Brazilian Journal of Probability and Statistics
2017, Vol. 31, No. 2, 338–352
DOI: 10.1214/16-BJPS316

G method in action: Fast exact sampling from set of permutations of order *n* according to Mallows model through Cayley metric

Udrea Păun

Romanian Academy

Abstract. Using G method, we give a fast exact (not approximate) Markovian method for sampling from $\mathbb{S}_n$, the set of permutations of order n, according to the Mallows model through Cayley metric (a model for ranked data). This method has something in common with the cyclic Gibbs sampler and something in common with the swapping method. The number of steps of our method is equal to the number of steps of swapping method, that is, $n-1$; 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). But, besides sampling, we can do other things for the Mallows model through Cayley metric—we compute the normalizing constant and, by Uniqueness theorem, certain important probabilities.

1 The basic result we need

In this section, we present the basic result from Păun (2010) we need.

Set

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

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

Definition 1.1. Let $\Delta_1, \Delta_2 \in \text{Par}(E)$. 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.

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

Key words and phrases. G method, exact sampling, Gibbs sampler in a generalized sense, swapping method, Cayley metric, Mallows model, normalizing constant, important probabilities.

Received March 2015; accepted February 2016.

Set

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

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

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

Set

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

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 (Păun (2010)). Let $P\in N_{m,n}$. Let $\Delta\in\operatorname{Par}(\langle m\rangle)$ and $\Sigma\in\operatorname{Par}(\langle n\rangle)$. 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 (Păun (2010)). Let $P\in N_{m,n}$. Let $\Delta\in\operatorname{Par}(\langle m\rangle)$ and $\Sigma\in\operatorname{Par}(\langle n\rangle)$. 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 $\Delta_1\in\operatorname{Par}(\langle m\rangle)$ and $\Delta_2\in\operatorname{Par}(\langle n\rangle)$. Set (see Păun (2010) for G_{Δ_1,Δ_2} and Păun (2011) 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*.

Below, we give the basic result from Păun (2010) we need.

Theorem 1.5 (Păun (2010)). *Let* $P_1 \in G_{(\langle m_1 \rangle),\Delta_2} \subseteq S_{m_1,m_2}$, $P_2 \in G_{\Delta_2,\Delta_3} \subseteq S_{m_2,m_3}, \dots, P_{n-1} \in G_{\Delta_{n-1},\Delta_n} \subseteq S_{m_{n-1},m_n}$, $P_n \in G_{\Delta_n,(\{i\})_{i\in\langle m_{n+1}\rangle}} \subseteq S_{m_n,m_{n+1}}$. *Then*

$$P_1 P_2 \cdots P_n$$

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

Proof. See Păun (2010). (Theorem 1.5 is part of Theorem 2.10 from Păun (2010); a generalization of Theorem 2.10 from Păun (2010) is Theorem 1.6 from Păun (2011).) □

2 The Markovian method

In this section, we present the Mallows model and our fast Markovian method for sampling exactly (not approximately) from $\mathbb{S}_n$, the set of permutations of order n, according to the Mallows model through Cayley metric. In addition to sampling, for this special Mallows model, we compute the normalizing constant and, by Uniqueness theorem, certain important probabilities.

Consider the group $(\mathbb{S}_n, \circ)$, where $\circ$ is the usual composition of functions. (u, v) is a transposition, $\forall u, v \in \langle n \rangle$, $u \neq v$. Set $(u, u) = \mathrm{Id}$, $\forall u \in \langle n \rangle$, where Id is the identity permutation.

Theorem 2.1. *Let* $n \geq 2$. *Let* $\sigma_0 \in \mathbb{S}_n$. *Let*

$$\mathbb{M}_{n,l} = \{\sigma_0 \circ (1, i_1) \circ (2, i_2) \circ \cdots \circ (l, i_l) \circ \sigma_l | i_1, i_2, \dots, i_l \in \langle n \rangle, 1 \leq i_1 \leq n, \\ 2 \leq i_2 \leq n, \dots, l \leq i_l \leq n, \sigma_l \in \mathbb{S}_n, \sigma_l(v) = v, \forall v \in \langle l \rangle\} \qquad \forall l \in \langle n-1 \rangle.$$

Then

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

Proof. Let $l \in \langle n-1 \rangle$. Since $(\mathbb{S}_n, \circ)$ is a group, we have $\mathbb{M}_{n,l} \subseteq \mathbb{S}_n$. Therefore, $|\mathbb{M}_{n,l}| \leq |\mathbb{S}_n| = n!$ ($|\cdot|$ is the cardinal). To finish the proof, we show that $|\mathbb{M}_{n,l}| = n!$.

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 $1 \leq i_1 \leq n$, $2 \leq i_2 \leq n, \dots, l \leq i_l \leq n$, it follows that $|\mathbb{M}_{n,l}|$ is at most equal to

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

We show that

$$\sigma_0 \circ (1, i_1) \circ (2, i_2) \circ \cdots \circ (l, i_l) \circ \sigma_l = \sigma_0 \circ (1, j_1) \circ (2, j_2) \circ \cdots \circ (l, j_l) \circ \tau_l$$

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

"$\Longleftarrow$" Obvious.

