Annals of the University of Craiova
Mathematics and Computer Science Series

Vol. XLVII Issue 1, June 2020

Volume Editors: Vicenţiu Rădulescu, Ionel Rovenţa, Mihaela Sterpu

Layout Editor: Mihai Gabroveanu

ISSN 1223-6934

Online ISSN 2246-9958

Web: http://inf.ucv.ro/∼ami/

Printed in Romania: Editura Universitaria, Craiova, 2020.

http://www.editurauniversitaria.ro

Annals of the University of Craiova, Mathematics and Computer Science Series
Volume 47(1), 2020, Pages 1–24
ISSN: 1223-6934

Ewens distribution on $\mathbb{S}_n$ is a wavy probability distribution with respect to n partitions

Udrea Păun

Abstract. We show that the Ewens distribution on $\mathbb{S}_n$, the set of permutations of order n, is a wavy probability distribution with respect to an order relation and n partitions which will be specified — the fact that the number of partitions is n is important. We then construct a Gibbs sampler in a generalized sense for the Ewens distribution. This chain leads

1) to a fast exact (not approximate) Markovian method for sampling from $\mathbb{S}_n$ according to the Ewens distribution and, as a result, to a fast exact method for sampling from $\mathbb{A}_n$, a set which will be specified, according to the Ewens sampling formula;

2) to the computation of normalization constant of Ewens distribution;

3) to the computation, by Uniqueness Theorem, of certain important probabilities for the Ewens distribution and, as a result, to upper bounds for the cumulative distribution function of number of cycles of permutation chosen from $\mathbb{S}_n$ according to the Ewens distribution.

Our sampling Markovian method has something in common with the swapping method. The number of steps of our sampling Markovian method is equal to the number of steps of swapping method, *i.e.*, $n-1$; moreover, both methods use the best probability distributions on sampling, the swapping method uses uniform probability distributions while our method uses almost uniform probability distributions (all the components of an almost uniform probability distribution are, here, identical, excepting at most one of them).

2010 Mathematics Subject Classification. 60J10, 60C05, 60E05, 62Dxx, 68U20, 92D15.
Key words and phrases. wavy probability distribution, Gibbs sampler in a generalized sense, G method, Ewens distribution, Ewens sampling formula, exact sampling, swapping method, normalization constant, important probabilities, cumulative distribution function.

1. Basic things, I

In this section, we present some basic things on nonnegative matrices, products of stochastic matrices, our hybrid Metropolis-Hastings chain, our Gibbs sampler in a generalized sense, and our wavy probability distributions.

Set

$$\operatorname{Par}(E) = \{\Delta \mid \Delta \text{ is a partition of } E\},$$

where E is a nonempty set. We shall agree that the partitions do not contain the empty set.

Definition 1.1. Let $\Delta_1, \Delta_2 \in \operatorname{Par}(E)$. We say that Δ_1 *is finer than* Δ_2 if $\forall V \in \Delta_1$, $\exists W \in \Delta_2$ such that $V \subseteq W$.

Write $\Delta_1 \preceq \Delta_2$ when Δ_1 is finer than Δ_2.

In this article, a vector is a row vector and a stochastic matrix is a row stochastic matrix.

Received May 10, 2019. Accepted May 17, 2020.

The entry (i, j) of a matrix Z will be denoted Z_{ij} or, if confusion can arise, $Z_{i\to j}$.
Set

$$\langle m\rangle = \{1, 2, ..., m\} \ \ (m \geq 1),$$
$$N_{m,n} = \{P \mid P \text{ is a nonnegative } m \times n \text{ matrix}\},$$
$$S_{m,n} = \{P \mid P \text{ is a stochastic } m \times n \text{ matrix}\},$$
$$N_n = N_{n,n},$$
$$S_n = S_{n,n}.$$

Let $P = (P_{ij}) \in N_{m,n}$. Let $\emptyset \neq U \subseteq \langle m\rangle$ and $\emptyset \neq V \subseteq \langle n\rangle$. Set the matrices

$$P_U = (P_{ij})_{i\in U, j\in\langle n\rangle}, \ P^V = (P_{ij})_{i\in\langle m\rangle, j\in V}, \text{ and } P_U^V = (P_{ij})_{i\in U, j\in V}.$$

Set

$$(\{i\})_{i\in\{s_1,s_2,...,s_t\}} = (\{s_1\}, \{s_2\}, ..., \{s_t\});$$
$$(\{i\})_{i\in\{s_1,s_2,...,s_t\}} \in \operatorname{Par}(\{s_1, s_2, ..., s_t\}) \ \ (t \geq 1).$$

E.g.,

$$(\{i\})_{i\in\langle n\rangle} = (\{1\}, \{2\}, ..., \{n\}).$$

Definition 1.2. Let $P \in N_{m,n}$. We say that P is a *generalized stochastic matrix* if $\exists a \geq 0$, $\exists Q \in S_{m,n}$ such that $P = aQ$.

Definition 1.3. ([8].) Let $P \in N_{m,n}$. Let $\Delta \in$Par$(\langle m\rangle)$ and $\Sigma \in$Par$(\langle n\rangle)$. We say that P is a $[\Delta]$-*stable matrix on* Σ if P_K^L is a generalized stochastic matrix, $\forall K \in \Delta, \forall L \in \Sigma$. In particular, a $[\Delta]$-stable matrix on $(\{i\})_{i\in\langle n\rangle}$ is called $[\Delta]$-*stable* for short.

Definition 1.4. ([8].) Let $P \in N_{m,n}$. Let $\Delta \in$Par$(\langle m\rangle)$ and $\Sigma \in$Par$(\langle n\rangle)$. We say that P is a Δ-*stable matrix on* Σ if Δ is the least fine partition for which P is a $[\Delta]$-stable matrix on Σ. In particular, a Δ-stable matrix on $(\{i\})_{i\in\langle n\rangle}$ is called Δ-*stable* while a $(\langle m\rangle)$-stable matrix on Σ is called *stable on* Σ for short. A stable matrix on $(\{i\})_{i\in\langle n\rangle}$ is called *stable* for short.

