*Applied Probability Trust* (28 January 2021)

# TWO POISSON LIMIT THEOREMS FOR THE COUPON COLLECTOR'S PROBLEM WITH GROUP DRAWINGS

JUDITH SCHILLING,* *Technische Universität Darmstadt*

NORBERT HENZE,** *Karlsruhe Institute of Technology (KIT)*

**Abstract**

In the collector's problem with group drawings, $s$ out of $n$ different types of coupon are sampled with replacement. In the uniform case, each $s$-subset of the types has the same probability of being sampled. For this case, we derive a Poisson limit theorem for the number of types that are sampled at most $c-1$ times, where $c \geq 1$ is fixed. In a specified approximate non-uniform setting, we prove a Poisson limit theorem for the special case $c = 1$. As corollaries, we obtain limit distributions for the waiting time for $c$ complete series of types in the uniform case and a single complete series in the approximate non-uniform case.

*Keywords:* Coupon collector's problem; group drawings; Poisson limit theorem

2010 Mathematics Subject Classification: Primary 60F05

Secondary 60G70

## 1. Introduction

The coupon collector's problem (CCP) has a long history, dating back to de Moivre, Laplace, and Euler, see, e.g. Sections 291, 448–461, 632–638, 775–781, 864, 910 and 971 of [37] for a detailed exposition of the early history of the CCP. The CCP has many applications (see, e.g., [8]), and it is of ongoing interest, as witnessed by the recent papers [1]–[4], [6]–[7], [9]–[13], [15]–[18], [20]–[27], [29]–[32], [34]–[35], and [38].

In the classical CCP, there are $n$ distinct types of coupon, and coupons are collected

* Postal address: Fachbereich Mathematik, Technische Universität Darmstadt, Schlossgartenstr. 7, D-64289 Darmstadt, Germany.

** Postal address: Institute of Stochastics, Karlsruhe Institute of Technology (KIT), Englerstr. 2, D-76133 Karlsruhe, Germany.

one at a time, independently of each other, with each type being equally probable. In this setting, let $W_{n,c}$ denote the number of coupons that must be sampled until each type has been sampled at least $c$ times, where $c \geq 1$ is a fixed number.

In what follows, we will adopt a conceptually equivalent interpretation, inasmuch as the $n$ types of coupons represent $n$ different cells, which are numbered from 1 to $n$, and the sampling of a type of coupon corresponds to the placement of a particle into a cell. In this setting, Erdös and Rényi [14] showed that, for each $x \in \mathbb{R}$,

$$\lim_{n\to\infty} \mathbb{P}\left(W_{n,c} \leq n\big\{\log n + (c-1)\log\log n - \log(c-1)! + x\big\}\right) = \exp\left(-\mathrm{e}^{-x}\right). \qquad (1)$$

In other words, $W_{n,c}/n - \log n - (c-1)\log\log n + \log(c-1)!$ has a Gumbel limit distribution as $n \to \infty$.

This paper deals with the CCP with group drawings, which means that at each stage of the 'placement process', $s$ particles are placed into $s$ different cells, which is henceforth called an *$s$-placement*. A standard applications is the collection of pictures of soccer players, where one can buy packages with $s = 5$ or $s = 6$ different pictures out of a total of several hundreds of pictures. The classical assumption in this more general setting is that each of the $\binom{n}{s}$ $s$-placements, which are subsets of size $s$ out of the $n$ cells, has the same probability. In what follows, the random variable of interest is $W_{n,s,c}$, which is the number of $s$-placements necessary to have at least $c$ particles in each cell. In other words, $W_{n,s,c}$ is the waiting time to see each type of coupon at least $c$ times. Under this assumption, [36], among other things, derived the probability distribution of $W_{n,s,1}$.

Suppressing the dependence on $c$,$s$ and $k$, define, for each $k \geq 1$ and $j \in \{1, \ldots, n\}$, $A_{n,j}$ to be the event that, after $k$ $s$-placements, cell $j$ contains at most $c - 1$ particles, and let

$$Z_{n,s,c}(k) := \sum_{j=1}^{n} \mathbf{1}\{A_{n,j}\} \qquad (2)$$

denote the number of cells that contain at most $c - 1$ particles after $k$ $s$-placements, where $\mathbf{1}\{\cdot\}$ denotes the indicator function.

For fixed $x \in \mathbb{R}$, let

$$k_{n,s,c}(x) = \left\lfloor \frac{n}{s}\left(\log n + (c-1)\log\log n - \log(c-1)! + x\right)\right\rfloor, \quad n > 1, \qquad (3)$$

where $\lfloor\cdot\rfloor$ denotes the floor function. In the special case $c=1$, Theorem 2 of [28] yields

$$Z_{n,s,1}(k_{n,s,1}(x)) \xrightarrow{\mathcal{D}} \mathrm{Po}(\mathrm{e}^{-x}), \quad x \in \mathbb{R}, \tag{4}$$

where denotes $\xrightarrow{\mathcal{D}}$ convergence in distribution, and $\mathrm{Po}(\lambda)$, $\lambda > 0$, is the Poisson distribution with parameter $\lambda$. Notice that this result gives (1) in the special case $c=1$, since $Z_{n,s,1}(k_{n,s,1}(x)) = 0$ if, and only if, $W_{n,1} \le n(x + \log n)$.

In this paper, we first generalize (4) to the case $c > 1$. This will be done in Section 2. In Section 3, we consider the special case $c = 1$, but relax the condition of a uniform distribution over all $s$-placements and consider, as $n \to \infty$, a sequence of distributions over all $s$-placements that approach the uniform distribution at the rate $O(1/\log n)$. For the sake of readability, parts of the proofs are deferred to Section 4.

## 2. A Poisson limit for the number of cells having at most $c-1$ particles

In this section, we generalize (4) to the case $c > 1$ and (1) to the case $s > 1$. Our main result is as follows.

**Theorem 2.1.** *Under the condition that each $s$-placement is equally probable, we have*

$$Z_{n,s,c}(k_{n,s,c}(x)) \xrightarrow{\mathcal{D}} Z \quad \text{as} \quad n \to \infty, \quad x \in \mathbb{R},$$

*where $k_{n,s,c}(x)$ is defined in (3), and $Z$ has the Poisson distribution $\mathrm{Po}(\mathrm{e}^{-x})$.*

Since $W_{n,s,c} \le k_{n,s,c}(x)$ if and only if $Z_{n,s,c}(k_{n,s,c}(x)) = 0$, Theorem 2.1 yields the following corollary.

