e.g., control variates, mixtures, common random numbers. Very high-dimensional (1000$^+$) integration is feasible as long as the integrand is reasonably well-behaved. Further developments are required to guide the design of efficient IS samplers. IS is not an all-purpose technique and lack of flexibility in the construction of samplers may complicate its application to ill-behaved integrands, for which alternative methods of integration such as MCMC might prove more useful. Nevertheless, IS has proved to be a remarkably useful technique for a wide range of applications including very high-dimensional integrations.

Bibliography

Bratley P, Fox B L, Schrage L E 1987 *A Guide to Simulation*, 2nd edn. Springer Verlag, New York

Danielsson J, Richard J F 1993 Accelerated Gaussian importance sampler with applications to dynamic latent variable models. *Journal of Applied Econometrics* **8**: s153–73

De Bruijn N G 1958 *Asymptotic Methods in Analysis*. North Holland, Amsterdam

Devroye L 1986 *Non-Uniform Random Variate Generation*. Springer Verlag, New York

DiMatteo I, Kadane J 2001 Vote tampering in a District Justice election in Beaver County, Pennsylvania. *Journal of the American Statistical Association* **96**: 510–18

Durbin J, Koopman S J 1997 Monte Carlo maximum likelihood estimation for non-Gaussian state space models. *Biometrika* **84**: 669–84

Evans M 1991 Adaptive importance sampling and chaining. *Contemporary Mathematics (Statistical Multiple Integration)* **115**: 137–42

Geweke J 1989 Bayesian inference in econometric models using Monte Carlo integration. *Econometrica* **57**: 1317–39

Geweke J 1996 Monte Carlo simulation and numerical integration. In: Amman H, Kendrick D, Rust J (eds.) *The Handbook of Computational Economics*. North Holland, Amsterdam, Vol. 1, pp. 731–800

Golub G H, Welsch J H 1969 Calculation of Gaussian quadrature rules. *Mathematics of Computation* **23**: 221–30

Gouriéroux C, Monfort A 1996 *Simulation Based Econometric Methods*. CORE Lecture series, Oxford University Press, Oxford, UK

Hammersley J M, Handscomb D C 1964 *Monte Carlo Methods*. Methuen, London

Harvey A C 1990 *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press, Cambridge, UK

Hendry D F 1984 Monte Carlo experimentation in econometrics. In: Griliches Z, Intriligator M D (eds.) *The Handbook of Econometrics*. North Holland, Amsterdam

Hesterberg T 1995 Weighted average importance sampling and defensive mixture distributions. *Technometrics* **37**: 185–94

Kahn H, Marshall A 1953 Methods of reducing sample size in Monte Carlo computations. *Journal of the Operations Research Society of America* **1**: 263–78

Kloek T, van Dijk H K 1978 Bayesian estimates of equation system parameters: Application of integration by Monte Carlo. *Econometrica* **46**: 1–19

Liesenfeld R, Richard J-F 2000 Univariate and multivariate stochastic volatility models: Estimation and diagnostics. Mimeo, Eberhard-Karls-Universität, Tübingen, Germany

Madras N, Piccioni M 1999 Importance sampling for families of distributions. *The Annals of Applied Probability* **9**: 1202–25

Owen A, Zhou Y 2000 Safe and effective importance sampling. *Journal of the American Statistical Association* **95**: 135–43

Pakes A, Pollard D 1989 Simulation and the asymptotics of optimization estimators. *Econometrica* **57**: 1027–57

Richard J-F, Zhang W 1998 Efficient high-dimensional Monte Carlo importance sampling. Mimeo, University of Pittsburgh, PA

Richard J-F, Zhang W 2000 Accelerated Monte Carlo integration: An application to dynamic latent variables models. In: Mariano R, Schuermann T, Weeks M (eds.) *Simulation-Based Inference in Econometrics: Methods and Applications*. Cambridge University Press, Cambridge, UK

Stern S 1997 Simulation-based estimation. *Journal of Economic Literature* **35**: 2006–39

Stroud A H, Secrest D H 1966 *Gaussian Quadrature Formulae*. Prentice-Hall, Englewood Cliffs, NJ