"$\Longrightarrow$" σ_0 can be removed. We remove it, so, we suppose that

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

It follows that

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

Therefore,

$$i_1 = j_1.$$

Since $i_1 = j_1$, removing $(1, i_1)$ and $(1, j_1)$, we have

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

It follows that

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

Therefore,

$$i_2 = j_2.$$

Proceeding in this way, we obtain

$$i_1 = j_1, \qquad i_2 = j_2, \qquad \ldots, \qquad i_l = j_l,$$

and, as a result of these equations,

$$\sigma_l = \tau_l.$$

We conclude that

$$|\mathbb{M}_{n,l}| = n!. \qquad \square$$

Theorem 2.1 says that we can work with $\mathbb{M}_{n,l}$ instead of $\mathbb{S}_n$, $\forall l \in \langle n-1\rangle$ (this fact will be used in Theorem 2.3).

Let $C(\sigma, \tau) =$ minimum number of transpositions required to bring σ to τ, $\forall \sigma, \tau \in \mathbb{S}_n$. C is a metric on $\mathbb{S}_n$, called the *Cayley metric* (see, e.g., Diaconis and Saloff-Coste (1998)).

Theorem 2.2. *Let $n \geq 2$. Let $\sigma_0 \in \mathbb{S}_n$. Consider on $\mathbb{S}_n$ the Cayley metric. Then*

$$C\big(\sigma_0 \circ (1, i_1) \circ (2, i_2) \circ \cdots \circ (l-1, i_{l-1}) \circ (l, j) \circ \sigma_l, \sigma_0\big)$$
$$= \begin{cases} C\big(\sigma_0 \circ (1, i_1) \circ (2, i_2) \circ \cdots \circ (l-1, i_{l-1}) \circ (l, k) \circ \sigma_l, \sigma_0\big) \\ \quad \text{if } j = k = l \text{ or } j, k > l, \\ C\big(\sigma_0 \circ (1, i_1) \circ (2, i_2) \circ \cdots \circ (l-1, i_{l-1}) \circ (l, k) \circ \sigma_l, \sigma_0\big) - 1 \\ \quad \text{if } j = l, k > l, \\ C\big(\sigma_0 \circ (1, i_1) \circ (2, i_2) \circ \cdots \circ (l-1, i_{l-1}) \circ (l, k) \circ \sigma_l, \sigma_0\big) + 1 \\ \quad \text{if } j > l, k = l, \end{cases}$$

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

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

Subcase 1.1. $j = k = l$. Obvious ($(l, j) = (l, k) = \mathrm{Id}$).

Subcase 1.2. $j, k > l$. It is known that $(u, v) \circ (u, v) = \mathrm{Id}$, $\forall (u, v)$, (u, v) is a transposition, $\psi_1 \circ \psi_2 = \psi_2 \circ \psi_1$, $\forall \psi_1, \psi_2 \in \mathbb{S}_n$, ψ_1 and ψ_2 are disjoint cycles, any permutation can be factored uniquely (leaving the order of factors aside) into a product of pair-wise disjoint cycles ("factored", "factors", and "product" are improper words), and any cycle of length s ($s \ge 2$) can be factored into a product of $s-1$ transpositions in s different ways. Since $\sigma_l(v) = v$, $\forall v \in \langle l\rangle$, it follows that, for any cycle of σ_l, any factorization of the cycle into a product of $s-1$ transpositions, where s is the length of cycle ("factorization" is an improper word), does not contain the transpositions $(1, i_1), (2, i_2), \ldots, (l-1, i_{l-1}), (l, j)$, and (l, k). So,

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

cannot be simplified. The equation we need prove is now obvious.

Case 2. $j = l$, $k > l$. Since $j = l$, we have $(l, j) = \mathrm{Id}$. Proceeding similar to Subcase 1.2,

$$(1, i_1) \circ (2, i_2) \circ \cdots \circ (l-1, i_{l-1}) \circ \sigma_l,$$
$$(1, i_1) \circ (2, i_2) \circ \cdots \circ (l-1, i_{l-1}) \circ (l, k) \circ \sigma_l$$

cannot be simplified. Therefore, the equation we need prove holds.

Case 3. $j > l$, $k = l$. Similar to Case 2. □

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

Let

$$\pi_\sigma = \frac{\theta^{d(\sigma,\sigma_0)}}{Z} \qquad \forall \sigma \in \mathbb{S}_n,$$

where $\theta \in \mathbb{R}^+$ (cases of interest: $0 < \theta \le 1$; $\theta > 1$), $\sigma_0 \in \mathbb{S}_n$ ($n \ge 1$), d is a metric on $\mathbb{S}_n$, and

$$Z = \sum_{\sigma \in \mathbb{S}_n} \theta^{d(\sigma,\sigma_0)}.$$

The probability distribution $\pi = (\pi_\sigma)_{\sigma \in \mathbb{S}_n}$ (on $\mathbb{S}_n$) is called the *Mallows model through metric d* (see Mallows (1957); see, e.g., also Critchlow (1985), Diaconis (2009), Diaconis and Saloff-Coste (1998), Fligner and Verducci (1993), and Marden (1995)). This is a model—an exponential model when $\theta \neq 1$—for ranked data (see the above references).

