ECONOMETRIC INSTITUTE

EXPERIMENTS WITH SOME ALTERNATIVES FOR SIMPLE
IMPORTANCE SAMPLING IN
MONTE CARLO INTEGRATION

H.K. VAN DIJK AND T. KLOEK

REPORT 8326/E

Erasmus

ERASMUS UNIVERSITY ROTTERDAM - P.O. BOX 1738 - 3000 DR ROTTERDAM - THE NETHERLANDS

EXPERIMENTS WITH SOME ALTERNATIVES FOR SIMPLE IMPORTANCE SAMPLING IN MONTE CARLO INTEGRATION

by H.K. van Dijk and T. Kloek

SUMMARY

Some alternatives for simple importance sampling [compare Kloek and van Dijk (1978) and van Dijk and Kloek (1980)] are investigated for the computation of posterior moments and densities. An importance sampling method that is based on a mixture of a finite number of multivariate normal densities is compared with simple importance sampling and with a method that is based on a combination of Monte Carlo and classical numerical integration. These methods are intended to handle econometric applications where a simple importance function that is a reasonable approximation to the posterior density is difficult to find. For illustrative purposes use is made of a small econometric model. The results include bivariate marginal densities of the importance functions and the posterior plotted in three dimensional figures.

Contents

KEYWORDS: MONTE CARLO; NUMERICAL INTEGRATION; SIMULTANEOUS EQUATION MODEL; POSTERIOR MOMENTS; BIVARIATE POSTERIOR DENSITIES; IMPORTANCE SAMPLING; COMPOSITION METHOD; FINITE MIXTURES; MIXED INTEGRATION

May 1983, revised October 1983

1. INTRODUCTION

In two earlier papers [Kloek and van Dijk (1978) and van Dijk and Kloek (1980)], we applied Monte Carlo integration (in particular, importance sampling) for the purpose of finding posterior moments and densities for parameters of econometric simultaneous equation models. (A simultaneous equation model is nonlinear in the sense that the expected values of the endogenous variables are nonlinear functions of the parameters of interest.) We made use of the multivariate Student density as an importance function. In these papers we emphasized the use of importance functions that fit reasonably well to the posterior distribution. So there exists a problem when the posterior is skew. In van Dijk and Kloek (1983a) we addressed this problem, but in a rather ad hoc way. For that reason we have started to investigate alternative approaches. We shall use the term simple importance sampling (SIS) for the approach used in our 1978 and 1980 papers.

The present paper contains some experiments with two alternative approaches to SIS. First, we propose a flexible importance function which consists of a mixture of a finite number of multivariate normal densities. This importance function is intended for cases where the dominating skewness is in one or two directions. More details are given in Section 2. Second, we make use of an alternative to simple importance sampling, where the basic idea is to transform the s-dimensional space of parameters of interest θ into another s-dimensional space of vectors $(\eta' \ \rho)$ where ρ is a scalar and η an (s-1)-vector, and the prime denotes transposition. The transformed posterior density of $(\eta' \ \rho)$ is decomposed as a conditional density of ρ given η and a marginal density of η. For ρ we take $\pm$ d, where d is a measure of the distance between a point $\theta^{(i)}$, generated at random, and a central point (θ^0) such as the posterior mode or a preliminary estimate of the posterior mean. For η we take the direction $(\theta^{(i)} - \theta^0)/\rho^{(i)}$ with one coordinate deleted in order to avoid degeneracy. After having performed the transformation described we draw a vector $\eta^{(i)}$ and apply classical numerical integration with respect to ρ given $\eta^{(i)}$. So, this method amounts to a combination of classical numerical integration and Monte Carlo, which we call mixed integration. For a motivation and details of this approach we refer to van Dijk and Kloek (1983b). This method is intended to handle cases where the posterior is skew in several directions.

In Section 3 we discuss the results that have been obtained with some experiments using the alternative approaches. As an example we make use of the same econometric model as in Kloek and van Dijk (1978). For details with respect to the choice of the prior density we refer to van Dijk and Kloek (1983a). Our conclusions are given in Section 4.

2. A NORMAL-PIECEWISE-UNIFORM IMPORTANCE FUNCTION

In this section we describe a flexible class of importance functions, which may be useful in cases where the surface of the posterior density of a nonlinear model is skew predominantly in one or two directions. The present approach was inspired by a technique for generating univariate random variables, called composition [compare Rubinstein (1981, Chapter 3), and the references cited there]. In this technique a composite density function is defined as a convex linear combination of a number of more elementary density functions in the following way. Let $p_1, \ldots, p_h$ be a set of nonnegative constants that sum to unity and let $\theta_1, \ldots, \theta_h$ be a set of random variables with distribution functions $F_1(\theta_1), \ldots, F_h(\theta_h)$ and density functions $I_1(\theta_1), \ldots, I_h(\theta_h)$. Define the random variable θ as equal to θ_j with probability p_j for $j = 1, \ldots, h$. The composite distribution function of θ is given in the point θ^* as

$$F(\theta^*) = P[\theta \leq \theta^*] = \sum_{j=1}^{h} p_j P[\theta_j \leq \theta^*] = \sum_{j=1}^{h} p_j F_j(\theta^*) \tag{1}$$

and the composite density function is given as

$$I(\theta) = \sum_{j=1}^{h} p_j I_j(\theta) \tag{2}$$

where we deleted the asterisk for notational convenience. In certain univariate cases this approach is very efficient [see Atkinson and Pearce (1976)]. In the literature the density (2) is also known as a mixture of a finite number of density functions with mixing parameters $p_1, \ldots, p_h$ [see Everitt and Hand (1981)].

One may extend the use of the composite density (2) to cases where θ is a vector of parameters. We have experimented with a composite density as an importance function where the elementary densities are multivariate normals. However, there are at least two problems with this approach. First, the number of parameters of a composite importance function $I(\theta)$ is very large. The estimation of the parameters of the multivariate normal densities $I_1(\theta)$, ..., $I_h(\theta)$ and the estimation of the mixing parameters p_1, ..., p_h is a far from trivial matter. Second, equation (2) implies that in order to evaluate $I(\theta)$ in a point $\theta^{(i)}$, every elementary function $I_j(\theta)$ has to be evaluated in $\theta^{(i)}$. This may be computationally expensive. So, there exists a need for simplification of the estimation of the large number of parameters without affecting the flexibility of the composite density more than marginally. Below we present a first attempt in this direction.

First, we note that the index variable j in equation (2) can be interpreted as a discrete random variable with probability mass function p_j and the elementary density functions can be interpreted as conditional density functions $I(\theta|j)$ with the index j as a conditioning variable. One may generalize this by assuming more general distributions of the conditioning variable. In the continuous case, for instance, one has

$$I(\theta) = \int I^C(\theta|z)I^M(z)dz \tag{3}$$

The composite density (3) is an infinite mixture of conditional densities $I^C(\theta|z)$ given the values of the continuous random variable z. The marginal density $I^M(z)$ of z, which is an auxiliary random variable, is called the mixing density. The Student density is an example of an infinite mixture of normal densities where the gamma-2 is the mixing density.

In our example we noted from a preliminary diagnostic analysis that the posterior density $p(\theta)$ is, roughly speaking, very skew in the direction of one component of θ. We make use of this information in the following way. Let θ be partitioned as

$$\theta = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} \tag{4}$$

where θ_1 contains s-1 components and θ_2 is a scalar and the skewness is supposed to be predominant in the direction of θ_2. The importance function $I(\theta)$ is factorized as the product of a conditional multivariate normal (CMN) density of θ_1 conditional upon θ_2 and a marginal piecewise-uniform (MPU) density of θ_2. The MPU distribution of θ_2 is defined as

$$F(\theta_2) = \begin{cases} 0 & \text{if } \quad \theta_2 \leq a_0 \\ P_{j-1} + \dfrac{p_j}{a_j - a_{j-1}}(\theta_2 - a_{j-1}) & \text{if } a_{j-1} \leq \theta_2 \leq a_j \\ & (j = 1, \ldots, h) \\ 1 & \text{if } a_h \leq \theta_2 \end{cases} \tag{5}$$

where

$$P_0 = 0,\ P_j = P_{j-1} + p_j\ (j = 1, \ldots, h),\ \sum_{j=1}^{h} p_j = P_h = 1$$

and the MPU density function of θ_2 is defined as

$$I^M(\theta_2) = \begin{cases} 0 & \text{if} \quad \theta_2 < a_0 \\ \dfrac{p_j}{a_j - a_{j-1}} & \text{if } a_{j-1} < \theta_2 < a_j \\ & (j = 1, \ldots, h) \\ 0 & \text{if } a_h < \theta_2 \end{cases} \tag{6}$$

Random drawings θ_2 can easily be generated by an inversion method because the distribution of the random variable $F(\theta_2)$ is uniform on [0, 1]. The CMN density of θ_1 conditional upon θ_2 is (for j = 1, ..., h) given as

$$\begin{aligned} I_j^C(\theta_1|\theta_2) &= (2\pi)^{-\frac{1}{2}(s-1)}|V_j|^{-\frac{1}{2}} \\ &\times \exp[-\tfrac{1}{2}(\theta_1 - \mu_j)'V_j^{-1}(\theta_1 - \mu_j)] \quad \text{if } a_{j-1} < \theta_2 < a_j \end{aligned} \tag{7}$$

where μ_j and V_j are the well known parameters of a multivariate normal density. Therefore the CMNMPU importance function can be written as

$$\begin{aligned} I(\theta) &= I_j^C(\theta_1|\theta_2)I^M(\theta_2) \\ &= I_j^C(\theta_1|\theta_2)\frac{p_j}{a_j - a_{j-1}} \quad \text{if } a_{j-1} < \theta_2 < a_j \end{aligned} \tag{8}$$

The parameters of (8) are μ_j, V_j, p_j, (j = 1, ..., h) and the interval bounds a_0, a_1, ..., a_h.