**Corollary 2.1.** *For each $x \in \mathbb{R}$, we have*

$$\lim_{n\to\infty} \mathbb{P}\left(W_{n,s,c} \le \frac{n}{s}\Big\{\log n + (c-1)\log\log n - \log(c-1)! + x\Big\}\right) = \exp\left(-\mathrm{e}^{-x}\right).$$

Notice that Corollary 2.1 indeed generalizes both (1) and (4). We mention that (1) may also be obtained from [20], see Remark 4.3 of [20].

Proof of Theorem 2.1. Put $k_n = k_{n,s,c}(x)$, and assume $n$ to be sufficiently large in order to have $k_n \ge 1$. For $\ell \in \{0,\dots,c-1\}$, let $A_{n,j}^{(\ell)}$ be the event that cell $j$ contains $\ell$ particles after $k_n$ $s$-placements, and define

$$Z_{n,s,c}^{(\ell)}(k_n) := \sum_{\ell=0}^{c-1} \mathbf{1}\{A_{n,j}^{(\ell)}\}.$$

Then $Z_{n,s,c}(k)$ figuring in (2), with $k$ replaced with $k_n$, takes the form

$$Z_{n,s,c}(k_n) = \sum_{\ell=0}^{c-1} Z_{n,s,c}^{(\ell)}(k_n).$$

By the definition of $k_n$, straightforward calculations yield

$$\mathbb{E}Z_{n,s,c}^{(\ell)}(k_n) = n\mathbb{P}(A_{n,1}^{(\ell)}) = n\binom{k_n}{\ell}\left(\frac{s}{n}\right)^{\ell}\left(1-\frac{s}{n}\right)^{k_n-\ell} = O\left((\log n)^{\ell-(c-1)}\right).$$

Invoking Markov's inequality, we have $Z_{n,s,c}^{(\ell)}(k) \xrightarrow{\mathbb{P}} 0$ as $n \to \infty$ if $\ell \le c-2$. Thus, putting $Z_n := Z_{n,s,c}^{(c-1)}(k_{n,s,c}(x))$, it remains to prove $Z_n \xrightarrow{\mathcal{D}} Z$ as $n \to \infty$.

The proof uses the method of moments. The distribution of $Z$ is uniquely determined by the sequence $(\mathbb{E}(Z^r))_{r\ge1}$, and we have $\mathbb{E}(Z^r) = \sum_{\ell=0}^{r} \left\{ {r \atop i} \right\} \mathbb{E}(Z^{\underline{i}})$, where, in a general notation, $t^{\underline{i}} = t(t-1)\cdots(t-i+1)$ denotes the $i$th descending factorial of $t \in \mathbb{R}$, and $\left\{ {r \atop i} \right\}$, $i = 0,\dots,r$, are the Stirling numbers of the second kind (see [19], p. 262). Since $\mathbb{E}(Z^{\underline{r}}) = \lambda^r$, where $\lambda = \mathrm{e}^{-x}$, and since the events $A_{n,j}^{(c-1)}$, $1 \le j \le n$, are exchangeable, we have $\mathbb{E}(Z_n^{\underline{r}}) = n^{\underline{r}}\,\mathbb{P}(A_{n,1}^{(c-1)} \cap \ldots \cap A_{n,r}^{(c-1)})$. In view of $n^{\underline{r}}/n^r \to 1$ as $n \to \infty$, it thus remains to prove

$$\lim_{n\to\infty} n^r\,\mathbb{P}(A_{n,1}^{(c-1)} \cap \ldots \cap A_{n,r}^{(c-1)}) = \lambda^r \tag{5}$$

for each $r \ge 1$. To this end, let $N_{n,j}$ be the number of particles in cell $j$ after $k_n$ $s$-placements, whence

$$\mathbb{P}(A_{n,1}^{(c-1)} \cap \ldots \cap A_{n,r}^{(c-1)}) = \mathbb{P}(N_{n,1} = c-1, \dots, N_{n,r} = c-1). \tag{6}$$

Notice that, if $r \ge 2$, more than one of the cells $1,\dots,r$ may receive a particle in each $s$-placement. We will see, however, that this event will be asymptotically negligible. For this purpose, let $B_i$ denote the number of all $k_n$ $s$-placements in which exactly $i$ of the cells $1,\dots,r$ receive a particle, where $i \in \{1,\dots,r\}$, and put $B := (B_1,\dots,B_r)$. Setting $N_j := N_{n,j}$ and $b := (b_1,\dots,b_r)$, we then have

$$\begin{aligned} &\mathbb{P}(N_1 = c-1, \dots, N_r = c-1) \\ =\ &\sum_{b_1\ge0}\cdots\sum_{b_r\ge0} \mathbb{P}(N_1 = c-1, \dots, N_r = c-1, B = b)\mathbf{1}\left\{\sum_{j=1}^{r} jb_j = r(c-1)\right\}. \end{aligned} \tag{7}$$

Notice that, under the condition $s < r$, $\mathbb{P}(N_1 = c-1, \dots, N_r = c-1, B = b)$ can only be positive if $b_{s+1} = \ldots = b_r = 0$. The above-mentioned asymptotic negligibility is a consequence of the following result, the proof of which is given in Section 4.

