On cell loads in an analogue of the generalized allocation scheme

Elena Khvorostyanskaya

Institute of Applied Mathematical Research
Karelian Research Center RAS
Petrozavodsk, Russia

III Russian Finnish Symposium on Discrete Mathematics
Petrozavodsk, Russia, September 15-18, 2014

Generalized allocation scheme

- V. F. Kolchin. *Random mapping.* New York, Springer, 1986.

Let $\eta_1, \ldots, \eta_N$ be integer nonnegative random variables and $\eta_1 + \ldots + \eta_N = n$. If there are integer nonnegative independent identically distributed random variables $\xi_1, \ldots, \xi_N$ such that for any integer $k_1, \ldots, k_N \leqslant 0$, $k_1 + \ldots + k_N = n$

$$\mathbf{P}\{\eta_1 = k_1, \ldots, \eta_N = k_N\} = \mathbf{P}\{\xi_1 = k_1, \ldots, \xi_N = k_N | \, \xi_1 + \ldots + \xi_N = n\}, \quad (1)$$

then the random variables $\eta_1, \ldots, \eta_N$ and $\xi_1, \ldots, \xi_N$ form a generalized allocation scheme.

One Poisson allocation scheme

- E.V. Khvorostyanskaya. On random Poisson allocation to cells // Transactions of Karelian Research Centre of Russian Academy of Science. 1. Mathematical Modeling and Information Technologies. Vol.4. Petrozavodsk: KarRC RAS, 2013. Pp. 112-116.
- E.V. Khvorostyanskaya, Y.L. Pavlov. Limit distributions of the maximum filling of cells in one allocation scheme // European Researcher, Vol.76, 61, 2014. Pp.1019-1027.

Let particles allocate to N cells numbered from 1 to N. Random variables $\xi_1, \ldots, \xi_N$ are equal to the numbers of particles in cells and they are independent and follow Poisson distribution

$$p_k = \mathbf{P}\{\xi = k\} = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, 2, \ldots$$

Let H be a subset of realizations that satisfy the condition

$$\xi_1 + \ldots + \xi_N \leqslant n \tag{2}$$

and let random variables $\eta_1, \ldots, \eta_N$ be equal to the numbers of particles in cells on the subset H.

An analogue of the generalised allocation scheme

$$\mathbf{P}\{\eta_1 = k_1, \ldots, \eta_N = k_N\}$$
$$= \mathbf{P}\{\xi_1 = k_1, \ldots, \xi_N = k_N \mid \xi_1 + \ldots + \xi_N \leqslant n\}, \tag{3}$$

- A.N. Chuprunov, I. Fazekas. An analogue of the generalised allocation scheme: limit theorems for the number of cells containing a given number of particles. // Discrete Mathematics and Applications, 2012, vol.22, iss.1, p. 101–122.
- A.N. Chuprunov, I. Fazekas. An analogue of the generalised allocation scheme: limit theorems for the maximum cell load. // Discrete Mathematics and Applications, 2012, vol.22, iss.3, p. 307–314.

$$\mathbf{P}\left\{\mu_r = k\right\} = \binom{N}{k} p_r^k \left(1-p_r\right)^{N-k} \frac{\mathbf{P}\left\{\zeta_{N-k}^{(r)} \leqslant n-kr\right\}}{\mathbf{P}\left\{\zeta_N \leqslant n\right\}}, \qquad (4)$$

where $p_r = \mathbf{P}\{\xi_1 = r\}$, $\zeta_N = \xi_1 + \ldots + \xi_N$, $\zeta_{N-k}^{(r)} = \xi_1^{(r)} + \ldots + \xi_{N-k}^{(r)}$, random variables $\xi_1^{(r)}, \ldots, \xi_N^{(r)}$ are independent and follow distributions

$$\mathbf{P}\left\{\xi_i^{(r)} = k\right\} = \mathbf{P}\left\{\xi_1 = k |\ \xi_1 \neq r\right\}, \quad i = 1, \ldots, N.$$

$$\mathbf{P}\left\{\eta_{(N)} \leqslant r\right\} = \left(1-P_r\right)^N \frac{\mathbf{P}\left\{\overset{\sim(r)}{\zeta}_N \leqslant n\right\}}{\mathbf{P}\left\{\zeta_N \leqslant n\right\}}, \qquad (5)$$

where $P_r = \mathbf{P}\left\{\xi_1 > r\right\}$, $\zeta_N = \xi_1 + \ldots + \xi_N$, $\overset{\sim(r)}{\zeta}_N = \overset{\sim(r)}{\xi}_1 + \ldots + \overset{\sim(r)}{\xi}_N$, random variables $\overset{\sim(r)}{\xi}_1, \ldots, \overset{\sim(r)}{\xi}_N$ are independent and follow distributions

$$\mathbf{P}\left\{\overset{\sim(r)}{\xi}_i = k\right\} = \mathbf{P}\left\{\xi_1 = k | \xi_1 \leqslant r\right\}, \quad i = 1, \ldots, N.$$

$$y = \frac{n - \lambda N}{\sqrt{\lambda N}}, \quad f_r(\lambda) = \frac{(\lambda - r)^2 p_r}{\lambda}$$

Theorem 1.

