Note on the Sampling Distribution for the Metropolis-Hastings Algorithm

Geweke, John

Tanizaki, Hisashi

(Citation)
Communications in Statistics: Theory and Methods,32(4):775-789

(Issue Date)
2003

(Resource Type)
journal article

(Version)
Accepted Manuscript

(URL)
https://hdl.handle.net/20.500.14094/90000129

Note on the Sampling Distribution for the Metropolis-Hastings Algorithm

John Geweke
Department of Economics, University of Iowa
Iowa City, IA 52242-1000, USA
john-geweke@uiowa.edu

and

Hisashi Tanizaki
Faculty of Economics, Kobe University
Kobe 657-8501, Japan
tanizaki@kobe-u.ac.jp

May 15, 2002

Abstract

The Metropolis-Hastings algorithm has been utilized in recent development of Bayes methods, which algorithm generates random draws from any target distribution. As well known, the Metropolis-Hastings algorithm requires the other auxiliary distribution, called the sampling distribution or the proposal distribution. In this paper, we introduce three sampling distributions, i.e., the independence chain, the random walk chain and the Taylored chain suggested by Geweke and Tanizaki (1999, 2001), and investigate which is the best sampling distribution.

1 Introduction

Use of Monte Carlo methods increases, especially in a Bayesian framework, as a personal computer progresses. In recent Bayes methods, random draws are generated from the posterior distribution. Consider the following situation. We want to generate random draws from a distribution, which is called the target distribution, but it is hard to sample directly from the target distribution. Suppose that it is easy to generate random draws from another appropriately chosen distribution, which is called the sampling distribution. Under the above setup, there are three representative random number generation methods which utilize the random draws generated from the sampling distribution, i.e., importance resampling, rejection sampling and the Metropolis-Hastings (MH) algorithm. In this paper, we focus on the MH algorithm and we see how the MH algorithm is influenced by the sampling distribution. There is a great amount of literature on choice of the sampling density. See, for example, Smith and Roberts (1993), Bernardo and Smith (1994), O'Hagan (1994), Tierney (1994), Chib and Greenberg (1995), Geweke (1996), Gamerman (1997), Robert and Casella (1999) and so on.

In this paper, three sampling distributions are introduced. One is known as the independence chain and another is called the random walk chain. In addition to these two sampling distributions, the density discussed in Geweke and Tanizaki (1999, 2001) is also examined, which utilizes the second-order Taylor series expansion of the target distribution, where the target density is approximated by the normal distribution, the exponential distribution or the uniform distribution. Hereafter, we call this sampling density the Taylored chain. Chib, Greenberg and Winkelmann (1998) also considered the Taylored chain, but their Taylored chain is based only on the normal distribution. The Taylored chain discussed in this paper utilizes the exponential and uniform distributions as well as the normal distribution, which is a substantial extension of Chib, Greenberg and Winkelmann (1998).

The sampling densities based on the independence chain and the random walk chain depend on hyper-parameters such as mean and variance. Therefore, choice of the hyper-parameters is one of the crucial problems for both the independence chain and the random walk chain. On the contrary, the sampling density based on the Taylored chain does not include any hyper-parameter. In addition, some simulation studies in Section 4 indicate that the Taylored chain performs quite good and therefore we see that the Taylored chain is practically more useful than the independence chain and the random walk chain.

2 A Review of the MH Algorithm

Let $f(x)$ be the target density function and $f_*(x|x_{i-1})$ be the sampling density function. The MH algorithm is one of the sampling methods to generate random draws from any target density $f(x)$, utilizing the sampling density $f_*(x|x_{i-1})$.

Let us define the acceptance probability by:

$$\omega(x_{i-1}, x^*) = \min\Big(\frac{q(x^*|x_{i-1})}{q(x_{i-1}|x^*)}, 1\Big) = \min\Big(\frac{f(x^*)/f_*(x^*|x_{i-1})}{f(x_{i-1})/f_*(x_{i-1}|x^*)}, 1\Big),$$

where $q(\cdot|\cdot)$ is defined as $q(x|z) = f(x)/f_*(x|z)$. By the MH algorithm, random draws from $f(x)$ are generated in the following way: (i) take the initial value of x as x_{-M}, (ii) given the $(i-1)$-th random draw, x_{i-1}, generate a random draw of x, i.e., x^*, from $f_*(x|x_{i-1})$ and compute the acceptance probability $\omega(x_{i-1}, x^*)$, (iii) set $x_i = x^*$ with probability $\omega(x_{i-1}, x^*)$ and $x_i = x_{i-1}$ otherwise, and (iv) repeat Steps (ii) and (iii) for $i = -M+1, -M+2, \cdots, N$. Thus, we can obtain a series of random draws x_{-M+1}, x_{-M+2}, $\cdots$, x_N. For sufficiently large M, x_1 is taken as a random draw from $f(x)$. Then, x_2 is also a random draw from $f(x)$ because x_2 is generated based on x_1. That is, x_i, $i = 1, 2, \cdots, N$, are regarded as the random draws generated from $f(x)$. There, the first M random draws, x_{-M}, x_{-M+1}, $\cdots$, x_0, are discarded, taking into account convergence of the Markov chain. M is sometimes called the burn-in period. $M = 1000$ is taken in the simulation studies of Section 4. See, for example, Geweke (1992) and Mengersen, Robert and Guihenneuc-Jouyaux (1999) for the convergence diagnostics.

The requirement for uniform convergence of the Markov chain is that the chain should be irreducible and aperiodic. See, for example, Roberts and Smith (1993). Let $C_i(x_0)$ be the set of possible values of x_i from starting point x_0. If there exist two possible starting values, say x^* and x^{**}, such that $C_i(x^*) \cap C_i(x^{**}) = \emptyset$ (i.e., empty set) for all i, then the same limiting distribution cannot be reached from both starting points. A Markov chain is said to be irreducible if there exists an i such that $P(x_i \in C|x_0) > 0$ for any starting point x_0 and any set C such that $\int_C f(x)\mathrm{d}x > 0$. The irreducible condition ensures that the chain can reach all possible x values from any starting point. Moreover, as another case in which convergence may fail, if there are two disjoint set C^1 and C^2 such that $x_{i-1} \in C^1$ implies $x_i \in C^2$ and $x_{i-1} \in C^2$ implies $x_i \in C^1$, then the chain oscillates between C^1 and C^2 and we again have $C_i(x^*) \cap C_i(x^{**}) = \emptyset$ for all i when $x^* \in C^1$ and $x^{**} \in C^2$. Accordingly, we cannot have the same limiting distribution in this case, either. It is called aperiodic if the chain does not oscillate between two sets C^1

and C^2 or cycle around a partition $C^1, C^2, \cdots, C^r$ of r disjoint sets for $r > 2$. See O'Hagan (1994) for the discussion above.

