Journal of the Korean Statistical Society (1999), **28**: *4, pp 443–455*

Monte Carlo Estimation of Multivariate Normal Probabilities

Man-Suk Oh[1] and Seung-Whan Kim[2]

ABSTRACT

A simulation-based approach to estimating the probability of an arbitrary region under a multivariate normal distribution is developed. In specific, the probability is expressed as the ratio of the unrestricted and the restricted multivariate normal density functions, where the restriction is given by the region whose probability is of interest. The density function of the restricted distribution is then estimated by using a sample generated from the Gibbs sampling algorithm.

Keywords: Gibbs sampler; Nonlinear constraints; Normal orthant probability; Density estimation.

1. INTRODUCTION

Consider a d-dimensional random vector $\mathbf{X} = (X_1, ..., X_d)$ following a multivariate normal distribution with mean μ and variance Σ. A problem that arises in many statistical application is that of computing the probability of a set

$$A = \{\mathbf{X} : \mathbf{a} \leq \mathbf{h}(\mathbf{X}) \leq \mathbf{b}\}, \tag{1.1}$$

where $\mathbf{a}$ and $\mathbf{b}$ are m-dimensional vectors such that $-\infty \leq \mathbf{a} < \mathbf{b} \leq \infty$ and $\mathbf{h}(\mathbf{X})$ is a m-dimensional vector of nonconstant functions of $\mathbf{X}$. Obviously, we exclude the case of $a_i = -\infty$ and $b_i = \infty$, where a_i and b_i are the i-th element of $\mathbf{a}$ and $\mathbf{b}$, respectively.

Except for a very few special cases, analytic expression of the above probability is not given and numerical estimation is in order. A simple method of estimating the probability is the Hit-Miss described in Rubinstein (1981), which generates samples from the multivariate normal distribution and estimate the

[1]Statistics Dept., Ewha Womans University, Seoul 120-750, Korea. This work was supported by KOSEF 981-0105-025-1.

[2]AID Consulting Institute, 625-11 Sinsa Dong, Gangnam Gu, Seoul, Korea.

probability by the relative frequency of samples falling in the region A. However, the Hit-Miss can be very inefficient if the probability is small because most of the samples are rejected, which is often the case especially in high dimension. There are more efficient but sophisticated Monte Carlo methods than the simple Hit-Miss. However, most of them are designed for simple functional forms of $\mathbf{h}$, such as linear functions of dimension less than or equal to d, see Geweke (1986, 1991), Evans and Swartz (1989), and Genz (1992). In situations where the region is defined by nonlinear functions or a combination of linear and nonlinear functions, there is no efficient and general method of estimating the probability.

Recently Markov Chain Monte Carlo method has become a very popular tool for handling difficult computational problems in statistics because it is easy to use and has wide range of application. Among the Markov Chain Monte Carlo methods, the Gibbs sampler (Gelfand and Smith, 1990) is particularly useful because it may convert a sample generation from a complicated multivariate distribution to a series of sample generations from simple univariate distributions. Thus, by using the Gibbs sampler one can easily generate samples of normally distributed random vector which is restricted to the region whose probability is of interest. Using these samples, we propose a simple and efficient method for estimating the probability of the region. Section 2 describes the sample generation and the estimation method in detail. Two illustrative examples are given in Section 3 and concluding remarks are given in Section 4.

2. THE EMTHOD

2.1. SAMPLE GENERATION

We describe the Gibbs sampling algorithm to generate multivariate normal samples restricted to the area of interest. We first introduce some notations. Let $N(\mu, \Sigma)$ denote the multivariate normal distribution with mean μ and variance Σ and let $N^A(\mu, \Sigma)$ denote the same normal distribution but restricted to the region A, where A is given in (1.1). Let $\mathbf{X} = (X_1, .., X_d)$ and $\mathbf{X}^A = (X_1^A, .., X_d^A)$ be random vectors following $N(\mu, \Sigma)$ and $N^A(\mu, \Sigma)$, respectively, and let $f(\mathbf{x})$ and $f^A(\mathbf{x}^A)$ be the density functions of $\mathbf{X}$ and $\mathbf{X}^A$, respectively. Finally, let $f_i^A(\cdot|x_j^A, j \neq i)$ be the conditional density function of X_i^A given $X_j^A = x_j^A, j \neq i$, let $f_i(\cdot|x_j, j \neq i)$ be the conditional distribution function of X_i given $X_j = x_j, j \neq i$, and let $P_i(C|x_j, j \neq i)$ be the conditional probability of $X_i \in C$ given $X_j = x_j, j \neq i$.

