An Algorithm for the Computation of Posterior Moments and Densities Using Simple Importance Sampling

Herman K. Van Dijk; J. Peter Hop; Adri S. Louter

The Statistician, Vol. 36, No. 2/3, Special Issue: Practical Bayesian Statistics. (1987), pp. 83-90.

Stable URL:
http://links.jstor.org/sici?sici=0039-0526%281987%2936%3A2%2F3%3C83%3AAAFTCO%3E2.0.CO%3B2-D

The Statistician is currently published by Royal Statistical Society.

The Statistician (1987) 36, pp. 83–90

An algorithm for the computation of posterior moments and densities using simple importance sampling*

HERMAN K. VAN DIJK, J. PETER HOP & ADRI S. LOUTER

Econometric Institute, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands

Abstract. In earlier work (van Dijk, 1984, Chapter 3) one of the authors discussed the use of Monte Carlo integration methods for the computation of the multivariate integrals that are defined in the posterior moments and densities of the parameters of interest of econometric models. In the present paper we describe the computational steps of one Monte Carlo method, which is known in the literature as importance sampling. Further, a set of standard programs is available, which may be used for the implementation of a simple case of importance sampling. The computer programs have been written in FORTRAN 77.

1 Introduction

In earlier work (see van Dijk, 1984, Chapter 3, and the references cited there) one of the authors discussed the use of Monte Carlo integration methods for the computation of the multivariate integrals that are defined in the posterior moments and the posterior densities of the parameters of interest of econometric models. In the present paper we describe the computational steps of one Monte Carlo method, mentioned in that work. This method is a simple application of a technique that is known in the literature as *importance sampling* (see Hammersley & Handscomb, 1964). Further, we have prepared a set of standard computer programs, which can be used for the application of importance sampling. The computer programs have been written in FORTRAN 77.

The multivariate integrals that we consider may be described briefly as follows. Let θ be an l-vector of parameters of interest and let $g(\theta)$ be an integrable function of θ. The posterior mean of $g(\theta)$ is defined as

$$Eg(\theta)=\frac{\int g(\theta)p(\theta)d\theta}{\int p(\theta)d\theta} \tag{1.1}$$

where $p(\theta)$ is a kernel of a posterior density function. That is, $p(\theta)$ is proportional and not equal to a density function and the denominator of (1.1) plays the role of integrating constant, similar to the role of $\sqrt{2\pi}$ in the case of the normal distribution. We assume that $g(\theta)p(\theta)$ is integrable on a certain region of integration. Simple examples of $g(\theta)$ are $g(\theta)=\theta$ and $g(\theta)=\theta\theta'$. Note that g may be a vector or a matrix. We emphasise that $g(\theta)$ may also be a complicated nonlinear function of θ such as the implied multipliers of the structural parameters of a simultaneous equation model (see, e.g. van Dijk & Kloek, 1980 and van Dijk, 1984, Chapter 4). There exist several other examples of nontrivial nonlinear functions of θ. For an example in the statistical

*A more extensive version of this paper, including the FORTRAN 77 programs, is available as Report 8625 of the Econometric Institute, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands. This offer expires two years after publication of the paper.

literature we refer to Kass (1985), and for some examples in the econometric literature we refer to van Dijk (1985), Zellner (1985), and Geweke (1986).

Monte Carlo (MC) integration methods make use of the property that generating a large sample of random numbers is very easy using a computer procedure. The value of an integral is then estimated in the sampling theory tradition using this set of random numbers. So, MC methods change the integration problem into a statistical estimation problem. A clear and concise introduction to Monte Carlo has been given by Hammersley & Handscomb (1964).

The contents of this paper has been organised as follows. In Section 2 we describe a Monte Carlo algorithm that is based on the principle of *importance sampling.* Some suggestions for further work are given in Section 3.

2 Simple importance sampling

In this section we discuss the application of a simple case of importance sampling to the computation of the integrals defined in posterior first-order and second order moments and the application of importance sampling to the computation of univariate and bivariate marginal posterior densities. For more details on the principle of importance sampling we refer to Hammersley & Handscomb (1964, Chapter 5) and van Dijk (1984, Chapter 3).

