Exploring heterogeneity in tumour data using Markov chain Monte Carlo

Mathisca C. M. de Gunst[1,*,†], Anup Dewanji[2] and E. Georg Luebeck[3]

[1]*Department of Mathematics, Free University of Amsterdam, De Boelelaan 1081 a, 1081 HV Amsterdam, Netherlands*
[2]*Applied Statistics Unit, Indian Statistical Institute, 203 B. T. Road, Kolkata 700 108, India*
[3]*Fred Hutchinson Cancer Research Center, Public Health Sciences Division, MP-665, 1100 Fairview Avenue North, Seattle, WA 98109-1024, U.S.A.*

SUMMARY

We describe a Bayesian approach to incorporate between-individual heterogeneity associated with parameters of complicated biological models. We emphasize the use of the Markov chain Monte Carlo (MCMC) method in this context and demonstrate the implementation and use of MCMC by analysis of simulated overdispersed Poisson counts and by analysis of an experimental data set on preneoplastic liver lesions (their number and sizes) in the presence of heterogeneity. These examples show that MCMC-based estimates, derived from the posterior distribution with uniform priors, may agree well with maximum likelihood estimates (if available). However, with heterogeneous parameters, maximum likelihood estimates can be difficult to obtain, involving many integrations. In this case, the MCMC method offers substantial computational advantages.

KEY WORDS: Markov chain Monte Carlo (MCMC); inter-individual variation; stochastic growth model; premalignant lesions; N-nitrosomorpholine (NNM)

1. INTRODUCTION

Stochastic models that explicitly incorporate information on the number and sizes of premalignant lesions are mathematically complex but promise to yield insights into fundamental aspects of carcinogenesis. For example, analyses of preneoplastic liver lesions in the rat using models that utilize the information of lesion number and sizes [1–3], are useful to characterize the role and potency of various putative tumour agents in experimental carcinogenesis.

Most analyses of such data assume that the model parameters are homogeneous among the animals, and therefore can be estimated without further complications by likelihood maximization. However, based on the inter-individual variation seen in some observations, there is also reason to believe that at least some of the parameters are heterogeneous (see reference [1], for example) between animals. The models incorporating this heterogeneity and, therefore, the analyses via maximum likelihood estimation, become much more complex and computationally challenging. The present paper suggests an alternative method of analysis via Markov chain Monte Carlo (MCMC) techniques in a Bayesian framework, which allows us to incorporate heterogeneity in a straightforward way.

In order to deal with heterogeneity between individuals, we may assume a probability distribution for each of the parameters suspected to be heterogeneous. The specific values of such parameters for one particular individual can then be seen as realizations from the corresponding probability distributions, and are independent of the values of those for other individuals. The method of maximum likelihood would proceed to consider likelihood contributions from individual subjects by integrating over the space of these parameters with respect to their probability distributions. This not only increases the number of parameters from that of the homogeneous case (the heterogeneity distribution for a positive parameter is usually assumed to be gamma or log-normal involving two parameters), but also increases the computational burden by having to calculate as many as n (the number of individuals) numerical integrals of high dimension (same as the number of heterogeneous parameters) for each evaluation of the likelihood. Often the stability of numerical integrations has to be checked carefully, involving additional computations. Moreover, with complicated parametric modelling, the likelihood surface may have multiple modes, affecting the search for the global maximum. In view of this, it seems worthwhile to investigate an alternative method of analysis when complex parametric models are used.

Since MCMC methods have been shown to perform quite well in situations where heterogeneity or, more generally, random effects, or other complex features are involved (see, for instance, references [4–6]), we suggest this as an alternative. We shall not focus on the mathematical details but try to give an understanding of the general idea in Section 2.2. For more details, we refer to references [7–10]. An MCMC procedure gives as output a long simulated chain of values for the set of parameters. It is obtained from a well-defined Markov process in such a way that the values of a particular parameter along the chain can be thought of as realizations from the corresponding marginal posterior distribution. Hence these values can be used to estimate any moment of the posterior distribution or to obtain highest posterior density credible intervals. Moreover, when a parametric functional is of interest rather than the individual parameters, MCMC methods allow one to calculate this functional for each set of parameters in the chain, thus having a chain of values for the functional itself, which in turn may yield posterior moment estimates or credible intervals for the functional. In situations where the number of observations is small, as is often the case for complex cancer data, the large sample normality of parameter estimates based on the method of maximum likelihood, and hence the validity of the information matrix based symmetric confidence intervals, is in doubt, a problem which does not arise when using MCMC.

For the sake of illustration, we chose to investigate the MCMC method for the analysis of simulated Poisson tumour count data under a simplified toy model, as well as for the analysis of real data of preneoplastic lesions in the rat. In particular, we consider an analysis based on a parametric two-stage model for carcinogenesis [11] with one or more of the model

parameters being heterogeneous between the animals. The MCMC-based technique presented here is, of course, quite general and can be used to investigate the effects of heterogeneity in other types of data, or with different models (see for example reference [6]).

In the following section we briefly describe the basics of Bayesian parameter estimation and of MCMC methods and how heterogeneity may be introduced. In Section 3, we discuss the application of MCMC by means of a simple toy example, comparing MCMC-based inferences with those obtained via the method of maximum likelihood. We illustrate this also by means of a real example of tumour data in Section 4. In Section 5 we deal with the computational issues related to the application of MCMC, detailing the stepwise construction of the sampler, for dealing with heterogeneity in general.

2. PRELIMINARIES

In this section we discuss parameter estimation in general and how heterogeneity between individuals may be incorporated in a model. We also describe some basics of the MCMC method.

2.1. *Parameter estimation*

Let us denote the vector of unknown parameters that we wish to estimate by θ. If the data are contained in the vector y, then it is common practice to estimate θ by maximizing the likelihood $l(y|\theta)$ with respect to θ. The likelihood can be written as

$$l(y|\theta) = \prod_{i=1}^{n} g(y_i|\theta) \tag{1}$$

where n is the number of (independent) animals, y_i denotes the data from the ith animal and $g(y_i|\theta)$ is the contribution of the ith animal to the likelihood. In most practical cases, there are no analytic expressions available for the maximum likelihood estimate (MLE) and its value needs to be determined numerically.

