International Journal of Pure and Applied Mathematics
Volume 118 No. 4 2018, 1021-1032
ISSN: 1311-8080 (printed version); ISSN: 1314-3395 (on-line version)
url: http://www.ijpam.eu
doi: 10.12732/ijpam.v118i4.16

ijpam.eu

REPEATED RANDOM ALLOCATIONS WITH INCOMPLETE INFORMATION

Vladimir G. Panov[1 §], Julia V. Nagrebetskaya[2]
[1]Institute of Industrial Ecology of Ural Branch of RAS
S. Kovalevskaya st., 20, Ekaterinburg, RUSSIA
[2]Ural Federal University,
V. Lenin pr., 51, Ekaterinburg, RUSSIA

Abstract: We consider random allocations into urns under the condition that this is done repeatedly. It is assumed that only information on the number of balls that are in each urn for all allocations is available, rather then corresponding occupation numbers. Three schemes of such allocations are introduced and analysed. The probability distribution functions, probability generating functions, and first and second order moments of corresponding random variables are presented. Maximum likelihood estimates of the parameters and hypothesis testing are discussed as well.

AMS Subject Classification: 05A15, 62H10, 62H12
Key Words: random allocations without information about occupation numbers, multivariate generation function, multinomial distribution, hypothesis testing, maximum likelihood estimates

1. Introduction

In allocation problems, balls (objects or particles) of different kinds are distributed into urns (boxes or cells) of different kinds. When the balls are randomly distributed into the urns following random variables may be naturally defined. The number ξ_0 of empty urns as well as the number ξ_i of urns occupied by i balls. These random variables were intensively studied by Charalambides [3], Kolchin et al. [9], Johnson and Kotz [6].

Received: March 2, 2018
Revised: May 11, 2018
Published: May 23, 2018

§Correspondence author

A collection of m balls of a general specification may be denoted by $(1^{q_1}2^{q_2} \ldots m^{q_m})$, where q_j is the number of different kinds of balls, each including exactly j like balls, $j = 1, 2, \ldots, m$, with $q_1+2q_2+\cdots+mq_m = m$. For example, (1^m) denotes the case where all m balls are of different kinds, i.e. a probability to fall into a certain urn is different for different balls. On the contrary, $\left(m^1\right)$ means that there is only one class of similar balls, i.e. all balls have the equal probabilities to fall into a certain urn.

General issues of urn model, its typical problems and methods for their solution are considered by Charalambides [3], Johnson and Kotz [6], Mahmoud [12]. A well-known and surprising fact is that many results in discrete probability theory can be derived from an appropriate urn model. Classical results described by Johnson and Kotz [6], Kotz and Balakrishnan [10] may be supplemented by recent applications in evolutionary processes presented by Benaïm et al. [2], in biology by Knoblauch et al. [8] and in physics by Niven and Grendar [13]. In addition, quite a long time there are important applications of urn models in the field of clinical trails, see, e.g. Hu and Rosenberger [5]. Of course, there are many other interesting applications of the urn models which we do not mention here.

However, a typical random allocation problem assumes that single distribution of the balls into urns is performed and related variables are studied. For example, probability distribution of variables ξ_i or their asymptotics are studied in the classical allocation problem formulation. Note that when balls are distributed once, we have information about how many balls are in each urn and this allows us to estimate effectively probability to put a certain kind ball into a certain urn.

We consider a random allocation scheme under which such an allocation occurs repeatedly (in general, an arbitrary finite number). We assume that a number of balls that fall into a certain urn will not be known, and instead we knew the number of common balls in each urn for all allocations under consideration.

As a typical example, consider the following problem. Let some number of objects with given (individual) characteristics be allocated several times into n classes. If rules for the objects distribution are unchanged for all allocations, then we can expect a certain probability distribution of the number of common objects in each class. On the other hand, if at least once the rule of object distribution into classes was different, then the number of common objects should be distributed differently and this can be examined by appropriate statistical methods.

