Adaptive Rejection Metropolis Sampling within Gibbs Sampling
Author(s): W. R. Gilks, N. G. Best and K. K. C. Tan
Source: *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, Vol. 44, No. 4 (1995), pp. 455-472
Published by: Wiley for the Royal Statistical Society
Stable URL: https://www.jstor.org/stable/2986138
Accessed: 12-10-2018 19:00 UTC

Appl. Statist. (1995)
44, *No.* 4, *pp.* 455–472

Adaptive Rejection Metropolis Sampling within Gibbs Sampling

By W. R. GILKS† and N. G. BEST

Medical Research Council Biostatistics Unit, Cambridge, UK

and K. K. C. TAN

Addenbrooke's Hospital, Cambridge, UK

[Received October 1993. Final revision October 1994]

SUMMARY

Gibbs sampling is a powerful technique for statistical inference. It involves little more than sampling from full conditional distributions, which can be both complex and computationally expensive to evaluate. Gilks and Wild have shown that in practice full conditionals are often log-concave, and they proposed a method of adaptive rejection sampling for efficiently sampling from univariate log-concave distributions. In this paper, to deal with *non*-log-concave full conditional distributions, we generalize adaptive rejection sampling to include a Hastings–Metropolis algorithm step. One important field of application in which statistical models may lead to non-log-concave full conditionals is population pharmacokinetics. Here, the relationship between drug dose and blood or plasma concentration in a group of patients typically is modelled by using non-linear mixed effects models. Often, the data used for analysis are routinely collected hospital measurements, which tend to be noisy and irregular. Consequently, a robust (*t*-distributed) error structure is appropriate to account for outlying observations and/or patients. We propose a robust non-linear full probability model for population pharmacokinetic data. We demonstrate that our method enables Bayesian inference for this model, through an analysis of antibiotic administration in new-born babies.

Keywords: Bayesian computation; Gibbs sampling; Markov chain Monte Carlo method; Metropolis algorithm; Pharmacokinetic model; Random variate generation

1. Introduction

Gibbs sampling (Geman and Geman, 1984; Gelfand and Smith, 1990) is a Markov chain Monte Carlo (MCMC) technique for drawing dependent samples from complex high dimensional distributions. In the Bayesian context, these distributions are usually posterior distributions of the model parameters, and samples produced by the Gibbs sampler can be used straightforwardly for Bayesian inference. A good introduction to Gibbs sampling is given by Casella and George (1992).

At each iteration of the Gibbs sampler, each parameter or set of parameters is updated in turn by sampling a new value from its full conditional distribution. The full conditional distribution of a parameter is its distribution conditional on the data

†*Address for correspondence*: Medical Research Council Biostatistics Unit, Institute of Public Health, Robinson Way, Cambridge, CB2 2SR, UK.
E-mail: wally.gilks@mrc-bsu.cam.ac.uk

 0035–9254/95/44455

and on the current values of all the other parameters. Thus, from one iteration to the next, full conditional distributions change as the conditioning parameters change. In a typical run of the Gibbs sampler, draws from millions of different full conditional distributions are required, so methods for constructing full conditional distributions and for sampling from them must be very efficient. Certain full conditionals reduce analytically to well-known distributions, for which special methods for efficient random variate generation are available. More usually no analytical reduction is possible. For log-concave distributions efficient random variate generation can be achieved through adaptive rejection sampling (ARS) (Gilks and Wild, 1992). Gilks and Wild showed that many full conditional distributions encountered in practice are log-concave. Dellaportas and Smith (1993) extended this result to generalized linear models with canonical link, and George *et al.* (1993) further extended this result to conjugate likelihood distributions.

However, not all models of practical importance yield log-concave full conditionals. One such example is the class of non-linear mixed effect models commonly used to estimate pharmacokinetic and pharmacodynamic parameters. This paper extends ARS to deal with distributions that are not log-concave by appending a Hastings–Metropolis algorithm step (Metropolis *et al.*, 1953; Hastings, 1970). In Section 2 we give an example of the population pharmacokinetic data and models which motivate this methodological development. We describe adaptive rejection Metropolis sampling (ARMS) in Section 3, and in Section 4 we present the results of a population pharmacokinetic analysis of the antibiotic gentamicin by using the ARMS within Gibbs sampling algorithm. In Section 5 we briefly discuss other approaches to sampling from non-log-concave full conditional distributions.

2. Population Pharmacokinetic Modelling

Population pharmacokinetics refers to the average behaviour and interindividual variation in the way that a drug is absorbed, distributed and eliminated in the subpopulation of patients for whom the drug is intended. Such information leads to a better understanding of how drugs are handled by the body and can be used to predict dosage requirements. The present example concerns the administration of the antibiotic gentamicin to treat serious infections in new-born babies. The data were originally analysed by Thomson *et al.* (1988) using the NONMEM software (see Section 2.3). The aim of the analysis is to estimate covariate-adjusted population values and between-patient variability for the pharmacokinetic parameters of interest.

2.1. *Data*

Typically, data used for population pharmacokinetic analysis are taken from the routine hospital records of patients being treated with the drug. For example, Fig. 1 shows the data for one baby who was treated with gentamicin in the present application. The arrows indicate the time of administration and magnitude of each dose of drug. Following each intravenous dose, the concentration of drug in the blood rises rapidly and then decreases over time until the next dose is administered. This results in an underlying blood concentration curve something like the broken curve in Fig. 1. Blood samples are taken from the patient at specific times to measure the

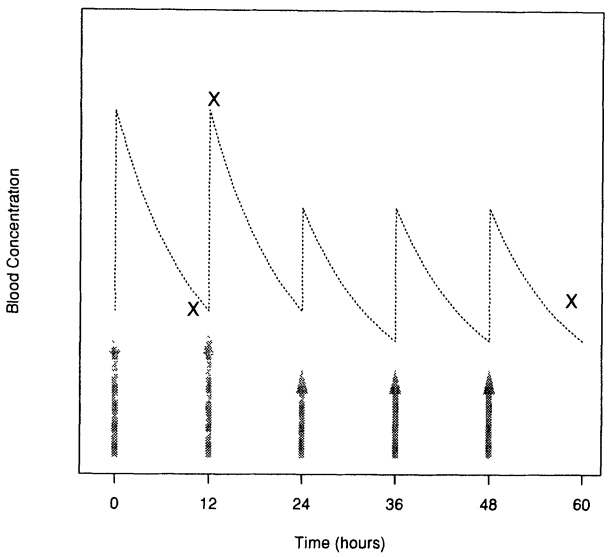

Fig. 1. Pharmacokinetic data for one patient

actual value of this concentration. These measurements are indicated by crosses in the figure.

