응용통계연구 제 7권 2호
1994년 9월 pp. 113-130

Further Development in the Integration of Multimodal Functions by Monte Carlo Importance Sampling[1)]

Man-Suk Oh[2)]

Abstract

The algorithm of Oh and Berger (1993) is extended to handle more general cases where the integrand $f(\theta)$ is not only multimodal but also skewed or has some undetected modes, each having curvature not much different from that of the nearest component. It runs Oh and Berger's algorithm in an iterative way, adding a component in each stage to the mixture importance function from previous stage for better approximation between $f(\theta)$ and the importance function.

1. Introduction

In Bayesian analysis, many posterior inferences are based on the posterior expectation of an appropriate function $\varphi(\theta)$ of the form,

$$E\varphi = \frac{\int \varphi(\theta) f(\theta) d\theta}{\int f(\theta) d\theta}, \tag{1}$$

where $f(\theta)$ is a product of the likelihood and a prior of θ so that $f(\theta) / \int f(\theta) d\theta$ is a posterior density. However, in most practical cases, the integrals in (1) can not be computed analytically and numerical integration is necessary. Recently, there has been significant development in this era. Among them are Monte Carlo importance sampling by van Dijk, Kloek, and Boender (1985), Geweke (1988, 1989), Evans (1989, 1991a, 1991b), Oh and Berger (1992, 1993), West(1991); Quadrature method by Naylor and Smith (1982, 1988), Smith et al. (1985); Sampling based method by Geman and Geman (1984), Tanner and Wong (1987), Gelfand and Smith (1990), Gelfand, Smith and Lee (1992), Gelfand et. al. (1990).

Among these schemes, Monte Carlo importance sampling (importance sampling, from here on) is known to be general, easy to apply, and effective in many cases. Importance sampling performs as follows: (i) Generate random samples $\theta_1, \ldots, \theta_n$ from a density

1) This paper was supported by NON DIRECTED RESEARCH FUND, Korea Research Foundation, 1993.

2) Department of Statistics, Ewha Womans University, Seo-Dae-Moon-Gu, Seoul, 120-750, KOREA.

function g, called the importance function (ii) estimate (1) by

$$\widehat{E\varphi} = \frac{\sum_{i=1}^{n} \varphi(\theta_i)\, w(\theta_i)}{\sum_{i=1}^{n} w(\theta_i)},$$

where the weight function $w(\theta)$ is $f(\theta)/g(\theta)$. The efficiency of importance sampling depends on the choice of importance function g whose desirable properties are (i) convenient Monte Carlo property, i.e., easy random generation (ii) having thicker tails than f, and (iii) good approximation to f. A common procedure for the selection of g is to first choose a parametric form satisfying (i) and (ii) and then choose paramters of the importance function for property (iii). The most crucial and difficult in many cases is the last step, choosing the parameters of g for a good approximation to f.

A good measure of approximation between f and g is the reciprocal of the squared variation coefficient of the weight function

$$CV^2(w) = \int \left(\frac{w(\theta)}{\int f(\theta) d\theta} - 1 \right)^2 g(\theta)\, d\theta. \qquad (2)$$

Thus, once (i) and (ii) are achieved, which is not difficult in most cases, it would be desirable to choose g which yields a small $CV^2(w)$.

Extensive research has been done on the choice of importance function (Geweke (1988), Evans (1991b), Oh and Berger (1992, 1993)). In particular, Oh and Berger (1993) suggested use of a mixture of multivariate t density functions as importance function to handle multimodal $f(\theta)$. A mixture density function $g(\theta)$ is written as

$$g(\theta) = \varepsilon_1 g_1(\theta) + \ldots + \varepsilon_m g_m(\theta)$$

where $0 < \varepsilon_i < 1$, $\sum_{i=1}^{m} \varepsilon_i = 1$ and $g_i(\theta)$ are density functions. Parameters ε_i are called mixing weights, $g_i(\theta)$ are called component density functions, and m is the number of components.

Oh and Berger's use of a mixture density to handle complicated f has some great advantages: First, it is easy to generate random samples from a mixture density function but also a mixture density function can have many different shapes such as unimodal, multimodal, and highly skewed. Second, as shown in Oh and Berger one can greatly reduce the error of numerical integration with almost no extra cost by use of stratification and control variates. To choose parameters of the mixture importance function $g(\theta)$, Oh and Berger (1993) set the number of components equal to the number of modes in $f(\theta)$

and used numerical minimization of a Monte Carlo estimate of $CV^2(w)$ with respect to mixing weights and parameters of component density functions.

In many practical cases, $f(\theta)$ is not only multimodal but also skewed or there are some undetected modes. But Oh and Berger's algorithm does not handle the above cases efficiently because it fixes the number of components in g equal to the number of (detected) modes in f. In other words, it uses only one component for each detected mode in f. Obviously it would be better to have extra components in the skewed area or near undetected modes.