Thus, the scheme under consideration makes it possible to verify invariance of rules by which the classification of objects is performed. Let us remember that we don't use any information about occupation numbers ξ_i, instead we take into consideration how many common objects are in each class after each of successive allocations.

Note that the number of common balls in urns must quickly decrease with each new distribution, so the problem in question would be more interesting when number m of balls is much more than the number of urns n. However, this problem can be considered for arbitrary ratio of number of balls and number of urns.

2. Notations and the problem statement

Let we have n urns and a supply of m balls (objects) of different kinds. Each ball is assigned to an urn chosen randomly, so that probability of assignment depends on the urn and the kind of a ball. We consider three schemes for the allocation of balls into urns which are based on a preliminary information about probabilities of putting m balls into n urns.

(A) No preliminary information is available, i.e. balls are distributed into urns randomly and equiprobably. In this case the probability that each ball puts into a certain urn is the same and is equal to $\frac{1}{n}$.

(B) There is a preliminary information that allows us to specify the probability to distribute balls into a certain urn, namely, the probability p_j of putting a ball into the jth urn, $\sum\limits_{j=1}^{n} p_j = 1$.

(C) The balls are classified by a certain property in such a way that for every ball from the class i its probability to be distributed into the jth urn is equal to p_{ij}, $\sum\limits_{j=1}^{n} p_{ij} = 1$. Let t be the number of classes of likely balls (relative to the classifying property) and s_i is the number of balls in the class i, $1 \le i \le t$, $\sum_{i=1}^{t} s_i = m$. Of course, empty class is excluded from the subsequent consideration, so we assume that $s_i \neq 0$, $i = 1, 2, \ldots, t$.

It is obvious that case (C) is general and is simplified to (A) or (B) with corresponding assumptions: $t = 1$ for case (B), and $t = 1$ and $p_{ij} = \frac{1}{n}$ for case (A). Therefore, we only consider the case (C) below, giving the corresponding formulas for cases (A) and (B) as well. Each of these cases will be referred to as a *balls distribution into urns rule.* We assume that balls are distributed into urns r times according to one of the aforementioned rules, and each time the allocation rule remains the same.

The problem we will be considering is to determine probability distribution for the number of common balls that will be in each of the urns after r allocations. Namely, let k_i be the number of common balls that are present in the ith urn for all r allocations. Thus, it is required to find the probability distribution of the random variable $\mathcal{K} = (k_1, k_2, \ldots, k_n)$.

To shorten we use multi-index notation. A set of values $(k_1, k_2, \ldots, k_n)$ that takes a random variate $\mathcal{K}$ is denoted by $\mathbf{k}$. Accordingly, $\mathbf{k}! = k_1!k_2! \cdots k_n!$ and $|\mathbf{k}| = \sum_{j=1}^{n} k_j$. For k_j common balls in jth urn and for a class i, $1 \leq i \leq t$, one can specify a number m_{ij}, $0 \leq m_{ij} \leq k_j$, of balls from among these k_j which are in the class i. Thus, one can write the following decomposition $k_j = (m_{1j}, m_{2j}, \ldots, m_{tj})$ and the equalities

$$\sum_{i=1}^{t} m_{ij} = k_j, \ \sum_{i=1}^{t} \sum_{j=1}^{n} m_{ij} = |\mathbf{k}|$$

3. Main results

The next theorem shows that the distribution of the random variable $\mathcal{K}$ is closely related to the multinomial one.