One may compare the case of the CMNMPU importance function (8) of the s-vector θ with the case where the composite importance function of θ, given in equation (2), contains s-dimensional conditional multivariate normal densities as elementary functions and the conditioning index variable is an auxiliary random variable. In contrast, in the CMNMPU approach we have partitioned θ [see (4)] and our conditional densities are (s-1)-dimensional rather than s-dimensional. As a result we have a composition (or mixture) of a finite number of conditional multivariate normal densities of the (s-1)-vector θ_1.

conditional upon the scalar θ_2. The marginal piecewise-uniform density of the continuous random variable θ_2 plays the role of the mixing density.

The CMNMPU approach has three advantages.

First, it simplifies the evaluation of the importance function $I(\theta)$. In a particular point $\theta^{(i)}$ one has to evaluate only one elementary normal density $I_j^C(\theta_1^{(i)}|\theta_2^{(i)})$ for a given value $\theta_2^{(i)}$ [compare equation (8)].

Second, the number of parameters of the CMNMPU density is smaller than the number of parameters of a composition of s-dimensional multivariate normal densities. This may simplify the estimation of the parameters, see below. However, the number of parameters of the CMNMPU density is larger than the number of parameters of a multivariate Student density. This gives the CMNMPU density flexibility. In our example of a three dimensional posterior density we make use of a CMNMPU density with $h = 20$ conditional bivariate normal densities and a total of 138 unrestricted parameters. [20×5 parameters (μ_j, V_j) of the bivariate normals; 19 unrestricted bounds a_j ($j = 1, \ldots, h-1$) since the end points a_0 and a_h are determined a priori; and 19 unrestricted parameters p_j, $j = 1, \ldots, h-1$]. More generally we make use of $h(s - 1 + (1/2)s(s-1)) + 2h - 2$ parameters if the model contains s parameters of interest and we use h elementary densities. By contrast, a composition of s-dimensional normal densities contains an additional $h(s+1)$ parameters.

Third, there are relatively simple algorithms to estimate the parameters of the CMNMPU importance function. This advantage is the most important one. One of these algorithms is described in the following four steps.

Step 1. Generate a random sample for θ of size N ($N = 2000$, say), by making use of the simple importance sampling method (SIS) described in van Dijk and Kloek (1980). Estimates of the parameters of the importance function of SIS have been discussed in Section 4 of that paper;

Step 2. Arrange the drawings of θ_2 in ascending order and divide them into h ($h = 20$, say) groups. Then the interval bounds a_j, $j = 1, \ldots, h-1$ are given by the random drawings $\theta_2^{(100)}$, $\theta_2^{(200)}$, ..., $\theta_2^{(1900)}$. The bounds a_0 and a_h are given <u>a priori</u> (or by a preliminary diagnostic analysis);

Step 3. Compute posterior first and second order moments of θ_1, by a standard Monte Carlo method, for the twenty regions constructed in step 2. Take these moments as estimates for μ_j, V_j, $j = 1, \ldots, 20$;

Step 4. Estimate p_j, $j = 1, \ldots, 20$, by a linear regression of the posterior density $p(\theta)$ on the importance function given in the right hand side of (8). The regression coefficients are

$$\hat{p}_j = \frac{\sum_{i=100(j-1)+1}^{100j} \frac{1}{a_j - a_{j-1}} I_j^C(\theta_1^{(i)}|\theta_2^{(i)}) p(\theta^{(i)})}{\sum_{i=100(j-1)+1}^{100j} [\frac{1}{a_j - a_{j-1}} I_j^C(\theta_1^{(i)}|\theta_2^{(i)}]^2} \tag{9}$$

Note that the summation is on points satisfying $a_{j-1} < \theta_2^{(i)} < a_j$.

One may perform a second round of parameter estimation by generating a sample of 2000 random drawings from the CMNMPU density (8) and reiterating steps 2 to 4. Since our starting point was a composition approach we shall call the approach consisting of steps 1-4 COM1 and the approach where a second round of parameter estimation has been performed COM2. In the next section we report some results using COM1 and COM2.

Note that as an alternative to Step 4 one may take the integral of the posterior density on the region $-\infty < \theta_1 < \infty$ and $a_{j-1} < \theta_2 < a_j$, and use this integral as estimator for the parameter p_j. Estimates of these integrals for j = 1, ..., h can be computed by the same standard Monte Carlo method as mentioned in Step 3.

Finally, we emphasize that our approach of estimating parameters of a multivariate composite density is only a first step. More research is needed in this area.

3. SOME EXPERIMENTAL RESULTS

In this section we discuss the results of some experiments with alternative Monte Carlo methods using a simple example. We take as an example the three dimensional marginal posterior density of the structural parameters β_1, β_2 and γ_2 of the Johnston model [see Kloek and van Dijk (1978) and van Dijk and Kloek (1983a)]. The prior on all structural parameters is the same as described in the latter reference. In this particular case we have taken a uniform prior on the interval (-2, +2) for the three parameters (β_1, β_2, γ_2). Further, we note that the marginal posterior density of (β_1, β_2, γ_2) is in our case of two stochastic equations equal to the concentrated likelihood function apart from a constant factor. We shall consider three methods: simple importance sampling (SIS); simplified composition in one and two rounds (COM1 and COM2), compare Section 2; and mixed integration (MIN).

Next, we describe briefly some problems that had to be solved in the design of the experiments. First, there exists the problem of the comparability of the Monte Carlo rounds for the different methods. We make use of the same starting value of the random number generator for all three methods. But the methods differ with respect to the importance function and as a consequence with respect to the way the random numbers are generated. Therefore we opted for the (crude) criterion of approximately equal CPU-time. All results of Tables 2 and 3 were taken after approximately 40 seconds CPU-time on a DEC 2060 computer. The computer programs were executed at different times of the day and night in order to verify the sensitivity with respect to the workload of the computing system. The results were only marginally affected. The 40 CPU-seconds gave 10,000 accepted random drawings for SIS and COM1 and COM2. The parameters of the CMNMPU-importance function were estimated with a sample of 2000 random drawings. The program for MIN was stopped after the 40 CPU-seconds, mentioned above, had been reached.

Second, we performed a preliminary diagnostic analysis. This analysis indicated that in our case several prior bounds could be reduced in absolute value without affecting the posterior results substantially. So we changed the lower bound of -2 for the three parameters (β_1, β_2, γ_2) into the vector (-2.0, -1.7, -.4) and the upper bound of +2 for the three parameters into (.8, .25, 1.0). In this way we reduce the numerical integration on the region where the posterior density is almost zero. This is an advantage in particular for MIN where one-dimensional integrals are computed on the intervals, bounded by the upper and lower bounds, mentioned above. Further, the diagnostic results indicated that the posterior is very skew in the direction of the parameter β_1. So β_1 was chosen as the parameter θ_2 of equation (4). We note that there is a minor notational problem with respect to β_1. In reporting the results we have taken the mixing parameter β_1 as the first element of the vector $\theta' := (\beta_1, \beta_2, \gamma_2)$, while in the theoretical discussion of Section 2 the last element of θ, denoted by θ_2, was taken as the mixing parameter.

Third, we had to specify the values of the parameters of the importance functions in the different Monte Carlo methods. As mentioned before we make use of a multivariate Student density as importance function in SIS and in the first step of COM1 and COM2. Further, we make use of a multivariate normal importance function in MIN. The location parameters (μ_1, μ_2, μ_3) correspond with the parameters of interest (β_1, β_2, γ_2). We shall consider two cases of parameter estimates. The degrees of freedom parameter of the Student function is fixed at unity in both cases. The location parameters μ and the covariance matrix V are different in the two cases. Case I consists of taking the posterior mode for μ and minus the inverse of the Hessian of the log posterior for V. We call this case the <u>local</u> approximation case. Case II consists of taking a preliminary estimate of the posterior mean for μ and a preliminary estimate of the posterior covariance matrix as estimate for V. This we call the <u>global</u> approximation case.