The complete data set used in this study consists of repeated gentamicin dose (in milligrams per kilogram) and blood concentration measurements on 113 babies. Further details are given in Table 1. Covariate information is also available for each baby as follows:

(a) $X_1 = 1$ if there is moderate or severe asphyxia at birth (measured by an APGAR score below 7) and $X_1 = 0$ otherwise;
(b) X_2, the reciprocal creatinine concentration (a measure of kidney function);
(c) $X_3 = 1$ for immature babies (post-conceptual age 34 weeks or less) and $X_3 = 0$ otherwise.

(The APGAR score is a universally recognized scoring system used to measure a baby's condition immediately after birth. It consists of assessing heart rate, respiratory effort, muscle tone, reflex response and colour and scoring 0, 1 or 2 for each according to the degree to which they are present. This gives a maximal score of 10 for a healthy baby.)

TABLE 1
Summary of the data

Variable	*Median*	*Range*
No. of gentamicin doses	7	2–24
No. of blood concentration measurements	2	1–9
Length of follow-up (days)	2.9	0.2–18.3

The advantage of using observational (as opposed to experimental) pharmacokinetic data is that the information is more representative of the population of patients who are actually being treated with a given drug, and is available on many more individuals than in the experimental setting. However, the measurements *per patient* are often sparse and irregularly spaced and inherently noisy due to variation in the timing of the routinely collected blood samples, analytical errors in measuring the drug concentration, patients forgetting or refusing to take their drugs at the correct time, and so on.

2.2. *Pharmacokinetic Model*

A simple diagrammatic representation of the pharmacokinetic process is shown in Fig. 2. The body is represented as a single homogeneous compartment, into which a dose of drug is administered via intravenous injection. The drug is then instantaneously distributed within the compartment and eventually eliminated from it.

The concentration of the drug in the compartment is described by the non-linear equation

$$y(t) = \frac{d}{V}\exp\left\{-\frac{C}{V}(t-s)\right\} \tag{1}$$

where d is the dose administered at time s, $y(t)$ is the concentration of drug in the blood at time $t > s$, and V and C are pharmacokinetic parameters called respectively the *volume of distribution* (a proportionality constant relating the total amount of drug in the body to the concentration of drug in the blood) and the *clearance* (the volume of blood which is cleared of drug per unit time). If more than one dose of drug is administered, the equation involves summation over previous doses (see Section 2.4). Since V and C are not directly measurable they must be estimated statistically to describe the pharmacokinetic characteristics of the drug.

2.3. *Estimation of Population Pharmacokinetic Parameters*

The estimation of population pharmacokinetic parameters is complicated by the sparse noisy data and the non-linearity of the models used. Various iterative approximate maximum likelihood approaches have been proposed (for example, see Steimer *et al.* (1984) for a critical review). The most widely used method is based on extended least squares (Beal, 1984) and is implemented in the NONMEM software package (Sheiner and Beal, 1980). However, these methods are somewhat restricted by the implicit assumption of normality or log-normality of the underlying parameter distributions. In addition, the NONMEM estimates have been shown to be biased when the pharmacokinetic parameters are highly correlated (Steimer *et al.*, 1984). Alternatively, some researchers have used nonparametric maximum likelihood methods to estimate population pharmacokinetic models (Mallet, 1986; Davidian and Gallant, 1992), whereas Racine-Poon (1985) adopted a Bayesian

Fig. 2. Simple diagrammatic representation of a one-compartment pharmacokinetic model

approach and used an EM-type numerical algorithm to perform the parameter estimation because of intractability of the resulting integrals. Wakefield *et al.* (1994) used Gibbs sampling, rather than numerical or analytical methods, to perform full Bayesian inference for population pharmacokinetic models. In this paper we extend the Gibbs sampling approach to a population pharmacokinetic model with a robust error distribution.

2.4. *Statistical Model*

For statistical inference, we need to consider how the data arise from the underlying pharmacokinetic model described in Section 2.2. Let d_{il} denote the lth drug dose for the ith patient, s_{il} the time of administration of that dose, y_{ij} the jth measured blood concentration for patient i and t_{ij} the time of that measurement. To provide robustness against the possibility of outliers (a common problem in pharmacokinetic data), we assume that $z_{ij} = \log y_{ij}$ follows a location- and scale-shifted Student t-distribution on $\nu = 10$ degrees of freedom (see for example DeGroot (1970), p. 42) with mean given by summing equation (1) over all previous doses,

$$E(z_{ij}) = \log\left[\sum_{l:t_{ij}>s_{il}} \frac{d_{il}}{V_i}\exp\left\{-\frac{C_{ij}}{V_i}(t_{ij}-s_{il})\right\}\right], \tag{2}$$

and variance σ^2. This model follows from equation (1) only under the assumption of linear kinetics (i.e. that the relationship between dose and concentration remains constant for all doses). Concentration is transformed to the logarithmic scale to stabilize the error variance, which tends to be greater when $E(y_{ij})$ is large. The clearance and volume for patient i at time t_{ij} are modelled deterministically:

$$\log C_{ij} = \alpha_i + \beta'(\mathbf{X}_{ij} - \bar{\mathbf{X}}) \tag{3}$$

and $\log V_i = \phi_i$, where α_i and ϕ_i are patient-specific random effects and $\mathbf{X}_{ij}$ is the vector of covariates $\{X_1, X_2, X_3\}$ for patient i at time t_{ij}. These covariates are centred on their means $\bar{\mathbf{X}}$ to reduce dependence between the elements of the regression coefficient vector β in the simulated Markov chain. We follow Thomson *et al.* (1988) in that V_i does not depend on covariates; however, our model could be extended to accommodate this. The random effects α_i and ϕ_i are assumed drawn from a bivariate normal population distribution with mean γ and covariance Σ. We assume vague conjugate prior distributions for the population parameters γ and Σ, for the regression parameters β and for the measurement error precision τ $(=1/\sigma^2)$. Thus, we have three stages in the model: stage 1,

