# Efficient Markov Chain Monte Carlo with Incomplete Multinomial Data

By KWANG WOO AHN and KUNG-SIK CHAN

*Department of Statistics and Actuarial Science, The University of Iowa, Iowa City, IA USA*

kwangwoo-ahn@uiowa.edu kung-sik-chan@uiowa.edu

Summary

We propose a new, block Gibbs sampling scheme for incomplete multinomial data. The new approach facilitates maximal blocking, thereby reducing serial dependency and speeding up the convergence of the Gibbs sampler. We compare the new method with the standard, non-block Gibbs sampler via a numerical example.

*Some key words*: Blocking; Gibbs Sampler; Dirichlet distribution; Epidemiology.

## 1. Introduction

Incomplete multinomial data abound in science. For example, in epidemiology, the strain of a pathogen infecting some trapped rodents may be unknown due to contamination problems of the ensuing blood tests, resulting in incomplete multinomial data; Ahn et al. (2007). Bayesian analysis with incomplete multinomial data may be carried out via Markov chain Monte Carlo. A standard approach is to consider the counts of each category as the latent complete data and impute these counts one by one and iteratively in the Monte Carlo, see Gelman et al. (2003). However, Ahn et al. (2007) noted that for missingness resulted from the presence of partially classified observations known to belong to one of several *disjoint* groups of categories, the posterior distribution is tractable with a Dirichlet prior. Here, we further develop this observation to derive a new block Gibbs sampling scheme for Bayesian analysis with incomplete multinomial data. The novelty of the new approach lies in updating counts of groups of categories instead of updating counts of individual categories, i.e. the new scheme promotes blocking in the Gibbs sampling.

The convergence properties of the Gibbs sampler have been extensively studied, see, e.g. Chan (1993), Liu et al. (1994), Tierney (1994), Rosenthal (1995), Amit (1996), and Roberts & Sahu (2001). Blocking in a Gibbs sampler generally speeds up its convergence to stationarity, and reduces serial dependence. Roberts & Sahu (1997) showed that for a Gaussian target distribution, convergence of a Gibbs sampler becomes faster with fewer blocks. Liu et al. (1994) showed that blocking may lead to a faster convergence of the Gibbs sampler by reducing the spectral norm of the underlying operator. We detail the new block Gibbs sampling scheme for incomplete multinomial data and show that it is theoretically superior than the non-block Gibbs sampling scheme in section 2. We compare the new method and the non-block procedure via a numerical example in section 3.

## 2. The Block Gibbs Sampler

Let $Y = (Y_1, Y_2, \dots, Y_k)$ be the vector of latent counts of $N$ independent and identically distributed observations each of which belongs to one of $k$ mutually exclusive categories, labeled from 1 to $k$, i.e. $Y_i$ is the count of the $i$-th category. Thus, $Y$ follows the multinomial distribution with parameters $(N, \theta)$ where $\theta = (\theta_1, \theta_2, \dots, \theta_k)$ is the vector of cell probabilities. Consider the case that $Y$ is unobservable because some observations are partially observable in that their categories are unknown and that they are only known to belong to some subset of categories. Specifically, let $A_j$ be distinct non-singleton proper subsets of $S = \{1, 2, \dots, k\}$, $j = 1, \dots, m$. The incomplete data consist of $X = (X_1, X_2, \dots, X_k, X_{A_1}, \dots, X_{A_m})$ where $X_i$ is the count of fully classified subjects that belong to category $i$, and $X_{A_j}$ is the count of the group of categories $A_j$, i.e. the count of partially classified subjects whose category belongs to $A_j$. We assume missing at random so that the probability density of $X$ given $\theta$ is given by

$$\pi(x|\theta) = \frac{n!}{x_1! \cdots x_k! x_{A_1}! \cdots x_{A_m}!} \theta_1^{x_1} \cdots \theta_k^{x_k} \Big( \sum_{i \in A_1} \theta_i \Big)^{x_{A_1}} \cdots \Big( \sum_{i \in A_m} \theta_i \Big)^{x_{A_m}},$$

where $n = \sum_{i=1}^k x_i + \sum_{i=1}^m x_{A_i}$ and $\sum_{i=1}^k \theta_i = 1$.