Let $N \to \infty$ *and* $y^2 \left(f_r(\lambda) + \sqrt{p_r / N} \right) \to 0$ *as* $y \to -\infty$ *and one of the following conditions*

1) $r \geqslant 0,\ \lambda \to \infty,\ \lambda^3/N \to 0;$
2) $r \to \infty,\ 0 < \lambda_1 \leqslant \lambda \leqslant \lambda_2 < \infty;$
3) $r \geqslant 2,\ \lambda \to 0;$
4) $r = 1,\ y \to \infty,\ \lambda \to 0,\ \lambda^2 N \to \infty.$

be satisfied. Then

$$\mathbf{P}\{\mu_r = k\} = \frac{(Np_r)^k}{k!} e^{-Np_r} (1 + o(1))$$

uniformly in the integers k *such that* $(k - Np_r)/\sqrt{Np_r}$ *lies in any finite fixed interval.*

Theorem 2.

Let $N \to \infty$, $0 < \lambda_1 \leqslant \lambda \leqslant \lambda_2 < \infty$ and let $r \geqslant 0$ be fixed and one of the following conditions

1) $y \to +\infty$;
2) $|y| \leqslant C < \infty$, $r - \lambda \to 0$;
3) $y \to -\infty$, $r - \lambda \to 0$, $y(r - \lambda) \to 0$, $y^2/\sqrt{N} \to 0$

be satisfied. Then

$$\mathbf{P}\{\mu_r = k\} = \frac{1}{\sqrt{2\pi N p_r (1 - p_r)}} e^{-u^2/2}(1 + o(1))$$

uniformly in the integers k such that $u = (k - Np_r)/\sqrt{Np_r(1 - p_r)}$ lies in any finite fixed interval.

Limit distributions of the number of cells containing a giving number of particles in the Poisson allocation scheme

Theorem 3.

Let $r = 1$ and $n \geqslant 1$ is fixed, $N \to \infty$, $\lambda \to 0$ in such a way that $\lambda N \leqslant C < \infty$. Then for nonnegative integer k

$$\mathbf{P}\left\{\mu_1 = k\right\} = \frac{(\lambda N)^k}{k!}\left(\sum_{i=0}^{n}\frac{(\lambda N)^i}{i!}\right)^{-1}(1 + o(1))$$

when $0 \leqslant k \leqslant n$ and $\mathbf{P}\left\{\mu_1 = k\right\} = 0$ if $k > n$.

Theorem 4.

Let N, $n \to \infty$, $\lambda N \to \infty$, $\lambda^3/N \to 0$, $y \geqslant -C > \infty$ *and* r *is chosen so that* $Np_{r-1} \to \infty$, $Np_r \to \alpha$ *where* α *is a positive constant. Then*

$$\mathbf{P}\left\{\eta_{(N)} = r-1\right\} \to e^{-\alpha}, \quad \mathbf{P}\left\{\eta_{(N)} = r\right\} \to 1 - e^{-\alpha}.$$

Theorem 5.

Let N, $n \to \infty$, $\lambda N \to \infty$, $\lambda^3/N \to 0$, $y \to \infty$, $y/N^{1/6} \to 0$ *and* $r \geqslant 3$ *is chosen so that* $y^2 r/N \to 0$, $Np_{r-1} \to \infty$, $Np_r \to \alpha$ *where* α *is a positive constant. Then*

$$\mathbf{P}\left\{\eta_{(N)} = r-1\right\} = exp\left\{-\frac{\alpha}{2}\left(1 + \left(1 + \frac{nr}{\lambda N} - r\right)^2\right)\right\} + o(1),$$

$$\mathbf{P}\left\{\eta_{(N)} = r\right\} = 1 - exp\left\{-\frac{\alpha}{2}\left(1 + \left(1 + \frac{nr}{\lambda N} - r\right)^2\right)\right\} + o(1).$$

Theorem 6.

