Liet. matem. rink., **44**, *spec. nr.*, 2004, 90–94

Harmonic Bernoulli strings and random permutations

Eugenijus MANSTAVIČIUS (VU)
e-mail: eugenijus.manstavicius@maf.vu.lt

Abstract. We examine fairly special b-harmonic Bernoulli strings appearing in n observations. It is shown that their count number can be used to define a random process converging to the Brownian motion as n tends to infinity. The proof is based upon the invariance principle for random permutations.

Keywords: Bernoulli string, invariance principle, Brownian motion, symmetric group.

This remark is stimulated by the recent paper [7] and preprint [9], where connections between Bernoulli strings and random permutations are discussed. These works also contain the historical background and motivation. Proving the invariance principle, we now demonstrate that the results obtained for random permutations can imply their analogs for the strings.

Let us observe the sequence $X_1, X_2, \dots$ of independent Bernoulli random variables with $p_i = P(X_i = 1) = 1 - P(X_i = 0) > 0$, $i \geqslant 1$. Let the variable Z_j count the number of occurrences of the j-string $(1, 0, \dots, 0, 1)$ containing $j-1$ consecutive zeros, $j \geqslant 1$. Sometimes (see, for instance [1]) such strings are called j-spacings of the sequence $X_1, X_2, \dots$. The indicator of the j-string, which started at time i, equals

$$Y_{ij} := X_i(1 - X_{i+1}) \dots (1 - X_{i+j-1})X_{i+j}$$

and

$$Z_j = \sum_{i \geqslant 1} Y_{ij}, \quad j \geqslant 1.$$

To assure $Z_j < \infty$ with probability one, it is necessary and sufficient to assume the condition

$$\sum_{i \geqslant 1} p_i \prod_{k=1}^{j-1} (1 - p_{i+k}) p_{i+j} < \infty. \tag{1}$$

In [7] and [9], a great attention is paid to the distribution of Z_j if $\{X_i\}$ is the b-harmonic Bernoulli sequence defined via $p_i = 1/(i + b)$, $b \geqslant 0$. The main result of [9] gives the expression of the factorial moments. For instance, if $b > 0$ and $u^{(s)} := u(u-1) \dots (u - s + 1)$, then

$$\mathbf{E}\big(Z_1^{(s_1)} \dots Z_N^{(s_N)}\big) = b \int_0^1 v^{b-1} \prod_{j=1}^{N} \Big(\frac{1 - v^j}{j}\Big)^{s_j} \, dv,$$

where $s_1, \dots, s_N \in \mathbb{N}$. If $b = 0$, then we have [2]

$$\mathbf{E}(Z_1^{(s_1)} \dots Z_N^{(s_N)}) = \prod_{j=1}^{N} \left(\frac{1}{j}\right)^{s_j}.$$

An interesting Bernoulli sequence satisfying (1) and closely connected with the theory of random permutations is given by $p_i = \theta/(\theta + i - 1)$, where $\theta > 0$, $i \geqslant 1$. This connection will be seen from the Feller Coupling Lemma below. By *the random permutation* we understand a permutation σ taken from the symmetric group S_N with the Ewens probability

$$P(\{\sigma\}) = \theta^{w(\sigma)}/\theta(\theta + 1) \dots (\theta + N - 1) =: \theta^{w(\sigma)}/\theta_{(N)}, \tag{2}$$

where $w(\sigma)$ denotes the number of independent cycles in the decomposition of σ. Let $k_j(\sigma)$ be the number of cycles of length j in it. Then $w(\sigma) = k_1(\sigma) + \dots + k_N(\sigma)$ and $\bar{k}(\sigma) := (k_1(\sigma), \dots, k_N(\sigma))$ is called the cycle structure vector. Its distribution is given (see [1] or [2]) by

$$P(\bar{k}(\sigma) = (k_1, \dots, k_N)) = \mathbf{1}\{1k_1 + \cdots Nk_N = N\} \frac{N!}{\theta_{(N)}} \prod_{j=1}^{N} \left(\frac{\theta}{j}\right)^{k_j} \frac{1}{k_j!}.$$

Note that the relation $1k_1(\sigma) + \dots + Nk_N(\sigma) = N$ implies dependence of the random variables $k_j(\sigma)$, $1 \leqslant j \leqslant N$.

