Appl. Statist. (1992)
41, *No.* 2, *pp.* 337–348

Adaptive Rejection Sampling for Gibbs Sampling

By W. R. GILKS†

Medical Research Council Biostatistics Unit, Cambridge, UK

and P. WILD

National Institute for Research and Safety in Occupational Health, Vandoeuvre, France

[Received March 1990. Final revision February 1991]

SUMMARY

We propose a method for rejection sampling from any univariate log-concave probability density function. The method is adaptive: as sampling proceeds, the rejection envelope and the squeezing function converge to the density function. The rejection envelope and squeezing function are piecewise exponential functions, the rejection envelope touching the density at previously sampled points, and the squeezing function forming arcs between those points of contact. The technique is intended for situations where evaluation of the density is computationally expensive, in particular for applications of Gibbs sampling to Bayesian models with non-conjugacy. We apply the technique to a Gibbs sampling analysis of monoclonal antibody reactivity.

Keywords: Adaptive rejection sampling; Bayesian inference; Gibbs sampling; Log-concave density; Non-conjugate Bayesian models; Simulation

1. Introduction

We present a black box technique for sampling from any univariate log-concave probability density function $f(x)$. Our method is based on rejection sampling and does not require a determination of the mode of $f(x)$. It is adaptive: the envelope function and the squeezing function (which form upper and lower bounds to $f(x)$) converge to the density $f(x)$ as sampling proceeds. The envelope and squeezing functions are piecewise exponentials. The adaptive nature of our technique enables samples to be drawn with few evaluations of $f(x)$; it will therefore be useful in situations where the evaluation of $f(x)$ is computationally expensive. We describe adaptive rejection sampling in Section 2.

Although not contained in Devroye (1986), adaptive rejection sampling has some points of contact with methods of Devroye (1986). In particular, in chapter 4.5, Devroye (1986) discusses rejection sampling using a sequence of envelope and squeezing functions which converge to the density; however, the approach is not adaptive as the bounding functions are determined in advance. In chapter 7.2,

†*Address for correspondence*: Medical Research Council Biostatistics Unit, Fair View Lodge, 5 Shaftesbury Road, Cambridge, CB2 2BW, UK.

 0035–9254/92/41337 $2.00

Devroye (1986) discusses non-adaptive black box methods for log-concave densities which require the location of the mode of the density; in particular he presents an algorithm for rejection sampling using a piecewise exponential envelope comprising three pieces, the centre piece touching the density at its mode. In chapter 7.3, Devroye (1986) discusses non-adaptive methods which take advantage of particular properties of unimodal densities. An exercise in chapter 8.2 of Devroye (1986) concerns an adaptive rejection sampling algorithm in which the rejection envelope is a histogram.

Adaptive rejection sampling grew out of an analysis of patterns of reactivity of monoclonal antibodies. Our data for this analysis were in the form of percentages, rounded to the nearest integer. This discretization would not have been a problem except that much of the data were concentrated at or close to the extremes of the percentage scale. The complexity of our model and data led us to a Gibbs sampling analysis (Geman and Geman, 1984), for which we needed to be able to sample efficiently from densities of complicated algebraic form. In Section 4 we present an analysis of these data using adaptive rejection sampling and Gibbs sampling.

Gibbs sampling is a Markovian updating scheme originally developed by Geman and Geman (1984) as a tool for image reconstruction. However, its applicability to statistical modelling has recently been demonstrated (Gelfand and Smith, 1990). The enormous potential of Gibbs sampling in complex statistical modelling is now being realized. Applications currently include Bayesian cluster analysis (Gilks *et al.*, 1989), changepoint problems (Carlin *et al.*, 1991), genetic linkage analysis (Mack *et al.*, 1990; Thomas, 1991), model selection from normal scale mixture densities (Carlin and Polson, 1991a), influence diagnostics (Carlin and Polson, 1991b), hierarchical models, variance components and errors in variables (Gelfand and Smith, 1990), missing data, ordered means and growth curve models (Gelfand *et al.*, 1990), outlier detection (Verdinelli and Wasserman, 1991), generalized linear models with random effects (Zeger and Karim, 1991) and analysis of frailty in survival analysis (Clayton, 1991). We are also aware of many other applications of Gibbs sampling, currently in the form of unpublished manuscripts. We describe Gibbs sampling in Section 3.