The vector of posterior first-order moments is obtained from (1.1) by defining $g(\theta)=\theta$. This yields

$$E\theta=\frac{\int\theta p(\theta)d\theta}{\int p(\theta)d\theta} \tag{2.1}$$

Suppose that it is not known how one can generate a sample of random drawings from a distribution with density equal or proportional to $p(\theta)$, but it is known how one can generate a random sample from a distribution with a density equal (or proportional) to $I(\theta)$, which is different from $p(\theta)$. Suppose further, that $I(\theta)$ is a reasonable approximation of $p(\theta)$. One can replace $p(\theta)$ in (2.1) by $w(\theta)I(\theta)$ where the weight function $w(\theta)$ is defined as $w(\theta)=p(\theta)/I(\theta)$. This yields

$$E\theta=\frac{\int\theta w(\theta)I(\theta)d\theta}{\int w(\theta)I(\theta)d\theta} \tag{2.2}$$

where $I(\theta)$ is restricted to be positive on the region of integration. $I(\theta)$ is known in the literature as *importance* function. In this paper we make use of a simple choice with respect to the class of importance functions. That is, we opt for the multivariate Student-t class of density functions [see Report 8625 for details on the generation of random drawings from this distribution]. We make use of the term *Simple Importance Sampling* (SIS) in this case. For more details on the choice of an importance function and for some alternatives to simple importance sampling we refer to van Dijk (1984, Chapter 3) and van Dijk and Kloek (1985).

Next, let $\theta^{(1)}, \ldots, \theta^{(N)}$ be a random sample from a distribution with a density function equal (or proportional) to $I(\theta)$. That is, $\theta^{(1)}, \ldots, \theta^{(N)}$ is a sequence of independently distributed random variables with a common distribution function. Let $\theta^{(i)}$ be the typical ith element of this sequence. Then the importance sampling estimator of the jth element of the vector $E\theta$ is given as

$$\hat{E}(\theta_j)=\frac{\frac{1}{N}\sum_{i=1}^{N}\theta_j^{(i)}w(\theta^{(i)})}{\frac{1}{N}\sum_{i=1}^{N}w(\theta^{(i)})} \qquad (j=1, \ldots, l) \tag{2.3}$$

$$=\sum_{j=1}^{N}\theta_j^{(i)}w^*(\theta^{(i)})$$

where

$$w^*(\theta^{(i)})=\frac{w(\theta^{(i)})}{\sum_{i=1}^{N}w(\theta^{(i)})}$$

This estimator may be interpreted as a weighted sample mean of the above mentioned random sample where $w^*(\theta^{(1)}), \ldots, w^*(\theta^{(N)})$ are the weights. The weighted sample mean is a good approximation of $E\theta_j$ if the sample size N is sufficiently large and the variation in the weights is bounded. In order to evaluate the numerical accuracy of the estimator (2.3), we are interested in the variation of $w(\theta)$ and of $\theta_j w(\theta)$, $j=1, \ldots, l$. More details on the numerical accuracy of the approximation (2.3) are given in Report 8625.

Second-order posterior moments can be computed in a similar way as first-order moments. An importance sampling estimator for the (j, k)th element of the second-order moment matrix is given as

$$\hat{E}(\theta_j\theta_k)=\frac{\frac{1}{N}\sum_{i=1}^{N}\theta_j^{(i)}\theta_k^{(i)}w(\theta^{(i)})}{\frac{1}{N}\sum_{i=1}^{N}w(\theta^{(i)})} \qquad (j, k=1, \ldots, l) \tag{2.4}$$

Practical details with respect to the computation of the estimators (2.3) and (2.4) are discussed below.