To handle these cases, the algorithm in this paper runs Oh and Berger (1993)'s minimization routine in an iterative way. In each stage, we add a component to the importance function from previous stage and then run a minimization routine to improve the new mixture importance function. The process stops when there is no significant reduction in the estimate of $CV^2(w)$. Note that the number of components in the mixture importance function increases as the process continues and this makes the importance function become a better approximation to f.

This paper is organized as follows. Initialization of mixing weights and parameters of the additional component for a minimization routine is described in Section 2.1, a linear interpolation with respect to mixing weights is mentioned in Section 2.2, and the algorithm is given in Section 2.3. A simple two dimensional example with contour plots of f and the importance functions from each stage of the algorithm is shown in Section 3, and summary and discussions in Section 4.

2. The Algorithm

2.1 Initialization of an additional component

As in Oh and Berger (1993), we consider in this paper the case where $g_i(\theta)$ are density functions of multivariate t distributions. Thus, to select $g(\theta)$, we need to choose the mixing weights, the degrees of freedom, the location and the scale parameters of each multivariate t component density.

2.1.1 Location, scale and the degrees of freedom

As mentioned in Section 1, a good measure of approximation between f and g is the

reciprocal of $CV^2(w)$ given in (2), assuming that the support of g covers that of f. Thus, it would be ideal to locate the additional component in a way to reduce $CV^2(w)$. One way of achieving it is to consider $CV^2(w)$ from previous stage and locate the additional component at a point where the integrand of $CV^2(w)$,

$$\left(\frac{w(\theta)}{\int f} - 1 \right)^2 g(\theta), \tag{3}$$

is large subject to $w(\theta)/\int f > 1$ and the curvature of f is large in positive direction. (The restriction $w(\theta)/\int f > 1$ is necessary since the additional component should be located where g is smaller than $f/\int f$.) But it is very difficult to find such a point especially in high dimensional space because the computation of curvature matrices at many points can be expensive. However, unless there is extremely sharp undetected mode in f, we may ignore curvature, simply consider (3), and use the point which maximizes (3) as initial location of the additional component.

As a result of considering (3) only, the algorithm does not add components in an ideal order. When there is an undetected sharp peak with large height, the algorithm may locate the first additional component near the sharp peak. But there can be another mode with smaller height but larger curvature so that addition of a component near it would have reduced $\widehat{CV}^2(w)$ more. This may cause a problem if the undetected sharp peak is too sharp to significantly reduce $\widehat{CV}^2(w)$ so that the algorithm stops before hitting the major mode. But in many cases extremely sharp peaks can be easily detected from the study of $f(\theta)$ (For instance from the study of the likelihood and the prior in Bayesian analysis). Moreover, an undetected extremely sharp peak is a problem in most efficient numerical integration algorithm.

If one still suspects about the existence of an extremely sharp undetected mode, one may run a few more stages even after the algorithm satisfies the stopping condition. Or one may look for a peak with the location of the additional component as a starting point and compute curvature at the peak to make sure the algorithm did not stop because of an extremely sharp peak.

Another justification of using the maximum of (3) for initial location of the additional component can be given as follows. The weight function $w(\theta) = f(\theta)/g(\theta)$ also

reflects approximation between f and g, hence we need to consider points where $w(\theta)$ is large, or equivalently where a standardized weight

$$\left(\frac{w(\theta)}{\int f} - 1 \right) \tag{4}$$

is large. Note that $w(\theta)/\int f$ is a normalized weight which has mean 1. However, in locating the additional component we need to also consider a standardized f, $f(\theta)/\int f$, as a measure of seriousness of (4) because large (standardized) $w(\theta)$ would be a more serious problem where $f(\theta)$ is large than where $f(\theta)$ is small. (There would be more observations of θ where $f(\theta)$ is large than where $f(\theta)$ is small.) This leads to the consideration of

$$\left(\frac{w(\theta)}{\int f} - 1 \right) \frac{f(\theta)}{\int f}. \tag{5}$$

Note that $f(\theta)/\int f$ is like a weight to (4). From simple calculations it can be shown that (3) is equal to

$$\left(\frac{w(\theta)}{\int f} - 1 \right) \frac{f(\theta)}{\int f} + \left(g(\theta) - \frac{f(\theta)}{\int f} \right) \tag{6}$$

and (5) is the first term of (6). The second term in (6) clearly reflects a linear difference between $f/\int f$ and g while the first term reflects a rational difference. Thus, it would be reasonable to choose a component to reduce a combination of the two terms.

Finding the maximum of (3) can be a difficult task especially because there can be many local maxima. But it is not necessary at all to find the exact maximum and a point which is near the maximum point would be sufficient for our purposes here. (Recall that our main purpose here is to significantly reduce $\widehat{CV}^2(w)$, i.e., significantly improve the importance function.) So we simply examine values of (3) at sample points generated in the previous stage and the point which yields the maximum among those sample points is selected as initial location of the additional component.