Theorem 1. *If the case (C) presents then the following equality holds*

$$\mathsf{P}(\mathcal{K} = \mathbf{k}) = \sum{}' \prod_{i=1}^{t} \binom{s_i}{\mathbf{m_i}\,(s_i - |\mathbf{m_i}|)} \prod_{j=1}^{n} \left(p_{ij}^{m_{ij}}\right)^r \left(1 - \sum_{j=1}^{n} p_{ij}^r\right)^{s_i - |\mathbf{m_i}|} \tag{1}$$

where summation performs over all decompositions of the numbers $(k_1, \ldots, k_n)$ *as described above. Multi-indices* $\mathbf{m_i}$ *is defined as follows* $\mathbf{m_i} = (m_{i1}, m_{i2}, \ldots, m_{in})$, $0 \leq m_{ij} \leq k_j$, *and*

$$\binom{s_i}{\mathbf{m_i}\,(s_i - |\mathbf{m_i}|)} = \binom{s_i}{m_{i1} m_{i2} \ \cdots \ m_{in} \ (s_i - |\mathbf{m_i}|)}$$

are multinomial coefficients.

Proof. Let numbers $1, 2, \ldots, n$ be assigned to the urns in an arbitrary way. Consider a decomposition $k_j = (m_{1j}, m_{2j}, \ldots, m_{tj})$ for a given value k_j of the common balls in the jth urn, $j = 1, 2, \dots, n$. The probability that in first urn there will be $k_1 = (m_{11}, m_{21}, \dots, m_{t1})$ common balls for all r allocations is

$$\binom{s_1}{m_{11}} p_{11}^{m_{11}} \binom{s_2}{m_{21}} p_{21}^{m_{21}} \cdots \binom{s_t}{m_{t1}} p_{t1}^{m_{t1}} \tag{2}$$

The similar probability for the second urn is equal to

$$\binom{s_1 - m_{11}}{m_{12}} p_{12}^{m_{12}} \binom{s_2 - m_{21}}{m_{22}} p_{22}^{m_{22}} \cdots \binom{s_t - m_{t1}}{m_{t2}} p_{t2}^{m_{t2}} \tag{3}$$

and for the last urn the probability that it has k_n common balls after r allocations is

$$\binom{s_1 - \sum_{j=1}^{n-1} m_{1j}}{m_{1n}} p_{1n}^{m_{1n}} \binom{s_2 - \sum_{j=1}^{n-1} m_{2j}}{m_{2n}} p_{2n}^{m_{2n}} \cdots \binom{s_t - \sum_{j=1}^{n-1} m_{tj}}{m_{tn}} p_{tn}^{m_{tn}} \tag{4}$$

Hence, the probability $\mathsf{P}(\mathcal{K} = \mathbf{k})$, $\mathbf{k} = (k_1, \dots, k_n)$, is equal to the product of (2)–(4) (after straightforward simplification)

$$\prod_{i=1}^{t} \binom{s_i}{m_{i1} m_{i2} \dots m_{in} \left(s_i - \sum_{j=1}^{n} m_{ij}\right)} \prod_{j=1}^{n} p_{ij}^{m_{ij}} \tag{5}$$

Thus, we have received that the urns are occupied with the number of common balls that are specified by the value $\mathbf{k}$. However, there are still $m - |\mathbf{k}|$ balls which should be distributed into the same urns in almost arbitrary way. Namely, the only restriction is that if a ball in a particular allocation falls into a certain urn then it must fall into another urn in the remaining allocations at least once, otherwise such a ball should be among common balls. Now we associate a tuple of integers $(a_1, a_2, \dots, a_r)$ with each ball a from the remaining $m - |\mathbf{k}|$ balls where a_j is the number of urn into which the ball a falls at jth allocation, $1 \le a_j \le n$. Since the ball a is not included in the number of common balls in any of the urns the equality $a_1 = a_2 = \cdots = a_r$ *should not* be satisfied.

The probability of obtaining the set $(a_1, a_2, \dots, a_r)$ depends on a class to which the ball a belongs. Namely, if the ball belongs to the class i then the probability of gathering the tuple $(a_1, a_2, \dots, a_r)$ is equal to $\prod_{j=1}^{r} p_{ia_j}$.

