sciences*, 22:221–229, 2008.
- [3] W. Feller. *An Introduction to Probability Theory and Its Applications*, volume I. John Wiley, New York, 3rd edition edition, 1968.
- [4] P. Flajolet, D. Gardy, and L. Thimonier. Birthday paradox, coupon collectors, caching algorithms and self-organizing search. *Discrete Applied Mathematics*, 39:207–229, 1992.

- [5] Lars Holst. Extreme value distributions for random coupon collector and birthday problems. *Extremes*, 4:129–145, 2001.
- [6] J. E. Kobza, S. H. Jacobson, and D. E. Vaughan. A survey of the coupon collector’s problem with random sample sizes. *Methodology And Computing In Applied Probability*, 9:573–584, 2007.
- [7] R. May. Coupon collecting with quotas. *The Electronic Journal of Combinatorics*, 15:N31, 2008.
- [8] D.J. Newman and L. S. Shepp. The double dixie cup problem. *The American Mathematical Monthly.*, 67(1):58–61, 1960.
- [9] V. G. Papanicolaou, G.E. Kokolakis, and S. Boneh. Asymptotics for the random coupon collector problem. *Journal of Computational and Applied Mathematics*, 93:95–105, 1998.