ALEXEY CHUPRUNOV and ISTVÁN FAZEKAS[1]

On the number of empty cells
in the allocation scheme
of indistinguishable particles

ABSTRACT. The allocation scheme of n indistinguishable particles into N different cells is studied. Let the random variable $\mu_0(n, K, N)$ be the number of empty cells among the first K cells. Let $p = \frac{n}{n+N}$. It is proved that $\frac{\mu_0(n,K,N)-K(1-p)}{\sqrt{Kp(1-p)}}$ converges in distribution to the Gaussian distribution with expectation zero and variance one, when $n, K, N \to \infty$ such that $\frac{n}{N} \to \infty$ and $\frac{n}{NK} \to 0$. If $n, K, N \to \infty$ so that $\frac{n}{N} \to \infty$ and $\frac{NK}{n} \to \lambda$, where $0 < \lambda < \infty$, then $\mu_0(n, K, N)$ converges in distribution to the Poisson distribution with parameter λ. Two applications of the results are given to mathematical statistics. First, a method is offered to test the value of n. Then, an analogue of the run-test is suggested with an application in signal processing.

1. Introduction and main results. The de Moivre–Laplace theorem and the Poisson limit theorem are widely known classical results in probability theory. For discrete probability models there are many other theorems for normal and Poisson approximation, see [7, 8] and [1]. In this paper, we

---

offer new normal and Poisson approximation theorems for allocations of indistinguishable particles into different cells.

Let n and N be integer numbers. The allocation scheme of n indistinguishable particles into N different cells can be described by the random variables $\eta_1, \ldots, \eta_N$ with joint distribution

$$(1.1) \qquad \mathbb{P}\{\eta_1 = k_1, \ldots, \eta_N = k_N\} = \frac{1}{\binom{n+N-1}{N-1}},$$

where $k_1, k_2, \ldots, k_N$ are non-negative integer numbers such that $k_1 + k_2 + \cdots + k_N = n$.

Let K be an integer number such that $0 < K \leq N$. Let r be a non-negative integer number. We will use the notation

$$\mu_r(n, K, N) = \sum_{i=1}^{K} \mathbb{I}_{\{\eta_i = r\}},$$

where $\mathbb{I}_A$ denotes the indicator of the set A. The random variable $\mu_r(n, K, N)$ is the number of cells among the first K cells containing precisely r particles.

Let us introduce the notation which we need in our results and will be used throughout the paper. $\xrightarrow{d}$ denotes convergence in distribution, γ is a Gaussian random variable with expectation zero and variance one, and Φ is the distribution function of γ. $\mathrm{o}(1)$ denotes a quantity converging to 0, and $\mathrm{O}(1)$ denotes a bounded quantity.

The main results of this paper are the following theorems. First we consider the asymptotic normality of the number of empty cells. We start with a local limit theorem.

Theorem 1.1. *Suppose that $n, K, N \to \infty$ such that $\frac{n}{N} \to \infty$ and $\frac{n}{NK} \to 0$. Let $p = \frac{n}{n+N}$ and $z = \frac{k - K(1-p)}{\sqrt{Kp(1-p)}}$. Then we have*

$$(1.2) \qquad \mathbb{P}(\mu_0(n, K, N) = k) = \frac{1}{\sqrt{2\pi K p(1-p)}} e^{-\frac{z^2}{2}} (1 + \mathrm{o}(1))$$

uniformly for $|z| < C$, where C is an arbitrary fixed positive number.

The next global limit theorem follows from Theorem 1.1.

Corollary 1.1. *Suppose that $n, K, N \to \infty$ such that $\frac{n}{N} \to \infty$ and $\frac{n}{NK} \to 0$. Let $p = \frac{n}{n+N}$. Then we have*

$$(1.3) \qquad \mathbb{P}\left(\frac{\mu_0(n, K, N) - K(1-p)}{\sqrt{Kp(1-p)}} < t\right) \to \Phi(t), \quad t \in \mathbb{R}.$$

The following theorem contains the Poisson limit for the number of empty cells:

Theorem 1.2. *Suppose that $n, K, N \to \infty$ such that $\frac{n}{N} \to \infty$ and $\frac{NK}{n} \to \lambda$, where $0 \le \lambda < \infty$. Then we have*

$$(1.4) \qquad \mathbb{P}(\mu_0(n, K, N) = k) = e^{-\lambda} \frac{\lambda^k}{k!}(1 + \mathrm{o}(1)), \quad k = 0, 1 \dots.$$

We mention that Theorem 1.1 of our paper contains a new result also in the case $K = N$. Here we list some known results. For $K = N$, the random variable $\mu_r(n, K, N)$ is denoted by $\mu_r(n, N)$. In Trunov [11] and in Timashev [10], limit theorems are proved for $\mu_r(n, N)$ in the allocation scheme of indistinguishable particles into different cells. In Theorem 2.1 of [11], a Gaussian limit theorem is proved for $\mu_0(n, N)$ in the case $C_1 < p < C_2$ for some $0 < C_1 < C_2 < 1$. In Theorem 2.3 of [11], a Gaussian limit theorem is proved for $\mu_0(n, N)$ in the case $p \to 0$. However, Theorem 1.1 of our paper concerns another case, that is the case of $p \to 1$.