Another way of finding parameter estimates is to use a Bayesian approach. Indeed, often biological considerations give us some ideas, or prior beliefs, about the unknown parameter vector θ, which can be expressed as the so-called prior density $p(\cdot)$ of θ. For instance, a parameter is known to be positive, or a biologically plausible range of values for a parameter is known. In the absence of such prior information, one can choose the corresponding $p(\cdot)$ to be uniform on a relatively wide interval, whereas strong ideas can result in a prior density that is more peaked and/or defined on a narrow interval. Ideas about correlations between the parameters can be taken into account in $p(\cdot)$ as well by having a joint prior density. The main goal of data analysis in a Bayesian context is to investigate the conditional distribution of θ given y, rather than to estimate a single value for θ. This distribution is known as the posterior distribution of the parameter vector, given y, and we shall denote it by $\pi(\cdot|y)$. We could, for instance, consider the expectation of θ under $\pi(\cdot|y)$, which is known as the Bayes estimate of θ with respect to the prior $p(\cdot)$. Prior and posterior distributions are related to each other via the likelihood as

$$\pi(\theta|y) = \frac{l(y|\theta)\,p(\theta)}{\int l(y|\theta)\,p(\theta)\,\mathrm{d}\theta} \propto l(y|\theta)\,p(\theta) \tag{2}$$

where $\propto$ means 'is proportional to'.

Let us now assume that one or more of the parameters are suspected to vary from individual to individual. Let ψ consist of these parameters and let h denote the probability distribution for ψ, which is characterized by an unknown parameter vector η. Values of ψ, assigned to different individuals, are considered independent realizations from this probability distribution h. Let ψ_i denote the realization of ψ for the ith individual. Write $\theta = (\psi, \theta_{-\psi})$, we are interested in estimating $\theta' = (\eta, \theta_{-\psi})$.

In principle, we may obtain the MLE of θ' by maximizing the likelihood

$$l(y|\theta') = \prod_{i=1}^{n} \int_{\psi_i} g(y_i|\psi_i, \theta_{-\psi}) h(\psi_i|\eta)\, \mathrm{d}\psi_i \tag{3}$$

with respect to θ'. When data arise from a biological system requiring complex models, numerical computation of the MLEs may be prohibitive in terms of computing effort. As mentioned above, this is the main reason we investigate the performance of MCMC methods in a Bayesian setting.

We assume priors $p(\theta_{-\psi})$ for $\theta_{-\psi}$ and $p(\eta)$ for η. For the MCMC methods, we consider simulation of the parameter vector $\theta'' = (\psi_1, \ldots, \psi_n, \theta_{-\psi}, \eta)$ from the corresponding posterior distribution $\pi(\theta''|y)$ given by

$$\begin{aligned} \pi(\theta''|y) &= \frac{\prod_{i=1}^{n} g(y_i|\psi_i, \theta_{-\psi}) h(\psi_i|\eta)\, p(\theta_{-\psi})\, p(\eta)}{\int_{\theta''} \prod_{i=1}^{n} g(y_i|\psi_i, \theta_{-\psi}) h(\psi_i|\eta)\, p(\theta_{-\psi})\, p(\eta)\, \mathrm{d}\theta''} \\ &\propto \prod_{i=1}^{n} g(y_i|\psi_i, \theta_{-\psi}) h(\psi_i|\eta)\, p(\theta_{-\psi})\, p(\eta) \end{aligned} \tag{4}$$

In general, it is difficult to calculate the posterior distribution $\pi(\cdot|y)$ analytically. However, if a sample out of $\pi(\cdot|y)$ is available, then a histogram of the sampled values would give us an idea of the shape of the posterior distribution. Then, also the expectation of (any function of) θ in the homogeneous case, or θ'' in the heterogeneous case, under π given y, can be estimated by the mean of (this function of) the sample values. The merit of MCMC methods is that they can generate such a sample for us.

2.2. *Markov chain Monte Carlo*

MCMC is a general method for generating a sample from a probability density known up to a proportionality constant (see references [7–9] for an introduction). We will see in Sections 3 and 4 that in our case the posterior distributions $\pi(\theta|y)$ and $\pi(\theta''|y)$ are only known up to a proportionality constant. Over the last decade, an enormous number of papers has been published on MCMC. For our purpose, a short description of the method with $\pi(\theta|y)$ will suffice.

An MCMC method generates successive values of θ, denoted by $\theta^{(1)}, \theta^{(2)}, \ldots$, from an irreducible Markov chain having π as its stationary distribution. An explicit formula for π is not necessary, as long as it is known up to a proportionality constant. The generated Markov chain serves as a, generally dependent, sample from π and based on this sample inferences from π can be made. The first m_0, say, values of the chain are discarded as the chain may not have reached stationarity by then. This m_0 is called the 'burn-in' period. Information on θ can

be obtained from the sample formed by $\theta^{(m_0+1)}, \ldots, \theta^{(m_0+m)}$, for some m. The general algorithm is the so-called Metropolis–Hastings algorithm [12, 13]; the well-known Gibbs sampler is a special case of this.

Specifically, given the current value $\theta^{(j)}$ of the Markov chain, a *candidate* value of θ, say θ^*, is sampled from some *proposal* distribution with density function $q(\cdot, \theta^{(j)})$, say, which may depend on the current value of the Markov chain. The value θ^* is then accepted with probability

$$A(\theta^*, \theta^{(j)}) = \min\left[1, \frac{\pi(\theta^*|y)q(\theta^{(j)}, \theta^*)}{\pi(\theta^{(j)}|y)q(\theta^*, \theta^{(j)})}\right] \tag{5}$$

If accepted, we set $\theta^{(j+1)} = \theta^*$; otherwise the current value is retained, that is, $\theta^{(j+1)} = \theta^{(j)}$. The usefulness of this Metropolis–Hastings algorithm is to be able to simulate a value from the posterior distribution $\pi(\theta|y)$ when it is either known only up to a proportionality constant or difficult to directly simulate from. The proposal distribution $q(\theta^*, \theta^{(j)})$ and the acceptance probability $A(\theta^*, \theta^{(j)})$ define a transition probability given by $p(\theta^{(j)}, \theta^*) = q(\theta^*, \theta^{(j)}) \times A(\theta^*, \theta^{(j)})$, by which a new value $\theta^{(j+1)}$ is selected given the current value $\theta^{(j)}$. Moreover, the following 'detailed balance' property:

$$\pi(\theta^{(j)}|y)\, p(\theta^{(j)}, \theta^{(j+1)}) = \pi(\theta^{(j+1)}|y)\, p(\theta^{(j+1)}, \theta^{(j)})$$

ensures that the corresponding stationary distribution is the same as $\pi(\theta|y)$; see reference [13] (and also reference [10]) for further details.

