BAYESIAN ANALYSIS OF LINEAR AND NONLINEAR POPULATION MODELS USING THE GIBBS SAMPLER

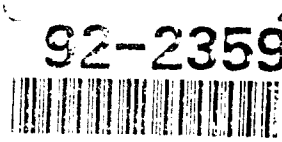
By
J.C. Wakefield
A.F.M. Smith
A. Racine-Poon
A.E. Gelfand

TECHNICAL REPORT No. 456
JULY 21, 1992

Prepared Under Contract
N00014-92-J-1254 (NR-042-267)
FOR THE OFFICE OF NAVAL RESEARCH

Herbert Solomon, Project Director

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305



BAYESIAN ANALYSIS OF LINEAR AND NONLINEAR POPULATION MODELS USING THE GIBBS SAMPLER

J C Wakefield[1]
A F M Smith[1]
A Racine-Poon[2]
A E Gelfand[3]

SUMMARY

A fully Bayesian analysis of linear and nonlinear population models has previously been unavailable, as a consequence of the seeming impossibility of performing the necessary numerical integrations in the complex multi-parameter structures typically arising in such models. It is demonstrated that, for a variety of linear and nonlinear population models, a fully Bayesian analysis can be implemented in a straightforward manner using the Gibbs sampler. The approach is illustrated with examples involving challenging problems of outliers and mean-variance relationships in population modelling.

Keywords: Population models; hierarchical models; Bayesian analysis; Gibbs sampler; population outliers; nonlinear models; mean-variance relationships.

(1) Department of Mathematics, Imperial College London
(2) Biometrics Section, CIBA-GEIGY AG, Basle
(3) Department of Statistics, University of Connecticut

1 INTRODUCTION

1.1 Population Models

Population models are widely used in biometrical growth analysis (see, for example, Berkey, 1982, Lange, Carlin and Gelfand, 1992), in pharmacokinetic studies as part of drug development procedures (see, for example, Beal and Sheiner, 1980, Lindstrom and Bates, 1990), and have a long history of use in educational research (Novick *et al*, 1972), econometrics (Swamy, 1970) and other fields. Related models are now increasingly used for multi-centre clinical trials (Skene and Wakefield, 1990) and for spatial epidemiology studies (Besag *et al*, 1991).

From a Bayesian perspective, such models are variations on and extensions of the following hierarchical structure. Let $\underset{\sim}{y}$ denote the totality of measurement data on I individuals in a designated population (for example, of patients, experimental animals, firms, *etc.*); let $\underset{\sim}{\theta}$ denote the parameters defining I underlying 'response' profiles (for example, weight versus age, drug concentration versus time after administration, profits versus structural variables defining a firm, *etc.*) and $\underset{\sim}{\phi}$ denote hyperparameters defining relationships among components of $\underset{\sim}{\theta}$. Then, population models correspond to the hierarchy of distributions

$$[\underset{\sim}{y}|\underset{\sim}{\theta}] \ , \ [\underset{\sim}{\theta}|\underset{\sim}{\phi}] \ , \ [\underset{\sim}{\phi}] \ , \tag{1.1}$$

where, adopting the notation of Gelfand and Smith (1990), joint, conditional and marginal densities for random quantities, $\underset{\sim}{u}$, $\underset{\sim}{v}$ are denoted, respectively, by $[\underset{\sim}{u}, \underset{\sim}{v}]$, $[\underset{\sim}{u}|\underset{\sim}{v}]$, $[\underset{\sim}{u}]$, $[\underset{\sim}{v}]$.

In the context of (1.1), interest may centre on inference for components of $\underset{\sim}{\theta}$ (*i.e.* relating to aspects of specific individual profiles), or for $\underset{\sim}{\phi}$ (*i.e.* relating to population characteristics), or on predictions of future observations from an already included individual or a new individual drawn from the same population. In all cases, the integrals required for a fully Bayesian analysis are typically not available in closed form and numerical or analytic approximation is required. Hitherto, however, no approximation approach has been found to be entirely satistfactory. The purpose of this paper is to demonstrate that a highly effective Bayesian computation strategy

for general population model analysis is available, based on the Gibbs sampler.

1.2 Structure of the paper

In Section 2, we consider in detail two population model examples (one linear, one non-linear), which pose challenging problems going beyond basic population analysis, by modelling and analyzing population outliers and mean-variance relationships. In Section 3, we provide a description of the Gibbs sampler approach to Bayesian calculations for hierarchical models. In Section 4, we analyse in detail the linear model example, exhibiting, in particular, a method for detecting population outliers. In Section 5, we analyse in detail the nonlinear model example. We show, in particular, that the inclusion of mean-variance relationships causes little additional computational difficulty with the Gibbs sampler approach. The key message in both Sections 4 and 5 is that the seemingly intractable calculations associated with the Bayesian analysis of population models do indeed become relatively straightforward under the Gibbs sampler approach. In Section 6, we put the Gibbs Sampler approach in perspective by commenting briefly on other available alternatives.

2 ILLUSTRATIVE EXAMPLES

2.1 A linear population biological growth example

Table 1 records dental measurements of the distance (in mm) from the centre of the pituitary to the pteryo-maxillary fissure in 11 girls and 16 boys at the ages of 8, 10, 12 and 14 years. Both in the original analysis (Potthoff and Roy, 1964) and in a Bayesian reanalysis (Fearn, 1975), a linear growth relationship between the dental measurement and age was assumed. This is also assumed in our subsequent analysis, together with homoscedastic normal errors within each separate population (girls, boys). Let x_{ij}, y_{ij} denote, respectively, the jth time point (using age 11 as origin) and associated measurement on the ith individual ($i = 1,\ldots,11$ for the population of girls, $i = 12,\ldots,27$ for the population of boys, $j = 1,\ldots,4$).