LEMMA 1 (Feller's Coupling). *Let $N \in \mathbb{N}$ and $\theta > 0$ be arbitrary. There exists a probability space on which the Bernoulli sequence $\{X_i\}$ with $p_i = \theta/(\theta + i - 1)$, $i \geqslant 1$ is defined together with a random permutation σ so that* (2) *holds and*

$$k_j(\sigma) = \sum_{i=1}^{N-j} Y_{ij} + X_{N-j+1}(1 - X_{N-j+2}) \dots (1 - X_N).$$

For a proof and applications of these relations, we refer to [1] and [2].

In the papers [3]–[6], [8], and others general functional limit theorems in the space $\mathbb{D}[0, 1]$ for the partial sum processes

$$\sum_{j \leqslant y_N(t)} h_{jN}(k_j(\sigma)) - \alpha_N(t)$$

were proved. Here $h_{jN}(k)$ is a real valued array, $j \leqslant N$, $N \in \mathbb{N}$, $k \geqslant 0$, y_N: $[0, 1] \to \{1, \dots, N\}$ is an appropriate nondecreasing mapping, $y_N(0) = 0$, $y_N(1) = N$, and $\alpha_N(t)$ is a centralizing function on $[0, 1]$. By the lemma above these results can be interpreted in terms of the Bernoulli string counts if $p_i = \theta/(\theta + i - 1)$. In this remark we show this possibility for the b-harmonic strings provided that $b \in \mathbb{N}$.

In what follows we take $p_i = 1/(i+b)$, $i \geqslant 1$, and $b \in \mathbb{Z}^+$. Let $U_{jn} = Y_{1j} + \dots + Y_{n-j,j}$, $1 \leqslant j \leqslant n$ be the number of j-strings in the sequence $X_1, \dots, X_n$. Set

$$W_n(t) = \frac{1}{\sqrt{\log n}} \Bigl(\sum_{j \leqslant nt} U_{jn} - t \log n \Bigr), \quad t \in [0,1],\ n \geqslant 2.$$

In what follows in the asymptotic relations we take $n \to \infty$.

THEOREM. *The process W_n weakly converges in the space $\mathbb{D}[0,1]$ to the standard Brownian motion.*

We will deduce the assertion from the following proposition.

LEMMA 2. *The process*

$$H_n(t) := \frac{1}{\sqrt{\log(n+b)}} \Bigl(\sum_{j \leqslant t(n+b)} k_j(\sigma) - t \log(n+b) \Bigr), \quad t \in [0,1],\ n \geqslant 2,$$

under the measure (2) *with arbitrary fixed $\theta > 0$, weakly converges in the space $\mathbb{D}[0,1]$ to the standard Brownian motion.*

This is a corollary of a result by J.C. Hansen [6]. See [3]–[5] for further generalizations.

Proof of Theorem. We apply Lemmas 1 and 2 for $\theta = 1$ taking also $N = n + b$. They also allow us to assume that the processes H_n and W_n are defined on a common probability space. So, it suffices to prove that

$$P_n(\varepsilon) := P\Bigl(\sup_{0 \leqslant t \leqslant 1} \bigl| H_n(t) - W_n(t) \bigr| \geqslant \varepsilon \Bigr) = \mathrm{o}(1)$$

for each $\varepsilon > 0$.