Apart from some regularity conditions that are generally met by most practical choices, the proposal distribution can have any form. It is desirable, however, to choose the proposal distribution such that candidate values can be generated quickly and that the acceptance probability can be easily computed. Some specific examples of proposal distributions, in the context of models that incorporate inter-individual variation, are given in Section 5.1. Note that if the proposal distribution is symmetric in θ^* and $\theta^{(j)}$, it will not be required in the above acceptance probability. In practice, the values of $\theta^{(j)}$ need not be updated as a whole; they may also be updated componentwise. The above acceptance probability then involves the corresponding full conditional (see Section 5.1) in place of π and the individual proposal distribution. A nice feature of the MCMC method is that it is generally quite easy to implement. A simple adaptive strategy for MCMC is discussed in Section 5. For a detailed discussion of MCMC methods, we refer to reference [10].

If, as is often the case, one is interested in the posterior distribution of a function $\phi(\theta)$ of θ under $\pi(\cdot|y)$, then the series $\phi(\theta^{(m_0+1)}), \ldots, \phi(\theta^{(m_0+m)})$ will be the focus and its sample distribution is an estimate of the posterior distribution of $\phi(\theta)$. In particular

$$\frac{1}{m}\sum_{j=1}^{m} \phi(\theta^{(m_0+j)}) \to E_{\pi|y}\phi(\theta) \quad \text{as } m \to \infty \tag{6}$$

and we can estimate the posterior expectation $E_{\pi|y}\phi(\theta) = \int \phi(\theta)\pi(\theta|y)\,\mathrm{d}\theta$ of $\phi(\theta)$ by $\bar{\phi}(\theta)_m = \frac{1}{m}\sum_{j=1}^{m} \phi(\theta^{(m_0+j)})$. This indicates how to estimate other moments of $\phi(\theta)$ since these are also the expectations of (different) functions of θ. Taking $\phi(\theta) = \theta_l$, the lth component of θ, we see that the marginal posterior distribution of θ_l can be investigated by considering the sample $\theta_l^{(m_0+1)}, \ldots, \theta_l^{(m_0+m)}$. It is customary to consider the histograms of these individual

parameter values (or functions thereof) giving estimates of the corresponding marginal posterior distributions and report the corresponding means or credible intervals. Note that in a Bayesian framework, the $100(1-p_\alpha)$ per cent credible interval for a parameter is defined by the two points having $p_\alpha/2$ and $1-p_\alpha/2$ probability, respectively, to the left of them in the corresponding marginal posterior distribution.

3. AN ILLUSTRATION WITH OVERDISPERSED POISSON COUNTS

Before applying the proposed MCMC simulation strategy to real data, we illustrate its use on simulated data. We chose to use overdispersed Poisson counts where the overdispersion is due to heterogeneity in the Poisson mean. The main interest here is to see how well the proposed strategy can reproduce the assumed heterogeneity, especially in situations when sample size is moderate and when the degree of heterogeneity is comparable to the model variance. For this purpose, we draw counts from a Poisson distribution with a mean that is randomly distributed. More specifically, we assume that the 'Poisson mean' factors into a dose-independent heterogeneous part (Λ) and a linear dose–response function of dose d that is homogeneous across individuals. Therefore, the 'Poisson mean' is given by $\Lambda \times (1 + \delta \times d)$ with Λ following a gamma or log-normal distribution and δ being the dose-related homogeneous parameter.

We would like to draw inferences with regard to the distribution of Λ (represented by its mean μ_Λ and its variance σ^2_Λ) and simultaneously estimate the slope parameter δ. In the notation of the preceding sections we have

$$\begin{aligned} \theta &= (\Lambda, \delta) \\ \psi &= \Lambda \\ \eta &= (\mu_\Lambda, \sigma^2_\Lambda) \\ \theta' &= (\mu_\Lambda, \sigma^2_\Lambda, \delta) \\ \theta'' &= (\Lambda_1, \ldots, \Lambda_n, \delta, \mu_\Lambda, \sigma^2_\Lambda) \end{aligned} \tag{7}$$

and we want to estimate θ'. For the priors of the components of θ' we assume independent uniform distributions with $0 < \theta' < D$, where the components of D are large (but fixed) positive numbers. For the example described here we use $D = (100, 100, 10)$. With this particular choice no rejections occurred in our simulations on grounds of a proposed value of θ' being larger than D.

A number of simulations were carried out with different choices for μ_Λ and σ_Λ, and with different numbers of animals. Since they consistently gave a similarly good agreement between MLEs and MCMC-based estimates, we only report the results of one (typical) simulated data set. For this simulation we assume four dose groups (doses 0, 0.1, 0.5 and 1.0) each with 50 animals, a unity slope parameter ($\delta = 1$), $\mu_\Lambda = 40$ and $\sigma_\Lambda = 4$. Although no general conclusions can be drawn, the case by case comparisons of estimates obtained using the MCMC method with those obtained via maximum likelihood for all simulations provide an indication of how well inferences with respect to the amount of heterogeneity predicted by these methods agree with one another.

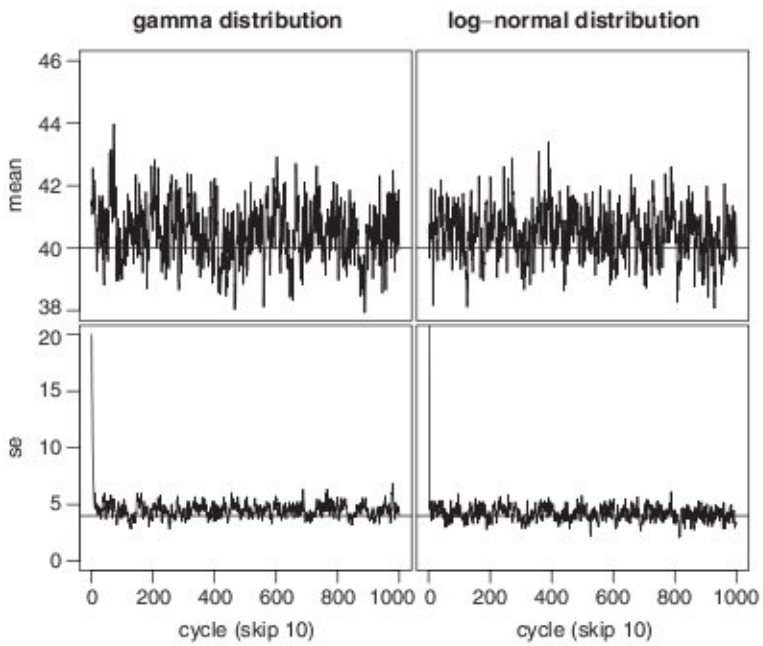

Figure 1. MCMC-based analysis of overdispersed Poisson data (see text). The mean of the gamma distributed Poisson mean μ_Λ was assumed to be 40 with a standard error of 4. The Markov chain was run for 10000 cycles.

