J. Appl. Probab. **60,** 1352–1366 (2023)
*doi:*10.1017/jpr.2023.6

APPROACHING THE COUPON COLLECTOR'S PROBLEM WITH GROUP DRAWINGS VIA STEIN'S METHOD

CARINA BETKEN,* ** AND
CHRISTOPH THÄLE,* *** *Ruhr University Bochum*

Abstract

We study the coupon collector's problem with group drawings. Assume there are n different coupons. At each time precisely s of the n coupons are drawn, where all choices are supposed to have equal probability. The focus lies on the fluctuations, as $n \to \infty$, of the number $Z_{n,s}(k_n)$ of coupons that have not been drawn in the first k_n drawings. Using a size-biased coupling construction together with Stein's method for normal approximation, a quantitative central limit theorem for $Z_{n,s}(k_n)$ is shown for the case that $k_n = (n/s)(\alpha \log(n) + x)$, where $0 < \alpha < 1$ and $x \in \mathbb{R}$. The same coupling construction is used to retrieve a quantitative Poisson limit theorem in the boundary case $\alpha = 1$, again using Stein's method.

Keywords: Central limit theorem; coupon collector's problem; Poisson limit theorem; size-biased coupling; Stein's method

2020 Mathematics Subject Classification: Primary 60C05
Secondary 60F05

1. Introduction

The coupon collector's problem is an old problem of probability theory which in its simplest form dates back to de Moivre, Laplace, and Euler [4, 6, 10]. While de Moivre used a die with s faces to pose the problem, Euler and Laplace used a lottery interpretation as motivation. However, a more recent example for a situation in which the coupon collector's problem occurs is the collection of pictures of the participating players of all teams before and during every World Cup. Typically, fans can buy the pictures in packages of five or six. Two natural questions which arise are: How many packages need to be bought to get the full or a specific portion of the full set of players? How many stickers are missing after buying k packages? The first question was studied, for example, in [3, 8, 9, 18]. In the work at hand we will deal with the latter of the two problems. The version of the coupon collector's problem we consider can be described as follows. Assume there are n different coupons. At each time we draw s of these n coupons, where we assume that each of the $\binom{n}{s}$ choices occurs with the same probability. We are then interested in the distribution of the number $Z_{n,s}(k_n)$ of coupons that have not been drawn in the first $k = k_n$ drawings. In a conceptually equivalent interpretation the n coupons are represented by n different cells numbered $1, \ldots, n$, and in each drawing we place s particles

Received 16 February 2022; revision received 12 January 2023.
* Postal address: Faculty of Mathematics, Ruhr University Bochum, 44780 Bochum, Germany.
** Email address: carina.betken@rub.de
*** Email address: christoph.thaele@rub.de

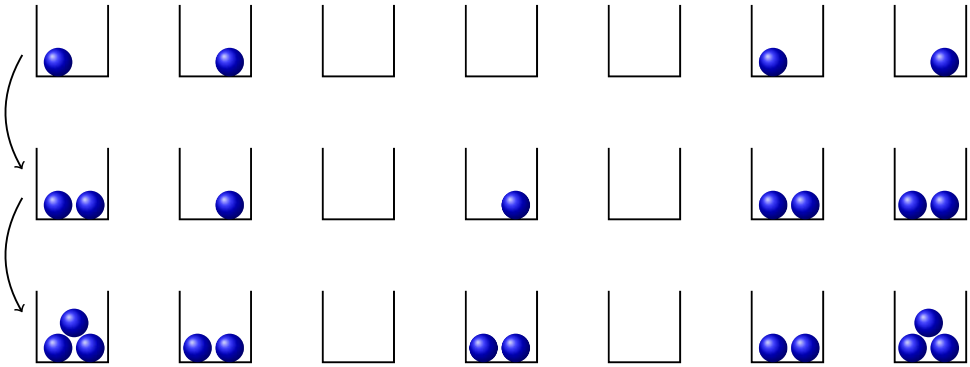

FIGURE 1. Illustration of the cell interpretation with $n = 7$ and $s = 4$ in the first three drawings.

into s distinct cells; see Fig. 1. We are then interested in the distribution of the number $Z_{n,s}(k_n)$ of empty cells after $k = k_n$ drawings.

The behaviour of $Z_{n,s}(k_n)$ for the different regimes of k_n has been the subject of numerous research works over the years. In [11] convergence towards a normal limit is proved for the sublinear and linear regime, i.e. for $k_n = o(n)$ and $k_n = \alpha n$, respectively. In [17] the author proves a central limit theorem for a generalised coupon collector's problem allowing for *random* package sizes S in the lower superlinear regime, i.e. for $k_n/n \to \infty$ and $\limsup_n k_n/(n \log(n)) < 1/\mathbb{E}[S]$ using a martingale representation. In [13] the method of moments is used to prove normal approximation for the case where $n[(k_n s)/n]^r \to \infty$ for all $r \in \mathbb{N}$ and $\mathbb{E}[Z_{n,s}(k_n)] \to \infty$, which covers our regime of normal approximation introduced below. However, no rates of convergence are given in either of the works mentioned so far. In [19] the authors deduce rates of convergence in the Kolmogorov distance towards a normal limit for $k_n = o(n \log(n))$ which are of order $1/\sqrt{\mathrm{var}(Z_{n,s}(k_n))}$ and thus optimal by a general result of Englund [5, p. 692], which shows that for integer-valued random variables such as $Z_{n,s}(k_n)$ the order of this rate cannot be improved. We complement these bounds in the case where k_n is assumed to be of the form $k_n = (n/s)(\alpha \log(n) + x)$ for some $\alpha \in (0, 1)$ and $x \in \mathbb{R}$. In the case $\alpha = 1$, i.e. if $k_n = (n/s)(\log(n) + x)$, the author in [12] uses the Stein–Chen method to prove convergence towards a Poisson limit for the number of cells containing exactly r particles in a more general setting allowing for multiple particles being placed in one cell at each step. The same question is also studied in [16], and a Poisson limit is deduced. The same work shows that for the special case $r = 0$ the condition of equally probable group drawings can be relaxed to a certain extent without losing the limiting Poisson distribution. In [2, Theorem 6.F] the authors also prove rates of convergence of order $\log(n)/n$ towards a Poisson distribution in this regime, which are optimal in view of [2, Theorem 3.D]. For $r = 0$ this covers our setting of Poisson approximation with the same rates, which are included here only for completeness and to demonstrate that both limit theorems can be based on the same coupling argument.

As explained above, there exists a sharp asymptotic distributional phase transition at $\alpha = 1$ in the sense that for $\alpha \in (0, 1)$ the random variable $Z_{n,s}(k_n)$ asymptotically follows a normal distribution, whereas for $\alpha = 1$ we obtain a Poisson limit. However, in both cases we use Stein's method in combination with the same size-biased coupling construction to prove upper bounds on the distance between $Z_{n,s}(k_n)$ and a Gaussian or Poisson random variable, respectively. Our results are presented in the next section, while the coupling construction is explained

https://doi.org/10.1017/jpr.2023.6 Published online by Cambridge University Press

in Section 3. The proof of the normal approximation result for $\alpha \in (0, 1)$ is the content of Section 4, while the Poisson limit theorem is derived in Section 5.