We now consider Bayesian analysis with the prior distribution for $\theta$ being the Dirichlet distribution with parameter $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_k)$, that is,

$$\pi(\theta|\alpha) \propto \prod_{i=1}^k \theta_i^{\alpha_i - 1}.$$

We first consider the special case that the $A$'s are disjoint. Define $A_0 = S - \bigcup_{j=1}^m A_j$. Then it can be verified that the posterior density function is given by

$$\begin{aligned}
\pi(\theta|x) &\propto \theta_1^{\alpha_1 + x_1 - 1} \cdots \theta_k^{\alpha_k + x_k - 1} \Big( \sum_{i \in A_1} \theta_i \Big)^{x_{A_1}} \cdots \Big( \sum_{i \in A_m} \theta_i \Big)^{x_{A_m}} \\
&\propto \prod_{i \in A_0} \Big( \frac{\theta_i}{\sum_{j \in A_0} \theta_j} \Big)^{\alpha_i + x_i - 1} \prod_{i \in A_1} \Big( \frac{\theta_i}{\sum_{j \in A_1} \theta_j} \Big)^{\alpha_i + x_i - 1} \cdots \prod_{i \in A_m} \Big( \frac{\theta_i}{\sum_{j \in A_m} \theta_j} \Big)^{\alpha_i + x_i - 1} \\
&\times \Big( \sum_{i \in A_0} \theta_i \Big)^{\sum_{j \in A_0} (\alpha_j + x_j - 1)} \Big( \sum_{i \in A_1} \theta_i \Big)^{x_{A_1} + \sum_{j \in A_1} (\alpha_j + x_j - 1)} \cdots \Big( \sum_{i \in A_m} \theta_i \Big)^{x_{A_m} + \sum_{j \in A_m} (\alpha_j + x_j - 1)}.
\end{aligned}$$

While the posterior distribution looks complex, it admits a simple representation. First, some notations. Let $A_p = \{p_1, \ldots, p_{n_p}\}$, $p = 0, \ldots, m$. Ahn et al. (2007) showed that the posterior distribution is tractable by re-parameterizing the model using the parameters defined by the group probabilities $U = (U_0, U_1, \ldots, U_m)^T$, where $U_j = \sum_{i \in A_j} \theta_i$, $j = 0, 1, \ldots, m$, and the conditional probabilities of individual cells within each group $V_{A_j} = (\theta_i / U_j, i \in A_j)^T, 0 \leq j \leq m$. Clearly, the sum of the components in $U$ and those of each $V_{A_j}$ are constrained to be 1. It is readily checked that $\theta$ and $(U, V_{A_j}, j = 0, \ldots, m)$ are equivalent parameterization as they bear a one-to-one relationship. Ahn et al. (2007) proved that the posterior distribution of $(U, V_{A_j}, j = 0, 1, \ldots, m)$ enjoys the following properties:

(1) The group probabilities $U$ and the random vectors of conditional probabilities of cells within each group $V_{A_j}, j = 0, \ldots, m$ are jointly independent.

(2) The vector of group probabilities $U = (\sum_{i \in A_0} \theta_i, \sum_{i \in A_1} \theta_i, \ldots, \sum_{i \in A_m} \theta_i)$ has the Dirichlet distribution with parameter vector $(\sum_{j \in A_0}(\alpha_j + x_j), x_{A_1} + \sum_{j \in A_1}(\alpha_j + x_j), \ldots, x_{A_m} + \sum_{j \in A_m}(\alpha_j + x_j))$.

(3) For the $p$th group, the conditional probability vector of the cells within the group $V_{A_p} = (\frac{\theta_{p_1}}{\sum_{j \in A_p} \theta_j}, \ldots, \frac{\theta_{p_{n_p}}}{\sum_{j \in A_p} \theta_j})$ follows the Dirichlet distribution with parameter vector $(\alpha_{p_1} + x_{p_1}, \ldots, \alpha_{p_{n_p}} + x_{p_{n_p}})$ for $p = 0, \ldots, m$.