Whereas the application of Gibbs sampling is straightforward for fully conjugate Bayesian models (Gelfand and Smith, 1990; Gelfand *et al.*, 1990), non-conjugacy can cause computational difficulties. In applying Gibbs sampling to the estimation of generalized linear models with random effects, Zeger and Karim (1991) deal with non-conjugacy by rejection sampling from a normal envelope centred at the mode of the sampling density. We show in Section 3 that adaptive rejection sampling is well suited to handling non-conjugacy in applications of Gibbs sampling, as it requires neither the mode of the sampling density nor a rejection envelope that corresponds to a standard density.

2. Adaptive Rejection Sampling

To set the scene we begin by describing standard (non-adaptive) rejection sampling.

2.1. *Non-adaptive Rejection Sampling*

Rejection sampling is a general method for sampling points independently from a density $f(x)$. The density need be specified only up to a constant of integration, i.e. rejection sampling may be performed by using $g(x)$ instead of $f(x)$, where $g(x) =$

$cf(x)$ for some possibly unknown value of c. This is particularly useful when $c = \int_D g(x)\,dx$ is not available in closed form (where D denotes the domain of $f(x)$, i.e. the set of x for which $f(x) > 0$).

To sample n points independently from $f(x)$ by rejection sampling, define an *envelope* function $g_u(x)$ such that $g_u(x) \geqslant g(x)$ for all x in D, and optionally define also a *squeezing* function $g_l(x)$ such that $g_l(x) \leqslant g(x)$ for all x in D. Then perform the following sampling step until n points have been accepted.

Sample a value x^* from $g_u(x)$, and sample a value w independently from the uniform(0, 1) distribution. If you have defined a $g_l(x)$-function, perform the following squeezing test: if

$$w \leqslant g_l(x^*)/g_u(x^*)$$

then accept x^*. Otherwise evaluate $g(x^*)$ and perform the following rejection test: if

$$w \leqslant g(x^*)/g_u(x^*)$$

then accept x^*; otherwise reject x^*. Repeat until n points have been accepted.

Rejection sampling is only useful if it is more efficient or convenient to sample from the envelope $g_u(x)$ than from the density $f(x)$ itself. In practice, finding a suitable $g_u(x)$ can be difficult and often involves locating the supremum of $g(x)$ in D by using a standard optimization technique.

2.2. *Adaptive Rejection Sampling*

For Gibbs sampling, usually only one sample is required from each density, although sampling from many thousands of different densities may be required. Moreover, when estimating a model involving non-conjugacy, evaluations of $g(x)$ may be computationally expensive. These points are elaborated in Section 3. In these circumstances rejection sampling may be very inefficient, since it may involve many thousands of optimizations, each involving several evaluations of a $g(x)$ function.

Adaptive rejection sampling reduces the number of evaluations of $g(x)$ in two ways. Firstly, through the assumption of log-concavity of $f(x)$, we avoid the need to locate the supremum of $g(x)$ in D. Secondly, after each rejection, the probability of needing to evaluate $g(x)$ further is reduced by updating the envelope and squeezing functions to incorporate the most recently acquired information about $g(x)$.

We now describe our method in more detail. We assume that D is connected, that $g(x)$ is continuous and differentiable everywhere in D and that $h(x) = \ln g(x)$ is concave everywhere in D (i.e. $h'(x) = dh(x)/dx$ decreases monotonically with increasing x in D). This definition of log-concavity admits both straight line segments in $h(x)$ and discontinuities in $h'(x)$. The continuous curve in Fig. 1 exemplifies a concave $h(x)$ in a domain D.

Suppose that $h(x)$ and $h'(x)$ have been evaluated at k abscissae in D: $x_1 \leqslant x_2 \leqslant \ldots \leqslant x_k$. Let $T_k = \{x_i; i = 1, \ldots, k\}$. We define the rejection envelope on T_k as $\exp u_k(x)$ where $u_k(x)$ is a piecewise linear *upper hull* formed from the tangents to $h(x)$ at the abscissae in T_k, in the manner of the upper broken curve of Fig. 1. For $j = 1, \ldots, k-1$ the tangents at x_j and x_{j+1} intersect at

$$z_j = \frac{h(x_{j+1}) - h(x_j) - x_{j+1}\,h'(x_{j+1}) + x_j\,h'(x_j)}{h'(x_j) - h'(x_{j+1})}. \tag{1}$$