In Chuprunov and Fazekas [2], Poisson limit theorems are proved for $\mu_r(n, K, N)$ in the allocation scheme of indistinguishable particles into different cells and also for other schemes of discrete probability theory. Concerning the case of distinguishable particles we mention the following. In Khakimullin and Enatskaia [5], limit theorems are obtained for $\mu_0(n, K, N)$ in the allocation scheme of distinguishable particles into different cells. Many papers deal with limit theorems for $\mu_r(n, N)$ in the allocation scheme of distinguishable particles into different cells, see Kolchin, Sevast'yanov and Chistyakov [8] and the references therein.

The method of the proofs. During the proofs we shall need the notion of the generalized allocation scheme introduced by V. F. Kolchin in [6]. Let $\xi_1, \xi_2, \dots, \xi_N$ be independent identically distributed integer valued random variables. The random variables $\eta_1, \dots, \eta_N$ are called a generalized allocation scheme of n particles into N cells if their joint distribution has the form

$$\mathbb{P}\{\eta_1 = k_1, \dots, \eta_N = k_N\} = \mathbb{P}\left\{\xi_1 = k_1, \dots, \xi_N = k_N \;\middle|\; \sum_{i=1}^N \xi_i = n\right\},$$

where $k_1, k_2, \dots, k_N$ are non-negative integer numbers such that $k_1 + k_2 + \cdots + k_N = n$.

Various models of discrete probability theory such as random forests, random permutations, random allocations, urn schemes are particular cases of the generalized allocation scheme. If $\xi_1, \xi_2, \dots, \xi_N$ are independent identically distributed geometrically distributed random variables with parameter $0 < p < 1$, then the generalized allocation scheme is an allocation scheme of n indistinguishable particles into N different cells [7].

In the proofs we will use the following formula which is true in any generalized allocation scheme. Let $p_r = \mathbb{P}\{\xi_i = r\}$, $r = 0, 1, \ldots$. Then we have

$$(1.5) \quad \mathbb{P}\{\mu_r(n, K, N) = k\} = \binom{K}{k}(p_r)^k(1 - p_r)^{K-k}\frac{\mathbb{P}\{\zeta_{N-k}^{\{r\}} = n - kr\}}{\mathbb{P}\{\zeta_N = n\}},$$

where $\zeta_{N-k}^{\{r\}} = \xi_1^{\{r\}} + \cdots + \xi_{K-k}^{\{r\}} + \xi_{K+1} + \cdots + \xi_N$, $\zeta_N = \xi_1 + \xi_2 + \cdots + \xi_N$, the random variables $\xi_1^{\{r\}}, \ldots, \xi_{K-k}^{\{r\}}, \xi_{K+1}, \ldots, \xi_N$ are independent, and the random variables $\xi_1^{\{r\}}, \xi_2^{\{r\}}, \ldots$ have distribution

$$(1.6) \qquad \mathbb{P}\{\xi_i^{\{r\}} = j\} = \mathbb{P}\{\xi_i = j \mid \xi_i \neq r\}, \quad j = 0, 1, 2 \ldots.$$

We mention that formula (1.5) was obtained for $K = N$ and for allocation scheme of distinguishable particles into different cells in Lemma 1 on p. 50 of [11]. Its generalization for the case of any generalized allocation scheme is given in Lemma 1.2.1 of [7]). The proof of (1.5) is similar to the proof of the lemmas mentioned above.

The proofs of our theorems are based on approximations of the expressions in (1.5). In order to estimate the binomial probability in (1.5), we use a certain version of the de Moivre–Laplace theorem in the case of Theorem 1.1 and the Poisson approximation theorem in the case of Theorem 1.2. The main difficulty during the proofs is to find proper approximations for the expressions $\mathbb{P}\{\zeta_N = n\}$ and $\mathbb{P}\{\zeta_N^{\{0\}} = n\}$ in the fractional in (1.5). In order to handle these expressions we used new local limit theorems.

We will apply equation (1.5) in the case when $\xi_1, \xi_2, \ldots, \xi_N$ are independent identically distributed random variables having geometric distribution with parameter p. So let

$$(1.7) \qquad p_k = \mathbb{P}(\xi_i = k) = (1 - p)p^k, \quad k = 0, 1, \ldots,$$

be the distribution of ξ_i, $e(p) = \mathbb{E}\xi_i$ the expectation of ξ_i, and $\sigma^2(p) = \mathbb{D}^2\xi_i$ the variance of ξ_i. Moreover, $e_0(p) = \mathbb{E}\xi_i^{\{0\}}$ is the expectation of $\xi_i^{\{0\}}$ and $\sigma_0^2(p) = \mathbb{D}^2\xi_i^{\{0\}}$ is the variance of $\xi_i^{\{0\}}$.

2. Applications.

Application to mathematical statistics. We will use Corollary 1.1 to study some analogue of the empty box test. Let us consider the allocation scheme of n indistinguishable particles into N different cells such that N is a known number but n is unknown. We want to check the hypothesis

$$H_0 : n = n_0$$

against the alternative hypothesis

$$H_1 : n = n_1,$$

where $n_0 < n_1$.

Our criterion is the following. Let k be the number of cells from the first K cells which are empty. Fix the level $0 < \alpha < 1$ and choose the critical value u_α such that $\mathbb{P}(\gamma > u_\alpha) = 1 - \alpha$. Let

$$\tilde{p}_0 = \frac{n_0}{N + n_0}, \quad \tilde{p}_1 = \frac{n_1}{N + n_1},$$

$$C_0 = K(1 - \tilde{p}_0) + u_\alpha \sqrt{K\tilde{p}_0(1 - \tilde{p}_0)}.$$

As $n_0 < n_1$, so $\tilde{p}_0 < \tilde{p}_1$.