Following the procedure outlined above, we generate MCMC samples from the posterior distribution of θ' by discarding the $(\Lambda_1,\ldots,\Lambda_n)$ part in the MCMC samples for θ''. In the case of the log-normal distribution, we use a multivariate normal proposal distribution for μ_Λ, σ_Λ and δ. For the gamma distribution, however, as explained in Section 5, we use a multivariate normal proposal distribution in terms of specific combinations of the shape and scale parameters a and b of the gamma distribution and δ. The respective covariance matrices for these two proposal distributions are determined adaptively as described in Section 5. Typical MCMC simulations using both gamma and log-normal distributed heterogeneity models for Λ are shown in Figure 1 (see also Table I). We see that both models reproduce the assumed heterogeneity in the respective data set well. However, when the assumed heterogeneity in simulated data is weak, that is, when σ_Λ is much smaller than the Poisson error ($\sqrt{\mu_\Lambda}$), both methods perform equally poorly (results not shown), although the estimates usually improve when the number of animals in the experiment is increased.

Since we use uniform priors on sufficiently large intervals for all parameters, Bayesian methods should give approximately the MLE at the mode of the joint posterior distribution. However, it is difficult to estimate the above mode based on the chain of values for the parameter vector as produced by the MCMC method. As mentioned in Section 2.2, we report only estimates of the marginal means with corresponding credible intervals. For the particular example described here with a gamma heterogeneity distribution, it is possible to obtain the likelihood in closed form. The individual likelihood contributions are integrals over the

Table I. Comparison of MLE- and MCMC-based parameter estimates (means and 95 per cent confidence or credibility regions) for simple Poisson count data with gamma-distributed Poisson-mean with mean 40, standard deviation 4 and with slope parameter (δ) 1. See text for more details. For the MCMC-based estimation a model with gamma distributed heterogeneity and a model with log-normally distributed heterogeneity were used. The MLEs were computed using the 'exact' integrations given in (8).

Parameter	Maximum likelihood	MCMC	
	Gamma	Gamma	Log-normal
μ_Λ	40.65 (39.05, 42.26)	40.45 [38.78, 42.11]	40.46 [39.03, 41.92]
σ_Λ	4.37 (3.39, 5.62)	4.47 [3.19, 5.56]	4.19 [2.96, 5.41]
δ	0.93 (0.82, 1.05)	0.94 [0.82, 1.08]	0.92 [0.82, 1.04]

parameter Λ. Let y_i be the observed count and d_i the dose for individual i. Defining $R_i \equiv (1 + \delta \times d_i)$, the integral in (3) becomes

$$\int_{\psi_i} g(y_i|\psi_i, \theta_{-\psi}) h(\psi_i|\eta)\, \mathrm{d}\psi_i = \int_0^\infty \mathrm{Pois}(y_i; \Lambda_i R_i)\gamma(\Lambda_i; a, b)\, \mathrm{d}\Lambda_i$$
$$= \frac{b^a}{(R_i + b)^{a+y_i}} \frac{\Gamma(a + y_i)}{\Gamma(a)\Gamma(y_i + 1)} R_i^{y_i} \qquad (8)$$

where $\mathrm{Pois}(y; \mu)$ denotes the probability of observing y under a Poisson distribution with mean μ, $\gamma(y; a, b)$ the density in y of a gamma distribution with shape parameter a and scale parameter b, and $\Gamma(\cdot)$ the usual gamma function. The parameters a, b and δ can then be estimated via maximum likelihood, and the mean and variance of Λ_i are given by $\mu_\Lambda = a/b$ and $\sigma_\Lambda^2 = a/b^2$, respectively.

Table I gives the MLEs and MCMC-based estimates for the mean and the standard error of the embedded heterogeneity distribution, and the estimates of the slope parameter δ for a typical data set generated with the model and the specific values as described above. The agreement between MLEs and MCMC-based estimates for the mean and the standard error of the embedded heterogeneity distribution is excellent. Also, for this particular case, both gamma and log-normal models provide very similar MCMC-based estimates. Although we cannot formally generalize these results to other situations, our simulations suggest that, given a particular model and sufficient data generated from it, the agreement between MLEs and MCMC-based estimates is usually remarkably good.

4. MCMC FOR THE ANALYSIS OF PRENEOPLASTIC LESIONS

We now illustrate the MCMC approach for incorporating inter-individual variation using real data. The data of interest are preneoplastic liver lesions, their number and sizes, in rats that were exposed chronically to a putative liver carcinogen, N-nitrosomorpholine (NNM). The purpose of the experiment was to obtain information about the mode of action of NNM

and to assess the carcinogenic potency of NNM. Before we discuss specific aspects of the analysis, we describe the experiment and discuss the model that is used for the analysis of the data.

Groups of female Lewis rats (age 14 weeks) were exposed to various doses of NNM in their drinking water. Within each dose group animals were sacrificed at different time points. A histological section from the liver of each animal was obtained and was examined for ATPase deficient lesions. These lesions are considered precursor lesions for hepatocellular carcinoma in the rat. Thus, for each animal, the data consist of the number and the sizes of ATPase deficient (two-dimensional) transections in each liver section. We make use of the first three dose groups only: 0 ppm, 0.1 ppm, and 1 ppm (excluding the 5, 10, 20 and 40 ppm groups). Our analysis therefore includes only 89 animals with a total of 372 transections. This choice is made for the sake of simplicity and allows us to ignore dose–response related heterogeneity. Individual sizes, although recorded in terms of transection areas, were converted into radii because the transections were mostly circular. The nature of these two-dimensional observations poses a stereological problem which is addressed using the method described in our earlier analysis [1].

In order to make inferences about the number and size distributions of these lesions, we employ a stochastic model that describes the initiation and clonal expansion processes of altered cells. We use the two-stage clonal expansion carcinogenesis model [11] which yields mathematical expressions for the number and the size distribution of such intermediate lesions on the pathway to malignancy [14]. These expressions depend on three biological parameters: the cell division rate α, the cell death rate β, and the rate of initiation ν of altered cells in a specific volume of normal tissue.