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

Below, we give the main result of this work.

Theorem 2.3. *Let* $n \geq 2$. *Let* $\pi = (\pi_\sigma)_{\sigma \in \mathbb{S}_n}$ *be the Mallows model through Cayley metric. Consider a Markov chain with state space* $\mathbb{S}_n$, *initial probability distribution* p_0, *and transition matrix* $P = P_1 P_2 \cdots P_{n-1}$, *where* P_l, $l \in \langle n-1 \rangle$, *are stochastic matrices on* $\mathbb{S}_n$,

$$(P_l)_{\sigma_0 \circ (1,i_1) \circ (2,i_2) \circ \cdots \circ (l,i_l) \circ \sigma_l \to \xi} = \begin{cases} \dfrac{\pi_{\sigma_0 \circ (1,i_1) \circ (2,i_2) \circ \cdots \circ (l-1,i_{l-1}) \circ (l,j) \circ \sigma_l}}{\sum_{l \leq k \leq n} \pi_{\sigma_0 \circ (1,i_1) \circ (2,i_2) \circ \cdots \circ (l-1,i_{l-1}) \circ (l,k) \circ \sigma_l}} & \text{if } \xi = \sigma_0 \circ (1, i_1) \circ (2, i_2) \circ \cdots \circ (l-1, i_{l-1}) \circ (l, j) \circ \sigma_l \\ & \text{for some } j, l \leq j \leq n, \\ 0 & \text{if } \xi \neq \sigma_0 \circ (1, i_1) \circ (2, i_2) \circ \cdots \circ (l-1, i_{l-1}) \circ (l, j) \circ \sigma_l, \\ & \forall j, l \leq j \leq n, \end{cases}$$

$\forall l \in \langle n-1 \rangle$—($(1, i_1) \circ (2, i_2) \circ \cdots \circ (l-1, i_{l-1})$ *vanishes when* $l = 1$), $\forall i_1, i_2, \ldots, i_l \in \langle n \rangle$, $1 \leq i_1 \leq n$, $2 \leq i_2 \leq n, \ldots, l \leq i_l \leq n$, $\forall \sigma_l \in \mathbb{S}_n$, $\sigma_l(v) = v$, $\forall v \in \langle l \rangle$, $\forall \xi \in \mathbb{S}_n$, *where* σ_0 *is the parameter from* $\mathbb{S}_n$ *of Mallows model through Cayley metric. Then*

$$P = e'\pi$$

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

Proof. Set

$$K_{(i_1,i_2,\ldots,i_l)} = \{\sigma_0 \circ (1, i_1) \circ (2, i_2) \circ \cdots \circ (l, i_l) \circ \sigma_l | \sigma_l \in \mathbb{S}_n, \sigma_l(v) = v, \forall v \in \langle l \rangle\},$$

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

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

(see Theorem 2.1). We show that

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

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

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

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

$$\begin{aligned}\omega &= \sigma_0 \circ (1,i_1) \circ (2,i_2) \circ \cdots \circ (l,i_l) \circ \sigma_l \\ &= \sigma_0 \circ (1,j_1) \circ (2,j_2) \circ \cdots \circ (l,j_l) \circ \tau_l,\end{aligned}$$

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 (σ_0 is removed, ...), we obtain

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

Therefore, we obtained a contradiction.

The above results lead to the fact that

$$(K_{(i_1,i_2,\ldots,i_l)})_{\substack{i_1,i_2,\ldots,i_l \in \langle n \rangle \\ 1 \le i_1 \le n \\ 2 \le i_2 \le n \\ \vdots \\ l \le i_l \le n}}$$

is a partition of $\mathbb{M}_{n,l}$ ($\mathbb{M}_{n,l} = \mathbb{S}_n$), $\forall l \in \langle n-1 \rangle$. Set the partitions (this can now be done)

$$\begin{aligned}\Delta_1 &= (\mathbb{S}_n), \\ \Delta_{l+1} &= (K_{(i_1,i_2,\ldots,i_l)})_{\substack{i_1,i_2,\ldots,i_l \in \langle n \rangle \\ 1 \le i_1 \le n \\ 2 \le i_2 \le n \\ \vdots \\ l \le i_l \le n}},\end{aligned}$$