Tierney L, Kadane J B 1986 Accurate approximations for posterior moments and marginal densities. *Journal of American Statistical Association* **81**: 82–6

Trotter H F, Tukey J W 1956 Conditional Monte Carlo for normal samples. In: Mayer H A (ed.) *Symposium on Monte Carlo Methods*. Wiley, New York

R. Liesenfeld and J.-F. Richard

Monte Carlo Methods and Bayesian Computation: Overview

1. Introduction

Monte Carlo simulation methods and, in particular, Markov chain Monte Carlo methods, play a large and prominent role in the practice of Bayesian statistics, where these methods are used to summarize the posterior distributions that arise in the context of the Bayesian prior–posterior analysis. Monte Carlo methods are used in practically all aspects of Bayesian inference, for example, in the context of prediction problems and in the computation of quantities, such as the marginal likelihood, that are used for comparing competing Bayesian models.

Suppose that $\pi(\psi \,|\, y, \mathcal{M}) \propto p(\psi \,|\, \mathcal{M}) f(y \,|\, \psi, \mathcal{M})$ is the posterior density for a set of parameters $\psi \in \mathcal{R}^d$ in a particular Bayesian model $\mathcal{M}$ defined by the prior density $p(\psi \,|\, \mathcal{M})$ and sampling density or likelihood function $f(y \,|\, \psi, \mathcal{M})$ and that interest centers on the posterior mean $\eta = \int_{\mathcal{R}^d} \psi \pi(\psi \,|\, y, \mathcal{M}) \, d\psi$. As an illustration of this general setting, consider binary data $y_i \in \{0, 1\}$ modeled by the probabilities $\Pr(y_i = 1 \,|\, \psi) = \Phi(x_i \psi)$, where Φ is the standard normal cumulative distribution function, x_i are covariates and ψ is an

unknown vector of parameters. The posterior distribution of ψ given n independent measurement $y = (y_1,...,y_n)$ is proportional to

$$\pi(\psi \mid y, \mathcal{M}) \propto p(\psi \mid \mathcal{M}) \prod_{i-1}^{n} \Phi(x_i'\psi)^{y_i}(1 - \Phi(x_i'\psi))^{(1-y_i)}$$

and the goal is to summarize this density.

Now suppose that a sequence of variates $\{\psi^{(1)},...,\psi^{(M)}\}$ drawn from $\pi(\psi \mid y, \mathcal{M})$ are available and one computes the sample average of these draws. Under any set of conditions that allow use of the strong law of large numbers, it follows that the sample average is simulation consistent, converging to the posterior mean as the simulation sample M goes to infinity. This then is the essence of the Monte Carlo method: to compute an expectation, or to summarize a given density, sample the density by any method that is appropriate and use the sampled draws to estimate the expectation and to summarize the density.

The sampling of the posterior distribution is, therefore, the central focus of Bayesian computation. One important breakthrough in the use of simulation methods was the realization that the sampled draws need not be independent, that simulation consistency can be achieved with correlated draws. The fact that the sampled variates can be correlated is of immense practical and theoretical importance and is the defining characteristic of Markov chain Monte Carlo methods, popularly referred to by the acronym MCMC, where the sampled draws form a Markov chain (see also *Markov Chain Monte Carlo Methods*). The idea behind these methods is simple and extremely general. In order to sample a given probability distribution, referred to as the target distribution, a suitable Markov chain is constructed with the property that its limiting, invariant distribution is the target distribution. Once the Markov chain has been constructed, a sample of draws from the target distribution is obtained by simulating the Markov chain a large number of times and recording its values. Within the Bayesian framework, where both parameters and data are treated as random variables, and inferences about the parameters are conducted conditioned on the data, the posterior distribution of the parameters provides a natural target for MCMC methods.