In general we may assume the parameters to be piecewise constant on $(0,t]$, that is, we assume constant values for the parameters α, β and ν on the k intervals $(\tau_0,\tau_1],(\tau_1,\tau_2]\ldots(\tau_{k-1},\tau_k]$, where $\tau_0=0$ is the time of birth, and $\tau_k=t$ is the time of observation. That is, the parameters α, β and ν take values α_j, β_j and ν_j, respectively, in the jth interval. The number and size (in terms of the number of constituent cells) distributions of non-extinct lesions are given in reference [15]. We summarize them here. The number of non-extinct premalignant clones is Poisson distributed with mean $\Lambda(t)$, given by

$$\Lambda(t)=\sum_{j=1}^{k}\frac{\nu_j}{\alpha_j}\log\left[\frac{\alpha_j e^{\{(\alpha_j-\beta_j)(\tau_j-\tau_{j-1})\}}-\tilde{\beta}_j}{\alpha_j-\tilde{\beta}_j}\right] \tag{9}$$

The probability $p(m,t)$ that at time t a non-extinct lesion contains m cells $(m>0)$ satisfies

$$p(m,t)=\frac{1}{\Lambda(t)}\sum_{j=1}^{k}P_j(m) \tag{10}$$

with $\Lambda(t)$ as in (9) and $P_j(m)$ given by

$$P_j(m)=\frac{1}{m}\left\{\left(\frac{\nu_j}{\alpha_j}-\frac{\nu_{j-1}}{\alpha_{j-1}}\right)\frac{\alpha_j-\tilde{\alpha}_j e^{\{-(\alpha_j-\beta_j)(\tau_j-\tau_{j-1})\}}}{\alpha_j-\tilde{\beta}_j e^{\{-(\alpha_j-\beta_j)(\tau_j-\tau_{j-1})\}}}\right\}^m,\quad j=1,\ldots,k \tag{11}$$

with $\nu_0/\alpha_0 = 0$. In (11) the $\tilde{\alpha}_j$'s are defined recursively by

$$\begin{aligned} \tilde{\alpha}_k &= \alpha_k \\ \tilde{\alpha}_j &= \alpha_j - \frac{(\alpha_j - \beta_j)}{(\alpha_{j+1} - \beta_{j+1})}(\alpha_{j+1} - \tilde{\alpha}_{j+1} e^{\{-(\alpha_{j+1}-\beta_{j+1})(\tau_{j+1}-\tau_j)\}}), \quad j = k-1, \ldots, 1 \end{aligned} \tag{12}$$

and in (9) and (11) the $\tilde{\beta}_j$'s are defined by

$$\begin{aligned} \tilde{\beta}_k &= \beta_k \\ \tilde{\beta}_j &= \tilde{\alpha}_j - (\alpha_j - \beta_j) \prod_{l=j+1}^{k} e^{\{-(\alpha_l - \beta_l)(\tau_l - \tau_{l-1})\}}, \quad j = k-1, \ldots, 1 \end{aligned} \tag{13}$$

In contrast to the earlier analyses of NNM data [1, 2], where the time interval for the analysis began at 14 weeks (after birth) when NNM was first administered, we now include the time interval from birth to start of the experiment. This refinement not only yields better data fits, but also yields more plausible cell division rates for the intermediate cells. Also, rather than directly estimating the cell division rate α and the cell death rate β, we estimate α together with the ratio β/α, which equals the asymptotic probability of extinction of an intermediate lesion. In our experience, this reparameterization improves the numerical convergence of MLE searches.

In preliminary analyses, we allow each model parameter (that is, α, β/α and ν) to assume different values before and after the start of the experiment. Thus, we consider two time intervals: $(0, \tau_1]$ and $(\tau_1, t]$ with $\tau_1 = 14$ weeks. Likelihood ratio tests reveal that only the two estimates of α on the two time intervals differ significantly from one another while the parameters β/α and ν are found to be similar for the two time intervals. The cell division rate in the liver lesions appears strongly increased in young animals up to 14 weeks of age.

The explicit inclusion of α_1 (the cell division rate before time τ_1) and of α_2 (the cell division rate after time τ_1) in our MCMC simulations is straightforward. Preliminary MCMC runs, however, indicate that much longer Markov chains are needed before stationarity is reached in this case. These runs also show that α_1 and α_2 are strongly correlated and mix poorly in the simulation. Better mixing of highly correlated parameters is usually achieved by suitable reparameterizations [10]. However, for the sake of simplicity, and to keep the focus of our illustration on the modelling of inter-individual variability, we fix the value of α in the period before the start of experiment at the MLE found. Specifically, we assume $\alpha_1 = 0.158$ per day (fixed) and $\alpha_2 = \alpha$ (unknown).

Because the lesions and their sizes can only be observed on histological sections, we apply the Wicksell method to transform the discrete size distribution into one for circular transection profiles. Similarly, we use the Fullman formula to transform the Poisson mean for the number of lesions in three-dimensions to the Poisson mean for the number of transections on two-dimensional sections. Because of this stereological problem, the likelihood function is complicated and the MLEs are determined numerically using general purpose optimization programs. See references [1, 3, 15] for more details.

To begin with, we assume that all parameters are homogeneous. Table II lists the maximum likelihood estimates and corresponding MCMC-based estimates of the marginal posterior means for the above model. The agreement between the MLE- and MCMC-based estimates again is excellent. Next, we allow ν to be heterogeneous assuming a gamma or log-normal

Table II. Comparison of maximum likelihood and MCMC-based estimates (means and 95 per cent confidence or credibility regions) assuming homogeneity.

Parameter	MLE	MCMC (10000 cycles)
α	0.046 (0.035, 0.061)	0.048 [0.036, 0.068]
β/α	0.844 (0.800, 0.885)	0.849 [0.800, 0.895]
ν	0.769 (0.614, 0.957)	0.787 [0.630, 1.040]

Table III. Comparison of maximum likelihood and MCMC-based estimates (means and 95 per cent confidence or credibility regions) assuming heterogeneity in the rate of initiation ν. MLE-based estimates were only obtained for gamma distributed ν. See text for details.

Parameter	MLE (gamma)	MCMC (gamma)	MCMC (log-normal)
α	0.043 (0.034, 0.054)	0.044 [0.023, 0.072]	0.048 [0.037, 0.072]
β/α	0.83 (0.79, 0.87)	0.82 [0.70, 0.90]	0.85 [0.81, 0.89]
μ_ν	0.70 (0.49, 0.99)	0.68 [0.35, 1.12]	1.03 [0.59, 1.88]
σ_ν	0.94 (0.64, 1.36)	1.01 [0.50, 1.76]	3.57 [1.18, 10.68]

distribution for the ν values of different individuals, as in Section 3. We then compare estimates of the parameters using maximum likelihood via exact integrations of the individual likelihood contributions (see (8)) with corresponding MCMC-based estimates (see Table III). Since closed form solutions for the integrals are not available for log-normal distributions of ν, maximum likelihood estimation is difficult in that case. However, note that MCMC simulations can easily be carried out. Figure 2 shows the obtained MCMC (posterior) samples for the case where ν has a log-normal distribution. The simulation was run for 20000 cycles. For simplicity we display only every 10th MCMC cycle. During the adaptive phase (see Section 5.2) the samples are marked by dots.