Univariate marginal posterior densities of θ_j, $j=1, \ldots, l$, can be approximated by so-called frequency *histograms* or frequency *polygones* using MC methods. We start by defining (a_{k-1}, a_k), $k=1, \ldots, K$, as a bounded interval for the parameter θ_j, $j=1, \ldots, l$. Further, let $d(\theta)$ be a dummy variable defined as

$$\begin{aligned} d(\theta)&=1 \quad \text{if } a_{k-1}<\theta_j<a_k \\ &=0 \quad \text{elsewhere} \end{aligned} \tag{2.5}$$

Then the posterior probability P_k, defined as $P_k=P[a_{k-1}<\theta_j<a_k]$, is given as

$$P_k=\frac{\int d(\theta)p(\theta)d\theta}{\int p(\theta)d\theta} \qquad (k=1, \ldots, K) \tag{2.6}$$

The probabilities $P_1, \ldots, P_K$ can be used for the construction of a frequency histogram. Further, the posterior density of θ_j evaluated at $\frac{1}{2}(a_{k-1}+a_k)$ can be approximated by $P_k/(a_k-a_{k-1})$ if the interval (a_{k-1}, a_k) is sufficiently small. This approximation of the posterior density at K points can be used for the construction of a frequency polygon.

Given that $I(\theta)$ is an importance function for $p(\theta)$ we can rewrite (2.6) as

$$P_k=\frac{\int d(\theta)w(\theta)I(\theta)d\theta}{\int w(\theta)I(\theta)d\theta} \tag{2.7}$$

An importance sampling estimator for (2.7) may be derived as follows. Let $\theta^{(1)}, \ldots, \theta^{(N)}$ be a random sample generated from a distribution with density $I(\theta)$. Further, let $\bar{\theta}^{(h)}=\theta^{(h(i))}$, where $h=h(i)$ is generated by the following rule

$$\begin{aligned} h(0) &= 0 \\ h(i) &= h(i-1)+d(i) \quad (i=1, \ldots, N) \end{aligned} \tag{2.8}$$

where

$$\begin{aligned} d(i) &= 1 \quad \text{if } a_{k-1}<\theta_j^{(i)}<a_k \\ &= 0 \quad \text{elsewhere} \end{aligned}$$

Finally, let N_1 be defined as $N_1=h(N)$. Then an importance sampling estimator for (2.7) is given as

$$\hat{P}_k=\frac{\frac{1}{N}\sum_{h=1}^{N_1} w(\bar{\theta}^{(h)})}{\frac{1}{N}\sum_{i=1}^{N} w(\theta^{(i)})} \tag{2.9}$$

The definition of the estimator $\hat{P}_k$ is a bit tedious, but the computation of $\hat{P}_k$ is very simple. In fact, one has only to determine the particular interval to which a weight $w(\theta^{(i)})$ belongs. This is especially simple when the interval width a_k-a_{k-1} is the same for all k. Let b be the common interval width for $k=1, \ldots, K$. Let r be a real number given as

$$r=(\theta_j^{(i)}-a_0)/b+1 \tag{2.10}$$

That is, r is a real number in the interval $[1, K+1]$. Truncate r at its decimal point in order to make r an integer, defined as ir. Then it follows that ir is the interval to which a particular weight $w(\theta^{(i)})$ belongs[1]. So, estimates for P_k are computed by adding the weights that belong in each interval and by dividing the sum of the weights in each interval by the total sum of the weights. Details are presented in the computer program. Minor modifications of the procedure described here are necessary when the intervals have inequal width. Further, the extreme values a_0 and a_K may be equal to minus and plus infinity. Finally, we note that the computation of bivariate marginal posterior densities proceeds in a similar way as the computation of the univariate marginal posterior densities.

Next, we discuss the structure of a computer program for Simple Importance Sampling (SIS). The different computational steps are shown in the flow diagram, given in Fig. 1. The computer program starts with statements that refer to the initial value of a random number generator and to initial zero-values. Note that we make use of an arrow sign (instead of an equality sign) in several statements. For instance, one interprets $S^{(0)}\leftarrow 0$ as: 'the value zero is assigned to the variable (or the array of variables) S with superindex 0.' The major part of the program refers to two so-called do loops. In the inner loop, with the index i, one has as typical statement

$$S^{(i)}\leftarrow S^{(i-1)}+g(\theta^{(i)})w(\theta^{(i)}) \quad (i=1, \ldots, N) \tag{2.11}$$

The symbol $S^{(i)}$ denotes the ith partial sum of a sequence of random function values, defined as

$$S^{(i)}=g(\theta^{(1)})w(\theta^{(1)})+g(\theta^{(2)})w(\theta^{(2)})+\ldots+g(\theta^{(i)})w(\theta^{(i)}) \tag{2.12}$$