**Lemma 1.** *We have* $\mathbb{P}(N_1 = c-1, \ldots, N_r = c-1, B = b) = O\left(n^{-r-\sum_{i=2}^{r}(i-1)b_i}\right)$.

We now put $\mathcal{S}_{n,r,s} := \mathbb{P}\left(N_1 = c-1, \ldots, N_r = c-1, B_1 = r(c-1), B_2 = \ldots = B_r = 0\right)$. In view of (5), (6), (7) and Lemma 1, we obviously have to show

$$\lim_{n\to\infty} n^r \mathcal{S}_{n,r,s} = \lambda^r. \tag{8}$$

Similar to the reasoning given in the proof of Lemma 1, it follows that

$$\mathcal{S}_{n,r,s} = \frac{k_n!}{(c-1)!^r (k_n - r(c-1))!} \left[\frac{\binom{n-r}{s-1}}{\binom{n}{s}}\right]^{r(c-1)} \left[\frac{\binom{n-r}{s}}{\binom{n}{s}}\right]^{k_n - r(c-1)}.$$

Now, $k_n!/(k_n - r(c-1))! = k_n^{r(c-1)}(1+o(1))$ and $\binom{n-r}{s-1}/\binom{n}{s} = (s/n)(1+o(1))$. Moreover,

$$\frac{\binom{n-r}{s}}{\binom{n}{s}} = \prod_{i=0}^{r-1}\left(1 - \frac{s}{n-i}\right),$$

and we obtain

$$\mathcal{S}_{n,r,s} = \left(\frac{k_n s}{n}\right)^{r(c-1)} \left[\prod_{i=0}^{r-1}\left(1 - \frac{s}{n-i}\right)\right]^{k_n - r(c-1)} \frac{(1+o(1))}{(c-1)!^r}.$$

We now take the logarithm of $\mathcal{S}_{n,r,s}$. By the definition of $k_n$, straightforward calculations yield

$$r(c-1)\log\left(\frac{k_n s}{n}\right) = r(c-1)\log\log n + o(1).$$

Since $\log t \le t-1$, $t > 0$, we further have

$$\begin{aligned}
(k_n - r(c-1))\textstyle\sum_{i=0}^{r-1} \log\left(1 - \frac{s}{n-i}\right) &\le -(k_n - r(c-1))\textstyle\sum_{i=0}^{r-1} \frac{s}{n-i} \\
&\le -(k_n - r(c-1)) \cdot \frac{rs}{n} \\
&= -rx - r\log n - r(c-1)\log\log n + r\log(c-1)! + o(1).
\end{aligned} \tag{9}$$

Upon exponentiating, it thus follows that

$$n^r \mathcal{S}_{n,r,s} \le \mathrm{e}^{-rx+o(1)}(c-1)!^r \exp\left[(\log\log n)\left(r(c-1) - r(c-1)\right)\right] \frac{1}{(c-1)!^r}.$$

We obtain

$$\limsup_{n\to\infty} n^r \mathcal{S}_{n,r,s} \le \mathrm{e}^{-rx} = \lambda^r. \tag{10}$$

In the same way, using the inequality $\log t \ge 1 - 1/t$, $t > 0$, to bound the left hand side of (9) from below, it follows that $\liminf_{n\to\infty} n^r \mathcal{S}_{n,r,s} \ge \mathrm{e}^{-rx} = \lambda^r$. The details are omitted. Together with (10), (8) follows, and the proof is completed.

## 3. A Poisson limit in the case $c = 1$ and non-uniform distributions

In this section, we consider the waiting time until each cell has received at least one particle, but allow for small deviations from the uniform distribution over all subsets of size $s$ of the $n$ cells. For the special case $s = 1$, this case has been treated by [29]. To have a somewhat more compact notation, we put $[n] := \{1, \ldots, n\}$ and let $[n]_s$ denote the system of all subsets of size $s$ of $[n]$. For $\mathcal{I} \in [n]_s$, we write $p_{n,\mathcal{I}}$ for the probability that each of the cells $i \in \mathcal{I}$ receives a particle in an $s$-placement. Our model is

$$p_{n,\mathcal{I}} = \frac{1}{\binom{n}{s}} \left(1 + \frac{\Delta_{n,\mathcal{I}}}{\log n}\right), \tag{11}$$

where $\sum_{\mathcal{I} \in [n]_s} \Delta_{n,\mathcal{I}} = 0$ and

$$\max_{\mathcal{I} \in [n]_s} |\Delta_{n,\mathcal{I}}| \le M \tag{12}$$

for some universal constant $M \in (0, \infty)$. Thus, the deviation from a uniform distribution over $[n]_s$ is of order $1/\log n$. To state the main theorem of this section let, for $h \in \{1, \ldots, s\}$ and $\{i_1, \ldots, i_h\} \subset [n]$

$$\overline{\Delta}_{n,i_1,\ldots,i_h} := \frac{1}{\binom{n-h}{s-h}} \sum_{\mathcal{I} \subset [n]_s : \mathcal{I} \supset \{i_1,\ldots,i_h\}} \Delta_{n,\mathcal{I}} \tag{13}$$

denote the average deviation from the uniform distribution over all $s$-placements $\mathcal{I}$ that occupy the cells $i_1, \ldots, i_h$. Notice that $\Delta_{n,\mathcal{I}} = \overline{\Delta}_{n,i_1,\ldots,i_h}$ if $s = h$. With $Z_{n,s,c}$ defined in (2) and $k_{n,s,c}(x)$ given in (3), we have the following result.

**Theorem 3.1.** *Under the assumptions made above and the condition*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \exp\left(-\frac{x + \log n}{\log n} \cdot \overline{\Delta}_{n,i}\right) =: c(x) < \infty, \tag{14}$$

*we have* $Z_{n,s,1}(k_{n,s,1}(x)) \xrightarrow{\mathcal{D}} Z$ *as* $n \to \infty$, $x \in \mathbb{R}$, *where* $Z$ *has the Poisson distribution* $\mathrm{Po}(\lambda)$, $\lambda = \mathrm{e}^{-x} c(x)$.

**Corollary 1.** *For the waiting time* $W_{n,s,1}$ *until each cell contains at least one particle, we have under the model* (11)

$$\lim_{n\to\infty} \mathbb{P}\left(W_{n,s,1} \le \frac{n}{s}\Big\{x + \log n\Big\}\right) = G\left(x - \log c(x)\right), \tag{15}$$

*where* $G(x) = \exp(-\mathrm{e}^{-x})$ *is the distribution function of the Gumbel distribution.*

Before we give a proof of Theorem 3.1, we present two examples for the model (11).