After the distribution of $\mathbf{k}$ common balls into n urns in the ith urn $s'_i = s_i - |\mathbf{m_i}|$ balls remain undistributed, $0 \le s'_i \le s_i$, $\sum_{i=1}^{t} s'_i = m - |\mathbf{k}|$. Let T_i be the rest of the class i after the allocation of $\mathbf{k}$ common balls, $|T_i| = s'_i$. Then the probability of the aforementioned distribution of rest balls is equal to the following product

$$\prod_{i=1}^{t}\prod_{a\in T_i}\left(\sum_{(a_1,\dots,a_r)}{}' \prod_{l=1}^{r} p_{ia_l}\right), \tag{6}$$

where the summation done over all tuples of integers $(a_1, \dots, a_r)$ with not all equal numbers $a_j,\ 1 \le a_j \le n$. Therefore this summation can be presented as follows

$$\sum_{(a_1,\dots,a_r)}{}' \prod_{l=1}^{r} p_{ia_l} = \sum_{\substack{(a_1,\dots,a_r)\\ 1\le a_j\le n\\ \neg(a_1=\dots=a_r)}} \prod_{l=1}^{r} p_{ia_l} = \sum_{\substack{(a_1,\dots,a_r)\\ 1\le a_j\le n}} \prod_{l=1}^{r} p_{ia_l} - \sum_{j=1}^{n} p_{ij}^{r}$$

Since, obviously, the following equality holds

$$\sum_{\substack{(a_1,\dots,a_r)\\ 1\le a_j\le n}} \prod_{l=1}^{r} p_{ia_l} = \left(\sum_{j=1}^{n} p_{ij}\right)^{r} = 1,$$

we obtain the expression

$$\prod_{i=1}^{t}\prod_{a\in T_i}\left(1-\sum_{j=1}^{n} p_{ij}^{r}\right)$$

for the probability of allocations the rest $m-|\mathbf{k}|$ balls into urns under condition that each ball doesn't fall into the same urn at the sequence of r allocations.

The latter equality takes the form

$$\prod_{i=1}^{t}\prod_{a\in T_i}\left(1-\sum_{j=1}^{n} p_{ij}^{r}\right) = \prod_{i=1}^{t}\left(1-\sum_{j=1}^{n} p_{ij}^{r}\right)^{s_i-|\mathbf{m_i}|},$$

since each ball $a \in T_i$ has the same probability p_{ij}.

Since common balls specified by the set $\mathbf{k}$ are preserved in every allocation, using (5), it follows that the probability to obtain $\mathbf{k}$ common balls into n urns after r allocations is equal to

$$\sum_{\substack{k_1=(m_{11},\dots,m_{t1})\\ \vdots \\ k_n=(m_{1n},\dots,m_{tn})}} \prod_{i=1}^{t} \binom{s_i}{m_{i1}m_{i2}\dots m_{in}\left(s_i-\sum_{j=1}^{n} m_{ij}\right)} \prod_{j=1}^{n}\left(p_{ij}^{m_{ij}}\right)^{r},$$

where the outer summation done over all decomposions of the integers $k_1, k_2, \ldots, k_n$ which are described above. Combining the latter formula with the probability of allocations of the rest $m - |\mathbf{k}|$ balls, we get the final formula (1). □

Remark. In the terms of number theory a sequence of integers $(m_{1j}, m_{2j}, \ldots, m_{tj})$ is a t-part (weak) composition of k_j [4]. Let us denote by $\mathcal{C}(n,k)$ the set of all k-part integer compositions of n, π_i a composition from $\mathcal{C}(s_i, n)$, and π_{ij} its jth part. Then (1) may be presented as follows

$$\mathsf{P}(\mathcal{K} = \mathbf{k}) = \sum_{(\pi_{1j},\ldots,\pi_{tj})\in\mathcal{C}(k_j,t)} \prod_{i=1}^{t} \binom{s_i}{\mathbf{m_i}\,(s_i - |\mathbf{m_i}|)} \prod_{j=1}^{n} \left(p_{ij}^{m_{ij}}\right)^r \left(1 - \sum_{j=1}^{n} p_{ij}^r\right)^{s_i - |\mathbf{m_i}|}, \quad (7)$$

where $\mathbf{m}_i = (\pi_{i1}, \ldots, \pi_{in})$, $i = 1, \ldots, t$, and the summation done over all $\pi_i \in \mathcal{C}(s_i, n)$, $i = 1, \ldots, t$ for those the condition $(\pi_{1j}, \ldots, \pi_{tj}) \in \mathcal{C}(k_j, t)$ holds.