Based on the preceding characterizations of the posterior distribution, Ahn et al. (2007) obtain the exact posterior mean and variance of $\theta$. Moreover, independent random realizations can be readily drawn from the posterior distribution.

Next, we study the general case that the $A$'s may overlap. The posterior distribution is no longer tractable although inference may be drawn via Gibbs sampling. Define $Z_{i|A_p}$ be the count of partially observed subjects included in the count of $A_p$ whose category is $i$ where $i = 1, \ldots, k$. Then, $x_{A_p} = \sum_{i=1}^{k} Z_{i|A_p}$. A popular implementation of Gibbs sampling (Gelman et al. 2003, p.533–539) runs as follows with the $j$th iterates obtained by the formulas: (the notation $\sim$ read as "is distributed as")

$$Y_i^{(j)} = x_i + \sum_{p=1}^{m} Z_{i|A_p},$$

$$\theta^{(j+1)} \sim \text{Dirichlet}(\alpha_1 + y_1^{(j)}, \ldots, \alpha_k + y_k^{(j)}),$$

where $(Z_{1|A_p}, \ldots, Z_{k|A_p})$ has a multinomial distribution with sample size $x_{A_p}$ and cell probabilities $(\theta_1 I_{A_p}(1), \ldots, \theta_k I_{A_p}(k))/\sum_{i\in A_p}\theta_i$ where $I_{A_p}(j) = 1$ if $j \in A_p$ and 0 otherwise. Note that this scheme updates the count of each category that falls in some $A$'s.

The tractability of the posterior distribution for the special case of disjoint $A$'s inspires the following new approach. First, observe that the standard approach makes use of the decomposition that each $A$ is a union of all singleton subsets of $A$, which then leads to a Gibbs sampling scheme that updates counts of individual categories belonging to some $A$'s; the collection of such categories constitute the finest building blocks making up the $A$'s. However, a set of coarser building blocks facilitates blocking in the Gibbs sampler, as explained below. Specifically, let $\{P_1, P_2, \ldots, P_d\}$ be a collection of mutually disjoint subsets of $\{1, 2, \ldots, k\}$ such that each $A$ is the union of some $P$'s and the union of $A$'s equal that of the $P$'s. The new Gibbs sampling scheme requires updating the counts of the $P$'s in each iteration. If the $P$'s are large, $d$ will be small, leading to more blocking in the Gibbs sampling. The coarsest such building blocks of the $A$'s can be described as follows. Recall that there are $m$ $A$'s, labeled as $A_1$ to $A_m$. Enumerate all non-empty subsets of $\{1, \ldots, m\}$ as $B_\ell$, $\ell = 1, \ldots, 2^m - 1$. It can be shown that the collection of sets

$$\mathcal{P} = \left\{P : P = \bigcap_{j\in B_\ell} A_j - \bigcup_{j\in B_\ell^c} A_j,\ \text{for some } \ell = 1, 2, \ldots, 2^m - 1, P \neq \emptyset\right\}$$

provides the coarsest *disjoint* building blocks of the $A$'s.

Indeed,

$$A_i = \bigcup_{\{B_\ell : i\in B_\ell\}} \Big(\bigcap_{j\in B_\ell} A_j - \bigcup_{j\in B_\ell^c} A_j\Big),$$

where the index of the union ranges over all $B_\ell$ that contains $i$. For the case that $m = 2$ so that there are only $A_1$ and $A_2$, the preceding results follow from the decomposition of

$A_1 \cup A_2$ as the disjoint union of $A_1 \cap A_2$, $A_1 - A_2$ and $A_2 - A_1$. The general case can then be proved by mathematical induction, and hence omitted.