Finally, we consider the case when all parameters α, β/α and ν are associated with heterogeneity. Comparisons of multiple MCMC runs suggest that much longer chains are needed than for the univariate case to obtain reliable estimates of marginal quantiles, especially for the quantiles in the tail region of the (marginal) posterior distributions. We generated two MCMC simulations, each with 100000 cycles. Both simulations yield very similar results. The estimated coefficient of variation (based on the MCMC samples) for the parameters ν, α and β/α are, respectively, 1.53, 0.23 and 0.02. This indicates that the heterogeneity associated with ν is the strongest, while the parameter α is only moderately heterogeneous and β/α does not appear to be significantly associated with inter-individual variability.

We have made no attempt to maximize the likelihood for this case by integrating numerically (say, via three-dimensional Gauss-quadrature) the individual likelihood contributions. An attempt to do so may require lengthy computations. For example, using just 20 Gauss mesh points in each dimension to numerically integrate the individual likelihood contributions, and over 100 function calls for the optimization procedure, more than $100 \times 20^3 = 800000$ likelihood evaluations would be required to estimate the MLEs. However, frequently many more function calls are needed to check for other modes and to assess the stability of the MLEs found. Furthermore, the numerical integrations may not readily yield stable results and may require additional searches for an optimal placement of integration mesh points. Also, fixing

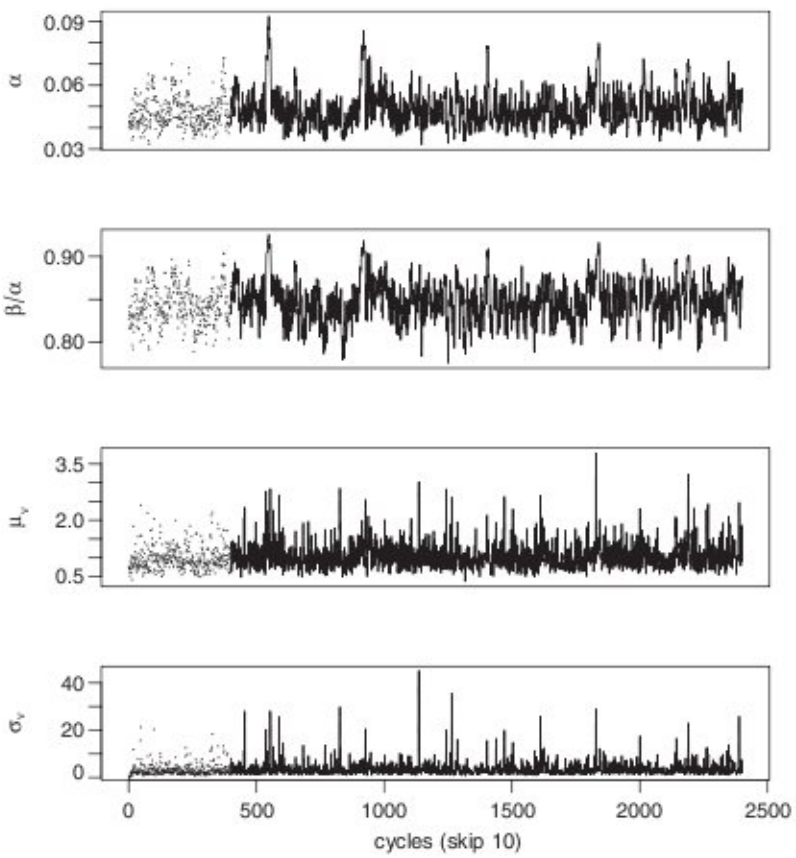

Figure 2. MCMC (posterior) samples for the NNM data assuming the rate of initiation v to be log-normally distributed. Also shown (dotted) are samples obtained during the adaptive phase.

the integration ranges in advance is problematic because the location of the MLEs is usually not known in advance. This problem can, at least in principle, be solved by use of an adaptive integration scheme.

In contrast, MCMC simulations have the advantage that they can be continued indefinitely, given the values of the current cycle and the random seed(s) of the sampler, while numerical integrations that do not have sufficient accuracy are usually discarded. Thus, the MCMC method provides a useful alternative to MLE based methods for exploring heterogeneity in complex models, especially in situations when integrations are complicated or when the integrals have to be evaluated numerically.

5. APPLICATION OF MCMC

Here we describe the general construction of an MCMC sampler designed to generate samples from posterior distributions with complicated likelihood functions. This sampler makes use of

the Metropolis–Hastings algorithm [7–10, 12, 13] to sample from the *full conditionals*. When these full conditionals are completely specified, samples can be drawn directly via Gibbs sampling, which is computationally simpler than the Metropolis–Hastings algorithm.

5.1. *Full conditionals*

Application of the Metropolis–Hastings algorithm requires specification (up to a constant multiplier) of the full conditionals of individual parameters or a block of parameters to be sampled simultaneously. The full conditional for a parameter is given (like a posterior distribution) by the conditional distribution of the parameter given the data y and the rest of the parameters. The full conditionals are usually available only up to the normalizing constant. In this case the Metropolis–Hastings algorithm can be used. It involves sampling from a suitable 'proposal' distribution and then accepting or rejecting the sampled value with certain probability. See references [7–10] for more details.

To be specific, let $\theta=(\theta_1,\ldots,\theta_K)$ denote the parameter vector with $l(y|\theta)$ being the likelihood and $p(\theta)$ the prior distribution for θ. Then, the joint posterior distribution is given by $\pi(\theta|y) \propto l(y|\theta)p(\theta)$. To begin with, assume that all parameters are homogeneous and that the prior distributions for the θ_k's are independent (that is, $p(\theta)=\prod_{k=1}^{K} p_k(\theta_k)$). Then, the full conditional for θ_k is given by

$$\pi_k(\theta_k|y,\theta_{-k}) \propto l(y|\theta)p_k(\theta_k)$$

for $k=1,\ldots,K$, where θ_{-k}, as usual, denotes the vector θ without θ_k. While θ can be updated componentwise using the full conditionals, we sometimes prefer to update the components of θ as a block in which case the corresponding full conditional is $\pi(\theta|y)$. This is done using a multivariate normal as the proposal distribution. A detailed description of the construction of such a proposal distribution is given in Section 5.2.