$$\log Z_{ij} = \log E(z_{ij}|\, V_i,\ C_{ij}) + \epsilon_{ij},$$
$$\epsilon_{ij} \sim t_\nu(0,\ \tau^{-1});$$

stage 2,

$$\log C_{ij} = \alpha_i + \beta'(\mathbf{X}_{ij} - \bar{\mathbf{X}}),$$
$$\log V_i = \phi_i,$$
$$\begin{pmatrix}\alpha_i \\ \phi_i\end{pmatrix} \sim N(\boldsymbol{\gamma},\ \boldsymbol{\Sigma});$$

stage 3,

$$\beta_m \sim N(0,\ 10\,000), \qquad \text{for } m = 1,\ 2,\ 3,$$

$$\gamma \sim N\left\{\begin{pmatrix} -3.1 \\ -0.76 \end{pmatrix}, \begin{pmatrix} 1000 & 0 \\ 0 & 100 \end{pmatrix}\right\},$$

$$\Sigma^{-1} \sim W\left\{2, \begin{pmatrix} 5 & -5 \\ -5 & 105 \end{pmatrix}\right\},$$

$$\tau \sim \text{Ga}(0.2,\ 10).$$

Here β_m is the mth element of vector $\boldsymbol{\beta}$, and Ga(0.2, 10) denotes a gamma distribution with mean 0.02 and variance 0.002. The values specified for the prior mean for $\boldsymbol{\gamma}$ and the prior precision matrix for $\boldsymbol{\Sigma}^{-1}$ were chosen according to previously published estimates of gentamicin and other similar antibiotic pharmacokinetic parameters (Zaske, 1992; Fattinger *et al.*, 1991).

Fig. 3 represents our statistical model in the form of a directed acyclic graph (Whittaker, 1990). Square nodes represent known quantities (data); round nodes represent unknown quantities (parameters); triangular nodes indicate deterministic relationships. The graph clarifies the conditional independence assumptions that are implicit in the model specification.

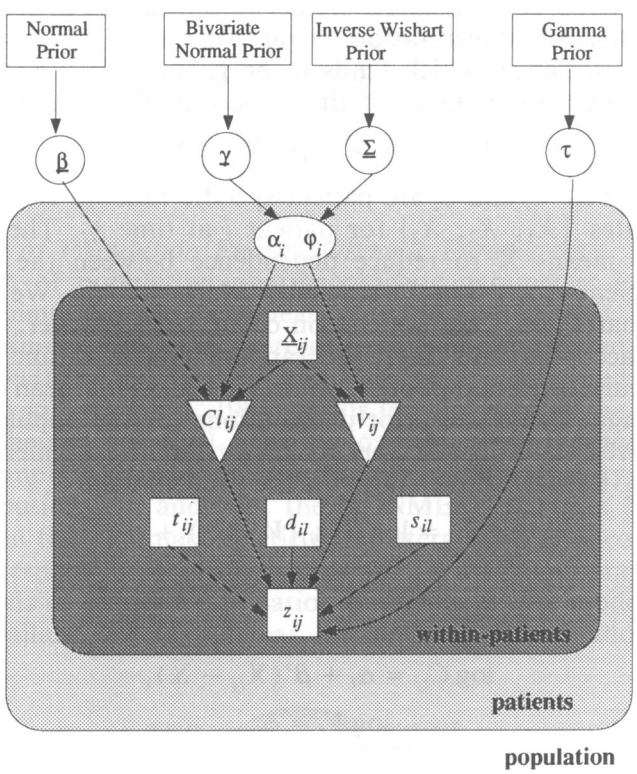

Fig. 3. Graphical representation of a pharmacokinetic model

2.5. *Full Conditional Distributions*

Estimation of the above model by using Gibbs sampling requires full conditional distributions to be constructed and sampled from, as discussed in Section 1. The full conditional distribution for any parameter in a directed acyclic graph is proportional to the product of all the terms in the model which contain it (hence full conditionals are usually known only up to a constant of proportionality—see Section 3.1). Thus, for example, the logarithm of the full conditional distribution for the mth element of the coefficient vector β is

$$\log f(\beta_m) \propto -\frac{1}{2}(\beta-\mu)'T(\beta-\mu) - \frac{\nu+1}{2}\left(\sum_{ij}\log\left[1+\frac{\tau}{\nu-2}\left\{z_{ij} - \log\left(\sum_{l:t_{ij}>s_{il}}\frac{d_{il}}{\exp\phi_i}\exp\left[-\frac{\exp\{\alpha_i+\beta'(\mathbf{X}_{ij}-\bar{\mathbf{X}})\}}{\exp\phi_i}(t_{ij}-s_{il})\right]\right)\right\}^2\right]\right) \tag{4}$$

where μ and T are the prior mean and precision matrix of β. Thus the non-linear model (2) and the t-distributed error structure lead to particularly complicated conditional probability densities which are not log-concave and are computationally very expensive to evaluate. To sample from this we modify the ARS algorithm by appending a Metropolis step. Thereby the scope of ARS is extended to non-log-concave distributions such as expression (4).

3. Adaptive Rejection Metropolis Sampling

Before describing ARMS we first describe its constituent methodologies: rejection sampling, ARS and the Metropolis algorithm.

3.1. *Rejection Sampling*

Rejection sampling (Ripley, 1987) is a method for drawing independent samples from a distribution (proportional to) $f(x)$. For this we require a sampling distribution $g(x)$ from which samples can be readily drawn and for which there is a finite constant m such that $mg(x) \geqslant f(x)$, $\forall x \in D$, where D denotes the domain of f. For practical purposes we also require that m be easily calculable. A single observation X_R from $f(x)$ is then drawn by the following algorithm:

step 1, sample X from $g(x)$;
step 2, sample U from uniform(0, 1);
step 3, if $U > f(X)/mg(X)$ then {
rejection step:
go back to step 1;}
else {
acceptance step:
set $X_R = X$;};
step 4, return X_R.

Further iterations of steps 1–4 will produce independent samples from f.

Note that rejection sampling does not involve evaluation of the integration constant $\int_D f(x)\,dx$. This is very convenient for sampling from full conditional

distributions, which are typically known only up to a constant of proportionality (for example, see expression (4) above).

The expected number of iterations of steps 1–3 is m, each iteration involving an evaluation of $f(X)$ at step 3. Thus unless m is small, which will be difficult to achieve in practice except in special cases, rejection sampling will involve many evaluations of $f(X)$. This is particularly critical when f is a full conditional distribution, as the evaluation of $f(X)$ at any point X can be very expensive computationally.