Markov chain sampling methods originate with the work of Metropolis et al. (1953) in statistical physics. A crucial and landmark extension of the method was made by Hastings (1970) leading to a method that is now called the Metropolis–Hastings algorithm. This algorithm was first applied to problems in spatial statistics and image analysis (Besag 1974). A resurgence of interest in MCMC methods started with the papers of D. Geman and S. Geman (1984) who developed an algorithm, a special case of the Metropolis method that later came to be called the Gibbs sampler, to sample a discrete distribution, Tanner and

Wong (1987) who proposed a MCMC scheme involving 'data augmentation' to sample posterior distributions in missing data problems, and Gelfand and Smith (1990) where the value of the Gibbs sampler was demonstrated for general Bayesian inference with continuous parameter spaces.

2. Markov Chain Monte Carlo Methods

2.1 The Metropolis–Hastings Algorithm

This powerful algorithm provides a general approach for producing a correlated sequence of draws from the target density $\pi(\psi)$ that may be difficult to sample by a classical independence method. As before, let $\pi(\psi)$ denote the density that one wishes to sample, where one can assume for simplicity that the density is absolutely continuous; and let $q(\psi, \psi')$ denote a source density for a candidate draw ψ' given the current value ψ in the sampled sequence. The density $q(\psi, \psi')$ is referred to as the proposal or candidate-generating density. Then, the M–H algorithm is defined by two steps: a first step in which a proposal value is drawn from the candidate-generating density and a second step in which the proposal value is accepted or rejected according to a probability given by

$$\alpha(\psi, \psi') = \begin{cases} \min\left[\dfrac{\pi(\psi')q(\psi', \psi)}{\pi(\psi)q(\psi, \psi')}, 1\right] & \text{if } \pi(\psi)q(\psi, \psi') > 0 \\ 1 & \text{otherwise} \end{cases}$$

(1)

If the proposal value is rejected, then the next sampled value is taken to be the current value; which means that when a rejection occurs the current value is repeated. Given the new value, the same two-step process is repeated and the cycle iterated a large number of times. A full expository discussion of this algorithm, along with a derivation of the method from the logic of reversibility, is provided by Chib and Greenberg (1995).

The M–H method delivers variates from π under quite general conditions. A weak requirement for a law of large numbers for sample averages based on the M–H output involve positivity and continuity of $q(\psi, \psi')$ for (ψ, ψ'), and connectedness of the support of the target distribution. In addition, if π is bounded then conditions for ergodicity, required to establish the central limit theorem, are satisfied. A discussion of the conditions and the underlying Markov chain theory that justifies these methods was first provided by Tierney (1994) and summarized recently by Robert and Casella (1999). Of course, the variates are from π only in the limit as the number of iterations becomes

large but, in practice, after an initial burn-in phase consisting of (say) n_0 iterations, the chain is assumed to have converged and subsequent values are taken as approximate draws from π. Theoretical calculation of the burn-in is not easy, so it is important that the proposal density be chosen to ensure that the chain makes large moves through the support of the target distribution without staying at one place for many iterations. Generally, the empirical behavior of the M–H output is monitored by the autocorrelation time of each component of ψ defined as

$$\left\{ 1 + 2 \sum_{s=1}^{\mathcal{M}} \left(1 - \frac{s}{\mathcal{M}} \right) \rho_{ks} \right\} \quad (2)$$

where ρ_{ks} is the sample autocorrelation at lag s for the kth component of ψ, and by the acceptance rate which is the proportion of times a move is made as the sampling proceeds. Because independence sampling produces an autocorrelation time that is theoretically equal to one, and Markov chain sampling produces autocorrelation times that are bigger than one, it is desirable to tune the M–H algorithm to get values close to one, if possible.

Given the form of the acceptance probability $\alpha(\psi, \psi')$ it is clear that the M–H algorithm does not require knowledge of the normalizing constant of $\pi(\cdot)$. Furthermore, if the proposal density satisfies the symmetry condition $q(\psi, \psi') = q(\psi', \psi)$, the acceptance probability reduces to $\pi(\psi')/\pi(\psi)$; hence, if $\pi(\psi') \geqslant \pi(\psi)$, the chain moves to ψ', otherwise it moves to ψ' with probability given by $\pi(\psi')/\pi(\psi)$. The latter is the algorithm originally proposed by Metropolis et al. (1953).