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

By hypothesis and Theorem 2.2, we have

$$\begin{aligned}&(P_l)_{\sigma_0 \circ (1,i_1) \circ (2,i_2) \circ \cdots \circ (l,i_l) \circ \sigma_l \to \xi} \\ &= \begin{cases} \dfrac{\theta^{C(\sigma_0 \circ (1,i_1) \circ (2,i_2) \circ \cdots \circ (l-1,i_{l-1}) \circ (l,j) \circ \sigma_l, \sigma_0)}}{\sum_{l \le k \le n} \theta^{C(\sigma_0 \circ (1,i_1) \circ (2,i_2) \circ \cdots \circ (l-1,i_{l-1}) \circ (l,k) \circ \sigma_l, \sigma_0)}} & \text{if } \xi = \sigma_0 \circ (1,i_1) \circ (2,i_2) \circ \cdots \circ (l-1,i_{l-1}) \circ (l,j) \circ \sigma_l \\ & \text{for some } j, l \le j \le n, \\ 0 & \text{if } \xi \ne \sigma_0 \circ (1,i_1) \circ (2,i_2) \circ \cdots \circ (l-1,i_{l-1}) \circ (l,j) \circ \sigma_l, \\ & \forall j, l \le j \le n, \end{cases}\end{aligned}$$

$$= \begin{cases} \dfrac{1}{1+(n-l)\theta} & \text{if } \xi = \sigma_0 \circ (1, i_1) \circ (2, i_2) \circ \cdots \circ (l-1, i_{l-1}) \circ (l, l) \circ \sigma_l, \\ \dfrac{\theta}{1+(n-l)\theta} & \text{if } \xi = \sigma_0 \circ (1, i_1) \circ (2, i_2) \circ \cdots \circ (l-1, i_{l-1}) \circ (l, j) \circ \sigma_l \\ & \text{for some } j, l < j \le n, \\ 0 & \text{if } \xi \ne \sigma_0 \circ (1, i_1) \circ (2, i_2) \circ \cdots \circ (l-1, i_{l-1}) \circ (l, j) \circ \sigma_l, \\ & \forall j, l \le j \le n, \end{cases}$$

$\forall l \in \langle n-1 \rangle$, $\forall i_1, i_2, \ldots, i_l \in \langle n \rangle$, $1 \le i_1 \le n$, $2 \le i_2 \le n, \ldots, l \le i_l \le n$, $\forall \sigma_l \in \mathbb{S}_n$, $\sigma_l(v) = v$, $\forall v \in \langle l \rangle$, $\forall \xi \in \mathbb{S}_n$. It follows that

$$P_l \in G_{\Delta_l, \Delta_{l+1}} \qquad \forall l \in \langle n-1 \rangle.$$

Since $P = P_1 P_2 \cdots P_{n-1}$, by Theorem 1.5, P is a stable matrix. Consequently, $\exists \psi$, ψ is a probability distribution on $\mathbb{S}_n$, such that

$$P = e' \psi.$$

It is easy to see that

$$\pi_\sigma (P_l)_{\sigma\tau} = \pi_\tau (P_l)_{\tau\sigma} \qquad \forall l \in \langle n-1 \rangle, \forall \sigma, \tau \in \mathbb{S}_n$$

($\mathbb{S}_n = \mathbb{M}_{n,l}$, $\forall l \in \langle n-1 \rangle$). This thing implies

$$\pi P_l = \pi \qquad \forall l \in \langle n-1 \rangle,$$

and, further,

$$\pi P = \pi.$$

Finally, we have

$$\pi = \pi P = \pi e' \psi = \psi,$$

so,

$$P = e' \pi. \qquad \square$$

We comment on Theorem 2.3 and its proof.

First, we can work with the chain with transition matrix P or, equivalently, with the chain with transition matrices $P_1, P_2, \ldots, P_{n-1}$, $P_1, P_2, \ldots, P_{n-1}, \ldots$ (the former chain is homogeneous while the latter one is nonhomogeneous when $n \ge 3$). We chose the first case. (For finite Markov chain theory, see, e.g., Iosifescu (2007).) Any 1-step of the chain with transition matrix P is performed via $P_1, P_2, \ldots, P_{n-1}$, that is, doing $n-1$ transitions: one using P_1, one using $P_2, \ldots,$ one using P_{n-1}. By Theorem 2.3, the chain with transition matrix P attains its stationarity at time 1 (to attain the stationarity, the chain with transition matrix P makes one step while the other chain makes $n-1$ steps (due to $P_1, P_2, \ldots, P_{n-1}$)).

We constructed the chain with transition matrix P being guided by G method, Theorem 1.5, our hybrid Metropolis–Hastings chain(s) from Păun (2011), and, especially, certain results and suggestions from Păun (2017). The chain with transition matrix P belongs to our collection of hybrid Metropolis–Hastings chains from Păun (2011) (this follows from $K_{(i_1,i_2,\ldots,i_{l+1})} \subset K_{(i_1,i_2,\ldots,i_l)}$, $\forall l \in \langle n-2\rangle$, $\forall i_1, i_2, \ldots, i_{l+1}$, $1 \le i_1 \le n$, $2 \le i_2 \le n, \ldots, l+1 \le i_{l+1} \le n$, etc.; for our collection, see also Păun (2017)). The chain with transition matrix P is a cyclic Gibbs sampler in a generalized sense because the state space is, here, $\mathbb{S}_n$, ratios used to define the transition probabilities of matrices P_l, $l \in \langle n-1\rangle$, are similar to those of (usual) cyclic Gibbs sampler with—keeping the finite framework—finite state space (this chain also belongs to our collection of hybrid Metropolis–Hastings chains from Păun (2011), see Păun (2017)), and matrices $P_1, P_2, \ldots, P_{n-1}$ are used cyclically. (For the Gibbs sampler, see, e.g., Madras (2002).)