Thus for $x \in [z_{j-1}, z_j]$ and $j = 1, \ldots, k$, we define

$$u_k(x) = h(x_j) + (x - x_j)\, h'(x_j) \tag{2}$$

where z_0 is the lower bound of D (or $-\infty$ if D is not bounded below) and z_k is the upper bound of D (or $+\infty$ if D is not bounded above). We also define

$$s_k(x) = \exp u_k(x) \Big/ \int_D \exp u_k(x')\, dx'. \tag{3}$$

Finally, we define the squeezing function on T_k as $\exp l_k(x)$, where $l_k(x)$ is a piecewise linear *lower hull* formed from the chords between adjacent abscissae in T_k, in the manner of the lower broken curve of Fig. 1. Thus for $x \in [x_j, x_{j+1}]$

$$l_k(x) = \frac{(x_{j+1} - x)\, h(x_j) + (x - x_j)\, h(x_{j+1})}{x_{j+1} - x_j} \tag{4}$$

for $j = 1, \ldots, k-1$. For $x < x_1$ or $x > x_k$ we define $l_k(x) = -\infty$.

Thus the rejection envelope and the squeezing function are piecewise exponential functions. The concavity of $h(x)$ ensures that $l_k(x) \leqslant h(x) \leqslant u_k(x)$ for all x in D.

To sample n points independently from $f(x)$ by adaptive rejection sampling, perform the following initialization step, and then perform the following sampling and updating steps alternately until n points have been accepted.

2.2.1. *Initialization step*

Initialize the abscissae in T_k. If D is unbounded on the left then choose x_1 such that $h'(x_1) > 0$. If D is unbounded on the right then choose x_k such that $h'(x_k) < 0$. Having

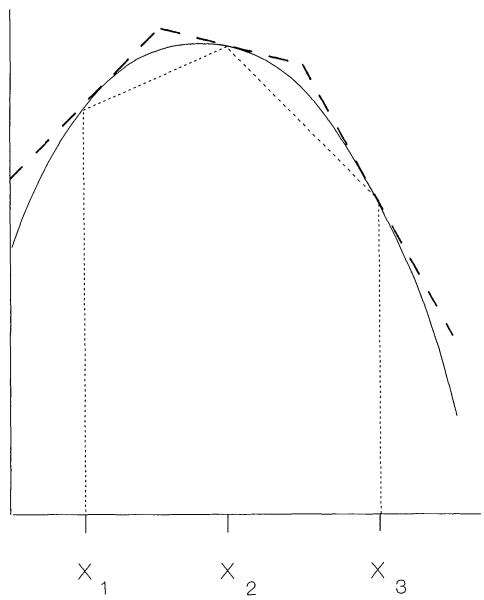

Fig. 1. A concave log-density $h(x)$, bounded on the left, showing upper and lower hulls based on three abscissae (x_1, x_2, x_3): ———, $h(x)$; – – –, upper hull; ·····, lower hull

defined k starting abscissae, calculate the functions $u_k(x)$, $s_k(x)$ and $l_k(x)$ from equations (2), (3) and (4) respectively.

2.2.2. *Sampling step*

Sample a value x^* from $s_k(x)$ and sample a value w independently from the uniform(0, 1) distribution. Perform the following squeezing test: if

$$w \leqslant \exp\{l_k(x^*) - u_k(x^*)\}$$

then accept x^*. Otherwise evaluate $h(x^*)$ and $h'(x^*)$ and perform the following rejection test: if

$$w \leqslant \exp\{h(x^*) - u_k(x^*)\}$$

then accept x^*; otherwise reject x^*.

2.2.3. *Updating step*

If $h(x^*)$ and $h'(x^*)$ were evaluated at the sampling step, include x^* in T_k to form T_{k+1}; relabel the elements of T_{k+1} in ascending order; construct the functions $u_{k+1}(x)$, $s_{k+1}(x)$ and $l_{k+1}(x)$ from equations (2), (3) and (4) respectively on the basis of T_{k+1}; increment k. Return to the sampling step if n points have not yet been accepted.

2.3. *Proof of Adaptive Rejection Sampling*

The proof that adaptive rejection sampling leads to independent samples from $f(x)$ is straightforward. Let x_r^* denote the rth sampled value of x, whether or not it was accepted or included in T_k. Let

$$\delta_r = \begin{cases} 0 & \text{if } x_r^* \text{ was accepted at the squeezing test,} \\ 1 & \text{if } x_r^* \text{ was accepted at the rejection test,} \\ 2 & \text{if } x_r^* \text{ was rejected.} \end{cases}$$

Let H_r denote the history of the process, up to and including the processing of x_r^*: so $H_r = \{(x_i^*, \delta_i); i = 1, \ldots, r\}$. Thus H_r defines the current upper and lower hulls.