Now, instead of using $Y$ as the complete data, we shall employ a different set of complete data defined by $W = \{X_1, \ldots, X_q, Z_P, P \in \mathcal{P}\}$ where $Z_P$ is the unobserved count of those partially classified data whose category belongs to $P$. Thanks to the disjointness of the $P$'s, it follows from earlier discussion in this section that the conditional distribution of $\theta$ given $W$ and $X$ is tractable; clearly this conditional distribution depends on $W$ only. The update of $W$ given $X$ and $\theta$ can be proceeded as follows. Corresponding to the count of a typical $A$, say $X_A$, it can be decomposed as the sum $X_A = \sum_{P \subset A, P \in \mathcal{P}} Z_{P|A}$ where $Z_{P|A}$ is the unobserved count of those observations, enumerated in the count $X_A$, whose category belongs to $P$. It is readily seen that given $X$ and $\theta$, $(Z_{P|A}, P \in \mathcal{P})$ has a multinomial distribution with sample size $X_A$ and cell probabilities $\sum_{i \in P \cap A} \theta_i / \sum_{i \in A} \theta_i$, $P \in \mathcal{P}$; these multinomial distributions of the $A$'s are jointly independent, given $X$ and $\theta$. Then compute $Z_P = \sum_{i=1}^m Z_{P|A}$, where the sum need only be taken over $P \subseteq A$. Thus, Gibbs sampling can be readily carried out by cycling the two steps of (i) updating $\theta$ given $W$ and (ii) updating $W$ given $\theta$ and $X$.

We now apply Theorem 5.1 of Liu et al. (1994) to show that the block Gibbs sampler is more efficient than its non-block counterpart. Below, the notation $[A|B]$ denotes the conditional distribution of $A$ given $B$ for any two random vectors $A$ and $B$. The block Gibbs sampler aims to draw Markov-chain realizations from $[\theta, F|X]$ where $F = (Z_{P|A_j}, P \in \mathcal{P}, j = 1, \ldots, m)$. It does so by drawing from $[\theta|F, X]$ and $[F|\theta, X]$ iteratively. On the other hand the non-block version attempts to draw Markov-chain realizations from $[\theta, G|X]$ where $G = (Z_{i|A_j}, i = 1, \ldots, k, j = 1, \ldots, m)$, by drawing from $[\theta|G, X]$ and $[G|\theta, X]$ iteratively. Because $Z_{P|A_j} = \sum_{i \in P} Z_{i|A_j}$ for $j = 1, \ldots, m$, we can reparameterize $G$ as $(F, H)$ for some $H$

| | Elapsed Time (sec.) | User CPU Time (sec.) |
|---|---|---|
| Block Gibbs Sampler | 30.53 | 30.19 |
| Non-Block Gibbs Sampler | 29.31 | 28.98 |

TABLE 1. Computing times of the block Gibbs sampler and the non-block version, with a computer running a *Mobile AMD Sempron(tm), Processor 3500+, 1.79 GHz, with 1.87 GB of RAM*

$\subset \{Z_{i|A_j}; i = 1, \ldots, k, j = 1, \ldots, m\}$. Thus, the non-block version draws dependent sample from $[\theta|(F,H), X]$ and $[(F,H)|\theta, X]$ iteratively. So it follows from Theorem 5.1 of Liu et al. (1994) that the operator norm of the underlying Markov chain for the block Gibbs sampler is less than or equal to that of the non-block version. Liu et al. (1994) showed that the operator norm equals the maximal correlation of consecutive iterates. Hence, the new block Gibbs sampler has a faster convergence rate and less autocorrelation.

## 3. Example

We study an artificial example motivated by epidemiology in which, based on blood tests, a number of trapped rodents were classified into 9 categories: Category 1 through 8 indicate that an examined rodent was infected by the bartonella bacterium of variants 1 through 8, respectively, and category 9 denotes the state of no detected bartonella. Suppose the counts are given by $x = (x_1, \ldots, x_9, x_{A_1}, x_{A_2}, x_{A_3}) = (20, 17, 15, 11, 8, 5, 10, 4, 655, 34, 21, 18)$, where $(x_1, \ldots, x_9)$ are cell counts of the completely classified observations, but $x_{A_1}$, $x_{A_2}$, and $x_{A_3}$ are counts of partially classified data, and where $A_1 = \{2,3\}$, $A_2 = \{4,5,6\}$, and $A_3 = \{5,6,7,8\}$. It is readily seen that the $A$'s are disjoint unions of some of the following sets: $P_1 = \{2,3\}, P_2 = \{4\}, P_3 = \{5,6\}$ and $P_4 = \{7,8\}$. Note that $A_1 =$