The following theorem gives the conditional distribution of X_i^A given $X_j^A, j \neq$

$i, j = 1, .., d$, which is necessary for the Gibbs sampling algorithm.

Theorem 2.1. *The conditional density function of* X_i^A *given* $X_j^A = x_j^A, j \neq i$, *is given by*

$$f_i^A(x_i^A | x_j^A, j \neq i) =$$
$$f_i(x_i^A | x_j^A, j \neq i) I(x_i^A \in A_i(x_j^A, j \neq i)) / P_i(A_i(x_j^A, j \neq i) | x_j^A, j = i),$$

where I *is the indicator function and*

$$A_i(x_j^A, j \neq i) = \{X_i; (x_1^A, .., x_{i-1}^A, X_i, x_{i+1}^A, .., x_d^A) \in A\}.$$

Proof: For simplicity, we will use the notation A_i for $A_i(x_j^A, j \neq i)$. Given $X_j^A = x_j^A, j \neq i$, the conditional density function of the i-th element X_i^A of $\mathbf{X}^A$ is given by

$$\begin{aligned} f_i^A(x_i^A | x_j^A, j \neq i) &\propto f^A(x_1^A, .., x_d^A) \\ &\propto f(x_1^A, .., x_d^A) I((x_1^A, .., x_d^A) \in A) \\ &\propto f_i(x_i^A | x_j^A, j \neq i) I(x_i^A \in A_i). \end{aligned}$$

Thus,

$$f_i^A(x_i^A | x_j^A, j \neq i) = \frac{f_i(x_i^A | x_j^A, j \neq i) I(x_i^A \in A_i)}{\int_{A_i} f_i(x_i^A | x_j^A, j \neq i) dx_i} = \frac{f_i(x_i | x_j^A, j \neq i) I(x_i^A \in A_i)}{P_i(A_i | x_j^A, j \neq i)}.$$

Since f_i is the conditional density function of X_i, it can be easily seen that $f_i(x_i | x_j^A, j \neq i)$ is the density function of a univariate normal distribution with mean $\alpha_i = \mu_i - \sum_{j \neq i}(\mu_j - x_j^A)\tau_{ij}/\tau_{ii}$ and variance $\beta_i = \tau_{ii}^{-1}$, where τ_{ij} is the (i, j)-th element of Σ^{-1}. Thus, the conditional distribution of X_i^A is a univariate normal distribution with mean α_i and variance β_i, but restricted to the region A_i. No matter how complicated the multi-dimensional region A is, the one-dimensional region A_i can be expressed as a union of disjoint intervals, so the conditional probability of $X_i \in A_i$ can be computed analytically. Therefore, efficient sample generation from the conditional distribution can be done by the inverse cdf method of Devroye (1986) or the mixed integration method of Geweke (1991).

Since the Gibbs sampler iteratively generates samples from conditional distributions, sample generation from $N^A(\mu, \Sigma)$ distribution can be done easily. In addition, because there is no waste of random samples once convergence is achieved in the Gibbs sampling algorithm, the algorithm may be efficiently applied even for the region with a very small probability.

2.2. ESTIMATION OF THE PROBABILITY

With a sample from the region of interest, we propose a method of estimating the probability of the region. We first give a simple lemma which relates the probability with the density function of $N^A(\mu, \Sigma)$.

