ON THE PROBABILITY OF THE EVENT: IN \( n \) GENERALIZED ALLOCATION SCHEMES THE VOLUME OF EACH CELL DOES NOT EXCEED \( r \).

A.I. AFONINA, I.R. KAYUMOV, A.N. CHUPRUNOV

Abstract. We consider \( n \) identical generalized schemes of allocating particles in cells. We study the probability of the event: for each generalized allocation scheme, there are at most \( r \) particles in each cell, where \( r \) is a given number. We obtain an asymptotic estimate for this probability and we consider the application of the obtained results to an antinoise coding.

Keywords: generalized allocation scheme, Cauchy integral, Hamming code.

Mathematics Subject Classification: 60K30

1. Introduction and main results

Let \( \xi, \xi_j, 1 \leq j \leq N \), be independent identically distributed integer-valued random variables. We recall [1] that random variables \( \eta_1, \ldots, \eta_N \) are called a generalized scheme of allocating \( m \) particles in \( N \) cells if their joint distribution is of the form

\[
P\{\eta_1 = k_1, \ldots, \eta_N = k_N\} = P\{\xi_1 = k_1, \ldots, \xi_N = k_N \mid \xi_1 + \cdots + \xi_N = m\},
\]

where \( k_1 + k_2 + \cdots + k_N = m \).

Many allocating schemes of the discrete probability theory such as the scheme of allocating discernible particles in cells, the scheme of allocating indiscernible particles in cells, random permutations, random forest are generalized allocating schemes (for generalized allocating schemes see [2]–[6]).

In what follows, as random variables \( \xi, \xi_j \), we shall consider random variables \( \xi = \xi(x), \xi_j = \xi_j(x), x > 0 \), with the distribution

\[
P(\xi(x) = k) = \frac{a_k x^k}{S(x)}, \quad k = 0, 1, 2, \ldots,
\]

where \( S(x) = \sum_{k=0}^{\infty} a_k x^k \) is the sum of a series with non-negative coefficients having a positive convergence radius \( R \). In this case we shall say that generalized allocating scheme \( \eta_1, \ldots, \eta_N \) is defined by function \( S(x) \).

Random variables \( \xi = \xi(x), \xi_j = \xi_j(x) \) were introduced in work [7]. In works [7–9], there were obtained limiting theorems for the sums of random variables \( \xi_j(x) \).

We consider the event \( A_N(m, r) \) that in a generalized scheme of allocating \( m \) particles in \( N \) cells each cell contains at most \( r \) particles:

\[
A_N(m, r) = \{\omega \in \Omega : \eta_1(\omega) \leq r, \ldots, \eta_N(\omega) \leq r\} = \{\omega \in \Omega : \max_{1 \leq i \leq N} \eta_i(\omega) \leq r\}.
\]

The probability of event \( A_N(m, r) \) has the following representation.
Lemma A. Let \( S_r(z) = \sum_{k=0}^{r} a_k z^k \) be a partial sum of series \( S(z) \), \( a_m(S^N) \) be the \( m \)-th coefficient in the expansion of function \((S(z))^N\), which is the \( N \)-th power of function \( S(z) \), \( a_m(S_r^N) \) be the \( m \)-th coefficient in the expansion of function \((S_r(z))^N\), which is the \( N \)-th power of function. Then

\[
P(A_N(m, r)) = \frac{a_m(S_r^N)}{a_m(S^N)} = \frac{1}{2\pi i} \oint_C \frac{S_r(z)^N}{z^{m+r}} \, dz,
\]

where \( C \) is a closed contour passed in the positive direction whose interior contains no zeroes of functions \( S \) and \( S_r \).

Let \( \eta_1, \ldots, \eta_N \), \( 1 \leq i \leq n \), be the sequence of independent identically distributed generalized schemes of allocating \( m \) particles in \( N \) cells, that is, such that the joint distribution is given by the formula

\[
P\{\eta_1 = k_1, \ldots, \eta_N = k_N\} = P\{\xi_1 = k_1, \ldots, \xi_N = k_N \mid \xi_1 + \cdots + \xi_N = m\},
\]

where \( k_1 + k_2 + \cdots + k_N = m \).