Hypothesis H_0 is accepted if $k \geq C_0$ and hypothesis H_1 is accepted if $k < C_0$. As H_0 is rejected if

$$\frac{k - K(1 - \tilde{p}_0)}{\sqrt{K\tilde{p}_0(1 - \tilde{p}_0)}} < u_\alpha,$$

therefore the probability of type I error is

$$\alpha(n_0, K, N) = \mathbb{P}\left(\frac{\mu_0(n, K, N) - K(1 - \tilde{p}_0)}{\sqrt{K\tilde{p}_0(1 - \tilde{p}_0)}} < u_\alpha\right).$$

So, by Corollary 1.1, as $n_0, K, N \to \infty$ such that $\frac{n_0}{N} \to \infty$, $\frac{n_0}{NK} \to 0$, then we have

$$\alpha(n_0, K, N) \to \alpha.$$

The probability of the type II error is

$$\beta(n_1, K, N) = \mathbb{P}\left(\frac{\mu_0(n, K, N) - K(1 - \tilde{p}_0)}{\sqrt{K\tilde{p}_0(1 - \tilde{p}_0)}} \geq u_\alpha\right),$$

but the probability should be calculated when hypothesis H_1 is true. By short calculation we can see that the event

$$\frac{\mu_0(n, K, N) - K(1 - \tilde{p}_0)}{\sqrt{K\tilde{p}_0(1 - \tilde{p}_0)}} \geq u_\alpha$$

is the same as

$$\frac{\mu_0(n, K, N) - K(1 - \tilde{p}_1)}{\sqrt{K\tilde{p}_1(1 - \tilde{p}_1)}} \geq \sqrt{\frac{KN}{n_1} \frac{n_1 - n_0}{n_0 + N}} - u_\alpha \sqrt{\frac{n_0}{n_1} \frac{n_1 + N}{n_0 + N}}.$$

Consider the right hand side of this inequality. We can assume that $\alpha < 0.5$, so $u_\alpha < 0$, therefore the second term is positive. The first term converges to ∞, if $\frac{KN}{n_1} \to \infty$ and $\frac{n_1}{n_0} \geq c_0 > 1$. Using Corollary 1.1, we can see that the left hand side of the above equality is asymptotically standard normal if $n_1, K, N \to \infty$ such that $\frac{n_1}{N} \to \infty$ and $\frac{n_1}{NK} \to 0$. Therefore the type II error $\beta(n_1, K, N)$ converges to 0, if $n_1, K, N \to \infty$ such that $\frac{n_1}{N} \to \infty$, $\frac{n_1}{NK} \to 0$ and $\frac{n_1}{n_0} \geq c_0 > 1$.

Application for runs. Runs play an important role in probability and statistics. There are well-known limit theorems for runs (see, e.g. [4]), moreover the Wald–Wolfowitz runs test is a famous non-parametric statistical test that checks a randomness hypothesis (see, e.g. [3]).

A usual way to imagine the allocation of n indistinguishable balls into a row of N distinct boxes is the following. Consider $n + N - 1$ digits, n of them are zeroes (the balls), $N - 1$ of them are ones (the barriers between two subsequent boxes). Fix an arrangement of the n zeroes (i.e. the balls), then insert the $N - 1$ ones (i.e. the barriers) amongst the zeroes. It can be made in $\binom{n+N-1}{N-1}$ different ways. Then the number of balls in the first box will be the number of zeroes before the first 1, the number of balls in the second box will be the number of zeroes between the first and the second 1 digits, etc. More precisely, let $\delta = (\delta_1, \delta_2, \ldots, \delta_{n+N-1})$ be a sequence of n zero digits and $N - 1$ one digits. Let the probability of any δ be

$$(2.1) \qquad \mathbb{P}(\delta) = \frac{1}{\binom{n+N-1}{N-1}}.$$

Let

$$\nu = (\nu_1, \ldots, \nu_N),$$

be a vector with $\nu_N = N + n$ and ν_i be the serial number of the ith digit 1 in the sequence δ, $1 \leq i \leq N - 1$. The vector $\eta = (\eta_1, \ldots, \eta_N)$, with coordinates

$$\eta_i = \nu_i - 1, \quad \text{for} \ \ i = 1, \quad \eta_i = \nu_i - \nu_{i-1} - 1 \ \ \text{for} \ \ 2 \leq i \leq N$$

is our previously defined allocation scheme of n indistinguishable particles into N different cells.

In a sequence $\delta = (\delta_1, \delta_2, \ldots, \delta_{n+N-1})$, a part of consecutive zeroes bordered by digits 1 is called a zero-run. Now, let $0 < K \leq N$. Let ξ be the number of zero-runs before the Kth 1 digit. Then $K - \xi = \mu_0(n, K, N)$. Therefore, from Corollary 1.1 and Theorem 1.2 we obtain the following corollaries.