Corollary 2. *If distribution of balls in urns occurs according to the case (B), then the following equality holds*

$$\mathsf{P}(\mathcal{K} = \mathbf{k}) = \frac{m!}{\mathbf{k}!\,(m - |\mathbf{k}|)!} \prod_{j=1}^{n} \left(p_j^{k_j}\right)^r \left(1 - \sum_{j=1}^{n} p_j^r\right)^{m - |\mathbf{k}|} \quad (8)$$

The proof follows from the fact that we should take $t = 1$ in the formula (1) for the case (B). Then $s_1 = m$ and $m_{1j} = k_j$. □

Corollary 3. *Under the condition of equiprobable distribution of balls into urns according to the rule (A) the following equality holds*

$$\mathsf{P}(\mathcal{K} = \mathbf{k}) = \frac{m!}{\mathbf{k}!\,(m - |\mathbf{k}|)!} \cdot \frac{1}{n^{r|\mathbf{k}|}} \cdot \left(1 - \frac{1}{n^{r-1}}\right)^{m - |\mathbf{k}|} \quad (9)$$

To prove this statement, one should take $p_i = \frac{1}{n}$ in equality (8). □

These corollaries show that in cases (A) and (B) distribution of the random variable $\mathcal{K}$ is multinomial. Namely, the probability distribution (9) corresponds to onefold distribution of m identical balls into $n + 1$ urns, with probability for a ball to fall into first $1, 2, \ldots, n$ urns is equal to $\frac{1}{n^r}$ and the probability to fall into the last $n + 1$th urn is equal to $1 - \frac{1}{n^{r-1}}$. For the case (B) probability for a ball to put into jth urn, $j = 1, 2, \dots, n$, is equal to p_j^r, and probability to fall into the last urn is $1 - \sum_{j=1}^n p_j^r$.

We characterize the general case (C) in terms of probability generating functions. For common references see Charalambides [3], Wilf [16], Johnson and Kotz [7].

Theorem 4. *The probability generating functions of the distributions* (9), (8) *and* (1) *are respectively*

$$g_A(\mathbf{x}) = \left(\sum_{j=1}^n \frac{x_j}{n^r} + x_{n+1} \left(1 - \frac{1}{n^{r-1}} \right) \right)^m \tag{10}$$

$$g_B(\mathbf{x}) = \left(\sum_{j=1}^n x_j p_j^r + x_{n+1} \left(1 - \sum_{j=1}^n p_j^r \right) \right)^m \tag{11}$$

$$g_C(\mathbf{x}) = \prod_{i=1}^t \left(\sum_{j=1}^n x_j p_{ij}^r + x_{n+1} \left(1 - \sum_{j=1}^n p_{ij}^r \right) \right)^{s_i}, \tag{12}$$

where variable $\mathbf{x} = (x_1, x_2, \dots, x_n, x_{n+1}) \in \mathbb{R}^{n+1}$.

Proof is a straightforward consequence of the aforementioned relationship between the random variable $\mathcal{K}$ and multinomial distribution in the cases (A) and (B). The case (C) requires some additional though obvious calculations. We see that in the case (C) the probability generating function g_C is a product of probability generating functions g_B for each class of the ball kinds. That is,

$$g_C(\mathbf{x}) = \prod_{i=1}^t g_{B_i}(\mathbf{x})$$

Hence, the random variable $\mathcal{K}$ in the case (C) is a sum of the random variables $\mathcal{K}_i$ of the case (B), where $i = 1, 2, \dots, t$ (the corresponding property of generating functions see, e.g. in Johnson and Kotz [7, p. 3] or Charalambides [3, p. 57]). □

Basically, Theorem 4 allows one to get full information about the variables $\mathcal{K}$ in the cases (A), (B) or (C). For instance, mean and variance of variables $\mathcal{K}$ presented in the following theorem.