Let $\Delta_1 \in$Par$(\langle m\rangle)$ and $\Delta_2 \in$Par$(\langle n\rangle)$. Set (see [8] for G_{Δ_1,Δ_2} and [9] for $\overline{G}_{\Delta_1,\Delta_2}$)

$$G_{\Delta_1,\Delta_2} = \{P \mid P \in S_{m,n} \text{ and } P \text{ is a } [\Delta_1]\text{-stable matrix on } \Delta_2\}$$

and

$$\overline{G}_{\Delta_1,\Delta_2} = \{P \mid P \in N_{m,n} \text{ and } P \text{ is a } [\Delta_1]\text{-stable matrix on } \Delta_2\}.$$

When we study or even when we construct products of nonnegative matrices (in particular, products of stochastic matrices) using G_{Δ_1,Δ_2} or $\overline{G}_{\Delta_1,\Delta_2}$, we shall refer this as the *G method*. *G* comes from the verb *to group* and its derivatives.

Below we give an important beautiful result on products of stochastic matrices.

Theorem 1.1. ([8].) *Let* $P_1 \in G_{(\langle m_1\rangle),\Delta_2} \subseteq S_{m_1,m_2}$, $P_2 \in G_{\Delta_2,\Delta_3} \subseteq S_{m_2,m_3}$, ..., $P_{n-1} \in G_{\Delta_{n-1},\Delta_n} \subseteq S_{m_{n-1},m_n}$, $P_n \in G_{\Delta_n,(\{i\})_{i\in\langle m_{n+1}\rangle}} \subseteq S_{m_n,m_{n+1}}$. *Then*

$$P_1P_2...P_n$$

is a stable matrix (*i.e., a matrix with identical rows, see Definition* 1.4).

Proof. See [8]. □

Definition 1.5. (See, *e.g.*, [16, p. 80].) Let $P \in N_{m,n}$. We say that P is a *row-allowable matrix* if it has at least one positive entry in each row.

Let $P \in N_{m,n}$. Set

$$\overline{P} = \left(\overline{P}_{ij}\right) \in N_{m,n},\ \overline{P}_{ij} = \begin{cases} 1 \text{ if } P_{ij} > 0, \\ 0 \text{ if } P_{ij} = 0, \end{cases}$$

$\forall i \in \langle m \rangle, \forall j \in \langle n \rangle$. We call $\overline{P}$ the *incidence matrix of* P (see, *e.g.*, [7, p. 222]).

In this article, the transpose of a vector x is denoted x'. Set $e = e(n) = (1, 1, ..., 1) \in \mathbb{R}^n$, $\forall n \geq 1$.

In this article, some statements on the matrices hold eventually by permutation of rows and columns. For simplification, further, we omit to specify this fact.

Warning! In this article, if a Markov chain has the transition matrix $P = P_1 P_2 ... P_s$, where $s \geq 1$ and P_1, P_2, ..., P_s are stochastic matrices, then any 1-step transition of this chain is performed via P_1, P_2, ..., P_s, *i.e.*, doing s transitions: one using P_1, one using P_2, ..., one using P_s.

Let S be a finite set with $|S| = r$, where $r \geq 2$ ($|\cdot|$ is the cardinal; for "$r \geq 2$", see below). Let $\pi = (\pi_i)_{i \in S}$ be a positive probability distribution on S. One way to sample approximately or, at best, exactly from S is by means of our hybrid Metropolis-Hastings chain from [9]. Below we define this chain.

Let E be a nonempty set. Set $\Delta \succ \Delta'$ if $\Delta' \preceq \Delta$ and $\Delta' \neq \Delta$, where Δ, $\Delta' \in \text{Par}(E)$.

Let Δ_1, Δ_2, ..., $\Delta_{t+1} \in \text{Par}(S)$ with $\Delta_1 = (S) \succ \Delta_2 \succ ... \succ \Delta_{t+1} = (\{i\})_{i \in S}$, where $t \geq 1$. ($\Delta_1 \succ \Delta_2$ implies $r \geq 2$.) Let Q_1, Q_2, ..., $Q_t \in S_r$, $Q_1 = \left((Q_1)_{ij}\right)_{i,j \in S}$, $Q_2 = \left((Q_2)_{ij}\right)_{i,j \in S}$, ..., $Q_t = \left((Q_t)_{ij}\right)_{i,j \in S}$, such that

(C1) $\overline{Q}_1$, $\overline{Q}_2$, ..., $\overline{Q}_t$ are symmetric matrices;

(C2) $(Q_l)_K^L = 0, \forall l \in \langle t \rangle - \{1\}, \forall K, L \in \Delta_l, K \neq L$ (this assumption implies that Q_l is a block diagonal matrix and Δ_l-stable matrix on $\Delta_l, \forall l \in \langle t \rangle - \{1\}$);

(C3) $(Q_l)_K^U$ is a row-allowable matrix, $\forall l \in \langle t \rangle$, $\forall K \in \Delta_l$, $\forall U \in \Delta_{l+1}$, $U \subseteq K$.

Define the matrices

$$P_l = \left((P_l)_{ij}\right)_{i,j \in S} \ (P_l \in S_r),$$

$$(P_l)_{ij} = \begin{cases} 0 & \text{if } j \neq i \text{ and } (Q_l)_{ij} = 0, \\ (Q_l)_{ij} \min\left(1, \frac{\pi_j (Q_l)_{ji}}{\pi_i (Q_l)_{ij}}\right) & \text{if } j \neq i \text{ and } (Q_l)_{ij} > 0, \\ 1 - \sum\limits_{k \neq i} (P_l)_{ik} & \text{if } j = i, \end{cases}$$

$\forall l \in \langle t \rangle$. Set $P = P_1 P_2 ... P_t$.

Theorem 1.2. ([9].) *Concerning P above we have $\pi P = \pi$ and $P > 0$.*

Proof. See [9]. □

By Theorem 1.2, $P^n \to e'\pi$ as $n \to \infty$. We call the Markov chain with transition matrix P the *hybrid Metropolis-Hastings chain.* In particular, we call this chain the *hybrid Metropolis chain* when Q_1, Q_2, ..., Q_t are symmetric matrices.

The next result is a corrected version of Theorem 2.1 from [14].