Different proposal densities give rise to specific versions of the M–H algorithm, each with the correct invariant distribution π. One family of candidate-generating densities is given by $q(\psi, \psi') = q(\psi' - \psi)$. The candidate ψ' is thus drawn according to the process $\psi' = \psi + z$, where z follows the distribution q, and is referred to as the *random walk M–H* chain. The random walk M–H chain is perhaps the simplest version of the M–H algorithm and is quite popular in applications. One has to be careful, however, in setting the variance of z because if it is too large it is possible that the chain may remain stuck at a particular value for many iterations while if it is too small the chain will tend to make small moves and move inefficiently through the support of the target distribution.

Hastings (1970) considers a second family of candidate-generating densities that are given by the form $q(\psi, \psi') = q(\psi')$. Tierney (1994) refers to this as an *independence M–H chain* because the candidates are drawn independently of the current location ψ. Chib and Greenberg (1994, 1995) implement such chains by matching the proposal density to the target at the mode by a multivariate normal or multivariate-t distribution with location given by the mode of the target and the dispersion given by inverse of the Hessian evaluated at the mode.

2.2 The Multiple-block M–H Algorithm

In applications when the dimension of ψ is large it may be preferable, as discussed by Hastings (1970), to construct the Markov chain simulation by first grouping the variables ψ into smaller blocks. For simplicity suppose that two blocks are adequate and that ψ is written as (ψ_1, ψ_2), with $\psi_k \in \Omega_k \subseteq \mathscr{R}^{d_k}$. In that case the M–H algorithm requires the specification of two proposal densities,

$$q_1(\psi_1, \psi_1' \mid \psi_2); \quad q_2(\psi_2, \psi_2' \mid \psi_1)$$

one for each block ψ_k, where the proposal density q_k may depend on the current value of the remaining block. Also, define

$$\alpha(\psi_1, \psi_1' \mid \psi_2) = \min \left\{ \frac{\pi(\psi_1', \psi_2) q_1(\psi_1', \psi_1 \mid \psi_2)}{\pi(\psi_1, \psi_2) q_1(\psi_1, \psi_1' \mid \psi_2)}, 1 \right\} \quad (3)$$

and

$$\alpha(\psi_2, \psi_2' \mid \psi_1) = \min \left\{ \frac{\pi(\psi_1, \psi_2') q_2(\psi_2', \psi_2 \mid \psi_1)}{\pi(\psi_1, \psi_2) q_2(\psi_2, \psi_2' \mid \psi_1)}, 1 \right\} \quad (4)$$

as the probability of move for block ψ_k conditioned on the other block. Then, in what may be called the multiple-block M–H algorithm, one cycle of the algorithm is completed by updating each block, say sequentially in fixed order, using a M–H step with the above probability of move, given the most current value of the other block.

The extension of this method to more than two blocks is straightforward.

2.3 The Gibbs Sampling Algorithm

Another MCMC method, which is a special case of the multiple-block M–H method, is called the Gibbs sampling method and was brought into statistical prominence by Gelfand and Smith (1990). An elementary introduction to Gibbs sampling is provided by Casella and George (1992). To describe this algorithm, suppose that the parameters are grouped into two blocks (ψ_1, ψ_2) and each block is sampled according to the *full conditional distribution* of block ψ_k, defined as the conditional distribution under π of ψ_k given the other block. In parallel with the multiple-block M–H algorithm, the most current value of the

other block is used in deriving the full conditional distribution. Derivation of these full conditional distributions is usually quite simple since, by Bayes' theorem, each full conditional distribution $\pi(\psi_1 | \psi_2)$ and $\pi(\psi_2 | \psi_1)$ are both proportional to $\pi(\psi_1, \psi_2)$, the joint distribution of the two blocks. In addition, the powerful device of data augmentation, due to Tanner and Wong (1987), in which latent or auxiliary variables are artificially introduced into the sampling, is often used to simplify the derivation and sampling of the full conditional distributions. For further details see *Markov Chain Monte Carlo Methods*.