We shall assume that condition \((A_k)\) is satisfied: \( a_0 = \cdots = a_{k-1} = 0, a_k = 1, a_{k+1} > 0 \).

Employing representation (1), in [13] we proved the following theorem.

Theorem B. Assume that condition \((A_0)\) is satisfied and \( m > r \). Then

\[
P(A_n, N(m, r)) = \exp \left[ -m(m-1)\cdots (m-r) \frac{a_{r+1}}{a_{r+1}^\alpha} \left( 1 + O \left( \frac{1}{N} \right) \right) \right],
\]

as \( n, N \to \infty \) so that \( \alpha = o(N^r) \), where \( \alpha \) the ratio of the amount of independent series of particles and the amount of cells, i.e.,

\[
\alpha = \alpha_n N = \frac{n}{N}.
\]

In particular, if \( m > r \), \( n, N \to \infty \) so that \( \frac{n}{N^r} \to \beta \), where \( \beta < \infty \), then

\[
P(A_n, N(m, r)) \to \exp \left[ -m(m-1)\cdots (m-r) \frac{a_{r+1}}{a_{r+1}^\alpha} \beta \right].
\]

The following generalization of Theorem B for arbitrary \( k \) is true.

Theorem 1. Assume that condition \((A_k)\) is satisfied and \( k < r < K < \infty \). We denote \( m_1 = m - kN \). Then

\[
P(A_{n,N}(m, r)) = \exp \left[ -m_1(m_1-1)\cdots (m_1-(r-k)) \frac{a_{r+1}}{a_{r+1}^\alpha} \left( 1 + O \left( \frac{1}{N} \right) \right) \right],
\]

uniformly in \( m_1 \in (k, K] \) as \( m, n, N \to \infty \) so that \( \alpha = o(N^{r-k}) \). In particular, if \( m, n, N \to \infty \) so that \( m_1 \) is fixed, \( m_1 > r \) and \( \frac{n}{N^{r-k}} \to \beta \), where \( \beta < \infty \), then

\[
P(A_{n,N}(m, r)) \to \exp \left[ -m_1(m_1-1)\cdots (m_1-(r-k)) \frac{a_{r+1}}{a_{r+1}^\alpha} \beta \right].
\]
Corollary. Assume that condition \((A_k)\) is satisfied for schemes \(\eta_1, \ldots, \eta_N\) and \(\eta_1^*, \ldots, \eta_N^*\), \(a_{k+1} = a_{k+1}', a_{r-k+1} = a_{r-k+1}', k < r < K < \infty\). Then

\[
P(A_{n,n}(m,r)) = P(A_{n,N}^*(m,r))^{(1+O(\frac{1}{N}))}
\]

uniformly in \(m_1 \in (k,K)\) as \(m,n,N \to \infty\) so that \(\alpha = o(N^{r-k})\).

To prove the corollary, it is sufficient to note that the probabilities \(P(A_{n,n}(m,r))\) and \(P(A_{n,N}^*(m,r))\) satisfy formula (3).

Remark 1. The random variable \(\eta_N = \max_{1 \leq i \leq N} \eta_i\) is called the maximal volume of a cell. Many works were devoted to studying the limiting behavior of the maximal volume of a cell \([2], [6], [17]\). As \(n = 1\), Theorem B and Theorem 1 can be regarded as the limiting theorem for the distribution function of random variable \(\eta_N\) in the case, when \(m_1\) is bounded and \(N \to \infty\).

Let \(0 \leq r_1 < r_2 \leq \infty\). Consider random variables \(\eta_1, \ldots, \eta_N\) having the joint distribution

\[
P\{\eta_1 = k_1, \ldots, \eta_N = k_N\} = P\{\eta_1 = k_1, \ldots, \eta_N = k_N \mid r_1 \leq \min_{1 \leq i \leq N} \eta_i, \max_{1 \leq i \leq N} \eta_i \leq r_2\},
\]

where \(k_1 + k_2 + \cdots + k_N = m, r_1 \leq k_i \leq r_2, 1 \leq i \leq N\).

