A Gibbs sampler in a generalized sense

UDREA PĂUN

ABSTRACT. We present, in the finite case, a generalization of the cyclic Gibbs sampler—the Gibbs sampler for short. The chain obtained we call the cyclic Gibbs sampler in a generalized sense—the Gibbs sampler in a generalized sense for short. The Gibbs sampler in a generalized sense belongs to our collection of hybrid Metropolis-Hastings chains from [U. Păun, A hybrid Metropolis-Hastings chain, Rev. Roumaine Math. Pures Appl. 56 (2011), 207–228] and we conjecture that it is the fastest chain in this collection in a sense which will be specified in article. We then present the wavy probability distributions, a generalization of the wavy probability distribution(s) of first type from [U. Păun, G method in action: from exact sampling to approximate one, Submitted]. We construct a Gibbs sampler in a generalized sense, a special one, for wavy probability distributions which is fast, attaining its stationarity in a finite number of steps, and, besides this, has other important properties—the computation of certain important probabilities iteratively and, for wavy probability distributions having normalization constant, the computation of this constant.

2010 Mathematics Subject Classification. 60J10, 60E99, 62Dxx, 68U20, 82B20.
Key words and phrases. G method, hybrid Metropolis-Hastings chain, Gibbs sampler, Gibbs sampler in a generalized sense, wavy probability distribution, Mallows model, Potts model, normalization constant, important probabilities, exact sampling.

1. Basic things

In this section, we present some basic things.

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 \( \Delta_1 \) is finer than \( \Delta_2 \) if \( \forall V \in \Delta_1, \exists W \in \Delta_2 \) such that \( V \subseteq W \).

Write \( \Delta_1 \preceq \Delta_2 \) when \( \Delta_1 \) is finer than \( \Delta_2 \).

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 \(Z_{ij}\) or, if confusion can arise, \(Z_{i \rightarrow j}\).

Set

\[ \langle m \rangle = \{ 1, 2, ..., m \} \ (m \geq 1) , \]
\[ \langle\langle m \rangle\rangle = \{ 0, 1, ..., m \} \ (m \geq 0) , \]
\[ N_{m,n} = \{ P \mid P \text{ is a nonnegative } m \times n \text{ matrix} \} , \]

Received April 15, 2016.
\[ S_{m,n} = \{ P \mid P \text{ is a stochastic } m \times n \text{ matrix} \}, \]
\[ N_n = N_{n,n}, \]
\[ S_n = S_{n,n}. \]

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
\[ P_U = (P_{ij})_{i \in U, j \in \langle n \rangle}, \quad P_V = (P_{ij})_{i \in \langle m \rangle, j \in V}, \quad \text{and} \quad P_{UV} = (P_{ij})_{i \in U, j \in V}. \]

Set
\[ (\{i\})_{i \in \{s_1,s_2,\ldots,s_t\}} = (\{s_1\}, \{s_2\}, \ldots, \{s_t\}); \]
\[ (\{i\})_{i \in \{s_1,s_2,\ldots,s_t\}} \in \text{Par}(\{s_1,s_2,\ldots,s_t\}). \]

E.g.,
\[ (\{i\})_{i \in \langle n \rangle} = (\{1\}, \{2\}, \ldots, \{n\}). \]

**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 \( \Delta \in \text{Par}(\langle m \rangle) \) and \( \Sigma \in \text{Par}(\langle n \rangle). \) We say that \( P \) is a \( [\Delta] \)-stable matrix on \( \Sigma \) if \( P_{K\ell}^L \) 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 \( \Delta \in \text{Par}(\langle m \rangle) \) and \( \Sigma \in \text{Par}(\langle n \rangle). \) We say that \( P \) is a \( \Delta \)-stable matrix on \( \Sigma \) if \( \Delta \) is the least fine partition for which \( P \) is a \([\Delta]\)-stable matrix on \( \Sigma. \) In particular, a \( \Delta \)-stable matrix on \( (\{i\})_{i \in \langle n \rangle} \) is called \( \Delta \)-stable while a \((\langle m \rangle)\)-stable matrix on \( \Sigma \) is called stable on \( \Sigma \) for short. A stable matrix on \((\{i\})_{i \in \langle n \rangle}\) is called stable for short.