As in the case of rejection sampling and importance resampling, note that $f(x)$ may be a kernel of the target density, or equivalently, $f(x)$ may be proportional to the target density. The same algorithm as Steps (i) – (iv) can be applied to the case where $f(x)$ is proportional to the target density.

Precision of the random draws based on the MH algorithm depends on the sampling density $f_*(\cdot|x_{i-1})$. As Chib and Greenberg (1995) pointed out, $f_*(x|x_{i-1})$ should be chosen so that the chain travels over the support of $f(x)$, which implies that the sampling density $f_*(x)$ should not have too large variance and too small variance, compared with the target density $f(x)$. See, for example, Smith and Roberts (1993), Bernardo and Smith (1994), O'Hagan (1994), Tierney (1994), Geweke (1996), Gamerman (1997), Robert and Casella (1999) and so on for the MH algorithm.

Based on Steps (i) – (iv), under some conditions the basic result of the MH algorithm is as follows:

$$\frac{1}{N}\sum_{i=1}^{N} g(x_i) \longrightarrow \mathrm{E}(g(x)) = \int g(x) f(x)\mathrm{d}x, \qquad \text{as } N \longrightarrow \infty,$$

where $g(\cdot)$ is a function, which is representatively taken as $g(x) = x$ for mean and $g(x) = (x - \overline{x})^2$ for variance. $\overline{x}$ denotes $\overline{x} = (1/N)\sum_{i=1}^{N} x_i$. It is easily shown that $(1/N)\sum_{i=1}^{N} g(x_i)$ is a consistent estimate of $\mathrm{E}(g(x))$, even though $x_1, x_2, \cdots, x_N$ are mutually correlated. Remember that clearly we have $\mathrm{Cov}(x_{i-1}, x_i) > 0$ because x_i is generated based on x_{i-1} in Step (iii).

As an alternative random number generation method to avoid the positive correlation, we can perform the case of $N = 1$ as in the above procedures (i) – (iv) N times in parallel, taking different initial values for x_{-M}. In this case, we need to generate $M + 1$ random numbers to obtain one random draw from $f(x)$. That is, N random draws from $f(x)$ are based on $N(1 + M)$ random draws from $f_*(x)$. Thus, we can obtain mutually independently distributed random draws. However, this alternative method is too computer-intensive, compared with the above procedures (i) – (iv). In this paper, Steps (i) – (iv) are utilized to obtain N random draws from $f(x)$.

3 Choice of Sampling Density

The sampling density has to satisfy the following conditions: (i) one can quickly and easily generate random draws from the sampling density and (ii) the sampling density should be distributed with the same range as the target density. Thus, the MH algorithm has the problem of specifying the sampling density, which is the crucial criticism. Several generic choices of the sampling density are discussed by Tierney (1994) and Chib and Greenberg (1995). In this paper we consider Sampling Densities I – III for the sampling density $f_*(x|x_{i-1})$.

3.1 Sampling Density I (Independence Chain)

One possibility of the sampling density is given by: $f_*(x|x_{i-1}) = f_*(x)$, where $f_*(\cdot)$ does not depend on x_{i-1}. This sampling density is called the independence chain. For example, it is possible to take $f_*(x) = N(\mu_*, \sigma_*^2)$. μ_* and σ_*^2 denote the hyper-parameters, which depend on a prior information. Or, when x lies on a certain interval, say (a, b), we can choose the uniform distribution $f_*(x) = 1/(b-a)$.

3.2 Sampling Density II (Random Walk Chain)

We may take the sampling density called the random walk chain, i.e., $f_*(x|x_{i-1}) = f_*(x - x_{i-1})$. Representatively, we can take $f_*(x|x_{i-1}) = N(x_{i-1}, \sigma_*^2)$, where σ_*^2 indicates the hyper-parameter.

3.3 Sampling Density III (Taylored Chain)

The alternative sampling distribution is based on approximation of the log-kernel (see Geweke and Tanizaki (1999, 2001)), which is a substantial extension of the Taylored chain discussed in Chib, Greenberg and Winkelmann (1998). Let $p(x) = \log(f(x))$, where $f(x)$ may denote the kernel which corresponds to the target density. Approximating the log-kernel $p(x)$ around x_{i-1} by the second-order Taylor series expansion, $p(x)$ is represented as:

$$p(x) \approx p(x_{i-1}) + p'(x_{i-1})(x - x_{i-1}) + \frac{1}{2}p''(x_{i-1})(x - x_{i-1})^2, \tag{1}$$

where $p'(\cdot)$ and $p''(\cdot)$ denote the first- and the second-derivatives. Depending on the values of $p'(x)$ and $p''(x)$, we have the following four cases, i.e., Cases 1 – 4. Here, we classify the four cases by (i) $p''(x) < -\epsilon$ in Case 1 or $p''(x) \geq -\epsilon$ in

Cases 2 – 4 and (ii) $p'(x) < 0$ in Case 2, $p'(x) > 0$ in Case 3 or $p'(x) = 0$ in Case 4. Introducing ϵ into the Taylored chain discussed in Geweke and Tanizaki (1999, 2001) is a new proposal in Geweke and Tanizaki (2002). Note that $\epsilon = 0$ is chosen in Geweke and Tanizaki (1999, 2001). To improve precision of random draws, ϵ should be a positive value, say $\epsilon = 0.1$, which will be discussed later in detail.

Case 1: $\boldsymbol{p''(x_{i-1}) < -\epsilon}$: Equation (1) is rewritten by:

$$p(x) \approx p(x_{i-1}) - \frac{1}{2}\Big(\frac{1}{-1/p''(x_{i-1})}\Big)\Big(x - (x_{i-1} - \frac{p'(x_{i-1})}{p''(x_{i-1})})\Big)^2 + r(x_{i-1}),$$

where $r(x_{i-1})$ is an appropriate function of x_{i-1}. Since $p''(x_{i-1})$ is negative, the second term in the right hand side is equivalent to the exponential part of the normal density. Therefore, $f_*(x|x_{i-1})$ is taken as $N(\mu_*, \sigma_*^2)$, where $\mu_* = x_{i-1} - p'(x_{i-1})/p''(x_{i-1})$ and $\sigma_*^2 = -1/p''(x_{i-1})$.