Corollary 2.1. *Let $n, K, N \to \infty$ such that $\frac{n}{N} \to \infty$ and $\frac{n}{NK} \to 0$. Then we have*

$$\mathbb{P}\left(\frac{Kp - \xi}{\sqrt{Kp(1-p)}} < t\right) \to \Phi(t), \quad t \in \mathbb{R}.$$

Corollary 2.2. *Suppose that $n, K, N \to \infty$ such that $\frac{n}{N} \to \infty$ and $\frac{NK}{n} \to \lambda$, where $0 \leq \lambda < \infty$. Then we have*

$$\mathbb{P}(K - \xi = k) = e^{-\lambda}\frac{\lambda^k}{k!}(1 + o(1)), \quad k = 0, 1 \ldots.$$

Now we turn to a simple application to image processing. Consider a digitalized black and white image of size $S \times T$ which contains a black

"curve" on the white background. The black pixels are coded by 1 and the white pixels are coded by 0. The above mentioned "curve" can be a non-random signal (say a letter or a number which can be hand written or typed), but it can be a random signal, too. We want to exclude the random signals, so we shall check randomness. Therefore, let δ be the row-major order of the $S \times T$ image. δ contains a lot of zeroes and only a few ones. So $\frac{n}{N}$ is large, therefore our main assumption is satisfied.

Using Corollary 2.1, we can check the hypothesis

$$H_0 : \text{the sequence of the ones and zeroes is random.}$$

Fix $0 < \alpha < 1$ and let v_α be a number so that $\mathbb{P}(|\gamma| < v_\alpha) = 1 - \alpha$. Then we accept H_0 if

$$\left| \frac{Kp - \xi}{\sqrt{Kp(1-p)}} \right| < v_\alpha.$$

Now, the probability of the type I error is

$$\alpha(n, K, N) = \mathbb{P}\left(\left| \frac{Kp - \xi}{\sqrt{Kp(1-p)}} \right| < v_\alpha \right),$$

where the probability is calculated assuming model (2.1). So, by Corollary 2.1, as $n, K, N \to \infty$ such that $\frac{n}{N} \to \infty$, $\frac{n}{NK} \to 0$, we have

$$\alpha(n, K, N) \to \alpha.$$

3. Auxiliary results and proofs. We will use the following version of the well-known Berry–Esséen inequality.

Lemma 3.1. *Suppose that ξ_i', $1 \leq i \leq N$, are independent random variables, $\sigma_i^2 = \mathbb{D}^2 \xi_i'$ is the variance of ξ_i', $1 \leq i \leq N$, and $\sigma^2 = \sum_{i=1}^{N} \sigma_i^2$. Then we have*

$$\sup_{t \in \mathbb{R}} \left| \mathbb{P}\left(\frac{1}{\sigma} \sum_{i=1}^{N} (\xi_i' - \mathbb{E}\xi_i') < t \right) - \Phi(t) \right| < 2c \left(\frac{\sum_{i=1}^{N} \mathbb{E}(\xi_i' - \mathbb{E}\xi_i')^4}{\sigma^4} \right)^{\frac{1}{2}},$$

where c is a constant.

We recall that $\zeta_N = \xi_1 + \xi_2 + \cdots + \xi_N$. Let $e_0(p)$ be the expectation of $\xi_i^{\{0\}}$ and let $\sigma_0^2(p)$ be the variance of $\xi_i^{\{0\}}$. Introduce notation:

(3.1) $$S_{KN} = \xi_1^{\{0\}} + \cdots + \xi_K^{\{0\}} + \xi_{K+1} + \ldots \xi_N.$$

Then $e_{KN}(p) = Ke_0(p) + (N - K)e(p)$ is the expectation of S_{KN} and $\sigma_{KN}^2(p) = K\sigma_0^2(p) + (N - K)\sigma^2(p)$ is the variance of S_{KN}.

Next lemma offers a Gaussian approximation for S_{KN}.

Lemma 3.2. *Let ξ_i be the geometrically distributed random variable defined by (1.7) and let $\xi_i^{\{0\}}$ be created from ξ_i by the rule (1.6). Let S_{KN} be the sum in (3.1). If $p \geq 1/2$, then*

$$(3.2) \qquad \sup_{t \in \mathbb{R}} \left| \mathbb{P}\left(\frac{S_{KN} - e_{KN}(p)}{\sigma_{KN}(p)} < t \right) - \Phi(t) \right| < 2c \left(\frac{38p^2}{N} \right)^{\frac{1}{2}}$$

and

$$(3.3) \qquad \sigma_{KN}^2(p) = N\sigma^2(p).$$

Proof. As ξ_i has geometric distribution, then its factorial moment is

$$\mathbb{E}\xi_i^{[k]} = \mathbb{E}\xi_i(\xi_i - 1)\cdots(\xi_i - k + 1) = k! \left(\frac{p}{1-p} \right)^k.$$

Therefore, we have

$$e(p) = \mathbb{E}\xi_i = \mathbb{E}\xi_i^{[1]} = \frac{p}{1-p},$$

$$\mathbb{E}\xi_i^2 = \mathbb{E}\xi_i^{[2]} + \mathbb{E}\xi_i^{[1]} = 2 \left(\frac{p}{1-p} \right)^2 + \frac{p}{1-p},$$

$$(3.4) \quad \mathbb{E}\xi_i^3 = \mathbb{E}\xi_i^{[3]} + 3\mathbb{E}\xi_i^{[2]} + \mathbb{E}\xi_i^{[1]} = 6 \left(\frac{p}{1-p} \right)^3 + 6 \left(\frac{p}{1-p} \right)^2 + \frac{p}{1-p},$$

$$\mathbb{E}\xi_i^4 = \mathbb{E}\xi_i^{[4]} + 6\mathbb{E}\xi_i^{[3]} + 7\mathbb{E}\xi_i^{[2]} + \mathbb{E}\xi_i^{[1]}$$

$$= 24 \left(\frac{p}{1-p} \right)^4 + 36 \left(\frac{p}{1-p} \right)^3 + 14 \left(\frac{p}{1-p} \right)^2 + \frac{p}{1-p}.$$