Since $b \in \mathbb{Z}^+$, the variable U_{jn} actually counts the j-strings in the sequence $\hat{X}_{b+1}, \dots, \hat{X}_{b+n}$, where $\hat{X}_{b+i} = X_i$, $i \geqslant 1$. If further

$$\hat{Y}_{ij} := \hat{X}_i \prod_{k=1}^{j-1} (1 - \hat{X}_{i+k}) \hat{X}_{i+j}, \quad i \geqslant 1,$$

then

$$U_{jn} = \sum_{i=1}^{n-j} \hat{Y}_{b+i,j}, \qquad k_j(\sigma) = \sum_{i=1}^{n+b-j} \hat{Y}_{ij} + \hat{Y}'_{n+b-j+1,j},$$

where $\hat{Y}'_{n+b-j+1,j} := \hat{X}_{n+b-j+1}(1 - \hat{X}_{n+b-j+2}) \dots (1 - \hat{X}_{n+b})$. Hence

$$\Delta_j := k_j(\sigma) - U_{jn} = \sum_{i=1}^{b} \hat{Y}_{ij} + \hat{Y}'_{n+b-j+1,j} \tag{3}$$

and

$$P_n(\varepsilon) \leqslant P\biggl(\sup_{0\leqslant t\leqslant 1}\biggl|\frac{1+\mathrm{O}(n^{-1})}{\sqrt{\log n}}\sum_{j\leqslant t(n+b)} k_j(\sigma) - \frac{1}{\sqrt{\log n}}\sum_{j\leqslant tn} U_{jn}\biggr| + \mathrm{O}\bigl(n^{-1/2}\bigr) \geqslant \varepsilon\biggr)$$

$$\leqslant P\biggl(\sum_{j\leqslant n}\Delta_j \geqslant (\varepsilon/4)\sqrt{\log n}\biggr) + P\biggl(\sup_{0\leqslant t\leqslant 1}\sum_{tn<j\leqslant t(n+b)} k_j(\sigma) \geqslant (\varepsilon/4)\sqrt{\log n}\biggr)$$

$$+ P\biggl(\sum_{j\leqslant n+b} k_j(\sigma) \geqslant (\varepsilon/4)n\sqrt{\log n}\biggr) + \mathrm{o}(1). \tag{4}$$

Since

$$\mathbf{E}\hat{Y}_{ij} = \frac{1}{(i+j-1)(i+j)}, \qquad \mathbf{E}\hat{Y}'_{n+b-j+1,j} = \frac{1}{n+b},$$

by (3) we have

$$\mathbf{E}\sum_{j\leqslant n}\Delta_j = \sum_{j\leqslant n}\sum_{i=1}^{b}\Bigl(\frac{1}{i+j-1} - \frac{1}{i+j}\Bigr) + \frac{n}{n+b} \leqslant b\sum_{j\leqslant n}\frac{1}{j(b+j)} + 1 = \mathrm{O}(1).$$

Thus, by virtue of Markov's inequality, the first probability on the right hand side of (4) is o(1). Since

$$\mathbf{E}k_j(\sigma) = \sum_{i=1}^{n+b-j}\Bigl(\frac{1}{i+j-1} - \frac{1}{i+j}\Bigr) + \frac{1}{n+b} = \frac{n+b-j}{j(n+b)} + \frac{1}{n+b} \tag{5}$$

and

$$\mathbf{E}\sum_{j\leqslant n+b} k_j(\sigma) = \frac{1}{n+b}\sum_{j\leqslant n+b}\frac{n+b-j}{j} + 1 = \mathrm{O}(\log n),$$

we again obtain the estimate o(1) for the last probability in (4).