Theorem 1.3. ([15].) *Consider a hybrid Metropolis-Hastings chain with state space S above ($|S| = r \geq 2$) and transition matrix $P = P_1P_2...P_t$, P_1, P_2, ..., P_t corresponding to Q_1, Q_2, ..., Q_t, respectively. Suppose that $\forall l \in \langle t \rangle$, $\forall i, j \in S$,*

$$(Q_l)_{ij} = \frac{\pi_j}{\sum\limits_{k \in S, (Q_l)_{ik} > 0} \pi_k} \quad \text{if } (Q_l)_{ij} > 0$$

(see above for Q_l, $l \in \langle t \rangle$, $\pi = (\pi_i)_{i \in S}$, ...). Then

$$(P_l)_{ij} = \begin{cases} 0 & \text{if } j \neq i \text{ and } (Q_l)_{ij} = 0, \\ (Q_l)_{ij} & \text{if } j \neq i \text{ and } \pi_j (Q_l)_{ji} \geq \pi_i (Q_l)_{ij} > 0, \\ \frac{\pi_j}{\sum\limits_{k \in S,\ (Q_l)_{jk} > 0} \pi_k} & \text{if } j \neq i \text{ and } \pi_j (Q_l)_{ji} < \pi_i (Q_l)_{ij}, \\ 1 - \sum\limits_{k \neq i} (P_l)_{ik} & \text{if } j = i, \end{cases}$$

$\forall l \in \langle t \rangle$, $\forall i, j \in S$. If, moreover,

$$\pi_i (Q_l)_{ij} = \pi_j (Q_l)_{ji}, \ \forall l \in \langle t \rangle, \forall i, j \in S,$$

then

$$P_l = Q_l, \ \forall l \in \langle t \rangle.$$

Proof. See [15]. □

We call the hybrid Metropolis-Hastings chain from Theorem 1.3 the *cyclic Gibbs sampler in a generalized sense* — the *Gibbs sampler in a generalized sense* for short.

Further, we consider that $S = \{s_1, s_2, ..., s_r\}$, where $r \geq 2$ ($|S| = r$). Equip S with an order relation, $\leqq$. Suppose that $s_1 \leqq s_2 \leqq ... \leqq s_r$. Let $\pi = (\pi_{s_i})_{i \in \langle r \rangle}$ be a positive probability distribution (on S). Let Δ_1, Δ_2, ..., Δ_{t+1} $\in$Par(S) with $\Delta_1 = (S) \succ \Delta_2 \succ ... \succ \Delta_{t+1} = (\{s_i\})_{i \in \langle r \rangle}$, where $t \geq 1$ and $(\{s_i\})_{i \in \langle r \rangle} = (\{s_1\}, \{s_2\}, ..., \{s_r\})$. ($\Delta_1 \succ \Delta_2$ implies $r \geq 2$.) Consider that $\Delta_l = \left(K_1^{(l)}, K_2^{(l)}, ..., K_{u_l}^{(l)}\right)$, $K_1^{(l)}$ having the first $\left|K_1^{(l)}\right|$ elements of S, $K_2^{(l)}$ having the next $\left|K_2^{(l)}\right|$ elements of S (this condition and the next ones vanish when $l = 1$), ..., $K_{u_l}^{(l)}$ having the last $\left|K_{u_l}^{(l)}\right|$ elements of S, $\forall l \in \langle t+1 \rangle$. Consider that

(c1) $\left|K_1^{(l)}\right| = \left|K_2^{(l)}\right| = ... = \left|K_{u_l}^{(l)}\right|$, $\forall l \in \langle t+1 \rangle$ with $u_l \geq 2$;

(c2) $r = r_1r_2...r_t$ with $r_1r_2...r_l = |\Delta_{l+1}|$, $\forall l \in \langle t-1 \rangle$, and $r_t = \left|K_1^{(t)}\right|$.

We have

$$K_v^{(l)} = \bigcup_{w \in D_{v,b_l} \cup \{vb_l\}} K_w^{(l+1)}, \ \forall l \in \langle t \rangle, \ \forall v \in \langle u_l \rangle,$$

where

$$b_l = \frac{|\Delta_{l+1}|}{|\Delta_l|}, \ \forall l \in \langle t \rangle,$$

and

$$D_{v,b_l} = \{(v-1) b_l + 1, \ (v-1) b_l + 2, \ ..., \ vb_l - 1\}, \ \forall l \in \langle t \rangle, \ \forall v \in \langle u_l \rangle.$$

Suppose that $\forall l \in \langle t\rangle$, $\forall v \in \langle u_l\rangle$, $\forall w \in D_{v,b_l}$, $\exists \alpha_w^{(l,v)} > 0$ such that

$$\pi_{s_{i+d_w^{(l,v)}}} = \alpha_w^{(l,v)} \pi_{s_i} \text{ (direct proportionality)}, \forall i \in \langle r\rangle \text{ with } s_i \in K_{(v-1)b_l+1}^{(l+1)},$$

which, using vectors, is equivalent to

$$(\pi_{s_i})_{i\in\langle r\rangle,\ s_i\in K_{w+1}^{(l+1)}} = \alpha_w^{(l,v)} (\pi_{s_i})_{i\in\langle r\rangle,\ s_i\in K_{(v-1)b_l+1}^{(l+1)}},$$

where

$$d_w^{(l,v)} = \left|K_{(v-1)b_l+1}^{(l+1)}\right| + \left|K_{(v-1)b_l+2}^{(l+1)}\right| + ... + \left|K_w^{(l+1)}\right|,$$

$\forall l \in \langle t\rangle$, $\forall v \in \langle u_l\rangle$, $\forall w \in D_{v,b_l}$.

Definition 1.6. (Based on Definition 3.1 from [14].) The probability distribution $\pi = (\pi_{s_i})_{i\in\langle r\rangle}$ having the above property (direct proportionality) we call the *wavy probability distribution* (*with respect to the order relation* $\leqq$ *and partitions* Δ_1, Δ_2, ..., Δ_{t+1}).

The wavy probability distributions of first type and those of second type from [10] are, according to Definition 1.6, wavy probability distributions (see also Example 3.1 in [14]). Below we give another simple example of wavy probability distribution.