Let \( \Delta_1 \in \text{Par}(\langle m \rangle) \) and \( \Delta_2 \in \text{Par}(\langle n \rangle). \) Set (see \[8\] for \( G_{\Delta_1,\Delta_2} \) 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 \( G_{\Delta_1,\Delta_2} \) 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 result.

**Theorem 1.1.** \[8\] Let \( P_1 \in G_{\langle \langle m_1 \rangle \rangle,\Delta_2} \subseteq S_{m_1,m_2}, \) \( P_2 \in G_{\Delta_2,\Delta_3} \subseteq S_{m_2,m_3}, \) \( \ldots, \) \( P_{n-1} \in G_{\Delta_{n-1},\Delta_n} \subseteq S_{m_{n-1},m_n}, \) \( P_n \in G_{\Delta_n,\langle \{i\}\rangle_{i \in \langle m_{n+1} \rangle}} \subseteq S_{m_{n+1},m_{n+1}}. \) Then
\[ P_1 P_2 \ldots P_n \]
is a stable matrix (i.e., a matrix with identical rows, see **Definition 1.4**).

**Proof.** See \[8\].

**Definition 1.5.** (See, e.g., \[14, p. 80\].) 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
\[
\overline{P} = (\overline{P}_{ij}) \in N_{m,n}, \overline{P}_{ij} = \begin{cases} 1 & \text{if } P_{ij} > 0, \\ 0 & \text{if } P_{ij} = 0, \end{cases}
\]
\( \forall i \in \langle m \rangle, \forall j \in \langle n \rangle \). We call \( \overline{P} \) the incidence matrix of \( P \) (see, e.g., [5, 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 \( S = \langle r \rangle \). (We work with this set for simplification; \( S \) can be any nonempty finite set.) Let \( \pi = (\pi_i)_{i \in S} = (\pi_1, \pi_2, ..., \pi_r) \) 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 \( \Delta, \Delta' \in \text{Par}(E) \).

Let \( \Delta_1, \Delta_2, ..., \Delta_{t+1} \in \text{Par}(S) \) 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 \( Q_1, Q_2, ..., Q_t \in S_r \) such that
\[
\begin{align*}
\text{(C1)} \; &\overline{Q}_1, \overline{Q}_2, ..., \overline{Q}_t \text{ are symmetric matrices;} \\
\text{(C2)} \; & (Q_l)_K^L = 0, \forall l \in \langle t \rangle - \{1\}, \forall K, L \in \Delta_l, K \neq L \text{ (this assumption implies that } Q_l \text{ is a block diagonal matrix and } \Delta_l\text{-stable matrix on } \Delta_l, \forall l \in \langle t \rangle - \{1\}); \\
\text{(C3)} \; & (Q_l)_K^U \text{ is a row-allowable matrix, } \forall l \in \langle t \rangle, \forall K \in \Delta_l, \forall U \in \Delta_{l+1}, U \subseteq K.
\end{align*}
\]

Although \( Q_l, l \in \langle t \rangle \), are not irreducible matrices if \( l \geq 2 \), we define the matrices \( P_l, l \in \langle t \rangle \), as in the Metropolis-Hastings case (see, e.g., [6, pp. 63–66] for this case), namely,
\[
P_l = (P_l)_{ij} \in S_r,
\]
\[
(P_l)_{ij} = \begin{cases} 0 & \text{if } j \neq i \text{ and } (Q_l)_{ij} = 0, \\ (Q_l)_{ij} \min \left( 1, \frac{\pi_j(Q_l)_{ij}}{\pi_i(Q_l)_{ij}} \right) & \text{if } j \neq i \text{ and } (Q_l)_{ij} > 0, \\ 1 - \sum_{k \neq i} (P_l)_{ik} & \text{if } j = i, \end{cases}
\]
\( \forall l \in \langle t \rangle \). 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]. \( \square \)

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.
2. Our Gibbs sampler in a generalized sense

In this section, we consider a generalization of the cyclic Gibbs sampler — the Gibbs sampler for short — which we call the cyclic Gibbs sampler in a generalized sense — the Gibbs sampler in a generalized sense for short. The Gibbs sampler in a generalized sense belongs to our collection of hybrid Metropolis-Hastings chains presented in Section 1 and, moreover, we conjecture that it is the fastest chain in this collection in a sense which will be specified in this section.