By (5), the probability on the right hand of (4) containing the supremum over t does not exceed

$$P\biggl(\max_{0<k\leqslant n}\sum_{k<j\leqslant k+bk/n} k_j(\sigma) \geqslant \frac{\varepsilon}{8}\sqrt{\log n}\biggr)$$

$$+ P\biggl(\max_{0<k\leqslant n+b}\sum_{kn/(n+b)<j\leqslant k} k_j(\sigma) \geqslant \frac{\varepsilon}{8}\sqrt{\log n}\biggr)$$

$$\leqslant \frac{8}{\varepsilon\sqrt{\log n}}\biggl(\sum_{k=0}^{n}\sum_{k<j\leqslant k+b}\mathbf{E}k_j(\sigma) + \sum_{k=0}^{n+b}\sum_{k-b<j\leqslant k}\mathbf{E}k_j(\sigma)\biggr)$$

$$= \mathrm{O}\biggl(\frac{1}{\sqrt{\log n}}\sum_{k=b}^{n}\sum_{k-b<j\leqslant k+b}\frac{1}{j}\biggr) + \mathrm{o}(1)$$

$$= \mathrm{O}\Big(\frac{1}{\sqrt{\log n}}\sum_{k=b}^{n}\Big(\frac{1}{2(k+b)} - \frac{1}{2(k-b)} + \mathrm{O}\Big(\frac{1}{k^2}\Big)\Big)\Big) + \mathrm{o}(1) = \mathrm{o}(1).$$

Gathering all estimates of the probabilities in (4), we have $P_n(\varepsilon) = \mathrm{o}(1)$ for each $\varepsilon > 0$. Theorem is proved.

Concluding Remark. The assertion of Theorem can be extended for the Bernoulli sequences $X_1, X_2, \ldots$ with

$$p_i = P(X_i = 1) = \frac{\theta}{\theta + i - 1 + b}, \quad i \geqslant 1,$$

where $\theta > 0$ and $b \in \mathbb{Z}^+$ are fixed. Nevertheless, the case $\theta < 1$ is much more involved.

References

1. R. Arratia, A.D. Barbour, S. Tavaré, Poisson process approximations for the Ewens sampling formula, *Ann. Appl. Probab.*, **2**(3), 519–535 (1992).
2. R. Arratia, A.D. Barbour, S. Tavaré, *Logarithmic Combinatorial Structures: a Probabilistic Approach*, EMS, Zürich (2003).
3. G.J. Babu, E. Manstavičius, Brownian motion and random permutations, *Sankhyā, A*, **61**(3), 312–327 (1999).
4. G.J. Babu, E. Manstavičius, Infinitely divisible limit processes for the Ewens sampling formula, *Lithuanian Math. J.*, **42**(3), 232–242 (2002).
5. G.J. Babu, E. Manstavičius, Processes with independent increments for the Ewens sampling formula, *Ann. Inst. Stat. Math.*, **54**(3), 607–620 (2002).
6. J.C. Hansen, A functional central limit theorem for the Ewens Sampling Formula, *J. Appl. Probab.*, **27**, 28–43 (1990).
7. A. Joffe, E. Marchand, F. Perron, P. Popadiuk, On sums of products of Bernoulli variables and random permutations, *J. Theor. Probab.*, **17**, 285–292 (2004).
8. E. Manstavičius, Functional limit theorem for sequences of mappings on the symmetric group, in: *Anal. Probab. Methods Number Theory*, A. Dubickas *et al.* (Eds.), TEV, Vilnius (2002), pp. 175–187.
9. J. Sethuraman, S. Sethuraman, *On Counts of Bernoulli Strings and Connections to Rank Orders and Random Permutations*, Internet, Preprint (2004).

REZIUMĖ

E. Manstavičius. Harmoninės Bernulio voros ir atsitiktiniai keitiniai

Nagrinėjamos atsitiktinės specialaus pavidalo nulių ir vienetų voros, gautos generuojant nepriklausomus Bernulio atsitiktinius dydžius, kurių sėkmių tikimybės sudaro b-harmoninę seką. Panaudojus tam tikros struktūros vorų, pasirodžiusių tarp pirmųjų n reikšmių, indikatorius, apibrėžiamas procesas ir įrodomas jo konvergavimas $\mathbb{D}[0, 1]$ erdvėje. Įrodyme pritaikomas atsitiktinių keitinių invariantiškumo principas.