2. Results

Denote by $(C_i)_{i=1,\dots,n}$ the collection of cells in the coupon collector's problem, and let $Z_{n,s}(k_n)$ be the number of empty cells after k_n drawings, i.e. $Z_{n,s}(k_n) = \sum_{j=1}^{n} \mathbf{1}_{E_{n,j}(k_n)}$, where $E_{n,j}(k) = \{|C_j| = 0 \text{ after } k_n \text{ drawings}\}$ and where $|C_j|$ stands for the number of particles in cell C_j. Moreover, for two random variables X and Y we denote by

$$d_{\mathrm{W}}(X, Y) := \sup_{h \in \mathrm{Lip}(1)} \left| \mathbb{E}[h(X)] - \mathbb{E}[h(Y)] \right|$$

the Wasserstein distance between X and Y, where the supremum runs over all Lipschitz functions $h : \mathbb{R} \to \mathbb{R}$ with Lipschitz constant less than or equal to one. We consider the case where $k_n = (n/s)(\alpha \log(n) + x)$ for fixed $s \in \mathbb{N}$, $x \in \mathbb{R}$, and $\alpha \in (0, 1)$. Note that since k_n denotes the number of drawings, we always assume implicitly that k_n is an integer; in particular we assume that n is large enough that $k_n \geq 0$. Furthermore, we define the centred and normalised random variables

$$\tilde{Z}_{n,s}(k_n) := \frac{Z_{n,s}(k_n) - \mathbb{E}[Z_{n,s}(k_n)]}{\sqrt{\mathrm{var}(Z_{n,s}(k_n))}}. \tag{1}$$

Throughout the paper we use the notation $C(x_1, x_2, \dots)$ to indicate that a constant $C \in (0, \infty)$ only depends on parameters $x_1, x_2, \dots$ of the model.

Theorem 1. *Put $k_n = (n/s)(\alpha \log(n) + x)$ for some $s \in \mathbb{N}$, $x \in \mathbb{R}$, and $\alpha \in (0, 1)$. Let $Z_{n,s}(k_n)$ be the number of empty cells after k_n drawings as introduced above, and denote by G a standard Gaussian random variable. Then there exist constants $C = C(x, \alpha) \in (0, \infty)$ and $N = N(s, x, \alpha) \in \mathbb{N}$ such that, for all $n \geq N$,*

$$d_{\mathrm{W}}(\tilde{Z}_{n,s}(k_n), G) \leq C\left(\frac{\sqrt{\log(n)}}{n^{\alpha}} + \frac{s^2}{n^{(1-\alpha)/2}} \right). \tag{2}$$

After this general bound we now consider the situation in which s behaves like a constant multiple of a non-negative power of n. In particular, this covers the case where s is constant. We denote by $[y]$ the integer part of a real number $y \in \mathbb{R}$.

Corollary 1. *In the situation of Theorem 1, suppose additionally that s is of the form $s = [s_0 n^{\beta}]$ for some $\beta \in \left[0, \frac{1-\alpha}{4}\right)$ and $s_0 \geq 1$. Then,*

$$d_{\mathrm{W}}(\tilde{Z}_{n,s}, G) \leq \begin{cases} C\sqrt{\log n}/n^{\alpha} & \text{for } \alpha \in \left(0, \frac{1}{3}\right], \\ Cn^{-(1-\alpha)/2+2\beta} & \text{for } \alpha \in \left(\frac{1}{3}, 1\right) \end{cases}$$

for $n \geq N$, where $C = C(x, \alpha, s_0) \in (0, \infty)$, $N = N(x, \alpha, s_0, \beta) \in (0, \infty)$. In particular, if $s \equiv s_0 \geq 1$ is constant,

$$d_{\mathrm{W}}(\tilde{Z}_{n,s}, G) \leq \begin{cases} C\sqrt{\log n}/n^{\alpha} & \text{for } \alpha \in \left(0, \frac{1}{3}\right], \\ Cn^{-(1-\alpha)/2} & \text{for } \alpha \in \left(\frac{1}{3}, 1\right). \end{cases}$$

Proof. The bounds are immediate from Theorem 1 by plugging in the particular choice for s. □

https://doi.org/10.1017/jpr.2023.6 Published online by Cambridge University Press

Remark 1.

(i) Since $n^{-(1-\alpha)/2}$ is of the same order as $1/\sqrt{\mathrm{var}(Z_{n,s}(k_n))}$, we believe that the rate in Corollary 1 is optimal in the regime $\alpha \in \left(\frac{1}{3}, 1\right)$ and if s is constant ($\beta = 0$). We leave it as an open problem to decide whether or not the rate is optimal for $\alpha \in \left(0, \frac{1}{3}\right]$. On the other hand, Remark 4 shows that in this situation the rate cannot be improved by arguments based on the general normal approximation bound (7) (although this does, of course, exclude the possibility of improving the bound by *other* methods).

(ii) We might ask whether the Wasserstein distance in Theorem 1 and Corollary 1 can be replaced by the Kolmogorov distance

$$d_K\left(\tilde{Z}_{n,s}(k_n), N\right) = \sup_{u\in\mathbb{R}} \left|\mathbb{P}\left(\tilde{Z}_{n,s}(k_n) \le u\right) - \mathbb{P}(N \le u)\right|.$$

As explained in Remark 3 in more detail, this is not possible by means of the size-biased coupling approach of Stein's method for normal approximation using our coupling construction. In fact, the resulting bound in this case does not even tend to zero with n. On the other hand, a presumably suboptimal bound follows directly from the fact that the Kolmogorov distance can always be bounded by the square root of the Wasserstein distance.

(iii) Remark 5 demonstrates that it is possible to make the constants C and N appearing in Theorem 1 (and thus Corollary 1) fully explicit in terms of the parameters x and α. However, since the resulting expressions are rather involved, we decided to present our results in a simplified form. A similar comment also applies to the constants appearing in Theorem 2.

The next result complements Theorem 1 by considering the case $\alpha = 1$ for which the upper bound (2) does not tend to zero as $n \to \infty$. As emphasised already, the result is known from [2] and is included here only for completeness. As above, for two random variables X and Y we denote by

$$d_{\mathrm{TV}}(X, Y) := \sup_A \left|\mathbb{P}(X \in A) - \mathbb{P}(Y \in A)\right|$$

the total variation distance between X and Y, where the supremum is taken over all Borel sets $A \subset \mathbb{R}$.

Theorem 2. *Put $k_n = (n/s)(\log(n) + x)$ for some $s \in \mathbb{N}$ and $x \in \mathbb{R}$. Let $Z_{n,s}(k_n)$ be the number of empty cells after k_n drawings, and denote by W a Poisson random variable with parameter $\lambda_n = \mathbb{E}[Z_{n,s}(k_n)]$. Then there exist constants $\tilde{C} = \tilde{C}(s, x) \in (0, \infty)$ and $\tilde{N} = \tilde{N}(s, x) \in \mathbb{N}$ such that, for all $n \ge \tilde{N}$,*

$$d_{\mathrm{TV}}(Z_{n,s}(k_n), W) \le \tilde{C} \log(n)/n.$$

Remark 2. It has been shown in [2] that the rate in Theorem 2 is optimal in the sense that we can find other constants $\widehat{C}, \widehat{N} \in (0, \infty)$, depending on s and x only, such that $d_{\mathrm{TV}}(Z_{n,s}(k_n), W) \ge \widehat{C} \log(n)/n$ for all $n \ge \widehat{N}$.