Theorem 2. Random variables \(\eta_1^*, \ldots, \eta_N^*\) are a generalized scheme of allocating \(m\) particles in \(N\) cells defined by the function \(S_{r_1r_2}(x) = \sum_{i=r_1}^{r_2} a_i x^i\).

Let \(0 < r_2 \leq \infty\). Consider random variables \(\eta_1^{(2)}, \ldots, \eta_N^{(2)}\) with joint distribution

\[
P\{\eta_1^{(2)} = k_1, \ldots, \eta_N^{(2)} = k_N\} = P\{\eta_1 = k_1, \ldots, \eta_N = k_N \mid \max_{1 \leq i \leq N} \eta_i \leq r_2\},
\]

where \(k_1 + k_2 + \cdots + k_N = m, r_1 \leq k_i \leq r_2, 1 \leq i \leq N\).

As \(r_1 = 0\), Theorem 2 implies

Corollary 1. Random variables \(\eta_1^{(2)}, \ldots, \eta_N^{(2)}\) are a generalized scheme of allocating \(m\) particles in \(N\) cells defined by the function \(S_{0r_2}(x) = \sum_{i=0}^{r_2} a_i x^i\).

Let \(0 \leq r_1 < \infty\). Consider random variables \(\eta_1^{(1)}, \ldots, \eta_N^{(1)}\) with joint distribution

\[
P\{\eta_1^{(1)} = k_1, \ldots, \eta_N^{(1)} = k_N\} = P\{\eta_1 = k_1, \ldots, \eta_N = k_N \mid \min_{1 \leq i \leq N} \eta_i \leq r_1\},
\]

where \(k_1 + k_2 + \cdots + k_N = m, r_1 \leq k_i, 1 \leq i \leq N\).

As \(r_2 = \infty\), Theorem 2 implies

Corollary 2. Random variables \(\eta_1^{(1)}, \ldots, \eta_N^{(1)}\) are a generalized scheme of allocating \(m\) particles in \(N\) cells defined by the function \(S_{r_1\infty}(x) = \sum_{i=r_1}^{\infty} a_i x^i\).

We apply Theorem 1 and Theorem 2 to studying the asymptotic behavior of the probability of event \(A_{n,N}(m,r,r_1,r_2)\) that each cell of each scheme contains at most \(r\) particles if we know that each cell of each scheme contains at least \(r_1\) and at most \(r_2\) particles, that is, the probability of event \(A_{n,N}(m,r)\) under the condition: each cell of each scheme contains at least \(r_1\) and at most \(r_2\) particles.

Theorem 3. Assume that \(r_1 < r < r_2, r < K < \infty\) and condition \((A_{r_1})\) is satisfied. We denote \(m_1 = m - r_1 N\). Then

\[
P(A_{n,N}(m,r,r_1,r_2)) = \exp \left[ \frac{-m_1(m_1 - 1) \cdots (m_1 - (r - r_1))}{N^{r-r_1-1}} a_{r+1} \frac{a_{r-r_1-1}}{a_{r-r_1+1}} \alpha \left( 1 + O\left( \frac{1}{N} \right) \right) \right],
\]
uniformly in \( r < m_1 < K \) as \( m, n, N \to \infty \) so that \( \alpha = o(N^{r-r_1}) \). In particular, if \( m, n, N \to \infty \) so that \( m_1 \) is fixed, \( m_1 > r \) and \( \frac{n}{N^{r-r_1}} \to \beta \), where \( \beta < \infty \), then

\[
P(A_{n,N}(m, r, r_1, r_2)) \to \exp \left[ -m_1(m_1 - 1) \cdots (m_1 - (r - r_1)) \frac{a_{r_1+1}}{a_{r_1+1} - r_1+1} \beta \right].
\]

Employing Corollary 1 of Theorem 2 instead of Theorem 2 in the proof of Theorem 3, we obtain