**Example 1.** Suppose that the set $[n]$ of cells is divided into $\ell$ pairwise disjoint blocks $B_{n,1},\ldots,B_{n,\ell}$, such that $[n]=B_{n,1}\cup\ldots\cup B_{n,\ell}$ and $\rho_{n,j}:=|B_{n,j}|\geq s$, $j=1,\ldots,\ell$, where $|B|$ denotes the number of elements of a finite set $B$. Without loss of generality, we let $B_{n,1}=\{1,\ldots,\rho_{n,1}\}$, $B_{n,2}=\{\rho_{n,1}+1,\ldots,\rho_{n,1}+\rho_{n,2}\}$ etc.

We assume that the proportions of the blocks stabilize as $n\to\infty$, i.e., we have

$$\lim_{n\to\infty}\frac{\rho_{n,j}}{n}=p_j,\quad j=1,\ldots,\ell \tag{16}$$

(say), where $p_j>0$ for each $j$ and $p_1+\ldots+p_\ell=1$. Furthermore, let $w_n:=(w_{n,1},\ldots,w_{n,\ell})$ be a sequence of 'weight vectors' in $\mathbb{R}^\ell$ such that $w_n\to w:=(w_1,\ldots,w_\ell)$ as $n\to\infty$. We assume that

$$\sum_{j=1}^{\ell}\rho_{n,j}w_{n,j}=0 \tag{17}$$

for each $n$, which implies $\sum_{j=1}^{\ell}p_jw_j=0$. In other words, the vectors $p=(p_1,\ldots,p_\ell)$ and $w$ are orthogonal. For $\mathcal{I}\in[n]_s$, let

$$\Delta_{n,\mathcal{I}}:=\sum_{j=1}^{\ell}w_{n,j}|\mathcal{I}\cap B_{n,j}|.$$

Notice that, in view of

$$\sum_{\mathcal{I}\in[n]_s}\Delta_{n,\mathcal{I}}=\sum_{j=1}^{\ell}w_{n,j}\sum_{\mathcal{I}\in[n]_s}|\mathcal{I}\cap B_{n,j}|=\sum_{j=1}^{\ell}w_{n,j}\binom{n-1}{s-1}\rho_{n,j}$$

and (17), we have $\sum_{\mathcal{I}\in[n]_s}\Delta_{n,\mathcal{I}}=0$. To determine $c(x)$ figuring in (14), notice that, by symmetry, $\overline{\Delta}_{n,i}$ is constant over $i$ within each of the blocks, and hence the left hand side of (14) (without the limit) takes the form

$$\sum_{j=1}^{\ell}\frac{\rho_{n,j}}{n}\exp\left(-\frac{x+\log n}{\log n}\cdot\overline{\Delta}_{n,j_1}\right),$$

where $j_1$ denotes the first element of block $B_{n,j}$. Without loss of generality, we choose the first block and notice that

$$\overline{\Delta}_{n,1}=\frac{1}{\binom{n-1}{s-1}}\sum_{k=1}^{\ell}w_{n,k}\sum_{\mathcal{I}\in[n]_s:\mathcal{I}\ni 1}|\mathcal{I}\cap B_{n,k}|.$$

We now consider the cases $k = 1$ and $k > 1$ separately. If we distinguish cases according to $m := |\mathcal{I} \cap B_{n,1}|$, it follows that

$$\sum_{\mathcal{I} \in [n]_s : \mathcal{I} \ni 1} |\mathcal{I} \cap B_{n,1}| = \sum_{m=1}^{s} m \binom{\rho_{n,1} - 1}{m-1} \binom{n - \rho_{n,1}}{s-m} = \binom{n-1}{s-1} \left(1 + (s-1)\frac{\rho_{n,1}}{n}\right).$$

Likewise, if $k > 1$ we have

$$\sum_{\mathcal{I} \in [n]_s : \mathcal{I} \ni 1} |\mathcal{I} \cap B_{n,k}| = \sum_{i=0}^{s-1} i \binom{\rho_{n,k}}{i} \binom{n - \rho_{n,k} - 1}{s-i-1} = \binom{n-1}{s-1} \frac{(s-1)\rho_{n,k}}{n-1}.$$

Summarizing, it follows that

$$\overline{\Delta}_{n,1} = w_{n,1} \left(1 + \frac{(s-1)\rho_{n,1}}{n}\right) + \sum_{k=2}^{\ell} w_{n,k} \frac{(s-1)\rho_{n,k}}{n}$$

and thus, in view of (16), $w_{n,j} \to w_j$ and $\sum_{k=1}^{\ell} p_k w_k = 0$, we deduce

$$\lim_{n\to\infty} \overline{\Delta}_{n,1} = w_1.$$

Since $(x + \log n)/\log n \to 1$ as $n \to \infty$, we obtain

$$c(x) = \sum_{j=1}^{\ell} p_j \mathrm{e}^{-w_j},$$

independently of $x$. Notice that $\sum_{j=1}^{\ell} p_j \mathrm{e}^{-w_j} \geq 1$ and thus $\log c(x) \geq 0$. This means that, at least under the sequence of alternative distributions to uniformity studied in this example, the limit distribution of $W_{n,s,1}$ is stochastically greater than under the assumption of a uniform distribution of all subsets of size $s$ of $[n]$.

**Example 2.** In the special case $s = 1$, a natural class of examples for the deviation (11) from the uniform distribution is given by

$$p_{n,j} := p_{n,\{j\}} = \frac{1}{n} \left(1 + \frac{1}{\log n} \cdot n \int_{(j-1)/n}^{j/n} m(t)\,\mathrm{d}t\right),$$

where $m : [0,1] \to \mathbb{R}$ is a continuous function satisfying $\int_0^1 m(t)\,\mathrm{d}t = 0$. Here, by the mean value theorem, we have $\Delta_{n,j} = \Delta_{n,\{j\}} = m(t_{n,j})$, where $(j-1)/n \leq t_{n,j} \leq j/n$. Since $\overline{\Delta}_{n,\{j\}} = \Delta_{n,j}$, the function $c(x)$ figuring in (14) is given by

$$c(x) = J_m := \int_0^1 \mathrm{e}^{-m(t)}\,\mathrm{d}t,$$

independent of $x$. Hence, we have (15), where $\log c(x) = \log J_m \geq 0$ by Jensen's inequality.