For the proof of both results we use Stein's method in combination with a size-biased coupling. We start by describing this coupling in the next section, which can be regarded as a particular instant of the construction in [14, p. 623] (choosing $p_{ni} = 1/n$ there).

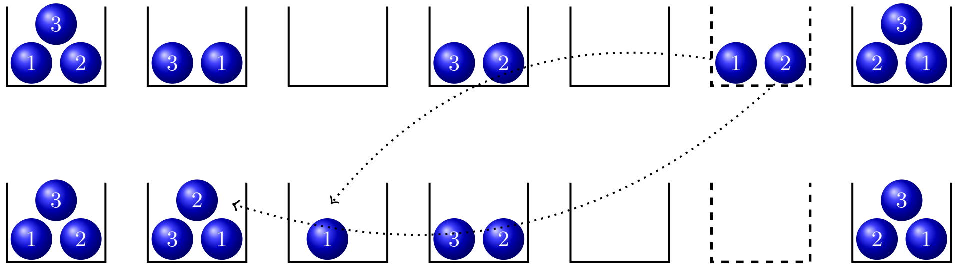

FIGURE 2. Illustration of the coupling construction (continuation of Fig. 1); the artificially isolated cell with label I is the dashed one.

3. Coupling construction

We are dealing with the coupon collector's problem with n coupons, which can be interpreted as n distinct cells. At each time step we place a fixed number $s \leq n$ of particles in s different cells. In order to keep track of when particles are placed into cells, we label all particles in the mth drawing with the letter m. Now, after k_n placements, we choose one of the n cells, which we denote by C_I, uniformly at random and take all particles out of it. For a particle labelled j taken from C_I we now choose one of the $n-s$ cells not containing a particle with label j uniformly and place the particle into it. We proceed in the same manner until all particles from cell I have been redistributed into the remaining $n-1$ cells; see Fig. 2.

Denote by $F^I_{n,j}$ the event that at least one particle from cell I is placed into cell j, and put $\bar{F}^I_{n,j} := \big(F^I_{n,j}\big)^c$. Furthermore, we define $E^I_{n,j}(k_n) := E_{n,j}(k_n) \cap \bar{F}^I_{n,j}$. For any $j \neq I$ we then have

$$\mathbb{P}\Big(E^I_{n,j}(k_n)\Big) = \sum_{\ell=0}^{k_n} \mathbb{P}\Big(E_{n,j}(k_n) \cap \bar{F}^I_{n,j} \cap \{|C_I| = \ell\}\Big)$$
$$= \sum_{\ell=0}^{k_n} \mathbb{P}\Big(\bar{F}^I_{n,j} \mid E_{n,j}(k_n) \cap \{|C_I| = \ell\}\Big)\mathbb{P}(|C_I| = \ell \mid E_{n,j}(k_n))\mathbb{P}\big(E_{n,j}(k_n)\big). \quad (3)$$

Note that by construction we have

$$\mathbb{P}(E_{n,j}(k_n)) = \Big(1 - \frac{s}{n}\Big)^{k_n}, \quad (4)$$

$$\mathbb{P}(|C_I| = \ell \mid E_{n,j}(k_n)) = \binom{k_n}{\ell}\left(\frac{s}{n-1}\right)^{\ell}\left(1 - \frac{s}{n-1}\right)^{k_n-\ell}, \quad (5)$$

since conditioning on the event $E_{n,j}(k_n)$ simply means that we can only place particles in $n-1$ instead of n cells. In addition,

$$\mathbb{P}\big(\bar{F}^I_{n,j} \mid E_{n,j}(k_n) \cap \{|C_I| = \ell\}\big) = \left(1 - \frac{1}{n-s}\right)^{\ell},$$

since for each of the particles from cell I the probability that it is placed into an originally empty cell is $1/(n-s)$. Putting these expressions back into (3), we obtain

$$\begin{aligned}\mathbb{P}\big(E^I_{n,j}(k_n)\big) &= \sum_{\ell=0}^{k_n}\binom{k_n}{\ell}\left(\frac{s}{n-1}\right)^{\ell}\left(1-\frac{s}{n-1}\right)^{k_n-\ell}\left(1-\frac{1}{n-s}\right)^{\ell}\left(1-\frac{s}{n}\right)^{k_n}\\ &= \left(1-\frac{s}{n-1}\right)^{k_n} = \mathbb{P}\big(E_{n,j}(k_n)\mid E_{n,i}(k_n)\big)\end{aligned} \tag{6}$$

for any $i \neq j$. We thus conclude that the random variable

$$Z^I_{n,s}(k_n) := 1 + \sum_{\substack{j=1\\ j\neq I}}^{n} \mathbf{1}_{E^I_{n,j}(k_n)}$$

has the $Z_{n,s}(k_n)$-size-biased distribution, meaning that

$$\mathbb{P}\big(Z^I_{n,s}(k_n)=y\big) = \frac{y}{\mathbb{E}[Z_{n,s}(k_n)]}\mathbb{P}(Z_{n,s}(k_n)=y), \qquad y\in\{0, 1, \ldots, n\};$$

see [1].

Remark 3. Theorem 5.6 in [1] provides a bound on the Kolmogorov distance between $\tilde{Z}_{n,s}(k_n)$ and a standard Gaussian random variable using a size-biased coupling. However, for this to yield a central limit theorem we need $|Z^I_{n,s} - Z_{n,s}| = o(\sqrt{n})$ almost surely as $n\to\infty$. For the coupling described above we have $|Z^I_{n,s} - Z_{n,s}| = n-s-1$ if in all k_n drawings the same s cells are filled and the remaining $n-s$ cells are filled when redistributing the k_n particles of one of the filled cells. Consequently, the result in [1] does not lead to a meaningful bound on the Komogorov distance, as explained in Remark 1(ii).

4. Proof of Theorem 1

Following [7, Theorem 1.1], the Wasserstein distance between $\tilde{Z}_{n,s} = \tilde{Z}_{n,s}(k_n)$ as defined in (1) and a standard Gaussian random variable G is bounded by

$$d_{\mathrm{W}}\big(\tilde{Z}_{n,s}, G\big) \le \frac{\lambda_n}{\sigma_n^2}\sqrt{\mathrm{var}\big(\mathbb{E}[Z_{n,s}-Z^I_{n,s}|\mathcal{C}_n(k_n)]\big)} + \frac{\lambda_n}{\sigma_n^3}\mathbb{E}\big[\big(Z_{n,s}-Z^I_{n,s}\big)^2\big], \tag{7}$$

where $\mathcal{C}_n(k_n)$ denotes the configuration of the n cells after k_n drawings, $\lambda_n = \mathbb{E}[Z_{n,s}(k_n)]$, and $\sigma_n^2 = \mathrm{var}(Z_{n,s}(k_n))$. In the next sections we further bound the right-hand side of (7) by dealing with the individual terms.