The assignment statement (2.11) indicates that one does not store a large sample of random numbers from which sample moments are computed (as is suggested by

[1] The case where i_n is exactly equal to $K+1$ is not important, since it has probability measure zero.

equation (2.12)), but one makes use of a statement that updates the value of $S^{(i)}$ after each accepted Monte Carlo drawing of $\theta^{(i)}$. After N random drawings one computes sample averages such as

$$\bar{S}_N \leftarrow \frac{S^{(N)}}{N} \tag{2.13}$$

The accuracy of the sample means may be studied by increasing the size of the sample from N to $2N$, $3N$, ..., MN. The outer loop of the program, with the index j, enables one to print results at each value of jN, with $j=1, \ldots, M$. This explains why we make use of the index jN at several places in the flow diagram. Given certain regularity conditions, the Monte Carlo estimator $\hat{E}(\theta)$ converges with probability one to $E(\theta)$.

Apart from the two do loops that refer to updating procedures, there are two other major computational procedures in the computer program for simple importance sampling. First, the user has to supply a procedure that describes the computation of the posterior kernel studied in each particular case. Second, we have opted for a procedure that generates random vectors $\theta^{(1)}, \ldots, \theta^{(N)}$ from a multivariate Student t distribution with λ degrees of freedom (λ is a positive integer).

We end the discussion of the structure of the computer program with two remarks. First, the computation of the weights $w(\theta^{(i)})$ is a key step in any importance sampling procedure. The distribution of the weights contains relevant information on the accuracy of importance sampling. The weights are nearly equal if the approximation of $I(\theta)$ to $p(\theta)$ is very accurate. In the opposite case one finds great variation in the weights. An approximation of the distribution of the weights $w(\theta^{(i)})$, $i=1, \ldots, N$ is computed as a byproduct in the computer program for simple importance sampling. Second, we assume that the region of integration is bounded. Therefore, we have inserted a rejection step in the program since the Student t density is defined on the entire region $\boldsymbol{R}^l$, where l is the dimension of the vector θ.

We emphasise that one can make use of the program SIS in an *sequential* way. That is, one starts with the posterior mode and minus the inverse of the Hessian matrix, evaluated at the mode, as location and scale parameters of the multivariate Student t importance function. After a round of Monte Carlo of, for instance, $N=2000$ random drawings one uses the posterior mean and the posterior covariance matrix as new starting values for the parameters of the Student t importance function.

We end this section with some examples. In van Dijk (1984, Chapter 4), the author discussed the computation of the prior and posterior moments of the structural parameters and the implied multipliers of a well-known simultaneous econometric model, which is known as Klein's Model I (see Klein, 1950). Two parameters in this model are particularly interesting, that is, the government-expenditures-multiplier and the tax-multiplier in the reduced form equation for national income, Y. The government-expenditures-multiplier indicates the change in Y when government expenditures, G, change with, for instance, one billion dollars and the tax-multiplier indicates the change in Y when taxes, T, change with such an amount. If one takes a uniform prior on the structural parameters of interest on a bounded region (in this case the unit region, see van Dijk (*op. cit.*)) then one can compute the bivariate prior density of the government- and tax-multipliers by direct simulation and simple rejection. The result is shown in Fig. 2(a, b and c). It is seen that the relevant interval for the tax multiplier is $[-4{\cdot}0, -0{\cdot}5]$ and the relevant interval for the government multiplier is $[0{\cdot}5, 4{\cdot}0]$. The results indicate that the prior distribution is skew, and that it has a long tail. The posterior results have been computed by SIS. These posterior results indicate that the data are quite informative and dominate the prior information strongly. For a more detailed discussion of the results we refer to van Dijk (1984).