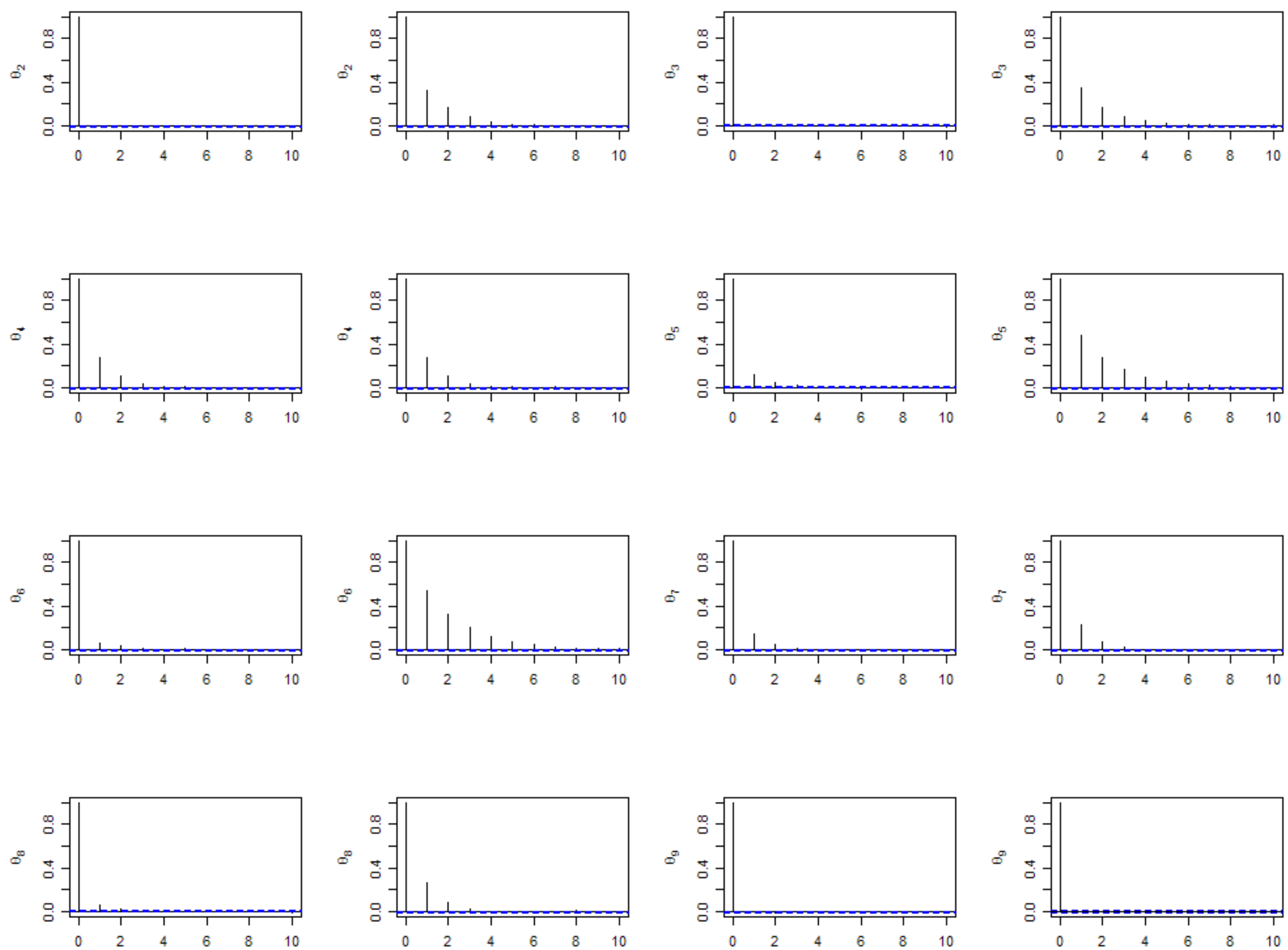

Figure 1. The sample auto-correlation function of the Gibbs iterates of $\theta_i, 2 \leq i \leq 9$. Plots in columns 1 and 3 are from the block Gibbs sampler while plots in columns 2 and 4 are from the non-block Gibbs sampler