**Theorem 2.1.** Consider a hybrid Metropolis-Hastings chain with state space \( S = \langle r \rangle \) and transition matrix \( P = P_1 P_2 ... P_l, P_1, P_2, ..., P_l \) corresponding to \( Q_1, Q_2, ..., Q_l \), respectively. Suppose that \( \forall l \in \langle t \rangle, \forall i, j \in S, \)

\[
(Q_l)_{ij} = \frac{\pi_j}{\sum_{k \in S, (Q_l)_{ik} > 0} \pi_k} \quad \text{if} \quad (Q_l)_{ij} > 0
\]

(see Section 1 again for \( Q_l, l \in \langle t \rangle \), \( \pi = (\pi_i)_{i \in S} \), \( ... \)). Then

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

**Proof.** Obvious. \( \square \)

We call the hybrid Metropolis-Hastings chain from Theorem 2.1 the *cyclic Gibbs sampler in a generalized sense* — the Gibbs sampler in a generalized sense for short. Recall that we work with \( S = \langle r \rangle \) for simplification; \( S \), here, can be any finite set with \( |S| \geq 2 \) (\(| \cdot |\) is the cardinal).

We use the convention that an empty term vanishes.

Let \( \pi \) be a positive probability distribution on \( S = \langle \langle h \rangle \rangle^n, h, n \geq 1 \) (more generally, \( S = \langle \langle h_1 \rangle \rangle \times \langle \langle h_2 \rangle \rangle \times ... \times \langle \langle h_n \rangle \rangle, h_1, h_2, ..., h_n, n \geq 1 \)). Let \( x = (x_1, x_2, ..., x_n) \in S \). Set

\[
x[k|l] = (x_1, x_2, ..., x_{l-1}, k, x_{l+1}, ..., x_n), \forall k \in \langle \langle h \rangle \rangle, \forall l \in \langle n \rangle.
\]

Obviously, \( x[k|l] \in S, \forall k \in \langle \langle h \rangle \rangle, \forall l \in \langle n \rangle \). The (usual) cyclic Gibbs sampler — the Gibbs sampler for short — on \( S = \langle \langle h \rangle \rangle^n, h, n \geq 1 \) (more generally, on \( S = \langle \langle h_1 \rangle \rangle \times \langle \langle h_2 \rangle \rangle \times ... \times \langle \langle h_n \rangle \rangle, h_1, h_2, ..., h_n, n \geq 1 \)) is a Markov chain with state space \( S \) and transition matrix \( P = P_1 P_2 ... P_n \), where

\[
(P_l)_{xy} = \begin{cases} 
0 & \text{if } y \neq x[k|l], \forall k \in \langle \langle h \rangle \rangle, \\
\frac{\pi_x[k|l]}{\sum_{j \in \langle \langle h \rangle \rangle} \pi_x[j|l]} & \text{if } y = x[k|l] \text{ for some } k \in \langle \langle h \rangle \rangle,
\end{cases}
\]

\( \forall l \in \langle n \rangle, \forall x, y \in S \). (For the Gibbs sampler, see, e.g., [2], [4, Chapter 5], and [6, pp. 69–81].)