Case 2: $\boldsymbol{p''(x_{i-1}) \geq -\epsilon}$ **and** $\boldsymbol{p'(x_{i-1}) < 0}$: Perform linear approximation of $p(x)$. Let x^+ be the nearest mode with $x^+ < x_{i-1}$. Then, $p(x)$ is approximated by a line passing between x^+ and x_{i-1}, which is written as:

$$p(x) \approx p(x^+) + \frac{p(x^+) - p(x_{i-1})}{x^+ - x_{i-1}}(x - x^+).$$

From the second term in the right hand side, the sampling density is represented as the exponential distribution with $x > x^+ - d$, i.e., $f_*(x|x_{i-1}) = \lambda \exp\big(-\lambda(x - (x^+ - d))\big)$ if $x^+ - d < x$ and $f_*(x|x_{i-1}) = 0$ otherwise, where λ is defined as:

$$\lambda = \left|\frac{p(x^+) - p(x_{i-1})}{x^+ - x_{i-1}}\right|.$$

d is a positive value, which will be discussed later. Thus, a random draw x^* from the sampling density is generated by $x^* = w + (x^+ - d)$, where w represents the exponential random variable with parameter λ.

Case 3: $\boldsymbol{p''(x_{i-1}) \geq -\epsilon}$ **and** $\boldsymbol{p'(x_{i-1}) > 0}$: Similarly, perform linear approximation of $p(x)$ in this case. Let x^{++} be the nearest mode with $x_{i-1} < x^{++}$. Approximation of $p(x)$ is exactly equivalent to that of Case 2. Taking into account $x < x^{++} + d$, the sampling density is written as: $f_*(x|x_{i-1}) = \lambda \exp\big(-\lambda((x^{++} + d) - x)\big)$ if $x < x^{++} + d$ and $f_*(x|x_{i-1}) = 0$ otherwise. Thus, a random draw x^* from the sampling density is generated by $x^* = (x^{++}+d)-w$, where w is distributed as the exponential random variable with parameter λ.

<u>Case 4:</u> $\boldsymbol{p''(x_{i-1}) \geq -\epsilon}$ **and** $\boldsymbol{p'(x_{i-1}) = 0}$: In this case, $p(x)$ is approximated as a uniform distribution at the neighborhood of x_{i-1}. As for the range of the uniform distribution, we utilize the two appropriate values x^+ and x^{++}, which satisfies $x^+ < x < x^{++}$. When we have two modes, x^+ and x^{++} may be taken as the modes. Thus, the sampling density $f_*(x|x_{i-1})$ is obtained by the uniform distribution on the interval between x^+ and x^{++}, i.e., $f_*(x|x_{i-1}) = 1/(x^{++} - x^+)$ if $x^+ < x < x^{++}$ and $f_*(x|x_{i-1}) = 0$ otherwise.

Thus, for approximation of the kernel, all the possible cases are given by Cases 1 – 4 depending on the values of $p'(\cdot)$ and $p''(\cdot)$. Moreover, in the case where x is a vector, applying the procedure above to each element of x, Sampling III is easily extended to multivariate cases. Finally, we discuss about ϵ and d in the following remarks.

Remark 1: ϵ in Cases 1 – 4 should be taken as an appropriate positive number. In the simulation studies of Section 4, we choose $\epsilon = 0.1$. It seems more natural to take $\epsilon = 0$, rather than $\epsilon > 0$. The reason why $\epsilon > 0$ is taken is as follows. Consider the case of $\epsilon = 0$. When $p''(x_{i-1})$ is negative and it is very close to zero, variance σ_*^2 in Case 1 becomes extremely large because of $\sigma_*^2 = -1/p''(x_{i-1})$. In this case, the obtained random draws are too broadly distributed and accordingly they become unrealistic, which implies that we have a lot of outliers. To avoid this situation, ϵ should be positive. It might be appropriate that ϵ should depend on variance of the target density, because ϵ should be small if variance of the target density is large. Thus, in order to reduce a number of outliers, $\epsilon > 0$ is recommended.

Remark 2: For d in Cases 2 and 3, note as follows. As an example, consider the unimodal density in which we have Cases 2 and 3. Let x^+ be the mode. We have Case 2 in the right hand side of x^+ and Case 3 in the left hand side of x^+. In the case of $d = 0$, we have the random draws generated from either Case 2 or 3. In this situation, the generated random draw does not move from one case to another. In the case of $d > 0$, however, the distribution in Case 2 can generate a random draw in Case 3. That is, for positive d, the generated random draw may move from one case to another, which implies that the irreducibility condition of the MH algorithm is guaranteed. In the simulation studies of Section 4, we take $d = 1/\lambda$, which represents the standard error of the exponential distribution with parameter λ.

4 Monte Carlo Results

4.1 Setup of the Simulation Studies

We have discussed the MH algorithm in Sections 2 and 3. In this section, we examine Sampling Densities I – III, which are introduced in Section 3. For Sampling Density I, in this paper $f_*(x|x_{i-1}) = f_*(x) = N(\mu_*, \sigma_*^2)$ is taken, where $\mu_* = -3$, -2, -1, 0, 1, 2, 3 and $\sigma_* = 1$, 2, 3, 4. As for Sampling Density II, we choose $f_*(x|x_{i-1}) = N(x_{i-1}, \sigma_*^2)$ for $\sigma_* = 1$, 2, 3, 4, 5, 6, 7. Sampling Density III utilizes the second-order Taylor series expansion of the target density, which is approximated by the normal distribution, the exponential distribution or the uniform distribution. We compare Sampling Densities I – III and examine which is the best sampling density.