**Corollary 1.** Assume that \( 0 < r < r_2 \) and condition \((A_0)\) is satisfied. Then

\[
P(A_{n,N}(m, r, 0, r_2)) = \exp \left[ \frac{-(m)(m-1) \cdots (m-r)}{N^{r-1}} \frac{a_{r+1}}{a_{r+1} - r+1} \alpha \left( 1 + O \left( \frac{1}{N} \right) \right) \right],
\]
as \( n, N \to \infty \) so that \( \alpha = o(N^r) \). In particular, if \( n, N \to \infty \) so that \( \frac{n}{N^r} \to \beta \), where \( \beta < \infty \), then

\[
P(A_{n,N}(m, r, 0, r_2)) \to \exp \left[ -m(m-1) \cdots (m-r) \frac{a_{r+1}}{a_{r+1} - r+1} \beta \right].
\]

**Remark 2.** It follows from Theorem 3 and Corollary 1 of Theorem 3 that the asymptotic behavior of the probability of event \( A_{n,N}(m, r, r_1, r_2) \) is independent of \( r_2 \).

Probability \( P(A_{n,N}(m, r)) \) has the following application in the theory of antinoise coding. Consider a code, which allows to fix at most \( r \) replacement kind errors in a block. A particular case of such code is Hamming code (for Hamming code see, for instance, [10]). Suppose that we have \( n \) messages. Each message has \( N \) blocks and contains \( m \) errors. We assume that the probability associated with different messages are independent and the errors are distributed in the blocks of the messages in accordance with some generalized allocating scheme. Interpreting errors as particles and cells as blocks, we observe that \( P(A_{n,N}(m, r)) \) is the probability of the even that all the errors in \( n \) messages are fixed.

In work [11], the convergence of probability \( P(A_{n,N}(m, r)) \) was studied in the case of the scheme of allocating discernible particles in different cells. In [12], [13] there was studied the convergence of probability \( P(A_{n,N}(m, r)) \) in the general case. In works [14], [15], the convergence of probabilities of some analogues of events \( A_{n,N}(m, r) \) was treated. As \( r = 1 \) in [16] there was studied the convergence of event \( A_{n,N}(m, r) \), in which the amount of particles in the blocks is random. The probabilities of events \( A_{n,N}(m, r) \) in some analogues of generalized allocating scheme were studied in [17].

Let us consider the application of Theorem 1 to some schemes of probability combinatorics.

**Random forests.** A random forest having \( N \) root and \( m \) non-root vertices is a generalized scheme of allocating \( m \) particles in \( N \) cells with the function

\[
S(z) = \frac{1^{n-1}}{1!} z + \frac{2^{n-1}}{2!} z^2 + \cdots \frac{r^{n-1}}{r!} z^r + \cdots ,
\]

that is, it corresponds to the case \( k = 1 \) (see [2]). This is the probability of the event that in \( n \) random forests, each consisting of \( N \) trees and \( m \) non-root vertices, each tree has at most \( r \) branches is equal

\[
P(A_{n,N}(m, r)) = \exp \left[ \frac{-(m - N)(m - N - 1) \cdots (m - N - (r - 1))(r + 1)^{r-1}}{N^{r-2}r!} \alpha \left( 1 + O \left( \frac{1}{N} \right) \right) \right],
\]
as \( n, N \to \infty \) so that \( \alpha = o(N^{r-1}) \). In particular, if \( m, n, N \to \infty \) so that \( m - N = m_1 \) is fixed and \( \frac{n}{N^{r-2}r!} \to \beta \), where \( \beta < \infty \), then

\[
P(A_{n,N}(m, r)) \to \exp \left[ -\frac{m_1(m_1 - 1) \cdots (m_1 - (r - 1))(r + 1)^{r-1}}{r!} \beta \right].
\]
Cycles in substitutions. A random substitution of degree $m$ containing exactly $N$ cycles is a generalized scheme of allocating $m$ particles in $N$ cells with the function $S(z) = -\ln(1-z)$ (see [2]). We note that $-\ln(1-z) = \sum_{k=1}^{\infty} \frac{z^k}{k}$. This is why the probability of the event that in $n$ random substitutions, each being of degree $m$ and consisting of $N$ cycles, each cycle has the length at most $r$, is equal to

$$
P(A_{n,N}(m,r)) = \exp\left[ \frac{-(m-N)(m-N-1)\cdots(m-N-(r-1))2^r}{N^{r-2}(r+1)} \alpha \left(1 + O\left(\frac{1}{N}\right)\right) \right],$$

as $n, N \to \infty$ so that $\alpha = o(N^{r-1})$. In particular, if $m, n, N \to \infty$ so that $m - N = m_1$ is fixed and $\frac{n}{N^{r-1}} \to \beta$, where $\beta < \infty$, then

$$
P(A_{n,N}(m,r)) \to \exp\left[ \frac{-m_1(m_1-1)\cdots(m_1-(r-1))2^r}{r+1} \beta \right].$$