Let $N,\ n \to \infty$, $\lambda / \ln N \to x$ and r is chosen so that $r > \lambda$ and $Np_r \to \alpha$ where x, α are positive constants and let $yN^{-1/6} \to 0$ as $y \to -\infty$. Then

$$\mathbf{P}\left\{\eta_{(N)} \leqslant r + k\right\} \to \exp\left\{-\frac{\alpha\gamma^{k+1}}{1-\gamma}\right\},$$

where k is a fixed integer and γ is a root of the equation
$\gamma + x(\ln\gamma - \gamma + 1) = 0$, $0 < \gamma < 1$.

Theorem 7.

Let $N,\ n \to \infty$, $\lambda^3/N \to 0$, $\lambda / \ln N \to \infty$ and let $yN^{-1/6} \to 0$ as $y \to -\infty$. Then

$$\mathbf{P}\left\{\frac{\eta_{(N)} - \lambda - \lambda u\left(\lambda^{-1}\left(\ln N - (\ln\ln N)/2\right)\right)}{\sqrt{\lambda/2\ln N}} + \frac{\ln 4\pi}{2} \leqslant z\right\} \to e^{-e^{-z}}$$

where $u(w)$ is a positive function defined on the interval $0 < w < \infty$ by the equation $-u + (1+u)\ln(1+u) = w$.

Theorem 8.

Let $N \to \infty$, n be a fixed number. Then

$$\mathbf{P}\left\{\eta_{(N)} = 0\right\} = \left(\sum_{k=o}^{n} \frac{(\lambda N)^k}{k!}\right)^{-1},$$

$$P\left\{\eta_{(N)} = 1\right\} = 1 - \left(\sum_{k=o}^{n} \frac{(\lambda N)^k}{k!}\right)^{-1} + o(1).$$

Theorem 9.

Let $n,\ N \to \infty$, $\lambda N \leqslant C < \infty$. Then

$$\mathbf{P}\left\{\eta_{(N)} = 0\right\} = e^{-\lambda N} + o(1), \quad \mathbf{P}\left\{\eta_{(N)} = 1\right\} = 1 - e^{-\lambda N} + o(1).$$

Random forests

Let $F_{N,n}$ be a set of different forests consisting of N rooted trees with roots $1, 2, \ldots, N$ and n nonroot labeled vertices. The uniform probability distribution is defined on $F_{N,n}$.

- Pavlov Yu. L. Limit theorems for the number of trees of a given size in a random forest. Mathematics of the USSR-Sbornik. 1977. V. 32. Iss. 3. P. 335–345.
- Pavlov Yu. L. The asymptotic distribution of maximum tree size in a random forest. Theory of Probability and its Applications. 1978. V. 22. Iss. 3. P. 509–520.
- Pavlov Yu. L. A case of limit distribution of the maximal volume on a tree in a random forest. Mathematical notes. 1979. V. 25. Iss. 5. P. 387–392.

Random forests

For a random forest from the set $F_{N,n}$ the generalized allocation scheme is formed by random variables $\eta_1, \dots, \eta_N$ that are equal to the numbers of non-root vertices of the trees and the independent random variables $\xi_1, \dots, \xi_N$ following distributions

$$\mathbf{P}\{\xi_i = k\} = \frac{(k+1)^k x^{k+1}}{(k+1)!\theta(x)}, \quad k = 0, 1, 2, \dots, \quad i = 1, \dots, N, \tag{6}$$

$$0 < x \leqslant e^{-1}, \quad \theta(x) = \sum_{k=1}^{\infty} \frac{k^{k-1}}{k!} z^k.$$

Galton-Watson forests

$G_N = \{\mu(t),\ t = 0, 1, \ldots\}$ is a subcritical or critical Galton-Watson process starting with N particles and a number of offspring of each particle is a random variable ξ with distribution

$$\mathbf{P}\{\xi = k\} = p_k, \quad k = 0, 1, 2, \ldots \tag{7}$$

A trajectory T_N of the process G_N is a directed forest consisting of N trees and m_k vertices with k outgoing arcs. The process G_N has the trajectory T_N with the probability

$$\mathbf{P}(T_N) = \prod_{k=0}^{\infty} p_k^{m_k}.$$

- Pavlov Yu. L. *Random forests.* VSP, Utrecht, 2000.