Let [|] generically denote a conditional probability density function. Then

$$[(x_{r+1}^* = x) \cap (\delta_{r+1} \neq 2) \mid H_r] = \exp h(x) \Big/ \int_D \exp u_k(x')\, dx'$$

and so

$$[x_{r+1}^* = x \mid H_r \cap (\delta_{r+1} \neq 2)] = \exp h(x) \Big/ \int_D \exp h(x')\, dx' = f(x)$$

which does not depend on H_r. Thus accepted values of x are drawn independently from $f(x)$.

2.4. *Efficiency*

Suppose that k evaluations of $h(x)$ and $h'(x)$ have been performed. Suppose also that the current x^* has not been accepted at the squeezing step. Then the probability that $x^* = x$ is proportional to $\exp u_k(x) - \exp l_k(x)$. Thus new evaluations of $h(x)$ and $h'(x)$ are most likely to occur at values of x where the rejection envelope and squeezing

function are most discrepant. Therefore the method tends to space evaluations of $h(x)$ and $h'(x)$ optimally.

The number of evaluations of $h(x)$ and $h'(x)$ needed to sample a value from $f(x)$ could be sensitive to the initial choice of T_k. We examine this empirically in Table 1, for the standard normal density and with two starting abscissae. Widely separated starting abscissae are only modestly detrimental, and the optimum starting abscissae are around -1 and $+1$. Asymmetry in the starting abscissae has little impact on the number of evaluations of $h(x)$ and $h'(x)$. In general we have found two starting abscissae ($k=2$) to be necessary and sufficient for computational efficiency.

Empirically, the number of evaluations of $h(x)$ and $h'(x)$ required to sample n points from $f(x)$ increases approximately in proportion to $\sqrt[3]{n}$, even for quite non-normal densities. To sample 100 points from the standard normal distribution about 15 evaluations of $h(x)$ and $h'(x)$ are required; to sample 1000 points about 30 evaluations of $h(x)$ and $h'(x)$ are required.

3. Adaptive Rejection Sampling and Gibbs Sampling

3.1. *Gibbs Sampling*

Complex data sets can often be modelled in the form of a number n of submodels, each submodel $m=1, \ldots, n$ expressing the conditional distribution $[\beta_m | \{\beta_i; i \in S_m\}]$ of parameter β_m conditional on a set of other parameters indexed by S_m. We shall refer to $[\beta_m | \{\beta_i; i \in S_m\}]$ as the *model conditional* for β_m. Usually model conditionals are defined hierarchically (Lindley and Smith, 1972). For notational convenience we shall regard data as fixed parameters. Gibbs sampling (Geman and Geman, 1984) is often a convenient method of obtaining the joint posterior distribution of the parameters of a hierarchical model, and of any marginal or conditional posteriors of interest (Gelfand and Smith, 1990; Gelfand *et al.*, 1990).

To perform Gibbs sampling, initial values are assigned to each free parameter. Then for each free parameter β_m in turn the current parameter value is replaced by a value drawn from the *full conditional* distribution of that parameter $[\beta_m | \{\beta_i; i=1,$

TABLE 1
Evaluations of h(x) required to sample one point from the standard normal density, using adaptive rejection sampling, for various starting abscissae †

Starting abscissae x_1	x_2	*Mean number‡ of evaluations of h(x)*	*Maximum number‡ of evaluations of h(x)*
−0.5	0.5	3.1	7
−1.0	1.0	2.8	6
−2.0	2.0	3.3	6
−5.0	5.0	4.4	8
−10.0	10.0	5.1	8
−9.0	1.0	4.3	8
−8.0	2.0	4.4	7
−7.0	3.0	4.5	8
−6.0	4.0	4.4	8

†1000 simulations.
‡Including evaluations at the starting abscissae x_1 and x_2.

$\ldots, m-1, m+1, \ldots, n\}]$, conditioning on *all* the remaining parameters (fixed or free). We shall denote this full conditional distribution by $[\beta_m|\ ]$. This resampling is repeated many times. Under fairly weak regularity conditions Geman and Geman (1984) show that this process generates samples from the joint posterior distribution of the free parameters, conditional on the fixed parameters (data). From these samples any posterior summaries of interest can be straightforwardly calculated (e.g. the mean, median, 5th and 95th centiles of the sampled values for each free parameter). For further details see Gelfand *et al.* (1990).