The Gibbs sampler on \( \langle \langle h \rangle \rangle^n \) belongs to our collection of hybrid Metropolis-Hastings chains, see [10] — this follows considering the partitions

\[
\Delta_1 = (S),
\]

\[
\Delta_{l+1} = (K_{(x_1, x_2, ..., x_l)})_{x_1, x_2, ..., x_l \in \langle \langle h \rangle \rangle}, \forall l \in \langle n \rangle,
\]

where

\[
K_{(x_1, x_2, ..., x_l)} = \{(y_1, y_2, ..., y_n) | (y_1, y_2, ..., y_n) \in S \text{ and } y_i = x_i, \forall i \in \langle l \rangle \}.
\]
\( \forall l \in \langle n \rangle, \forall x_1, x_2, ..., x_l \in \langle \langle h \rangle \rangle \) (so, \( \Delta_1 = (S) \succ \Delta_2 \succ ... \succ \Delta_{n+1} = (\{x\})_{x \in S} \)), and matrices \( Q_l, l \in \langle n \rangle \),
\[
Q_l = P_l, \forall l \in \langle n \rangle .
\]

Obviously – now, it is obvious –, the Gibbs sampler on \( \langle \langle h \rangle \rangle^n \) is a special case of our Gibbs sampler in a generalized sense on \( \langle \langle h \rangle \rangle^n \).

**Remark 2.2.** The matrices \( Q_l, l \in \langle n \rangle \), of Gibbs sampler on \( \langle \langle h \rangle \rangle^n \) have the property:
\[
(Q_l)_{ij}^U \text{ has in each row just one positive entry,}
\]
\[
\forall l \in \langle n \rangle, \forall K \in \Delta_l, \forall U \in \Delta_{l+1} \text{ with } U \subseteq K .
\]

Therefore, a subcollection – an interesting subcollection – of our collection of Gibbs samplers in a generalized sense is that of the Gibbs samplers in a generalized sense having the above property, i.e., satisfying the condition (c4) from [9].

Let \( P = (P_{ij}) \in N_{m,n} \) – \( P \) is called 0-1 matrix if \( P_{ij} \in \{1\}, \forall i \in \langle m \rangle, \forall j \in \langle n \rangle \).

We conjecture that the Gibbs sampler in a generalized sense is the fastest chain in our collection of hybrid Metropolis-Hastings chains in a sense which is specified below.

**Conjecture 2.3.** Fix \( S = \langle r \rangle \). Fix \( \Delta_1, \Delta_2, ..., \Delta_{t+1} \in \text{Par}(S) \) with \( \Delta_1 = (S) \succ \Delta_2 \succ ... \succ \Delta_{t+1} = (\{i\})_{i \in S} \), where \( t \geq 1 \). Fix the 0-1 matrices \( V_1, V_2, ..., V_t \in N_r \) having the properties (C1)-(C3) from Section 1 – we extend these properties for matrices from \( N_r \). Fix the target probability distribution \( \pi = (\pi_i)_{i \in S} \) (on \( S \)). Fix \( w \), a probability distribution on \( S \). Consider the subcollection of hybrid Metropolis-Hastings chains corresponding to \( \pi \) (each chain from this subcollection is constructed using \( \pi \)) with \( Q_1, Q_2, ..., Q_t \in S_r \) having the properties (C1)-(C3) and, moreover,
\[
\forall l \in \langle t \rangle
\]
\[
Q_l = V_t,
\]
(these equations simply say that all the chains from this subcollection have the same neighbor system; \( V_1, V_2, ..., V_t \) are fixed, not \( Q_1, Q_2, ..., Q_t \) – the latter are only fixed for each chain from subcollection). Then the Gibbs sampler in a generalized sense from this subcollection (there exists a unique Gibbs sampler in a generalized sense in this subcollection) is the fastest chain in this subcollection, i.e., for any hybrid Metropolis-Hastings chain from this subcollection,
\[
\| p_n - \pi \|_1 \leq \| s_n - \pi \|_1, \forall n \geq 1 ,
\]
where \( p_n \) = the probability distribution of Gibbs sampler in a generalized sense (from this subcollection) at time \( n \), \( s_n \) = the probability distribution of hybrid Metropolis-Hastings chain at time \( n \), \( \forall n \geq 0 \), \( p_0 = s_0 = w \) (all the chains from this subcollection have the same initial probability distribution, \( w \)).

Obviously, the word “fastest” from the above conjecture refers to Markov chains strictly, not to computers. The running time of our hybrid chains on a computer is another matter (the computational cost per step is the main problem; on a computer, a step of a Markov chain can be performed or not).