4.1. Expectation and variance

We start by bounding from above the expectation, and from below the variance, of $Z_{n,s}$. First, we note that, for $k_n = \frac{n}{s}(\alpha\log(n)+x)$,

$$\mathbb{P}(E_{n,j}(k_n)) = \left(1-\frac{s}{n}\right)^{k_n} = \left(\left(1-\frac{s}{n}\right)^{n/s}\right)^{\alpha\log(n)+x} \sim \mathrm{e}^{-(\alpha\log(n)+x)} = \frac{\mathrm{e}^{-x}}{n^{\alpha}}, \tag{8}$$

https://doi.org/10.1017/jpr.2023.6 Published online by Cambridge University Press

where we write $f(n) \sim g(n)$ for two functions $f, g : \mathbb{N} \to \mathbb{R}$ if $\lim_{n\to\infty} [f(n)/g(n)] = 1$. Thus,

$$\mathbb{E}[Z_{n,s}(k_n)] = \sum_{j=1}^{n} \mathbb{P}(E_{n,j}(k_n)) \sim \mathrm{e}^{-x} n^{1-\alpha}. \tag{9}$$

In particular, $\mathbb{P}(E_{n,j}(k_n)) \le e^{-x} n^{-\alpha}$ for all $n \ge 1$. Similarly, we obtain

$$\begin{aligned}\left(1-\frac{s}{n}\right)^{k_n} - \left(1-\frac{s}{n-1}\right)^{k_n} &= \left(1-\frac{s}{n}\right)^{k_n}\left(1-\left(\frac{(n-1-s)n}{(n-1)(n-s)}\right)^{k_n}\right)\\ &\le \left(1-\frac{s}{n}\right)^{k_n}\left(-k_n\left(\frac{(n-1-s)n}{(n-1)(n-s)}-1\right)\right)\\ &= \left(1-\frac{s}{n}\right)^{k_n} k_n \,\frac{(n-1)(n-s)-(n-1-s)n}{(n-1)(n-s)}\\ &= \left(1-\frac{s}{n}\right)^{k_n} \frac{k_n s}{(n-1)(n-s)}\\ &\le \frac{2\mathrm{e}^{-x}(\alpha \log(n) + x)}{n^{1+\alpha}} \end{aligned} \tag{10}$$

for all $n \ge n_1$ for some $n_1 = n_1(\alpha, s) \in \mathbb{N}$, where we have used that

$$1 - z^k \le -k(z-1) \tag{11}$$

for $z \in (0, 1)$. Using that $\mathrm{var}(\mathbf{1}_A) = \mathbb{P}(A)(1-\mathbb{P}(A))$ and $\mathrm{Cov}\,(\mathbf{1}_A, \mathbf{1}_B) = \mathbb{P}(A\cap B) - \mathbb{P}(A)\mathbb{P}(B) = \mathbb{P}(B)(\mathbb{P}(A \mid B) - \mathbb{P}(A))$, we see that

$$\begin{aligned}\mathrm{var}(Z_{n,s}) &= \sum_{j=1}^{n} \mathrm{var}\big[\mathbf{1}_{E_{n,j}(k_n)}\big] + \sum_{j=1}^{n}\sum_{\substack{i=1\\ i\neq j}}^{n} \mathrm{Cov}\,\big[\mathbf{1}_{E_{n,j}(k_n)}, \mathbf{1}_{E_{n,i}(k_n)}\big]\\ &= \sum_{j=1}^{n} \mathbb{P}(E_{n,j}(k_n))(1-\mathbb{P}(E_{n,j}(k_n)))\\ &\quad + \sum_{j=1}^{n}\sum_{\substack{i=1\\ i\neq j}}^{n} [\mathbb{P}(E_{n,i}(k_n))(\mathbb{P}(E_{n,j}(k_n) \mid E_{n,i}(k_n)) - \mathbb{P}(E_{n,j}(k_n)))]\\ &= \left[1-\left(1-\frac{s}{n}\right)^{k_n} + (n-1)\left(\left(1-\frac{s}{n-1}\right)^{k_n} - \left(1-\frac{s}{n}\right)^{k_n}\right)\right]\mathbb{E}[Z_{n,s}].\end{aligned}$$

Applying (10), we conclude that there exists a constant $c_1(x) \in (0, \infty)$ such that, for $n \ge n_1$, we have the lower variance bound

$$\mathrm{var}(Z_{n,s}) \ge \left(1 - \frac{\mathrm{e}^{-x}}{n^{\alpha}} - \frac{2\mathrm{e}^{-x}\,(\alpha \log(n) + x)}{n^{\alpha}}\right)\mathbb{E}[Z_{n,s}] \ge c_1(x)\,\mathbb{E}[Z_{n,s}]. \tag{12}$$

https://doi.org/10.1017/jpr.2023.6 Published online by Cambridge University Press

4.2. Bounding $\mathrm{var}\big(\mathbb{E}\big[Z_{n,s} - Z^I_{n,s}|_n(k_n)\big]\big)$

For the variance of the conditional expectation on the right-hand side in (7) we obtain

$$\begin{aligned}
&\mathrm{var}\big(\mathbb{E}\big[Z_{n,s} - Z^I_{n,s}|\mathcal{C}_n(k_n)\big]\big) = \mathrm{var}\left(\mathbb{E}\left[\sum_{j=1}^{n} \mathbf{1}_{E_{n,j}(k_n)\cap F^I_{n,j}} \mid \mathcal{C}_n(k_n)\right]\right)\\
&= \mathbb{E}\left[\left(\sum_{j=1}^{n} \mathbb{E}\Big[\mathbf{1}_{E_{n,j}(k_n)\cap F^I_{n,j}} \mid \mathcal{C}_n(k_n)\Big]\right)^2\right] - \mathbb{E}\left[\sum_{j=1}^{n} \mathbb{E}\Big[\mathbf{1}_{E_{n,j}(k_n)\cap F^I_{n,j}} \mid \mathcal{C}_n(k_n)\Big]\right]^2\\
&= \mathbb{E}\left[\sum_{j=1}^{n} \mathbb{E}\Big[\mathbf{1}_{E_{n,j}(k_n)\cap F^I_{n,j}} \mid \mathcal{C}_n(k_n)\Big]^2\right]\\
&+ \left(\mathbb{E}\left[\sum_{j=1}^{n}\sum_{\substack{i=1\\ i\neq j}}^{n} \mathbb{E}\Big[\mathbf{1}_{E_{n,j}(k_n)\cap F^I_{n,j}} \mid \mathcal{C}_n(k_n)\Big]\mathbb{E}\Big[\mathbf{1}_{E_{n,i}(k_n)\cap F^I_{n,i}} \mid \mathcal{C}_n(k_n)\Big]\right] - \left[\sum_{j=1}^{n} \mathbb{E}\Big[\mathbf{1}_{E_{n,j}(k_n)\cap F^I_{n,j}}\Big]\right]^2\right)\\
&=: T_1 + T_2. \qquad (13)
\end{aligned}$$

We start by dealing with T_1. Note that, conditionally on the event $E_{n,j}(k_n)$,

$$\mathbb{E}\Big[\mathbf{1}_{F^I_{n,j}} \mid \mathcal{C}_n(k_n)\Big] = 1 - \left(1 - \frac{1}{n-s}\right)^{|C_I|}, \qquad (14)$$

so that, with (11), we obtain

$$T_1 = \mathbb{E}\left[\sum_{j=1}^{n} \mathbf{1}_{E_{n,j}(k_n)}\left(1 - \left(1 - \frac{1}{n-s}\right)^{|C_I|}\right)^2\right] \leq \frac{1}{(n-s)^2}\mathbb{E}\left[\sum_{j=1}^{n} \mathbf{1}_{E_{n,j}(k_n)} \mid C_I|^2\right].$$