For each sampling density, the first and second moments are estimated based on the N random draws through the MH algorithm. The three sampling densities are compared by two criteria. For one criterion, we examine whether the estimated moments are close to the theoretical values, i.e., whether the estimated moments are biased, which is called the $\overline{g(x)}$ criterion in this paper. For another criterion, we check whether the estimated moments have small standard errors or not, which is called the $\mathrm{Se}(\overline{g(x)})$ criterion in this paper. The notations $\overline{g(x)}$ and $\mathrm{Se}(\overline{g(x)})$ will be discussed subsequently. When we compare two consistent estimates, the estimate with small variance is clearly preferred to that with large one. Consider dividing the N random draws into N_2 partitions, where each partition consists of N_1 random draws. That is, $N_1 \times N_2 = N$ holds. Let $g(x_i)$ be a function of the i-th random draw, x_i, which function is taken as $g(x) = x^k$, $k = 1, 2$, in this paper. Define $\overline{g_i(x)}$ and $\overline{g(x)}$ as follows:

$$\overline{g_i(x)} = \frac{1}{N_1} \sum_{j=1}^{N_1} g(x_{(i-1)N_1+j}),$$

$$\overline{g(x)} = \frac{1}{N_1 \times N_2} \sum_{j=1}^{N_1 \times N_2} g(x_j) = \frac{1}{N_2} \sum_{i=1}^{N_2} \Big(\frac{1}{N_1} \sum_{j=1}^{N_1} g(x_{(i-1)N_1+j}) \Big) = \frac{1}{N_2} \sum_{i=1}^{N_2} \overline{g_i(x)},$$

for $N = N_1 \times N_2$, where $N = 10^7$, $N_1 = 10^4$, $N_2 = 10^3$ and $g(x) = x^k$, $k = 1, 2$, are taken. Thus, $\overline{g_i(x)}$ represents the arithmetic average of $g(x_{(i-1)N_1+j})$, $j = 1, 2, \cdots, N_1$, and $\overline{g(x)}$ denotes the arithmetic average of $g(x_i)$, $i = 1, 2, \cdots, N_1 \times N_2$.

If $\overline{g_i(x)}$ is independent of $\overline{g_j(x)}$ for all $i \neq j$, the standard error of $\overline{g(x)}$ is given by:

$$\mathrm{Se}(\overline{g(x)}) = \sqrt{\frac{1}{N_2}\sum_{i=1}^{N_2}\Big(\overline{g_i(x)} - \overline{g(x)}\Big)^2}.$$

However, since it is clear that $\overline{g_i(x)}$ is not independent of $\overline{g_j(x)}$ for all $i \neq j$, $\mathrm{Se}(\overline{g(x)})$ does not show the exact standard error of $\overline{g(x)}$. But still we can take $\mathrm{Se}(\overline{g(x)})$ as a rough measure of the standard error of the estimated moment. In order to compare precision of the random draws based on the three sampling densities in the $\overline{g(x)}$ and $\mathrm{Se}(\overline{g(x)})$ criteria, we take the following five target distributions, i.e., bimodal, $t(5)$, logistic, LaPlace and Gumbel distributions.

Bimodal Distribution: $f(x) = \frac{1}{2}N(1, 1^2) + \frac{1}{2}N(-1, 0.5^2)$,
where the first and second moments are given by: $\mathrm{E}(X) = 0$ and $\mathrm{E}(X^2) = 1.625$.

***t*(5) Distribution:** $f(x) \propto \Big(1 + \frac{x^2}{k}\Big)^{-(k+1)/2}$,
where $\mathrm{E}(X) = 0$ and $\mathrm{E}(X^2) = k/(k-2)$. $k = 5$ is taken in the simulation studies.

Logistic Distribution: $f(x) = \frac{e^x}{(1 + e^x)^2}$,
where the first and second moments are $\mathrm{E}(X) = 0$ and $\mathrm{E}(X^2) = \pi^2/3$.

LaPlace Distribution: $f(x) = \frac{1}{2}\exp(-|x|)$,
where the first two moments are given by $\mathrm{E}(X) = 0$ and $\mathrm{E}(X^2) = 2$.

Gumbel Distribution: $f(x) = e^{-(x-\alpha)}\exp(-e^{-(x-\alpha)})$,
where mean and variance are $\mathrm{E}(X) = \alpha + \gamma$ and $\mathrm{V}(X) = \pi^2/6$. $\gamma = 0.5772156599$ is known as Euler's constant. Since we take $\alpha = -\gamma$ in this paper, the first and second moments are $\mathrm{E}(X) = 0$ and $\mathrm{E}(X^2) = \pi^2/6$.

4.2 Results and Discussion

It is very easy to generate random draws from the densities shown above without utilizing the MH algorithm. However, in this section, to see whether the sampling densities work well or not, using the MH algorithm we consider generating random draws from the above target distributions.

The results are in Tables 1 – 6. In order to obtain the i-th random draw (i.e., x_i), we take $N(\mu_*, \sigma_*^2)$ for Sampling Density I (Tables 1 – 5), where $\mu_* = -3, -2, -1, 0, 1, 2, 3$ and $\sigma_* =1, 2, 3, 4$ are taken, and $N(x_{i-1}, \sigma_*^2)$ for Sampling Density II (Table 6), where $\sigma_* =1, 2, 3, 4, 5, 6, 7$ is chosen. Sampling Density III (Table 7) does not have the hyper-parameters such as μ_* and σ_*. For each sampling density, the two moments $\mathrm{E}(X^k)$, $k = 1, 2$, are estimated, generating N random draws, where $N = 10^7$ is taken. Note that the estimates of the two moments are given by $(1/N)\sum_{i=1}^{N} x_i^k$, $k = 1, 2$.

The estimated moments should be close to the theoretical values, which are given by $\mathrm{E}(X) = 0$ and $\mathrm{E}(X^2) = 1.625$ in the bimodal distribution, $\mathrm{E}(X) = 0$ and $\mathrm{E}(X^2) = 5/3$ in the $t(5)$ distribution, $\mathrm{E}(X) = 0$ and $\mathrm{E}(X^2) = 3.290$ in the logistic distribution, $\mathrm{E}(X) = 0$ and $\mathrm{E}(X^2) = 2$ in the LaPlace distribution and $\mathrm{E}(X) = 0$ and $\mathrm{E}(X^2) = 1.645$ in the Gumbel distribution. Thus, for all the five target distributions taken in this section, $\mathrm{E}(X) = 0$ is chosen. In each table, the values in the parentheses indicate $\mathrm{Se}(\overline{g(x)})$, where $g(x) = x$ in the case of $\mathrm{E}(X)$ and $g(x) = x^2$ for $\mathrm{E}(X^2)$. AP denotes the acceptance probability (%) corresponding to each sampling density. Note that AP is obtained by the ratio of the number of the cases where x_i is updated for $i = 1, 2, \cdots, N$ (i.e., $x_i \neq x_{i-1}$) relative to the number of random draws (i.e., $N = 10^7$).