Table 1 Here

For both the girl and boy populations the first stage model of (1.1) takes the form

$$\prod_i \prod_j [y_{ij}|\underset{\sim}{\theta}_i, \tau] = \prod_i \prod_j [y_{ij}|\alpha_i, \beta_i, \tau] \qquad (1.2)$$
$$= \prod_i \prod_j N(y_{ij}|\alpha_i + \beta_i x_{ij}, \tau) \quad ,$$

where τ denotes the common normal measurement precision (reciprocal variance) and $\underset{\sim}{\theta}_i = (\alpha_i, \beta_i)$ denotes the intercept and slope for the ith individual's straight-line growth curve.

Fearn (1975) takes as the second stage (population) distribution for the $\underset{\sim}{\theta}_i$'s a bivariate normal distribution (separately, for each of the girl and boy populations), so that

$$\prod_i [\underset{\sim}{\theta}_i|\underset{\sim}{\phi}] = \prod_i N(\underset{\sim}{\theta}_i|\underset{\sim}{\mu}, \underset{\sim}{\Sigma}) \quad ,$$

with $\underset{\sim}{\phi} = (\underset{\sim}{\mu}, \underset{\sim}{\Sigma})$, where $\underset{\sim}{\mu} = (\mu_1, \mu_2)$, $E(\alpha_i) = \mu_1$, $E(\beta_i) = \mu_2$, so that the individual straight-line growth curves are, in effect, regarded as distributed around a 'mean' population growth curve, $\mu_1 + \mu_2 x$ with population variation described by the 2 x 2 covariance matrix $\underset{\sim}{\Sigma}$. (We digress to note that since the β_i are positive in this case it might be more reasonable to assume log β_i to be normally distributed. Such a refinement is not, in fact, important in this example and so we shall not pursue it, in order to keep this initial exposition as simple as possible. We shall illustrate such a transformation in our second, nonlinear, example.) In what follows, we shall denote μ_1 by α_G (α_B) for the girl (boy) populations and μ_2 by β_G (β_B).

Since information from individuals within each population is effectively 'pooled' to give population 'mean' inferences, it can be important in such studies to guard against an aberrant or 'outlying' individual unduly influencing the population inference. Proceeding naively, examining, for example, the pooled population of girls and boys, one might plot the least-squares estimates of intercepts and slopes, as in Figure 1. Should one conclude from the plot that the boy labelled 24 is a 'slope outlier'? Or that the boy labelled 21 is an 'intercept outlier'? We seek a modelling analysis

strategy which will provide both a coherent outlier detection diagnostic and direct inferences which accommodate the effect of any outliers present.

Figure 1 Here

The strategy we shall adopt is to replace the population bivariate normal assumption for the $\underset{\sim}{\theta}_i$'s by a bivariate Student-t assumption (see, for example, Smith, 1983, O'Hagan, 1987, for general discussions of modelling with heavy-tailed distributions). The hierarchical model is then completed by assigning priors to τ, $\underset{\sim}{\mu}$ and $\underset{\sim}{\Sigma}$. As we shall show in Section 4, analysis of this model (perhaps surprisingly) is still easily implemented via the Gibbs sampler and provides a novel form of graphical diagnostic for second-stage outliers in hierarchical models.

The main inference questions in this study relate to differences in growth between the girl and boy populations. We shall provide illustrative analyses of this in Section 4, taking into account the outlier issue discussed above.

2.2 A nonlinear population pharmacokinetic example

Table 2 presents pharmacokinetic data on the plasma concentration of the drug Cadralazine in 10 cardiac failure patients at various times after the administration of a single dose of 30mg.

Table 2 Here

The starting point for modelling the first stage of the hierarchy in this case is the one-compartment nonlinear model for individual plasma concentrations (y_{ij}) against time (x_{ij}) (see Racine *et al*, 1986), which implies that

$$\text{plasma concentration} = 30 \times \alpha_i^{-1} \exp(-\beta_i \times \text{time}) \quad , \qquad (2.1)$$

where α_i, β_i (> 0) are, respectively, the volume of distribution and elimination rate for individual i, $i = 1,\ldots,10$.

Measurement variance is certainly related to underlying concentration level in studies such as this, so that a simple additive homoscedastic normal error assumption for the first stage distribution is inappropriate. We shall illustrate possible models and their subsequent analyses by considering the following three intra-individual error structures.

As a first possibility, letting $\underset{\sim}{\psi}_i = (\alpha_i, \beta_i)$ and denoting the right-hand-side of (2.1) by $\eta_{ij}(\underset{\sim}{\psi}_i)$, we assume that

$$\log y_{ij} = \log \eta_{ij}(\underset{\sim}{\psi}_i) + \varepsilon_{ij} \quad ,$$

with independent normal errors having zero mean and constant variance τ_i^{-1}.

A second modelling possibility is to assume that

$$y_{ij} = \eta_{ij}(\underset{\sim}{\psi}_i) + \varepsilon_{ij} \quad ,$$

with independent normal errors having zero mean and variances given by

$$[\eta_{ij}(\underset{\sim}{\psi}_i)]^{\gamma} \tau_i^{-1} \quad , \tag{2.2}$$

so that $\gamma \geq 0$ indexes a power law relationship between the variance and the mean. We shall refer to these two variance models as the lognormal model and the power model.

In Section 5 we shall show, in fact, that neither of these formulations is adequate for the Cadralazine data. We now describe a third, more complex error model.

Careful study of data resulting from the analytical assay technique revealed that the variance became approximately constant for low concentrations, but increased as a function of the mean for larger concentrations. To model this behaviour directly would require a 'cut-off' point for concentrations, below which the variance was assumed constant and above which the power model was used. Such a model requires additional parameters, so instead the power model (2.2) was used, but it was assumed that the error distribution is a truncated normal distribution with $y_{ij} \geq 0$. This model reproduces the correct behaviour and also ensures that predictive

distributions will always produce non-negative concentrations.