$P_1, A_2 = P_2 \cup P_3$ and $A_3 = P_3 \cup P_4$. So the updating of the $Z_P$'s are straightforward: $Z_{\{2,3\}} \equiv x_{\{2,3\}}$, $Z_4 = Z_{4|\{4,5,6\}} = \text{Bin}(x_{\{4,5,6\}}, \theta_4/(\theta_4+\theta_5+\theta_6))$, $Z_{\{5,6\}|\{4,5,6\}} = x_{\{4,5,6\}} - Z_4$, $Z_{\{5,6\}|\{5,6,7,8\}} = \text{Bin}(x_{\{5,6,7,8\}}, (\theta_5+\theta_6)/\sum_{i=5}^{8}\theta_i)$. Hence, $Z_{\{5,6\}} = x_{\{4,5,6\}} - Z_4 + Z_{\{5,6\}|\{5,6,7,8\}}$ and $Z_{\{7,8\}} = x_{\{5,6,7,8\}} - Z_{\{5,6\}|\{5,6,7,8\}}$. Below, we write $Z$ for the vector consisting of $Z_P, P \in \mathcal{P}$.

Thus, the complete-data posterior density function is given by

$$\begin{aligned}
\pi(\theta|x, Z) \propto\ & {\theta_1}^{20}{\theta_2}^{17}{\theta_3}^{15}{\theta_4}^{11+Z_{4|\{4,5,6\}}}{\theta_5}^{8}{\theta_6}^{5}{\theta_7}^{10}{\theta_8}^{4}{\theta_9}^{655} \\
& \times (\theta_2+\theta_3)^{34}(\theta_5+\theta_6)^{21-Z_{4|\{4,5,6\}}+Z_{\{5,6\}|\{5,6,7,8\}}}(\theta_7+\theta_8)^{18-Z_{\{5,6\}|\{5,6,7,8\}}} \\
\propto\ & \left(\frac{\theta_2}{\theta_2+\theta_3}\right)^{17}\left(\frac{\theta_3}{\theta_2+\theta_3}\right)^{15}\left(\frac{\theta_5}{\theta_5+\theta_6}\right)^{8}\left(\frac{\theta_6}{\theta_5+\theta_6}\right)^{5}\left(\frac{\theta_7}{\theta_7+\theta_8}\right)^{10}\left(\frac{\theta_8}{\theta_7+\theta_8}\right)^{4} \\
& \times {\theta_1}^{20}(\theta_2+\theta_3)^{66}{\theta_4}^{11+Z_{4|\{4,5,6\}}}(\theta_5+\theta_6)^{34-Z_{4|\{4,5,6\}}+Z_{\{5,6\}|\{5,6,7,8\}}}(\theta_7+\theta_8)^{32-Z_{\{5,6\}|\{5,6,7,8\}}}{\theta_9}^{655}.
\end{aligned}$$

Since the $P$'s are disjoint, the conditional distributions of the $\theta$'s given $x$ and $Z$ can be characterized as follows:

$$\begin{aligned}
\frac{\theta_2}{\theta_2+\theta_3} &\sim \text{Beta}(18, 16), \\
\frac{\theta_5}{\theta_5+\theta_6} &\sim \text{Beta}(9, 6), \\
\frac{\theta_7}{\theta_7+\theta_8} &\sim \text{Beta}(11, 5), \\
(\theta_1, \theta_2+\theta_3, \theta_4, \theta_5+\theta_6, \theta_7+\theta_8, \theta_9) &\sim \text{Dirichlet}(21, 68, 12+Z_{4|\{4,5,6\}}, \\
&\qquad 36 - Z_{4|\{4,5,6\}} + Z_{\{5,6\}|\{5,6,7,8\}}, 34 - Z_{\{5,6\}|\{5,6,7,8\}}, 656).
\end{aligned}$$