$\overline{g(x)}$ Criterion: According to the $\overline{g(x)}$ criterion, we can observe the following results. For Sampling Density I (Tables 1 – 5), when σ_* is small and μ_* is far from zero, the estimated moments are very different from the true values. Since we have $\mathrm{E}(X) = 0$, $\mathrm{E}(X^2)$ is equivalent to variance of X. When σ_* is larger than $\sqrt{\mathrm{E}(X^2)}$, the first and second moments are very close to the true values. For example, we take Table 1 (the bimodal distribution). When $\sigma_* = 2, 3, 4$, the estimates of the first moment are -0.001 to 0.002 and those of the second one are from 1.623 to 1.627. Thus, the cases of $\sigma_* = 2, 3, 4$ perform much better than those of $\sigma_* = 1$. In Tables 2 – 5, the similar results are obtained. Therefore, for Sampling Density I we can conclude from Tables 1 – 5 that variance of the sampling density should be larger than that of the target density. As for Sampling Density II (Table 6), the estimated moments are close to the true moments for all $\sigma_* =1, 2, 3, 4, 5, 6, 7$. For Sampling Density III (Table 6), all the estimates of the two moments are very close to the true values. Thus, by the $\overline{g(x)}$ criterion, Sampling Densities II and III are quite good. Moreover, Sampling Density I also performs good, provided that

σ_*^2 is larger than variance of the target density and μ_* is close to mean of the target density.

Se($\overline{g(x)}$) Criterion: Next, we focus on the Se($\overline{g(x)}$) criterion. In Tables 1 – 7, the values in the parentheses indicate the corresponding standard errors. In Table 1, the cases of $\mu_* = -1$, 0, 1 and $\sigma_* = 2$, 3 show a quite good performance, and especially the case of $(\mu_*, \sigma_*) = (0, 2)$ gives us the smallest standard errors of the estimated moments. In Table 2, the cases of $\mu_* = -1$, 0, 1 and $\sigma_* = 3$, 4 are acceptable, and especially the case of $(\mu_*, \sigma_*) = (0, 3)$ show the best performance. In Table 3, the cases of $\mu_* = -1$, 0, 1 and $\sigma_* = 3$ are quite good, and especially the case of $(\mu_*, \sigma_*) = (0, 3)$ indicate the best performance. In Table 4, the cases of $\mu_* = -1$, 0, 1 and $\sigma_* = 3$ give us good results, and especially the case of $(\mu_*, \sigma_*) = (0, 3)$ is better than any other cases. In Table 5, the cases of $\mu_* = -2$, -1, 0, 1, 2 and $\sigma_* = 3$, 4 are quite good, and especially $(\mu_*, \sigma_*) = (0, 3)$ is the best. Thus, for Sampling Density I, the sampling density should have the same mean as the target density and the sampling density has to be distributed more broadly than the target density. However, we find from Tables 1 – 5 that too large variance of the sampling density yields poor estimates of the moments. In Table 6 (Sampling Density II), the smallest standard errors of the estimated moments are given by $\sigma_* = 3$ for the bimodal distribution and $\sigma_* = 4$ for the $t(5)$, logistic, LaPlace and Gumbel distributions. As a result, we can see from the tables that Sampling Density I shows the best performance when μ_* and σ_* are properly chosen, compared with Sampling Densities II and III. For example, the smallest standard errors in Table 1 are 0.020 for $\mathrm{E}(X)$ and 0.025 for $\mathrm{E}(X^2)$, those of Bimodal in Table 6 are given by 0.029 for $\mathrm{E}(X)$ and 0.040 for $\mathrm{E}(X^2)$, and the standard errors in Table 7 are 0.063 for $\mathrm{E}(X)$ and 0.047 for $\mathrm{E}(X^2)$. Thus, Sampling Density I gives us the best results.

AP and Se($\overline{g(x)}$): We compare AP and Se($\overline{g(x)}$). In Table 1, at $(\mu_*, \sigma_*) = (0, 1)$, which is the point that gives us the maximum acceptance probability (i.e., 66.98%), the standard errors of the estimated first and second moments are given by 0.029 and 0.065. The standard errors of the estimated moments at $(\mu_*, \sigma_*) = (0, 1)$ are larger than those at $(\mu_*, \sigma_*) = (0, 2)$, because the standard errors of the estimated moments at $(\mu_*, \sigma_*) = (0, 2)$ are given by 0.020 and 0.025. Therefore, judging from the standard error criterion, the case of $(\mu_*, \sigma_*) = (0, 2)$ is preferred to that of $(\mu_*, \sigma_*) = (0, 1)$. In Table 2, the largest AP is given by $(\mu_*, \sigma_*) = (0, 1)$ while the smallest standard errors are around $(\mu_*, \sigma_*) = (0, 3)$. Moreover, in Table

Table 1: Bimodal Distribution (Sampling Density I)

μ_* \ σ_*		1	2	3	4
−3	E(X)	−0.167 (0.798)	0.001 (0.055)	0.000 (0.032)	−0.001 (0.031)
	E(X^2)	1.289 (0.832)	1.625 (0.084)	1.626 (0.045)	1.623 (0.045)
	AP	4.18	19.65	28.76	27.52
−2	E(X)	0.006 (0.589)	0.000 (0.032)	0.000 (0.026)	−0.001 (0.028)
	E(X^2)	1.626 (1.260)	1.625 (0.044)	1.625 (0.036)	1.626 (0.042)
	AP	20.21	37.44	37.87	31.90
−1	E(X)	0.002 (0.143)	0.000 (0.020)	−0.001 (0.023)	−0.001 (0.027)
	E(X^2)	1.634 (0.388)	1.627 (0.029)	1.626 (0.034)	1.626 (0.039)
	AP	54.09	56.89	44.14	34.69
0	E(X)	0.001 (0.029)	0.000 (0.020)	0.000 (0.024)	0.000 (0.028)
	E(X^2)	1.626 (0.065)	1.626 (0.025)	1.625 (0.031)	1.624 (0.040)
	AP	66.98	62.91	45.81	35.40
1	E(X)	0.002 (0.048)	0.000 (0.024)	0.000 (0.026)	0.000 (0.029)
	E(X^2)	1.627 (0.036)	1.624 (0.028)	1.625 (0.034)	1.626 (0.039)
	AP	53.74	53.47	42.48	33.93
2	E(X)	0.012 (0.202)	0.002 (0.036)	0.001 (0.031)	0.001 (0.033)
	E(X^2)	1.621 (0.151)	1.627 (0.035)	1.625 (0.038)	1.627 (0.041)
	AP	24.55	36.25	35.37	30.56
3	E(X)	0.030 (0.809)	0.001 (0.056)	0.001 (0.038)	0.000 (0.035)
	E(X^2)	1.618 (0.499)	1.624 (0.054)	1.626 (0.044)	1.626 (0.046)
	AP	8.09	20.30	26.47	25.90