For the (population) second stage of the hierarchical model we define $\underset{\sim}{\theta}_i = (\log \alpha_i, \log \beta_i)$ and assume that the the $\underset{\sim}{\theta}_i$'s follow a bivariate normal distribution with mean $\underset{\sim}{\mu} = (\mu_1, \mu_2)$ and covariance matrix $\underset{\sim}{\Sigma}$. The hierarchical model is then completed by assigning priors to the elements of $\underset{\sim}{\tau}$, $\underset{\sim}{\mu}$, $\underset{\sim}{\Sigma}$ and, for the second and third of the above models, γ. We shall show in Section 5 that the Gibbs sampler again permits relatively straightforward implementation, despite the complications of nonlinearity, mean-variance relationships and parameter transformations in specifying a population distribution.

In pharmacokinetic studies the main questions relate to population inferences and/or inferences for future concentration levels. Though not relevant to our particular example, the former question may relate to the identification of important covariates such as age, weight and sex. Interest may focus on nonlinear functions of the population parameters, for example, the so-called clearance parameter (defined for an individual by $\alpha_i\beta_i$) and the elimination half-life (defined for an individual by $\log 2/\beta_i$). Issues arise here concerning the best summaries of the appropriate population analogue (mean? mode? median? quantiles?). In fact, from a Bayesian perspective the single most relevant summary will often be the predictive distribution (for example, of the half-life) for a new individual from the population, or from a subpopulation (typically conditioning on covariates). A predictive distribution for concentration can, in particular, enable the benefits of a candidate dosage regimen to be investigated.

3. THE GIBBS SAMPLER

Suppose that the joint probability structure for a collection of random variables $U_1, \ldots, U_k$ is such that the joint density $[U_1, \ldots, U_k]$ is uniquely determined by the full conditional densities $[U_s | U_r, r \neq s]$, $s = 1, \ldots, k$. Suppose that samples of U_s can be generated efficiently from $[U_s | U_r, r \neq s]$ given specified values of the conditioning variables, U_r, $r \neq s$. An algorithm for extracting information from these full conditional distributions in order to estimate the marginal distributions, $[U_s]$, $s = 1, \ldots, k$, has been discussed by Hastings (1970) and Geman and Geman (1984). This so-called Gibbs sampler

algorithm, further developed and illustrated in Gelfand and Smith (1990) and Gelfand *et al.* (1990), is a Markovian updating scheme which proceeds as follows.

Given arbitrary staring values, $U_1^{(0)},\ldots,U_k^{(0)}$, for the k random variables, we generate a random variate $U_1^{(1)}$ from $[U_1|U_2^{(0)},\ldots,U_k^{(0)}]$, followed by $U_2^{(1)}$ from $[U_2|U_1^{(1)}, U_3^{(0)},\ldots,U_k^{(0)}]$, and so on up to $U_k^{(1)}$ from $[U_k|U_1^{(1)},\ldots,U_{k-1}^{(1)}]$. This completes one iteration of the sampling scheme. After t such iterations, we would arrive at a joint sample $(U_1^{(t)},\ldots,U_k^{(t)})$. As $t \to \infty$, Geman and Geman show (under rather mild regularity conditions) that this tends in distribution to a variable having the joint distribution $[U_1,\ldots,U_k]$. Suppose that we have m independent realizations of $[U_1,\ldots,U_k]$. Such realizations could arise from the reiplication of the iterative cycle m times or from the use of a smaller number of such cycles from which, for t sufficiently large, realizations are extracted a number of iterations apart, the gap being large enough to ensure the 'independence' of the subsequent samples. See for example, Raftery and Lewis (1992). We note that the parameterization of the model is relevant to the choice of Gibbs strategy. See Hills and Smith (1992) for general comment, Wakefield (1992) for comments specific to the application in Section 5. The replicate values for $U_s, U_{s1},\ldots,U_{sm}$, say, can be regarded as a sample of size m from $[U_s]$. Based on the m iid k-tuples $(U_{1j},\ldots,U_{kj})$, $j = 1,\ldots,m$, the marginal density $[U_s]$ can be approximated by the finite mixture density

$$[\hat{U}_s] = m^{-1} \sum_{j=1}^{m} [\hat{U}_s | U_r = U_{rj},\ r \neq s] \quad ,$$

if the right-hand side is explicitly available. Alternatively, an estimate $[\hat{U}_s]$ could be obtained using a kernel density estimate based on $U_{s1},\ldots,U_{sm}$. Moreover, if interest centres on the marginal distribution for a random variable $V = g(U_1,\ldots,U_k)$, for an arbitrary function g, we note that evaluation of $g(U_{1j},\ldots,U_{kj})$, $j = 1,\ldots,m$, directly provides a sample $V_1,\ldots,V_m$, so that $[\hat{V}]$ is immediately available from a kernel density estimate. See Gelfand and Smith, 1990, for further discussion.

Turning specifically now to Bayesian applications, suppose that $\underset{\sim}{\psi} = (\psi_1, \ldots, \psi_k)$ is a parameter vector of interest and that, given $h(\underset{\sim}{\psi}) \propto [\underset{\sim}{\psi}|\text{data}]$ we wish to evaluate $[\psi_i|\text{data}]$, for some or all of $i = 1, \ldots, k$. Successively regarding $h(\underset{\sim}{\psi})$ as a function of ψ_i for fixed ψ_j, $j \neq i$, immediately identifies functions $h_i(\psi_i|\psi_j,\ j \neq i)$, $i = 1, \ldots, k$, which are such that

$$h_i(\psi_i|\psi_j,\ j \neq i) \propto [\psi_i|\psi_j,\ j \neq i,\ \text{data}] \quad .$$

From knowledge of $h_i(\psi_i|\psi_j,\ j \neq i)$ we can sample from $[\psi_i|\psi_j,\ j \neq i,\ \text{data}]$, so that the Gibbs sampling algorithm is seen to provide a general solution to the problem of calculating marginal densities given the specification of a likelihood and prior.

In fact, very general methods (including ratio-of-uniforms and envelope rejection techniques) are available for random variate generation, given only the form, up to proportionality, of the density function (see, for example, Ripley, 1987). We shall describe a recent adaptation of the ratio-of-uniforms method in Section 5. If inferences are only required in the form of summaries of posterior densities, for example, moments or quantiles, these are obtained in a straightforward manner from the samples generated by the Gibbs sampler (see Gelfand and Smith, 1991).