Second, to define transition probabilities of P_l, $l \in \langle n-1\rangle$ fixed, we used states from $\mathbb{M}_{n,l}$. So, using P_l, the chain passes from a state, say, γ of $\mathbb{M}_{n,l}$ to a state, say, δ of $\mathbb{M}_{n,l}$ also. For P_{l+1}, when $l+1 \le n-1$, we need states from $\mathbb{M}_{n,l+1}$, so, when we run the chain, we must rewrite δ using the generators of $\mathbb{M}_{n,l+1}$.

Third, there exists a case, a happy case, for which rewriting the states from Second is not necessary, namely, when $\sigma_l = \mathrm{Id}$. So, to avoid rewriting the states, we consider the chain with initial probability distribution p_0 with $(p_0)_{\sigma_0} = 1$ (warning! here we have σ_0 and above we have σ_l). Since $P = e'\pi$, we have

$$p_0 P^m = \pi \qquad \forall m \ge 1, \forall p_0,\ p_0 = \text{initial probability distribution.}$$

So, for the initial probability distribution p_0 with $(p_0)_{\sigma_0} = 1$, the above equations hold as well. From $\sigma_0 = \sigma_0 \circ (1,1) \circ \mathrm{Id} \in \mathbb{M}_{n,1}$ ($\sigma_1 = \mathrm{Id}$), the chain passes in one of the states

$$\begin{aligned} \sigma_0 = \sigma_0 \circ (1,1) &= \sigma_0 \circ (1,1) \circ \mathrm{Id} \in \mathbb{M}_{n,1}, \\ \sigma_0 \circ (1,2) &= \sigma_0 \circ (1,2) \circ \mathrm{Id} \in \mathbb{M}_{n,1}, \\ &\vdots \\ \sigma_0 \circ (1,n) &= \sigma_0 \circ (1,n) \circ \mathrm{Id} \in \mathbb{M}_{n,1}. \end{aligned}$$

Suppose that it passed in the state $\sigma_0 \circ (1,2)$. From $\sigma_0 \circ (1,2) = \sigma_0 \circ (1,2) \circ (2,2) \circ \mathrm{Id} \in \mathbb{M}_{n,2}$ ($\sigma_2 = \mathrm{Id}$), the chain passes in one of the states

$$\begin{aligned} \sigma_0 \circ (1,2) = \sigma_0 \circ (1,2) \circ (2,2) &= \sigma_0 \circ (1,2) \circ (2,2) \circ \mathrm{Id} \in \mathbb{M}_{n,2}, \\ \sigma_0 \circ (1,2) \circ (2,3) &= \sigma_0 \circ (1,2) \circ (2,3) \circ \mathrm{Id} \in \mathbb{M}_{n,2}, \\ &\vdots \\ \sigma_0 \circ (1,2) \circ (2,n) &= \sigma_0 \circ (1,2) \circ (2,n) \circ \mathrm{Id} \in \mathbb{M}_{n,2}. \end{aligned}$$

Suppose that it passed in the state $\sigma_0 \circ (1,2) \circ (2, n-1)$, etc. Therefore, the states are generated proceeding similar to the swapping method, the difference being that, here, we use the probability distributions

$$\left(\frac{1}{1+(n-l)\theta}, \frac{\theta}{1+(n-l)\theta}, \frac{\theta}{1+(n-l)\theta}, \ldots, \frac{\theta}{1+(n-l)\theta}\right), \qquad l \in \langle n-1 \rangle,$$

instead of uniform probability distributions. (For the swapping method, see, e.g., Devroye (1986), pp. 645–646.) The above probability distributions, the former being almost uniform probability distributions—we call them *almost uniform probability distributions* because each of these probability distributions has identical components, excepting at most one of them (all the components are identical when $\theta = 1$)—and the latter, those of swapping method, being uniform probability distributions, are, concerning the implementation, the best ones. To see that this is also true for almost uniform probability distributions, we split each almost uniform probability distribution into two blocks,

$$\left(\frac{1}{1+(n-l)\theta}\right), \qquad \left(\frac{\theta}{1+(n-l)\theta}, \frac{\theta}{1+(n-l)\theta}, \ldots, \frac{\theta}{1+(n-l)\theta}\right).$$

If

$$X > \frac{1}{1+(n-l)\theta}, \qquad X \sim U(0,1),$$

further, we work with the latter block, which, by normalization, leads to the uniform probability distribution

$$\left(\frac{1}{n-l}, \frac{1}{n-l}, \ldots, \frac{1}{n-l}\right).$$

Therefore, our exact sampling Markovian method, having $n-1$ steps, is simple and good.

