Liet. matem. rink., **44**, *spec. nr.*, 2004, 95–100

The law of iterated logarithm for the Ewens sampling formula

Jolita NORKŪNIENĖ (VU, VIKO)
e-mail: jolita.norkuniene@mif.vu.lt

1. Introduction

Let σ be a permutation in the symmetric group $\mathbb{S}_n$ and

$$\sigma = \kappa_1 \kappa_2 \ldots \kappa_\omega \tag{1}$$

be its unique representation via independent cycles κ, where $\omega = \omega(\sigma)$ denotes the number of cycles. Let k_j, $1 \leqslant j \leqslant n$ be the number of cycles of length j in representation (1). Then $\omega(\sigma) = k_1 + \cdots + k_n$. The vector $\bar{k} = (k_1, \ldots, k_n)$ is called the structure vector of σ and it satisfies the relation

$$k_1 + 2k_2 + \cdots + nk_n = n. \tag{2}$$

We define a measure on the symmetric group $\mathbb{S}_n$ by

$$\nu_{n,\theta}(\sigma) = \frac{\theta^{\omega(\sigma)}}{\theta_{(n)}}$$

for each $\sigma \in \mathbb{S}_n$, where $\theta > 0$ and $\theta_{(n)} = \theta(\theta+1)\ldots(\theta+n-1)$. Then the probability to take σ with the structure vector $\bar{k}$ equals

$$\nu_{n,\theta}(k_1, \ldots, k_n) = \frac{n!}{\theta_{(n)}} \prod_{j=1}^{n} \left(\frac{\theta}{j}\right)^{k_j} \frac{1}{k_j!}. \tag{3}$$

Formula (3) is called the Ewens sampling formula. It firstly appeared in the statistical models of population genetics investigated by Ewens (1972). More about the Ewens sampling formula and its connection to population genetics one can find in monographs by Ewens (1979) and Kingman (1980) (see [1] for more information). It is known, that the asymptotic distribution of $k_j(\cdot)$ for a fixed $j \geqslant 0$ is Poisson with parameter θ/j. The relation (2) makes $k_j(\cdot)$ dependent random variables, satisfying the conditioning relation

$$\nu\big(k_1(\sigma) = k_1, \ldots, k_n(\sigma) = k_n\big) = P\big(\xi_1 = k_1, \ldots, \xi_n = k_n | \zeta = n\big) \tag{4}$$

for $\bar{k} = (k_1, \ldots, k_n) \in \mathbb{Z}^{+n}$. Here ξ_j, $1 \leqslant j \leqslant n$ are independent Poisson random variables with $E\xi_j = \theta/j$ and $\zeta = 1\xi_1 + \cdots + n\xi_n$. Dependence of $k_j(\cdot)$ is rather strong for $\varepsilon n \leqslant j < n$. The strong convergence and the law of iterated logarithm for random permutations was established by E. Manstavičius [4]. Functional limit theorems for partial sum processes were investigated by Babu and Manstavičius [2], [3]. Using the methods of probabilistic number theory and technics of E. Manstavičius we investigate the strong convergence of random variables $k_j(\sigma)$ with distribution (3) satisfying the conditioning relation (4).

2. Results

Let $\mathbb{G}$ be an additive abelian group. A map $h\colon \mathbb{S}_n \to \mathbb{G}$ is called an additive function if it satisfies the relation

$$h(\sigma) = \sum_{j=1}^{n} h_j\bigl(k_j(\sigma)\bigr) \tag{5}$$

for each $\sigma \in \mathbb{S}_n$, where $h_j(0) = 0$ and $h_j(k)$, $j \geqslant 1$, $k \geqslant 1$ is some double sequence in $\mathbb{G}$. Let

$$h(\sigma, m) := \sum_{j=1}^{m} h_j\bigl(k_j(\sigma)\bigr), \quad A(m) := \theta \sum_{j=1}^{m} \frac{a_j}{j}, \quad B(m) := \theta \sum_{j=1}^{m} \frac{a_j^2}{j}. \tag{6}$$

Here $a_j := h_j(1)$. Also let ξ_j be independent Poisson random variables defined above,

$$\Xi_n = h_1(\xi_1) + \cdots + h_n(\xi_n), \quad S_n = a_1\xi_1 + \cdots + a_n\xi_n. \tag{7}$$

As in [4], we compare distributions of $h(\sigma, n)$ with distribution of Ξ_n or S_n.