Fourth, there exists a problem with respect to the choice of the parameter estimates of μ and V in the global case. One may apply the different Monte Carlo methods in a two-stage approach in the sense that the posterior moments obtained in the first (local) round of each different method are used as parameter estimates of the importance functions in the second (global) round. As a consequence a poor approximation of the posterior moments in the first round of a particular approach influences the posterior results in the second round. In order to avoid different effects of large sampling errors in different Monte Carlo methods, we decided to take the same set of posterior estimates in the second round for all alternative Monte Carlo methods. Further, we decided to take a large sample of 100,000 Monte Carlo drawings, using the COM2 approach, in the first round. This is a very large sample and we performed a sensitivity analysis with respect to the sample size in the following sense. The posterior first and second order moments from the COM2 (local) approach, using 10,000 random drawings (instead of 100,000 drawings) were taken as parameter estimates of μ and V for all alternative methods in the global case. The results of Tables 2 and 3 were not very much affected. We emphasize that it is attractive to have a rather small sample of random drawings (say N = 1000) in the first round of Monte Carlo. Further research is needed to decide upon the trade-off between sample size and desired accuracy of the preliminary estimates of μ and V in the first round.

Fifth, Table 1 gives the two cases of estimates of importance function parameters. These estimates indicate three major differences between the local and global case. First, the modes and means of β_1 and β_2 differ considerably, which indicates skewness. Second, the posterior standard deviations of β_1 and β_2 are for the global case roughly eight times as large as for the local case. This indicates leptokurtosis. Third, the correlation between β_2 and γ_2 is positive in the global case but negative in the local case. This is an indication that the contours of the posterior density are not concentric ellipsoids as is the case in linear models. This concludes our discussion of the main points of the experimental design.

The results on the numerical errors of the posterior mean estimates of β_1, β_2 and γ_2 are presented in Tables 2 and 3 using the alternative Monte Carlo methods. We take two measures of numerical error. First, we take the ratio (× 100) of the standard deviation of the Monte Carlo estimate of the posterior mean and the posterior standard deviation. This relative numerical error is chosen because we are more interested to estimate a posterior mean accurately if the posterior variance is small, than if it is large. Second, a quadratic loss function (× 100) is evaluated around the large sample estimate of the posterior mean reported in Table 1, with the inverse of the posterior covariance matrix as the matrix of weights. This function has been tentatively chosen as a summary statistic for high dimensional cases.

TABLE 1

ESTIMATES OF IMPORTANCE FUNCTION PARAMETERS

Means μ_1	μ_2	μ_3	
Posterior mode (I)	.46	.09	.36
Posterior* mean (II)	-.60	-.31	.31
Standard deviations	v_1	v_2	v_3
Local approximation in mode (I)	.10	.04	.11
Posterior standard deviations (II)	.79	.33	.15
Correlations	r_{12}	r_{13}	r_{23}
Local approximation in mode (I)	.88	.17	-.16
Posterior correlations (II)	.92	.19	.33

* These posterior moments have an absolute numerical error which is less than .005 at the five per cent significance level.

TABLE 2

RELATIVE NUMERICAL ERRORS OF POSTERIOR MEAN ESTIMATES

	SIS	MIN	COM1	COM2	Best method
	Case I (local approximation)				
β_1	9.92	5.51	1.95	1.52	COM2
β_2	10.02	6.22	2.19	1.88	COM2
γ_2	5.56	4.47	2.96	2.30	COM2
	Case II (global approximation)				
β_1	2.88	2.05	1.63	1.62	COM2
β_2	2.64	1.90	1.75	1.69	COM2
γ_2	2.12	1.71	1.55	1.57	COM1

TABLE 3

SQUARED ERROR LOSSES OF POSTERIOR MEAN ESTIMATES*

	SIS	MIN	COM1	COM2	Best method
Case I	22.23	8.62	.48	1.27	COM1
Case II	.12	.62	.03	.01	COM2

* The squared error loss function $L(\bar{\theta}, E(\theta)) = (\bar{\theta} - E(\theta))'W(\bar{\theta} - E(\theta))$, where $\bar{\theta}$ is the Monte Carlo estimate of $E(\theta)$ and W is the inverse of the posterior covariance matrix.

We make the following remarks on the numerical results in Tables 2 and 3.

(1) According to the results of the relative numerical error COM2 performs best with one exception, but the differences with COM1 are very small. The results for MIN are between those for COM and SIS for all cases. Note that the results of the squared error loss functions are in some cases different from the results of the relative error criterion. For instance, in Table 3 MIN does poorly in case II. Since the summary statistics give sometimes conflicting evidence we shall investigate the approximation of the marginal posterior densities in some cases.

(2) The local approximations of μ and V are poor starting values for SIS and MIN. There is a substantial gain in computational efficiency for these methods when the parameter estimates of the global case are taken as values of the importance function parameters.

(3) The results of the different methods are sensitive for different reasons. With respect to SIS we note that the results are sensitive for the prior bound of $\beta_1 = -2.0$. When the value of this bound is relaxed SIS performs worse. That is, SIS is sensitive to the degree of skewness of the problem. We shall comment on this below. The computational efficiency of MIN is sensitive for the accuracy level of the one-dimensional classical numerical integration. We took an iterative 16-point Gauss-Legendre formula for the numerical integration with an accuracy level of three digits, because the Monte Carlo estimates of the posterior means have usually no more than a two-digit accuracy. Further, it has been mentioned before (see the second problem of the design of the experiments) that it is an advantage for MIN to change the upper and lower bounds of the region of integration in such a way that one-dimensional integration on intervals where the posterior density is almost zero can be avoided. The results of COM1 and COM2 are, of course, dependent upon the preliminary diagnostic analysis that indicated that in our case the skewness was in the direction of β_1. Here MIN appears rather robust since it is not dependent on a priori or diagnostic knowledge with respect to the direction of skewness (compare COM1 and COM2) and it is not dependent on the degree of skewness (compare SIS). Only a few hundred random drawings were sufficient to indicate the direction(s) of skewness.

Next, univariate and bivariate marginal posterior densities and univariate and bivariate marginal importance functions are shown in Figures 1 to 6 for simple importance sampling based on the local and global cases and for the two-step composition method for the global case. The marginal

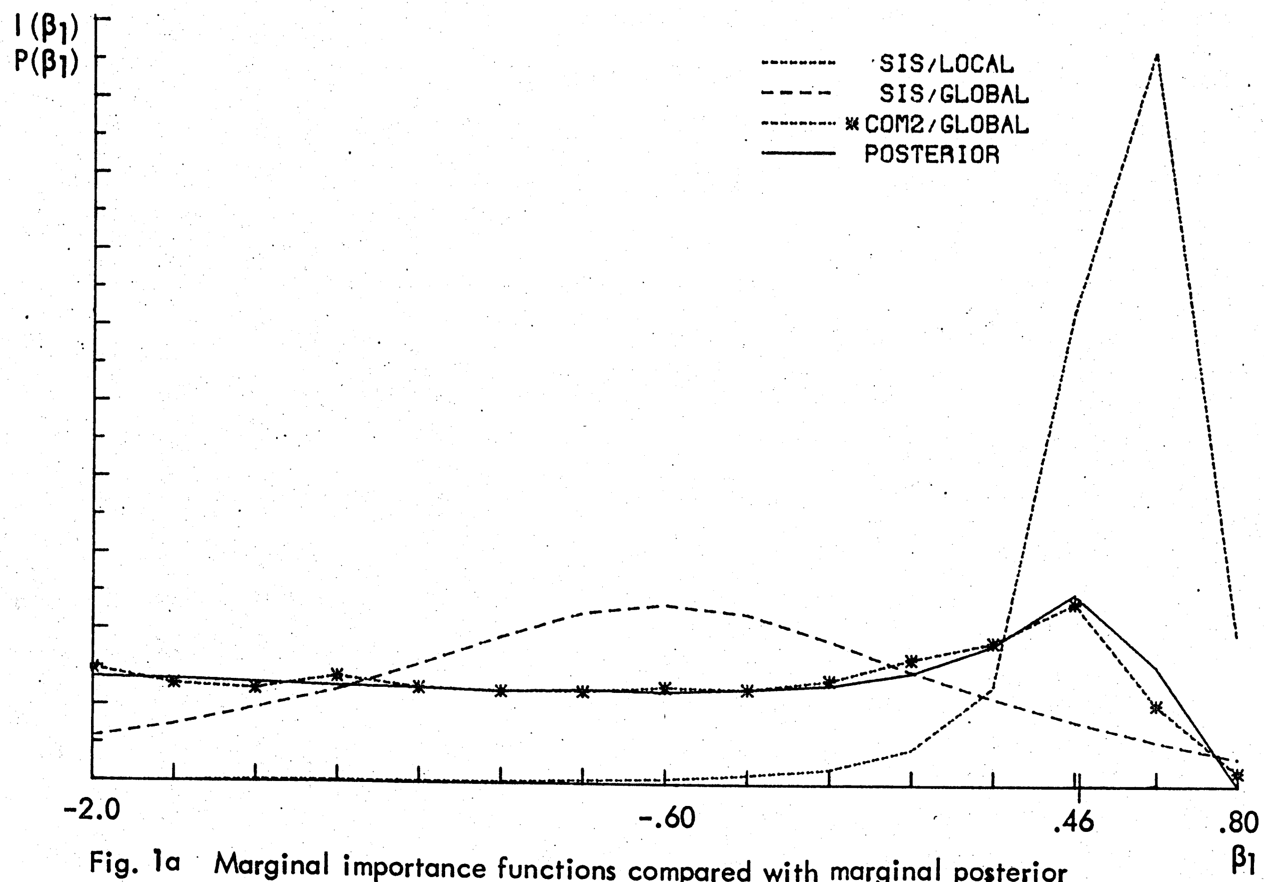


Fig. 1a Marginal importance functions compared with marginal posterior density of β_1