Consequently, the variance of ξ_i is

$$(3.5) \quad \sigma^2(p) = \mathbb{E}\xi_i^{[2]} + \mathbb{E}\xi_i^{[1]} - (\mathbb{E}\xi_i^{[1]})^2 = \left(\frac{p}{1-p} \right)^2 + \frac{p}{1-p} = \frac{p}{(1-p)^2}.$$

Using (3.4), we obtain

$$\mathbb{E}(\xi_i - \mathbb{E}\xi_i)^4 = \mathbb{E}\xi_i^4 - 4\mathbb{E}(\xi_i^3)(\mathbb{E}\xi_i) + 6(\mathbb{E}\xi_i^2)(\mathbb{E}\xi_i)^2 - 3(\mathbb{E}\xi_i)^4$$

$$= 24 \left(\frac{p}{1-p} \right)^4 + 36 \left(\frac{p}{1-p} \right)^3 + 14 \left(\frac{p}{1-p} \right)^2 + \frac{p}{1-p}$$

$$(3.6) \qquad\qquad - 4 \left(6 \left(\frac{p}{1-p} \right)^3 + 6 \left(\frac{p}{1-p} \right)^2 + \frac{p}{1-p} \right) \left(\frac{p}{1-p} \right)$$

$$+ 6 \left(2 \left(\frac{p}{1-p} \right)^2 + \frac{p}{1-p} \right) \left(\frac{p}{1-p} \right)^2 - 3 \left(\frac{p}{1-p} \right)^4$$

$$= 9 \left(\frac{p}{1-p} \right)^4 + 18 \left(\frac{p}{1-p} \right)^3 + 10 \left(\frac{p}{1-p} \right)^2 + \frac{p}{1-p}.$$

So we have

$$(3.7) \qquad \mathbb{E}(\xi_i - \mathbb{E}\xi_i)^4 \leq 38 \left(\frac{p}{1-p}\right)^4 \quad \text{as} \quad \frac{p}{1-p} \geq 1.$$

We can see that the distribution of $\xi_i^{\{0\}}$ is the same as that of $\xi_i + 1$. Therefore, $e_0(p) = \mathbb{E}\xi_i^{\{0\}} = \mathbb{E}\xi_i + 1 = \frac{1}{1-p}$. The central moments of $\xi_i^{\{0\}}$ and ξ_i are the same. Consequently, the variance of $\xi_i^{\{0\}}$ is

$$(3.8) \qquad \sigma_0^2(p) = \frac{p}{(1-p)^2}$$

and for the fourth central moment we have

$$(3.9) \qquad \mathbb{E}(\xi_i^{\{0\}} - \mathbb{E}\xi_i^{\{0\}})^4 \leq 38 \left(\frac{p}{1-p}\right)^4.$$

Using (3.5), (3.7), (3.8), (3.9) and Lemma 3.1, we obtain (3.2). Equality (3.3) follows from equalities (3.5) and (3.8). $\square$ $\qquad\qquad$ $\square$

Lemma 3.3. *Let ξ_i be the geometrically distributed random variable defined by (1.7) and let $\xi_i^{\{0\}}$ be created from ξ_i by the rule (1.6). Let S_{KN} be the sum in (3.1). Let $N \to \infty$ and $p \geq C > 0$. Then we have*

$$(3.10) \qquad \sigma_{KN}(p)\mathbb{P}\{S_{KN} = l\} - \frac{1}{\sqrt{2\pi}}e^{-\frac{(l-e_{KN}(p))^2}{2\sigma_{KN}^2(p)}} \to 0$$

uniformly for $0 \leq K \leq N$ and $l = 0, 1, 2, \ldots$. The statement includes the case of $p \to 1$.

Proof. We will use the following notation. $\phi(t) = \frac{1-p}{1-pe^{it}}$ is the characteristic function of ξ_i, $\phi^c(t) = \phi(t)e^{-ite(p)}$ is the characteristic function of $\xi_i - \mathbb{E}\xi_i$, $\phi_0(t) = e^{it}\frac{1-p}{1-pe^{it}}$ is the characteristic function of $\xi_i^{\{0\}}$, $\phi_0^c(t) = \phi_0(t)e^{-ite_0(p)}$ is the characteristic function of $\xi_i^{\{0\}} - \mathbb{E}\xi_i^{\{0\}}$, and $\phi_{KN}(t)$ is the characteristic function of $\frac{S_{KN} - e_{KN}(p)}{\sigma_{KN}(p)}$. We know that $\phi_0^c(t) = \phi^c(t)$.