Conjecture 2.3 is supported by [11]-[13]; each of these articles is based on a Gibbs sampler in a generalized sense which is fast – we refer, here, both the Markov chains and computers – and, moreover, has other interesting properties. (For more information, see Section 3.)
3. Wavy probability distributions

In this section, we construct a class of probability distributions which contains the wavy probability distribution(s) of first type from [10] as a special case. The probability distributions from this class we call the wavy probability distributions.

Let \( S = \langle r \rangle \). Let \( \pi = (\pi_i)_{i \in S} \) be a positive probability distribution on \( S \). Let \( \Delta_1, \Delta_2, \ldots, \Delta_{t+1} \in \text{Par}(S) \) with \( \Delta_1 = (S) > \Delta_2 > \ldots > \Delta_{t+1} = \langle \{ i \} \rangle_{i \in S} \), where \( t \geq 1 \).

\( \Delta_1 > \Delta_2 \) implies \( r \geq 2 \). Consider that \( \Delta_l = (K^{(l)}_1, K^{(l)}_2, \ldots, K^{(l)}_{u_l}) \), \( K^{(l)}_1 \) having the first \( |K^{(l)}_1| \) elements of \( S \), \( K^{(l)}_2 \) having the next \( |K^{(l)}_2| \) elements of \( S \) (this condition and the next ones vanish when \( l = 1 \)), \ldots, \( K^{(l)}_{u_l} \) having the last \( |K^{(l)}_{u_l}| \) elements of \( S \), \( \forall l \in (t + 1) \). Consider that

\[
(\text{c1}) \ |K^{(l)}_1| = |K^{(l)}_2| = \ldots = |K^{(l)}_{u_l}|, \forall l \in (t + 1) \text{ with } u_l \geq 2;
\]

\[
(\text{c2}) \ r = r_1r_2\ldots r_t \text{ with } r_1r_2\ldots r_t = |\Delta_{t+1}|, \forall l \in (t - 1), \text{ and } r_t = |K^{(l)}_1| \text{ (this condition is compatible with } \Delta_1 > \Delta_2 > \ldots > \Delta_{t+1}).
\]

The above conditions are noted (c1) and (c2) as in [9]. The condition (c2) is superfluous because it follows from (c1) and \( \Delta_1 > \Delta_2 > \ldots > \Delta_{t+1} \).

We have

\[
K^{(l)}_v = \bigcup_{s \in D_{v,b_l} \cup \{ v b_l \}} K^{(l+1)}_s, \forall l \in (t), \forall v \in (u_l),
\]

where

\[
b_l = \frac{|\Delta_{l+1}|}{|\Delta_l|}, \forall l \in (t),
\]

and

\[
D_{v,b_l} = \{(v - 1) b_l + 1, (v - 1) b_l + 2, \ldots, v b_l - 1\}, \forall l \in (t), \forall v \in (u_l).
\]

Suppose that \( \forall l \in (t), \forall v \in (u_l), \forall s \in D_{v,b_l}, \exists \alpha^{(l,v)}_s > 0 \) such that

\[
\pi_i d^{(l,v)}_s = \alpha^{(l,v)}_s \pi_i \text{ (direct proportionality), } \forall i \in K^{(l+1)}_{(v-1)b_l+1},
\]

where

\[
d^{(l,v)}_s = |K^{(l+1)}_{(v-1)b_l+1}| + |K^{(l+1)}_{(v-1)b_l+2}| + \ldots + |K^{(l+1)}_s|,
\]

\( \forall l \in (t), \forall v \in (u_l), \forall s \in D_{v,b_l} \).

**Definition 3.1.** The probability distribution \( \pi = (\pi_i)_{i \in S} \) having the above property we call the *wavy probability distribution*.