Lemma 2.1. *For an arbitrary point* $\mathbf{x}^A \in A$,

$$P(\mathbf{X} \in A) = \frac{f(\mathbf{x}^A)}{f^A(\mathbf{x}^A)}.$$

Proof: Straightforward from $f^A(\mathbf{x}^A) = f(\mathbf{x}^A)I(\mathbf{x}^A \in A)/P(\mathbf{X} \in A)$. □

Since f is known, the problem of estimating the probability becomes to estimating the density function f^A at one point. In general, estimation of density function is not easier than estimation of probability. In this situation, however, efficient sample generation from the distribution corresponding to the desired density function can be done. Moreover, in the conditional distribution of X_i^A, the restricted region A_i can be expressed as a union of disjoint intervals and hence the probability of A_i can be computed. This makes the conditional density function of each element X_i^A of $\mathbf{X}^A$ be given in closed form expression.

Oh (1999) proposed a simple method for estimating posterior density functions in Bayesian analysis, when a posterior sample is given and the conditional posterior density functions are given in closed forms. Clearly, the distribution $N^A(\mu, \Sigma)$ and the density function f^A meets the conditions for Oh (1999)'s method, if we consider $N^A(\mu, \Sigma)$ as a posterior distribution and f^A as the corresponding posterior density function. Thus, we can apply Oh (1999)'s method to estimate $f^A(\mathbf{x}^A)$ and then estimate the probability. The following theorem gives the resulting estimate.

Theorem 2.2. *For a given point* $\mathbf{x}^A \in A$, *define*

$$\hat{P}(\mathbf{X} \in A) = f(\mathbf{x}^A)/\hat{f^A}(\mathbf{x}^A),$$

where

$$\hat{f^A}(\mathbf{x}^A) = \frac{1}{n}\sum_{k=1}^{n} g(\mathbf{x}^A, \mathbf{X}_k^A) \tag{2.1}$$

$$g(\mathbf{x}^A, \mathbf{X}^A) = \Pi_{i=1}^{d} \frac{f_i(x_i^A | X_{(i-1)}^A, x_{-(i)}^A) I(x_i^A \in A_i(X_{(i-1)}^A, x_{-(i)}^A)}{P_i(A_i(X_{(i-1)}^A, x_{-(i)}^A) | X_{(i-1)}^A, x_{-(i)}^A)},$$

$X^A_{(i-1)} = (X^A_1, .., X^A_{i-1})$, $x^A_{-(i)} = (x^A_{i+1}, .., x^A_d)$, *and* $\{\mathbf{X}^A_k = (X^A_{1k}, .., X^A_{dk}),\ k = 1, .., n\}$ *is a sample of* $\mathbf{X}^A$ *obtained from the Gibbs sampling algorithm. Under some regularity conditions,* $\hat{P}(\mathbf{X} \in A) \to P(\mathbf{X} \in A)$ *almost surely as the sample size* n *gets large.*

Proof: Considering $N^A(\mu, \Sigma)$ as a posterior distribution and f^A as the corresponding posterior density function, from Oh (1999),

$$f^A(\mathbf{x}^A) = E[\Pi_{i=1}^d f_i^A(x_i^A | X^A_{(i-1)}, x^A_{-(i)})],$$

where the expectation is taken with respect to $\mathbf{X}^A$. Plugging the conditional density function given in Theorem 2.1 into the above equation yields

$$f^A(\mathbf{x}^A) = E[g(\mathbf{x}^A, \mathbf{X}^A)],$$

Taking the sample average of $g(\mathbf{x}^A, \mathbf{X}^A)$ gives the estimate of $f^A(\mathbf{x}^A)$ and the convergence result follows from Tierney (1994). □

From the delta method, one can estimate the variance of $\hat{P}(\mathbf{X} \in A)$ by

$$\hat{Var}(\hat{P}(\mathbf{X} \in A)) = \left(\frac{\hat{P}(\mathbf{X} \in A)}{\hat{f}^A(\mathbf{x}^A)}\right)^2 \hat{Var}(\hat{f}^A(\mathbf{x}^A)).$$