Thus Gibbs sampling requires specification of the *full* conditional distribution $[\beta_m|\ ]$ for each parameter β_m. For a hierarchical model the full conditional for β_m can be expressed in terms of the model conditionals:

$$[\beta_m|\ ] \propto [\beta_m|\{\beta_i; i \in S_m\}] \prod_{\{j: m \in S_j\}} [\beta_j|\{\beta_i; i \in S_j\}] \tag{5}$$

where the product is over all submodels which condition on β_m. Here proportionality implies that the full conditional distribution for β_m differs from the right-hand side of expression (5) only by a multiplicative term which does not depend on β_m, but which will in general depend on other $\{\beta_i; i \neq m\}$. Unless there is conjugacy between each of the producted terms in expression (5), the full conditional will not correspond to a common distribution and it will not be possible to derive a closed form for the proportionality constant in expression (5). Moreover, since it is the product of possibly many terms, expression (5) will be computationally expensive to evaluate repeatedly. Section 4 provides an example in which over 150 terms form the product in expression (5); in other applications several thousand terms could be involved.

3.2. *Log-concavity*

Now, most commonly used densities are concave on the logarithmic scale, with respect to both random variable and distributional parameters (see Table 2). Moreover, when this is not so, the log-density may be concave with respect to a suitably transformed random variable or parameter (taking account of the Jacobian if transforming the random variable) (Table 2). In particular if $[x|\mu, \sigma]$ is any density parameterized in terms of a location parameter μ and scale parameter σ, such that

$$[x|\mu, \sigma] = \frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right)$$

for some function $f(z)$, and if $\ln f(z)$ is concave with respect to its argument z, then the logarithm of the density $[x|\mu, \sigma]$ is concave with respect to x, μ and $\tau = \sigma^{-1}$, but not necessarily with respect to σ or other functions of σ.

Therefore, taking logarithms in expression (5), often all the terms in

$$h(\beta_m) = \ln[\beta_m|\{\beta_i; i \in S_m\}] + \sum_{\{j: m \in S_j\}} \ln[\beta_j|\{\beta_i; i \in S_j\}] \tag{6}$$

will be concave with respect to β_m, and consequently also $h(\beta_m)$ will be concave with respect to β_m, being the sum of concave terms. In these circumstances adaptive rejection sampling can be used to sample efficiently from $h(\beta_m)$. For adaptive rejection sampling of β_m we require of $[\beta_j|\{\beta_i; i \in S_j\}]$ in equation (6) only that it be continuous, differentiable and log-concave with respect to β_m. Thus $[\beta_j|\{\beta_i; i \in S_j\}]$

TABLE 2
Log-concavity for common probability density functions f(x) †

f(x)	*Parameters*	*log f(x) concave with respect to*	*log f(x) not concave with respect to*
Normal	Mean μ, variance σ^2	x, μ, $1/\sigma$, $\log \sigma$	σ
Log-normal	Location μ, scale σ^2	$\log x$, μ, $1/\sigma$, $\log \sigma$	x, σ
Exponential	Rate λ	x, $\log x$, λ	
Gamma	Index r, rate λ	$\log x$, x (if $r \geq 1$), λ, r	x (if $r < 1$)
Beta	Shape a, b	$\text{logit}(x)$, x (if $a, b \geq 1$), a, b	
Double exponential	Location α, scale β	x, α, $1/\beta$, $\log \beta$	β
Weibull	Shape b, scale a	$\log x$, x (if $b \geq 1$), a, b	
Logistic	Location α, scale β	x, α, $1/\beta$	β
Pareto	Shape θ, bound x_0	$\log x$, $1/x$ (if $\theta \geq 1$), x_0, $\log x_0$, θ	x
Gumbel or extreme value	Location α, scale β	x, α, $1/\beta$	β
t	Degrees of freedom k		x
F	Degrees of freedom m, n	$\log x$	x
χ^2	Degrees of freedom k	$\log x$, x (if $k \geq 2$), k	x (if $k < 2$)
Bernoulli	Proportion p	p, $\text{logit}(p)$	
Binomial	Proportion p, index r	p, $\text{logit}(p)$	
Poisson	Rate λ	λ	
Geometric	Proportion p	p, $\text{logit}(p)$	
Negative binomial	Proportion p, index r	p, $\text{logit}(p)$	

†In considering concavity of $\log f(x)$ with respect to transformations of the random variable x, the Jacobian of the transformation was taken into account. The table also contains common discrete distributions with continuous parameters. The notation follows that of Mood *et al.* (1974).

could be a discrete distribution of β_i. We have included some common discrete distributions in Table 2.