THEOREM 1. *Let $\alpha(m), \beta(m)$ be real sequences, $\beta(m) > 0$, $\beta(m) \uparrow \infty$, as $m \to \infty$. Then the following relations are equivalent:*

$$\lim_{x\to\infty} \overline{\lim_{n\to\infty}} \nu_{n,\theta}\Bigl(\max_{x\leqslant m\leqslant n} \beta(m)^{-1}\bigl|h(\sigma, m) - \alpha(m)\bigr| \geqslant \varepsilon\Bigr) = 0 \tag{8}$$

for each $\varepsilon > 0$ and

$$\beta(n)^{-1}\bigl(S_n - \alpha(n)\bigr) \to 0 \quad P-a.s. \tag{9}$$

COROLLARY 1. *If $\beta(m) \to \infty$ and the series*

$$\sum_{j=1}^{\infty} \frac{|a_j|^p}{j\beta(j)^p}$$

converges for some $1 \leqslant p \leqslant 2$, then relation (8) *holds with $\alpha(m) = A(m)$.*

For the next theorem, we need more definitions. Let $Z_n := (S_n - A(n))/\beta(n)$. We write $Z_n \Rightarrow [-1; 1]$ if the sequence Z_n is relatively compact and the set of limit points is the interval $[-1; 1]$ with probability one. Denote $\beta(n) = (2B(n)LLB(n))^{1/2}$, where $Lu := \log\max(u, e)$, $u \in \mathbb{R}$, and

$$f(\sigma, m) = \frac{h(\sigma, m) - A(m)}{\beta(m)}.$$

We now state the law of iterated logarithm.

THEOREM 2. *Suppose $B(n) \to \infty$ and there exists a sequence $r = r(n) \to \infty$ such that $B(n) = o(\beta(r))$. Then the following assertions are equivalent:*

$$Z_n \Rightarrow [-1, 1] \quad P - a.s. \tag{10}$$

and

$$\lim_{x\to\infty} \overline{\lim_{n\to\infty}} \nu_{n,\theta}\Big(\max_{x\leqslant m\leqslant n} |f(\sigma, m)| \geqslant 1 + \delta\Big) = 0$$

but

$$\lim_{x\to\infty} \underline{\lim_{n\to\infty}} \nu_{n,\theta}\Big(\min_{x\leqslant m\leqslant n} |f(\sigma, m) - b| < \delta\Big) = 1$$

for each $b \in [-1, 1]$ and $\delta > 0$.

THEOREM 3. *Let $j(\sigma, 1) < \ldots < j(\sigma, s)$ be all different lengths of cycles in decomposition* (1), *$s = s(\sigma)$.*
Then

$$\lim_{x\to\infty} \overline{\lim_{n\to\infty}} \nu_{n,\theta}\Big(\max_{x\leqslant m\leqslant s} \frac{|\log j(\sigma, k) - k|}{(2kLLk)^{1/2}} \geqslant 1 + \delta\Big) = 0$$

and

$$\lim_{x\to\infty} \underline{\lim_{n\to\infty}} \nu_{n,\theta}\Big(\min_{x\leqslant m\leqslant s} \Big|\frac{\log j(\sigma, k) - k}{(2kLLk)^{1/2}} - b\Big| < \delta\Big) = 1$$

for each $b \in [-1, 1]$ and for each $\delta > 0$.

For the proofs of Theorems 1, 2 and 3, we need some auxiliary results.

FUNDAMENTAL LEMMA (R. Arratia, A.D. Barbour and S. Tavaré [1]). *Let ξ_j, $j \geqslant 1$ be the Poisson random variables defined above. Then*

$$\nu_n\Big(\big(k_1(\sigma), \ldots, k_b(\sigma)\big) \in A\Big) - P\big((\xi_1, \ldots, \xi_b) \in A\big) = O(n^{-1}b)$$

uniformly in $A \subset \mathbb{Z}^{+b}$.