Conditional Poisson Galton–Watson forests

We consider a subcritical or critical Galton-Watson process with N initial particles and Poisson offspring distribution

$$p_k = \mathbf{P}\{\xi = k\} = \frac{\lambda^{k-1}}{k!} e^{-\lambda}, \quad k = 0, 1, 2, \ldots \tag{8}$$

Independent random variables $\xi_1, \ldots, \xi_N$ are equal to the sizes of trees in the Galton-Watson forest and follow Borel-Tanner distribution

$$p_k(\lambda) = \mathbf{P}\{\xi_i = k\} = \frac{(\lambda k)^{k-1}}{k!} e^{-\lambda k}, \quad k = 1, 2, \ldots, \quad 0 < \lambda \leqslant 1. \tag{9}$$

We consider the subset of the Galton-Watson forests such that

$$\xi_1 + \ldots + \xi_N \leqslant n.$$

Random variables $\eta_1, \ldots, \eta_N$ are equal to the sizes of trees in a conditional Poisson Galton-Watson forest. Let μ_r be a random variable equal to the number of trees with r vertices in a conditional Poisson Galton-Watson forest.

Theorem 10.

Let $N \to \infty$ and one of the following conditions

1) $r \to \infty$, $(1-\lambda)N \to \gamma$, $0 \leqslant \gamma < \infty$, $n/N^2 \geqslant C > 0$;
2) $r \to \infty$, $\lambda \geqslant \lambda_1 > 0$, $(1-\lambda)N \to \infty$, $\sqrt{1-\lambda}\,(N - n(1-\lambda)) \leqslant C\sqrt{N}$, $C \geqslant 0$;
3) $r \geqslant 3$, $\lambda \to 0$, $N\lambda^3 \to \infty$, $N - n(1-\lambda) \leqslant C\sqrt{\lambda N}$,, $C \geqslant 0$;
4) $r = 2$, $\lambda \to 0$, $N\lambda^6 \to \infty$, $(n(1-\lambda) - N)/\sqrt{\lambda N} \to \infty$

be satisfied. Then

$$\mathbf{P}\{\mu_r = k\} = \frac{(Np_r(\lambda))^k}{k!} e^{Np_r(\lambda)}(1 + o(1))$$

uniformly in the integers k such that $(k - Np_r(\lambda))/\sqrt{Np_r(\lambda)}$ lies in any finite fixed interval.

Theorem 11.

Let $N \to \infty$, $Np_r(\lambda)(1 - p_r(\lambda)) \to \infty$ *and one of the following conditions*

1) $r = 1, 2$, $\lambda \to 0$, $N\lambda^{2r+2} \to \infty$, $(n(1-\lambda) - N)/\sqrt{\lambda N} \to \infty$;
2) $r \geqslant 3$, $\lambda \to 0$, $N\lambda^3 \to \infty$, $N - n(1-\lambda) \leqslant C\sqrt{\lambda N}$,, $C \geqslant 0$;
3) $r \to \infty$, $0 < \lambda_1 \leqslant \lambda \leqslant \lambda_2 < 1$, $N - n(1-\lambda) \leqslant C\sqrt{N}$, $C \geqslant 0$;
4) $r \geqslant 1$ *is fixed*, $0 < \lambda_1 \leqslant \lambda \leqslant \lambda_2 < 1$, $(n(1-\lambda) - N)/\sqrt{N} \to \infty$;
5) $r \geqslant 2$ *is fixed*, $\lambda \to 1 - r^{-1}$, $|n(1-\lambda) - N|/\sqrt{N} \leqslant C$, $C \geqslant 0$;
6) $\lambda \to 1$, $(1-\lambda)N \to \infty$, $\sqrt{1-\lambda}(N - n(1-\lambda)) \leqslant C\sqrt{N}$, $C \geqslant 0$;
7) $(1-\lambda)N \to \gamma$, $0 \leqslant \gamma < \infty$, $n/N^2 \geqslant C > 0$;
8) $r \geqslant 2$ *is fixed*, $\lambda = 1 - r^{-1}$, $(n - Nr)/\sqrt{N} \to -\infty$, $n = Nr + o\left(N^{2/3}\right)$

be satisfied. Then

$$\mathbf{P}\{\mu_r = k\} = \frac{1 + o(1)}{\sqrt{2\pi Np_r(\lambda)(1 - p_r(\lambda))}} e^{-u^2/2}$$

uniformly in the integers k *such that* $u = (k - Np_r(\lambda))/\sqrt{Np_r(\lambda)(1 - p_r(\lambda))}$ *lies in any finite fixed interval.*

Let $\eta_{(N)}$ be a random variable equal to the maximum tree size in a conditional Poisson Galton–Watson forest.

Theorem 12.