Denoting by D^I_m the event that a particle is placed into cell C_I in the mth drawing for some $m \in \{1, \ldots, k_n\}$, we see that

$$\begin{aligned}
\mathbb{E}\big[\mathbf{1}_{E_{n,j}(k_n)} \mid C_I|^2\big] &= \mathbb{E}\left[\mathbf{1}_{E_{n,j}(k_n)}\left(\sum_{m=1}^{k_n} \mathbf{1}_{D^I_m}\right)^2\right]\\
&= \mathbb{E}\left[\sum_{m=1}^{k_n} \mathbf{1}_{E_{n,j}(k_n)}\mathbf{1}_{D^I_m}\right] + \mathbb{E}\left[\sum_{m=1}^{k_n}\sum_{\substack{r=1\\ r\neq m}}^{k_n} \mathbf{1}_{E_{n,j}(k_n)}\mathbf{1}_{D^I_m}\mathbf{1}_{D^I_r}\right]\\
&= \mathbb{P}(E_{n,j}(k_n))\left(\sum_{m=1}^{k_n} \mathbb{P}\Big(D^I_m \mid E_{n,j}(k_n)\Big) + \sum_{m=1}^{k_n}\sum_{\substack{r=1\\ r\neq m}}^{k_n} \mathbb{P}\Big(D^I_m \cap D^I_r \mid E_{n,j}(k_n)\Big)\right).
\end{aligned}$$

Since all s-placements occur with the same probability, and conditioning on the event $E_{n,j}(k_n)$ simply means that we can only place particles into $n-1$ cells, we have $\mathbb{P}\big(D^I_m \mid E_{n,j}(k_n)\big) =$

https://doi.org/10.1017/jpr.2023.6 Published online by Cambridge University Press

$s/(n-1)$. As the drawings are independent of each other,

$$\sum_{m=1}^{k_n} \mathbb{P}\big(D_m^I \mid E_{n,j}(k_n)\big) + \sum_{m=1}^{k_n} \sum_{\substack{r=1\\ r\neq m}}^{k_n} \mathbb{P}\big(D_m^I \cap D_r^I \mid E_{n,j}(k_n)\big)$$
$$= \sum_{m=1}^{k_n} \mathbb{P}\big(D_m^I \mid E_{n,j}(k_n)\big) + \sum_{m=1}^{k_n} \sum_{\substack{r=1\\ r\neq m}}^{k_n} \mathbb{P}\big(D_m^I \mid E_{n,j}(k_n)\big)\mathbb{P}\big(D_r^I \mid E_{n,j}(k_n)\big) = \frac{k_n s}{n-1} + \frac{k_n(k_n-1)s^2}{(n-1)^2}.$$

Using (8), there exists a constant $c_2(x) \in (0, \infty)$ such that we can bound T_1 by

$$T_1 \leq \left(1-\frac{s}{n}\right)^{k_n} \frac{n}{(n-s)^2}\left(\frac{k_n s}{n-1} + \frac{k_n(k_n-1)s^2}{(n-1)^2}\right)$$
$$\leq 6(s+1)^2 \mathrm{e}^{-x} \frac{(\alpha \log n + x)^2}{n^{1+\alpha}} \leq c_2(x) \frac{s^2\alpha^2 \log (n)^2}{n^{1+\alpha}}. \tag{15}$$

To deal with the term T_2, first note that, by (4) and (5),

$$\mathbb{E}\Big[\mathbf{1}_{E_{n,j}(k_n)\cap F_{n,j}^I}\Big] = \mathbb{P}(E_{n,j}(k_n))\mathbb{P}\big(F_{n,j}^I \mid E_{n,j}(k_n)\big)$$
$$= \mathbb{P}(E_{n,j}(k_n)) \sum_{\ell=0}^{k_n} \mathbb{P}\big(F_{n,j}^I \mid E_{n,j}(k_n) \cap \{|C_I| = \ell\}\big)\mathbb{P}\big(|C_I| = \ell \mid E_{n,j}(k_n)\big)$$
$$= \left(1-\frac{s}{n}\right)^{k_n} \sum_{\ell=0}^{k_n} \left(1-\left(1-\frac{1}{n-s}\right)^{\ell}\right) \binom{k_n}{\ell} \left(\frac{s}{n-1}\right)^{\ell}\left(1-\frac{s}{n-1}\right)^{k_n-\ell}$$
$$= \left(1-\frac{s}{n}\right)^{k_n} \left(1-\sum_{\ell=0}^{k_n} \left(1-\frac{1}{n-s}\right)^{\ell} \binom{k_n}{\ell} \left(\frac{s}{n-1}\right)^{\ell}\left(1-\frac{s}{n-1}\right)^{k_n-\ell}\right)$$
$$= \left(1-\frac{s}{n}\right)^{k_n} \left(1-\left(1-\frac{s}{(n-s)(n-1)}\right)^{k_n}\right). \tag{16}$$

Using (14) and slightly adapting (5), the first part of T_2 can be handled in the following way:

$$\mathbb{E}\Big[\mathbb{E}\Big[\mathbf{1}_{E_{n,j}(k_n)\cap F_{n,j}^I} \mid \mathcal{C}_n(k_n)\Big]\,\mathbb{E}\Big[\mathbf{1}_{E_{n,i}(k_n)\cap F_{n,i}^I} \mid \mathcal{C}_n(k_n)\Big]\Big]$$
$$= \mathbb{E}\left[\mathbf{1}_{E_{n,i}(k_n)}\mathbf{1}_{E_{n,j}(k_n)}\left(1-\left(1-\frac{1}{n-s}\right)^{|C_I|}\right)^2\right]$$
$$= \sum_{\ell=0}^{k_n} \left(1-\left(1-\frac{1}{n-s}\right)^{\ell}\right)^2 \mathbb{P}(|C_I| = \ell \mid E_{n,i}(k_n) \cap E_{n,j}(k_n))\,\mathbb{P}(E_{n,i}(k_n) \cap E_{n,j}(k_n))$$
$$= \left(1-\frac{s}{n}\right)^{k_n}\left(1-\frac{s}{n-1}\right)^{k_n} \sum_{\ell=0}^{k_n} \left(1-\left(1-\frac{1}{n-s}\right)^{\ell}\right)^2 \binom{k_n}{\ell} \left(\frac{s}{n-2}\right)^{\ell}\left(1-\frac{s}{n-2}\right)^{k_n-\ell}$$

$$
\begin{aligned}
&= \left(1-\frac{s}{n}\right)^{k_n}\left(1-\frac{s}{n-1}\right)^{k_n} \\
&\quad \times \sum_{\ell=0}^{k_n}\left(1-2\left(1-\frac{1}{n-s}\right)^{\ell}+\left(1-\frac{1}{n-s}\right)^{2\ell}\right)\binom{k_n}{\ell}\left(\frac{s}{n-2}\right)^{\ell}\left(1-\frac{s}{n-2}\right)^{k_n-\ell} \\
&= \left(1-\frac{s}{n}\right)^{k_n}\left(1-\frac{s}{n-1}\right)^{k_n} \\
&\quad \times \left(1-2\left(1-\frac{s}{(n-2)(n-s)}\right)^{k_n}+\left(1-\frac{2s}{(n-2)(n-s)}+\frac{s}{(n-2)(n-s)^2}\right)^{k_n}\right),
\end{aligned}
$$

independently of i and j. Combining this with (16) yields

$$
\begin{aligned}
T_2 \le n^2\left(1-\frac{s}{n}\right)^{2k_n}\Bigg[&2\left(\left(1-\frac{s}{(n-s)(n-1)}\right)^{k_n}-\left(1-\frac{s}{(n-s)(n-2)}\right)^{k_n}\right) \\
&+\left(1-\frac{2s}{(n-s)(n-2)}+\frac{s}{(n-2)(n-s)^2}\right)^{k_n}-\left(1-\frac{s}{(n-s)(n-1)}\right)^{2k_n}\Bigg].
\end{aligned}
$$