Let $L(I)$ be the linear space of real functions $g(\cdot)$ with $\sup_{t\in I} |g(t)| < \infty$, $I \subset \mathbb{R}$. Denote $\theta \wedge 1 := \min(\theta, 1)$.

LEMMA 1 (G.J. Babu, E. Manstavičius [3]). *Let $h(\sigma,t)$, $t\in I\subset\mathbb{R}$, be a set of real valued additive functions defined by* (5), *where $h_j(k,\cdot)\in L(I)$, $k\geqslant 0$, $h_j(0,t)=0$, for $j\leqslant n$, $t\in I$ and $\Xi_n(t)=h_1(\xi_1,t)+\cdots+h_n(\xi_n,t)$. Then*

$$\nu_{n,\theta}\Big(\sup_{t\in I}\big|h(\sigma,t)-a(t)\big|\Big)\leqslant C(\theta)\Big(P^{\theta\wedge 1}\Big(\sup_{t\in I}\big|\Xi_n(t)-a(t)\big|\geqslant u/3\Big)+n^{-\theta}\Big)$$

for each function $a(\cdot)\in L(I)$, $u\geqslant 0$ and a positive constant $C(\theta)$ depending only on θ.

3. Proofs

Proof of Theorem 1. As in [4], at first we notice, that it is suffices to consider the linear function

$$\hat{h}(\sigma,m):=\sum_{j=1}^{m}a_jk_j(\sigma).$$

Following E. Manstavičius and G.J. Babu (see the proof of Theorem 1 in [2]), we have

$$\nu_{n,\theta}\Big(\max_{x\leqslant m\leqslant n}\beta(m)^{-1}\big|h(\sigma,m)-\hat{h}(\sigma,m)\big|\geqslant\varepsilon\Big)$$

$$\leqslant\nu_{n,\theta}\Big(\sum_{j\leqslant n}\big|h_j(k_j(\sigma))-a_jk_j(\sigma)\big|\geqslant\varepsilon\beta(x)\Big)=\mathrm{o}(1),$$

From the Lemma 9.2.5 [6], one can see that relation (9) is equivalent to

$$P\Big(\sup_{m\geqslant x}\beta(m)^{-1}\big|S(m)-\alpha(m)\big|\geqslant\varepsilon\Big)$$

$$=\lim_{n\to\infty}P\Big(\sup_{x\leqslant m\leqslant n}\beta(m)^{-1}\big|S(m)-\alpha(m)\big|\geqslant\varepsilon\Big)=\mathrm{o}(1)$$

for each $\varepsilon>0$ and $x\to\infty$. From the conditioning relation (4) and Lemma 1 we have

$$\nu_{n,\theta}\Big(\max_{x\leqslant m\leqslant n}\beta(m)^{-1}\big|h(\sigma,m)-\alpha(m)\big|\geqslant\varepsilon\Big)$$

$$\ll P^{1\wedge\theta}\Big(\sup_{x\leqslant m\leqslant n}\beta(m)^{-1}\big|S(m)-\alpha(m)\big|\geqslant\varepsilon/3\Big)+n^{-\theta}.$$

Thus from (9) we've obtained (8). Now using the fundamental lemma, we have

$$\nu_{n,\theta}\Big(\max_{x\leqslant m\leqslant r}\beta(m)^{-1}\big|h(\sigma,m)-\alpha(m)\big|\geqslant\varepsilon\Big)$$

$$=P\Big(\sup_{x\leqslant m\leqslant r}\beta(m)^{-1}\big|S_m-\alpha(m)\big|\geqslant\varepsilon\Big)+\mathrm{o}(1).$$

for each $r=r(n)$, $r\to\infty$, $r=\mathrm{o}(n)$. Taking limits with respect to n, later with respect to x, from (8) we deduce (9). Theorem is proved.