Let $N \to \infty$, $N\sqrt{1-\lambda} - n(1-\lambda)^{3/2} \leqslant C\sqrt{\lambda N}$, $0 \leqslant C < \infty$, $r \geqslant 5$, $Np_{r-1}(\lambda) \to \infty$, $Np_r(\lambda) \to \alpha$ *where* α *is a positive constant. Then*

$$\mathbf{P}\left\{\eta_{(N)} = r\right\} = 1 - e^{-\alpha} + o(1), \qquad \mathbf{P}\left\{\eta_{(N)} = r-1\right\} = e^{-\alpha} + o(1).$$

Theorem 13.

Let $N \to \infty$, $0 < \lambda_1 \leqslant \lambda \leqslant \lambda_2 < 1$, $N - n(1-\lambda) \leqslant C\sqrt{N}$, $0 \leqslant C < \infty$, $Np_r(\lambda) \to \alpha$ *where* α *is a positive constant. Then for any fixed* k

$$\mathbf{P}\left\{\eta_{(N)} \leqslant r+k\right\} \to \exp\left\{-\frac{\alpha\left(\lambda e^{1-\lambda}\right)^{k+1}}{1-\lambda e^{1-\lambda}}\right\}.$$

Theorem 14.

Let $N \to \infty$, $\lambda \to 1$, $(1-\lambda)N \to \infty$, $N\sqrt{1-\lambda} - n(1-\lambda)^{3/2} \leqslant C\sqrt{N}$, $0 \leqslant C < \infty$. *Then for any fixed* $z > 0$

$$\mathbf{P}\left\{\eta_{(N)}\beta - u \leqslant z\right\} \to e^{-e^{-z}}$$

where $\beta = -\ln\lambda + \lambda - 1$ *and* u *is chosen so that* $N\sqrt{\beta}u^{-3/2}e^{-u} = \sqrt{2\pi}$.

Theorem 15.

Let $N \to \infty$, $(1-\lambda)N \to \gamma$, $0 < \gamma < \infty$, $n/N^2 \geqslant C > 0$. *Then for any fixed* $z > 0$

$$\mathbf{P}\left\{\frac{\eta_{(N)}}{N^2} \leqslant z\right\} = e^{-Q(z,\gamma)} \frac{\int\limits_0^{n/N^2} g_z(x)dx}{\int\limits_0^{n/N^2} g(x)dx}(1+o(1))$$

where

$$Q(z,\gamma) = \sqrt{\frac{2}{\pi z}} e^{-z\gamma^2/2} - \gamma\sqrt{\frac{2}{\pi}} \int\limits_{\gamma\sqrt{z}}^{\infty} e^{-y^2/2} dy,$$

$$g(x) = \frac{1}{x\sqrt{2\pi x}} \exp\left\{\gamma - \frac{\gamma^2 x}{2} - \frac{1}{2x}\right\},$$

$g_z(x)$ *is a density of the distribution with the characteristic function*

$$\exp\left\{\gamma - \sqrt{\gamma^2 - 2it} + Q(z,\gamma)\right\}\left(1 - \frac{1}{\sqrt{2\pi}} \int\limits_z^{\infty} y^{-3/2} e^{ity - \gamma^2 y/2} dy\right).$$

Theorem 16.

Let $N \to \infty$, $(1-\lambda)N = o(1)$, $n/N^2 \geqslant C > 0$. Then the following assertions hold.

1. *If $n/N^2 \to \alpha$, $0 < \alpha < \infty$ then for any fixed $z > 0$*

$$\mathbf{P}\left\{\frac{\eta_{(N)}}{N^2} \leqslant z\right\} \to 1 + \frac{\sum\limits_{k=1}^{[\alpha/z]} \frac{(-1)^k}{k!} \int\limits_{kz}^{\alpha} I_k(z,y)dy}{\int\limits_{0}^{\alpha} y^{-3/2} e^{-1/2y} dy}$$

where

$$I_k(z,y) = \int\limits_{x_k(z,y)} \frac{\exp\left\{-1/2\left(y - x_1 - \ldots - x_k\right)\right\} dx_1 \ldots dx_k}{(2\pi)^{k/2}\left(x_1 \ldots x_k\left(y - x_1 - \ldots - x_k\right)\right)^{3/2}},$$

$$x_k(z,y) = \{x_i \geqslant z,\ i = 1, \ldots, k,\ x_1 + \ldots + x_k \leqslant y\}, \quad k = 1, 2, \ldots$$

2. *If $n/N^2 \to \infty$ then for any fixed $z > 0$*

$$\mathbf{P}\left\{\frac{\eta_{(N)}}{N^2} \leqslant z\right\} \to 1 + \frac{1}{\sqrt{2\pi}} \sum_{k=1}^{\infty} \frac{(-1)^k}{k!} \int\limits_{kz}^{\infty} I_k(z,y)dy.$$