Now, there exists $n_2=n_2(\alpha, s)$ such that, for all $n\ge n_2$,

$$
\begin{aligned}
\left(1-\frac{s}{(n-s)(n-1)}\right)^{k_n}-\left(1-\frac{s}{(n-s)(n-2)}\right)^{k_n} &\le 2\frac{\alpha\log(n)+x}{n^2}, \\
\left(1-\frac{2s}{(n-s)(n-2)}+\frac{s}{(n-2)(n-s)^2}\right)^{k_n}-\left(1-\frac{s}{(n-s)(n-1)}\right)^{2k_n} &\le 2\frac{\alpha\log(n)+x}{n^2}.
\end{aligned}
$$

Combining this with (8) we can conclude that there exists a constant $c_3(x)\in(0,\infty)$ such that, for all $n\ge \max\{3, n_2\}$,

$$
T_2 \le c_3(x)\frac{\alpha\log(n)}{n^{2\alpha}}. \tag{17}
$$

Putting the bounds in (15) and (17) into (13), we see that there exists a constant $c_4(x)\in(0,\infty)$ such that, for all $n\ge\max\{3, n_2\}$,

$$
\begin{aligned}
\operatorname{var}\big(\mathbb{E}\big[Z_{n,s}-Z^I_{n,s}\mid \mathcal{C}_n(k_n)\big]\big) &\le c_2(x)\frac{s^2\alpha^2\log(n)^2}{n^{1+\alpha}}+c_3(x)\frac{\alpha\log(n)}{n^{2\alpha}} \\
&\le c_4(x)\left(\frac{s^2\alpha^2\log(n)^2}{n^{1+\alpha}}+\frac{\alpha\log(n)}{n^{2\alpha}}\right).
\end{aligned} \tag{18}
$$

Remark 4. Using (13) and the exact expressions for the probabilities appearing there, it can be shown that for constant s the order of (18) is optimal, in the sense that we can find constants $c, C\in(0,\infty)$ only depending on s and x such that

$$
c\frac{\alpha\log(n)}{n^{2\alpha}} \le \operatorname{var}\big(\mathbb{E}\big[Z_{n,s}-Z^I_{n,s}\mid \mathcal{C}_n(k_n)\big]\big) \le C\frac{\alpha\log(n)}{n^{2\alpha}}
$$

for sufficiently large n.

https://doi.org/10.1017/jpr.2023.6 Published online by Cambridge University Press

4.3. Bounding $\mathbb{E}\big[\big(Z_{n,s}-Z^I_{n,s}\big)^2\big]$

For the second term in (7) it remains to bound $\mathbb{E}\big[\big(Z_{n,s}-Z^I_{n,s}\big)^2\big]$. We have

$$\begin{aligned}\mathbb{E}\big[\big(Z_{n,s}-Z^I_{n,s}\big)^2\big] &\le \mathbb{E}\Bigg[\Bigg(\sum_{\substack{j=1\\ j\neq I}}^{n}\mathbf{1}_{E_{n,j}(k_n)\cap F^I_{n,j}}-1\Bigg)^2\Bigg]\\ &=\mathbb{E}\Bigg[\Bigg(\sum_{\substack{j=1\\ j\neq I}}^{n}\mathbf{1}_{E_{n,j}(k_n)\cap F^I_{n,j}}\Bigg)^2\Bigg]-2\mathbb{E}\Bigg[\sum_{\substack{j=1\\ j\neq I}}^{n}\mathbf{1}_{E_{n,j}(k_n)\cap F_{n,j}}\Bigg]+1\\ &\le \mathbb{E}\Bigg[\sum_{\substack{j=1\\ j\neq I}}^{n}\sum_{\substack{i=1\\ i\neq j,I}}^{n}\mathbf{1}_{E_{n,j}(k_n)\cap F^I_{n,j}}\mathbf{1}_{E_{n,i}(k_n)\cap F^I_{n,i}}\Bigg]+1, \end{aligned}\tag{19}$$

where the -1 in the first line comes from the artificially isolated cell with index I. For the first sum, note that

$$\mathbb{P}\big(E_{n,i}(k_n)\cap F^I_{n,i}\cap E_{n,j}(k_n)\cap F^I_{n,j}\big)=\mathbb{P}\big(F^I_{n,i}(k_n)\cap F^I_{n,j}\mid E_{n,j}(k_n)\cap E_{n,i}\big)\mathbb{P}\big(E_{n,j}(k_n)\cap E_{n,j}\big),$$

where $\mathbb{P}(E_{n,j}(k_n)\cap E_{n,j}(k_n))$ can be bounded using (6) and (8). To bound the remaining probability, we observe that, similarly to the considerations in (16), we have

$$\begin{aligned}&\mathbb{P}\big(F^I_{n,j}\mid E_{n,i}(k_n)\cap E_{n,j}(k_n)\big)\\ &\quad=\sum_{\ell=0}^{k_n}\mathbb{P}\big(F^I_{n,j}\mid E_{n,i}(k_n)\cap E_{n,j}(k_n)\cap\{|C_I|=\ell\}\big)\mathbb{P}\big(|C_I|=\ell\mid E_{n,i}(k_n)\cap E_{n,j}(k_n)\big)\\ &\quad=\sum_{\ell=0}^{k_n}\binom{k_n}{\ell}\left(1-\left(1-\frac{1}{n-s}\right)^{\ell}\right)\left(\frac{s}{n-2}\right)^{\ell}\left(1-\frac{s}{n-2}\right)^{k_n-\ell}\\ &\quad=1-\left(1-\frac{s}{(n-2)(n-s)}\right)^{k_n},\end{aligned}$$

independently of the choice of i and j. So, we are left to deal with $\mathbb{P}\big(F^I_{n,i}\mid E_{n,i}(k_n)\cap E_{n,j}(k_n)\cap F^I_{n,j}\big)$. For $i\neq j$,

$$\mathbb{P}\big(F^I_{n,i}\mid E_{n,j}(k_n)\cap E_{n,i}(k_n)\cap F^I_{n,j}\big)\le\mathbb{P}\big(F^I_{n,i}\mid E_{n,i}(k_n)\cap E_{n,j}(k_n)\big),$$

since the event $F^I_{n,j}$ implies that at least one particle from cell I is placed into cell C_j, reducing the chances of cell C_i receiving a particle. Hence,

$$\mathbb{P}\big(F^I_{n,i}\cap F^I_{n,j}\mid E_{n,i}(k_n)\cap E_{n,j}(k_n)\big)\le\left(1-\left(1-\frac{s}{(n-2)(n-s)}\right)^{k_n}\right)^2,$$