Because we consider the cases where f is skewed or f have undetected modes each having curvature not much different from that of the nearest component, we match the scale parameter of the additional component with that of the nearest component. Since tail

area is relatively less important (there would be fewer samples in the area), Oh and Berger used the same degrees of freedom for each component. If one decides to use different degrees of freedom for different components, it would be reasonable to use the degrees of freedom of the nearest component as those of the additional component because the nearest component is likely to cover the skewed area or the undetected modal area and hence to incorporate tail rates of the area. Thus here we use the degrees of freedom of the nearest component as those of the additional component.

2.1.2 The Mixing weights

Once initial location, scale and degrees of freedom of the additional component are determined, i.e., the additional component density function is chosen, we need to choose a mixing weight to combine it with the importance function from the previous stage.

In Oh and Berger(1993), mixing weights were chosen to match the heights of f and g at each mode. The same idea can not be applied here since the additional component may not be located at a mode. But we can use a similar idea: Consider an overall distance between f and the new mixture importance function, and choose mixing weights to minimize it. To be specific, consider

$$\min_{\varepsilon} \| f(\theta) - [(1-\varepsilon) g_p(\theta) + \varepsilon g_a(\theta)] \|_2^2 \tag{7}$$

$$= \min_{\varepsilon} \int [f(\theta) - ((1-\varepsilon) g_p(\theta) + \varepsilon g_a(\theta))]^2 d\theta , \tag{8}$$

where g_p is the (mixture) importance function from previous stage and g_a is the additional component density function. (Note that $(1-\varepsilon) g_p(\theta) + \varepsilon g_a(\theta)$ is the mixture importance function in the current stage.) It is easy to show that (8) is minimized at ε^{**}, where

$$\varepsilon^{**} = \frac{\int (f(\theta) - g_p(\theta))(g_a(\theta) - g_p(\theta)) d\theta}{\int (g_a(\theta) - g_p(\theta))^2 d\theta} . \tag{9}$$

An estimate of (9) can be easily obtained using random samples from previous stage.

2.2 A Linear Interpolation with respect to the Mixing Weights

In Oh and Berger (1993), we used the number of random samples from the ith

component equal to the integer part of the total number of random samples times the mixing weight of the ith component. So the estimate of $CV^2(w)$ is not actually a continuous function of mixing weights $\{\varepsilon_i\}$ but a step function of it, with a very small step size. Thus it is possible that the minimization routine is stabilized before reaching the minimum of $\widehat{CV}^2(w)$. But from our experience this did not cause a serious problem unless the initial importance function is very bad. We do not know the exact reason but possible reasons may be (i) a little early stop of the routine with respect to $\{\varepsilon_i\}$ can be adjusted by appropriate choice of the location and the scale parameters (ii) the discretization makes very little difference in $\widehat{CV}^2(w)$, in other words, the step size is very small and (iii) it does not change the decreasing (increasing) direction of $\widehat{CV}^2(w)$ so the routine always moves in the direction of reducing $\widehat{CV}^2(w)$.

When we know all the modes in f, the initial importance function selected as suggested in Oh and Berger (1993) is likely to be a reasonable choice. Thus smoothing with respect to $\{\varepsilon_i\}$ seems not necessary. However, when an additional component is added, the initial mixture importance function would not be as good as in the above case. For instance, when there is only one component in a skewed area, the scale of the component would be quite large to cover the skewed area. So when an additional component with the same scale is mixed with the existing component, the scale of the mixture would be much larger than necessary to cover the skewed area. Clearly it would be better to shrink the scale of the two components but this is not done in the initialization step. Therefore, in cases we are considering here the initial importance function may take longer to reach a good importance function and there is more risk of stabilizing too early. To get around this problem we suggest a simple smoothing with respect to ε_i such as a linear interpolation.

Of course a linear interpolation in $m-1$ dimensional space (there are $m-1$ independent ε_i's) requires some extra computations but it would result in a better performance (i.e., better importance function) thus would eventually save cost in the actual importance sampling. Therefore, we recommend to use a linear interpolation especially when one doesn't feel comfortable with the initial importance function.

Actually, the simplest way to handle the discretization problem mentioned above is to make the moving step size of $\{\varepsilon_i\}$ in a minimization subroutine always larger than the step size of $\{\varepsilon_i\}$ in $\widehat{CV}^2(w)$. But this requires full knowledge about the minimization subroutine one will use. With NAG subroutine we used here, we did not know how to

control the moving step size.

2.3 The Algorithm

Suppose that $f(\theta)$ has m detected modes. An iterative runs of Oh and Berger (1993), adding a component in each stage as described in Section 2.1, leads to the following algorithm.