Now, $\hat{Var}(\hat{f}^A(\mathbf{x}^A))$ is given by

$$\hat{Var}(\hat{f}^A(\mathbf{x}^A)) = \frac{1}{n}\{\frac{1}{n}\sum_{i=1}^n [g^2(\mathbf{x}^A, \mathbf{X}^A_k) - [\hat{f}^A(\mathbf{x}^A)]^2\},$$

if independent samples are used. Sometimes, one obtains autocorrelated samples from one long-run of the Gibbs sampling algorithm. In such a case the variance estimate is given by

$$\hat{Var}(\hat{f^A}(\mathbf{x}^A) = \frac{1}{n}[\Omega_0 + \sum_{s=1}^q 2(1 - s/(1+q))\Omega_s], \qquad (2.2)$$

where

$$\Omega_s = \frac{1}{n-s}\sum_{k=1}^{n-s} g(\mathbf{x}^A, \mathbf{X}^A_k) g(\mathbf{x}^A, \mathbf{X}^A_{k+s}) - [\hat{f^A}(\mathbf{x}^A)]^2,$$

and q is the lag size where the autocorrelation of $\{g(\mathbf{x}^A, \mathbf{X}^A_k)\}$ is negligible. See Chib (1995) and Oh (1999) for more details on the variance estimate.

3. EXAMPLES

3.1. EXAMPLE 1

For a simple example, we consider the normal orthant probability $P(\mathbf{X} \geq 0)$ with $\mu = 0$. The normal orthant probability is required in various statistical area, for instance in measuring association of two categorical data. In general, however, the normal orthant probability is unknown and is very small in high dimension.

For illustrative purposes, we consider the case of $\sigma_{ii} = 1$ for $i = 1, .., d$ and $\sigma_{ij} = 0.5$ for all $i \neq j$, where σ_{ij} is the (i,j)-th element of covariance matrix Σ. In this case, the true orthant probability is known to be 0.5^d (Johnson and Kotz, 1972). But to see the performance of the proposed method, we estimated the probability and its standard error for various sample size n and dimension d.

As the point $\mathbf{x}^A$ at which the density functions are to be evaluated, we selected $\mathbf{x}^A = 0$ which has the largest density function value among the points in the region A. In sample generation from the Gibbs sampler we used the warm-up size 100 which seemed to be large enough in all cases considered here from a rough convergence check of the Gibbs sampler. After the warm-up, autocorrelated samples were taken from successive cycles of the Gibbs sampler, hence the variance was estimated from (2.2).

All the computations in this paper were done by using the Fortran programming language Power Station 4.0 in personal computer with a Pentium processor. Uniform random numbers are generated from the IMSL routine RNUN and standard normal cdf's were computed by using the IMSL routine ANORDF.

We first illustrated the estimation results with sample size 10 000 in Table 3.1 (a) and then results with sample size 100 000 in Table 3.1(b). In the two tables, columns 2 and 3 shows the estimate and the standard error, and columns 4 and 5 shows the lower and upper limits of a 95% confidence interval for the probability. The true probability is given in column 6 and computing time in column 7. To see the results more clearly we plotted the estimates, the confidence intervals, and the true probabilities for each dimension in Figure 3.1 (a) and (b). From the tables and figures, it is clear that the estimate from the proposed method converges to the true probability very quickly. With sample size 10 000, the estimate gets very close to the true value. Since the proposed method generates samples only from the region of interest, standard error of the estimate and computing time do not increase dramatically as the dimension increases, unlike most other numerical estimation methods. With sample size 100 000 the estimates are almost equal to the true values and the confidence intervals are very narrow, but the computing time seems to be still within acceptable range.

Table 3.1: Estimates of the normal orthant probability.