Thus, the block Gibbs sampling algorithm iterates are given by:

$$\begin{aligned}
{Z_{4|\{4,5,6\}}}^{(j)} &\sim \text{Bin}(21, {\theta_4}^{(j)}/\sum_{i=4}^{6}{\theta_i}^{(j)}), \\
{Z_{\{5,6\}|\{5,6,7,8\}}}^{(j)} &\sim \text{Bin}(18, ({\theta_5}^{(j)}+{\theta_6}^{(j)})/\sum_{i=5}^{8}{\theta_i}^{(j)}),
\end{aligned}$$

| | $\theta_1$ | $\theta_2$ | $\theta_3$ | $\theta_4$ | $\theta_5$ | $\theta_6$ | $\theta_7$ | $\theta_8$ | $\theta_9$ |
|---|---|---|---|---|---|---|---|---|---|
| Block Gibbs Sampling | 1.000 | 1.000 | 1.000 | 0.528 | 0.708 | 0.806 | 0.709 | 0.851 | 1.000 |
| Non-Block Gibbs Sampling | 1.000 | 0.440 | 0.413 | 0.525 | 0.305 | 0.268 | 0.609 | 0.571 | 0.964 |

TABLE 2. Comparison of the efficiency for the block Gibbs sampler and the non-block Gibbs sampler.

$$U_1{}^{(j+1)} = \left(\frac{\theta_2}{\theta_2+\theta_3}\right)^{(j+1)} \sim \text{Beta}(18, 16),$$

$$U_2{}^{(j+1)} = \left(\frac{\theta_5}{\theta_5+\theta_6}\right)^{(j+1)} \sim \text{Beta}(9, 6),$$

$$U_3{}^{(j+1)} = \left(\frac{\theta_7}{\theta_7+\theta_8}\right)^{(j+1)} \sim \text{Beta}(11, 5),$$

$$(V_1{}^{(j+1)}, \ldots, V_6{}^{(j+1)}) = (\theta_1{}^{(j+1)}, (\theta_2+\theta_3)^{(j+1)}, \theta_4{}^{(j+1)}, (\theta_5+\theta_6)^{(j+1)}, (\theta_7+\theta_8)^{(j+1)}, \theta_9{}^{(j+1)})$$

$$\sim \text{Dirichlet}(21, 68, 12 + Z_{4|\{4,5,6\}}{}^{(j)}, 36 - Z_{4|\{4,5,6\}}{}^{(j)} + Z_{\{5,6\}|\{5,6,7,8\}}{}^{(j)},$$

$$34 - Z_{\{5,6\}|\{5,6,7,8\}}{}^{(j)}, 656),$$

$$\theta_1{}^{(j+1)} = V_1{}^{(j+1)}, \qquad \theta_2{}^{(j+1)} = U_1{}^{(j+1)}V_2{}^{(j+1)}, \qquad \theta_3{}^{(j+1)} = (1-U_1)^{(j+1)}V_2{}^{(j+1)},$$

$$\theta_4{}^{(j+1)} = V_3{}^{(j+1)}, \qquad \theta_5{}^{(j+1)} = U_2{}^{(j+1)}V_4{}^{(j+1)}, \qquad \theta_6{}^{(j+1)} = (1-U_2)^{(j+1)}V_4{}^{(j+1)},$$

$$\theta_7{}^{(j+1)} = U_3{}^{(j+1)}V_5{}^{(j+1)}, \qquad \theta_8{}^{(j+1)} = (1-U_3)^{(j+1)}V_5{}^{(j+1)}, \qquad \theta_9{}^{(j+1)} = V_6{}^{(j+1)}.$$