Stage 1: Initialize the mixture importance function with m components and run the minimization routine as described in Oh and Berger. Let the obtained importance function be $g_{p(\theta)}$ and $k=2$

Stage k $(k \geq 2)$:

* Step 1: Compute

$$\left(\frac{w_p(\theta_{i,p})}{\overline{w_p}} - 1 \right)^2 g_{p(\theta_{i,p})} \tag{10}$$

for each of $\theta_{i,p}, \cdots, \theta_{n,p}$, satisfying $w_p(\theta_{i,p}) / \overline{w_p} - 1 > 0$, where $\theta_{1,p}, \cdots, \theta_{n,p}$ are random samples from previous stage, $w_p(\theta) = f(\theta)/g_p(\theta)$ and $\overline{w_p} = \sum_{i=1}^{n} w_p(\theta_{i,p})/n$. Let $\mu^* = \theta_{l,p}$, where $\theta_{l,p}$ maximizes (10) for $l \in \{1, \ldots, n\}$.

* Step 2: Find a component $ll \in \{1, \ldots, m\}$. which minimizes

$$(\mu^* - \mu_{i,p})' T_{i,p}'^{-1} T_{i,p}^{-1} (\mu^* - \mu_{i,p}) \quad , \tag{11}$$

where $\mu_{i,p}$, is the location and $T_{i,p} T_{i,p}'$ is the scale parameters of the ith component in $g_p(\theta)$. Let $T^* = T_{ll,p}$ and $\alpha^* = \alpha_{ll}$.

* Step 3: Let $g_\alpha(\theta)$ be the density function of multivariate t distribution with α^*, μ^*, and $T^* T^{*'}$ as the degrees of freedom, location, and scale parameters, respectively.

* Step 4: Let

$$\varepsilon^* = \frac{\sum_{i=1}^{n} (f(\theta_{i,p}) - (g_p(\theta_{i,p}))(g_\alpha(\theta_{i,p}) - (g_p(\theta_{i,p}))}{\sum_{i=1}^{n} (g_\alpha(\theta_{i,p}) - (g_p(\theta_{i,p}))^2}$$

(Note that ε^* is an estimate of ε^{**} given in (9).)

* Step 5: Run the minimization routine of Oh and Berger with

$$(1-\varepsilon^*)g_p(\theta) + \varepsilon^* g_a(\theta)$$

as the initial mixture importance function and with $m = m + 1$.

* Step 6: If there is no significant reduction in the estimate of $CV^2(w)$, stop and use $g_p(\theta)$ as importance function in the actual importance sampling. Otherwise, continue iterating with the importance function obtained in the current stage as g_p and with $k = k + 1$.

3. Examples

3.1 Example 1

To illustrate efficiency of the algorithm proposed here, a two dimensional example, where f is skewed in two different directions, is given in this section. Let $f(\theta)$ be a density function of the form

$$f(\theta) = 0.30f_1(\theta) + 0.30f_2(\theta) + 0.25f_3(\theta) + 0.15f_4(\theta),$$

where $f_1, \cdots f_4(\theta)$ are density functions of $N((0.0, 0.0)', 0.67I)$, $N((3.0, 2.5)', I)$ and $N((0.0, 2.5)', I)$, $N((2.0, 0.0)', I)$, respectively. Thus, $f(\theta)$ is actually a mixture density functions and its integral is known. But for illustrative purposes let us apply the algorithm here. As shown in Figure 1, $f(\theta)$ has two modes

$$\hat{\theta}_1 = (0.1200, 0.0250), \quad \hat{\theta}_2 = (2.9555, 2.4340)$$

and is skewed in the directions of $(2.0, 0.0)'$ and $(0.0, 2.5)'$. Because there are two modes Oh and Berger's algorithm was applied with $m = 2$ and $\hat{\theta}_1$, $\hat{\theta}_2$, as initial locations of the component density functions.

Figures of initial importance function and the importance function after Oh and Berger's algorithm are given in Figures 2 and 3, respectively. One can see some improvement but the skewed area of $f(\theta)$ is not well taken care of. In stage 2, a component density function was added to the importance function from the first stage and Oh and Berger's minimization routine was run, according to the algorithm described in Section 2. The

resulting importance function is shown in Figure 4 and a significant improvement can be seen.

In Stage 3, a component was added again and the minimization routine was applied. There was a significant improvement from Stage 2 in the estimate of $CV^2(w)$. The importance function became the one in Figure 5, now a very good approximation of $f(\theta)$. At stage 4 the estimate of $CV^2(w)$ was not reduced much from that in Stage 3. Therefore, we stopped the process and the importance function obtained from Stage 3 is suggested as importance function in the actual Importance Sampling.

Now, to illustrate efficiency of the algorithm when there are undetected modes, let us start with a unimodal importance function, i.e., let us assume that we detected only one mode $\hat{\theta}_1 = (0.1200, 0.0250)$ of $f(\theta)$. Obviously the initial importance function is a very poor approximation of f (Figure 6). The algorithm started with $m = 1$ and stopped after 5 stages and the importance function at Stage 4 will be taken. Figures 7 to 10 shows the importance functions obtained from each stages. The final importance function is not as good as the one in Figure 5 (the final importance function when we started with two components) but considering that we started with a very poor approximation, the algorithm successfully improved the importance function even in this case.