To examine the structure of the Gibbs sampler in the context of population hierarchical models, consider first a very general Bayesian hierarchical model having K stages. In an obvious notation, the joint distribution of data, Y, and the parameters $\omega_1, \ldots, \omega_k$ from each of the stages is given by

$$[Y|\omega_1]*[\omega_1|\omega_2]*\ldots*[\omega_{K-1}|\omega_K] \quad .$$

Suppose now that inferential interest centres on $[\omega_i|Y]$, $i = 1, \ldots, K$ and that we wish to calculate these using the Gibbs sampler, so that we need to sample iteratively from the full conditionals $[\omega_s|Y,\ \omega_r,\ r \neq s]$.

Examination of the hierachical structure (a Markov random field with an 'adjacent' neighbourhood system) reveals that

$$[\omega_s|Y,\ \omega_r,\ r \neq s] = \begin{cases} [\omega_1|Y,\ \omega_2] & s = 1 \\ [\omega_s|Y,\ \omega_{s-1},\ \omega_{s+1}] & 2 \leq s \leq K-1 \\ [\omega_K|Y,\ \omega_{K-1}] & s = K \end{cases},$$

so that there is typically considerable simplification in the forms of the full conditional distributions, greatly facilitating the implementation of the Gibbs sampler. In the following two sections, we shall illustrate the power and relative simplicity of the latter by considering the two particular population problems introduced earlier.

4. ANALYSIS OF THE LINEAR GROWTH EXAMPLE

We begin, in Section 4.1, by reviewing the case where the population (second-stage) distribution is assumed to be normal, so that each population follows the hierarchical normal-linear model (Lindley and Smith, 1972). We then, in Section 4.2, extend this linear case to incorporate the outlier accommodation modelling discussed in Section 2.1 by means of a Student-t population distribution. An illustrative analysis of the data presented in Section 2.1 is given in Section 4.3.

4.1 A normal-linear population model

To illustrate the structure (1.1) in the case of the normal-linear hierarchical model, suppose that the first-stage model has the form

$$\prod_{i=1}^{I} \prod_{j=1}^{n_i} [y_{ij}|\underset{\sim}{\theta}_i,\ \tau] = \prod_{i=1}^{I} N(\underset{\sim}{y}_i|\underset{\sim}{X}_i\underset{\sim}{\theta}_i,\ \tau^{-1}\underset{\sim}{I}_{n_i}) \quad ,$$

so that, $\underset{\sim}{y}_i = (y_{i1}, \ldots, y_{in_i})$, the ith of I observation vectors is modelled as linear structure $\underset{\sim}{X}_i\underset{\sim}{\theta}_i$, $\underset{\sim}{\theta}_i$, $p \times 1$, with conditionally independent homoscedastic normal errors having variances τ^{-1}. Suppose further that the second-stage structure is given by

$$\prod_{i=1}^{I} [\underset{\sim}{\theta}_i|\underset{\sim}{\phi}] = \prod_{i=1}^{I} N(\underset{\sim}{\theta}_i|\underset{\sim}{\mu},\ \underset{\sim}{\Sigma}) \quad ,$$

so that the first-stage regression parameter vectors are taken to be a random sample from a 'population distribution', $N(\underset{\sim}{\mu}, \underset{\sim}{\Sigma})$, together with

$$[\tau] = G(\tau | \tfrac{1}{2}\nu_0, \tfrac{1}{2}\nu_0\tau_0) \quad ,$$

corresponding to an inverse-gamma prior for the common variance τ^{-1}.

Finally, suppose that the prior specification is completed by assuming ν_0, τ_0 known and taking the third-stage of the hierarchy to have the form

$$[\underset{\sim}{\phi}] = [\underset{\sim}{\mu}][\underset{\sim}{\Sigma}^{-1}] = N(\underset{\sim}{\mu} | \underset{\sim}{\eta}, \underset{\sim}{C}) W(\underset{\sim}{\Sigma}^{-1} | (\rho \underset{\sim}{R})^{-1}, \rho) \quad ,$$

with $\underset{\sim}{\eta}$, $\underset{\sim}{C}$, ρ and $\underset{\sim}{R}$ known, and W denoting a Wishart distribution. The Wishart distribution is used here for convenience (sampling from it is straightforward). The parameters ρ and $\underset{\sim}{R}$ are chosen apriori: $\rho = p$ is most non-informative in the sense that its distribution is flattest; the matrix $\underset{\sim}{R}$ is chosen to be an approximate estimate of $\underset{\sim}{\Sigma}$.

Defining $\underset{\sim}{y} = (\underset{\sim}{y}_1, \ldots, \underset{\sim}{y}_I)$, $\underset{\sim}{\theta} = (\underset{\sim}{\theta}_1, \ldots, \underset{\sim}{\theta}_I)$, $\bar{\underset{\sim}{\theta}} = I^{-1} \sum_i \underset{\sim}{\theta}_i$, $\underset{\sim}{D}_i^{-1} = \tau \underset{\sim}{X}_i^T \underset{\sim}{X}_i + \underset{\sim}{\Sigma}^{-1}$, $\underset{\sim}{V}^{-1} = I\underset{\sim}{\Sigma}^{-1} + \underset{\sim}{C}^{-1}$, and treating $\underset{\sim}{\psi} = (\underset{\sim}{\theta}_1, \ldots, \underset{\sim}{\theta}_I, \underset{\sim}{\mu}, \underset{\sim}{\Sigma}, \tau)$ as an unknown parameter of dimension $p(I + 1) + \tfrac{1}{2}p(p + 1) + 1$, it is easily seen that a Gibbs sampler is defined by the following conditional distributions:

$$[\underset{\sim}{\theta}_i | \underset{\sim}{y}, \underset{\sim}{\mu}, \underset{\sim}{\Sigma}^{-1}, \tau, \underset{\sim}{\theta}_j, j \neq i] = N(\underset{\sim}{\theta}_i | \underset{\sim}{D}_i (\tau \underset{\sim}{X}_i^T \underset{\sim}{y}_i + \underset{\sim}{\Sigma}^{-1} \underset{\sim}{\mu}), \underset{\sim}{D}_i) \quad , \; i = 1, \ldots, I \quad ,$$