Next, we assume ψ, consisting of some of the θ_k's, to be associated with heterogeneity. For simplicity, let ψ be scalar containing only one of the θ_k's. We will consider two commonly used distributions for heterogeneity. First, let ψ be gamma distributed with shape parameter a and scale parameter b. Thus, for the ith subject, ψ_i is a realization from this distribution. The second distribution for ψ is log-normal with parameters μ and σ^2. Thus, $\log\psi_i$ is normally distributed with expectation μ and variance σ^2 in this case. As in Section 3, let $g(y_i|\psi_i,\theta_{-\psi})$ denote the likelihood contribution for subject i with observation y_i. Note that $\theta_{-\psi}$ is the vector of parameters that does not contain ψ. The full conditional for ψ_i is then given by

$$\pi(\psi_i|y,\theta_{-\psi},\psi_{-i},a,b) \propto \gamma(\psi_i;a,b)g(y_i|\psi_i,\theta_{-\psi})$$

for the gamma distribution and

$$\pi(\psi_i|y,\theta_{-\psi},\psi_{-i},\mu,\sigma) \propto \rho(\psi_i;\mu,\sigma)g(y_i|\psi_i,\theta_{-\psi})$$

for the log-normal distribution where ψ_{-i} is the vector $(\psi_1,\ldots,\psi_{i-1},\psi_{i+1},\ldots,\psi_n)$, $\gamma(\cdot;a,b)$ is the density of the gamma distribution and $\rho(\psi_i;\mu,\sigma)$ is the density for the log-normal distribution. Note that these full conditionals depend on neither ψ_{-i} nor the priors of $(\theta_{-\psi},a,b)$ (or $(\theta_{-\psi},\mu,\sigma)$).

Generating samples from these full conditionals is difficult because the likelihood contribution g is not of a simple form. However, as mentioned before, the Metropolis–Hastings algorithm can be used for this purpose. Calculation of the Metropolis acceptance probabilities is usually simplified when the assumed heterogeneity distributions are used as proposals. In practice, the ψ_i's may be updated as a block.

With uniform prior for $\theta_{-\psi}$, the corresponding full conditional can be seen to be

$$\pi(\theta_{-\psi}|y,\psi_1,\ldots,\psi_n,\eta) \propto \prod_{i=1}^{n} g(y_i|\psi_i,\theta_{-\psi})$$

where η, using the notation of Section 2.1, is either (a,b) for the gamma heterogeneity distribution or (μ,σ) for the log-normal. For the sampling of $\theta_{-\psi}$, we also need to employ the Metropolis–Hastings algorithm. Here too we use a multivariate normal distribution as the proposal (see Section 5.2).

For computational convenience we reparameterize the gamma distribution in terms of the inverse mean $c=b/a$ and the shape parameter a. The gamma density is then given by

$$\gamma(\psi_i;a,ac) = a\Gamma^{-1}(a)(a\psi_i)^{a-1}c^a \exp(-\psi_i ac)$$

The full conditional for c (assuming a uniform prior for c) can now be obtained as

$$\pi(c|y,\theta_{-\psi},\psi_1,\ldots,\psi_n,a) = \gamma\left(c;na+1,\sum_{i=1}^{n}(a\psi_i)\right)$$

where the right-hand side is, as before, the gamma density with shape parameter $(na+1)$ and scale parameter $\sum_{i=1}^{n}(a\psi_i)$. Hence, successive values of c can be sampled directly using Gibbs sampling.

For the shape parameter a, the full conditional is given by

$$\pi(a|y,\theta_{-\psi},\psi_1,\ldots,\psi_n,c) \propto \prod_{i=1}^{n}\gamma(\psi_i;a,ac)$$

Therefore, we need to use the Metropolis–Hastings algorithm for sampling a. We use a uniform distribution for the proposal.

For the log-normal distribution, we sample the parameters μ and $\xi=1/\sigma^2$. Again, assuming that these two parameters have uniform priors, we get the full conditional for μ as

$$\pi(\mu|y,\theta_{-\psi},\psi_1,\ldots,\psi_n,\xi) = \phi\left(\mu;\sum_{i=1}^{n}\log(\psi_i)/n,\sigma^2/n\right)$$

where $\phi(\cdot;\mu_0,\sigma_0^2)$ denotes the normal density function with mean μ_0 and variance σ_0^2. Similarly, the full conditional for ξ is

$$\pi(\xi|y,\theta_{-\psi},\psi_1,\ldots,\psi_n,\mu) = \gamma\left(\xi;n/2+1,\sum_{i=1}^{n}(\log(\psi_i)-\mu)^2/2\right)$$

Thus, with this parameterization, and with uniform priors, both the parameters for the log-normal distribution can be sampled by Gibbs sampling.

5.2. *MCMC steps*

Generally, one can sample successive values of $(\psi_1, \ldots, \psi_n, \theta_{-\psi}, \eta)$ using the full conditionals of Section 5.1 and some initial values that may be arbitrary. We, however, follow a specific strategy as described below. Our strategy is a modification of the 'adaptive kernel switching' approach to achieve MCMC acceleration proposed by Gelfand and Sahu [16]. The idea is first to carry out a few pilot runs that will provide an estimate of the correlations between the successively sampled parameter values. As a modification, we consider estimating correlations between the successive increments as explained later. These estimates can be used to define a multivariate normal distribution as the proposal distribution for $\theta_{-\psi}$. This approach attempts to improve 'mixing' by anticipating the correlations among parameters as the Markov chain progresses.

In order to initiate the process, we first assume homogeneity in the whole of $\theta = (\psi, \theta_{-\psi})$. If the maximum likelihood estimate of θ can be determined, it can be used as an initial estimate for the MCMC. Otherwise, one can start with an arbitrary estimate. We then carry out the following steps.

1. *Pilot run* 1. Using the full conditionals for the homogeneous case (as given in the beginning of Section 5.1) and uniform priors for the components of θ, sample m_1 values of θ. Use the Metropolis–Hastings algorithm for this purpose with a multivariate normal distribution (with the current value as the mean and the negative Hessian matrix evaluated at the initial value as the covariance matrix) as the proposal.
2. Compute the sample covariance matrix of the successive increments of these m_1 values of θ. Denote it by Σ_1.
3. *Pilot run* 2. Sample m_2 values of θ as in step 1 above, but with the covariance matrix for the proposal distribution replaced by $s\Sigma_1$, where s is a suitable scale factor ranging between 2 and 3. As in step 2 above, compute the sample covariance matrix of the successive increments of these m_2 values of θ. Denote it by Σ_2.
4. *Final MCMC run.* The heterogeneity parameters are brought in at this stage. As the initial values of $(\psi_1, \ldots, \psi_n, \theta_{-\psi}, \eta)$ for the final run, take ψ_i (for all $i = 1, \ldots, n$) equal to the last (m_2th) value of ψ in step 3 above, for $\theta_{-\psi}$ take its last value, and some arbitrary value for η. Now carry on the MCMC sampling using the full conditionals of Section 5.1. For sampling $\theta_{-\psi}$, the proposal distribution is multivariate normal with its current value as the mean and s times the corresponding submatrix of Σ_2 as the covariance matrix.