In particular, it can be seen that the iterates of $\theta_1, \theta_2, \theta_3$ and $\theta_9$ are independent processes. The block Gibbs sampling scheme essentially generates dependent realizations from the conditional distribution of $(Z_{\{2,3\}}, Z_{\{5,6\}}, Z_{\{7,8\}}, \theta)$ given $X = x$ where $Z_{\{2,3\}} = x_{A_1}$ is constant. In contrary, for the approach that uses the counts of each category as the complete data, the Gibbs sampler can be viewed as generating dependent data from the conditional distribution of $(Z_2, Z_3, Z_5, Z_6, Z_7, Z_8, \theta)$ given $X = x$; below, this scheme will be referred to as the non-block Gibbs sampler. It is then clear that the new approach facilitates maximal blocking.

We compare the block Gibbs sampler with the non-block version by drawing 110000 samples with the initial value of $\theta$ set to be $(1/9, \ldots, 1/9)$. The first 10000 transient samples were deleted from subsequent analysis. The posterior mean (unreported) of $\theta$ for both methods were found to be very close. Table 1 displays the CPU times needed for the simulation, which shows that the two methods are roughly equally fast. Also the Auto-Correlation Function (ACF) plots (Figure 1) show that the iterates from the block Gibbs sampler are much less serially correlated than the non-block version. As expected from theory, for the block Gibbs sampler, the iterates of $\theta_2$ and $\theta_3$ are serially uncorrelated although they are serially correlated for the unblocked version. The ACFs of the iterates of $\theta_1$ for both methods are omitted from Figure 1 as they should and indeed found to be consistent with the pattern induced by white noise. Table 2 compares the efficiency of both algorithms, with the efficiency of a sampling scheme w.r.t. a scaler parameter defined as $1/(1 + 2\sum_{i=1}^{\infty} \rho_i)$, where $\rho_i$ are the lag-$i$ auto-correlations of the samples of the parameter of interest. In table 2, the efficiency is estimated by replacing $\rho_i$ by their sample analogues; non-significant correlations are replaced by zero in the formula. Note that the efficiency is close to 1 for an almost independent sampling scheme, but it is much less than 1 for a strongly, positively dependent sampling scheme. Table 2 shows that the block Gibbs sampler is clearly much more efficient than the non-block version. The authors gratefully acknowledge the National Science Foundation, U.S.A. for partial support.

## References


Ahn, K. W., Chan, K.-S., Bai, Y. & Kosoy, M. (2007). A note on Bayesian inference with incomplete multinomial data with applications for assessing the spatio-temporal variation in pathogen-variant diversity. Tech. rep., The University of Iowa, http://www.stat.uiowa.edu/techrep/tr381.pdf.

Amit, Y. (1996). Convergence properties of the Gibbs sampler for perturbations of Gaussians. *Ann. Statist.* **24,** 122–140.

Chan, K. S. (1993). Asymptotic behavior of the Gibbs sampler. *J. Amer. Statist. Assoc.* **88,** 320–326.

Gelman, A., Carlin, J. B., Stern, H. S. & Rubin, D. B. (2003). *Bayesian Data Analysis*. Boca Roton, FL: Chapman and Hall, 2nd ed.

Liu, J. S., Wong, W. H. & Kong, A. (1994). Covariance structure of the Gibbs sampler with applications to the comparisons of estimators and augmentation schemes. *Biometrika* **81,** 27–40.

Roberts, G. O. & Sahu, S. K. (1997). Updating schemes, correlation structure, blocking and parameterization for the Gibbs sampler. *J. Roy. Statist. Soc. Ser. B* **59,** 291–317.

Roberts, G. O. & Sahu, S. K. (2001). Approximate predetermined convergence properties of the Gibbs sampler. *J. Comput. Graph. Statist.* **10,** 216–229.

Rosenthal, J. S. (1995). Minorization conditions and convergence rates for Markov Chain Monte Carlo. *J. Amer. Statist. Assoc.* **90,** 558–566.

Tierney, L. (1994). Markov Chains for exploring posterior distributions (with discussion). *Ann. Statist.* **22,** 1701–1762.