3.2. Example 2

As a second example, the $f(\theta)$ of Example 1 in Oh and Berger(1992) is considered. The figure of $f(\theta)$, given in Figure 11, shows that $f(\theta)$ is highly skewed in the direction of $(0, 0)$ but the mode and the inverse Hessian at the mode does not detect the skewness. Hence the initial importance function is a bad approximation to $f(\theta)$, as shown in Figure 12, and $\widehat{CV}^2(w)$ with this importance function is 1.062. The algorithm in this paper was applied to improve the initial importance function. After the first stage we obtained a little better approximation as shown in Figure 13 and $\widehat{CV}^2(w)$ was also improved to 0.62834. (Note that this value of $\widehat{CV}^2(w)$ is even about three times smaller than that obtained in the adaptive algorithm of Oh and Berger(1992) with one component.) After the second stage, the importance function became very close to $f(\theta)$ as shown in Figure 14. And the resulting $\widehat{CV}^2(w)$ was 0.3003, about one third of the initial $\widehat{CV}^2(w)$. The process stopped at the third stage since there was not significant

improvement in the importance function and the importance function from the second stage would be chosen as importance function for the actual importance sampling.

From Figures 11 to 14, we can see that the algorithm successfully improved the importance function even when $f(\theta)$ is highly skewed and the skewness is not detected initially.

4. Summary and discussions

The algorithm in this article suggested a way to handle complicated $f(\theta)$. It takes advantages of using a mixture importance function as described in Oh and Berger (1993) but also brings more flexibility by increasing the number of components if necessary. It efficiently takes care of (unimodal or multimodal) highly skewed $f(\theta)$ and $f(\theta)$ with undetected modes. For $f(\theta)$ with undetected modes, the curvature of $f(\theta)$ at the undetected modes should not be significantly different from the curvature of the nearest component because the scale of the nearest component is used to initialize that of the additional component. So when there are undetected modes in $f(\theta)$ some additional conditions are required for efficiency of the algorithm. Moreover, a caution should be given when there is an undetected extremely sharp peak, as mentioned in Section 2.1. However, undetected modes (especially extremely sharp peaks) is a problem in any numerical integration scheme, not a particular problem in the algorithm of this paper.

When one does not know whether curvatures of undetected modes satisfy the conditions described above, we suggest using the largest scale parameter of components in the previous stage for the initialization of the additional component because a larger scale is safer than a smaller scale in importance sampling. When in fact the true curvature of the undetected mode is much smaller than the largest scale parameter, the use of largest scale would slow down the convergence of the algorithm. But this is a price to be paid because of lack of information about $f(\theta)$.

A linear interpolation with respect to the mixing weights is suggested for a better performance. It requires some extra computations but would be compensated for by saving cost in the actual importance sampling, because of a better importance function.

Some modifications of the algorithm may be helpful. For instance, if $f(\theta)$ is skewed in two different directions then it would be more desirable to add two components at the same time rather than to add one at a time. But implementation of the above idea is difficult since it is very difficult to determine the direction of skewness in high dimensional space.

References

[1] Evans, M., and Gilula, Z., and Guttman, I. (1989). Latent Class Analysis of Two-way Contingency Tables by Bayesian Methods, *Biometrika*, Vol. 76, 557-563.

[2] Evans, M. (1991a). Chaining via Annealing, *Annals of Statistics*, Vol. 19, 382-393.

[3] Evans, M. (1991b). Adaptive Importance Sampling and Chaining, *American Mathematical Society Contemporary Mathematics*, eds. N. Flournoy and R. Tsutakawa.

[4] Everitt, B.S., and Hand, D.J. (1981). *Finite Mixture Distributions*, London: Chapman and Hall.

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

[6] Gelfand, A., Smith, A.F.M., and Lee, T.M. (1992). Bayesian Analysis of Constrained Parameter and Truncated Data Problems Using Gibbs Sampling, *Journal of the American Statistical Association*, Vol. 87, 523-532.

[7] Gelfand, A., Hills, S., Racine-Poon, A., and Smith, A.F.M. (1990). Illustration of Bayesian Inference in Normal Data Models Using Gibbs Sampling, *Journal of the American Statistical Association*, Vol. 85, 972-985.

[8] Geman, S., and Geman, D. (1984). Stochastic Relaxation, Gibbs Distributions and Bayesian Restoration of Images, *IEEE Transactionson Pattern Analysis and Machine Inteligence*, Vol. 6, 721-741.

[9] Geweke, J. (1988). Antithetic Acceleration of Monte Carlo Integration in Bayesian Inference, *Journal of Econometrics*, Vol. 38, 73--89.

[10] Geweke, J. (1989). Bayesian Inference in Econometric Models Using Monte Carlo Integration, *Econometrica*, Vol. 46, 1--20.

[11] Naylor, J.C., and Smith, A.F.M. (1982). Application of a Method for the Efficient Computation of Posterior Distributions, *Applied Statistics*, Vol. 31, 214--225.