Let

$$z = \frac{l - e_{KN}(p)}{\sigma_{KN}(p)}.$$

The inversion formula for an integer valued random variable X is

$$\mathbb{P}\{X = k\} = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-itk} \phi_X(t)dt,$$

where $\phi_X(t)$ is the characteristic function of X. Therefore, by short calculation:

$$\mathbb{P}\{S_{KN} = l\} = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-it\sigma_{KN}(p)z} \phi_{KN}(\sigma_{KN}(p)t)dt$$

$$= \frac{1}{2\pi\sigma_{KN}(p)} \int_{-\sigma_{KN}(p)\pi}^{\sigma_{KN}(p)\pi} e^{-itz} \phi_{KN}(t)dt.$$

Choose $0 < \varepsilon < 1$ and $B > 0$. Using

$$\frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}} = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-ixz}e^{-\frac{x^2}{2}}\,dx,$$

we obtain

$$(3.11) \quad R_N = 2\pi\left(\sqrt{N}\sigma(p)\mathbb{P}\{S_{KN} = l\} - \frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}\right) = I_1 + I_2 + I_3 + I_4,$$

where

$$I_1 = \int_{|x|<B} e^{-ixz}\phi_{KN}(x)dx - \int_{|x|<B} e^{-ixz}e^{-\frac{x^2}{2}}\,dx,$$

$$I_2 = -\int_{|x|>B} e^{-ixz}e^{-\frac{x^2}{2}}\,dx,$$

$$I_3 = \int_{B<|x|\le\varepsilon\sqrt{N}\sigma(p)} e^{-ixz}\left(\phi^c\left(\frac{x}{\sigma_{KN}(p)}\right)\right)^N dx$$

$$I_4 = \int_{\varepsilon\sqrt{N}\sigma(p)<|x|\le\pi\sqrt{N}\sigma(p)} e^{-ixz}\left(\phi^c\left(\frac{x}{\sigma_{KN}(p)}\right)\right)^N dx.$$

Since, by Lemma 3.2,

$$S_{KN} \xrightarrow{d} \gamma \quad \text{as} \quad N \to \infty,$$

so $\phi_{KN}(x) \to e^{-\frac{x^2}{2}}$, therefore

$$(3.12) \qquad\qquad\qquad I_1 \to 0$$

for any fixed $B > 0$.

Since

$$|I_2| \le \int_{|x|>B} e^{-\frac{x^2}{2}}\,dx,$$

therefore

$$(3.13) \qquad\qquad |I_2| \to 0 \quad \text{as} \quad B \to \infty.$$

We need the following formula for the characteristic function:

$$|\phi(t)| = |\phi^c(t)| = \left|\frac{1-p}{1-pe^{it}}\right| = \frac{1-p}{\sqrt{(1-p\cos(t))^2 + p^2\sin^2(t)}}$$

$$= \sqrt{\frac{(1-p)^2}{1-2p\cos(t)+p^2}} = \sqrt{\frac{(1-p)^2}{(1-p)^2 + 2p(1-\cos(t))}}$$

$$= \sqrt{\frac{1}{1 + \frac{2p(1-\cos(t))}{(1-p)^2}}}.$$

Therefore, we obtain

$$
|I_3| \leq \int_{B < |x| \leq \varepsilon\sqrt{N}\sigma(p)} \left| e^{-ixz} \left(\phi^c \left(\frac{x}{\sqrt{N}\sigma(p)} \right) \right)^N \right| dx
$$

$$
= \int_{B < |x| \leq \varepsilon\sqrt{N}\sigma(p)} \left(\frac{1}{1 + \frac{2p\left(1 - \cos\left(\frac{x}{\sqrt{N}\sigma(p)}\right)\right)}{(1-p)^2}} \right)^{N/2} dx
$$

$$
= \int_{B < |x| \leq \varepsilon\sqrt{N}\sigma(p)} \exp\left(-\frac{N}{2} \ln\left(1 + \frac{2p\left(1 - \cos\left(\frac{x}{\sqrt{N}\sigma(p)}\right)\right)}{(1-p)^2} \right) \right) dx
$$

$$
= \int_{B < |x| \leq \varepsilon\sqrt{N}\sigma(p)} \exp\left(-\frac{N}{2}(1 + o(1)) \frac{2p\left(1 - \cos\left(\frac{x}{\sqrt{N}\sigma(p)}\right)\right)}{(1-p)^2} \right) dx
$$

$$
= \int_{B < |x| \leq \varepsilon\sqrt{N}\sigma(p)} \exp\left(-\frac{N}{2}(1 + o(1))(1 + O(\varepsilon)) \frac{2p\frac{1}{2}\left(\frac{x}{\sqrt{N}\sigma(p)}\right)^2}{(1-p)^2} \right) dx
$$

$$
= \int_{B < |x| \leq \varepsilon\sqrt{N}\sigma(\alpha)} \exp\left(-\frac{1}{2}(1 + O(\varepsilon)) \frac{px^2}{(1-p)^2\sigma^2(p)} \right) dx
$$

$$
= \int_{B < |x| \leq \varepsilon\sqrt{N}\sigma(p)} \exp\left(-\frac{1}{2}(1 + O(\varepsilon))x^2 \right) dx.
$$

Consequently, we have

(3.14) $$ |I_3| \to 0 \quad \text{as} \quad B \to \infty. $$

Since

$$
|\phi(x)| \leq \sqrt{\frac{1}{1 + \frac{2p(1 - \cos(\varepsilon))}{(1-p)^2}}}, \quad \varepsilon \leq x \leq \pi,
$$

therefore, we have

$$
|I_4| \leq 2\pi\sqrt{N}\sigma(p) \left(\frac{1}{1 + \frac{2p(1 - \cos(\varepsilon))}{(1-p)^2}} \right)^{N/2}
$$

$$
\leq C\sqrt{N} \left(1 + \frac{2p(1 - \cos(\varepsilon))}{(1-p)^2} \right)^{-\frac{N-1}{2}}.
$$

Therefore,

(3.15) $$ |I_4| \to 0. $$

Relations (3.12), (3.13), (3.14), (3.15), and (3.11) imply (3.10). □

Lemma 3.4. *Assume that the conditions of Theorem 1.1 or the conditions of Theorem 1.2 are satisfied. Then we have*

(3.16)
$$\frac{\mathbb{P}(S_{(K-k)(N-k)} = n)}{\mathbb{P}(S_N = n)} = 1 + o(1).$$

In the case of Theorem 1.1, this convergence is uniform for $|z| < C$, where $z = \frac{k - K(1-p)}{\sqrt{Kp(1-p)}}$.