3.3. *Multivariate Full Conditionals*

Adaptive rejection sampling, as we have described it, permits only univariate sampling. However, if β_m in formulae (5) and (6) is multivariate, then (univariate) adaptive rejection sampling can still be used. To see this note that, for each element β_{mk} of β_m, the univariate full conditional $[\beta_{mk} |\ ]$ is proportional to the multivariate full conditional $[\beta_m |\ ]$. Therefore, if $[\beta_m |\ ]$ is log-concave with respect to β_m (which it may be, by invoking the argument of Section 3.2), then $[\beta_{mk} |\ ]$ will be log-concave with respect to β_{mk}. Thus the Gibbs sampler can be implemented to update each element β_{mk} in turn, using adaptive rejection sampling with $h(\beta_{mk}) = \ln[\beta_m |\ ]$.

4. Application to Monoclonal Antibody Reactivity

We have applied adaptive rejection sampling to a Gibbs sampling analysis of monoclonal antibody reactivity. The data that we examine originate from flow cytometry experiments in which 13 monoclonal antibodies were tested against 15 normal and malignant human cell types. The data form part of a much larger study designed to discover antigens involved in small cell lung cancers (Souhami *et al.*, 1991). Each of the 13 antibodies studied here is known to react with cells expressing

the neural cell adhesion molecule (NCAM), a molecule which is expressed on the surface of normal peripheral nerve cells, but which also appears on malignant small cell lung cancer cells. In normal nerve cells the molecule is thought to attach the nerve to other cells, thereby fixing it in position; in malignant cells its role is as yet unclear. The NCAM molecule is also present, in varying densities, on other normal and malignant cell types (e.g. brain cells and neuroblastoma cells) but is absent on many cell types.

The recognition of an antigen (a molecule expressed on the surface of a cell) by an antibody depends on the fraction antibody binding (FAB) portion of the antibody being the correct shape to lock on to the antigen. Since different monoclonal antibodies recognizing the same antigen will in general have slightly differently shaped FAB portions, we might expect such antibodies to differ somewhat in their ability to bind to the antigen. The purpose of the analysis presented here is to quantify the nature and extent of this variability. Lack of variability could indicate that the NCAM molecule has only one epitope (molecular feature) accessible to antibody.

4.1. *Model*

In each flow cytometry experiment, reactivity was measured as the percentage of cells reacting with antibody. Percentages were reported rounded to the nearest integer. Let y_{ijr} denote the percentage reactivity recorded in the rth experiment testing antibody i against target cell type j (for $i=1, \ldots, I; j=1, \ldots, J; r=1, \ldots, n_{ij}$). We might proceed by ignoring the rounding and constructing a linear model of $\text{logit}(y_{ijr}/100)$, where $\text{logit}(x)$ denotes $\log\{x/(1-x)\}$. However, with these data the rounding is important, as there are many (rounded) reactivities of 0.0 or 100.0. Therefore, instead we suppose that y_{ijr} is related to an underlying logistic distribution with location parameter μ_{ij} and scale parameter τ_y as follows:

$$[y_{ijr} \mid \mu_{ij}, \tau_y] = \frac{1}{1+\exp\{-\tau_y(b_{ijr}-\mu_{ij})\}} - \frac{1}{1+\exp\{-\tau_y(a_{ijr}-\mu_{ij})\}} \tag{7}$$

where

$$a_{ijr} = \text{logit}\left(\frac{y_{ijr}-0.5}{100}\right) \qquad \text{if } y_{ijr} > 0,$$

$$= -\infty \qquad \text{if } y_{ijr} = 0,$$

and

$$b_{ijr} = \text{logit}\left(\frac{y_{ijr}+0.5}{100}\right) \qquad \text{if } y_{ijr} < 100,$$

$$= +\infty \qquad \text{if } y_{ijr} = 100.$$

For the location parameter we specify

$$\mu_{ij} = \beta_0 + \beta_{1j} + \beta_{2i}. \tag{8}$$

This model specifies a basic pattern of reactivity ($\beta_0+\beta_{1j}$) across cell types, reflecting different densities of cell surface expression of the NCAM molecule. The basic pattern is adjusted to allow for differences in affinity for the NCAM molecule (β_{2i}).