[12] Naylor, J.C., and Smith, A.F.M. (1988). Econometric Illustrations of Novel Numerical Integration Strategies for Bayesian Inference, *Journal of Econometrics*, Vol. 38, 103-125.

[13] Oh, M.S., and Berger, J.O. (1992). Adaptive Importance Sampling in Monte Carlo Integration, *Journal of Statistical Computing and Simulation*, Vol. 41, 143-168.

[14] Oh, M.S., and Berger, J.O. (1993). Integration of Multimodal Functions by Monte Carlo Importance Function, *Journal of the American Statistical Association*, Vol. 88 450-456.

[15] Smith, A.F.M., Sken, A.M., Shaw, J.E.H., Naylor, J.C., and Dransfield, M. (1985). The Implementation of the Bayesian Paradigm, *Communications in Statistics - Theory and Method*, Vol. 14, 1079-1102.

[16] Tanner M., and Wong W. (1987). The Calculation of Posterior Distributions by Data Augmentation, *Journal of the American Statistical Association*, Vol. 82, 528-550.

[17]Titterlington, D.M., Smith, A.F.M., and Makov, U.E. (1985). *Statistical Analysis of Finite Mixture Distributions*, New York: Wiley.

[18] West, M. (1991). Bayesian Computation: Monte Carlo Density Estimation, *Tech. Report 90-A10,* Duke University, Institute of Statistics and Decision Sciences.

[19] van Dijk, H.K., and Kloek, T., and Boender, C.G.E. (1985). Posterior Moments Computed by Mixed Integration, *Journal of Econometrics*, Vol. 29, 3-18.