**Remark 1.** In the special case $s = 1$, Theorem 3.1 may be derived from Theorem 2.1 of [29]. Putting $p_{n,i} = p_{n,\{i\}}$, Theorem 2.1 of [29] requires the existence of sequences $(a_n)$ and $(b_n)$ such that $a_n/b_n \to 0$ as $n \to \infty$ and

$$\sum_{i=1}^{n} \exp\big(-p_{n,i}(b_n + xa_n)\big) \to g(x) \text{ as } n \to \infty \tag{18}$$

($x \in \mathbb{R}$), where $g(\cdot)$ is a nonincreasing function with $g(x) \to \infty$ as $x \to -\infty$ and $g(x) \to 0$ as $x \to \infty$. Then $(W_{n,1,1} - b_n)/a_n \xrightarrow{\mathcal{D}} Y$ as $n \to \infty$, where $Y$ has distribution function $\mathbb{P}(Y \leq x) = \mathrm{e}^{-g(x)}$, $x \in \mathbb{R}$.

For the special choice $a_n = n$, $b_n = n \log n$ and $p_{n,i} = (1 + \Delta_{n,i}/\log n)/n$, we have

$$\exp\big(-p_{n,i}(b_n + xa_n)\big) = \frac{1}{n}\mathrm{e}^{-x} \exp\left(-\frac{x + \log n}{\log n}\Delta_{n,i}\right).$$

Hence (14) implies (18), where $g(x) = c(x)\mathrm{e}^{-x}$.

Proof of Theorem 3.1. For the sake of brevity, put $Z_n = Z_{n,s,1}(k_{n,s,1}(x))$ and $k_n = k_{n,s,1}(x)$. Just as in the proof of Theorem 2.1, we employ the method of moments. The reasoning, however, is more delicate since the events $A_{n,1}, \ldots, A_{n,n}$ figuring in (2) are no longer exchangeable. Notice that, in this section, $A_{n,j}$ is the event that cell $j$ has not been occupied after $k_n$ $s$-placements. The $r$th factorial moment of $Z_n$ is given by

$$\mathbb{E}(Z_n^{\underline{r}}) = r! \sum_{1 \leq i_1 < \ldots < i_r \leq n} \mathbb{P}(A_{n,i_1} \cap \ldots \cap A_{n,i_r}),$$

and hence we have to prove

$$\lim_{n\to\infty} \sum_{1 \leq i_1 < \ldots < i_r \leq n} \mathbb{P}(A_{n,i_1} \cap \ldots \cap A_{n,i_r}) = \frac{\mathrm{e}^{-rx} c(x)^r}{r!}, \quad r \geq 1. \tag{19}$$

To obtain suitable bounds for $\mathbb{P}(A_{n,i_1} \cap \ldots \cap A_{n,i_r})$, we put

$$\alpha_{n,r} := \tfrac{s}{n}\left(r + \tfrac{rM}{\log n}\right), \tag{20}$$

where $M$ is given in (12). Moreover, let $R_{n,1} := \overline{R}_{n,1} := 0$ and, for $r \geq 2$ and

$h \in \{2,\dots,r\}$ with $\overline{\Delta}_{n,i_1,\dots,i_r}$ given in (13), set

$$R_{n,r} := \sum_{h=2}^{r}(-1)^{h-1}\mathbf{s}_{n,h}\Big(\tbinom{r}{h} + \tfrac{1}{\log n}\textstyle\sum_{1\le \ell_1<\ldots<\ell_h\le r}\overline{\Delta}_{n,i_{\ell_1},\dots,i_{\ell_h}}\Big), \tag{21}$$

$$\overline{R}_{n,r} := \sum_{h=2}^{r}\mathbf{s}_{n,h}\Big(\tbinom{r}{h} + \tfrac{M}{\log n}\tbinom{r}{h}\Big), \tag{22}$$

where

$$\mathbf{s}_{n,h} := \frac{n}{s}\cdot\frac{\binom{n-h}{s-h}}{\binom{n}{s}} = \frac{(s-1)\cdots(s-h+1)}{(n-1)\cdots(n-h+1)}. \tag{23}$$

Furthermore, if $r \ge 2$ we define

$$\beta_{n,r} := \sum_{h=2}^{r}\Big(\mathbf{s}_{n,h}\tbinom{r}{h}(x+\log n) + \mathbf{s}_{n,h}\tfrac{x+\log n}{\log n}\textstyle\sum_{1\le \ell_1<\ldots<\ell_h\le r}\overline{\Delta}_{n,i_{\ell_1},\dots,i_{\ell_h}}\Big), \tag{24}$$

$$\gamma_{n,r} := \exp\Big(-\tfrac{\alpha^+_{n,r}}{1-\alpha^+_{n,r}}\Big(rx + r\log n + \tfrac{x+\log n}{\log n}rM\Big)\Big), \tag{25}$$

where

$$\alpha^+_{n,r} := \tfrac{s}{n}\Big(r + \tfrac{rM}{\log n} + \overline{R}_{n,r}\Big). \tag{26}$$

Finally, we put $\beta_{n,1} := 0$ and $\alpha^+_{n,1} := \alpha_{n,1}$. The following lemma states bounds on $\mathbb{P}(A_{n,i_1}\cap\ldots\cap A_{n,i_r})$. Its proof is given in Section 4.

**Lemma 2.** *We have*

$$\mathbb{P}(A_{n,i_1}\cap\ldots\cap A_{n,i_r}) \le \frac{\mathrm{e}^{-rx}}{n^r}\cdot\exp\Big(-\frac{x+\log n}{\log n}\sum_{\ell=1}^{r}\overline{\Delta}_{n,i_\ell}\Big)\cdot \mathrm{e}^{\alpha_{n,r}+\beta_{n,r}}$$

$$\mathbb{P}(A_{n,i_1}\cap\ldots\cap A_{n,i_r}) \ge \frac{\mathrm{e}^{-rx}}{n^r}\cdot\exp\Big(-\frac{x+\log n}{\log n}\sum_{\ell=1}^{r}\overline{\Delta}_{n,i_\ell}\Big)\cdot\gamma_{n,r}\cdot \mathrm{e}^{-(x+\log n)\overline{R}_{n,r}}$$
$$\cdot\exp\Big(-\tfrac{\alpha^+_{n,r}}{1-\alpha^+_{n,r}}(x+\log n)\overline{R}_{n,r}\Big).$$