Fourth, using the equation $P = e'\pi$, we can compute the normalizing constant Z. Set $\Delta \succ \Delta'$ if $\Delta' \preceq \Delta$ and $\Delta \neq \Delta'$, where Δ, $\Delta' \in \mathrm{Par}(E)$, E is a nonempty set. The partitions $\Delta_1 = (\mathbb{S}_n)$, $\Delta_2, \ldots, \Delta_{n-1}$, $\Delta_n = (\{\sigma\})_{\sigma \in \mathbb{S}_n}$ from the proof of Theorem 2.3 have the property: $\Delta_1 \succ \Delta_2 \succ \cdots \succ \Delta_n$ (recall that $K_{(i_1,i_2,\ldots,i_{l+1})} \subset K_{(i_1,i_2,\ldots,i_l)}$, $\forall l \in \langle n-2 \rangle$, $\forall i_1, i_2, \ldots, i_{l+1}$, $1 \le i_1 \le n$, $2 \le i_2 \le n, \ldots, l+1 \le i_{l+1} \le n$). P_l is a block diagonal matrix (eventually by permutation of rows and columns), $\forall l \in \langle n-1 \rangle - \{1\}$, and Δ_l-stable matrix on Δ_l, $\forall l \in \langle n-1 \rangle$ (see First again—the fact that the chain with transition matrix P belongs to our collection of hybrid Metropolis–Hastings chains from Păun (2011)). Moreover, P_l is a Δ_l-stable matrix on Δ_{l+1}, $\forall l \in \langle n-1 \rangle$. Due to these facts, using $P = e'\pi$, it is easy to compute π_{σ_0} (hint: use, directly, $\mathbb{S}_n \supset K_{(1)} \supset K_{(1,2)} \supset \cdots \supset K_{(1,2,\ldots,n-1)} = \{\sigma_0\}$ or, indirectly, the operator $(\cdot)^{-+}$ from Păun (2010)); we have

$$\pi_{\sigma_0} = \frac{1}{1+(n-1)\theta} \cdot \frac{1}{1+(n-2)\theta} \cdot \cdots \cdot \frac{1}{1+\theta}.$$

Since, on the other hand,

$$\pi_{\sigma_0} = \frac{\theta^0}{Z} = \frac{1}{Z},$$

we obtain

$$Z = (1+\theta)(1+2\theta)\cdots\big(1+(n-1)\theta\big).$$

This result is known (see, e.g., Diaconis (2009)), but our computation method is new, simple, and interesting.

Fifth, using Uniqueness theorem from Păun (2017) (the presentation of this result is too long, so, we omit to give it here), we can compute certain important probabilities of the Mallows model through Cayley metric. Indeed, by Uniqueness theorem we have

$$P(K_{(i_1)}) = \sum_{\sigma \in K_{(i_1)}} \pi_\sigma = \begin{cases} \dfrac{1}{1+(n-1)\theta}, & \text{if } i_1 = 1, \\ \dfrac{\theta}{1+(n-1)\theta}, & \text{if } 1 < i_1 \le n. \end{cases}$$

Note that

$$K_{(i_1)} = \{\sigma \,|\, \sigma \in \mathbb{S}_n, \sigma(1) = \sigma_0(i_1)\} \qquad \forall i_1, 1 \le i_1 \le n$$

($K_{(i_1)}$ is the set of permutations from $\mathbb{S}_n$, each permutation having the first component equal to $\sigma_0(i_1)$, the i_1th component of σ_0). Further, by Uniqueness theorem we have

$$\frac{P(K_{(i_1,i_2)})}{P(K_{(i_1)})} = \frac{\sum_{\sigma \in K_{(i_1,i_2)}} \pi_\sigma}{\sum_{\sigma \in K_{(i_1)}} \pi_\sigma} = \begin{cases} \dfrac{1}{1+(n-2)\theta}, & \text{if } i_2 = 2, \\ \dfrac{\theta}{1+(n-2)\theta}, & \text{if } 2 < i_2 \le n, \end{cases}$$

$\forall i_1, 1 \le i_1 \le n$, so,

$$P(K_{(i_1,i_2)}) = \begin{cases} \dfrac{1}{[1+(n-1)\theta][1+(n-2)\theta]}, & \text{if } i_1 = 1, i_2 = 2, \\ \dfrac{\theta}{[1+(n-1)\theta][1+(n-2)\theta]}, & \text{if } i_1 = 1, 2 < i_2 \le n \\ & \text{or } 1 < i_1 \le n, i_2 = 2, \\ \dfrac{\theta^2}{[1+(n-1)\theta][1+(n-2)\theta]}, & \text{if } 1 < i_1 \le n, 2 < i_2 \le n. \end{cases}$$

Note that

$$K_{(i_1,i_2)} = \begin{cases} \{\sigma \,|\, \sigma \in \mathbb{S}_n, \sigma(1) = \sigma_0(i_1), \sigma(2) = \sigma_0(1)\}, & \text{if } i_2 = i_1, \\ \{\sigma \,|\, \sigma \in \mathbb{S}_n, \sigma(1) = \sigma_0(i_1), \sigma(2) = \sigma_0(i_2)\}, & \text{if } i_2 \ne i_1, \end{cases}$$

$\forall i_1,\ 1 \le i_1 \le n,\ \forall i_2,\ 2 \le i_2 \le n$. To compute $P(K_{(i_1,i_2,i_3)})$, etc., we use (see Uniqueness theorem)

$$\frac{P(K_{(i_1,i_2,\ldots,i_u)})}{P(K_{(i_1,i_2,\ldots,i_{u-1})})} = \frac{\sum_{\sigma\in K_{(i_1,i_2,\ldots,i_u)}}\pi_\sigma}{\sum_{\sigma\in K_{(i_1,i_2,\ldots,i_{u-1})}}\pi_\sigma} = \begin{cases} \dfrac{1}{1+(n-u)\theta}, & \text{if } i_u = u,\\ \dfrac{\theta}{1+(n-u)\theta}, & \text{if } u < i_u \le n, \end{cases}$$

$\forall i_1, 1 \le i_1 \le n, \forall i_2, 2 \le i_2 \le n, \ldots, \forall i_{u-1}, u-1 \le i_{u-1} \le n\ (3 \le u \le n-1)$.