Example 1.1. Let $S = \langle 9\rangle$. Let $\leqq = \leq$. Let

$$\pi = \left(\frac{a}{Z}, \frac{a^3}{Z}, \frac{a^4}{Z}, \frac{a^3}{Z}, \frac{a^5}{Z}, \frac{a^6}{Z}, \frac{a^{10}}{Z}, \frac{a^{12}}{Z}, \frac{a^{13}}{Z}\right),$$

a probability distribution on S, where $a > 0$ and

$$Z = a + a^3 + a^4 + a^3 + a^5 + a^6 + a^{10} + a^{12} + a^{13}$$

(the normalization constant). Let

$$\Delta_1 = (S) = (\langle 9\rangle),$$

$$\Delta_2 = (\{1, 2, 3\}, \{4, 5, 6\}, \{7, 8, 9\}),$$

$$\Delta_3 = (\{i\})_{i\in\langle 9\rangle}$$

$$(\Delta_1 \succ \Delta_2 \succ \Delta_3;\ \ |\{1, 2, 3\}| = |\{4, 5, 6\}| = |\{7, 8, 9\}| = 3).$$

First, we consider Δ_1 and Δ_2. We have

$$\pi_4 = a^2\pi_1,\ \pi_5 = a^2\pi_2,\ \pi_6 = a^2\pi_3,$$

which, using vectors, is equivalent to

$$(\pi_4, \pi_5, \pi_6) = a^2 (\pi_1, \pi_2, \pi_3)$$

(the proportionality factor is a^2), and

$$\pi_7 = a^9\pi_1,\ \pi_8 = a^9\pi_2,\ \pi_9 = a^9\pi_3,$$

which is equivalent to

$$(\pi_7, \pi_8, \pi_9) = a^9 (\pi_1, \pi_2, \pi_3)$$

(the proportionality factor is a^9). Second, we consider Δ_2 and Δ_3. We have

$$\pi_2 = a^2\pi_1$$

(here, we do not use vectors anymore; the proportionality factor is a^2),

$$\pi_3 = a^3\pi_1$$

(the proportionality factor is a^3),

$$\pi_5 = a^2\pi_4,$$

$$\pi_6 = a^3\pi_4$$

(the proportionality factors are a^2 and a^3, respectively),

$$\pi_8 = a^2\pi_7,$$

$$\pi_9 = a^3\pi_7$$

(the proportionality factors are also a^2 and a^3, respectively). Consequently, π is a wavy probability distribution on S (neither of the first type nor of the second type because $S = \langle 9 \rangle$ and, moreover, $\leqq = \leq$).

The next result is another important result.

Theorem 1.4. (Based on Theorem 3.1 from [14].) *Let* $\pi = (\pi_{s_i})_{i\in\langle r\rangle}$ *be a wavy probability distribution* (*on* S) *with respect to the order relation* $\leqq$ *and partitions* Δ_1, Δ_2, ..., Δ_{t+1} — *for* S, $\leqq$, ..., *see Definition* 1.6 *and above this definition. Consider a Markov chain with state space* S *and transition matrix* $P = P_1P_2...P_t$ $(t \geq 1)$, *where* (*we again use the notation from Definition* 1.6 *and above this definition*)

$$(P_l)_{s_{i+d_w^{(l,v)}}\to\xi} = \begin{cases} \dfrac{\pi_{s_{i+d_u^{(l,v)}}}}{\sum\limits_{z\in\{0\}\cup D_{v,b_l}} \pi_{s_{i+d_z^{(l,v)}}}} & \text{if } \xi = s_{i+d_u^{(l,v)}} \text{ for some } u \in \{0\}\cup D_{v,b_l}, \\ 0 & \text{if } \xi \neq s_{i+d_u^{(l,v)}},\ \forall u \in \{0\}\cup D_{v,b_l}, \end{cases}$$

$\forall l \in \langle t\rangle$, $\forall v \in \langle u_l\rangle$, $\forall i \in \langle r\rangle$ *with* $s_i \in K^{(l+1)}_{(v-1)b_l+1}$, $\forall w \in \{0\}\cup D_{v,b_l}$, $\forall \xi \in S$, *setting* $d_0^{(l,v)} = 0$, $\forall l \in \langle t\rangle$, $\forall v \in \langle u_l\rangle$. *Then this chain is a Gibbs sampler in a generalized sense and*

$$P = e'\pi$$

(*therefore, this chain attains its stationarity at time* 1, *its stationary probability distribution* (*limit probability distribution*) *being, obviously,* π).

Proof. It is easy to see that

$$\pi_{s_i}(P_l)_{s_is_j} = \pi_{s_j}(P_l)_{s_js_i},\ \forall l \in \langle t\rangle,\ \forall i,j \in \langle r\rangle.$$

Taking — Theorem 1.3 together with the above equations, definitions of matrices P_l, $l \in \langle t\rangle$, ... suggest to take so —

$$Q_l = P_l,\ \forall l \in \langle t\rangle,$$

we obtain that the above Markov chain is a Gibbs sampler in a generalized sense. For the proof of equation $P = e'\pi$ — this equation follows from Theorems 1.1 and 1.2 —, see the proof of Theorem 3.1 in [14]. □

Theorem 1.4 leads to the next result.

Theorem 1.5. (Based on Theorem 3.2 from [14].) *Let* $\pi = (\pi_{s_i})_{i\in\langle r\rangle}$ *be a wavy probability distribution* (*on* S) *with respect to the order relation* $\leqq$ *and partitions* Δ_1, Δ_2, ..., Δ_{t+1} — *for* S, $\leqq$, ..., *see Definition* 1.6 *and above this definition. Suppose that*

$$\pi_{s_i} = \frac{\nu_{s_i}}{Z},\ \forall i \in \langle r\rangle,$$

where

$$Z=\sum_{i\in\langle r\rangle}\nu_{s_i},$$

Z is the normalization constant (π is a positive probability distribution, so, $\nu_{s_i}\in\mathbb{R}^+$, $\forall i\in\langle r\rangle$, and, as a result, $Z\in\mathbb{R}^+$). Then

$$Z=\nu_{s_1}\prod_{l\in\langle t\rangle}\left(1+\sum_{w\in D_{1,b_l}}\alpha_w^{(l,1)}\right).$$

Proof. See the proof of Theorem 3.2 from [14]. □

2. Basic things, II

In this section, we present the Ewens distribution, Ewens sampling formula, and, in connection with these, some basic things on permutations.