We first show (19) for $r=1$. Since $\beta_{n,1} = 0 = \overline{R}_{n,1}$, Lemma 2 yields

$$\sum_{i=1}^{n}\mathbb{P}(A_{n,i}) \le \mathrm{e}^{-x}\frac{1}{n}\sum_{i=1}^{n}\exp\Big(-\frac{x+\log n}{\log n}\sum_{\ell=1}^{r}\overline{\Delta}_{n,i_\ell}\Big)\cdot \mathrm{e}^{\alpha_{n,1}},$$

$$\sum_{i=1}^{n}\mathbb{P}(A_{n,i}) \ge \mathrm{e}^{-x}\frac{1}{n}\sum_{i=1}^{n}\exp\Big(-\frac{x+\log n}{\log n}\sum_{\ell=1}^{r}\overline{\Delta}_{n,i_\ell}\Big)\cdot\gamma_{n,1}.$$

In view of condition (14) and the fact that $\alpha_{n,1}\to 0$ and $\gamma_{n,1}\to 1$ as $n\to\infty$, (19) follows for $r=1$. For $r\ge 2$, Lemma 2 gives

$$\sum_{1\le i_1<\ldots<i_r\le n}\mathbb{P}(A_{n,i_1}\cap\ldots\cap A_{n,i_r}) \le \frac{\mathrm{e}^{-rx}}{n^r}\cdot \mathrm{e}^{\alpha_{n,r}+\beta_{n,r}}\sum_{1\le i_1<\ldots<i_r\le n}\exp\Big(-\frac{x+\log n}{\log n}\sum_{\ell=1}^{r}\overline{\Delta}_{n,i_\ell}\Big).$$

Now, $\alpha_{n,r} \to 0$ as $n \to \infty$, and straightforward algebra shows that also $\beta_{n,r} \to 0$ as $n \to \infty$. Hence,

$$\limsup_{n\to\infty} \sum_{1\le i_1<\ldots<i_r\le n} \mathbb{P}(A_{n,i_1} \cap \ldots \cap A_{n,i_r}) \\ \le \mathrm{e}^{-rx} \limsup_{n\to\infty} \frac{1}{n^r} \sum_{1\le i_1<\ldots<i_r\le n} \exp\Big( -\frac{x+\log n}{\log n} \sum_{\ell=1}^{r} \overline{\Delta}_{n,i_\ell}\Big). \tag{27}$$

In the same way, one can use the second inequality of Lemma 2. Together with $\gamma_{n,r} \to 1$, $\log n \overline{R}_{n,r} \to 0$ and $\alpha^+_{n,r} \to 0$, where $\gamma_{n,r}$, $\overline{R}_{n,r}$ and $\alpha^+_{n,r}$ are defined in (25), (22) and (26), respectively, one obtains the reverse inequality in (27), with each lim sup replaced by lim inf. We will now show that

$$\lim_{n\to\infty} \frac{1}{n^r} \sum_{1\le i_1<\ldots<i_r\le n} \exp\Big( -\frac{x+\log n}{\log n} \sum_{\ell=1}^{r} \overline{\Delta}_{n,i_\ell}\Big) = \frac{c(x)^r}{r!}, \tag{28}$$

which completes the proof. To this end, notice that the sum figuring in (28) equals

$$\frac{1}{r!} \sum_{i_1=1}^{n} \cdots \sum_{i_r=1}^{n} \exp\Big( -\frac{x+\log n}{\log n} \sum_{\ell=1}^{r} \overline{\Delta}_{n,i_\ell}\Big) - \frac{1}{r!} \sum_{I} \exp\Big( -\frac{x+\log n}{\log n} \sum_{\ell=1}^{r} \overline{\Delta}_{n,i_\ell}\Big),$$

where $\sum_I$ denotes summation over all $(i_1, \ldots, i_r) \in \{1, \ldots, n\}^r$ such that the cardinality of $\{i_1, \ldots, i_r\}$ is at most $r-1$. Since this sum is of order $O(n^{r-1})$, the assertion follows from (14).

## 4. Proofs of auxiliary results

Proof of Lemma 1. Put $b_+ := b_1 + \ldots + b_r$, and recall $r(c-1) = \sum_{j=1}^r jb_j$. For each $s$-placement, the probability of occupying exactly $i$ of the cells $1, \ldots, r$ with a particle is given by $\binom{n-r}{s-i}/\binom{n}{s}$, and the probability of not occupying any of these cells is $\binom{n-r}{s}/\binom{n}{s}$. By independence of events related to different $s$-placements, and counting cases, we obtain

$$\mathbb{P}(N_1 = c-1, \ldots, N_r = c-1, B = b) = \prod_{i=1}^{r} \left[\frac{\binom{n-r}{s-i}}{\binom{n}{s}}\right]^{b_i} \left[\frac{\binom{n-r}{s}}{\binom{n}{s}}\right]^{k_n-b_+} \frac{k_n!}{b_1!\cdots b_r!(k_n-b_+)!}. \tag{29}$$

Now, $\binom{n-r}{s-i}/\binom{n}{s} = O(n^{-i})$ as $n \to \infty$, and thus the product over $i$ figuring on the right hand side of (29) is of the order $O(n^{-\sum_{i=1}^r ib_i})$. Taking logarithms and using

$\log t \le t-1$, $t>0$, and the definition of $k_n$, some calculations yield

$$\left[\frac{\binom{n-r}{s}}{\binom{n}{s}}\right]^{k_n-b_+} = O\left(n^{-r}(\log n)^{-r(c-1)}\right).$$

In view of

$$\frac{k_n!}{b_1!\cdots b_r!(k_n-b_+)!} = \frac{k_n^{b_+}(1+o(1))}{b_1!\cdots b_r!} = O\left(n^{b_+}(\log n)^{b_+}\right),$$

$-r-\sum_{i=1}^r ib_i + b_+ = -r - \sum_{i=2}^r (i-1)b_i$ and $b_+ - r(c-1) \le 0$, the assertion follows.