$$[\underset{\sim}{\mu} | \underset{\sim}{y}, \underset{\sim}{\theta}, \underset{\sim}{\Sigma}^{-1}, \tau] = N(\underset{\sim}{\mu} | \underset{\sim}{V} (I\underset{\sim}{\Sigma}^{-1} \bar{\underset{\sim}{\theta}} + \underset{\sim}{C}^{-1} \underset{\sim}{\eta}), \underset{\sim}{V}) \quad ,$$

$$[\underset{\sim}{\Sigma}^{-1} | \underset{\sim}{y}, \underset{\sim}{\theta}, \underset{\sim}{\mu}, \tau] = W(\underset{\sim}{\Sigma}^{-1} | [\sum_i (\underset{\sim}{\theta}_i - \underset{\sim}{\mu})(\underset{\sim}{\theta}_i - \underset{\sim}{\mu})^T + \rho \underset{\sim}{R}]^{-1}, I + \rho) \quad ,$$

$$[\tau | \underset{\sim}{y}, \underset{\sim}{\theta}, \underset{\sim}{\mu}, \underset{\sim}{\Sigma}^{-1}] = G(\tau | \tfrac{1}{2}(\nu_0 + n), \tfrac{1}{2}[\sum_i (\underset{\sim}{y}_i - \underset{\sim}{X}_i \underset{\sim}{\theta}_i)^T (\underset{\sim}{y}_i - \underset{\sim}{X}_i \underset{\sim}{\theta}_i) + \nu_0 \tau_0]) \quad .$$

Generation of random variates is straightforwardly achieved for the normal and gamma distributions: generation for the Wishart distribution is achieved using the algorithm given in Odell and Fieveson (1966). See Gelfand *et al* (1990) for further details and an application.

4.2 A Student-linear population model

To robustify the second-stage of the hierarchical model given in Section 2.3, suppose that we wish to replace the assumption that the $\underset{\sim}{\theta}_i$ are a random sample from an $N(\underset{\sim}{\mu}, \underset{\sim}{\Sigma})$ distribution by the assumption that they are a random sample from $St_\nu(\underset{\sim}{\mu}, \underset{\sim}{\Sigma})$, a multivariate Student-t distribution with mean $\underset{\sim}{\mu}$, covariance matrix $\underset{\sim}{\Sigma}$ and degrees of freedom ν.

One way of representing such an assumption in the structure (1.2), is to take

$$\prod_{i=1}^{I} [\underset{\sim}{\theta}_i|\underset{\sim}{\phi}] = \prod_{i=1}^{I} N(\underset{\sim}{\theta}_i|\underset{\sim}{\mu}, \lambda_i^{-1}\underset{\sim}{\Sigma}) \quad ,$$

so that now $\underset{\sim}{\phi} = (\underset{\sim}{\mu}, \underset{\sim}{\Sigma}, \lambda_1, \ldots, \lambda_I)$, and then subsequently to assume that

$$[\underset{\sim}{\phi}] = [\underset{\sim}{\mu}][\underset{\sim}{\Sigma}][\lambda_1]\ldots[\lambda_I] \quad ,$$

where $[\lambda_i]$ is defined by $[\nu\lambda_i] = G(\nu\lambda_i|\tfrac{1}{2}\nu, \tfrac{1}{2})$ $(= \chi^2_\nu)$. See, also, Racine-Poon (1992). The remaining hierarchical structure is defined exactly as in Section 2.3. For references establishing that $[\underset{\sim}{\theta}_i|\underset{\sim}{\mu}, \underset{\sim}{\Sigma}, \lambda_i] = N(\underset{\sim}{\theta}_i|\underset{\sim}{\mu}, \lambda_i^{-1}\underset{\sim}{\Sigma})$ and $[\nu\lambda_i] = G(\nu\lambda_i|\tfrac{1}{2}\nu, \tfrac{1}{2})$ generate $[\underset{\sim}{\theta}_i|\underset{\sim}{\mu}, \underset{\sim}{\Sigma}] = St_\nu(\underset{\sim}{\theta}_i|\underset{\sim}{\mu}, \underset{\sim}{\Sigma})$, see Johnson and Kotz (1972, Chapter 27, §3-4).

Defining $\underset{\sim}{C}_i^{-1} = \tau\underset{\sim}{X}_i^T\underset{\sim}{X}_i + \lambda_i\underset{\sim}{\Sigma}^{-1}$, $\underset{\sim}{U}^{-1} = \underset{\sim}{\Sigma}^{-1}\sum_i \lambda_i + \underset{\sim}{C}^{-1}$, $\underset{\sim}{\lambda} = (\lambda_1, \ldots, \lambda_I)$, the full conditionals defining the Gibbs sampler become

$$[\underset{\sim}{\theta}_i|\underset{\sim}{y}, \underset{\sim}{\mu}, \underset{\sim}{\Sigma}^{-1}, \tau, \underset{\sim}{\lambda}, \underset{\sim}{\theta}_j, j\neq i] = N(\underset{\sim}{\theta}_i|\underset{\sim}{C}_i(\tau\underset{\sim}{X}_i^T\underset{\sim}{y}_i + \lambda_i\underset{\sim}{\Sigma}^{-1}\underset{\sim}{\mu}), \underset{\sim}{C}_i) \quad i = 1, \ldots, I \quad ,$$

$$[\underset{\sim}{\mu}|\underset{\sim}{y}, \underset{\sim}{\theta}, \underset{\sim}{\Sigma}^{-1}, \tau, \underset{\sim}{\lambda}] = N(\underset{\sim}{\mu}|\underset{\sim}{U}(\underset{\sim}{\Sigma}^{-1}\sum_i \lambda_i\underset{\sim}{\theta}_i + \underset{\sim}{C}^{-1}\underset{\sim}{\eta}), \underset{\sim}{U})$$