We begin with some basic things on permutations in connection with the Ewens distribution and Ewens sampling formula.

Consider the group $(\mathbb{S}_n,\circ)$, where $\mathbb{S}_n$ is the set of permutations of order n $(n\geq 1)$ and $\circ$ is the usual composition of functions. $(u_1,u_2,...,u_k)$ is a cycle of length k, where k, u_1, u_2, ..., $u_k\in\langle n\rangle$, $u_s\neq u_t,\forall s,t\in\langle k\rangle$, $s\neq t$; (u_1) is a degenerate (improper) cycle and (u_1,u_2) is a transposition. Set $(u)=$Id, $\forall u\in\langle n\rangle$, where (u) is a degenerate cycle, $\forall u\in\langle n\rangle$, and Id is the identity permutation.

Setting $(u,u)=$Id, $\forall u\in\langle n\rangle$, we have the following result.

Theorem 2.1. (Similar to Theorem 2.1 from [11].) *Let $n\geq 2$. Let*

$$\mathbb{E}_{n,l}=\{(1,i_1)\circ(2,i_2)\circ...\circ(l,i_l)\circ\sigma_l \mid i_1,i_2,...,i_l\in\langle n\rangle,\ 1\leq i_1\leq n,$$
$$2\leq i_2\leq n,...,\ l\leq i_l\leq n,\ \sigma_l\in\mathbb{S}_n,\ \sigma_l(v)=v,\ \forall v\in\langle l\rangle\},\ \forall l\in\langle n-1\rangle.$$

Then

$$\mathbb{E}_{n,l}=\mathbb{S}_n,\ \forall l\in\langle n-1\rangle.$$

Proof. (Similar to the proof of Theorem 2.1 from [11].) Let $l\in\langle n-1\rangle$. Since $(\mathbb{S}_n,\circ)$ is a group, we have $\mathbb{E}_{n,l}\subseteq\mathbb{S}_n$. Therefore, $|\mathbb{E}_{n,l}|\leq|\mathbb{S}_n|=n!$. To finish the proof, we show that $|\mathbb{E}_{n,l}|=n!$.

The number of permutations $\sigma_l\in\mathbb{S}_n$ with $\sigma_l(v)=v$, $\forall v\in\langle l\rangle$, is equal to $(n-l)!$. Since $1\leq i_1\leq n$, $2\leq i_2\leq n$, ..., $l\leq i_l\leq n$, it follows that $|\mathbb{E}_{n,l}|$ is at most equal to

$$n(n-1)...(n-l+1)[(n-l)!]=n!.$$

We show that

$$(1,i_1)\circ(2,i_2)\circ...\circ(l,i_l)\circ\sigma_l=(1,j_1)\circ(2,j_2)\circ...\circ(l,j_l)\circ\tau_l$$

if and only if

$$i_k=j_k,\ \forall k\in\langle l\rangle,\ \text{and}\ \sigma_l=\tau_l,$$

where i_1, j_1, i_2, j_2, ..., i_l, $j_l\in\langle n\rangle$, $1\leq i_1$, $j_1\leq n$, $2\leq i_2$, $j_2\leq n$, ..., $l\leq i_l$, $j_l\leq n$, $\sigma_l,\tau_l\in\mathbb{S}_n$, $\sigma_l(v)=\tau_l(v)=v$, $\forall v\in\langle l\rangle$.

" $\Longleftarrow$ " Obvious.

" $\Longrightarrow$ " We have

$$[(1,i_1)\circ(2,i_2)\circ...\circ(l,i_l)\circ\sigma_l](1)=[(1,j_1)\circ(2,j_2)\circ...\circ(l,j_l)\circ\tau_l](1).$$

Therefore,

$$i_1 = j_1.$$

Since $i_1 = j_1$, we have

$$(2, i_2) \circ ... \circ (l, i_l) \circ \sigma_l = (2, j_2) \circ ... \circ (l, j_l) \circ \tau_l.$$

It follows that

$$[(2, i_2) \circ ... \circ (l, i_l) \circ \sigma_l] (2) = [(2, j_2) \circ ... \circ (l, j_l) \circ \tau_l] (2) .$$

Therefore,

$$i_2 = j_2.$$

Proceeding in this way, we obtain

$$i_1 = j_1,\ i_2 = j_2,\ ...,\ i_l = j_l,$$

and, as a result of these equations,

$$\sigma_l = \tau_l.$$

We conclude that

$$|\mathbb{E}_{n,l}| = n!.$$

□

Theorem 2.1 says that we can work with $\mathbb{E}_{n,l}$ instead of $\mathbb{S}_n$, $\forall l \in \langle n-1 \rangle$ (this fact will be used in Section 3 (Theorem 3.1, ...) and Section 4 (Theorem 4.1, ...)).

Let $\psi = (u_1, u_2, ..., u_k)$ be a (proper or not) cycle ($\psi \in \mathbb{S}_n$, $1 \leq k \leq n$). We call u_1, u_2, ..., u_k the *cyclic elements of (cycle)* ψ. *E.g.*, the cyclic elements of cycle

$$(1, 2, 4) = \begin{pmatrix} 1234 \\ 2431 \end{pmatrix} = (2431) \in \mathbb{S}_4$$

are 1, 2, 4 while the cyclic element (this is not a proper cyclic element) of cycle

$$(2) = \begin{pmatrix} 1234 \\ 1234 \end{pmatrix} = (1234) = \mathrm{Id} \in \mathbb{S}_4$$

is 2 (not 1, 3, or 4). We call $\{u_1, u_2, ..., u_k\}$ the *set* (or *orbit*) *of cyclic elements of (cycle)* ψ.

Let $N(\sigma)$ be the number of pair-wise disjoint cycles of permutation σ, where $\sigma \in \mathbb{S}_n$. *E.g.*, $N(\mathrm{Id}) = n$ because Id$= (1) \circ (2) \circ ... \circ (n)$ ((1), (2), ..., (n) are degenerate cycles).