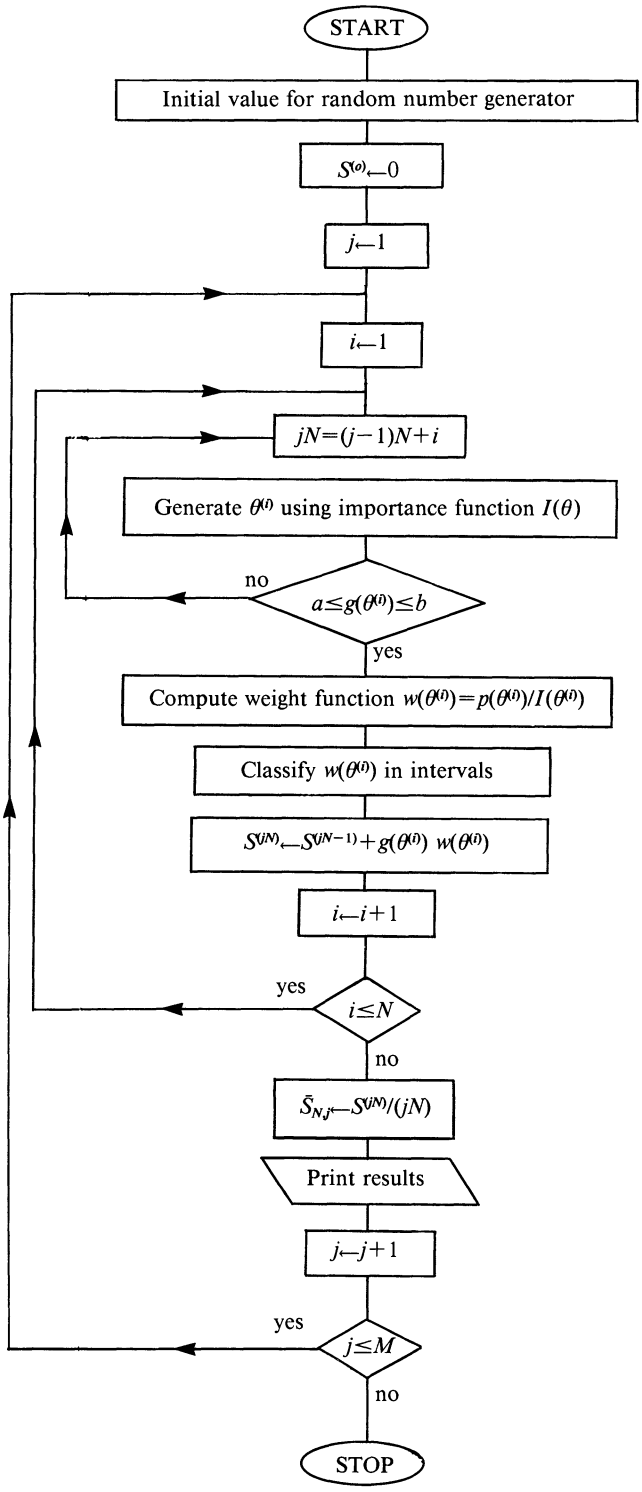

Fig. 1. Flow diagram for simple importance sampling.