$$[\underset{\sim}{\Sigma}^{-1}|\underset{\sim}{y}, \underset{\sim}{\theta}, \underset{\sim}{\mu}, \tau, \underset{\sim}{\lambda}] = W(\underset{\sim}{\Sigma}^{-1}|[\sum_i \lambda_i(\underset{\sim}{\theta}_i - \underset{\sim}{\mu})(\underset{\sim}{\theta}_i - \underset{\sim}{\mu})^T + \rho\underset{\sim}{R}]^{-1}, I + \rho)$$

$$[\tau|\underset{\sim}{y}, \underset{\sim}{\theta}, \underset{\sim}{\mu}, \underset{\sim}{\Sigma}^{-1}, \underset{\sim}{\lambda}] = G(\tau|\tfrac{1}{2}(\nu_0 + n), \tfrac{1}{2}[\sum_i \underset{\sim}{y}_i - \underset{\sim}{X}_i\underset{\sim}{\theta}_i)^T(\underset{\sim}{y}_i - \underset{\sim}{X}_i\underset{\sim}{\theta}_i) + \nu_0\tau_0])$$

$$[\lambda_i|\underset{\sim}{y}, \underset{\sim}{\theta}, \underset{\sim}{\mu}, \underset{\sim}{\Sigma}^{-1}, \tau, \lambda_j, j \neq i] = G(\nu_i\lambda_i|\tfrac{1}{2}(\nu + p), \tfrac{1}{2}) \quad ,$$

where

$$\nu_i = (\underset{\sim}{\theta}_i - \underset{\sim}{\mu})^T \underset{\sim}{\Sigma}^{-1} (\underset{\sim}{\theta}_i - \underset{\sim}{\mu}) + \nu \quad .$$

Generation of all the required random variates is straightforward, thus providing - via the Gibbs sampler - a fully Bayesian implementation of the hierarchical linear model with Student-t second stage. By using the same device of scale mixtures of normals, the first stage of the hierarchy could also be taken to be Student-t, thus providing an analysis robust to both data outliers and outlying individuals in the population. More generally, the device used here leads to a straightforward Gibbs sampler implementation for any robustifying distribution defined as a scale mixture of normals (see Andrews and Mallows, 1974, Carlin and Polson, 1992).

4.3 Illustrative analysis

Using the model structure outlined in Section 4.2, the data in Table 1 was analysed with a number of second stage assumptions. The first analysis assumed that the (α, β)-pairs from each of the girl and boy populations arose from separate bivariate t-distributions with 2 degrees of freedom. We denote this model by St_2. For this example the values chosen for the hyper-parameters were

$$\nu_0 = 0, \ \underset{\sim}{C}^{-1} = \underset{\sim}{0}, \ \rho = 2, \ \underset{\sim}{R} = \begin{bmatrix} 1 & 0 \\ 0 & .1 \end{bmatrix} \quad .$$

The following Gibbs strategy was used: 25 cycles were run initially for 30 iterations before being increased to 50 cycles. These were run for 30 iterations also before being increased to 100 cycles for 100 iterations.

One of the principal aims of this illustration is the detection of outlying individuals using the St_ν model. The scale parameter λ_i is a good global indicator of outliers. The prior expectation of λ_i is 1, so that a λ_i value substantially below 1 indicates that the ith individual parameter vector (α_i, β_i) is likely to be far away from the population mean $\underset{\sim}{\mu}$. Since the Mahalanobis distance is effectively used here to measure the distance from $\underset{\sim}{\mu}$, λ_i provides only a global diagnostic for outliers. To investigate further the specific elements of $\underset{\sim}{\theta}_i$ for which the particular individual is outlying, one needs to examine moment summaries or graphical displays of α_i, β_i, marginally or jointly using the generated samples. Figure 2 displays the box plots of

the posterior sample values of the λ_i's, for each of the 16 boys. It is clear that boy 21 is likely to be an outlier. Boy number 24 is not, however, as extreme. Further investigation of the marginal posterior plots of α_i and β_i enables us to conclude that boy 21 is an intercept outlier, whereas boy 24 is a possible slope outlier.

Figure 2 Here

To compare the influence of the outliers on the overall inferences, the data set of the boys was reanalyzed using the normal model, first with the full data set (N0), then with boy 21 removed (N1) and then with both boys 21 and 24 removed (N2). The median, 5% and 95% posterior sample percentiles of α_B, β_B, and τ^{-1} (the population intercept, slope, and measurement variance), along with the sample mean of the population covariance matrix Σ are summarized in Table 3 for the various models for the boys, and for the St_2 model for the girls. The boy population parameter inferences shift in the expected directions with the various normal models; *e.g.*, the intercept in N0 is higher than that of N1 and N2. The population variance of both the intercept and slope are higher in the N0 model than in the other normal models. The results for the St_2 model lie, as expected, between N0 and N2.

Table 3 Here

The main objective in this problem is to make inferences concerning the difference in dental growth between boys and girls. For this comparison, the St_2 model was chosen for both groups. Let $\delta(t)$ be the difference in the dental measurement of boys and girls at age t, given by

$$\delta(t) = (\alpha_B + \beta_B(t-11)) - (\alpha_G + \beta_G(t-11)) \quad .$$

One can easily obtain posterior samples of $\delta(t)$ by direct substitution of the corresponding values of the generated samples of α_B, β_B and α_G, β_G. Figure 3 displays box plots of the differences at ages 8, 10, 12 and 14 years. It is quite clear that the boys have higher dental measurements and the differences

become increasingly larger as a function of age.

Figure 3 Here

We have demonstated here that the Gibbs sampler provides a straightforward means both for inference summaries and for diagnostic checks. Once the basic posterior samples are obtained for the original model parameters, the effort required to perform additional analyses is small.

5. ANALYSIS OF THE NONLINEAR PHARMACOKINETIC EXAMPLE

We begin in Section 5.1 by identifying the forms of the full conditionals for the power model specification (2.2) given in Section 2.2. In Section 5.2 we carry out an analysis of the data presented in Table 2 and discuss the two alternative model specifications of Sections 2.2.