Theorem 2.2. *Let* $n \geq 2$. *Then*

$$N((1, i_1) \circ (2, i_2) \circ ... \circ (l-1, i_{l-1}) \circ (l, j) \circ \sigma_l) =$$

$$= \begin{cases} N((1, i_1) \circ (2, i_2) \circ ... \circ (l-1, i_{l-1}) \circ (l, k) \circ \sigma_l) & \text{if } j = k = l \text{ or } j, k > l, \\ N((1, i_1) \circ (2, i_2) \circ ... \circ (l-1, i_{l-1}) \circ (l, k) \circ \sigma_l) + 1 & \text{if } j = l, k > l, \\ N((1, i_1) \circ (2, i_2) \circ ... \circ (l-1, i_{l-1}) \circ (l, k) \circ \sigma_l) - 1 & \text{if } j > l, k = l, \end{cases}$$

$\forall l \in \langle n-1\rangle$, $\forall i_1, i_2, ..., i_{l-1}, j, k \in \langle n\rangle$, $1 \le i_1 \le n$, $2 \le i_2 \le n, ..., l-1 \le i_{l-1} \le n$, $l \le j$, $k \le n$, $\forall \sigma_l \in \mathbb{S}_n$, $\sigma_l(v) = v$, $\forall v \in \langle l\rangle$ $((1,i_1) \circ (2,i_2) \circ ... \circ (l-1, i_{l-1})$, *etc. vanish when* $l = 1$).

Proof. *Case* 1. $j = k = l$ or $j,\ k > l$.

Subcase 1.1. $j = k = l$. Obvious ($(l,j) = (l,k) =$Id).

Subcase 1.2. $j,\ k > l$. Since $\sigma_l(v) = v$, $\forall v \in \langle l\rangle$, $\exists u \in \langle n-l\rangle$, $\exists \gamma_1, \gamma_2, ..., \gamma_u \in \mathbb{S}_n$, $\gamma_1, \gamma_2, ..., \gamma_u$ are pair-wise disjoint cycles and $\lfloor \gamma_w \rceil \ge 1$, $\forall w \in \langle u\rangle$, where $\lfloor \gamma_w \rceil$ is the length of cycle γ_w, $\forall w \in \langle u\rangle$, such that (the cycles of length 1 are not omitted)

$$\sigma_l = (1) \circ (2) \circ ... \circ (l) \circ \gamma_1 \circ \gamma_2 \circ ... \circ \gamma_u.$$

Since $j,\ k > l$, $\exists s, t \in \langle u\rangle$ such that j is a cyclic element of γ_s and k is a cyclic element of γ_t. It follows that

$$(l,j) \circ \sigma_l = (1) \circ (2) \circ ... \circ (l-1) \circ \xi_1^{(1)} \circ \xi_2^{(1)} \circ ... \circ \xi_u^{(1)},$$

where

$$\xi_z^{(1)} = \begin{cases} \gamma_z & \text{if } z \ne s, \\ \text{the cycle whose set of cyclic elements} & \\ \text{contains } l,\ j, \text{ and the cyclic elements of } \gamma_s & \text{if } z = s, \end{cases}$$

$\forall z \in \langle u\rangle$ (obviously, $\xi_1^{(1)}, \xi_2^{(1)}, ..., \xi_u^{(1)}$ are pair-wise disjoint cycles), and

$$(l,k) \circ \sigma_l = (1) \circ (2) \circ ... \circ (l-1) \circ \varphi_1^{(1)} \circ \varphi_2^{(1)} \circ ... \circ \varphi_u^{(1)},$$

where

$$\varphi_z^{(1)} = \begin{cases} \gamma_z & \text{if } z \ne t, \\ \text{the cycle whose set of cyclic elements} & \\ \text{contains } l,\ k, \text{ and the cyclic elements of } \gamma_t & \text{if } z = t, \end{cases}$$

$\forall z \in \langle u\rangle$. Consequently,

$$N((l,j) \circ \sigma_l) = N((l,k) \circ \sigma_l)$$

Further, we consider the permutations

$$(l-1, i_{l-1}) \circ (l,j) \circ \sigma_l \text{ and } (l-1, i_{l-1}) \circ (l,k) \circ \sigma_l.$$

If $i_{l-1} = l-1$, from (recall that $(x,x) =$Id, $\forall x \in \langle n\rangle$)

$$N((l-1, i_{l-1}) \circ (l,j) \circ \sigma_l) = N((l,j) \circ \sigma_l) =$$
$$= N((l,k) \circ \sigma_l) = N((l-1, i_{l-1}) \circ (l,k) \circ \sigma_l),$$

we obtain

$$N((l-1, i_{l-1}) \circ (l,j) \circ \sigma_l) = N((l-1, i_{l-1}) \circ (l,k) \circ \sigma_l).$$

If $i_{l-1} > l-1$, then $\exists s, t \in \langle u\rangle$ such that i_{l-1} is a cyclic element of $\xi_s^{(1)}$ and, on the other hand, i_{l-1} is a cyclic element of $\varphi_t^{(1)}$. It follows that

$$(l-1, i_{l-1}) \circ (l,j) \circ \sigma_l = (1) \circ (2) \circ ... \circ (l-2) \circ \xi_1^{(2)} \circ \xi_2^{(2)} \circ ... \circ \xi_u^{(2)},$$

where

$$\xi_z^{(2)} = \begin{cases} \xi_z^{(1)} & \text{if } z \ne s, \\ \text{the cycle whose set of cyclic elements} & \\ \text{contains } l-1,\ i_{l-1}, \text{ and the cyclic elements of } \xi_s^{(1)} & \text{if } z = s, \end{cases}$$

$\forall z \in \langle u \rangle$, and

$$(l-1, i_{l-1}) \circ (l,k) \circ \sigma_l = (1) \circ (2) \circ ... \circ (l-2) \circ \varphi_1^{(2)} \circ \varphi_2^{(2)} \circ ... \circ \varphi_u^{(2)},$$

where

$$\varphi_z^{(2)} = \begin{cases} \gamma_z^{(1)} & \text{if } z \neq t, \\ \text{the cycle whose set of cyclic elements} & \\ \text{contains } l-1,\ i_{l-1},\ \text{and the cyclic elements of } \gamma_t^{(1)} & \text{if } z = t, \end{cases}$$