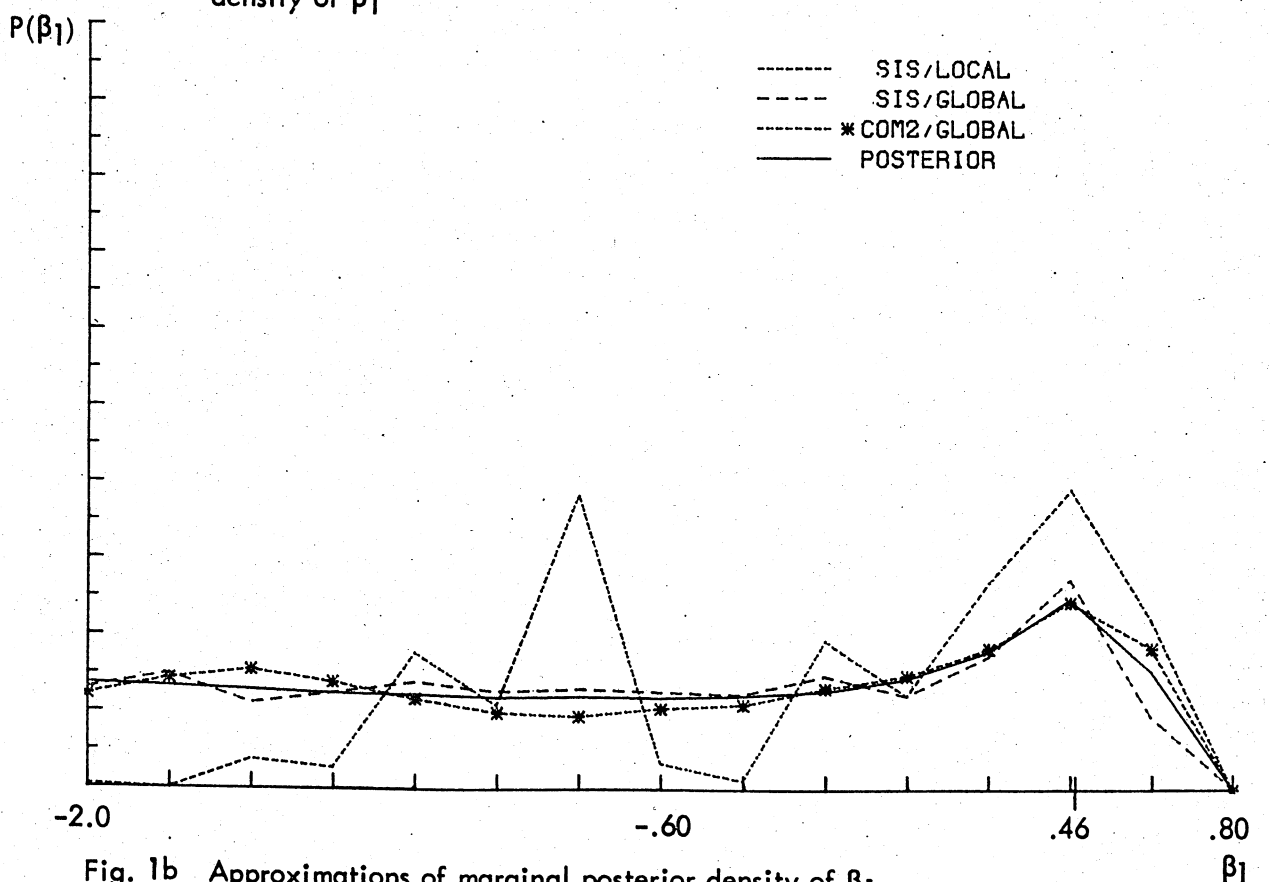


Fig. 1b Approximations of marginal posterior density of β_1

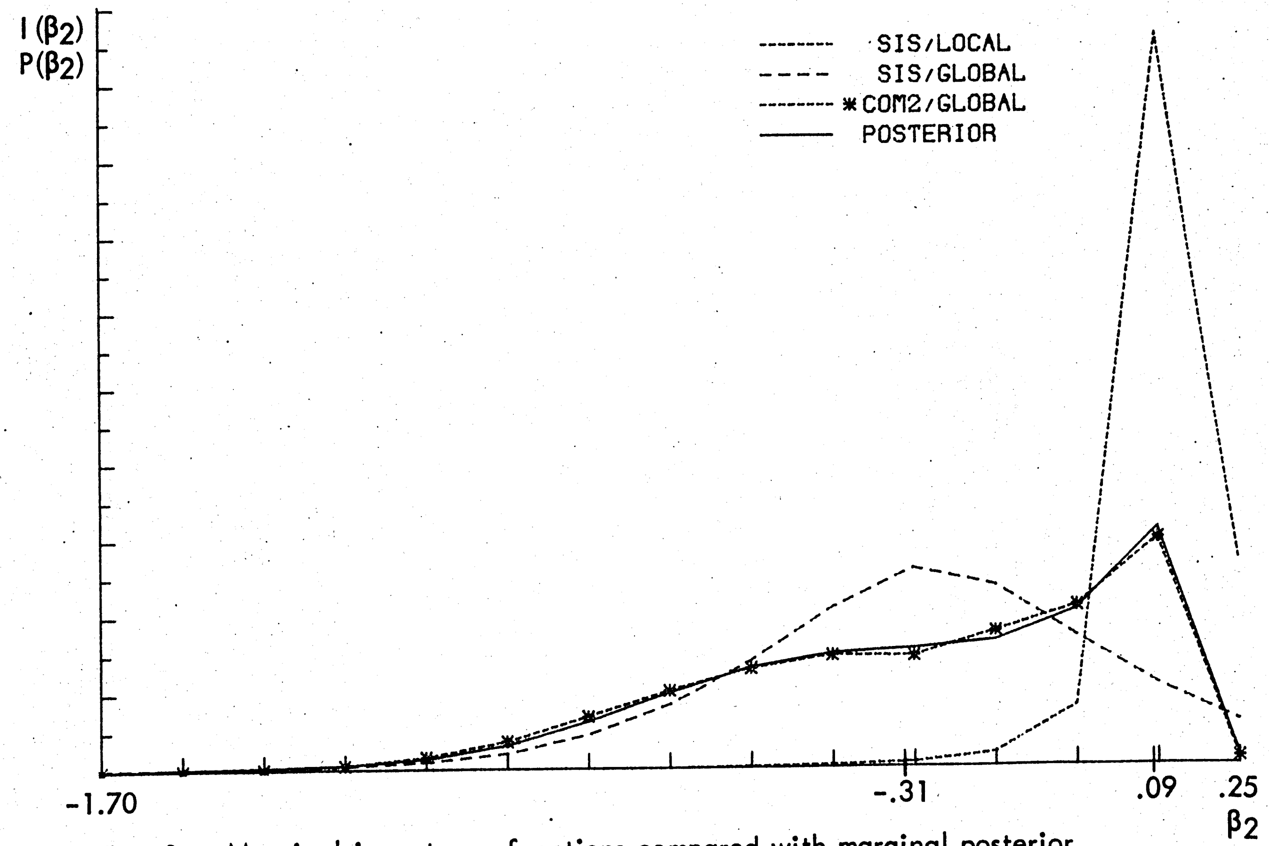


Fig. 2a Marginal importance functions compared with marginal posterior density of β_2