Theorem 5. *Mean and variance of the random variable* $\mathcal{K} = (K_1, \dots, K_n)$ *in the cases (A)–(C) are as follows* $(j = 1, 2, \dots, n)$

$$(A) \quad \mathsf{E}(K_j) = \frac{m}{n^r} \qquad \mathsf{Var}(K_j) = \frac{m(n^r - 1)}{n^{2r}}$$

$$(B)\quad \mathsf{E}(K_j) = m \cdot p_j^r \qquad \mathsf{Var}(K_j) = mp_j^r(1 - p_j^r)$$

$$(C)\quad \mathsf{E}(K_j) = m \cdot \sum_{i=1}^{t} p_{ij}^r \qquad \mathsf{Var}(K_j) = m \sum_{i=1}^{t} p_{ij}^r \left(1 - \sum_{i=1}^{t} p_{ij}^r\right)$$

The covariance between two variables K_i and K_j, $i \neq j$ are given by

$$(A)\quad \mathsf{Cov}(K_i, K_j) = -\frac{m}{n^{2r}}$$

$$(B)\quad \mathsf{Cov}(K_i, K_j) = -mp_i^r p_j^r$$

$$(C)\quad \mathsf{Cov}(K_i, K_j) = -m \cdot \sum_{l=1}^{t} p_{li}^r \cdot \sum_{l=1}^{t} p_{lj}^r$$

As one would expect, mean value of common balls in a particular urn after r distributions decreases as rth power of the probability to fall into this urn.

Remark. From these expressions for expectation of the variables K_j we see that at the values $m < p_j^r$ (for case (B)) one should expect that number of common balls in jth urn will be equal to zero. This determines magnitude of the numbers m, r and p_j so that the expected values of the number of common balls in the urns are not equal to zero.

4. Parameter estimation and hypothesis testing

Since the probability distribution of random variables $\mathcal{K}$ is known (see Theorem 1 and Corollaries 2 and 3), one could easily calculate likelihood estimates for the probabilities p_j or p_{ij}. Consider the main case (8) of repeated random allocations and introduce the following notation. Let $\mathcal{K}^{(l)}$ be a random variable $\mathcal{K}$ after l random allocations, $l = 2, 3, \dots, r$, $\mathcal{K}^{(l)} = (K_1^{(l)}, \dots, K_n^{(l)})$. In this case, obviously, the inequality $K_j^{(l)} \geq K_j^{(l+1)}$, $l = 2, 3, \dots, r-1$ holds. The observable values of the variable $\mathcal{K}^{(l)}$ are denoted by $\mathbf{k}^{(l)} = (k_1^{(l)}, \dots, k_n^{(l)})$.. Denote by $\mathbf{p} = (p_1, p_2, \dots, p_n)$ the vector of probabilities p_j, $j = 1, 2, \dots, n$, which should be estimated by a given sample. Let a function

$$L(\mathbf{p}) = \prod_{l=2}^{r} \mathsf{P}\left(\mathcal{K}^{(l)} = \mathbf{k}^{(l)}\right),$$

be the likelihood, where probabilities $\mathsf{P}\left(\mathcal{K}^{(l)} = \mathbf{k}^{(l)}\right)$ are calculated according to (8).