A few comments are in order. The sample covariance matrix, as in steps 2 and 3, is based on the successive increments since they are likely to be less autocorrelated and, at least at the initial stage, have similar distributions. Excessive rejection in the Metropolis–Hastings algorithm can be avoided by reducing the scale s in step 3. Recommended acceptance rates range between 30 and 70 per cent [10]. The lengths of the two pilot runs (m_1 and m_2, respectively) depend on the particular problem. Our experience with the cancer data analysed here suggest that at least several hundred cycles are needed. Moreover, convergence of the chain may improve when additional pilot runs are generated. The final run (step 4) may be monitored using diagnostic tools, such as the GIBBSIT program [17], in order to determine a sufficient number of cycles for computing expectations under the posterior. We also recommend the

generation of several independent MCMC simulations to better gauge the effects of 'burn-in' and to better assess the convergence of the chains.

6. CONCLUDING REMARKS

Because the biological parameters in our models are positive and in general bounded from above by physiological limits, one may often have some information on their prior distributions. However, in the absence of any such knowledge, we find it natural to work with uniform prior distributions (see Sections 3–5).

The MCMC technique, as an alternative to maximum likelihood estimation, has a computational advantage, specially when dealing with between-individual heterogeneity (see Sections 3 and 4). The MCMC approach to incorporate heterogeneity, although described for tumour data in particular, is also applicable to other types of data and different models for analysing them (see reference [6], for example). The MCMC method can also deal with within-individual heterogeneity, which will make the maximum likelihood calculation almost intractable. For example, in our tumour data problem of Section 4, there may be variation in the rate parameters for different tumours within a tissue depending on their location.

Although we suggest a rather generic strategy for the implementation of the MCMC method (see Section 5.2), one can work with different ones that are more suitable for the problem of interest. However, our implementation works well for a variety of models and data sources, and usually leads to rapid convergence and well 'mixed' Markov chains.

The MCMC method also allows one to work with incomplete data (common in biological problems) through the 'data augmentation' technique. The idea is to sample the 'missing' part of the data as well in addition to the different parameters in each MCMC cycle. In the context of our problem in Section 4, the number and size distributions of the preneoplastic lesions in three dimensions forms the missing part. By sampling these data also (from the corresponding full conditionals), one can work with a simpler likelihood avoiding the stereological complications (see reference [1]).

ACKNOWLEDGEMENTS

We thank Dr Julian Besag (University of Washington, Seattle) for introducing the MCMC method to us, and Dr Michael Schwarz (University of Tübingen) for making the NNM data available to us. We also thank Dr Suresh Moolgavkar (Fred Hutchinson Cancer Research Center, Seattle) for his helpful comments, and an anonymous referee for useful suggestions resulting in an improved paper. We acknowledge support from NIH grants RO1 CA47658, PO1 CA76466, and travel grants awarded to AD and EGL from the Dutch Organization for Scientific Research (NWO).

REFERENCES

1. Moolgavkar SH, Luebeck EG, de Gunst MCM, Port RE, Schwarz M. Quantitative analysis of enzyme-altered foci in rat hepatocarcinogenesis experiments. I. Single agent regimen. *Carcinogenesis* 1990; **11**:1271–1278.
2. Dewanji A, Luebeck EG, Moolgavkar SH. A biologically based model for the analysis of premalignant foci of arbitrary shape. *Mathematical Biosciences* 1996; **135**:55–68.
3. De Gunst MCM, Luebeck EG. A method for parametric estimation of the number and size distribution of cell clusters from observations in a section plane. *Biometrics* 1998; **54**:100–112.
4. Gilks WR, Wang CC, Yvonnet B, Coursaget P. Random-effects models for longitudinal data using Gibbs sampling. *Biometrics* 1993; **49**(2):441–453.

5. Gilks WR, Clayton DG, Spiegelhalter DJ, Best NG, McNeil AJ, Sharples LD, Kirby AJ. Modelling complexity: application of Gibbs sampling in medicine. *Journal of the Royal Statistical Society, Series B* 1993; **55**:39–52.
6. Dunson DB, Dinse GE. Distinguishing effects on tumour multiplicity and growth rate in chemoprevention experiments. *Biometrics* 2000; **56**(4):1068–1075.
7. Besag J, Green PJ, Higdon D, Mengersen K. Bayesian computation and stochastic systems. *Statistical Science* 1995; **10**:3–41.
8. Geyer CJ. Practical Markov chain Monte Carlo (with discussion). *Statistical Science* 1992; **7**:473–511.
9. Smith AFM, Roberts GO. Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society, Series B* 1993; **55**:3–24.
10. Gilks WR, Spiegelhalter DJ, Richardson S. *Markov Chain Monte Carlo in Practice*. Chapman and Hall: London, 1996; 118–119.
11. Moolgavkar SH, Luebeck EG. Two-event model for carcinogenesis: biological, mathematical, and statistical considerations. *Risk Analysis* 1990; **10**:323–341.
12. Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. Equation of state calculations by fast computing machine. *Journal of Chemical Physics* 1953; **21**:1087–1091.
13. Hastings WK. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* 1970; **57**:97–109.
14. Dewanji A, Venzon DJ, Moolgavkar SH. A stochastic two-stage model for cancer risk assessment. II. The number and size of premalignant clones. *Risk Analysis* 1989; **9**:179–187.
15. Luebeck EG, de Gunst MCM. A stereological method for the analysis of cellular lesions in tissue sections using 3-dimensional cellular automata. In *Mathematical and Computer Modelling*, Yakovlev AY, Moolgavkar SH (eds). Elsevier: 2001; 1387–1400.
16. Gelfand AE, Sahu SK. On Markov chain Monte Carlo acceleration. *Journal of Computer Graphics and Statistics* 1994; **3**:261–276.
17. Raftery AE, Lewis SM. Implementing MCMC. In *Markov Chain Monte Carlo in Practice*, Gilks WR, Spiegelhalter DJ, Richardson S (eds). Chapman and Hall: London, 1996; 118–119.