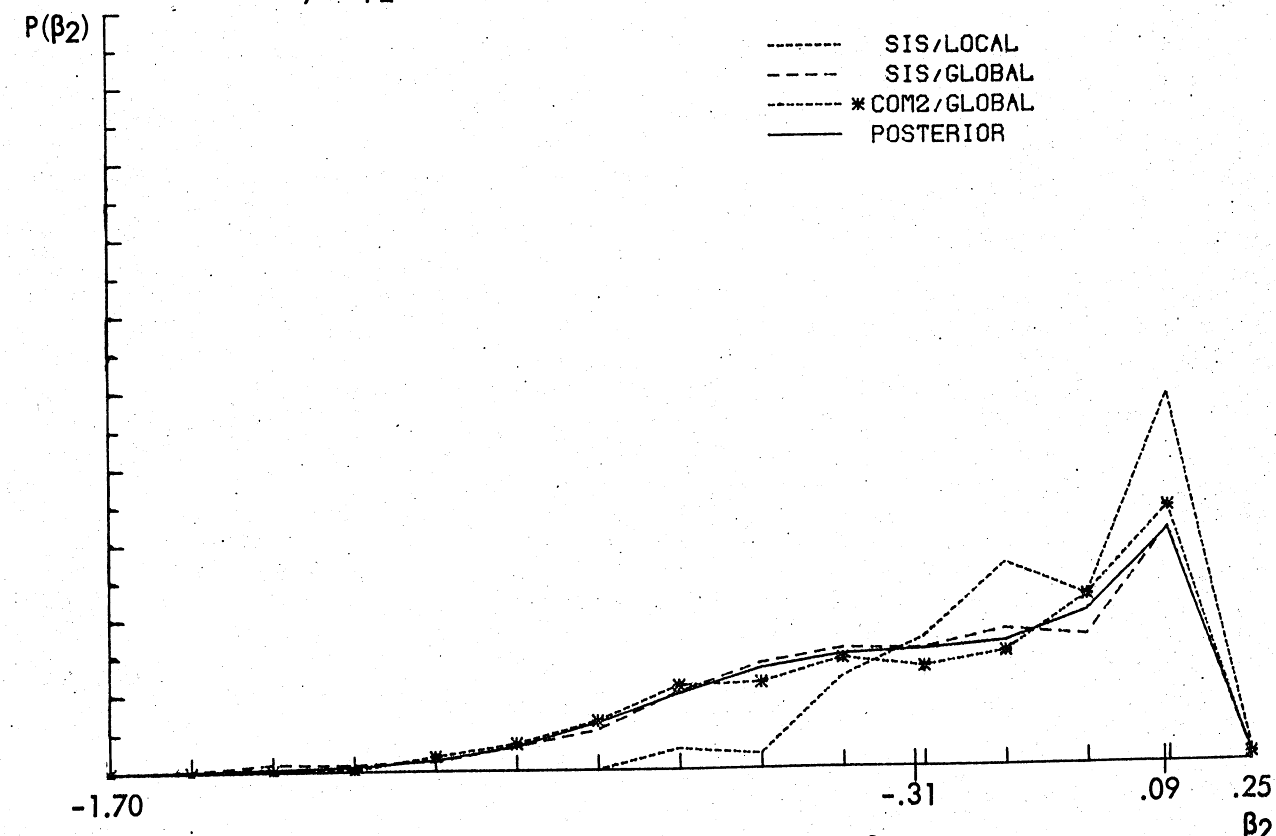


Fig. 2b Approximations of marginal posterior density of β_2

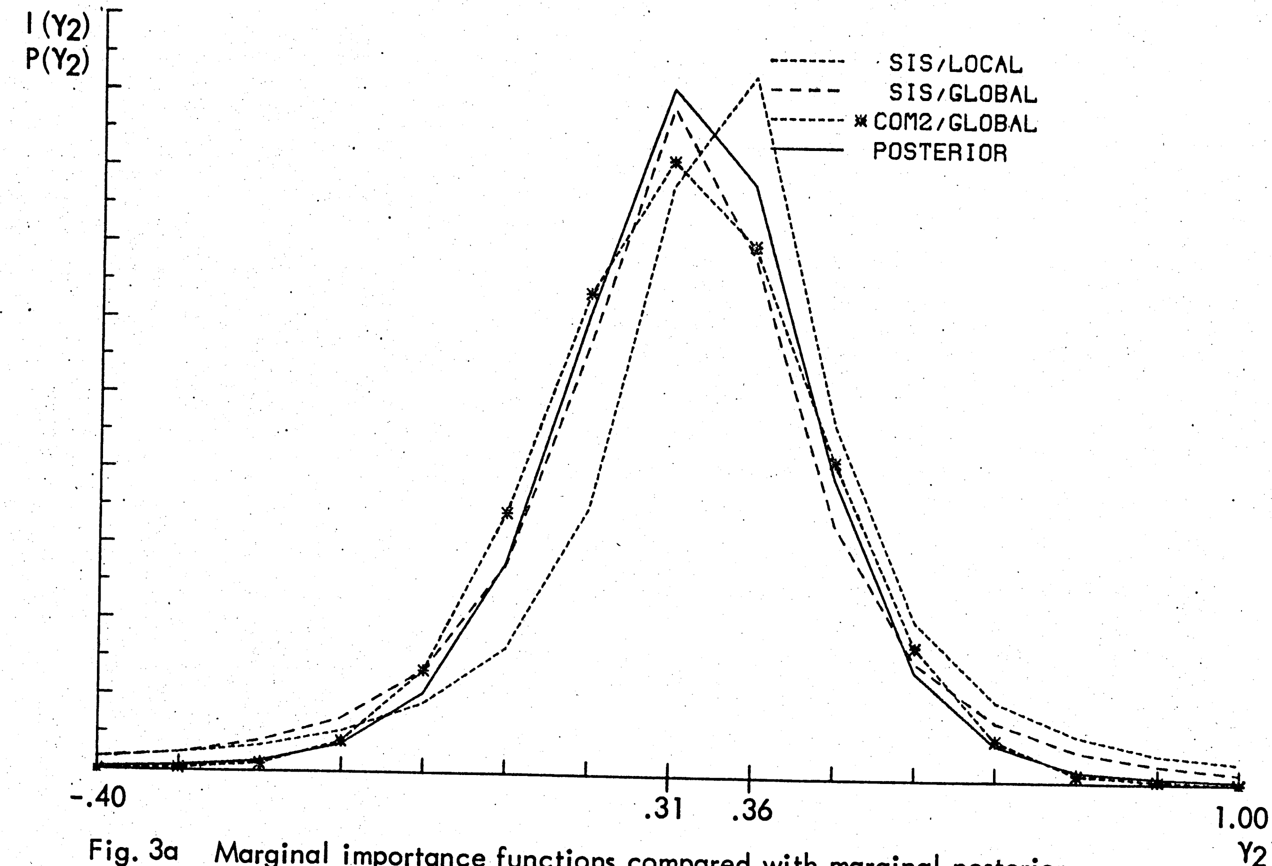


Fig. 3a Marginal importance functions compared with marginal posterior density of Y_2