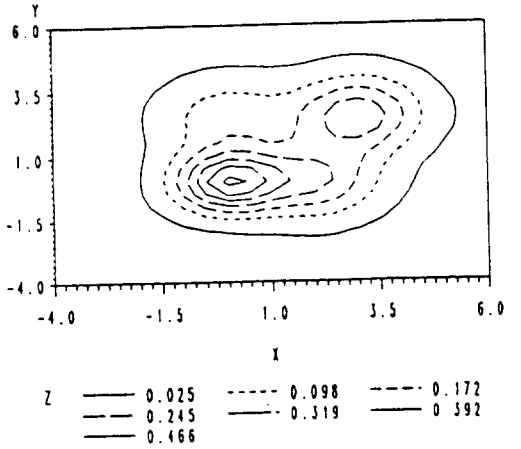

Fig. 1 : Contour Plot of $f(\theta)$ in Example 1

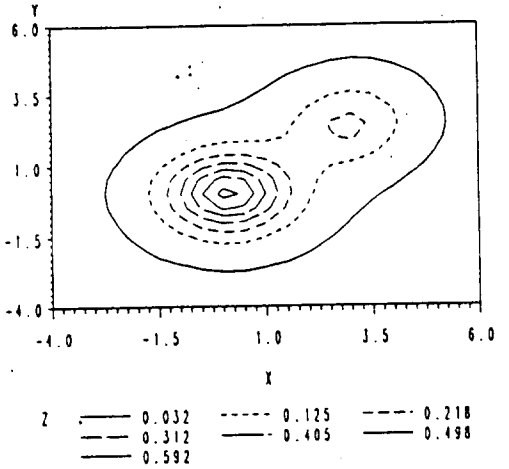

Fig. 2 : Initial Importance Function with Two Components

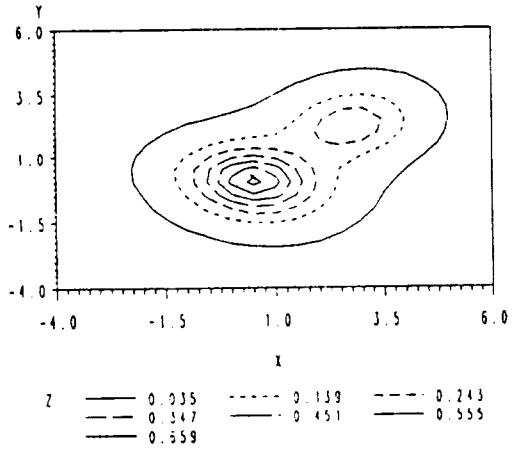

Fig. 3 : Importance Function From Stage 1

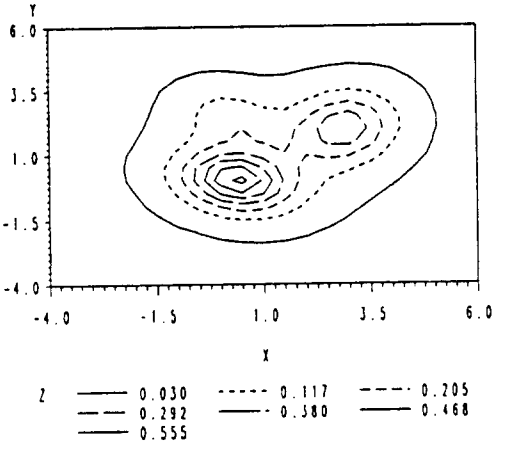

Fig. 4 : Importance Function From Stage 2

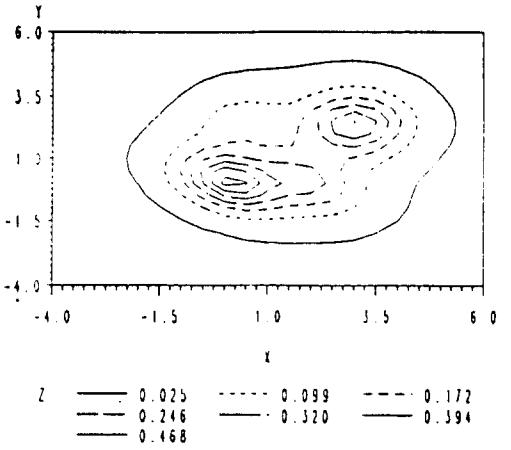

Fig. 5 : Importance Function From Stage 3

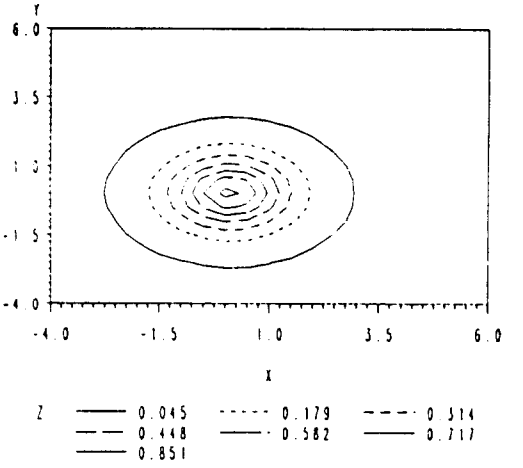

Fig. 6 : Initial Importance Function with One Components

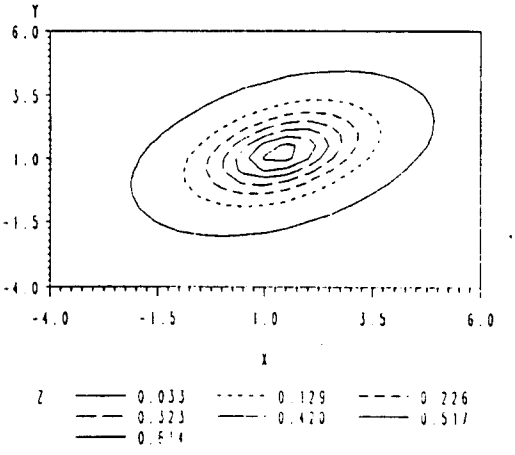

Fig. 7 : Importance Function From Stage 1

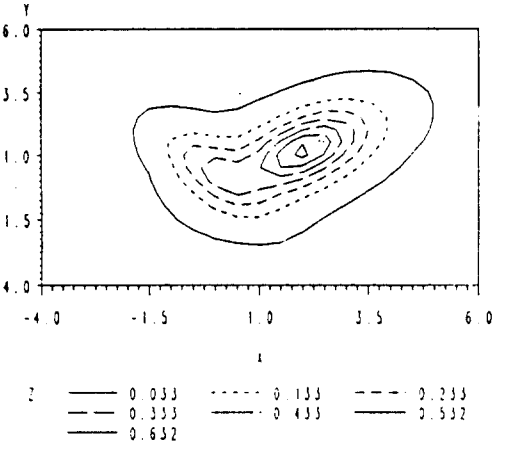

Fig. 8 : Importance Function From Stage 2

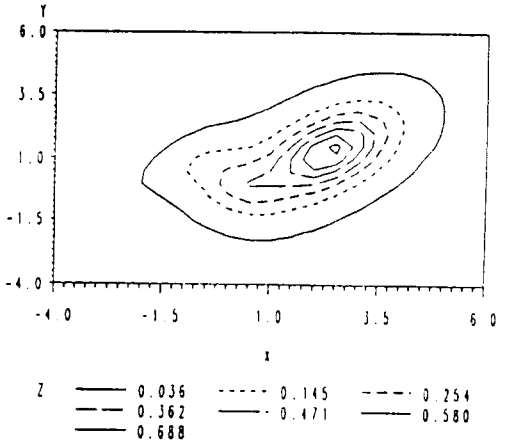

Fig. 9 : Importance Function From Stage 3

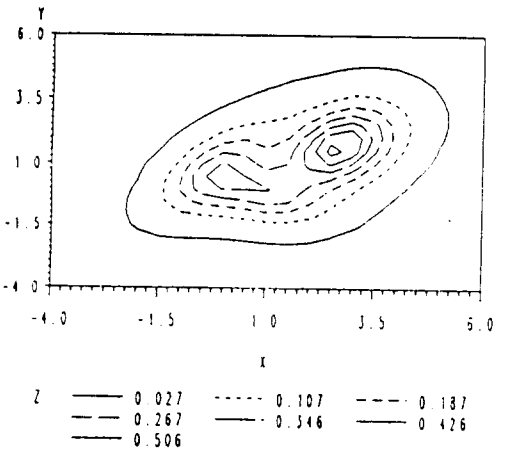

Fig. 10: Importance Function From Stage 4