To complete the model we need to specify priors for β_0, β_{1j}, β_{2i} and τ_y. We set

$$[\beta_0 | \alpha_0, \tau_0] \sim N(\alpha_0, \tau_0^{-1}), \tag{9}$$

$$[\beta_{1j} | \tau_1] \sim N(0, \tau_1^{-1}), \tag{10}$$

$$[\beta_{2i} | \tau_2] \sim N(0, \tau_2^{-1}), \tag{11}$$

$$[\tau_y | \rho_y, \lambda_y] \sim G(\rho_y, \lambda_y), \tag{12}$$

where $N(\alpha, \tau^{-1})$ denotes a normal distribution with mean α and variance τ^{-1}, and $G(\rho, \lambda)$ denotes a gamma distribution with index parameter ρ and scale parameter λ. We set $\alpha_0 = -1.0$, $\tau_0 = 0.1$, $\rho_y = 2.0$ and $\lambda_y = 4.0$ to give fairly flat priors for β_0 and τ_y. We do not fix the remaining hyperparameters τ_1 and τ_2, as the data contain information about them: indeed, these hyperparameters are the focus of this analysis. Instead we specify fairly flat hyperpriors

$$[\tau_k | \rho_k, \lambda_k] \sim G(\rho_k, \lambda_k) \tag{13}$$

for $k = 1, 2$, setting $\rho_k = 2.0$ and $\lambda_k = 4.0$.

4.2. *Gibbs Sampling*

To estimate the hierarchical model (7)–(13) by Gibbs sampling the full conditional distribution for each free model parameter is required. The full conditional for β_0 is

$$[\beta_0 | \ ] \propto \exp\left\{ -\frac{1}{2}\tau_0(\beta_0 - \alpha_0)^2 \right\} \prod_{ijr} \left[\frac{1}{1 + \exp\{-\tau_y(b_{ijr} - \mu_{ij})\}} - \frac{1}{1 + \exp\{-\tau_y(a_{ijr} - \mu_{ij})\}} \right] \tag{14}$$

where each data point contributes one term to the product. Expressions similar to expression (14) also hold for the full conditional distributions of β_{1j}, β_{2i} and τ_y (with suitable restrictions on the subscripts of the product operator and, for τ_y, the first term in expression (14) being replaced by the gamma prior (12)).

The full conditional distribution for β_0 in expression (14) does not simplify; thus sampling β_0 could be time consuming. Fortunately, each of the many terms in expression (14) is concave on the logarithmic scale with respect to β_0, and so adaptive rejection sampling can be used. Similar considerations apply to the full conditionals for β_{1j}, β_{2i} and τ_y.

The full conditional for τ_1 is straightforward:

$$[\tau_1 | \ ] \sim G\left(\rho_1 + \frac{1}{2}J, \ \lambda_1 + \frac{1}{2}\sum_{j=1}^{J} \beta_{1j}^2 \right) \tag{15}$$

with a similar expression for the full conditional for τ_2.

We performed 1000 iterations of the Gibbs sampler. At each iteration, for each of the parameters β_0, β_{1j}, β_{2i} and τ_y, we used the 15th and 85th centiles of the sampling density $s_k(x)$ from the previous iteration as starting values for adaptive rejection sampling.

4.3. *Results*

For each of the parameters sampled by adaptive rejection sampling, on average only three evaluations of $h(x)$ were required at each iteration (including the two starting evaluations), and on only 5% of iterations were more than four evaluations of $h(x)$ required.

Convergence in distribution was achieved within 10 iterations. Therefore the posterior summaries in Table 3 were based on iterations 11–1100. The results suggest that there is substantial variability in the expression of NCAM among targets (τ_1^{-1}), but relatively little variability in antibody affinity for the NCAM molecule (τ_2^{-1}). There is some evidence that antibody 12 has low affinity to NCAM as the 90% support interval for $\beta_{2,12}$ does not cover 0.

5. Conclusions

We have shown that adaptive rejection sampling can be used as a black box routine for efficiently sampling from complex densities, in particular those arising in applications of Gibbs sampling to the analysis of hierarchical Bayesian models involving non-conjugacy. In the context of Gibbs sampling, we suggest the use of centiles from the sampling densities $s(x)$ from one iteration to provide starting abscissae for the next iteration, as described in Section 4.