5.1 A nonlinear population model

Suppose that the first-stage of the model has the form

$$\prod_{i=1}^{I}\prod_{j=1}^{n_i}[y_{ij}|\underset{\sim}{\theta}_i,\ \tau_i,\ \gamma] = \prod_{i=1}^{I}\prod_{j=1}^{n_i} N(y_{ij}|\eta_{ij}(\underset{\sim}{\theta}_i),\ \eta_{ij}(\underset{\sim}{\theta}_i)^{\gamma}\tau_i^{-1}) \ ,$$

where $\eta_{ij}(\underset{\sim}{\theta}_i)$ is a nonlinear function. Suppose further that the second and third-stage structures are as given in Section 4.1, with, additionally, a uniform prior for γ over some suitably assessed interval $0 \le \gamma \le c$, with c known. As in that section, the Gibbs sampler can again be defined by the full conditional distributions

$$[\underset{\sim}{\theta}_i|\underset{\sim}{y},\ \underset{\sim}{\mu},\ \underset{\sim}{\Sigma}^{-1},\ \underset{\sim}{\tau},\ \gamma,\ \underset{\sim}{\theta}_j,\ j\neq i] \qquad i = 1,\ldots,I$$

$$[\underset{\sim}{\mu}|\underset{\sim}{y},\ \underset{\sim}{\theta},\ \underset{\sim}{\Sigma}^{-1},\ \underset{\sim}{\tau},\ \gamma]$$

$$[\underset{\sim}{\Sigma}^{-1}|\underset{\sim}{y},\ \underset{\sim}{\theta},\ \underset{\sim}{\mu},\ \underset{\sim}{\tau},\ \gamma]$$

$$[\tau_i|\underset{\sim}{y},\ \underset{\sim}{\theta},\ \underset{\sim}{\mu},\ \underset{\sim}{\Sigma}^{-1},\ \gamma,\ \tau_j,\ j\neq i] \qquad i = 1,\ldots,I$$

$$[\gamma|\underset{\sim}{y},\ \underset{\sim}{\theta},\ \underset{\sim}{\mu},\ \underset{\sim}{\Sigma},\ \underset{\sim}{\tau}] \ .$$

The conditional distributions for $\underset{\sim}{\mu}$ and $\underset{\sim}{\Sigma}^{-1}$ are identical to those described in Section 4.1. The conditional distribution for τ_i is similar with

$$[\tau_i | \underset{\sim}{y}, \underset{\sim}{\theta}, \underset{\sim}{\mu}, \underset{\sim}{\Sigma}^{-1}, \gamma, \tau_j, j \neq i] = Ga\left[\tau_i \middle| \tfrac{1}{2}(\nu_0 + n_i), \tfrac{1}{2}\left[\nu_0\tau_0 + \sum_{j=1}^{n_i} \frac{(y_{ij} - \eta_{ij}(\underset{\sim}{\theta}_i))^2}{\eta_{ij}(\underset{\sim}{\theta}_i)^{\gamma}}\right]\right].$$

For each $i = 1, \ldots, I$, the conditional distribution of the p-vector $\underset{\sim}{\theta}_i$ is given by

$$[\underset{\sim}{\theta}_i | \underset{\sim}{y}, \underset{\sim}{\mu}, \underset{\sim}{\Sigma}^{-1}, \underset{\sim}{\tau}, \gamma, \underset{\sim}{\theta}_j, j \neq i] \propto$$

$$\prod_{j=1}^{n_i} \left[\frac{\tau_i}{\eta_{ij}(\underset{\sim}{\theta}_i)^{\gamma}}\right]^{\frac{1}{2}} \exp\left[-\frac{\tau_i}{2} \sum_{j=1}^{n_i} \frac{(y_{ij} - \eta_{ij}(\underset{\sim}{\theta}_i))^2}{\eta_{ij}(\underset{\sim}{\theta}_i)^{\gamma}}\right]$$

$$\times \exp\left[-\tfrac{1}{2}(\underset{\sim}{\theta}_i - \underset{\sim}{\mu})^T \underset{\sim}{\Sigma}^{-1}(\underset{\sim}{\theta}_i - \underset{\sim}{\mu})\right] .$$

Ignoring the final term gives the conditional form for γ on the range $0 \leq \gamma \leq c$. To generate from the conditional distributions of $\underset{\sim}{\theta}_i$ and γ we use the generalized ratio-of-uniforms technique as described in Wakefield, Gelfand and Smith (1991) and summarized here in an Appendix. For a range of examples, this method has been found to be considerably more reliable than the normal approximation rejection techniques used in a related setting by Zeger and Karim (1991).

5.2 Illustrative analysis

Using the power model structure outlined in Section 5.1 the data in Table 2 was analyzed using the following hyperameters

$$V_0 = 0, \ \underset{\sim}{C}^{-1} = \underset{\sim}{0}, \ \rho = 2 \ , \ \underset{\sim}{R} = \begin{bmatrix} .09 & 0 \\ 0 & .09 \end{bmatrix} .$$

$\underset{\sim}{R}$ was chosen in the following manner. We require an approximate estimate of

the variance-covariance matrix of the log α and log β population distribution. The off-diagonal element is chosen to be zero. Without loss of generality consider log α. In many pharmacokinetic applications we have some idea of the magnitude of the coefficient of variation of the α's. Now if the variance of the α's is small we have

$$\log \alpha \approx \log E(\alpha) + \frac{(\alpha - E(\alpha))}{E(\alpha)} \quad ,$$

and

$$\text{var } (\log \alpha) \approx \frac{\text{var}(\alpha)}{E(\alpha)^2} \quad .$$

Consequently the square of the coefficient of variation gives an estimate of the variance of log α. In this example the coefficient of variation for both α and β was estimated to be 30%.