Proof. First we show that

(3.17)
$$N - k = N(1 + o(1)).$$

When the conditions of Theorem 1.1 are valid, then $n/N \to \infty$, so $p = \frac{n}{n+N} \to 1$. As $z = \frac{k - K(1-p)}{\sqrt{Kp(1-p)}}$, therefore we have

$$N - k = N\left(1 + \frac{K}{N}(1-p) + z\frac{\sqrt{Kp(1-p)}}{N}\right).$$

As $|z| < C$, $K \leq N$, and $p \to 1$, we obtain (3.17). On the other hand, when the conditions of Theorem 1.2 are satisfied, then k is a fixed number. So we also have (3.17).

Now, using Lemma 3.3, equation (3.17) and the formula $p = \frac{n}{n+N}$, we obtain

(3.18) $\dfrac{\mathbb{P}(S_{(K-k)(N-k)} = n)}{\mathbb{P}(S_N = n)}$

$$= \frac{\frac{1}{\sqrt{2\pi}(N-k)\sigma(p)}\left(\exp\left(-\frac{\left(n - (N-K)\frac{\frac{n}{n+N}}{1 - \frac{n}{n+N}} - (K-k)\frac{\frac{n}{n+N}}{1 - \frac{n}{n+N}}\frac{1}{\frac{n}{n+N}}\right)^2}{2(N-k)\sigma^2(p)}\right) + o(1)\right)}{\frac{1}{\sqrt{2\pi}N\sigma(p)}\left(\exp\left(-\frac{\left(n - N\frac{\frac{n}{n+N}}{1 - \frac{n}{n+N}}\right)^2}{2N\sigma^2(p)}\right) + o(1)\right)}$$

$$= \frac{N}{N-k}\frac{\exp\left(-\frac{\left(K\frac{\frac{n}{n+N}}{1 - \frac{n}{n+N}}\left(1 - \frac{1}{\frac{n}{n+N}}\right) + k\frac{\frac{n}{n+N}}{1 - \frac{n}{n+N}}\frac{1}{\frac{n}{n+N}}\right)^2}{2(N-k)\sigma^2(p)}\right) + o(1)}{\exp\left(-\frac{0}{2N\sigma^2(p)}\right) + o(1)}$$

$$= \exp\left(-\frac{\left(K\frac{\frac{n}{n+N}}{1 - \frac{n}{n+N}}\left(1 - \frac{1}{\frac{n}{n+N}}\right) + k\frac{\frac{n}{n+N}}{1 - \frac{n}{n+N}}\frac{1}{\frac{n}{n+N}}\right)^2}{2(N-k)\sigma^2(p)}\right) + o(1)$$

$$= \exp\left(-\frac{\left(-K + k\frac{n+N}{N}\right)^2}{2(N-k)\frac{\frac{n}{n+N}}{\left(1-\frac{n}{n+N}\right)^2}}\right) + o(1) = \exp\left(-\frac{\left(-K + k\frac{n+N}{N}\right)^2}{2(N-k)\frac{n(N+n)}{N^2}}\right) + o(1).$$

Now we prove (3.16) separately for Theorem 1.1 and Theorem 1.2. First, let the conditions of Theorem 1.1 be valid. Then we have

$$(3.19) \quad k = K(1-p) + z\sqrt{Kp(1-p)} = K\frac{N}{n+N} + z\sqrt{K\frac{nN}{(n+N)^2}}, \quad |z| < C.$$

Using equations (3.19) and (3.18), we obtain

$$(3.20)$$

$$\frac{\mathbb{P}(S_{(K-k)(N-k)} = n)}{\mathbb{P}(S_N = n)}$$

$$= \exp\left(-\frac{\left(-K + \left(K\frac{N}{n+N} + z\sqrt{K\frac{nN}{(n+N)^2}}\right)\frac{n+N}{N}\right)^2}{2(N-k)\frac{n(n+N)}{N^2}}\right) + o(1)$$

$$= \exp\left(-\frac{z^2 K\frac{n}{N}}{2(N-k)\frac{n(n+N)}{N^2}}\right) + o(1)$$

$$= \exp\left(-\frac{z^2 KN}{2(N-k)(n+N)}\right) + o(1)$$

$$= 1 + o(1).$$

In the last step we used the formula $\frac{N-k}{K} \cdot \frac{n+N}{N} \to \infty$, which follows from conditions $\frac{n}{N} \to \infty$ and $|z| \leq C$.