(a) $n = 10,000$

dimension	estimate	S.E.	lower limit	upper limit	true value
2	0.33230	0.00117	0.32997	0.33467	0.33333
5	0.16305	0.00207	0.15890	0.16721	0.16667
10	0.08946	0.00312	0.08320	0.09571	0.09091
15	0.06169	0.00443	0.05281	0.07056	0.06250
20	0.03766	0.00612	0.02540	0.04991	0.04762
25	0.03291	0.00826	0.01639	0.04943	0.03846

(b) $n = 100,000$

dimension	estimate	S.E.	lower limit	upper limit	true value
2	0.33316	0.00036	0.33242	0.33389	0.33333
5	0.16528	0.00070	0.16386	0.16669	0.16667
10	0.09165	0.00101	0.08962	0.09360	0.09091
15	0.06147	0.00141	0.05864	0.06430	0.06250
20	0.05275	0.00212	0.04850	0.05701	0.04762
25	0.04084	0.00349	0.03385	0.04783	0.03846

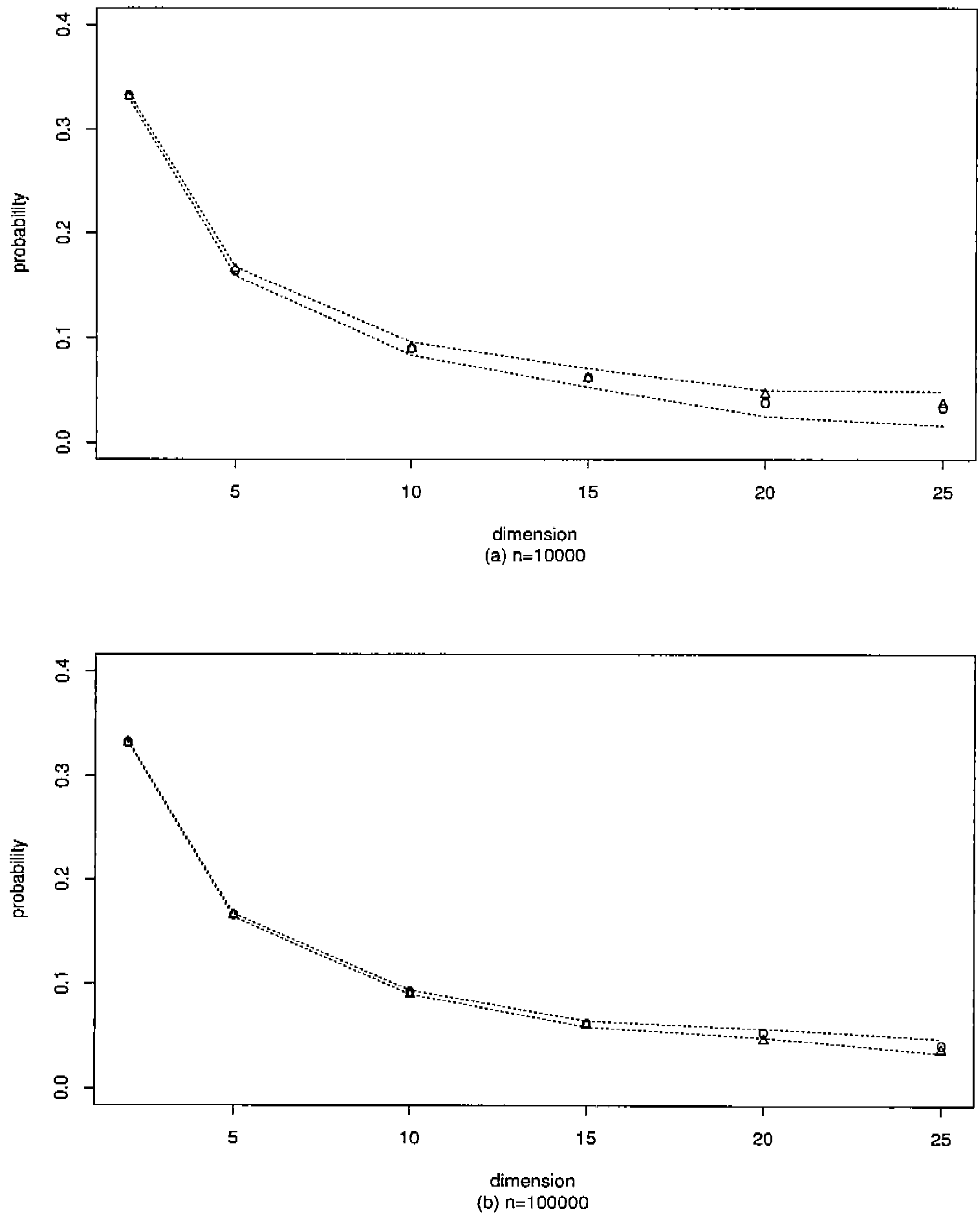

Figure 3.1: The Normal Orthant Probability Estimates (dots: estimates, dotted lines: 95% confidence interval, triangle: true value).

3.2. EXAMPLE 2

When $\mathbf{y}$ follows a d-dimensional normal distribution with unknown mean vector θ and known variance Σ, an interesting question is whether $H_0 : \theta = 0$ or $H_1 : \theta \geq 0$. In a Bayes test of the above hypotheses, a proper prior density $\rho(\theta)$ is assigned to the area $\theta \geq 0$ and decision is often made depending on the value of the Bayes factor B which is given by

$$B = \frac{f(\mathbf{y}|\theta = 0)}{\int f(\mathbf{y}|\theta)\rho(\theta)d\theta},$$

where $f(\mathbf{y}|\theta)$ is the density function of data $\mathbf{y}$ given θ.

Since the Bayes factor is sensitive to the choice of the prior ρ, it is often interesting to find the lower bound of the Bayes factor over some reasonable classes of density functions for ρ. One of the reasonable classes for ρ would be the class of unimodal symmetric density functions with mode at 0. Oh (1998) shows that the lower bound of B over the above class is given by