To obtain a simplified representation for the likelihood function we take a logarithm of $L(\mathbf{p})$ and omit constants (terms which do not depend on the estimated parameters). Therefore, we get the following equality (preserving the notation for $L(\mathbf{p})$)

$$\ln L(\mathbf{p}) = \sum_{l=2}^{r} \sum_{j=1}^{n} \left(k_j^{(l)} \cdot l \ln p_j \right) + \sum_{l=2}^{r} \left\{ \left(m - |\mathbf{k}^{(l)}| \right) \ln \left(1 - \sum_{j=1}^{n} p_j^l \right) \right\}$$

Thus, we have the following system for the maximum likelihood estimates p_a after differentiation the latter equations by $p_a,\ a = 1, 2, \dots, n$.

$$\sum_{l=2}^{r} \left\{ k_a^{(l)} \left(1 - \sum_{j=1}^{n} p_j^l \right) \right\} = \sum_{l=2}^{r} \left(m - |\mathbf{k}^{(l)}| \right) p_a^l, \quad a = 1, \dots, n \qquad (13)$$

Clearly, that (13) is a system of polynomial equations of degree r which can be solved in many ways, see e.g. Basu et al. [1], Sturmfels [15], as well as there are general mathematical software which can solve such systems with or without user intervention (*Mathematica*, Maple, Bertini etc.).

Remark. Due to the fact that number of balls that fall into each urn at the first and the remaining allocations is not known the available information for determining the parameters of a ball's distribution rule, i.e. probabilities p_j, is significantly less than for classical one-time allocation. We can now use only information about the common number of balls in each urn at successive random allocations. Nevertheless, equations (13) show that estimates of these probabilities can be found by solving a system of polynomial equations.

Example. Consider the estimation of probabilities p_j for simulated data. The following parameters were used to simulate the repeated allocations $n = 3, p_1 = 1/2, p_2 = 1/3, p_3 = 1/6$. For every $m = 30, 50, 100$ ten samples were generated and appropriate solutions of system (13) were found for $r = 2, 3, 4, 5$ and 6 repetitions. For every pair (m, r) observed values of the common balls in urns 1, 2 and 3 are registered and system (13) is solved. In all cases there was a unique solution of system (13) in the segment [0;1] which is the estimate of unknown probabilities p_1, p_2, p_3. We show below the pivot table for the mean values of these estimates in dependence on m and r.

Let us denote $p_j^{(l)}$ the probability that a ball falls into jth urn at lth allocation, $l = 1, 2, \dots, r$. Consider for the case (B) statistical hypothesis testing that probabilities $p_j^{(l)}$ don't change at successive allocations. This can be thought of as invariance of the distribution rules at the successive allocations. More

	$m = 30$			$m = 50$			$m = 100$		
r	p_1	p_2	p_3	p_1	p_2	p_3	p_1	p_2	p_3
2	0.443	0.368	0.156	0.418	0.421	0.178	0.446	0.378	0.122
3	0.431	0.369	0.163	0.425	0.413	0.163	0.435	0.379	0.132
4	0.430	0.371	0.159	0.430	0.407	0.163	0.435	0.382	0.130
5	0.426	0.379	0.158	0.428	0.405	0.163	0.432	0.389	0.129
6	0.425	0.384	0.157	0.427	0.403	0.163	0.429	0.394	0.129

Table 1: MLE estimates for probabilities p_j for different sample size m and repetitions r

exactly we test that probabilities $p_j^{(l)}$ depend only on j, not on l.

$$\mathsf{H}_0: \; p_j^{(1)} = p_j^{(2)} = \cdots = p_j^{(r)} = p_j, \quad j = 1, 2, \ldots, n$$

To check this hypothesis one could estimate how much observed value of the random variable $\mathcal{K}$ differs from its predicted value described in Theorem 3. Actually, this means estimation of tail probabilities of the variable $\mathcal{K}$ which can be carry out using Theorem 1. For instance, if observed numbers of common balls in urns are $(k_1, k_2, \ldots, k_n)$, $k_j \in \{0, 1, \ldots, m\}$ and $k_j \le \mathsf{E}(K_j) = m \cdot p_j^r$ then in order to test hypothesis H_0 one should calculate the probability

$$\mathsf{P}\left(K_1 \le k_1, \ldots, K_n \le k_n\right) = \mathsf{F}(k_1, \ldots, k_n)$$

where F is the probability distribution function of the random variable $\mathcal{K}$. In other cases corresponding probability can be obtained by the direct summation of probabilities $\mathsf{P}(\mathcal{K} = \mathbf{k})$ from Theorem 1 and Corollaries 2 and 3.