Now, let the conditions of Theorem 1.2 be valid. Then we have

$$\frac{\left(-K + k\frac{n+N}{N}\right)^2}{2(N-k)\frac{n(N+n)}{N^2}} \leq \frac{(K)^2}{(N-k)\frac{n(N+n)}{N^2}} + \frac{\left(k\frac{n+N}{N}\right)^2}{(N-k)\frac{n(N+n)}{N^2}}$$

$$= \frac{K}{N-k}\frac{NK}{n}\frac{N}{N+n} + k^2\frac{n+N}{(N-k)n} = o(1).$$

Therefore, using (3.18), we obtain

$$(3.21) \qquad \frac{\mathbb{P}(S_{(K-k)(N-k)} = n)}{\mathbb{P}(S_N = n)} = \exp\left(-o(1)\right) + o(1) = 1 + o(1).$$

We see that (3.20) and (3.21) imply (3.16). □ □

We will use the following version of the de Moivre–Laplace theorem.

Lemma 3.5. *Let A_{Ki}, $1 \leq i \leq K$, $K \in \mathbb{N}$, be an array of row-wise independent events having the same probability within rows. Let $s = s(K) = \mathbb{P}(A_{Ki})$*

denote these probabilities. Let $\mathbb{I}_{Ki}$ denote the indicator of the event A_{Ki}. Suppose $K \to \infty$ such that $Ks(1-s) \to \infty$. Then we have

$$(3.22) \qquad \sqrt{Ks(1-s)}\mathbb{P}\left(\sum_{i=1}^{K}\mathbb{I}_{Ki} = k\right) - \frac{1}{\sqrt{2\pi}}e^{-\frac{(k-Ks)^2}{2Ks(1-s)}} \to 0$$

uniformly for those values of $k = 0, 1, 2, \ldots$ for which $\left|\frac{k-K(1-s)}{\sqrt{Ks(1-s)}}\right| < C$.

For the proof see e.g. [9].

Proof of Theorem 1.1. We can see that

$$Kp(1-p) = K\frac{n}{N+n}\frac{N}{N+n} = \frac{KN}{n}(1+o(1)) \to \infty.$$

Therefore, we can apply Lemma 3.5 with $s = 1-p$. Consider the right hand side of (1.5). There $p_0 = 1-p$. So, for the binomial probability in (1.5), relation (3.22) implies

$$\binom{K}{k}(p_0)^k(1-p_0)^{K-k} \to \frac{1}{\sqrt{2\pi Kp(1-p)}}e^{-\frac{z^2}{2}},$$

where $z = \frac{k-K(1-p)}{\sqrt{Kp(1-p)}}$ and this convergence is uniform for $|z| < C$. For the other part of (1.5), by (3.16), we have

$$\frac{\mathbb{P}\{\zeta_{N-k}^{\{r\}} = n - kr\}}{\mathbb{P}\{\zeta_N = n\}} \to 1.$$

So we obtain (1.2).   □

Proof of Theorem 1.2. Recall that in the right hand side of (1.5), $p_0 = 1-p$. Observe that

$$K(1-p) = K\frac{N}{N+n} = \frac{KN}{n}(1+o(1)) = \lambda(1+o(1)).$$

Therefore, using the well-known Poisson limit theorem for the binomial probability in (1.5), we obtain

$$\binom{K}{k}(p_0)^k(1-p_0)^{K-k} \to e^{-\lambda}\frac{\lambda^k}{k!}.$$

For the other part of (1.5), by (3.16), we have

$$\frac{\mathbb{P}\{\zeta_{N-k}^{\{r\}} = n - kr\}}{\mathbb{P}\{\zeta_N = n\}} \to 1.$$

So we obtain (1.4).   □

References

[1] Barbour, A. D., Holst, L., Janson, S., *Poisson Approximation*, Oxford University Press, Oxford, 1992.

[2] Chuprunov, A. N., Fazekas, I., *Poisson limit theorems for the generalized allocation scheme*, Ann. Univ. Sci. Budapest, Sect. Comp. **49** (2019), 77–96.

[3] Gibbons, J. D., *Nonparametric Statistical Inference*, McGraw-Hill, New York, 1971.

[4] Gordon, L., Schilling, M. F., Waterman, M. S., *An extreme value theory for long head runs*, Probab. Theory Related Fields **72** (1986), 279–287.

[5] Khakimullin, E. R., Enatskaya, N. Yu., *Limit theorems for the number of empty cells*, Diskret. Mat. **9** (2) (1997), 120–130 (Russian); translation in Discrete Math. Appl. **7** (2) (1997), 209–219.

[6] Kolchin, V. F., *A class of limit theorems for conditional distributions*, Litovsk. Mat. Sb. **8** (1968), 53–63 (Russian).

[7] Kolchin, V. F., *Random Graphs*, Cambridge University Press, Cambridge, 1999.

[8] Kolchin, V. F., Sevast'yanov, B. A., Chistyakov, V. P., *Random Allocations*, V. H. Winston & Sons, Washington D. C., 1978.

[9] Rényi, A., *Probability Theory*, Elsevier, New York, 1970.

[10] Timashev, A. N., *Asymptotic Expansions in Probabilistic Combinatorics*, TVP Science Publishers, Moscow, 2011 (Russian).

[11] Trunov, A. N., *Limit theorems in the problem of distributing identical particles in different cells*, Proc. Steklov Inst. Math. **177** (1988), 157–175.

Alexey Chuprunov
Department of Math. Anal.
Kazan Federal University
Kremlevskaya 35, room 503
420008 Kazan
Russia
e-mail: `achuprunov@mail.ru`

István Fazekas
Faculty of Informatics
University of Debrecen
P.O. Box 400
4002 Debrecen
Hungary
e-mail: `fazekas.istvan@inf.unideb.hu`