To define the wavy probability distribution, we considered \( S = \langle r \rangle \) \( (r \geq 2) \) for simplification; moreover, \( S = \langle r \rangle \) was equipped with the usual order relation \( < (1 < 2 < \ldots < r) \) for simplification too. \( S \) can be any finite set for which the conditions (c1) and (c2) hold, setting \( r = |S| \). In this case, supposing that \( \pi = (\pi_i)_{i \in S} \) is given, to see if \( \pi \) is a wavy probability distribution on \( S \), we must consider, excepting \( S = \langle r \rangle \) equipped with the usual order relation \( < \), order relations on \( S \) or bijective functions (i.e., coding functions; for coding functions, see, e.g., [3, p. 643]) from \( S \) to \( |S| \), equipping the latter set with the usual order relation \( < \).
Recall that \( \mathbb{R}^+ = \{x \mid x \in \mathbb{R} \text{ and } x > 0\} \). Recall that in [10] the probability distribution
\[
\pi = (c_0, c_0a, ..., c_0a^w, c_1, c_1a, ..., c_1a^w, ..., c_h, c_ha, ..., c_ha^w) \\
(\text{the sequence } c_0, c_0a, ..., c_0a^w, c_1, c_1a, ..., c_1a^w, ..., c_h, c_ha, ..., c_ha^w \text{ appears } (h+1)^{n-t-1} \text{ times if } 0 \leq t < n \text{ and } c_0, c_0a, ..., c_0a^w \text{ only appears if } t = n) \text{ on } (\langle h \rangle)^n \text{ equipped with the lexicographic order, where } w = (h+1)^t - 1, 0 \leq t \leq n, c_0, c_1, ..., c_h, a \in \mathbb{R}^+, \text{ was called the wavy probability distribution of first type. This probability distribution is, according to Definition 3.1, a wavy probability distribution. Below we illustrate this in a simple case.}

**Example 3.1.** Consider \( S = \langle \langle 1 \rangle \rangle^3 \) (\( h = 1, n = 3 \)) equipped with the lexicographic order and
\[
\pi = (c_0, c_0a, c_0a^2, c_0a^3, c_1, c_1a, c_1a^2, c_1a^3) .
\]
\( \pi \) is a wavy probability distribution (according to Definition 3.1). To show this, we consider (see Section 2, where we showed that the (usual) Gibbs sampler belongs to our collection of hybrid Metropolis-Hastings chains) the sets
\[
K_{(0)} = \{(0, 0, 0), (0, 0, 1), (0, 1, 0), (0, 1, 1)\}, \\
K_{(1)} = \{(1, 0, 0), (1, 0, 1), (1, 1, 0), (1, 1, 1)\}, \\
K_{(0,0)} = \{(0, 0, 0), (0, 0, 1)\}, K_{(0,1)} = \{(0, 1, 0), (0, 1, 1)\}, \\
K_{(1,0)} = \{(1, 0, 0), (1, 0, 1)\}, K_{(1,1)} = \{(1, 1, 0), (1, 1, 1)\}, \\
K_{(0,0,0)} = \{(0, 0, 0)\}, K_{(0,0,1)} = \{(0, 0, 1)\}, ..., K_{(1,1,1)} = \{(1, 1, 1)\}
\]
and partitions
\[
\Delta_1 = \langle S \rangle = \langle \langle 1 \rangle \rangle^3 , \\
\Delta_2 = \langle K_{(0)}, K_{(1)} \rangle , \\
\Delta_3 = \langle K_{(0,0)}, K_{(0,1)}, K_{(1,0)}, K_{(1,1)} \rangle , \\
\Delta_4 = \langle K_{(0,0,0)}, K_{(0,0,1)}, ..., K_{(1,1,1)} \rangle .
\]
Obviously,
\[
\Delta_1 \supset \Delta_2 \supset \Delta_3 \supset \Delta_4 , \\
|K_{(0)}| = |K_{(1)}| = 4, \\
|K_{(0,0)}| = |K_{(0,1)}| = |K_{(1,0)}| = |K_{(1,1)}| = 2, \\
|K_{(0,0,0)}| = |K_{(0,0,1)}| = ... = |K_{(1,1,1)}| = 1, \\
\pi_{(1,0,0)} = \frac{c_1}{c_0} \pi_{(0,0,0)}, \pi_{(1,0,1)} = \frac{c_1}{c_0} \pi_{(0,0,1)}, \\
\pi_{(1,1,0)} = \frac{c_1}{c_0} \pi_{(0,1,0)}, \pi_{(1,1,1)} = \frac{c_1}{c_0} \pi_{(0,1,1)}
\]
(the proportionality factor is \( \frac{c_1}{c_0} \)),
\[
\pi_{(0,1,0)} = a^2 \pi_{(0,0,0)}, \pi_{(0,1,1)} = a^2 \pi_{(0,0,1)}
\]
(the proportionality factor is \( a^2 \)),
\[
\pi_{(1,1,0)} = a^2 \pi_{(1,0,0)}, \pi_{(1,1,1)} = a^2 \pi_{(1,0,1)}
\]
On the other hand, by Theorem 1.2 we have $\pi(0,0,1) = a\pi(0,0,0)$, $\pi(0,1,1) = a\pi(0,1,0)$, $\pi(1,0,1) = a\pi(1,0,0)$, $\pi(1,1,1) = a\pi(1,1,0)$ (the proportionality factor is also $a$ in all these last cases). Therefore, $\pi$ is a wavy probability distribution (according to Definition 3.1).