LEMMA 2. *Let $b_n \to 0$, $n \to \infty$, $1 \leqslant s \leqslant n$ and $\varepsilon > 0$. Then*

$$\nu_{n,\theta}\left(\max_{s\leqslant m\leqslant n}\left|\sum_{j\leqslant m} a_j k_j(\sigma) - A(m)\right| \geqslant \varepsilon\right)$$

$$\leqslant C_1(\varepsilon)\left(b_s^2 B(s) + \theta \sum_{s\leqslant j\leqslant n} \frac{b_j^2 a^2(j)}{j}\right)^{\theta\wedge 1} + Cn^{-\theta}.$$

Proof. Using the conditioning relation, Lemma 1 and Theorem 3.3.13 of [6] we have

$$\nu_{n,\theta}\left(\max_{s\leqslant m\leqslant n} b_m\left|\sum_{j\leqslant m} a_j k_j(\sigma) - A(m)\right| \geqslant \varepsilon\right)$$

$$= P\left(\max_{s\leqslant m\leqslant n} b_m\left|\sum_{j\leqslant m} a_j \xi_j - A(m)\right| \geqslant \varepsilon | \zeta = m\right)$$

$$\leqslant C P^{\theta\wedge 1}\left(\max_{s\leqslant m\leqslant n} b_m|S_m - A(m)| \geqslant \frac{\varepsilon}{3}\right) + Cn^{-\theta}$$

$$= C P^{\theta\wedge 1}\left(\max_{s\leqslant m\leqslant n} b_m|S_m - ES_m| \geqslant \frac{\varepsilon}{3}\right) + Cn^{-\theta}$$

$$\leqslant C_1(\varepsilon)\left(b_s^2 B(s) + \theta \sum_{s\leqslant j\leqslant n} \frac{b_j^2 a^2(j)}{j}\right)^{\theta\wedge 1} + Cn^{-\theta}.$$

Lemma 2 is proved.

Proof of Theorem 2. The desired assertion one can obtain from Theorem 1, fundamental lemma and Lemma 2. See the proof of Theorem 2 in [4] for details.

Proof of Theorem 3. At the beginning, we apply Theorem 2 to the additive function

$$h(\sigma, m) = s(\sigma, m) := \sum_{l\leqslant m} k_l^0(\sigma),$$

which is the count of all different cycle lengths in decomposition (1), $1 \leqslant m \leqslant n$, $0^0 := 0$.

In this case we have

$$\lim_{x\to\infty} \overline{\lim_{n\to\infty}}\, \nu_{n,\theta}\left(\max_{x\leqslant m\leqslant n} \frac{|s(\sigma,m) - \log m|}{(2\log m LLLm)^{1/2}} \geqslant 1 + \delta\right) = 0$$

and

$$\lim_{x\to\infty} \underline{\lim_{n\to\infty}}\, \nu_{n,\theta}\left(\min_{x\leqslant m\leqslant n} \left|\frac{s(\sigma,m) - \log m}{(2\log m LLLm)^{1/2}} - b\right| < \delta\right) = 1$$

for each $b \in [-1, 1]$ and $\delta > 0$. If $s(\sigma, m) = k$, then we have the relation $k = s(\sigma, j(\sigma, k))$. Now we see that last two assertions are also satisfied for $j(\sigma, k)$. The proof of Theorem 3 is completed.

References

1. R. Arratia, A.D. Barbour, S. Tavare, Logarithmic combinatorial structures: a probabilistic approach, in: *EMS Monographs in Math.*, The EMS Publishing House, Zürich (2003).
2. G.J. Babu, E. Manstavičius, Brownian motion for random permutations, *Sankhya Ser. A*, **61**(3), 312–327 (1999).
3. G.J. Babu, E. Manstavičius, Limit processes with independent increments for the Ewens sampling formula, *Ann. Inst. Statist. Math.*, **54**(3), 607–620 (2002).
4. E. Manstavičius, The law of iterated logarithm for random permutations, *Lith. Math. J.*, **38**, 160–171 (1999).
5. E. Manstavičius, Stochastic processes with independent increments for random mappings, *Lith. Math. J.*, **39**(4), 393–407 (1999).
6. V.V. Petrov, *Sums of Independent Random Variables*, Nauka, Moscow (1972) (in Russian).

REZIUMĖ

J. Norkūnienė. Kartotinio logaritmo dėsnis Evenso formulei

Šiame straipsnyje nagrinėjamas kombinatorinių struktūrų, aprašytų Evenso formule, stiprusis konvergavimas, įrodytas kartotinio logaritmo dėsnis.