$$\underline{B} = \frac{exp[-\frac{1}{2}\mathbf{y}'\Sigma^{-1}\mathbf{y}]}{\sup_k \frac{1}{V(k)} \int_{\theta \geq 0, \theta'\Sigma^{-1}\theta \leq k^2} exp[-\frac{1}{2}(\theta - \mathbf{y})'\Sigma^{-1}(\theta - \mathbf{y})]d\theta},$$

where $V(k)$ is the volume of the region $A = \{\theta; \theta \geq 0,\ \theta'\Sigma^{-1}\theta \leq k^2\}$.

The integral in the denominator can not obtained analytically. But it can be expressed as $(2\pi)^{d/2}|\Sigma|^{1/2}$ times $P(\theta \in A)$, where θ follows $N(\mathbf{y}, \Sigma)$ distribution. Thus, the problem becomes estimation of the probability $P(\theta \in A)$. Note that in this example θ plays the same role as $\mathbf{X}$ in Section 2.

If we denote the (i,j)-th element of Σ^{-1} by τ_{ij} and let $\theta^A = (\theta_1^A, .., \theta_d^A)$ follow $N^A(\mathbf{y}, \Sigma)$, the normal distribution restricted to the region A, then the conditional distribution of θ_i^A is restricted normal distribution $N(\alpha_i, \beta_i)I(0 \leq \mu_i \leq c_i)$, where $\alpha_i = y_i - \sum_{j \neq i}(\theta_j^A - y_j)\tau_{ij}/\tau_{ii}$, $\beta_i = \tau_{ii}^{-1}$, and

$$c_i = \frac{1}{\tau_{ii}}\sqrt{k^2 - \sum_{j,k \neq i} \tau_{jk}\theta_j^A\theta_k^A - (\sum_{j \neq i} \tau_{ij}\theta_j^A)^2/\tau_{ii}} - \sum_{j \neq i} \tau_{ij}\theta_j^A/\tau_{ii}.$$

Thus,

$$f^A(\theta_i^A|\theta_j^A, j \neq i) = \frac{phi((\theta_i^A - \alpha_i)/\sqrt{\beta_i})I(0 \leq \theta_i^A \leq c_i)}{\Phi((c_i - \alpha_i)/\sqrt{\beta_i}) - \Phi(-\alpha_i/\sqrt{\beta_i})},$$

where ϕ and Φ are respectively the standard normal density and distribution functions. Note that the region $A = \{\theta; \theta \geq 0,\ \theta'\Sigma^{-1}\theta \leq k^2\}$ is converted to a fixed interval $(0, c_i)$ in the conditional distribution of θ_i^A so that the sample generation is easy and the conditional density function is given in closed form.