The Mallows model through Cayley metric and that through Kendall metric and Potts model on the tree are three interesting examples of wavy probability distributions. It is easy to show these statements using [11]-[13]. (For the first two models (models for ranked data), see also [6, Chapter 6]; for the last model (a model used in statistical physics and other fields), see also [7].)

Below we give the main result of this section.

**Theorem 3.1.** Let $S = \langle r \rangle ^{\ast}$. Let $\pi = (\pi_i)_{i \in S}$ be a wavy probability distribution (on $S$). Consider a Gibbs sampler in a generalized sense with state space $S$ and transition matrix $P = P_1P_2\ldots P_t$ ($t \geq 1$), where (we use the notation from the definition of wavy probability distribution)

$$(P_i)_{i + d_s^{(l,v)} \rightarrow \xi} = \begin{cases} \frac{\pi_{i + d_s^{(l,v)}}}{\sum_{u \in \{0\} \cup D_{v,b_1}} \pi_{i + d_u^{(l,v)}}} & \text{if } \xi = i + d_u^{(l,v)} \text{ for some } u \in \{0\} \cup D_{v,b_1}, \\ 0 & \text{if } \xi \neq i + d_u^{(l,v)}, \forall u \in \{0\} \cup D_{v,b_1}, \end{cases}$$

$\forall l \in \langle t \rangle, \forall v \in \langle u_l \rangle, \forall i \in K_{(v-1)b_1+1}^{(l+1)}, \forall s \in \{0\} \cup D_{v,b_1}, \forall \xi \in S$, setting $d_0^{(l,v)} = 0, \forall l \in \langle t \rangle, \forall v \in \langle u_l \rangle$. Then

$$P = e^t \pi$$

(therefore, the chain attains its stationarity at time 1, its stationary probability distribution being, obviously, $\pi$).

**Proof.** We have

$$(P_i)_{i + d_s^{(l,v)} \rightarrow \xi} = \begin{cases} \frac{1}{1 + \sum_{u \in D_{v,b_1}} a_u^{(l,v)}} & \text{if } \xi = i, \\ \frac{\alpha_u^{(l,v)}}{1 + \sum_{u \in D_{v,b_1}} \alpha_u^{(l,v)}} & \text{if } \xi = i + d_u^{(l,v)} \text{ for some } u \in D_{v,b_1}, \\ 0 & \text{if } \xi \neq i + d_u^{(l,v)}, \forall u \in \{0\} \cup D_{v,b_1}, \end{cases}$$

$\forall l \in \langle t \rangle, \forall v \in \langle u_l \rangle, \forall i \in K_{(v-1)b_1+1}^{(l+1)}, \forall s \in \{0\} \cup D_{v,b_1}, \forall \xi \in S$. It follows that $P_i \in G_{\Delta_l, \Delta_{l+1}}, \forall l \in \langle t \rangle$ (for $\Delta_l, l \in \langle t + 1 \rangle$, see the definition of wavy probability distribution). Since $P = P_1P_2\ldots P_t$, by Theorem 1.1, $P$ is a stable matrix. Consequently, $\exists \psi, \psi$ is a probability distribution on $S$, such that

$$P = e^t \psi.$$ 

On the other hand, by Theorem 1.2 we have

$$\pi P = \pi.$$
Finally, we have
\[ \pi = \pi P = \pi e' \psi = \psi, \]
so,
\[ P = e' \pi. \]


Below, in Theorem 3.2 and Remarks 3.3 and 3.4, we give three important applications of Theorem 3.1.