Table 2: $t(5)$ Distribution (Sampling Density I)

μ_* \ σ_*		1	2	3	4
−3	E(X)	−0.122 (0.608)	−0.003 (0.063)	−0.001 (0.028)	−0.002 (0.026)
	E(X^2)	1.274 (0.587)	1.643 (0.316)	1.661 (0.093)	1.665 (0.088)
	AP	6.42	20.13	28.09	27.26
−2	E(X)	−0.008 (0.438)	−0.001 (0.035)	−0.001 (0.023)	0.000 (0.024)
	E(X^2)	1.503 (0.994)	1.643 (0.203)	1.662 (0.107)	1.668 (0.081)
	AP	19.33	37.83	37.34	31.93
−1	E(X)	−0.016 (0.149)	−0.001 (0.027)	0.000 (0.019)	−0.002 (0.022)
	E(X^2)	1.503 (0.524)	1.646 (0.186)	1.662 (0.079)	1.667 (0.082)
	AP	50.13	57.07	44.37	35.13
0	E(X)	−0.003 (0.157)	0.001 (0.026)	0.000 (0.019)	0.000 (0.022)
	E(X^2)	1.541 (0.747)	1.650 (0.213)	1.661 (0.070)	1.666 (0.075)
	AP	92.80	65.95	47.02	36.28
1	E(X)	0.014 (0.195)	0.001 (0.027)	0.000 (0.019)	0.000 (0.022)
	E(X^2)	1.510 (0.699)	1.647 (0.205)	1.658 (0.078)	1.667 (0.076)
	AP	50.08	57.04	44.38	35.15
2	E(X)	0.060 (0.313)	0.003 (0.035)	0.001 (0.023)	0.001 (0.024)
	E(X^2)	1.399 (0.678)	1.642 (0.185)	1.660 (0.088)	1.667 (0.079)
	AP	19.69	37.85	37.36	31.98
3	E(X)	0.138 (0.597)	0.005 (0.052)	0.001 (0.029)	0.000 (0.026)
	E(X^2)	1.276 (0.581)	1.628 (0.219)	1.655 (0.090)	1.663 (0.088)
	AP	6.47	20.11	28.07	27.27

Table 3: Logistic Distribution (Sampling Density I)

$\mu_* \backslash \sigma_*$		1	2	3	4
−3	E(X)	−0.369 (0.824)	0.002 (0.180)	−0.001 (0.037)	−0.002 (0.033)
	E(X^2)	2.337 (0.580)	3.308 (0.978)	3.292 (0.111)	3.296 (0.107)
	AP	14.67	25.94	37.51	38.29
−2	E(X)	−0.147 (0.669)	−0.002 (0.070)	−0.002 (0.029)	−0.001 (0.028)
	E(X^2)	2.651 (1.183)	3.279 (0.371)	3.293 (0.091)	3.295 (0.101)
	AP	29.77	44.99	50.90	45.13
−1	E(X)	−0.063 (0.431)	−0.001 (0.038)	−0.001 (0.024)	−0.002 (0.025)
	E(X^2)	2.861 (1.505)	3.283 (0.234)	3.294 (0.083)	3.293 (0.096)
	AP	53.80	70.45	61.56	49.86
0	E(X)	−0.001 (0.338)	0.001 (0.031)	0.000 (0.022)	−0.001 (0.027)
	E(X^2)	2.908 (1.463)	3.291 (0.237)	3.289 (0.076)	3.289 (0.093)
	AP	70.65	88.26	65.65	51.52
1	E(X)	0.060 (0.450)	0.002 (0.043)	0.000 (0.023)	0.000 (0.026)
	E(X^2)	2.841 (1.424)	3.292 (0.293)	3.287 (0.079)	3.290 (0.096)
	AP	53.89	70.41	61.54	49.85
2	E(X)	0.218 (0.554)	0.004 (0.068)	0.001 (0.030)	0.001 (0.028)
	E(X^2)	2.552 (0.985)	3.273 (0.339)	3.286 (0.091)	3.286 (0.102)
	AP	30.56	44.99	50.89	45.14
3	E(X)	0.404 (0.785)	0.004 (0.169)	0.002 (0.040)	0.002 (0.034)
	E(X^2)	2.336 (0.626)	3.254 (0.957)	3.292 (0.114)	3.288 (0.114)
	AP	14.87	25.93	37.52	38.29

Table 4: LaPlace Distribution (Sampling Density I)

$\mu_* \backslash \sigma_*$		1	2	3	4
−3	E(X)	−0.188 (0.599)	−0.001 (0.085)	−0.001 (0.029)	−0.002 (0.027)
	E(X^2)	1.453 (0.578)	1.998 (0.431)	2.001 (0.094)	2.003 (0.093)
	AP	8.32	21.20	29.01	28.40
−2	E(X)	−0.052 (0.469)	−0.001 (0.043)	−0.002 (0.023)	0.000 (0.024)
	E(X^2)	1.692 (1.010)	1.997 (0.235)	2.001 (0.078)	2.006 (0.085)
	AP	20.88	38.44	38.54	33.27
−1	E(X)	−0.039 (0.204)	0.000 (0.029)	−0.001 (0.020)	−0.002 (0.022)
	E(X^2)	1.754 (0.715)	2.000 (0.189)	2.004 (0.071)	2.003 (0.085)
	AP	49.54	57.49	45.86	36.60
0	E(X)	−0.011 (0.249)	0.000 (0.021)	0.000 (0.019)	0.000 (0.023)
	E(X^2)	1.845 (1.151)	2.001 (0.155)	2.000 (0.069)	2.000 (0.083)
	AP	83.97	67.08	48.62	37.79
1	E(X)	0.029 (0.259)	0.000 (0.025)	0.000 (0.019)	0.001 (0.023)
	E(X^2)	1.766 (0.915)	2.001 (0.137)	1.998 (0.072)	2.002 (0.083)
	AP	49.45	57.48	45.86	36.62
2	E(X)	0.108 (0.359)	0.002 (0.044)	0.001 (0.023)	0.001 (0.024)
	E(X^2)	1.598 (0.782)	1.999 (0.224)	1.997 (0.080)	2.000 (0.087)
	AP	21.31	38.45	38.54	33.28
3	E(X)	0.201 (0.579)	0.005 (0.071)	0.001 (0.031)	0.001 (0.027)
	E(X^2)	1.444 (0.564)	1.975 (0.308)	1.998 (0.096)	2.000 (0.097)
	AP	8.35	21.18	29.01	28.39