For simple illustration, we choose $k = 2$, $\Sigma = I_d$, the d-dimensional identity matrix, and $\mathbf{y} = (t/\sqrt{d})\mathbf{1}$ for some constant t, where $\mathbf{1}$ is the vector of all elements 1. Table 3.2 presents estimates of $P(\theta \in A)$ and their standard errors for some selected values of d and t. In this example, we selected the point which is closest to the origin as $\mathbf{x}^A$ since it has the largest density function value among the points in the region A.

Table 3.2: Estimates of the probability in Example 2

dimension (sample size)	t	estimate	SE	time (sec)	estimate (Hit-Miss)	sample size (Hit-Miss)	time(sec) (Hit-Miss)
(10000)	0.5	3.50E-01	6.98E-04	0	3.48E-01	467703	0
	1	3.79E-01	7.42E-04	0	3.79E-01	427135	0
	2	3.76E-01	1.03E-03	0	3.77E-01	222116	0
	3	3.43E-01	1.87E-03	0	3.45E-01	64389	1
	4	2.97E-01	3.36E-03	0	2.94E-01	18471	0
5(20000)	0.5	2.78E-02	3.01E-04	8	2.75E-02	297636	7
	1	3.25E-02	2.72E-04	7	3.24E-02	425105	10
	2	3.42E-02	2.56E-04	7	3.50E-02	502579	11
	3	3.33E-02	4.05E-04	7	3.33E-02	196026	4
	4	2.94E-02	8.35E-04	8	2.89E-02	40892	1
10(30000)	0.5	1.08E-04	1.28E-05	23	1.03E-04	662026	28
	1	1.31E-04	7.04E-06	23	1.50E-04	2646509	114
	2	1.48E-04	3.83E-06	23	1.54E-04	10039766	434
	3	1.45E-04	3.87E-06	22	1.46E-04	9725135	421
	4	1.32E-04	7.43E-06	21	1.30E-04	2381505	103
15(40000)	0.5	2.08E-07	5.79E-08	47	?	62004876	3844
	1	2.21E-07	3.62E-08	47	?	168505648	10447
	2	2.09E-07	1.34E-08	47	?	1165140224	72238
	3	2.08E-07	9.86E-09	45	?	2135097472	136646
	4	1.82E-07	1.74E-08	43	?	599941888	37196
20(50000)	0.5	4.35E-10	9.18E-11	81	?	51528884224	4328426
	1	1.63E-10	4.30E-11	81	?	88208449536	7409509
	2	1.75E-10	3.22E-11	80	?	1.69E+11	14190875
	3	1.44E-10	1.31E-11	79	?	8.44E+11	70922556
	4	1.35E-10	2.57E-11	73	?	2.03E+11	17093970

To see the efficiency of the proposed method, we also consider estimates from the Hit-Miss algorithm which seems to be the only alternative method for computing the probability in this example. From the table, it is clearly seen that the estimate gets close to 0 very fast as the dimension increases. So it is meaningless to compare the standard errors across dimensions and one may require small standard errors for small estimates. Thus, we increased the sample size by 10 000 as the dimension increases by 5 except for dimension 2, and it seemed to give a reasonable size of the standard error. Next, to compare the proposed method with the Hit-Miss, we ran the Hit-Miss until it achieved the same standard error in each case and observed the required sample size and computing time. The Hit-Miss results are given in the last two columns of Table 3.2. Obviously, the computing time for the Hit-Miss blows up so that the Hit-Miss becomes infeasible, as dimension gets larger than 10. The reason is that the probability is very small in high dimension, yielding too high rejection rate in the Hit-Miss algorithm. In contrast, the computing time for the proposed method is within some reasonable range even in high dimension. Thus, even repeated runs of the proposed method is possible which is often required in simulation study.