3. Implementation Issues

In implementing a MCMC method it is important to assess the performance of the sampling algorithm to determine the rate of mixing and the size of the burn-in (both having implications for the number of iterations required to get reliable answers). A large literature has now emerged on these issues, for example Tanner (1996, Sect. 6.3), Cowles and Carlin (1996), Gamerman (1997, Sect. 5.4) and Robert and Casella (1999).

Writing at the turn of the millenium, the more popular approach is to utilize the sampled draws to assess both the performance of the algorithm and its approach to the stationary, invariant distribution. Several such relatively informal methods are now available. Perhaps the simplest and most direct are autocorrelation plots (and autocorrelation times) of the sampled output. Slowly decaying correlations indicate problems with the mixing of the chain. It is also useful in connection with M–H Markov chains to monitor the acceptance rate of the proposal values with low rates implying 'stickiness' in the sampled values and thus a slower approach to the invariant distribution.

Somewhat more formal sample-based diagnostics are also available in the literature, as summarized in the routines provided by Best et al. (1995) and now packaged in the BOA program. Although these diagnostics often go under the name 'convergence diagnostics' they are in principle approaches that detect *lack* of convergence. Detection of convergence based entirely on the sampled output, without analysis of the target distribution, is extremely difficult and perhaps impossible. Cowles and Carlin (1996) discuss and evaluate thirteen such diagnostics (for example, those proposed by Geweke (1992), Gelman and Rubin (1992), Zellner and Min (1995), amongst others) without arriving at a consensus.

While implementing MCMC methods it is important to construct samplers that mix well (mixing being measured by the autocorrelation times) because such samplers can be expected to converge more quickly to the invariant distribution. As a general rule, sets of parameters that are highly correlated should be treated

as one block when applying the multiple-block M–H algorithm. Otherwise, it would be difficult to develop proposal densities that lead to large moves through the support of the target distribution and the sampled draws would tend to display autocorrelations that decay slowly. Liu (1994) and Liu et al. (1994) discuss possible strategies in the context of three-block Gibbs MCMC chains. Roberts and Sahu (1997) provide further discussion of the role of blocking in the context of Gibbs Markov chains.

4. Additional Uses of MCMC Methods in Bayesian Inference

MCMC methods provide an elegant solution to the problem of prediction in the Bayesian context. The goal is to summarize the Bayesian predictive density defined as

$$f(y_f | y, \mathcal{M}) = \int f(y_f | y, \mathcal{M}, \theta) \pi(\theta | y, \mathcal{M}) \, d\theta, \quad (5)$$

where $f(y_f | y, \mathcal{M}, \theta)$ is the conditional density of future observations y_f given $(y, \mathcal{M}, \theta)$. In general, the predictive density is not available in closed form. However, this is not a problem when a sample of draws from the posterior density is available. If one simulates

$$y_f^{(i)} \sim f(y_f | y, \mathcal{M}, \theta^{(j)}) \quad (6)$$

for each sampled draw $\theta^{(j)}$, then the collection of simulated values $\{y_f^{(1)}, \ldots, y_f^{(M)}\}$ is a sample from the Bayes prediction density $f(y_f | y, \mathcal{M})$. The simulated sample can be summarized in the usual way by the computation of sample averages and quantiles.

MCMC methods have also been widely applied to the problem of the model choice. Suppose that there are K possible models $\mathcal{M}_1, \ldots, \mathcal{M}_K$ for the observed data defined by the sampling densities $\{f(y | \theta_k, \mathcal{M}_k)\}$ and proper prior densities $\{p(\theta_k | \mathcal{M}_k)\}$ and the objective is to find the evidence in the data for the different models. In the Bayesian approach this question is answered by placing prior probabilities $\Pr(\mathcal{M}_k)$ on each of the K models and using the Bayes calculus to find the posterior probabilities $\{\Pr(\mathcal{M}_1 | y), \ldots, \Pr(\mathcal{M}_K | y)\}$ conditioned on the data but marginalized over the unknowns θ_k (Jeffreys 1961 and Kass and Raftery 1995). Specifically, the posterior probability of $\mathcal{M}_k$ is given by the expression

$$\Pr(\mathcal{M}_k | y) = \frac{\Pr(\mathcal{M}_k) m(y | \mathcal{M}_k)}{\sum_{i=1}^{K} \Pr(\mathcal{M}_i) m(y | \mathcal{M}_i)}$$
$$\propto \Pr(\mathcal{M}_k) m(y | \mathcal{M}_k),$$
$$(k \leqslant K)$$

$$(7)$$

where $m(y \,|\, \mathcal{M}_k)$ is the marginal likelihood of $\mathcal{M}_k$.