3.2. *Adaptive Rejection Sampling*

ARS reduces the number of evaluations of $f(X)$ in rejection sampling by improving the sampling density $g(x)$ after each rejection so that m decreases monotonically. The improvement is made by incorporating into $g(x)$ information about $f(x)$ obtained at each of the previously rejected points. For univariate log-concave densities this can be done by the method of Gilks and Wild (1992), or alternatively by the method of Gilks (1992). We describe the latter. For this, the domain D of f is an interval of the real line, densities are with respect to Lebesgue measure, and we define log-concavity of f as

$$\ln f(a) - 2\ln f(b) + \ln f(c) < 0 \qquad \forall\, a, b, c \in D \text{ such that } a < b < c. \tag{5}$$

This definition does not assume continuity in derivatives of f and includes, for example, linear and piecewise linear continuous functions.

Let $S_n = \{x_i;\ i = 0, \ldots, n+1\}$ denote a *current* set of abscissae in ascending order, where x_0 and x_{n+1} are the possibly infinite lower and upper limits of D. For $1 \leqslant i \leqslant j \leqslant n$ let $L_{ij}(x;\ S_n)$ denote the straight line through points $[x_i,\ \ln f(x_i)]$ and $[x_j,\ \ln f(x_j)]$, and for other (i, j) let $L_{ij}(x;\ S_n)$ be undefined.

Define a piecewise linear function $h_n(x)$:

$$h_n(x) = \min[L_{i-1,i}(x;\ S_n),\ L_{i+1,i+2}(x;\ S_n)], \qquad x_i \leqslant x < x_{i+1}, \tag{6}$$

where we notationally suppress the dependence of $h_n(x)$ on S_n. Here we establish the convention that if b is undefined then $\min(a,\ b) = \min(b,\ a) = a$. As a consequence of the assumed log-concavity of $f(x)$, $h_n(x)$ is an envelope for $\ln f(x)$, i.e. $h_n(x) \geqslant \ln f(x)$ everywhere in D. This is illustrated for $n = 4$ in Fig. 4. We can now perform rejection sampling with the sampling distribution given by

$$g_n(x) = \frac{1}{m_n} \exp h_n(x) \tag{7}$$

where

$$m_n = \int \exp h_n(x)\, \mathrm{d}x.$$

Note that $g_n(x)$ is a piecewise exponential distribution and can be sampled directly (Gilks and Wild, 1992).

The important feature of the sampling distribution $g_n(x)$ defined in equation (7) is that it can be updated each time that $f(X)$ is evaluated. We have then the following algorithm for ARS:

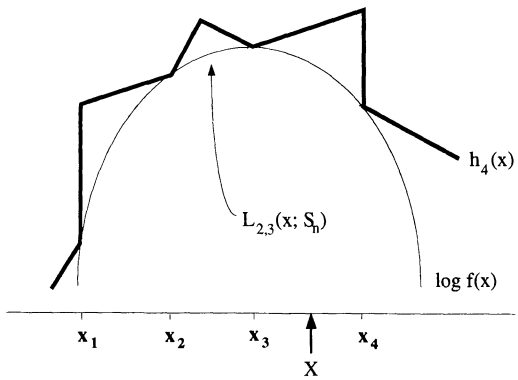

Fig. 4. Adaptive rejection function $h_4(x)$ (——) constructed from equation (6) for a log-concave function $f(x)$: X is sampled from $h_4(x)$

step 0, initialize n and S_n;
step 1, sample X from $g_n(x)$;
step 2, sample U from uniform(0, 1);
step 3, if $U > f(X)/\exp h_n(X)$ then {
rejection step:
set $S_{n+1} = S_n \cup \{X\}$;
relabel points in S_{n+1} in ascending order;
increment n and go back to step 1;}
else {
acceptance step:
set $X_A = X$;};
step 4, return X_A.

The relabelling in step 3 is for notational consistency with equation (6). At each iteration of ARS the number of points of contact between $\ln f(x)$ and $h_n(x)$ is increased by 1, thereby reducing m_n and decreasing the probability of rejection at step 3. This is illustrated in Fig. 5. Further iterations of steps 1–4 will produce independent samples from f, while $h_n(x)$ is continually improving, making rejections increasingly less likely.

This method works straightforwardly if domain D is bounded on the left and right. If D is unbounded on the left, starting abscissae should be chosen so that the gradient of $L_{1,2}(x; S_n)$ is positive. Similarly, if D is unbounded on the right then the gradient of $L_{n-1,n}(x; S_n)$ should be negative.

The average number of iterations of ARS required to accept one point depends on the initial S_n and on f. However, we have found that, starting with $n = 3$, on average just two or three iterations are typically required, although with very poor starting values more may be necessary (Gilks, 1992). If f is a full conditional distribution, the envelope function $\exp h_n(x)$ from the previous iteration of the Gibbs sampler may be used to construct approximate 5%, 50% and 95% centiles of f, for use as starting abscissae, although in many applications fixed starting abscissae will be adequate.

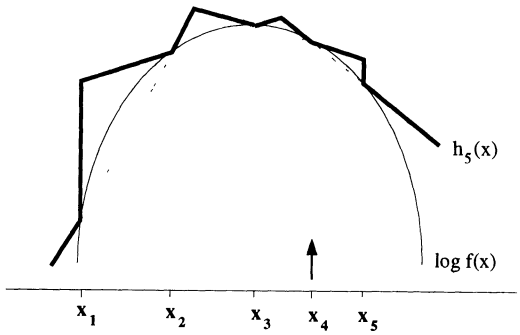

Fig. 5. Updated current set S_5 and rejection function $h_5(x)$ (——) after incorporating X in Fig. 4

For densities $f(x)$ which are not log-concave, ARS cannot be used as $h_n(x)$ may not be an envelope for $\ln f(x)$. To deal with non-log-concave densities we propose to append a Hastings–Metropolis algorithm step to ARS. We first briefly describe the Hastings–Metropolis algorithm.