Although adaptive rejection sampling is conceptually simple, care must be taken in its implementation to avoid numerical problems when sampling from densities which are extremely concentrated or skewed. We have written a Fortran program (available on request) to perform adaptive rejection sampling, which behaves well under

TABLE 3
Posterior summaries based on iterations 11–1000 of the Gibbs sampler for model (7)–(13)

Parameter	*Mean*	*Standard deviation*	*5th centile*	*95th centile*
τ_1	0.13	0.05	0.06	0.21
τ_2	1.59	0.55	0.80	2.65
τ_y	0.87	0.04	0.81	0.93
$\sigma_1 = 1/\sqrt{\tau_1}$	2.95	0.56	2.19	4.02
$\sigma_2 = 1/\sqrt{\tau_2}$	0.83	0.16	0.61	1.11
$\sigma_y = 1/\sqrt{\tau_y}$	1.07	0.02	1.04	1.11
$\beta_{2,1}$	0.23	0.37	−0.39	0.85
$\beta_{2,2}$	0.03	0.39	−0.62	0.63
$\beta_{2,3}$	0.42	0.38	−0.21	1.06
$\beta_{2,4}$	0.06	0.36	−0.52	0.60
$\beta_{2,5}$	−0.15	0.37	−0.75	0.44
$\beta_{2,6}$	−0.26	0.36	−0.83	0.34
$\beta_{2,7}$	−0.04	0.36	−0.64	0.53
$\beta_{2,8}$	0.06	0.37	−0.54	0.72
$\beta_{2,9}$	−0.26	0.37	−0.37	0.88
$\beta_{2,10}$	−0.13	0.36	−0.69	0.46
$\beta_{2,11}$	0.21	0.36	−0.39	0.78
$\beta_{2,12}$	−0.66	0.36	−1.27	−0.09
$\beta_{2,13}$	0.10	0.37	−0.50	0.74

extreme conditions. In particular, this algorithm can be used straightforwardly to sample from truncated distributions.

Acknowledgements

We are grateful to Professor Peter Beverley for permitting us the use of data from the Second International Workshop on Small-cell Lung Cancer Antigens, and to referees for drawing our attention to the work of Devroye (1986).

References

Carlin, B. P., Gelfand, A. E. and Smith, A. F. M. (1991) Hierarchical Bayesian analysis of changepoint problems. *Appl. Statist.*, **41**, 389–405.

Carlin, B. P. and Polson, N. G. (1991a) Bayesian model choice in nonconjugate settings. *Can. J. Statist.*, to be published.

——(1991b) An expected utility approach to influence diagnostics. *J. Am. Statist. Ass.*, to be published.

Clayton, D. G. (1991) A Monte Carlo method for Bayesian inference in frailty models. *Biometrics*, to be published.

Devroye, L. (1986) *Non-uniform Random Variate Generation*, 1st edn. New York: Springer.

Gelfand, A. E., Hills, S. E., Racine-Poon, A. and Smith, A. F. M. (1990) Illustration of Bayesian inference in normal data models using Gibbs sampling. *J. Am. Statist. Ass.*, **85**, 972–985.

Gelfand, A. E. and Smith, A. F. M. (1990) Sampling based approaches to calculating marginal densities. *J. Am. Statist. Ass.*, **85**, 398–409.

Geman, S. and Geman, D. (1984) Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. Pattn Anal. Mach. Intell.*, **6**, 721–741.

Gilks, W. R., Oldfield, L. and Rutherford, A. (1989) Statistical analysis. In *Leucocyte Typing IV: White Cell Differentiation Antigens* (eds W. Knapp, B. Dorken, W. R. Gilks, E. P. Rieber, R. E. Schmidt, H. Stein and A. E. G. Kr. von dem Borne), pp. 6–12. Oxford: Oxford University Press.

Lindley, D. V. and Smith, A. F. M. (1972) Bayes estimates for the linear model (with discussion). *J. R. Statist. Soc.* B, **34**, 1–41.

Mack, W., Langholz, B. and Thomas, D. C. (1990) Survival models for familial aggregation of cancer. *Environ. Hlth Perspect.*, **87**, 27–35.

Mood, A. M., Graybill, F. A. and Boes, D. C. (1974) *Introduction to the Theory of Statistics*, 3rd edn. Tokyo: McGraw-Hill Kogakusha.

Souhami, R., Gilks, W. R., Bobrow, L. and Beverley, P. (1991) *Br. J. Cancer*, **63**, suppl. XIV.

Thomas, D. C. (1991) Fitting genetic data using Gibbs Sampling: an application to nevus counts in 38 Utah kindreds. *Cytgenet. Cell Genet.*, to be published.

Verdinelli, I. and Wasserman, L. (1991) Bayesian analysis of outlier problems using the Gibbs sampler. *Statist. Comput.*, **1**, in the press.

Zeger, S. L. and Karim, M. R. (1991) Generalized linear models with random effects; a Gibbs sampling approach. *J. Am. Statist. Ass.*, **86**, 79–86.