4. CONCLUDING REMARKS

We have proposed a simulation-based method for estimating the probability of an arbitrary region which may be defined by complicated nonlinear constraints, under multivariate normal distributions. The method generates samples from the multivariate normal distribution restricted to the region whose probability is of interest by using the Gibbs sampling algorithm. With the sample, the density function of the restricted normal distribution is estimated by the sample average of an appropriate function, by using Oh (1999)'s method. Then the probability is estimated by the ratio of density functions of the unrestricted and the restricted normal distributions, evaluated at one point.

The method has some great features. First, the method is very general in that it can handle any form of constraint in A, whereas most analytic or numerical approximation methods developed so far require some specific forms of constraints. Second, it is efficient even when the probability is very small since samples are generated only from the region of interest. Moreover, the samples can be used not only for estimation of the probability but also for other statistical inference. Third, it is very easy to use. All that actually required for the method are sample generations from univariate restricted normal distributions and computations of the corresponding density functions, which can be done easily with existing statistical software.

The point $\mathbf{x}^A$ in the estimate (2.1) can be chosen arbitrarily. But in practical application of the method, the accuracy of the estimate is sensitive to the choice of $\mathbf{x}^A$. Unfortunately, there is no obvious rule for optimal $\mathbf{x}^A$. From experience, however, we suggest a few guidelines for selection of $\mathbf{x}^A$. First, a point $\mathbf{x}^A$ from a tail area of $\mathbf{x}^A$ is not good. It seems to highly underestimate the probability and its variance. Second, a point $\mathbf{x}^A$ near the mode of f^A seems to be good. Thus, when the mode is known we suggest to choose the mode as $\mathbf{x}^A$. When it is difficult to locate the mode, which is often the case in high dimensions, we suggest to try several different $\mathbf{x}^A$s at which f is relatively high, and take the best estimate or the average of estimates as a final result.

Finally, further research interest would be computation of the probability of an arbitrary region under a distribution whose density function is known only up to a functional form.

References

Chib, S. (1995). "Marginal Likelihood from the Gibbs Output," *Journal of the American Statistical Association*, 90, 1313-1321.

Devroye, M. (1986). *Non-Uniform Random Generation*, Springer-Verlag, New-York.

Evans, M. and Swartz, T. (1989). "Monte Carlo Computation of Some Multivariate Normal Probabilities," *Journal of Statistical Computation and Simulation*, 30, 117-128.

Gelfand, A.E. and Smith, A.F.M. (1990). "Sampling-Based Approaches to Calculating Marginal Densities," *Journal of the American Statistical Association*, Vol. 85, 398-409.

Genz, A. (1992). "Numerical Computation of Multivariate Normal Probabilities," *Journal of Computational and Graphical Statistics*, Vol. 1, 141-150.

Geweke, M. (1986). "Exact Inference in the Inequality Constrained Normal Linear Regression Model," *Journal of Applied Econometrics*, Vol. 1, 127-141.

Geweke, M. (1991). "Efficient Simulation from the Multivariate Normal and Student-t Distributions Subject to Linear Constraints," *Computing Science and Statistics (Proceedings of the 23rd Symposium on the Interface).*

Johnson, N.L. and Kotz, S.(1972). *Continuous Multivariate Distributions*, John Wiley, New York.

Oh, M-S (1998). "A Bayes Test for Simple versus One-sided Hypotheses on Multivariate Normal Mean," *Communications in Statistics-Theory and Methods*, to appear.

Oh, M-S. (1999). "Estimation of Posterior Density Functions from a Posterior Sample," *Computational Statistics and Data Analysis*, in revision.

Rubinstein, R.Y. (1981). *Simulation and the Monte Carlo Method*, Wiley & Sons, New-York.

Tierney, L. (1994). "Markov Chains for Exploring Posterior Distributions," *The Annals of Statistics*, 22, 1701-1762.