https://doi.org/10.1017/jpr.2023.6 Published online by Cambridge University Press

and we finally obtain, for $n \geq s$,

$$\begin{aligned}
\mathbb{E}\Bigg[\sum_{\substack{j=1\\ j\neq I}}^{n}\sum_{\substack{i=1\\ i\neq j,I}}^{n}\mathbf{1}_{E_{n,j}(k_n)\cap F^I_{n,j}}\mathbf{1}_{E_{n,i}(k_n)\cap F^I_{n,i}}\Bigg]
&= \sum_{\substack{j=1\\ j\neq I}}^{n}\sum_{\substack{i=1\\ i\neq j,I}}^{n}\mathbb{P}\big(F^I_{n,i}\cap F^I_{n,j} \mid E_{n,i}(k_n)\cap E_{n,j}(k_n)\big)\mathbb{P}(E_{n,i}(k_n)\cap E_{n,j}(k_n))\\
&\leq (n-1)(n-2)\left(1-\left(1-\frac{s}{(n-2)(n-s)}\right)^{k_n}\right)^2\left(1-\frac{s}{n}\right)^{2k_n}\\
&\leq (n-1)(n-2)\left(k_n\frac{s}{(n-2)(n-s)}\right)^2\left(1-\frac{s}{n}\right)^{2k_n}\\
&\leq (n-1)(n-2)\left(\frac{3(s+1)(\alpha\log(n)+x)}{n}\right)^2\left(\frac{\mathrm{e}^{-x}\alpha\log(n)}{n^{\alpha}}\right)^2\\
&\leq 9(s+1)^2\mathrm{e}^{-2x}\frac{(\alpha\log(n)+x)^2\alpha^2\log(n)^2}{n^{2\alpha}},
\end{aligned}$$

where we used (11) to arrive at the third line. Plugging this back into (19), we conclude that there exists a constant $c_5(x)\in(0,\infty)$ such that, for all $n\geq s$,

$$\mathbb{E}\big[\big(Z_{n,s}-Z^I_{n,s}\big)^2\big]\leq c_5(x)s^2. \tag{20}$$

Combining the normal approximation bound (7) with the estimates (9), (12), (18), and (20), we arrive at

$$\begin{aligned}
d_{\mathrm{W}}\big(\tilde{Z}_{n,s}, G\big) &\leq \frac{\lambda_n}{c_1(x)\lambda_n}\sqrt{c_4(x)\left(\frac{s^2\alpha^2\log(n)^2}{n^{1+\alpha}}+\frac{\alpha\log(n)}{n^{2\alpha}}\right)}+\frac{\lambda_n}{(c_1(x)\lambda_n)^{3/2}}c_5(x)s^2\\
&\leq \frac{\sqrt{c_4(x)}\,\alpha}{c_1(x)}\frac{s\log n}{n^{(1+\alpha)/2}}+\frac{\sqrt{c_4(x)}\sqrt{\alpha}}{c_1(x)}\frac{\sqrt{\log(n)}}{n^{\alpha}}+\frac{c_5(x)}{c_1(x)^{3/2}}\frac{\sqrt{2}s^2}{n^{(1-\alpha)/2}\mathrm{e}^{-x/2}}\\
&\leq C\left(\frac{\sqrt{\log(n)}}{n^{\alpha}}+\frac{s^2}{n^{(1-\alpha)/2}}\right)
\end{aligned}$$

for some constant $C=C(x,\alpha)\in(0,\infty)$ and for all $n\geq N:=\max\{n_1, n_2, 3, s, \mathrm{e}^{-x}\}$. Here, we used that $\mathbb{E}[Z_{n,s}]\geq\frac{1}{2}\mathrm{e}^{-x}n^{1-\alpha}$ for $n\geq 2$ in the second step. Note that in the last step we also used that

$$\frac{s\log n}{n^{(1+\alpha)/2}}\leq\frac{s^2}{n^{(1-\alpha)/2}},$$

which is equivalent to $\log n\leq sn^{\alpha}$ and thus automatically satisfied. This proves Theorem 1. □

Remark 5. The constants C and N in Theorem 1 can be made fully explicit in terms of the parameters x and α. To see this, we start by noting that in (8) the asymptotic equivalence '$\sim$'

https://doi.org/10.1017/jpr.2023.6 Published online by Cambridge University Press

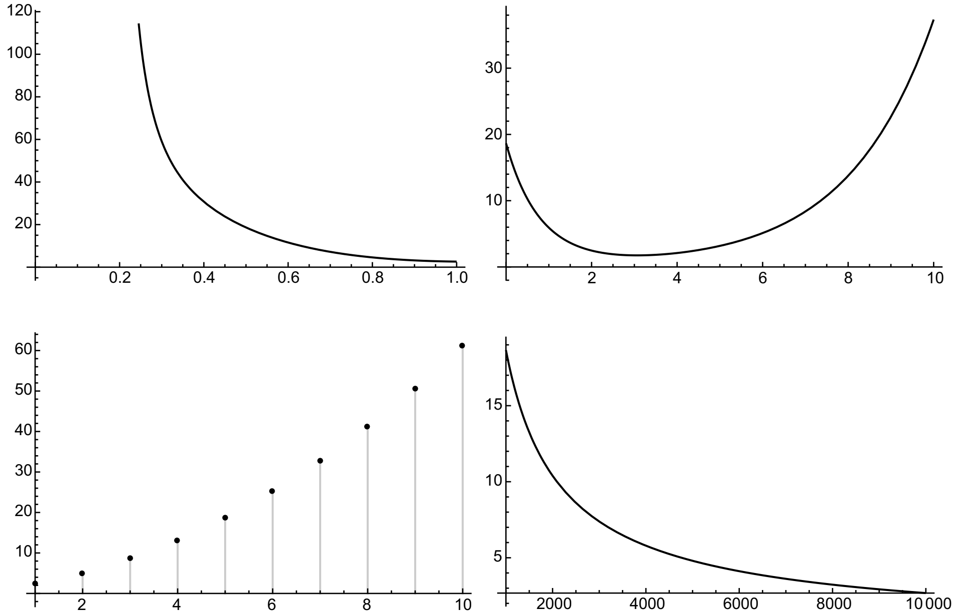

FIGURE 3. Dependence of the normal approximation bound on the parameters α, x, s, and n. Top left: $\alpha \in \left[\frac{1}{3}, 1\right)$, while $x = 0$, $s = 5$, $n = 1000$. Top right: $x \in [0, 10]$, while $\alpha = \frac{1}{2}$, $s = 5$, $n = 1000$. Bottom left: $s \in \{1, \dots, 10\}$, while $\alpha = \frac{1}{2}$, $x = 0$, $n = 1000$. Bottom right: $n \in \{1000, \dots, 10\,000\}$, while $\alpha = \frac{1}{2}$, $x = 0$, $s = 5$.