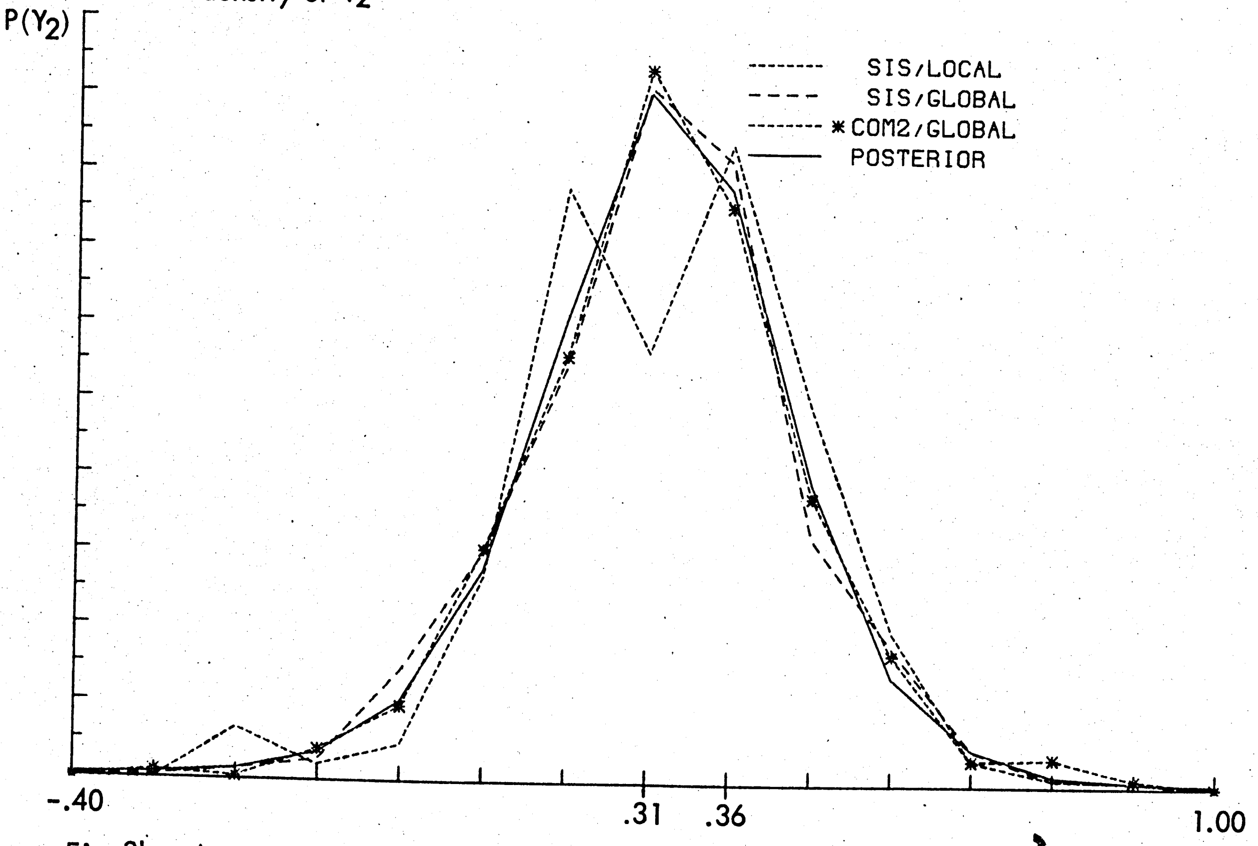
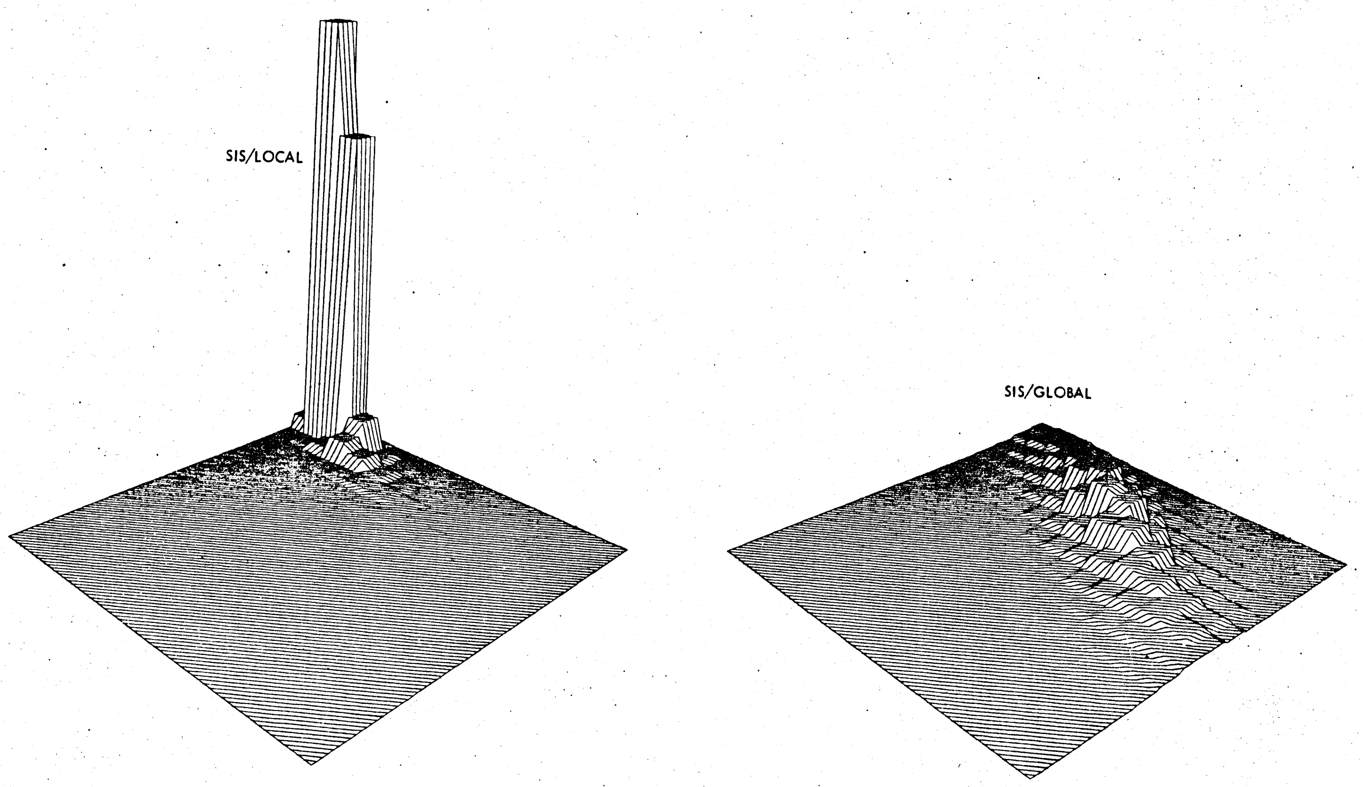


Fig. 3b Approximations of marginal posterior density of Y_2



Mode $(\beta_1, \beta_2) = (.46, .09)$

Mean $(\beta_1, \beta_2) = (-.60, -.31)$

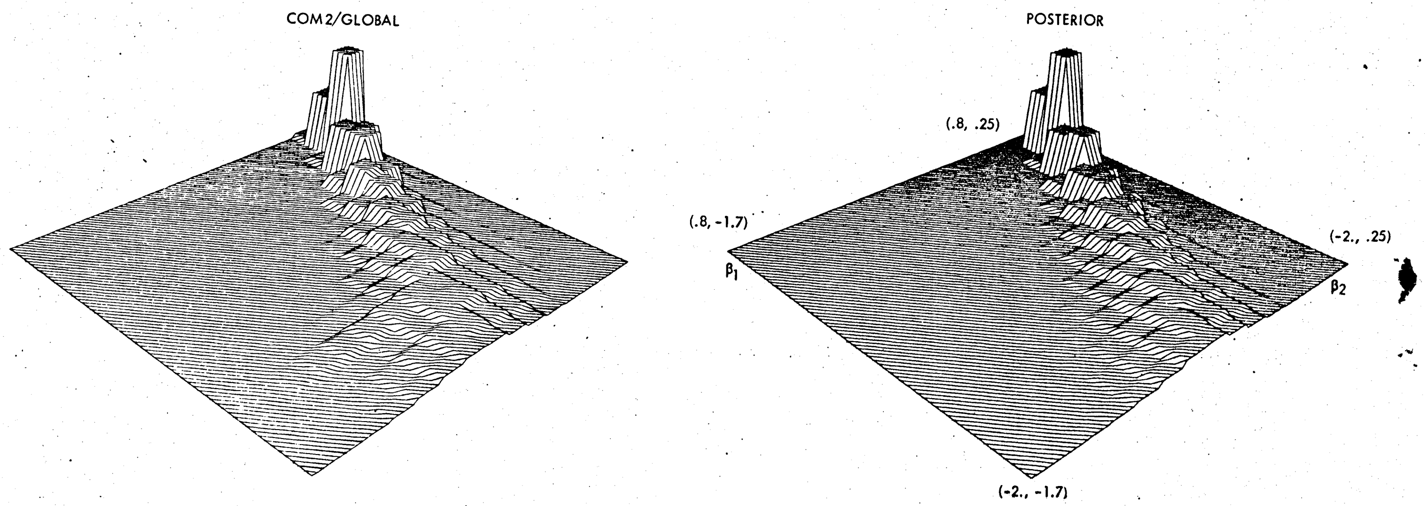
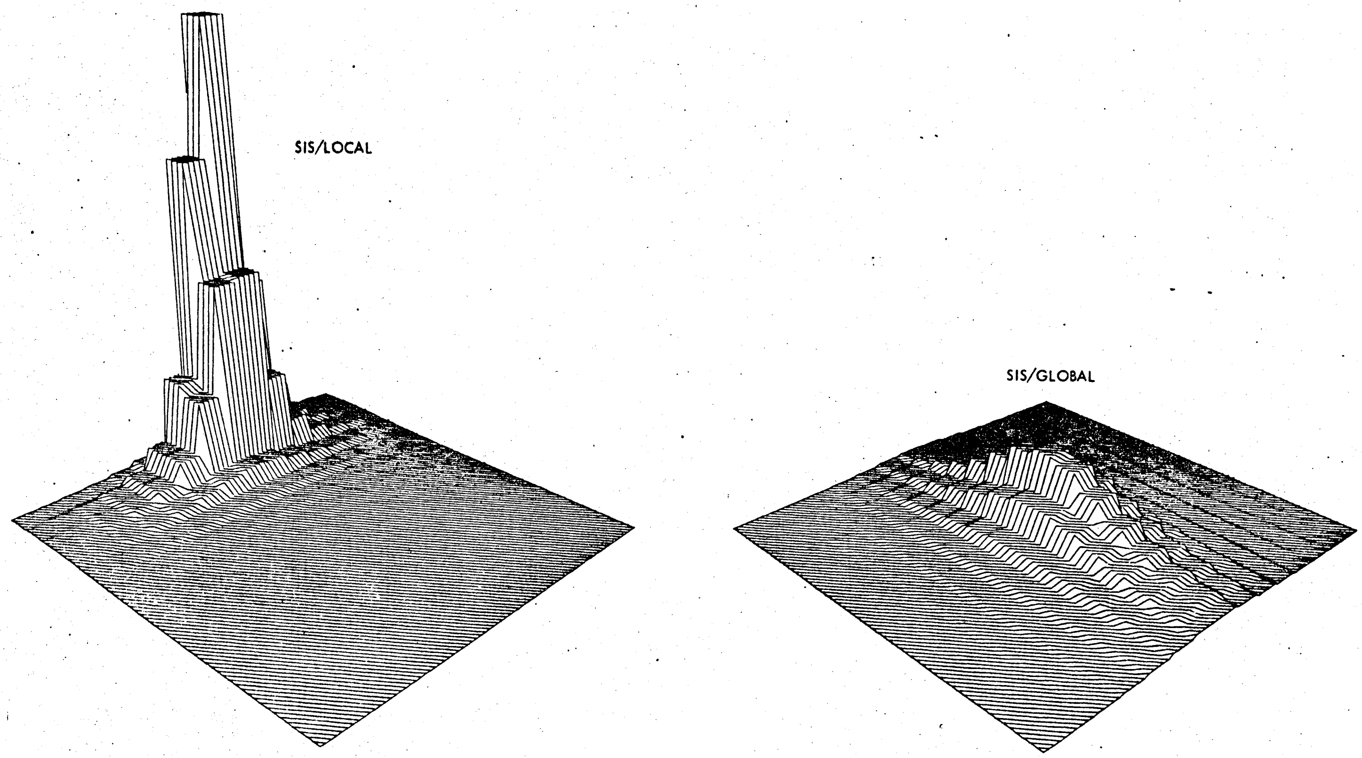