$\forall z \in \langle u \rangle$. Consequently, if $i_{l-1} > l-1$, then

$$N\left((l-1, i_{l-1}) \circ (l,j) \circ \sigma_l\right) = N\left((l-1, i_{l-1}) \circ (l,k) \circ \sigma_l\right).$$

Finally, for $i_{l-1} \geq l-1$ ($i_{l-1} = l-1$ or $i_{l-1} > l-1$), we have

$$N\left((l-1, i_{l-1}) \circ (l,j) \circ \sigma_l\right) = N\left((l-1, i_{l-1}) \circ (l,k) \circ \sigma_l\right).$$

Proceeding in this way for

$$(l-2, i_{l-2}) \circ (l-1, i_{l-1}) \circ (l,j) \circ \sigma_l \text{ and } (l-2, i_{l-2}) \circ (l-1, i_{l-1}) \circ (l,k) \circ \sigma_l,$$

for

$$(l-3, i_{l-3}) \circ (l-2, i_{l-2}) \circ (l-1, i_{l-1}) \circ (l,j) \circ \sigma_l \text{ and}$$
$$(l-3, i_{l-3}) \circ (l-2, i_{l-2}) \circ (l-1, i_{l-1}) \circ (l,k) \circ \sigma_l,$$

$$\vdots$$

for

$$(1, i_1) \circ (2, i_2) \circ ... \circ (l-1, i_{l-1}) \circ (l,j) \circ \sigma_l \text{ and}$$
$$(1, i_1) \circ (2, i_2) \circ ... \circ (l-1, i_{l-1}) \circ (l,k) \circ \sigma_l,$$

we obtain (finally)

$$N\left((1, i_1) \circ (2, i_2) \circ ... \circ (l-1, i_{l-1}) \circ (l,j) \circ \sigma_l\right) =$$
$$= N\left((1, i_1) \circ (2, i_2) \circ ... \circ (l-1, i_{l-1}) \circ (l,k) \circ \sigma_l\right).$$

Case 2. $j = l$, $k > l$. In this case, we have $(l,j) = (l,l) = (l)$. Further, we proceed in a way similar to that used in Subcase 1.2 — finally, we obtain

$$N\left((1, i_1) \circ (2, i_2) \circ ... \circ (l-1, i_{l-1}) \circ (l,j) \circ \sigma_l\right) =$$
$$= N\left((1, i_1) \circ (2, i_2) \circ ... \circ (l-1, i_{l-1}) \circ (l,k) \circ \sigma_l\right) + 1.$$

Case 3. $j > l$, $k = l$. Similar to Case 2 — finally, we obtain

$$N\left((1, i_1) \circ (2, i_2) \circ ... \circ (l-1, i_{l-1}) \circ (l,j) \circ \sigma_l\right) =$$
$$= N\left((1, i_1) \circ (2, i_2) \circ ... \circ (l-1, i_{l-1}) \circ (l,k) \circ \sigma_l\right) - 1.$$

□

Recall that $\mathbb{R}^+ = \{x \mid x \in \mathbb{R} \text{ and } x > 0\}$.

Let

$$\pi_\sigma = \frac{\theta^{N(\sigma)}}{Z},\ \forall \sigma \in \mathbb{S}_n,$$

where $\theta \in \mathbb{R}^+$ and

$$Z = \sum_{\sigma \in \mathbb{S}_n} \theta^{N(\sigma)}$$

$(n \geq 1)$. Z is known;

$$Z = \theta(\theta+1)\dots(\theta+n-1)$$

(see also Comment 5 from Section 4 — a new computation method for Z is given there). The probability distribution $\pi = (\pi_\sigma)_{\sigma\in\mathbb{S}_n}$ (on $\mathbb{S}_n$) is called the *Ewens distribution,* see, *e.g.*, [1] and [4]. This probability distribution is called so because, from it, we can obtain the Ewens sampling formula, a formula for a probability distribution on

$$\mathbb{A}_n = \{(m_1, m_2, ..., m_n) \mid (m_1, m_2, ..., m_n) \in \mathbb{N}^n \text{ and } m_1 + 2m_2 + ... + nm_n = n\}$$

$(n \geq 1)$. The Ewens sampling formula is

$$P(\{(m_1, m_2, ..., m_n)\}) = \frac{n!}{\theta(\theta+1)\dots(\theta+n-1)} \prod_{j=1}^{n} \frac{\theta^{m_j}}{j^{m_j} m_j!},$$

$\forall (m_1, m_2, ..., m_n) \in \mathbb{A}_n$, where P is the probability on $(\mathbb{A}_n, \mathcal{P}(\mathbb{A}_n))$ ($\mathcal{P}(\mathbb{A}_n)$ is the power set of $\mathbb{A}_n$; $(\mathbb{A}_n, \mathcal{P}(\mathbb{A}_n))$ is a measurable space),

$$P(B) = \begin{cases} \sum\limits_{(m_1,m_2,...,m_n)\in B} \frac{n!}{\theta(\theta+1)\dots(\theta+n-1)} \prod\limits_{j=1}^{n} \frac{\theta^{m_j}}{j^{m_j} m_j!} & \text{if } \emptyset \neq B \subseteq \mathbb{A}_n, \\ 0 & \text{if } B = \emptyset, \end{cases}$$

$P(\{(m_1, m_2, ..., m_n)\})$ is the probability of $\{(m_1, m_2, ..., m_n)\}$, and $\theta \in \mathbb{R}^+$, see [5], see, *e.g.*, also [2], [6], and [17]. This formula is used in genetics and other fields. Below we derive this formula from the formula of Ewens distribution, $\pi_\sigma = \frac{\theta^{N(\sigma)}}{Z}$, $\forall \sigma \in \mathbb{S}_n$.

Let $\sigma \in \mathbb{S}_n$. σ can be written as a composition of pair-wise disjoint cycles. Let $k_i(\sigma)$ be the number of pair-wise disjoint cycles of length i of σ, where $i \in \langle n \rangle$. The vector $k(\sigma) = (k_1(\sigma), k_2(\sigma), ..., k_n(\sigma))$ is called the *cycle structure vector of* σ (see, *e.g.*, [1]).