can be replaced by '$\leq$', and that $\mathbb{E}[Z_{n,s}] \geq \frac{1}{2}\mathrm{e}^{-x} n^{1-\alpha}$ for $n \geq 2$. With this inequality we can conclude from the computations in Section 4.1 that

$$\frac{\lambda_n}{\sigma_n^2} \leq \left(1 - \frac{\mathrm{e}^{-x}}{n^\alpha} - \frac{2\mathrm{e}^{-x}(\alpha \log(n) + x)}{n^\alpha}\right)^{-1},$$
$$\frac{\lambda_n}{\sigma_n^3} \leq \left(1 - \frac{e^{-x}}{n^\alpha} - \frac{2e^{-x}(\alpha \log(n) + x)}{n^\alpha}\right)^{-3/2} \left(\frac{1}{2}\mathrm{e}^{-x} n^{1-\alpha}\right)^{-1/2}.$$

Furthermore, an inspection of Sections 4.2 and 4.3 shows that

$$\operatorname{var}\big(\mathbb{E}\big[Z_{n,s} - Z_{n,s}^I \mid \mathcal{C}_n(k_n)\big]\big) \leq 6(s+1)^2 \mathrm{e}^{-x} \frac{(\alpha \log(n) + x)^2}{n^{1+\alpha}} + 4\mathrm{e}^{-2x} \frac{\alpha \log(n) + x}{n^{2\alpha}},$$
$$\mathbb{E}\big[\big(Z_{n,s} - Z_{n,s}^I\big)^2\big] \leq 9(s+1)^2 \mathrm{e}^{-2x} \frac{(\alpha \log(n) + x)^2 \alpha^2 \log(n)^2}{n^{2\alpha}} + 1,$$

which in view of (7) leads to a fully explicit error bound (which is valid whenever the resulting expression is positive). In particular, the dependence of this bound on the parameters s, x, α, and n can be studied as demonstrated in Fig. 3 (note that in contrast to the Kolmogorov distance, which is bounded by 1, the Wasserstein distance can take arbitrarily large values).

5. Proof of Theorem 2

As already mentioned in the introduction, rates of convergence towards a Poisson limit in the case $\alpha = 1$ can be concluded from [2, Theorem 6.F]. Nevertheless, we give an alternative self-contained proof using the coupling from Section 3.

Proof of Theorem 2. It follows from [15, Theorem 4.13] that the total variation distance between the number of empty cells in the coupon collector's problem and a Poisson random variable W with parameter $\lambda_n = \mathbb{E}[Z_{n,s}(k_n)]$ can be bounded using the following inequality:

$$d_{\mathrm{TV}}(Z_{n,s}(k_n),\, W) \le \min\{1,\, \lambda_n\}\, \mathbb{E}\big[Z_{n,s}(k_n) + 1 - Z^I_{n,s}(k_n)\big]. \tag{21}$$

For the expectation on the right, the definitions of $Z_{n,s}(k_n)$ and $Z^I_{n,s}(k_n)$ yield

$$\begin{aligned}\mathbb{E}\big[Z_{n,s}(k_n) + 1 - Z^I_{n,s}(k_n)\big] &= \mathbb{E}[\mathbf{1}_{E_{n,I}(k_n)}] + \mathbb{E}\Bigg[\sum_{j\neq I} \mathbf{1}_{E_{n,j}(k_n)} - \mathbf{1}_{E^I_{n,j}(k_n)}\Bigg] \\ &= \left(1-\frac{s}{n}\right)^k + (n-1)\left(\left(1-\frac{s}{n}\right)^k - \left(1-\frac{s}{n-1}\right)^k\right).\end{aligned}$$

Combining (8) and (10) for $\alpha = 1$ with (21) yields

$$\begin{aligned}d_{\mathrm{TV}}(Z_{n,s}(k_n),\, W) &\le \min\{1,\, \lambda_n\}\, \mathbb{E}\big[Z_{n,s}(k_n) + 1 - Z^I_{n,s}(k_n)\big] \\ &= \left(1-\frac{s}{n}\right)^k + (n-1)\left(\left(1-\frac{s}{n}\right)^k - \left(1-\frac{s}{n-1}\right)^k\right) \\ &\le \frac{\mathrm{e}^{-x}}{n}(\log(n) + x + 1)\end{aligned}$$

for $n \ge \tilde{N} := \max\left\{n_1,\, \mathrm{e}^{-(x+1)}\right\}$, which completes the proof of Theorem 2. □

Acknowledgements

The authors would like to thank Chinmoy Bhattacharjee for pointing us to the paper of Englund [5]. Moreover, we would like to thank two anonymous referees for helping us to improve our text.

Funding information

This work has been supported by the DFG priority program SPP 2265 *Random Geometric Systems*.

Competing interests

There were no competing interests to declare which arose during the preparation or publication process of this article.

References

[1] ARRATIA, R. GOLDSTEIN, L. AND KOCHMAN, F. (2019). Size bias for one and all. *Prob. Surv.* **16**, 1–61.
[2] BARBOUR, A. D., HOLST, L. AND JANSON, S. (1992). *Poisson Approximation*. Oxford University Press.
[3] BAUM, L. E. AND BILLINGSLEY, P. (1965). Asymptotic distributions for the coupon collector's problem. *Ann. Math. Statist.* **36**, 1835–1839.

[4] DE MOIVRE, A. (1718). The Doctrine of Chances. W. Pearson, London.
[5] ENGLUND, G. (1981). A remainder term estimate for the normal approximation in classical occupancy. *Ann. Prob.* **9**, 684–692.
[6] EULER, L. (1785). Solutio quarundam quaestionum difficiliorum in calculo probabilium. *Opuscula Analytica*, Vol. **2**, 331–346.
[7] GOLDSTEIN, L. AND RINOTT, Y. (1996). Multivariate normal approximations by Stein's method and size bias couplings. *J. Appl. Prob.* **33**, 1–17.
[8] JOHNSON, N. L. AND KOTZ, S. (1977). *Urn Models and Their Application: An Approach to Modern Discrete Probability Theory*. Wiley, Chichester.
[9] KOLCHIN, V. F., SEVASTIANOV, B. A. AND CHISTIAKOV, V. P. (1978). *Random Allocations*. Halsted Press, Washington.
[10] LAPLACE, P. S. (1812). *Théorie Analytique des Probabilités*. Courcier, Paris.
[11] MAHMOUD, H. (2010). Gaussian phases in generalized coupon collection. *Adv. Appl. Prob.* **42**, 994–1012.
[12] MIKHAILOV, V. G. (1977). A Poisson limit theorem in the scheme of group disposal of particels. *Theory Prob. Appl.* **22**, 152–156.
[13] MIKHAILOV, V. G. (1980). Asymptotic normality of the number of empty cells for group allocation of particles. *Theory Prob. Appl.* **25**, 82–90.
[14] NEAL, P. (2008). The generalised coupon collector problem. *J. Appl. Prob.* **45**, 621–629.
[15] ROSS, N. (2011). Fundamentals of Stein's method. *Prob. Surv.* **8**, 210–293.
[16] SCHILLING, J. AND HENZE, N. (2021). Two Poisson limit theorems for the coupon collector's problem with group drawings. *J. Appl. Prob.* **58**, 966–977.
[17] SMYTHE, R. T. (2011). Generalized coupon collection: The superlinear case. *J. Appl. Prob.* **48**, 189–199.
[18] STADJE, W. (1990). The collector's problem with group drawings. *Adv. Appl. Prob.* **22**, 866–882.
[19] VATUTIN, V. A. AND MIKHAILOV, V. G. (1983). Limit theorems for the number of empty cells in an equiprobable scheme for group allocation of particles. *Theory Prob. Appl.* **27**, 734–743.