Fig. 4 Bivariate marginal importance functions and bivariate marginal posterior density of (β_1, β_2)



Mode $(\beta_1, \gamma_2) = (.46, .36)$

Mean $(\beta_1, \gamma_2) = (-.60, .31)$

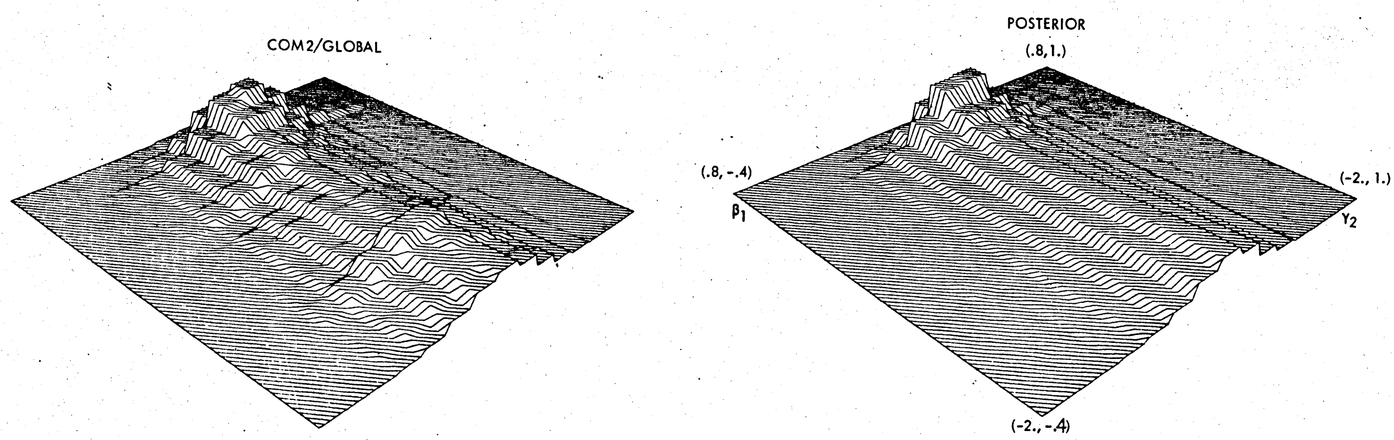
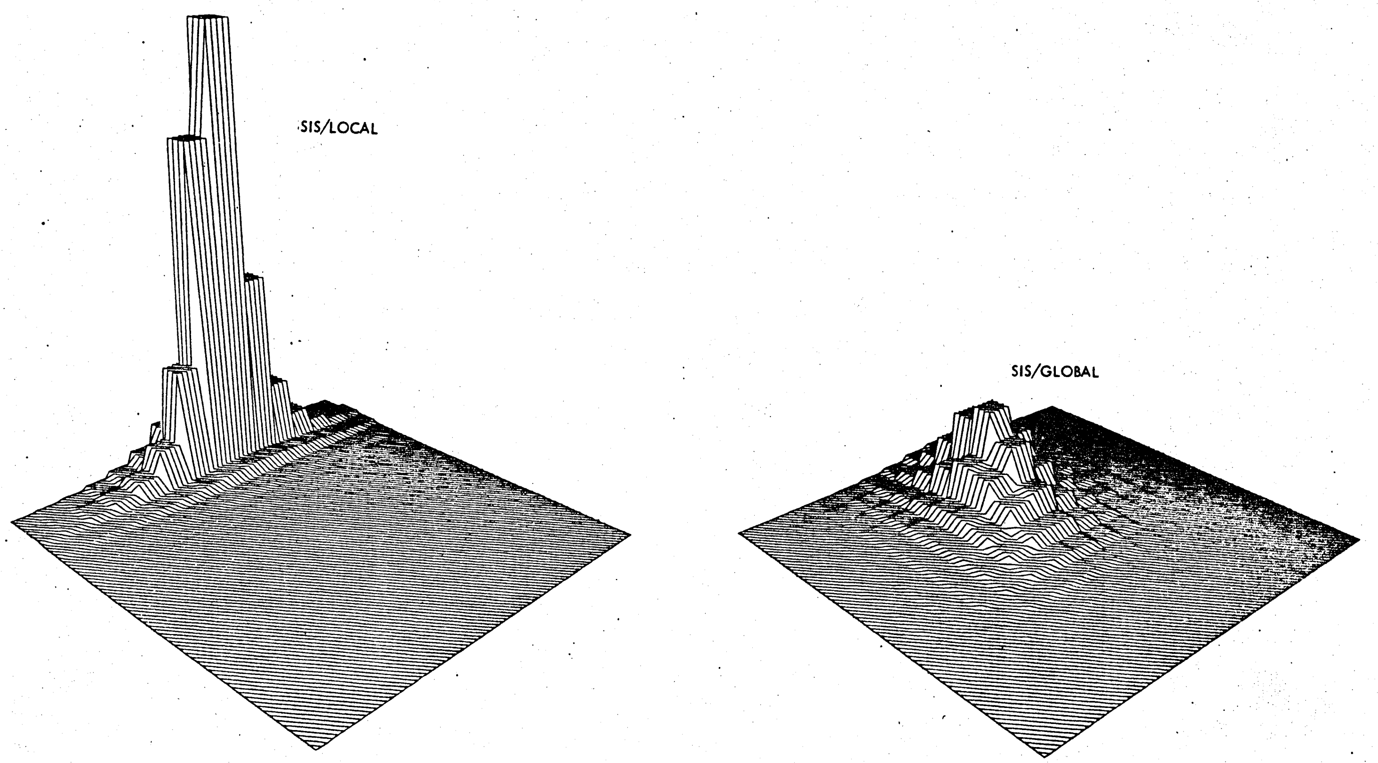


Fig. 5 Bivariate marginal importance functions and bivariate marginal posterior density of (β_1, γ_2)



Mode (β_2, γ_2) = .09, .36

Mean (β_2, γ_2) = (-.31, .31)