Random partitions. An equiprobable partition of an integer positive number $m$ in $N$ ordered terms not exceeding $k$ is the generalized allocating scheme with the function

$$S(z) = \frac{z^k}{1-z} = \sum_{i=k}^{\infty} z^i$$

(see [2]), i.e., it satisfies condition $(A_k)$. This is why the probability of the event that in $n$ independent random partitions of an integer positive number $m$ into $N$ ordered terms not exceeding $k$, each element of the partition does not exceed $r$, is equal to

$$
P(A_{n,N}(m,r)) = \exp\left[ \frac{-(m-kN)(m-kN-1)\cdots(m-kN-(r-k))}{N^{r-k-1}} \alpha \left(1 + O\left(\frac{1}{N}\right)\right) \right],$$

as $m, n, N \to \infty$ so that $\alpha = o(N^{r-k})$. In particular, if $m > r$, $n, N \to \infty$ so that $m-kN = m_1$ is fixed, $\frac{n}{N^{r-k}} \to \beta$, where $\beta < \infty$, then

$$
P(A_{n,N}(m,r)) \to \exp\left[-m_1(m_1-1)\cdots(m_1-(r-k))\beta \right].$$

2. Proofs

Proof of Lemma A. We have

$$
P(A_{m,N}(r)) = \frac{P\{\xi_1 \leq r, \xi_2 \leq r, \ldots, \xi_N \leq r \mid \xi_1 + \xi_2 + \cdots + \xi_N = m\}}{P\{\xi_1 + \xi_2 + \cdots + \xi_N = m\}}$$

$$= \frac{\sum P\{\xi_1 = k_1\}P\{\xi_2 = k_2\} \cdots P\{\xi_N = k_N\}}{\sum \{ (k_i) : k_1+k_2+\cdots+k_N=m, k_i \leq r, 1 \leq i \leq N \} P\{\xi_1 = k_1\}P\{\xi_2 = k_2\} \cdots P\{\xi_N = k_N\}}$$

$$= \frac{\sum \{ (k_i) : k_1+k_2+\cdots+k_N=m, k_i \leq r, 1 \leq i \leq N \} \frac{a_{k_1}x^{k_1}}{S(x)} \frac{a_{k_2}x^{k_2}}{S(x)} \cdots \frac{a_{k_N}x^{k_N}}{S(x)}}{a_m(S^N_r)} = \frac{a_m(S^N_r)}{a_m(S^N)}.$$ 

The first identity in (1) is proven. Since

$$a_m(S^N) = \frac{1}{2\pi i} \oint_{C} \frac{S(z)^N}{z^{m+1-2}} dz, \quad a_m(S^N_r) = \frac{1}{2\pi i} \oint_{C} \frac{S_r(z)^N}{z^{m+1}} dz,$$

the second identity in (1) holds true. The proof is complete. \qed
Proof of Theorem 1. By (1) the probability of event \( A_{n,N} \) has the following representation

\[
P(A_{n,N}) = P\left( \bigcap_{i=1}^{n} \{ \eta_i \leq r, \ldots, \eta_N \leq r \} \right) = \left( \frac{1}{2\pi} \int_{C} \frac{S_1(z)^N}{z^{m+1}} \, dz \right)^n.
\]

Hence, employing (5) and (2), where \( m \) and \( r \) are replaced by \( m - kN \) and \( r - k \), respectively, we obtain

\[
P(A_{n,N}(m,r)) = \left( \frac{1}{2\pi} \int_{C} \frac{\sum_{i=1}^{k} a_i z^i}{z^{m+k} + 1} \, dz \right)^n = \exp \left[ \frac{-m_1(m_1 - 1) \cdots (m_1 - (r - k))}{N^{r-k-1}} \frac{a_{r+1}}{a_{k+1}} \alpha \left( 1 + O \left( \frac{1}{N} \right) \right) \right].
\]