A central problem in estimating the marginal likelihood is that it is an integral of the sampling density over the prior distribution of θ_k. Thus, MCMC methods, which deliver sample values from the posterior density, cannot be used to directly average the sampling density. In addition, taking draws from the prior density to do the averaging produces an estimate that is simulation consistent but highly inefficient because draws from the prior density are not likely to be in high density regions of the sampling density $f(y \,|\, \theta_k, \mathcal{M}_k)$. A natural way to correct this problem is by the method of importance sampling. If we let $h(\theta_k \,|\, \mathcal{M}_k)$ denote a suitable importance sampling function, then the marginal likelihood can be estimated as

$$\hat{m}_I(y \,|\, \mathcal{M}_k) = M^{-1} \sum_{j-1}^{M} \frac{f(y \,|\, \theta_k^{(j)}, \mathcal{M}_k) p(\theta_k^{(j)} \,|\, \mathcal{M}_k)}{h(\theta_k^{(j)} \,|\, \mathcal{M}_k)}$$
$$\theta_k^{(j)} \sim h(\theta_k^{(j)} \,|\, \mathcal{M}_k) \quad (j \leqslant M) \qquad (8)$$

This method is useful when it can be shown that the ratio is bounded, which can be difficult to check in practice, and when the sampling density is not expensive to compute; this, unfortunately, is often not so. Gelfand and Dey (1994) propose an estimator that is based on the posterior simulation output. Another method for finding the marginal likelihood is due to Chib (1995). The marginal likelihood, by virtue of being the normalizing constant of the posterior density, can be expressed as

$$m(y \,|\, \mathcal{M}_k) = \frac{f(y \,|\, \theta_k, \mathcal{M}_k) p(\theta_k \,|\, \mathcal{M}_k)}{\pi(\theta_k \,|\, y, \mathcal{M}_k)} \qquad (9)$$

which is an identity in θ_k. Based on this expression an estimate of the marginal likelihood on the log-scale is given by

$$\log \hat{m}(y \,|\, \mathcal{M}_k) = \log f(y \,|\, \theta_k^*, \mathcal{M}_k) + \log p(\theta_k^* \,|\, \mathcal{M}_k)$$
$$- \log \hat{\pi}(\theta_k^* \,|\, y, \mathcal{M}_k) \qquad (10)$$

where θ_k^* denotes an arbitrarily chosen point and $\hat{\pi}(\theta_k^* \,|\, y, \mathcal{M}_k)$ is the estimate of the posterior density at that single point. To estimate this posterior ordinate Chib (1995) utilizes the Gibbs output in conjunction with a decomposition of the ordinate into marginal and conditional components. Chib and Jeliazkov (2001) extend that approach for output produced by the Metropolis–Hastings algorithm. A comparison of different estimators of the marginal likelihood is provided by Chen and Shao (1997).

When one is presented with a large collection of candidate models, each with parameters $\theta_k \in B_k \subseteq R^{d_k}$, direct fitting of each model to find the marginal likelihood can be computationally expensive. In such cases it may be more fruitful to utilize model space–parameter space MCMC algorithms that eschew direct

fitting of each model for an alternative simulation of a 'mega model' where a model index random variable, denoted as $\mathcal{M}$, taking values on the integers from 1 to K, is sampled in tandem with the parameters. The posterior distribution of $\mathcal{M}$ is then computed as the frequency of times each model is visited. Methods for doing this have been proposed by Carlin and Chib (1995) and Green (1995) which are discussed in article 2.1.76. Both methods are closely related as shown by Dellaportas et al. (1998) and Godsill (1998). The problem of variable selection in regression has also been tackled in this context starting with George and McCulloch (1993).