Remark. Calculation of tail probabilities of multinomial distribution was considered by Johnson and Kotz [7], Olkin and Sobel [14], see also interesting paper by Levin [11].

References

[1] S. Basu, R. Pollack, M.-F. Roy, *Algorithms in Real Algebraic Geometry*,ser. Algorithms and Computation in Mathematics, **10**. Springer, Berlin, Heidelberg, Germany (2006). DOI: 10.1007/3-540-33099-2.

[2] M. Benaïm, S. J. Schreiber, P. Tarrés, Generalized urn models of evolutionary processes, *Ann. Appl. Probab.*, **14**, No. 3 (2004), 1455–1478. DOI: 10.1214/105051604000000422.

[3] C. A. Charalambides, *Combinatorial methods in discrete distributions*. John Wiley & Sons, Inc., Hoboken, NJ, USA (2005). DOI: 10.1002/0471733180.

[4] S. Eger, Restricted weighted integer compositions and extended binomial coefficients, *Journal of Integer Sequences*, **16** (2013), Article 13.1.3.

[5] F. Hu, W. F. Rosenberger, *The Theory of Response-Adaptive Randomization in Clinical Trials*. John Wiley & Sons, Inc., Hoboken, NJ, USA (2006). DOI: 10.1002/047005588x.

[6] N. L. Johnson, S. Kotz, *Urn Models and Their Applications*. John Wiley & Sons, New-York, USA (1977). ISBN-13: 978-0471446309.

[7] N. L. Johnson, S. Kotz, N. Balakrishnan, *Discrete Multivariate Distributions*. John Wiley & Sons, Inc., Hoboken, NJ, USA (1997). ISBN-13: 978-9812530745.

[8] K. Knoblauch, M. Neitz, J. Neitz, An urn model of the development of l/m cone ratios in human and macaque retinas, *Visual Neuroscience*, **23**, No. 3–4 (2006), 387–394. DOI: 10.1017/s0952523806233157.

[9] V. F. Kolchin, B. A. Sevast'yanov, and V. P. Chistyakov, *Random allocations*, ser. Scripta series in mathematics. Washington: V. H. Winston & Sons, USA (1978). ISBN-13: 978-0470993941.

[10] S. Kotz, N. Balakrishnan, Advances in Urn Models during the Past Two Decades. In *Advances in Combinatorial Methods and Applications to Probability and Statistics*. Boston, MA: Birkhäuser Boston, USA (1997), 203–257. DOI: 10.1007/978-1-4612-4140-9_14.

[11] B. Levin, A representation for multinomial cumulative distribution functions, *Ann. Stat.*, **9**, No. 5 (1981), 1123–1126. DOI: 10.1214/aos/1176345593.

[12] H. M. Mahmoud, *Pólya Urn Models*, ser. CRC texts in statistical science series. **76**, Chapman & Hall/CRC, USA (2009). DOI:10.1201/9781420059847.

[13] R. K. Niven, M. Grendar, Generalized classical, quantum and intermediate statistics and the Pólya urn model, *Phys. Lett. A*, **373** (2010), 645–675. DOI: 10.1016/j.physleta.2008.12.025.

[14] I. Olkin and M. Sobel, Integral expressions for tail probabilities of the multinomial and negative multinomial distributions, *Biometrika*, **52**, No. 1/2 (1965), 167–179. DOI: 10.2307/2333822.

[15] B. Sturmfels, *Solving Systems of Polynomial Equations*. AMS, USA (2002). DOI: 10.1090/cbms/097.

[16] G. S. Wilf, *Generatingfunctionology*. A. K. Peters/CRC Press, USA (2005). DOI: 10.1201/b10576.