Estimate (3) is proven. It implies (4). This completes the proof of Theorem 1. \( \square \)

Proof of Theorem 2. We consider independent identically distributed random variables \( \xi'_i(x) \) with the distribution \( P(\xi'_i(x) = j) = \frac{a_{xj}}{S_{r_1 r_2}(x)} \), \( r_1 \leq j \leq r_2 \). Let \( k_1 + k_2 + \cdots + k_N = m, \ 1 \leq k_i \leq r_2, \ 1 \leq i \leq N \). We have

\[
P\left\{ \xi'_1 = k_1, \ldots, \xi'_N = k_N \right\} = \prod_{i=1}^{N} a_{k_i} \frac{x^{k_i}}{S(x)}
\]

\[
= \frac{\prod_{i=1}^{N} a_{k_i} x^{k_i}}{S_{r_1 r_2}(x)} \sum_{k_1+k_2+\cdots+k_N=m} \prod_{i=1}^{N} \frac{a_{k_i} x^{k_i}}{S(x)}
\]

\[
= \frac{\prod_{i=1}^{N} a_{k_i} x^{k_i}}{S_{r_1 r_2}(x)} \sum_{k_1+k_2+\cdots+k_N=m, r_1 \leq k_i \leq r_2} \prod_{i=1}^{N} a_{k_i} x^{k_i}
\]

\[
= \frac{\prod_{i=1}^{N} a_{k_i} x^{k_i}}{S_{r_1 r_2}(x)} \sum_{k_1+k_2+\cdots+k_N=m} \prod_{i=1}^{N} \frac{a_{k_i} x^{k_i}}{S_{r_1 r_2}(x)}
\]

\[
= \prod_{i=1}^{N} a_{k_i} \frac{x^{k_i}}{S(x)}
\]

The proof is complete. \( \square \)

Proof of Corollary 1 of Theorem 2. Since

\[
\{ \max_{1 \leq i \leq N} \eta_i \leq r_2 \} = \{ \min_{1 \leq i \leq N} \eta_i, \ \max_{1 \leq i \leq N} \eta_i \leq r_2 \},
\]
the distribution of random vector \( \eta^{(2)}_1, \ldots, \eta^{(2)}_N \) coincides with the distribution of random vector \( \eta_1, \ldots, \eta_N \) corresponding to the case \( r_1 = 0 \). Hence, we can apply Theorem 2 and this completes the proof.

\( \square \)

**Proof of Theorem 3.** By Theorem 2 the identity

\[
\mathbb{P}(A_{n,N}(m,r_1,r_2)) = \left( \mathbb{P}\{ \max_{1 \leq i \leq N} \eta_i \leq r \mid r_1 \leq \min_{1 \leq i \leq N} \eta_i, \max_{1 \leq i \leq N} \eta_i \leq r_2 \} \right)^n
\]

\[
= \left( \mathbb{P}\{ \max_{1 \leq i \leq N} \xi_i' \leq r \mid \xi_1' + \cdots + \xi_N = m \} \right)^n
\]

holds true. Applying Theorem 1 to the last expression in the above identity, we complete the proof.

\( \square \)

**BIBLIOGRAPHY**

Afonina Alexandra Igorevna,
Kazan Federal University,
Lobachevskii Institute of Mathematics and Mechanics,
Kremlevskaya str. 35,
420008, Kazan, Russia
E-mail: sanyagirl89@mail.ru

Kayumov Il’giz Rifatovich,
Kazan Federal University,
Lobachevskii Institute of Mathematics and Mechanics,
Kremlevskaya str. 35,
420008, Kazan, Russia
E-mail: ikayumov@kpfu.ru

Chuprunov Alexei Nikolaevich,
Kazan Federal University,
Lobachevskii Institute of Mathematics and Mechanics,
Kremlevskaya str. 35,
420008, Kazan, Russia
E-mail: Alexey.Chuprunov@ksu.ru