Table 5: Gumbel Distribution (Sampling Density I)

μ_* \ σ_*		1	2	3	4
−3	E(X)	−0.151 (0.679)	0.000 (0.090)	0.000 (0.029)	−0.001 (0.029)
	E(X^2)	1.100 (0.822)	1.637 (0.457)	1.646 (0.078)	1.642 (0.074)
	AP	3.79	19.76	29.21	27.43
−2	E(X)	−0.063 (0.516)	0.000 (0.047)	0.000 (0.025)	0.000 (0.026)
	E(X^2)	1.334 (1.214)	1.643 (0.298)	1.644 (0.065)	1.647 (0.070)
	AP	18.91	39.78	38.17	31.77
−1	E(X)	−0.048 (0.214)	0.000 (0.029)	−0.001 (0.022)	−0.001 (0.026)
	E(X^2)	1.407 (0.674)	1.647 (0.195)	1.646 (0.058)	1.644 (0.065)
	AP	55.87	59.50	44.26	34.51
0	E(X)	−0.017 (0.130)	0.001 (0.023)	0.000 (0.022)	0.000 (0.026)
	E(X^2)	1.533 (0.571)	1.649 (0.155)	1.645 (0.055)	1.644 (0.064)
	AP	80.23	63.99	45.72	35.20
1	E(X)	−0.006 (0.075)	0.000 (0.023)	0.000 (0.024)	0.000 (0.028)
	E(X^2)	1.595 (0.360)	1.645 (0.073)	1.645 (0.057)	1.644 (0.065)
	AP	45.50	52.93	42.26	33.74
2	E(X)	0.006 (0.193)	0.002 (0.032)	0.000 (0.028)	0.001 (0.030)
	E(X^2)	1.627 (0.361)	1.647 (0.059)	1.645 (0.062)	1.647 (0.069)
	AP	20.65	35.44	35.13	30.42
3	E(X)	0.034 (0.717)	0.001 (0.049)	0.002 (0.035)	0.000 (0.033)
	E(X^2)	1.643 (0.607)	1.647 (0.083)	1.647 (0.071)	1.647 (0.075)
	AP	8.55	19.98	26.36	25.80

Table 6: Sampling Density II

σ_*		Bimodal	$t(5)$	Logistic	LaPlace	Gumbel
1	E(X)	0.001 (0.046)	−0.001 (0.054)	0.000 (0.088)	−0.002 (0.062)	0.001 (0.054)
	E(X^2)	1.625 (0.055)	1.676 (0.318)	3.299 (0.284)	2.008 (0.222)	1.649 (0.176)
	AP	71.12	72.21	80.81	69.95	72.75
2	E(X)	0.000 (0.033)	−0.002 (0.037)	−0.002 (0.054)	−0.001 (0.039)	0.000 (0.035)
	E(X^2)	1.625 (0.043)	1.673 (0.193)	3.295 (0.175)	2.003 (0.138)	1.646 (0.107)
	AP	53.76	53.11	65.04	52.32	53.36
3	E(X)	−0.001 (0.029)	−0.001 (0.031)	−0.002 (0.045)	−0.002 (0.035)	0.000 (0.032)
	E(X^2)	1.625 (0.040)	1.671 (0.142)	3.298 (0.155)	2.005 (0.125)	1.646 (0.092)
	AP	41.58	41.01	53.18	41.18	40.95
4	E(X)	0.001 (0.032)	−0.002 (0.030)	−0.002 (0.042)	−0.002 (0.032)	0.000 (0.033)
	E(X^2)	1.626 (0.044)	1.666 (0.128)	3.298 (0.140)	2.005 (0.112)	1.646 (0.090)
	AP	33.29	33.05	44.42	33.64	32.78
5	E(X)	0.001 (0.035)	−0.001 (0.030)	−0.002 (0.041)	−0.002 (0.032)	0.000 (0.035)
	E(X^2)	1.626 (0.047)	1.675 (0.137)	3.293 (0.141)	2.006 (0.115)	1.645 (0.091)
	AP	27.55	27.55	37.87	28.28	27.15
6	E(X)	0.000 (0.036)	0.000 (0.032)	−0.003 (0.041)	−0.001 (0.033)	0.002 (0.036)
	E(X^2)	1.627 (0.049)	1.679 (0.134)	3.297 (0.144)	2.002 (0.120)	1.649 (0.093)
	AP	23.42	23.56	32.84	24.32	23.11
7	E(X)	0.000 (0.040)	−0.002 (0.033)	0.000 (0.042)	−0.001 (0.035)	0.001 (0.037)
	E(X^2)	1.625 (0.054)	1.667 (0.125)	3.297 (0.148)	2.001 (0.123)	1.645 (0.099)
	AP	20.31	20.50	28.92	21.28	20.06

Table 7: Sampling Density III

	Bimodal	$t(5)$	Logistic	LaPlace	Gumbel
E(X)	0.001 (0.063)	0.000 (0.069)	−0.001 (0.049)	0.000 (0.039)	0.001 (0.054)
E(X^2)	1.626 (0.047)	1.654 (0.261)	3.284 (0.259)	2.001 (0.094)	1.644 (0.164)
AP	75.03	68.13	74.66	63.01	59.42