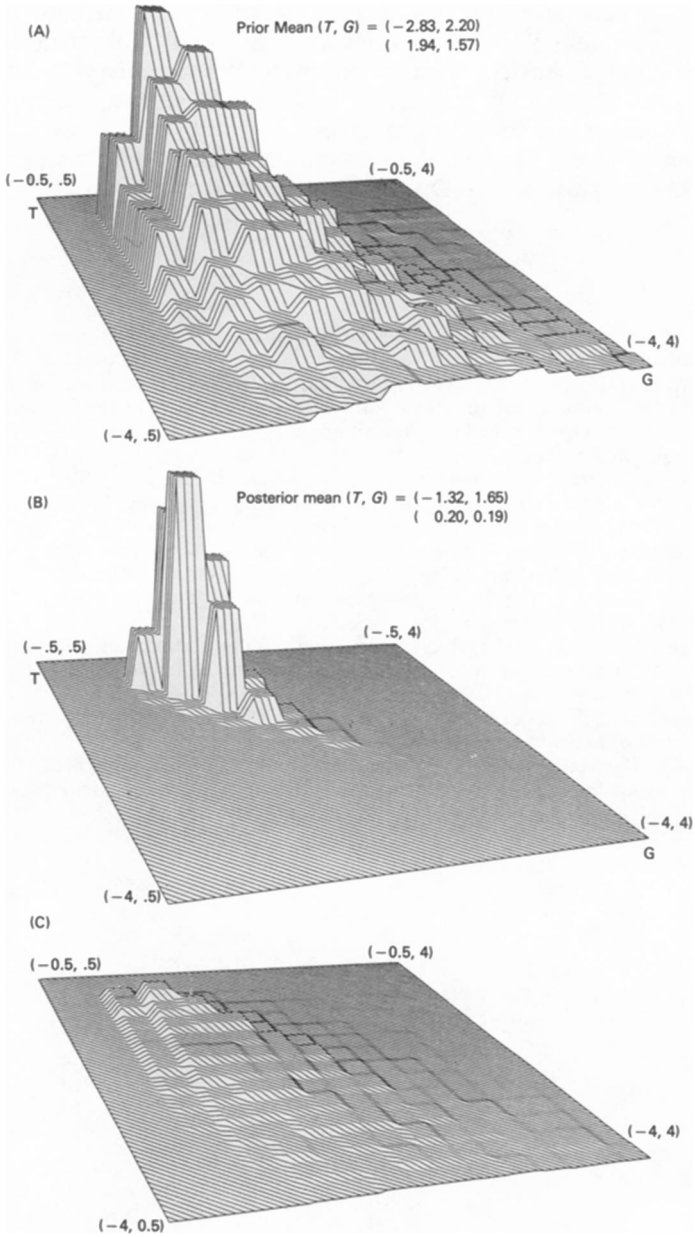

Fig. 2.(A) Bivariate marginal prior densities of short-run government and tax multipliers for Klein's Model 1. (B) Bivariate marginal posterior densities of short-run government and tax multipliers for Klein's Model 1. (C) Scaled bivariate marginal prior densities.

3 Remarks

The computer program for simple importance sampling is a first step towards the development of standard software for Bayesian analysis of econometric and statistical

models. Further developments in this area are needed. An immediate extension is to construct a family of importance functions that is more flexible than the symmetric multivariate Student t density. Some preliminary experiments with an importance function that consists of a finite mixture of conjugate densities appear promising. The results will be reported in a forthcoming paper.

Another topic of recent research is to prepare a standard Fortran program for the method of mixed integration (van Dijk, Kloek & Boender, 1985).

References

GEWEKE, J. (1986) Exact inference in the inequality constrained normal linear regression model, *Journal of Applied Econometrics*, 1, pp. 127–141.

HAMMERSLEY, J.M. & HANDSCOMB, D.C. (1964) *Monte Carlo Methods* (Methuen, London).

KASS, R.E. (1985) Inferences about principal components and related quantities using a numerical delta method and posteriors calculated by simulation, *Technical Report 346* (Department of Statistics, Carnegie-Mellon University).

KLEIN, L.R. (1950) *Economic Fluctuations in the United States, 1921–1941* (Wiley, New York).

VAN DIJK, H.K. (1984) Posterior analysis of econometric models using Monte Carlo integration, *Doctoral dissertation* (Erasmus University Rotterdam).

VAN DIJK, H.K. (1985) A note on simple algebraic relations between structural and reduced form parameters, *Report 8549* (Econometric Institute, Erasmus University, Rotterdam).

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

VAN DIJK, H.K. & KLOEK, T. (1985) Experiments with some alternatives for simple importance sampling in Monte Carlo integration (with discussion), in: BERNARDO, J.M., DEGROOT, M.H., LINDLEY, D.V. & SMITH, A.F.M. (Eds) *Bayesian Statistics 2* (North-Holland, Amsterdam).

VAN DIJK, H.K., KLOEK, T. & BOENDER, C.G.E. (1985) Posterior moments computed by mixed integration, *Journal of Econometrics*, 29, pp. 3–18.

ZELLNER, A. (1985) Further results on Bayesian minimum expected loss (MELO) estimates and posterior distributions for structural coefficients, Manuscript (H.G.B. Alexander Research Foundation, Graduate School of Business, University of Chicago).

References

Exact Inference in the Inequality Constrained Normal Linear Regression Model

John Geweke

Journal of Applied Econometrics, Vol. 1, No. 2. (Apr., 1986), pp. 127-141.

Stable URL:

http://links.jstor.org/sici?sici=0883-7252%28198604%291%3A2%3C127%3AEIITIC%3E2.0.CO%3B2-F