Finally, to illustrate Theorem 2.3, its proof, and the above comments, we give an example.

Example 2.4. Consider the Mallows model on $\mathbb{S}_3$ through Cayley metric with $\sigma_0 = (213)$. By Theorem 2.3, we have (the rows and columns of P_1 and P_2 are labeled in lexicographic order)

$$P_1 = \begin{array}{c} (123)\\(132)\\(213)\\(231)\\(312)\\(321) \end{array}\overset{\begin{array}{cccccc}(123) & (132) & (213) & (231) & (312) & (321)\end{array}}{\begin{pmatrix} \frac{\theta}{1+2\theta} & 0 & \frac{1}{1+2\theta} & 0 & \frac{\theta}{1+2\theta} & 0\\ 0 & \frac{\theta}{1+2\theta} & 0 & \frac{1}{1+2\theta} & 0 & \frac{\theta}{1+2\theta}\\ \frac{\theta}{1+2\theta} & 0 & \frac{1}{1+2\theta} & 0 & \frac{\theta}{1+2\theta} & 0\\ 0 & \frac{\theta}{1+2\theta} & 0 & \frac{1}{1+2\theta} & 0 & \frac{\theta}{1+2\theta}\\ \frac{\theta}{1+2\theta} & 0 & \frac{1}{1+2\theta} & 0 & \frac{\theta}{1+2\theta} & 0\\ 0 & \frac{\theta}{1+2\theta} & 0 & \frac{1}{1+2\theta} & 0 & \frac{\theta}{1+2\theta} \end{pmatrix}}$$

and

$$P_2 = \begin{array}{c} (123)\\(132)\\(213)\\(231)\\(312)\\(321) \end{array}\overset{\begin{array}{cccccc}(123) & (132) & (213) & (231) & (312) & (321)\end{array}}{\begin{pmatrix} \frac{1}{1+\theta} & \frac{\theta}{1+\theta} & & & & \\ \frac{1}{1+\theta} & \frac{\theta}{1+\theta} & & & & \\ & & \frac{1}{1+\theta} & \frac{\theta}{1+\theta} & & \\ & & \frac{1}{1+\theta} & \frac{\theta}{1+\theta} & & \\ & & & & \frac{1}{1+\theta} & \frac{\theta}{1+\theta}\\ & & & & \frac{1}{1+\theta} & \frac{\theta}{1+\theta} \end{pmatrix}}.$$

Since, for P_1, we have

$$(123) = (213) \circ (1,2) \circ \mathrm{Id} \in \mathbb{M}_{3,1},$$
$$(132) = (213) \circ (1,2) \circ (2,3) \in \mathbb{M}_{3,1},$$
$$(213) = (213) \circ (1,1) \circ \mathrm{Id} \in \mathbb{M}_{3,1},$$
$$(231) = (213) \circ (1,1) \circ (2,3) \in \mathbb{M}_{3,1},$$
$$(312) = (213) \circ (1,3) \circ \mathrm{Id} \in \mathbb{M}_{3,1},$$
$$(321) = (213) \circ (1,3) \circ (2,3) \in \mathbb{M}_{3,1},$$

it follows that

$$K_{(1)} = \{(213), (231)\}, \qquad K_{(2)} = \{(123), (132)\},$$
$$K_{(3)} = \{(312), (321)\}.$$

Since, for P_2, we have

$$(123) = (213) \circ (1,2) \circ (2,2) \circ \mathrm{Id} \in \mathbb{M}_{3,2},$$
$$(132) = (213) \circ (1,2) \circ (2,3) \circ \mathrm{Id} \in \mathbb{M}_{3,2},$$
$$(213) = (213) \circ (1,1) \circ (2,2) \circ \mathrm{Id} \in \mathbb{M}_{3,2},$$
$$(231) = (213) \circ (1,1) \circ (2,3) \circ \mathrm{Id} \in \mathbb{M}_{3,2},$$
$$(312) = (213) \circ (1,3) \circ (2,2) \circ \mathrm{Id} \in \mathbb{M}_{3,2},$$
$$(321) = (213) \circ (1,3) \circ (2,3) \circ \mathrm{Id} \in \mathbb{M}_{3,2},$$

it follows that

$$K_{(1,2)} = \{(213)\}, \qquad K_{(1,3)} = \{(231)\},$$
$$K_{(2,2)} = \{(123)\}, \qquad K_{(2,3)} = \{(132)\},$$
$$K_{(3,2)} = \{(312)\}, \qquad K_{(3,3)} = \{(321)\}.$$