5. An Example

Markov chain Monte Carlo methods have proved enormously popular in Bayesian statistics (see for example, Gilks et al. 1996) where these methods have opened up vistas that were unimaginable fifteen years ago.

As an illustration of how MCMC methods are designed in a Bayesian context, we return to the binary data problem discussed in the introduction and describe the approach of Albert and Chib (1993) that has found many applications. Recall that the posterior distribution of interest is proportional to

$$\pi(\beta \,|\, y, \mathcal{M}) \propto p(\beta \,|\, \mathcal{M}) \prod_{i-1}^{n} p_i^{y_i} (1 - p_i)^{(1-y_i)} \qquad (11)$$

where $p_i = \Phi(x_i' \psi)$. To analyze this probit model, and other models for binary and ordinal data, Albert and Chib introduce latent variables and write $z_i = x_i' \psi + \varepsilon_i$ where ε_i is $\mathcal{N}(0, 1)$ and let $y_i = I[z_i > 0]$ where $I[\cdot]$ is the indicator function. This two-equation specification is equivalent to the binary probit regression model since $\Pr(y_i = 1 \,|\, \psi) = \Pr(z_i > 0 \,|\, \psi) = \Phi(x_i' \psi)$, as required. The Albert–Chib algorithm now consists of simulating the joint posterior distribution of (ψ, Z), where $Z = (z_1, ..., z_n)$. The MCMC simulation is implemented by sampling the conditional distributions $[\psi \,|\, y, Z]$ and $[Z \,|\, y, \psi]$. The first of these distributions depends only on Z (and not on y) and is obtained by standard Bayesian results on normal linear models. The second factors into a product of independent distributions $[z_i \,|\, y_i, \psi]$. The knowledge of y_i merely restricts the support of the normal distribution $z_i \,|\, \psi$; if $y_i = 0$, then $z_i \leqslant 0$ and if $y_i = 1$, then $z_i > 0$. Suppose that *a priori* $\psi \sim \mathcal{N}_k(\psi_0, C_0)$, where ψ_0 and C_0 are known hyperparameters. Then, a sample of ψ and Z from the posterior distribution is obtained by iterating on the following steps:

(a) sample ψ from the distribution $\mathcal{N}_k(\hat{\psi}, C)$, where $\hat{\psi} = C(C_0^{-1} \psi_0 + \sum_{i-1}^{n} x_i z_i)$, $C = (B_0^{-1} + \sum_{i-1}^{n} x_i x_i')^{-1}$;

(b) sample z_i from $\mathcal{TN}_{(-\infty, 0)}(x_i' \psi, 1)$ if $y_i = 0$ and from $\mathcal{TN}_{(0, \infty)}(x_i' \psi, 1)$ if $y_i = 1$, where $\mathcal{TN}_{(a,b)}(\mu, \sigma^2)$ denotes the $\mathcal{N}(\mu, \sigma^2)$ distribution truncated to the region (a, b).

At this time, considerable work on MCMC methods and applications continues to emerge, a testament to the importance and viability of MCMC methods in Bayesian analysis.

See also: Bayesian Statistics; Markov Chain Monte Carlo Methods; Markov Models and Social Analysis

Bibliography

Albert J H, Chib S 1993 Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association* **88**: 669–79

Besag J 1974 Spatial interaction and the statistical analysis of lattice systems (with discussion). *Journal of the Royal Statistical Society*, B Met **36**: 192–236

Best N G, Cowles M K, Vines S K 1995 CODA: Convergence diagnostics and output analysis software for Gibbs sampling. *Technical report*, Cambridge MRC Biostatistics Unit

Carlin B P, Chib S 1995 Bayesian model choice via Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society*, B Met **57**: 473–84

Casella G, George E I 1992 Explaining the Gibbs sampler. *American Statistician* **46**: 167–74

Chen M-H, Shao Q-M 1997 On Monte Carlo methods for estimating ratios of normalizing constants. *Annals of Statistics* **25**: 1563–94

Chib S 1995 Marginal likelihood from the Gibbs output. *Journal of the American Statistical Association* **90**: 1313–21