3.3. *Hastings–Metropolis Algorithm*

The Metropolis algorithm (Metropolis *et al.*, 1953) is, like the Gibbs sampler, an MCMC method. We describe the generalization of the algorithm given by Hastings (1970), which requires a *proposal distribution* $q(\ |z)$ from which samples X can be drawn for any z in D. The algorithm runs as follows:

step 0, set starting value x_0; set iteration counter $i = 0$;
step 1, sample X from $q(x|x_i)$;
step 2, sample U from uniform(0, 1);
step 3, if $U > \min\left\{1, \dfrac{f(X)\,q(x_i|X)}{f(x_i)\,q(X|x_i)}\right\}$ {
rejection step:
set $x_{i+1} = x_i$;}
else {
acceptance step:
set $x_{i+1} = X$;};
step 4, increment i and go back to step 1.

After suitably many iterations of this algorithm, the samples $\{x_i\}$ can be considered to be dependent samples from $f(x)$.

Tierney (1991) suggested the use of the Hastings–Metropolis algorithm within Gibbs sampling to sample from full conditional distributions. Indeed, this was the original form of the Metropolis algorithm (Metropolis *et al.*, 1953). For this x_0 should be the value of x at the start of the current Gibbs iteration, and x_1 will be the new value for x. Just one iteration of steps 1–4 suffices to preserve the stationary distribution of the Gibbs chain. However, this chain may be slower to converge through rejections at step 3.

3.4. *Adaptive Rejection Metropolis Sampling*

We noted in Section 3.2 that ARS cannot be used to sample from non-log-concave distributions. To sample from such distributions, we could abandon rejection sampling in favour of the Hastings–Metropolis algorithm, applied to update one parameter (or one set of parameters) at a time (Section 3.3). However, to avoid high probabilities of rejection (and hence slower convergence of the chain) it may be helpful to adapt the proposal density q to the shape of the full conditional density f (Gelman, 1992). Since ARS provides a way of adapting a function to f, we propose to use ARS to create a good proposal density. We then append to ARS a single Hastings–Metropolis step, thus creating an ARMS within Gibbs chain. However, unlike ARS, ARMS will not produce independent samples from f. ARMS is an adaptive generalization of the rejection sampling chain proposed by Tierney (1991).

Let (x, y) denote the complete set of variables being sampled by the Gibbs sampler. As before, x is the current variable to be sampled from its full conditional density (proportional to) $f(x)$, where we notationally suppress the conditioning on y. Let X_{cur} denote the current value of x at a given iteration of the Gibbs sampler. The aim then is to replace X_{cur} with a new value X_{M} from f.

For ARMS we construct a function $h_n(x)$ which is slightly more complex than in expression (6):

$$h_n(x) = \max[L_{i,i+1}(x, S_n), \min\{L_{i-1,i}(x, S_n), L_{i+1,i+2}(x, S_n)\}], \quad x_i \leqslant x \leqslant x_{i+1}, \tag{8}$$

where, if b is undefined, $\min(a, b) = \min(b, a) = \max(a, b) = \max(b, a) = a$. In general, $h_n(x)$ will not be an envelope of $\ln f(x)$, as illustrated in Fig. 6. The sampling density $g_n(x)$ is then given by equation (7) as before. Starting abscissae for ARMS must be independent of X_{cur}, as discussed below. The algorithm for ARMS then runs as follows:

step 0, initialize n and S_n independently of X_{cur};
step 1, sample X from $g_n(x)$;
step 2, sample U from uniform(0, 1);

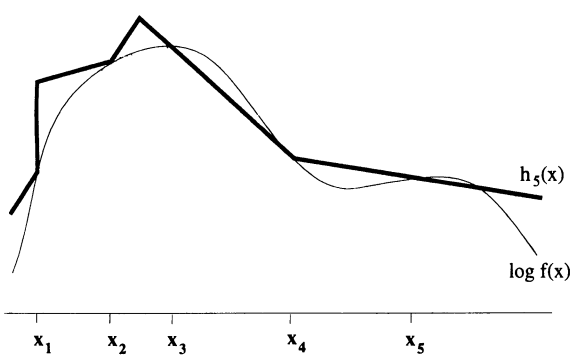

Fig. 6. Adaptive rejection function $h_5(x)$ (——) for a non-log-concave function $f(x)$, constructed from equation (8)

step 3, if $U > f(X)/\exp h_n(X)$ then {
ARS rejection step:
set $S_{n+1} = S_n \cup \{X\}$;
relabel points in S_{n+1} in ascending order;
increment n and go back to step 1;}
else {
ARS acceptance step:
set $X_A = X$;};
step 4, sample U from uniform(0, 1);

$$\textit{step 5, if } U > \min\left[1, \frac{f(X_A)\min\{f(X_{cur}), \exp h_n(X_{cur})\}}{f(X_{cur})\min\{f(X_A), \exp h_n(X_A)\}}\right] \text{ then } \{$$

Hastings–Metropolis rejection step:
set $X_M = X_{cur}$;}
else {
Hastings–Metropolis acceptance step:
set $X_M = X_A$;};
step 6, return X_M.

When f is log-concave then h_n in equation (8) reduces to expression (6) and is an envelope for $\ln f$, so step 5 will always accept. Thus ARMS reduces to ARS for log-concave densities.

The proof that ARMS preserves the stationary distribution of the Gibbs sampler is an application of the auxiliary variables method (Besag and Green, 1993). The whole of the following argument conditions on y, so we shall not express this condition explicitly. Let N denote the value of n on reaching step 4. Then S_N determines the final state of the $h_n(x)$-function. Conditionally on S_N and X_{cur}, X_A is a sample from

$$q(x|X_{cur}, S_N) \propto \min\{f(x), \exp h_N(x)\}. \tag{9}$$

As the starting abscissae for ARMS are drawn independently of X_{cur}, the right-hand side of expression (9) does not depend on X_{cur}. Thus we can consider $q(x|X_{cur}, S_N)$ as an independence Hastings–Metropolis proposal density, where S_N represents auxiliary variables. Let $P(X_M|X_{cur}, S_N)$ be the Markov transition function (for moving from X_{cur} to X_N) associated with the proposal density $q(x|X_{cur}, S_N)$ and the acceptance–rejection function in step 5 above. Then it is straightforward to show that the detailed balance equation

$$f(X_{cur})\,P(X_M|X_{cur}, S_N) = f(X_M)\,P(X_{cur}|X_M, S_N) \tag{10}$$

holds for every S_N, and by integrating equation (10) with respect to X_{cur} we see that X_M is independent of S_N and has density f. Thus X_M is a sample from the full conditional for x, so ARMS within Gibbs sampling preserves the stationary distribution of the Gibbs chain.