3, the largest AP is achieved at $(\mu_*, \sigma_*) = (0, 2)$ and the minimum standard errors are given by $(\mu_*, \sigma_*) = (0, 3)$. In Table 4, the maximum AP is $(\mu_*, \sigma_*) = (0, 1)$ and the smallest standard errors are around $(\mu_*, \sigma_*) = (0, 3)$. In Table 5, the largest AP is achieved at $(\mu_*, \sigma_*) = (0, 1)$ and the minimum standard errors are given by $(\mu_*, \sigma_*) = (0, 3)$. The acceptance probabilities which give us the smallest standard error are 62.91 in Table 1, 47.02 in Table 2, 65.65 in Table 3, 48.62 in Table 4 and 45.47 in Table 5, while the maximum acceptance probabilities are given by 66.98, 92.80, 70.65, 83.97 and 80.23, respectively. Thus, for the independence chain (Sampling Density I) we can conclude that the sampling density which gives us the maximum acceptance probability is not necessarily the best choice. We should choose the larger value than the σ_* which gives us the maximum acceptance probability. For Sampling Density II, the acceptance probability decreases as σ_* increases for all the target distributions (see Table 6). As discussed above, the smallest standard errors of the estimated moments are given by $\sigma_* = 3$ for Bimodal, $\sigma_* = 4$ for $t(5)$, $\sigma_* = 4, 5$ for Logistic, $\sigma_* = 4$ for LaPlace and $\sigma_* = 3, 4$ for Gumbel (see Table 7), where the acceptance probabilities are from 32.78 to 44.42. These acceptance probabilities are consistent with the results obtained in previous research (for example, see Carlin and Louis (1996, p.175), Chen, Shao and Ibrahim (2000, p.24), Gelman, Roberts and Gilks (1995), Besag, Green, Higdon and Mengersen (1995) and Gamerman (1997, p.165)). For all the five distributions, the standard errors at the optimum σ_* in Sampling Density II are larger than those in Sampling Density I. Therefore, if the appropriate sampling density is taken, Sampling Density I is better than Sampling Density II.

Final Comments: From Tables 1 – 7, Sampling Densities I and II with the appropriate hyper-parameters are better than Sampling Density III from the Se$(\overline{g(x)})$ criterion, but Sampling Densities I and II are not too different from Sampling Density III. The problem of Sampling Densities I and II is that we have to choose the hyper-parameters, which problem is a crucial criticism. If we take the hyper-parameters which give us the least standard errors of the estimated moments, it extremely takes a lot of time because we have to compute the standard errors for many combinations of the hyper-parameters. Since Sampling Density III does not have any hyper-parameter, it is easy to use Sampling Density III in practice and it can avoid the criticism of choosing the hyper-parameters included in Sampling Densities I and II. In addition, Sampling Density III shows quite good performance for all the five target densities. Therefore, Sampling Density III is more useful than Sampling Densities I and II.

5 Summary

In this paper, we have considered choice of the sampling density in the MH algorithm. The three sampling densities have been examined, i.e., the independence chain (Sampling Density I), the random walk chain (Sampling Density II) and the Taylored chain (Sampling Density III). Through the simulation studies based on the five target distributions, we have obtained the following results.

(i) Sampling Density I indicates the best performance when the appropriate hyper-parameters are chosen. That is, a scale parameter (σ_*) of Sampling Density I should be 1.5 – 2.5 times larger than that of the target density and a location parameter (μ_*) of Sampling Density I should be close to that of the target density.

(ii) Sampling Density II is optimal when the acceptance rate is about 30% – 50%, which result is consistent with past studies.

(iii) Sampling Density III shows a good performance for all the simulation studies. Moreover, because it does not depend on any hyper-parameter, Sampling Density III is more useful than Sampling Densities I and II in practical use.

References

Besag, J., Green, P., Higdon, D. and Mengersen, K., 1995, “Bayesian Computation and Stochastic Systems,” *Statistical Science*, Vol.10, No.1, pp.3 – 66 (with discussion).

Bernardo, J.M. and Smith, A.F.M., 1994, *Bayesian Theory*, John Wiley & Sons.

Carlin, B.P. and Louis, T.A., 1996, *Bayes and Empirical Bayes Methods for Data Analysis*, Chapman & Hall.

Chen, M.H., Shao, Q.M. and Ibrahim, J.G., 2000, *Monte Carlo Methods in Bayesian Computation*, Springer-Verlag.

Chib, S. and Greenberg, E., 1995, "Understanding the Metropolis-Hastings Algorithm," *The American Statistician*, Vol.49, No.4, pp.327 – 335.

Chib, S., Greenberg, E. and Winkelmann, R., 1998, "Posterior Simulation and Bayes Factors in Panel Count Data Models," *Journal of Econometrics*, Vol.86, No.1, pp.33 – 54.

Gelman, A., Roberts, G.O. and Gilks, W.R., 1995, "Efficient Metropolis Jumping Rules," in *Bayesian Statistics, Vol.5*, pp.599 – 607, edited by Bernardo, J.M., Berger, J.O., Dawid, A.P. and Smith, A.F.M., Oxford University Press.

Gamerman, D., 1997, *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*, Chapman & Hall.

Gentle, J.E., 1998, *Random Number Generation and Monte Carlo Methods*, Springer-Verlag.

Geweke, J., 1992, "Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments," in *Bayesian Statistics, Vol.4*, pp.169 – 193 (with discussion), edited by Bernardo, J.M., Berger, J.O., Dawid, A.P. and Smith, A.F.M., Oxford University Press.

Geweke, J., 1996, "Monte Carlo Simulation and Numerical Integration," in *Handbook of Computational Economics*, Vol.1, edited by Amman, H.M., Kendrick, D.A. and Rust, J., pp.731 – 800, North-Holland.

Geweke, J. and Tanizaki, H., 1999, "On Markov Chain Monte-Carlo Methods for Nonlinear and Non-Gaussian State-Space Models," *Communications in Statistics, Simulation and Computation*, Vol.28, No.4, pp.867 – 894.

Geweke, J. and Tanizaki, H., 2001, "Bayesian Estimation of State-Space Model Using the Metropolis-Hastings Algorithm within Gibbs Sampling," *Computational Statistics and Data Analysis*, Vol.37, No.2, pp.151-170.

Mengersen, K.L., Robert, C.P. and Guihenneuc-Jouyaux, C., 1999, "MCMC Convergence Diagnostics: A Reviewww," in *Bayesian Statistics, Vol.6*, pp.514 – 440 (with discussion), edited by Bernardo, J.M., Berger, J.O., Dawid, A.P. and Smith, A.F.M., Oxford University Press.

O'Hagan, A., 1994, *Kendall's Advanced Theory of Statistics*, Vol.2B (Bayesian Inference), Edward Arnold.

Robert, C.P. and Casella, G., 1999, *Monte Carlo Statistical Methods*, Springer-Verlag.

Smith, A.F.M. and Roberts, G.O., 1993, "Bayesian Computation via Gibbs Sampler and Related Markov Chain Monte Carlo Methods," *Journal of the Royal Statistical Society*, Ser.B, Vol.55, No.1, pp.3 – 23.

Tierney, L., 1994, “Markov Chains for Exploring Posterior Distributions,” *The Annals of Statistics*, Vol.22, No.4, pp.1701 – 1762.