Proof of Lemma 2. By independence and the inclusion-exclusion principle,

$$\mathbb{P}(A_{n,i_1}\cap\ldots\cap A_{n,i_r}) = \left[1-\sum_{h=1}^r \sum_{1\le \ell_1<\ldots<\ell_h\le r}(-1)^{h-1}P_{n,i_{\ell_1},\ldots,i_{\ell_h}}\right]^{k_n},$$

where, by (11) and (13),

$$\begin{aligned}
P_{n,i_{\ell_1},\ldots,i_{\ell_h}} &= \sum_{\mathcal{I}\subset[n]_s:\mathcal{I}\supset\{i_{\ell_1},\ldots,i_{\ell_h}\}} p_{n,\mathcal{I}} = \sum_{\mathcal{I}\subset[n]_s:\mathcal{I}\supset\{i_{\ell_1},\ldots,i_{\ell_h}\}} \frac{1}{\binom{n}{s}}\Big(1+\frac{\Delta_{n,\mathcal{I}}}{\log n}\Big)\\
&= \frac{\binom{n-h}{s-h}}{\binom{n}{s}}\Big(1+\frac{\overline{\Delta}_{n,i_{\ell_1},\ldots,i_{\ell_h}}}{\log n}\Big).
\end{aligned}$$

Hence

$$\mathbb{P}(A_{n,i_1}\cap\ldots\cap A_{n,i_r}) = \left[1-\left(\sum_{h=1}^r(-1)^{h-1}\frac{\binom{n-h}{s-h}}{\binom{n}{s}}\left(\binom{r}{h}+\frac{1}{\log n}\sum_{1\le\ell_1<\ldots<\ell_h\le r}\right)\right)\right]^{k_n}.$$

From (21) and the inequality $\log t\le t-1$, $t>0$, we obtain

$$\begin{aligned}
\log\mathbb{P}(A_{n,i_1}\cap\ldots\cap A_{n,i_r}) &= k_n\log\left(1-\frac{s}{n}\left(r+\frac{1}{\log n}\sum_{\ell=1}^r\overline{\Delta}_{n,i_\ell}+R_{n,r}\right)\right) \qquad (30)\\
&\le -\frac{k_ns}{n}\left(r+\frac{1}{\log n}\sum_{\ell=1}^r\overline{\Delta}_{n,i_\ell}+R_{n,r}\right)\\
&= -\frac{k_ns}{n}\left(r+\frac{1}{\log n}\sum_{\ell=1}^r\overline{\Delta}_{n,i_\ell}\right)-\frac{k_ns}{n}R_{n,r}. \qquad (31)
\end{aligned}$$

Using (12) and (20), the first summand of (31) is bounded from above by

$$-\frac{k_ns}{n}\left(r+\frac{1}{\log n}\sum_{\ell=1}^r\overline{\Delta}_{n,i_\ell}\right)\le -rx-r\log n-\frac{x+\log n}{\log n}\sum_{\ell=1}^r\overline{\Delta}_{n,i_\ell}+\alpha_{n,r}.$$

For the second summand we have

$$-\frac{k_n s}{n} R_{n,r} \leq k_n \sum_{h=2}^{r} \frac{\binom{n-h}{s-h}}{\binom{n}{s}} \left( \binom{r}{h} + \frac{1}{\log n} \sum_{1 \leq \ell_1 < \ldots < \ell_h \leq r} \overline{\Delta}_{i_{\ell_1},\ldots,i_{\ell_h}} \right).$$

Moreover, with $\mathbf{s}_{n,h}$ defined in (23), it follows that

$$\begin{aligned} & k_n \frac{\binom{n-h}{s-h}}{\binom{n}{s}} \left( \binom{r}{h} + \frac{1}{\log n} \sum_{1 \leq \ell_1 < \ldots < \ell_h \leq r} \overline{\Delta}_{i_{\ell_1},\ldots,i_{\ell_h}} \right) \\ \leq \quad & \mathbf{s}_{n,h} \binom{r}{h} (x + \log n) + \mathbf{s}_{n,h} \frac{x + \log n}{\log n} \sum_{1 \leq \ell_1 < \ldots < \ell_h \leq r} \overline{\Delta}_{i_{\ell_1},\ldots,i_{\ell_h}}. \end{aligned}$$

Thus $-k_n s R_{n,r}/n \leq \beta_{n,r}$, where $\beta_{n,r}$ is defined in (24), and the upper bound of Lemma 2 follows. The lower bound is obtained if we apply the inequality $\log t \geq 1 - 1/t$, $t > 0$, to the right hand side of (30) and proceed similarly. The details are omitted.

**Remark 2.** In the special case $s = 1$, the method of moments can be dispensed with, and proofs can be simplified by using the Stein–Chen method and coupling, see e.g. the proof of Theorem 2.1 of [29]. In the setting of Section 3, a suitable coupling would arise if we throw each of the particles which have fallen into cell $j$ independently of each other into one of the other cells, according to the probability $p_{n,i}/(1 - p_{n,j})$, $i \neq j$ (see, e.g., Section 3 and Theorem 2.3 of [5]). Such a coupling, however, is not possible if $s > 1$ since, with each $s$-placement, $s$ different cells are occupied, and the removal of each particle of cell $j$ would mean to add $s - 1$ further particles in order to be able to start an $s$-placement according to the conditional distribution of (11) given that cell $j$ must not be occupied.

**Remark 3.** This paper is based on a part of the first author's doctoral dissertation [33], written under the supervision of the second author.

**Acknowledgements:** The authors would like to thank an anonymous referee and an editor for helpful remarks that improved the presentation.

## References

[1] Anceaume, E., Busnel, Y., and Sericola, B. (2015). New results on a generalized coupon collector problem using Markov chains. *J. Appl. Probab.*, *52*, 405–418.

[2] Anceaume, E., Busnel, Y., Schulte-Geers, E., and Sericola, B. (2016). Optimization results for a generalized coupon collector problem. *J. Appl. Probab.*, *53*, 622–629.

[3] Aki, S., and Hirano, K. (2013). Coupon collector's problems with statistical applications to rankings. *Ann. Inst. Statist. Math.*, *65*, 571–587.

[4] Angus, J.E. (2013). A variation on the coupon collecting problem. *Comm. Statist. Simulation Comput.*, *42*, 2197–2202.