This proof depends critically on the independence of S_N and X_{cur}. If starting abscissae were allowed to depend on X_{cur} we would need to replace $f(\ )$ in step 5 and equations (9) and (10) by the awkward density $f(\ |S_N)$, where

$$f(x|S_N) \propto f(x)\,\mathrm{pr}(S_N|x = X_{cur}).$$

There is no need to iterate through steps 1–6 before updating y. However, the probability of moving away from X_{cur} can be increased through a fixed number of additional iterations of ARMS, setting $X_{cur} = X_M$, $n = N$ and $S_n = S_N$ in step 1 at the second and subsequent ARMS iterations. This is permissible because, as noted above, S_N and X_M are independent at the end of step 6, and hence X_{cur} and S_n will be independent at the start of the next ARMS iteration, as required by the theory. However, the stationary distribution of the ARMS within Gibbs chain would be affected if the decision to use extra iterations was allowed to depend in any way on previous rejections at step 5.

The probability of rejection at step 5 can be reduced through choosing good starting abscissae in S_n. An effective way of doing this is to base starting abscissae on the $\exp h_N(x)$ function from the previous Gibbs iteration, as described for ARS in Section 3.2. This strategy for choosing starting abscissae is valid because $h_N(x)$ from the previous Gibbs iteration is independent of X_{cur}. This follows because X_{cur} is identically X_M from the previous Gibbs iteration, X_M and S_N are independent after each iteration of ARMS (as noted above) and S_N defines $h_N(x)$ completely.

This method works straightforwardly if domain D is bounded on the left and right. For D unbounded on the left or right, starting abscissae should be chosen as described for ARS in Section 3.2.

4. Estimation of Gentamicin Pharmacokinetics by using the Adaptive Rejection Metropolis Sampling within Gibbs Sampler

The population pharmacokinetic model for the gentamicin data discussed in Section 2 was estimated by using the ARMS within Gibbs algorithm. The Gibbs sampler was run for 15000 iterations on each of two separate occasions, using different overdispersed starting values for each run. These starting values were selected arbitrarily to be 'too small' for one run and 'too large' for the other, on the basis of physiological considerations and prior information from previous studies of gentamicin pharmacokinetics in infants. For ARMS, for each parameter, six initial abscissae were used, based on the 5%, 30%, 45%, 55%, 70% and 95% centiles of the $h_N(x)$-function from the previous Gibbs iteration (see Section 3.4 for a justification of this). The whole process took a total of 10 h on a Sun SPARC workstation (model 30).

The non-log-concavity of the full conditionals to be sampled from is illustrated in Fig. 7, which shows the logarithm of the full conditional distribution for the regression parameter β_1 at one randomly chosen iteration of the ARMS within Gibbs sampler. Consecutive sampled values for β_1 and β_2 from each run are plotted as time series in Fig. 8. Visual inspection suggests that the chains from the second run converged quite rapidly, but those from the first took over 10000 iterations to stabilize. The method of Gelman and Rubin (1992) was used to assess convergence, and on this basis the last 3000 iterations from each run were pooled to form a sample for posterior inference.

4.1. *Performance of Adaptive Rejection Metropolis Sampling*

Table 2 describes the operating characteristics of ARMS for a selection of the model parameters. The number of density evaluations reported is the number

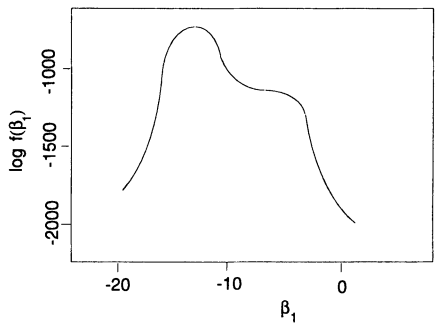

Fig. 7. Full conditional distribution for β_1 at a randomly chosen iteration of the ARMS within Gibbs sampler

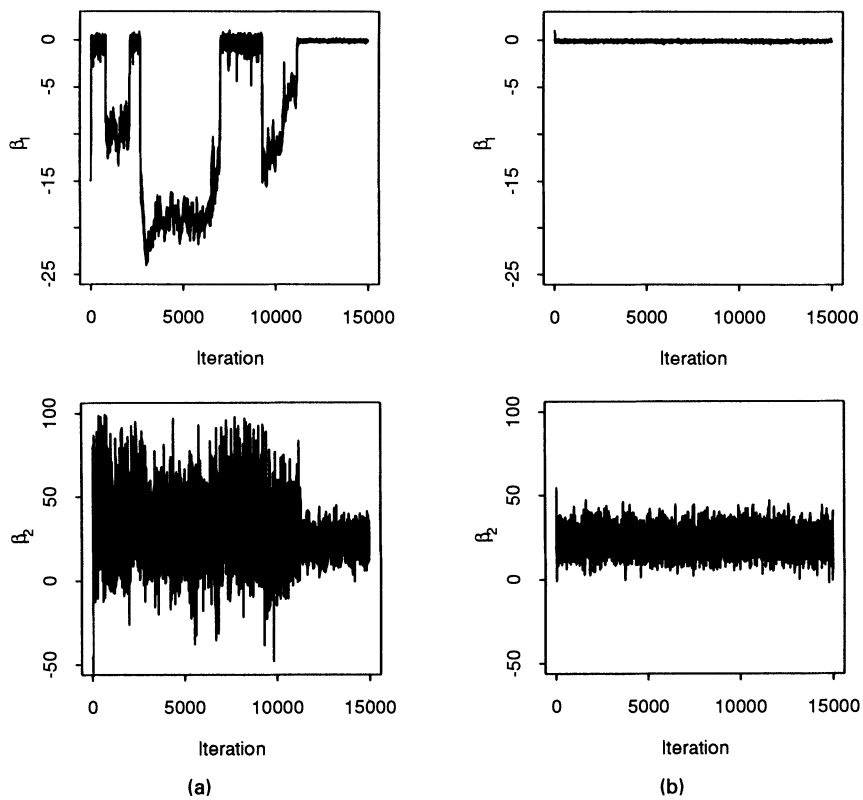

Fig. 8. Consecutively sampled values for two of the regression coefficients (β_1 and β_2) for (a) run 1 and (b) run 2 of the Gibbs sampler

of evaluations of $f(X)$ required to sample one point from the full conditional, including the six evaluations at the starting abscissae. A further indicator of the performance of ARMS is given in the final column of Table 2, which shows the proportion of iterations at which rejection occurred at step 5 of ARMS. This is important as high rejection rates slow convergence of the ARMS within Gibbs chain.