Note that $k(\sigma) \in \mathbb{A}_n$. Let $(m_1, m_2, ..., m_n) \in \mathbb{A}_n$. We have (see, *e.g.*, also [1]), using the Cauchy formula on permutations (see, *e.g.*, [18]-[19]),

$$P_{\mathbb{S}_n}(\{\sigma \mid \sigma \in \mathbb{S}_n \text{ and } k(\sigma) = (m_1, m_2, ..., m_n)\}) =$$

$$= \sum_{\sigma\in\mathbb{S}_n,\ k(\sigma)=(m_1,m_2,...,m_n)} \pi_\sigma = \sum_{\sigma\in\mathbb{S}_n,\ k(\sigma)=(m_1,m_2,...,m_n)} \frac{\theta^{m_1+m_2+...+m_n}}{Z} =$$

$$= \frac{\theta^{m_1+m_2+...+m_n}}{Z} \cdot |\{\sigma \mid \sigma \in \mathbb{S}_n \text{ and } k(\sigma) = (m_1, m_2, ..., m_n)\}| =$$

$$= \frac{\theta^{m_1+m_2+...+m_n}}{Z} \cdot \left(n! \prod_{j=1}^{n} \frac{1}{j^{m_j} m_j!} \right) = \frac{n!}{\theta(\theta+1)\dots(\theta+n-1)} \prod_{j=1}^{n} \frac{\theta^{m_j}}{j^{m_j} m_j!} =$$

$$= P(\{(m_1, m_2, ..., m_n)\}),$$

therefore, we obtained the Ewens sampling formula from the formula of Ewens distribution, where — it is obvious or almost obvious — $P_{\mathbb{S}_n}$ is the probability on $(\mathbb{S}_n, \mathcal{P}(\mathbb{S}_n))$,

$$P_{\mathbb{S}_n}(A) = \begin{cases} \sum\limits_{\sigma\in A} \pi_\sigma & \text{if } \emptyset \neq A \subseteq \mathbb{S}_n, \\ 0 & \text{if } A = \emptyset. \end{cases}$$

Above we were forced to use "P" with the subscript $\mathbb{S}_n$. When no confusion can arise, we use "P" for probability.

3. A basic property of Ewens distribution

In this section, we show that the Ewens distribution on $\mathbb{S}_n$ is a wavy probability distribution with respect to an order relation and n partitions which will be specified — recall that the fact that the number of partitions is n is important.

Let $n \geq 2$. Set

$$W_{(i_1,i_2,...,i_l)} = \{(1,i_1) \circ (2,i_2) \circ ... \circ (l,i_l) \circ \sigma_l \mid \sigma_l \in \mathbb{S}_n, \sigma_l(v) = v, \forall v \in \langle l \rangle\},$$

$\forall l \in \langle n-1 \rangle$, $\forall i_1, i_2, ..., i_l \in \langle n \rangle$, $1 \leq i_1 \leq n$, $2 \leq i_2 \leq n, ..., l \leq i_l \leq n$.

Theorem 3.1.

$$\left(W_{(i_1,i_2,...,i_l)}\right)_{\substack{i_1,i_2,...,i_l \in \langle n \rangle \\ 1 \leq i_1 \leq n \\ 2 \leq i_2 \leq n \\ \vdots \\ l \leq i_l \leq n}}$$

is a partition of $\mathbb{S}_n$ $(n \geq 2)$, $\forall l \in \langle n-1 \rangle$.

Proof. We have

$$\bigcup_{\substack{i_1,i_2,...,i_l \in \langle n \rangle \\ 1 \leq i_1 \leq n \\ 2 \leq i_2 \leq n \\ \vdots \\ l \leq i_l \leq n}} W_{(i_1,i_2,...,i_l)} = \mathbb{E}_{n,l} = \mathbb{S}_n, \ \forall l \in \langle n-1 \rangle$$

(see Theorem 2.1).

Now, we show that

$$W_{(i_1,i_2,...,i_l)} \cap W_{(j_1,j_2,...,j_l)} = \emptyset$$

if $\exists u \in \langle l \rangle$ such that $i_u \neq j_u$, where $l \in \langle n-1 \rangle$, $i_1, j_1, i_2, j_2, ..., i_l, j_l \in \langle n \rangle$, $1 \leq i_1$, $j_1 \leq n$, $2 \leq i_2$, $j_2 \leq n, ..., l \leq i_l$, $j_l \leq n$. Suppose that $\exists u \in \langle l \rangle$ with $i_u \neq j_u$ such that

$$W_{(i_1,i_2,...,i_l)} \cap W_{(j_1,j_2,...,j_l)} \neq \emptyset.$$

Let $\omega \in W_{(i_1,i_2,...,i_l)} \cap W_{(j_1,j_2,...,j_l)}$. We have

$$\omega = (1,i_1) \circ (2,i_2) \circ ... \circ (l,i_l) \circ \sigma_l = (1,j_1) \circ (2,j_2) \circ ... \circ (l,j_l) \circ \tau_l,$$

where $\sigma_l, \tau_l \in \mathbb{S}_n$, $\sigma_l(v) = \tau_l(v) = v, \forall v \in \langle l \rangle$. Proceeding as in the proof of Theorem 2.1, we obtain

$$i_1 = j_1, \ i_2 = j_2, \ ..., \ i_l = j_l, \ \sigma_l = \tau_l.$$

Therefore, we obtained a contradiction. □

Set the partitions (this can now be done)

$$\Delta_1 = (\mathbb{S}_n),$$

$$\Delta_{l+1} = \left(W_{(i_1,i_2,...,i_l)}\right)_{\substack{i_1,i_2,...,i_l \in \langle n \rangle \\ 1 \leq i_1 \leq n \\ 2 \leq i_2 \leq n \\ \vdots \\ l \leq i_l \leq n}},$$

$\forall l \in \langle n-1 \rangle$. Obviously, we have $\Delta_n = (\{\sigma\})_{\sigma \in \mathbb{S}_n}$.