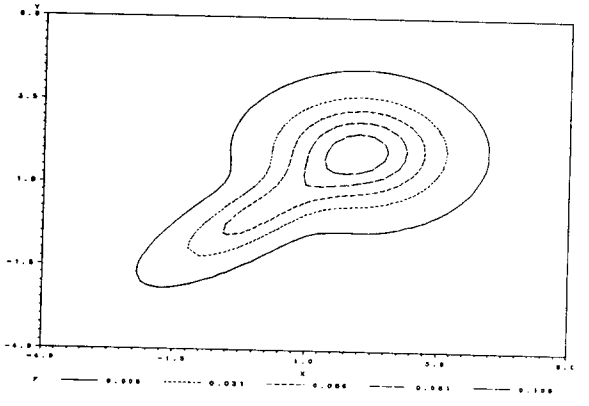

Fig. 11: Contour Plot of f in Example 2

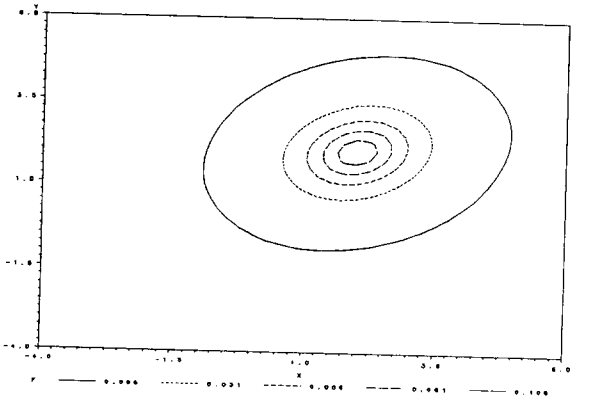

Fig. 12: Initial Importance Function

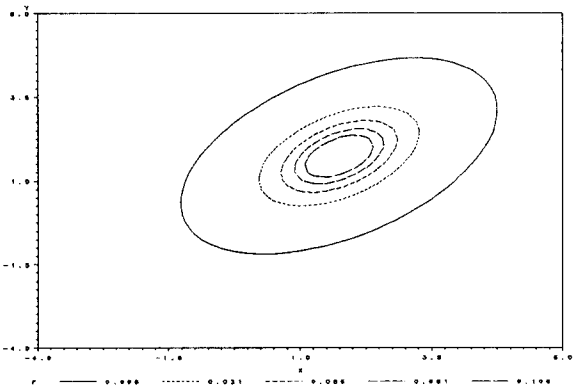

Fig. 13: Importance Function From Stage 1

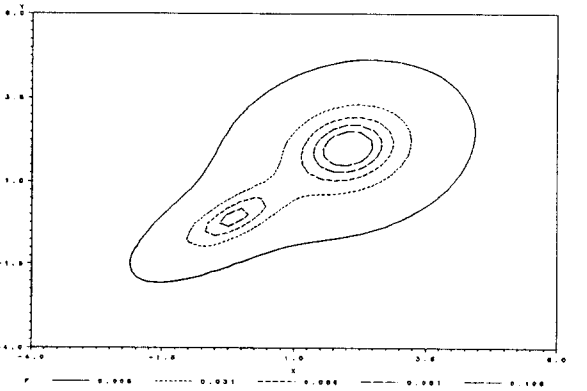

Fig. 14: Importance Function From Stage 2

다봉 함수의 다차원 적분을 위한 몬테카를로 기법의 개선[3)]

오만숙[4)]

요약

본 논문에서는 Oh and Berger (1993) 의 알고리즘을 확장하여, Oh and Berger 의 알고리즘으로 효율적 처리가 어려웠던 복잡한 모양을 가진 다차원 함수의 적분에 보다 더 일반적으로 적용될 수 있는 알고리즘을 제시한다. 예를 들면 다봉 함수이면서 동시에 기울어진 모양을 갖는 함수나 모든 극대점들이 다 파악되지 못한 경우 등이다. 제시된 알고리즘은 Oh and Berger 의 알고리즘을 단계적으로 수행해 가면서 각 단계 마다 새로운 부확률밀도 함수 (component density function) 를 합성 밀도 함수 (mixture importance function) 형태인 중요 함수 (importance function) 에 더해 감으로써 중요 함수를 적분 함수에 접근시킨다.

3) 본 논문은 1993년 학술재단 신진과제 연구비의 지원에 의한 것임.
4) (120-750) 서울시 서대문구 대현동 이화여자대학교 통계학과.