TABLE 2
Operating characteristics of ARMS for selected model parameters

Parameter	*Mean no. of density evaluations*		*95% interval for the no. of density evaluations*		*No.† of iterations at which Metropolis step rejected*	
	Run 1	*Run 2*	*Run 1*	*Run 2*	*Run 1*	*Run 2*
α_1	8.4	8.4	(8, 12)	(8, 12)	0	0
ϕ_1	8.5	8.5	(8, 13)	(8, 13)	0	1
β_1	10.3	10.2	(8, 16)	(8, 16)	16	8
β_2	8.4	8.3	(8, 11)	(8, 11)	0	0
β_3	10.5	10.4	(8, 17)	(8, 17)	8	5
τ	10.0	9.6	(8, 13)	(8, 14)	0	1

†Out of a total of 15000 iterations.

4.2. *Posterior Distributions*

The mean and 95% credible intervals for the posterior distribution of the main parameters are given in Table 3. These results are in line with other published values for gentamicin pharmacokinetic parameters (Thomson *et al.*, 1988). The regression coefficients associated with clearance seem reasonable, with the negative values obtained for β_1 and β_3 corresponding to the expected impairment in gentamicin clearance associated with low APGAR score and immature babies. Similarly, β_2, the effect of reciprocal creatinine concentration, is large and positive which suggests that gentamicin clearance depends on kidney function, in accordance with the known renal elimination of the drug. However, the root residual error variance of 0.38 corresponds to a coefficient of variation of about 38%, which is rather high. This could indicate a lack of model fit: we are currently investigating this.

TABLE 3
Posterior means and 95% credible intervals for selected parameters

Parameter	*Posterior mean*	*95% credible interval*
$\exp\{\alpha_1 + \beta(\mathbf{X}_{1,2} - \bar{\mathbf{X}}\}$: C for patient 1 at time $t_{1,2}$ ($\mathrm{l\,kg^{-1}\,h^{-1}}$)	0.044	(0.031, 0.061)
$\exp\phi_1$: V for patient 1 ($\mathrm{l\,kg^{-1}}$)	0.49	(0.39, 0.61)
$\exp\gamma_1$: population mean of random effects for C ($\mathrm{l\,kg^{-1}\,h^{-1}}$)	0.049	(0.045, 0.053)
$\exp\gamma$: population mean of random effects for V ($\mathrm{l\,kg^{-1}}$)	0.49	(0.43, 0.57)
$\sqrt{\Sigma_{11}} \times 100\%$: approximate coefficient of variation in C	28.2%	(18.9%, 50.2%)
$\sqrt{\Sigma_{22}} \times 100\%$: approximate coefficient of variation in V	13.4%	(5.0%, 48.4%)
ρ: correlation between C and V	0.23	(−0.36, 0.70)
β_1: coefficient for APGAR score	−0.10	(−0.26, 0.06)
β_2: coefficient for reciprocal creatinine	22.9	(10.4, 35.2)
β_3: coefficient for immature babies	−0.22	(−0.37, −0.07)
$\sqrt{(1/\tau)}$: square root of residual error variance	0.38	(0.34, 0.44)

5. Discussion

5.1. *Methodology*

We have shown how ARS can be generalized, through addition of a Hastings–Metropolis step, to sample from non-log-concave distributions encountered in applications of Gibbs sampling. ARMS can also be used straightforwardly in other MCMC methods that require the sampling of full conditional distributions, such as the hit-and-run algorithm (Belisle *et al.*, 1990) and adaptive direction sampling (Gilks and Roberts, 1994).

ARMS will be most computationally efficient when few rejections at the Hastings–Metropolis step are encountered, and this will happen when full conditionals are nearly log-concave. In many situations this is indeed so, where non-log-concavity occurs well into the tails of full conditionals. For log-concave full conditionals, ARMS reduces to ARS. Thus ARMS provides a way of retaining the efficiency of ARS while accommodating non-log-concavity when it is present. As we have shown in the analysis of a robust non-linear pharmacokinetic model, ARMS can accommodate quite severe non-log-concavity. In difficult problems the use of a larger number of ARMS starting abscissae, together with other strategies discussed in Section 3.4, should help to reduce high Hastings–Metropolis rejection rates.

Other approaches to dealing with awkward full conditional distributions have been proposed. The ratio of uniforms method has been successfully used to sample from univariate full conditionals in difficult pharmacokinetic problems (Wakefield *et al.*, 1991, 1994). Sampling from full conditional distributions can be avoided altogether by updating parameters instead via the Hastings–Metropolis algorithm (see Section 3.3). This can be applied to update single parameters, sets of parameters or all parameters simultaneously. Hastings–Metropolis proposal distributions can be designed to approximate full conditional distributions (Gelman, 1992) as in ARMS, but this is not necessary to obtain samples from the required stationary distribution (Tierney, 1991). Indeed, more efficient Markov chain simulation can sometimes be obtained with proposal distributions which differ substantially from full conditionals (Besag and Green, 1993). In special situations where full conditional distributions are all of the same algebraic form, e.g. in lattice-based image analysis models, exploratory work may help to identify efficient proposal distributions. However, as yet there are no generally applicable methods for identifying efficient proposal distributions; conversely there is a real danger of employing very inefficient proposal distributions. Thus, the use of proposal distributions which approximate full conditional distributions may be a relatively safe option, and ARMS is one way of constructing these. Gelfand and Lee (1993) compared several methods of dealing with awkward full conditional distributions, although their performance data for ARS are somewhat at odds with those of Gilks and Wild (1992).

A disadvantage of ARMS is that it relies on single-parameter updating. Multivariate updating through use of carefully tailored Hastings–Metropolis proposal distributions can produce a rapidly mixing Markov chain. In problems with a small or moderate number of parameters, the adaptive approach of Mueller (1992) may work. However, as noted above, generally applicable methods for constructing efficient proposal distributions have yet to be devised. Generally, single-parameter updating can be made more efficient through sensible parameterization of the

model, to reduce posterior correlations. In particular, Wakefield (1993) noted that the volume and clearance parameters in the pharmacokinetic model of Section 2 are approximately independent.