[5] Barbour, A.D., and Holst, L. (1989). Some applications of the Stein-Chen method for proving Poisson convergence. *Adv. in Appl. Probab.*, *21*, 74–90.

[6] Berenbrink, P., and Sauerwald, Th. (2009). The weighted coupon collector's problem and applications. Computing and combinatorics, 449–458, Lecture Notes in Comput. Sci., 5609, Springer, Berlin.

[7] Berenbrink, P., Elsässer, R., Friedetzky, T., Nagel, L., and Sauerwald, Th. (2011). Faster coupon collecting via replication with applications in gossiping. Mathematical foundations of computer science 2011, 72–83. Lecture Notes in Comput. Sci., 6907, Springer, Heidelberg.

[8] Boneh, A., and Hofri, M. (1997). The coupon-collector problem revisited - a survey of engineering problems and computational methods. *Commun. Statist. – Stoch. Models*, *13*, 39–66.

[9] Brown, M., Peköz, E., and Ross, S.M. (2008). Coupon collecting. *Probab. Engrg. Inform. Sci.*, *22*, 221–229.

[10] Brown, M., and Ross, S.M. (2016). Optimality results for coupon collection. *J. Appl. Probab.*, *53*, 930–937.

[11] Dobson, G., and Tezcan, T. (2015). Optimal sampling strategies in the coupon collector's problem with unknown population size. *Ann. Oper. Res.*, *233*, 77–99.

[12] Doumas, A., and Papanicolaou, V. (2012). The coupon collector's problem revisited: asymptotics of the variance. *Adv. in Appl. Probab.*, *44*, 166–195.

[13] Doumas, A., and Papanicolaou, V. (2013). Asymptotics of the rising moments for the coupon collector's problem. *Electron. J. Probab.*, *18*, No. 41, 15 pp.

[14] Erdős, P., and Rényi, A. (1961). On a classical problem of probability theory. *Magyar Tud. Akad. Mat. Kutató Int. Közl.*, *6*, 215–220.

[15] Falgas-Ravry, V., Larsson, J., and Markström, K. (2020). Speed and concentration of the covering time for structured coupon collectors. *Adv. in Appl. Probab.*, *52*, 433–462.

[16] Ferrante, M., and Saltalamacchia, M. (2014). The coupon collector's problem. *MATerials MATemàtics*, pages 01–35.

[17] Fu, J.C., and Lee, W.-C. (2017). On coupon collector's and Dixie cup problems under fixed and random sample size sampling schemes. *Ann. Inst. Statist. Math.*, *69*, 1129–1139.

[18] Glavaš, L., and Mladenović, P. (2018). New limit results related to the coupon collector's problem. *Studia Sci. Math. Hungar.*, *55*, 115–140.

[19] Graham, R., Knuth, D., and Patashnik, O. (1995). *Concrete Mathematics*. Addison-Wesley Publishing Company, Boston.

[20] Ilienko, A. (2019). Convergence of point processes associated with coupon collector's and Dixie cup problems. *Electron. Commun. Probab.*, *24*, Paper No. 51, 9 pp.

[21] Inoue, K., and Aki, S. (2008). Methods for studying generalized birthday and coupon collection problems. *Comm. Statist. Simulation Comput.*, *37*, 844–862.

[22] Jocković, J., and Mladenović, P. (2011). Coupon collector's problem and generalized Pareto distributions. *J. Statist. Plann. Infer.*, *141*, 2348–2352.

[23] Jocković, J., and Mladenović, P. (2014). Coupon collector's problem and its extensions in extreme value framework. *Stat. Interface*, *7*, 381–388.

[24] Kella, O., and Stadje, W. (2008). A collector's problem with renewal arrival processes. *J. Appl. Probab.*, *45*, 610–620.

[25] Mahmoud, H.M. (2010). Gaussian phases in generalized coupon collection. *Adv. in Appl. Probab.*, *42*, 994–1012.

[26] Mahmoud, H.M., and Smythe, R.T. (2012). On the joint behavior of types of coupons in generalized coupon collection. *Adv. in Appl. Probab.*, *44*, 429–451.

[27] May, R. (2008). Coupon collecting with quotas. *Electron. J. Combin.*, *15*, No.1, Note 31, 7 pp.

[28] Mikhailov, V.G. (1977). A Poisson limit theorem in the scheme of group disposal of particels. *Th. Probab. Applic.*, *22*, 152–156.

[29] Neal, P. (2008). The generalised coupon collector problem. *J. Appl. Probab.*, *45*, 621–629.

[30] Pósfai, A., and Csörgő, S. (2009). Asymptotic approximations for coupon collectors. *Studia Sci. Math. Hungar.*, *46*, 61–96.

[31] Pósfai, A. (2009). Poisson approximation in a Poisson limit theorem inspired by coupon collecting. *J. Appl. Probab.*, *46*, 585–592.

[32] Ross, S.M., and Wu, D.T. (2013). A generalized coupon collecting model as a parsimonious optimal stochastic assignment model. *Ann. Oper. Res.*, *208*, 133–146.

[33] Schilling, J. (2020). Untersuchungen zur Asymptotik und zum Erwartungswert im verallgemeinerten Coupon-Collector-Problem (A study on the asymptotics and the expected value in the generalized coupon collector's problen (in German)). *Doctoral dissertation, Karlsruhe Institute of Technology (KIT)*, 128pp. DOI: 10.5445/IR/1000117773

[34] Shank, N., and Yang, H. (2013). Coupon collector problem for non-uniform coupons and random quotas. *Electron. J. Combin.*, *20*, No. 2, Paper 33, 10 pp.

[35] Smythe, R.T. (2011). Generalized coupon collection: the superlinear case. *J. Appl. Probab.*, *48*, 189–199.

[36] Stadje, W. (1990). The collector's problem with group drawings. *Adv. in Appl. Probab.*, *22* 866–882.

[37] Todhunter, I. (1865). *A History of the Mathematical Theory of Probability from the Time of Pascal to that of Laplace.* Cambrige, London, Macmillan and co.

[38] Xu, W., and Tang, A.K. (2011). A generalized coupon collector problem. *J. Appl. Probab.*, *48*, 1081–1094.