Further, we have

$$\Delta_1 = (\mathbb{S}_3),$$
$$\Delta_2 = (K_{(1)}, K_{(2)}, K_{(3)}),$$
$$\Delta_3 = (K_{(1,2)}, K_{(1,3)}, K_{(2,2)}, K_{(2,3)}, K_{(3,2)}, K_{(3,3)}).$$

Obviously, $\Delta_1 = (\mathbb{S}_3) \succ \Delta_2 \succ \Delta_3 = (\{\sigma\})_{\sigma \in \mathbb{S}_3}$. It is easy to see that $P_1 \in G_{\Delta_1,\Delta_2}$, $P_2 \in G_{\Delta_2,\Delta_3}$, and $\pi_\sigma (P_l)_{\sigma\tau} = \pi_\tau (P_l)_{\tau\sigma}$, $\forall l \in \langle 2 \rangle$, $\forall \sigma, \tau \in \mathbb{S}_3$. By Theorem 2.3 or direct computation, $P = e'\pi$. Since $\pi_{\sigma_0} = \frac{1}{Z}$, it is easy to see, using $P = e'\pi$, that $Z = (1 + \theta)(1 + 2\theta)$. Obviously, P_2 is a block diagonal matrix and Δ_2-stable matrix on Δ_2. Moreover, P_2 is a Δ_2-stable matrix, see Definition 1.4. P_1 is a stable matrix both on Δ_1 and on Δ_2. By Uniqueness theorem from Păun (2017) or direct computation, we have

$$P(K_{(1)}) = \frac{1}{1+2\theta}, \qquad P(K_{(2)}) = P(K_{(3)}) = \frac{\theta}{1+2\theta},$$

$$P(K_{(1,2)}) = \frac{1}{(1+\theta)(1+2\theta)},$$

$$P(K_{(1,3)}) = P(K_{(2,2)}) = P(K_{(3,2)}) = \frac{\theta}{(1+\theta)(1+2\theta)},$$

$$P(K_{(2,3)}) = P(K_{(3,3)}) = \frac{\theta^2}{(1+\theta)(1+2\theta)}.$$

If the initial state of chain is σ_0, $\sigma_0 = (213)$, then from this state the chain passes in one of the states $(213) \circ (1, 1)$, $(213) \circ (1, 2)$, $(213) \circ (1, 3)$. Suppose that it passed in state $(213) \circ (1, 3)$. $(213) \circ (1, 3) = (312)$. From (312) the chain passes in one of the states $(312) \circ (2, 2)$, $(312) \circ (2, 3)$. Suppose that it passed in state $(312) \circ (2, 3)$. $(312) \circ (2, 3) = (321)$. (321) is the state selected from $\mathbb{S}_3$ with our method, having, here, two $(3 - 1 = 2)$ steps.

Acknowledgments. The author wishes to express his thanks to the referees for their interest in the publication of this article.

References

Critchlow, D. E. (1985). *Metric Methods for Analyzing Partially Ranked Data*. *Lecture Notes in Statistics* **34**. Berlin: Springer. MR0818986

Devroye, L. (1986). *Non-Uniform Random Variate Generation*. New York: Springer. Available at http://cg.scs.carleton.ca/~luc/rnbookindex.html. MR0836973

Diaconis, P. (2009). The Markov chain Monte Carlo revolution. *Bull. Amer. Math. Soc. (N.S.)* **46**, 179–205. MR2476411

Diaconis, P. and Saloff-Coste, L. (1998). What do we know about the Metropolis algorithm? *J. Comput. System Sci.* **57**, 20–36. MR1649805

Fligner, M. A. and Verducci, J. S., eds. (1993). *Probability Models and Statistical Analyses for Ranking Data*. *Lecture Notes in Statistics* **80**. New York: Springer. MR1237197

Iosifescu, M. (2007). *Finite Markov Processes and Their Applications*. Mineola, NY: Dover Publications, Inc. Corrected reprint of the 1980 original. MR2663628

Madras, N. (2002). *Lectures on Monte Carlo Methods*. Providence, RI: American Mathematical Society. MR1870056

Mallows, C. L. (1957). Non-null ranking models, I. *Biometrika* **44**, 114–130. MR0087267

Marden, J. I. (1995). *Analyzing and Modeling Rank Data*. London: Chapman & Hall. MR1346107

Păun, U. (2010). G_{Δ_1,Δ_2} in action. *Rev. Roumaine Math. Pures Appl.* **55**, 387–406. MR2778116

Păun, U. (2011). A hybrid Metropolis–Hastings chain. *Rev. Roumaine Math. Pures Appl.* **56**, 207–228. MR2952638

Păun, U. (2017). *G* method in action: From exact sampling to approximate one. *Rev. Roumaine Math. Pures Appl.* To appear.

Gheorghe Mihoc–Caius Iacob Institute
of Mathematical Statistics
and Applied Mathematics
Romanian Academy
Calea 13 Septembrie nr. 13
050711 Bucharest 5
Romania
E-mail: paun@csm.ro