5.2. *Application*

The simple one-compartment pharmacokinetic model employed in this study illustrates the general problem of statistical inference with non-linear mixed effects models that are commonly encountered in pharmacokinetic and pharmacodynamic analysis. Important extensions to this model include the assumption of a t-distribution or a mixture density for the population mean parameters to 'downweight' outlying individuals as well as outlying measurements or to allow for multimodality respectively. More complex structural models such as a two-compartment pharmacokinetic model or a combined pharmacokinetic–pharmacodynamic model are also of major scientific interest. For each model, it should be straightforward to apply the ARMS within Gibbs algorithm described here to estimate the required model parameters.

5.3. *Computer Programs*

Computer programs for ARMS (written in C) and for ARS (written in Fortran) are available from the authors on request.

Acknowledgements

We thank the Editor and two referees for many valuable comments which helped to improve the analysis and the manuscript. We are also grateful to Dr Alison Thomson for providing us with the gentamicin data set.

References

Beal, S. L. (1984) Population pharmacokinetic data and parameter estimation based on their first two statistical moments. *Drug Metab. Rev.*, **15**, 173–193.

Belisle, C., Romeijn, H. and Smith, R. (1990) Hit-and-run algorithms for generating multivariate distributions. *Technical Report 90-18*. Department of Industrial and Operations Engineering, University of Michigan, Ann Arbor.

Besag, J. and Green, P. J. (1993) Spatial statistics and Bayesian computation. *J. R. Statist. Soc.* B, **55**, 25–37.

Casella, G. and George, E. I. (1992) Explaining the Gibbs sampler. *Am. Statistn*, **46**, 167–174.

Davidian, M. and Gallant, A. (1992) Smooth non-parametric maximum likelihood estimation for population pharmacokinetics with application to quinidine. *J. Pharm. Biopharm.*, **20**, 529–556.

DeGroot, M. E. (1970) *Optimal Statistical Decisions*. New York: McGraw-Hill.

Dellaportas, P. and Smith, A. F. M. (1993) Bayesian inference for generalized linear and proportional hazards models via Gibbs sampling. *Appl. Statist.*, **42**, 443–459.

Fattinger, K., Vozeh, S., Olafsson, A., Vlcek, J., Wenk, M. and Follath, F. (1991) Netilmicin in the neonate: population pharmacokinetic analysis and dosing recommendations. *Clin. Pharmacol. Therpeut.*, **50**, 55–65.

Gelfand, A. E. and Lee, T.-M. (1993) Discussion on The Gibbs sampler and other Markov chain Monte Carlo methods. *J. R. Statist. Soc.* B, **55**, 72–73.

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

Gelman, A. (1992) Iterative and non-iterative simulation algorithms. In *Computing Science and Statistics* (ed. H. J. Newton), pp. 433–438. Fairfax Station: Interface Foundation of North America.

Gelman, A. and Rubin, D. B. (1992) Inference from iterative simulation using multiple sequences. *Statist. Sci.*, **7**, 457–472.

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.

George, E. I., Makov, U. E. and Smith, A. F. M. (1993) Conjugate likelihood distributions. *Technical Report*. University of Chicago, Chicago.

Gilks, W. R. (1992) Derivative-free adaptive rejection sampling for Gibbs sampling. In *Bayesian Statistics 4* (eds J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith), pp. 641–649. Oxford: Clarendon.

Gilks, W. R., Roberts, G. O. and George, E. I. (1994) Adaptive direction sampling. *Statistician*, **43**, 179–189.

Gilks, W. R. and Wild, P. (1992) Adaptive rejection sampling for Gibbs sampling. *Appl. Statist.*, **41**, 337–348.

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

Mallet, A. (1986) A maximum likelihood estimation method for random coefficient regression models. *Biometrika*, **73**, 645–656.

Metropolis, M., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. and Teller, E. (1953) Equations of state calculations by fast computing machines. *J. Chem. Phys.*, **21**, 1087–1093.

Mueller, P. (1992) Alternatives to the Gibbs sampling scheme. *Technical Report*. Institute of Statistics and Decision Sciences, Duke University, Durham.

Racine-Poon, A. (1985) A Bayesian approach to nonlinear random effects models. *Biometrics*, **41**, 1015–1024.

Ripley, B. D. (1987) *Stochastic Simulation*. New York: Wiley.

Sheiner, L. B. and Beal, S. L. (1980) Evaluation of methods for estimating population pharmacokinetic parameters. I, Michaelis–Menten model: clinical pharmacokinetic data. *J. Pharm. Biopharm.*, **8**, 553–571.

Steimer, J., Mallet, A., Golmard, J. and Boisvieux, J. (1984) Alternative approaches to estimation of population pharmacokinetic parameters: comparison with nonlinear mixed-effects model. *Drug Metab. Rev.*, **15**, 265–292.

Thomson, A. H., Way, S., Bryson, S. M., McGovern, E. M., Kelman, A. W. and Whiting, B. (1988) Population pharmacokinetics of gentamicin in neonates. *Dev. Pharmacol. Ther.*, **11**, 173–179.

Tierney, L. (1991) Exploring posterior distributions using Markov chains. In *Computer Science and Statistics: Proc 23rd Symp. Interface* (ed. E. Keramidas), pp. 563–570. Fairfax Station: Interface Foundation.

Wakefield, J. (1993) Discussion on The Gibbs sampler and other Markov chain Monte Carlo methods. *J. R. Statist. Soc.* B, **55**, 56–57.

Wakefield, J. C., Gelfand, A. E. and Smith, A. F. M. (1991) Efficient generation of random variates via the ratio-of-uniforms method. *Statist. Comput.*, **1**, 129–133.

Wakefield, J. C., Smith, A. F. M., Racine-Poon, A. and Gelfand, A. E. (1994) Bayesian analysis of linear and non-linear population models by using the Gibbs sampler. *Appl. Statist.*, **43**, 201–222.

Whittaker, J. (1990) *Graphical Models in Applied Multivariate Statistics*. Chichester: Wiley.

Zaske, D. E. (1992) Aminoglycosides. In *Applied Pharmacokinetics: Principles of Therapeutic Drug Monitoring* (eds W. E. Evans, J. J. F. Schentag and W. J. Jusko), ch. 14. Spokane: Applied Therapeutics Inc.