**Theorem 3.2.** Let \( S = \langle r \rangle \). Let \( \pi = (\pi_i)_{i \in S} \) be a wavy probability distribution (on \( S \)). Suppose that
\[ \pi_i = \frac{\nu_i}{Z}, \forall i \in S, \]
where
\[ Z = \sum_{i \in S} \nu_i, \]
\( Z \) is the normalization constant (\( \nu_i \in \mathbb{R}^+, \forall i \in S, \) so, \( Z \in \mathbb{R}^+ \)). Then (we use the notation from the definition of wavy probability distribution and Theorem 3.1)
\[ Z = \nu_1 \prod_{l \in \langle t \rangle} \left( 1 + \sum_{w \in D_1, b_l} \alpha_w^{(l,1)} \right). \]

**Proof.** We know that
\[ \pi_1 = \frac{\nu_1}{Z}. \]
On the other hand, by Theorem 3.1, using the equation \( P = e' \pi \), we can compute \( \pi_1 \) because \( S = K_1^{(1)} \supset K_1^{(2)} \supset \ldots \supset K_1^{(t+1)} = \{1\} \), \( P_l \) is a block diagonal matrix, \( \forall l \in \langle t \rangle - \{1\} \), and \( P_l \in G_{\Delta_l, \Delta_{l+1}}, \forall l \in \langle t \rangle \) (moreover, \( P_l \) is a \( \Delta_l \)-stable matrix on \( \Delta_{l+1}, \forall l \in \langle t \rangle \)); we obtain
\[ \pi_1 = \frac{1}{\prod_{l \in \langle t \rangle} \left( 1 + \sum_{w \in D_1, b_l} \alpha_w^{(l,1)} \right)}. \]
So,
\[ Z = \nu_1 \prod_{l \in \langle t \rangle} \left( 1 + \sum_{w \in D_1, b_l} \alpha_w^{(l,1)} \right). \]

Each of articles [11]-[13] contains a special case of Theorem 3.2 (in Theorem 3.2, \( S = \langle r \rangle \) for simplification ...). Moreover, in [13], by \( Z \) (this computation way using \( Z \) is known, see, e.g., [1, p. 6]) it is computed the expectation for the Potts model on the tree.
Remark 3.3. When the transition probabilities from Theorem 3.1 can be computed, this theorem gives an exact Markovian method to sample from $S$ according to the wavy probability distribution $\pi$. Since $P = P_1 P_2 ... P_t$, this method has $t$ steps and it is fast when the transition probabilities from Theorem 3.1 can be computed fast. See [11]-[13] for examples.

Remark 3.4. The important probabilities $P \left( K_v^{(l)} \right)$, $l \in \{t + 1\}$, $v \in \{u_l\}$, can be computed for wavy probability distributions iteratively. Indeed, this can be done using Theorem 3.1 and Uniqueness Theorem from [10]. See [11]-[13] for examples.

The Gibbs sampler in a generalized sense is good, very good, at least for wavy probability distributions when its transition probabilities from Theorem 3.1 can be computed fast (no problem for the models from [11]-[13]). This is our final conclusion.

References


(Udrea Păun) ROMANIAN ACADEMY ”GHEORGHE MIHOC - CAIUS IACOB” INSTITUTE OF MATHEMATICAL STATISTICS AND APPLIED MATHEMATICS
CALEA 13 SEPTEMBRIE NR. 13, 050711 BUCHAREST, ROMANIA
E-mail address: paun@csn.ro