Chib S, Greenberg E 1994 Bayes inference in regression models with ARMA(p, q) errors. *Journal of Econometrics* **64**: 183–206

Chib S, Greenberg E 1995 Understanding the Metropolis–Hastings algorithm. *American Statistician* **49**: 327–35

Chib S, Jeliazkov I 2001 Marginal likelihood from the Metropolis–Hastings output. *Journal of the American Statistical Association* **96**(453): 276–81

Cowles M K, Carlin B 1996 Markov Chain Monte Carlo convergence diagnostics: A comparative review. *Journal of the American Statistical Association* **91**: 883–904

Dellaportas P, Forster J J, Ntzoufras I 1998 On Bayesian model and variable selection using MCMC. *Technical report*, University of Economics and Business, Greece

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

Gelfand A E, Dey D (1994) Bayesian Model choice: Asymptotics and exact calculations. *Journal of the Royal Statistical Society* **B56**: 501–14

Gelfand A E, Smith A F M 1990 Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association* **85**: 398–409

Gelman A, Rubin D B 1992 Inference from iterative simulation using multiple sequence. *Statistical Science* **4**: 457–72

Geman D, Geman S 1984 Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **12**: 609–28

George E I, McCulloch R E 1993 Variable selection via Gibbs sampling. *Journal of the American Statistical Association* **88**: 881–9

Geweke J 1992 Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In: Bernardo J M, Berger J O, Dawid A P, Smith A F M (eds.) *Bayesian Statistics*. New York: Oxford University Press, pp. 169–93

Gilks W R, Richardson S, Spiegelhalter D J 1996 *Markov Chain Monte Carlo in Practice*. Chapman and Hall, London

Godsill S J 1998 On the relationship between model uncertainty methods. *Technical report*, Signal Processing Group, Cambridge University

Green P J 1995 Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* **82**: 711–32

Hastings W K 1970 Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**: 97–109

Jeffreys H 1961 *Theory of Probability*, 3rd edn. Oxford University Press, New York

Kass R E, Raftery A E 1995 Bayes factors. *Journal of the American Statistical Association* **90**: 773–95

Liu J S 1994 The collapsed Gibbs sampler in Bayesian computations with applications to a gene regulation problem. *Journal of the American Statistical Association* **89**: 958–66

Liu J S, Wong W H, Kong A 1994 Covariance structure of the Gibbs Sampler with applications to the comparisons of estimators and augmentation schemes. *Biometrika* **81**: 27–40

Mengersen K L, Tweedie R L 1996 Rates of convergence of the Hastings and Metropolis algorithms. *Annals of Statistics* **24**: 101–21

Metropolis N, Rosenbluth A W, Rosenbluth M N, Teller A H, Teller E 1953 Equations of state calculations by fast computing machines. *Journal of Chemical Physics* **21**: 1087–92

Muller P, Erkanli A, West M 1996 Bayesian Curve fitting using multivariate normal mixtures. *Biometrika* **83**: 67–79

Robert C P, Casella G 1999 *Monte Carlo Statistical Methods*. Springer, New York

Roberts G O, Sahu S K 1997 Updating schemes, correlation structure, blocking, and parametization for the Gibbs sampler. *Journal of the Royal Statistical Society*, B Met **59**: 291–317

Tanner M A 1996 *Tools for Statistical Inference*, 3rd edn. Springer, New York

Tanner M A, Wong H W 1987 The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association* **82**: 528–40

Tierney L 1994 Markov chains for exploring posterior distributions (with discussion). *Annals of Statistics* **22**: 1701–62

Zellner A, Min C K 1995 Gibbs sampler convergence criteria. *Journal of the American Statistical Association* **90**: 921–7

S. Chib

Montesquieu, Charles, the Second Baron of (1689–1755)

1. Biography and Works

Montesquieu, generally held to be the 'chief theoretician of the separation of powers in western constitutional thought' (Stubbe-Da Luz 1998, p. 7), was born in La Brède, south of Bordeaux, on January 18, 1689 as Charles-Louis de Secondat. His family, which owned extensive vineyards, was part of the aristocratic

10009