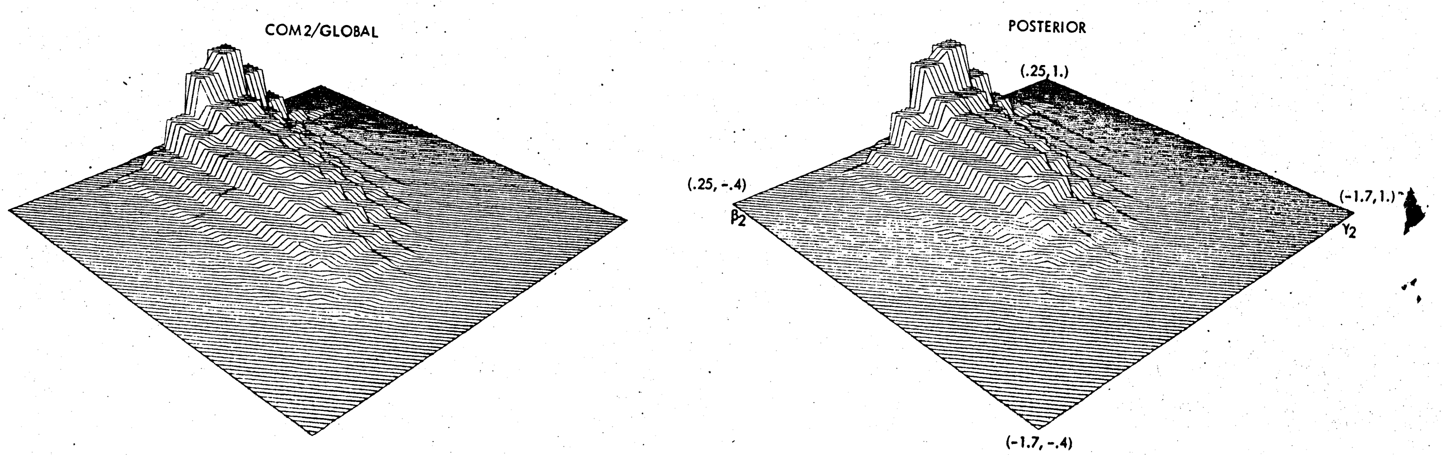


Fig. 6 Bivariate marginal importance functions and bivariate marginal posterior density of (β_2, γ_2)

posterior densities from the mixed integration method require additional numerical calculations. For diagnostic purposes one may use a rough approximation procedure. We have deleted these results in order to save space.

The approximations to the marginal posterior densities of β_1, β_2 and γ_2 are presented in Figures 1b to 3b for a sample of 2000 random drawings. Especially Figure 1b illustrates the poor approximation of the SIS/LOCAL importance function for the posterior of β_1. The spikes in the middle of this figure are caused by very large values of the ratio $p(\theta)/I(\theta)$. The Student importance function in the SIS/LOCAL approach is considerably below the posterior density in the long tail for β_1. This phenomenon has shown up in several other experiments. The bivariate densities illustrate clearly the poor approximation of the Student density in the local case. The CMNMPU density appears to approximate the posterior very well in the global case. Finally, we note that the shapes of the different bivariate posterior densities confirm the remarks about skewness, leptokurtosis and nonlinearity which we made from an inspection of the point estimates of Table 1 above.

4. CONCLUSIONS

In this paper we have reported on the results of some experiments with alternative Monte Carlo methods for simple importance sampling. The results must be interpreted with care since the experiments were limited and were performed on one model only. However, preliminary experiments with Klein's Model I, which involves nine-dimensional numerical integration, using the same methods appear to confirm the following. First, considerable skewness causes gross approximation errors for simple importance sampling, but a minor case of skewness can be dealt with. Second, poor values of the importance function parameters yield also gross approximation errors for simple importance sampling. Mixed integration appears more robust and can be used in a mechanical way. Our experience with the computation of the posterior moments of the Klein-Goldberger model which involves thirty-dimensional numerical integration indicates that mixed integration is feasible in high dimensional cases where skewness occurs in several directions. In contrast, simple importance sampling does not appear to give reliable results in this case [compare van Dijk and Kloek (1982)]. Third, the simplified composition approach, which makes use of the CMNMPU importance function, is computationally efficient once the direction of skewness is known.

Finally, we emphasize that more research is needed in diagnostic analysis in order to investigate which Monte Carlo alternative is suitable and that more research is needed in the area of estimating the parameters of mixtures of multivariate density functions.

ACKNOWLEDGEMENT

We wish to thank A.S. Louter for his assistance with the preparation of the necessary computer programs.

REFERENCES

ATKINSON, A.C. and PEARCE, M.C. (1976). The computer generation of Beta, Gamma and Normal random variables. Journal of the Royal Statistical Society A, 139, 431-448.

EVERITT, B.S. and HAND, D.J. (1981). Finite mixture distributions. London: Chapman and Hall.

KLOEK T. and VAN DIJK, H.K. (1978). Bayesian estimates of equation system parameters; an application of integration by Monte Carlo. Econometrica, 46, 1-19. Reprinted in: A. ZELLNER, ed. (1980) Bayesian analysis in econometrics and statistics. Amsterdam: North-Holland.

RUBINSTEIN, R.Y. (1981). Simulation and the Monte Carlo method. New York: Wiley.

VAN DIJK, H.K. and KLOEK, T. (1980). Further experience in Bayesian analysis using Monte Carlo integration. Journal of Econometrics, 14, 307-328.

VAN DIJK, H.K. and KLOEK, T. (1982). Posterior moments of the Klein-Goldberger model. Report 8223/E, Econometric Institute, Erasmus University Rotterdam. To appear in: P.K. GOEL and A. ZELLNER eds. Bayesian inference and decision techniques with applications. Amsterdam: North-Holland.

VAN DIJK, H.K. and KLOEK, T. (1983a). Monte Carlo analysis of skew posterior distributions: An illustrative econometric example. The Statistician, 32, 216-223.

VAN DIJK, H.K. and KLOEK, T. (1983b). Posterior moments computed by Mixed Integration. Report 8318/E, Econometric Institute, Erasmus University Rotterdam.

LIST OF REPORTS 1983

8300 "Publications of the Econometric Institute Second Half 1982: List of Reprints 300-334, Abstracts of Reports".

8301/S "Constraint Analysis and Hypothesis Generation: An example of systematic preliminary investigation", by J.C.F.L. Groen, A.R. Thurik and D. van der Wijst.

8302/S "Dominated Variation and Related Concepts and Tauberian Theorems for Laplace Transforms", by L. de Haan and U. Stadtmüller.

8303/S "Further Study on the Use of Retail Floorspace and its Efficiency", by A.R. Thurik and J. Koerts.

8304/E "On the Utility Functions of the Indirect Addi-Log Budget Allocation Model", by J. van Daal.

8305/M "On the Difference between Perfect Powers", by J.W.M. Turk.

8306/S "Regression Coëfficient Estimation in Small Linear Models with Unknown Autocorrelation: Evaluation and Recommendation", by C. Dubbelman and E.A. de Groot.

8307/O "A Hierarchical Scheduling Problem with a Well-Solvable Second Stage", by J.B.G. Frenk, A.H.G. Rinnooy Kan and L. Stougie.

8308/M "The Serre Problem for Discrete Hodge Algebras", by A.C.F. Vorst.

8309/S "A Spectral Representation for Max-Stable Processes", by L. de Haan.

8310/M "The Cusp Catastrophe in the Urban Retail Model", by J.F. Kaashoek and A.C.F. Vorst.

8311/O "A Framework for the Probabilistic Analysis of Hierarchical Planning Systems", by J.K. Lenstra, A.H.G. Rinnooy Kan and L. Stougie.

8312/M "Almost Stabilizability Subspaces and High Gain Feedback", by J.M. Schumacher.

8313/I "User Manual to Subroutine Package Minmax for the Dimensioning of a Telephone Network with Nonlinear Blocking Probability Constraints", by V. Stern.

8314/O "Bounds and Heuristics for Capacitated Routing Problems: Part I", by M. Haimovich and A.H.G. Rinnooy Kan.

8315 "Publications of the Econometric Institute First Half 1983"

8316/M "The Fixed Divisor of a Polynomial", by J.W.M. Turk.

8317/O "Stochastic Methods for Global Optimization", by A.H.G. Rinnooy Kan and G.T. Timmer.

8318/E "Posterior moment Computed by Mixed Integration", by H.K. van Dijk and T. Kloek.

8319/O "Bayesian Stopping Rules for a Class of Stochastic Global Optimization Methods", by C.G.E. Boender and A.H.G. Rinnooy Kan.

8320/O "Nonparametric Bayesian Estimation of a Discrete Probability Distribution with Unknown Domain", by C.G.E. Boender and A.H.G. Rinnooy Kan.

8321/M "Brocardian Geometry Revisited or Some Remarkable Inequalities", by R.J. Stroeker and H.J.T. Hoogland.

8322/O "Bayesian Multinomial Estimation of Animal Population Size", by C.G.E. Boender and A.H.G. Rinnooy Kan.

8323/M "On the Square Roots of the Dirac Delta Distribution", by B. Hanzon.

8324/O "Scheduling Theory Since 1981: An Annotated Bibliography", by J.K. Lenstra and A.H.G. Rinnooy Kan.

8325/M "Zeros at Infinity for Afine Nonlinear Control Systems", by H. Nijmeijer and J.M. Schumacher.

8326/E "Experiments with Some Alternatives for Simple Importance Sampling in Monte Carlo Integration", by H.K. van Dijk and T. Kloek.