The upper bound for γ, c was chosen to be 5 in this example. The Gibbs strategy was as follows. Good initial estimates were found and from these 10 cycles were run for 400 iterations. Samples were extracted from each cycle, 10 iterations apart from iteration 310 onwards to give 100 realizations from the posterior. The marginal distribution for γ was found to be located at approximately 0.6. As we noted in Section 2.2 one of the principal aims of studies of this kind is the computation of predictive distributions for concentrations. The model described in Section 5.1 was found to be inadequate in providing such predictive distributions. The value of γ was not large enough to ensure that the predictions avoided assigning significant probabilities to negative concentrations. The lognormal intra-individual error specification described in Section 2.2 would, of course, always produce positive predictive concentrations. However, the lognormal model also does not fit these data, as is clear from the following. The lognormal model $y = \eta(\underset{\sim}{\theta})e^{\varepsilon}$ with $\varepsilon \sim N(0, \sigma^2)$ can, for small σ^2, be approximated by

$$y \approx \eta(\underset{\sim}{\theta}) \ (1 + \varepsilon) \quad ,$$

yielding $E(Y) \approx \eta(\underset{\sim}{\theta})$ and $\text{Var}(\underset{\sim}{Y}) \approx \eta^2\sigma^2$. This corresponds to $\gamma = 2$ in our power model error specification . Consequently, if the lognormal error model were correct and the above approximation were accurate we would expect the marginal distribution for γ to be located close to the value 2. For error

variances and parameter values similar to those for this dataset, lognormal data were simulated and the marginal distribution for γ was indeed found to be close to 2 and not 0.6 as in this example. We conclude that the lognormal specification is not adequate for this example and so do not include here details of the Gibbs sampler conditional forms. We note, however, that the conditional form for $\underset{\sim}{\theta}_i$ is the only element of the sampler to change and generation via the ratio-of-uniforms method is again possible.

The truncated normal model outlined in Section 2.2 produces the following first stage

$$\prod_{i=1}^{I} \prod_{j=1}^{n_i} [y_{ij} | \underset{\sim}{\theta}_i, \tau_i, \gamma]$$

$$\propto \prod_{i=1}^{I} \prod_{j=1}^{n_i} \left[1 - \Phi \left[\frac{-\eta_{ij} \tau_i^{1/2}}{\eta_{ij}^{\gamma/2}} \right] \right]^{-1} \left[\frac{\tau_i}{\eta_{ij}^{\gamma}} \right]^{1/2}$$

$$\exp \left[-\frac{\tau_i}{2} \left[\frac{y_{ij} - \eta_{ij}}{\eta_{ij}^{\gamma}} \right]^2 \right] .$$

for $y_{ij} \geq 0$. Here $\Phi(.)$ denotes the cumulative distribution function of the standard normal distribution. In terms of the Gibbs sampler for the power model, specified in the previous Section, the conditional distributions for $\underset{\sim}{\theta}$, $\underset{\sim}{\tau}$ and γ are affected by this change. For $\underset{\sim}{\theta}_i$ and γ the ratio-of-uniforms can still be used though there is an additional computational expense in the numerical calculation of $\Phi(.)$. Previously the generation for the τ_i's was straightforward. This is no longer the case so, again, the ratio-of-uniforms technique (for log τ_i) was utilized and the analysis successfully implemented. Figure 4 shows a predictive distribution for concentration for patient 2 at 32 hours.

Figure 4 Here

For population inferences, we note the following interesting issue. Our second stage assumption is that the (α_i, β_i)-pairs are lognormally distributed

with mean $[\mu_1, \mu_2]$ and covariance elements $[\Sigma_{11}, \Sigma_{12}, \Sigma_{22}]$. To summarize the population distribution of, say, the α_i's we therefore have a number of options. We could choose to calculate the mean, the mode, the median, or more generally the quantiles of order q of the distribution, given, respectively, by $\exp(\mu_1 + \Sigma_{11}/2)$, $\exp(\mu_1 - \Sigma_{11})$, $\exp(\mu_1)$, and $\exp(\mu_1 + n_q \Sigma_{11}^{1/2})$, where n_q is the quantile of order q of an $N(0, 1)$ distribution. For a posterior sample from the marginal distribution of μ_1 and Σ_{11} we can easily generate any of these quantities. Recall that we may also be interested in the clearance Ω and the half-live λ, where for the ith individual $\Omega_i = \alpha_i\beta_i$ and $\lambda_i = \log 2/\beta_i$. It is straightforward to make inferences about these quantities since the (Ω_i, λ_i)-pairs also have a lognormal distribution with mean $(\mu_1 + \mu_2, \log(\log 2) - \mu_2)$ and covariance elements $(\Sigma_{11} + \Sigma_{22} + 2\Sigma_{12}, -\Sigma_{12} - \Sigma_{22}, \Sigma_{22})$. Figure 5 shows a bivariate plot of the medians for Ω and λ, with corresponding (histogram) density estimates.

Figure 5 Here

---

6. DISCUSSION AND RELATION TO OTHER WORK

Two alternative approaches for approximating posterior distributions for hierarchical models, such as those of Section 2, are the EM-type approximations (Racine-Poon, 1985; Racine-Poon and Smith, 1990), and the Laplace approximation method (see, for example, Tierney and Kadane, 1985; Kass and Steffey 1989).

The EM type approximation treats the individual level effects, $\underset{\sim}{\theta}_i$, as missing data and uses the EM algorithm to obtain the mode of the joint posterior distribution of the hyperparameters of θ_i (in our terminology, the population parameter $\underset{\sim}{\phi}$). The marginal posterior density for a population level parameter is approximated by the full conditional distribution for this parameter, with estimates from the EM algorithm replacing the conditioning parameters. Since the exact marginal posterior distribution is a continuous mixture of these full conditional forms, we might expect such an approximation to be poor (see Gelfand *et al*, 1990, for evidence that this is the case).

More specifically, let us consider the models in Sections 2.1 and 2.2 omitting the variance power transformation for simplicity. For the student-linear population model interest would often focus upon the posterior distribution of $\underset{\sim}{\mu}$. One attempt at using the EM approximation is to take as the estimates of the posterior distribution of $\underset{\sim}{\mu}$, the full conditional for $\underset{\sim}{\mu}$, substituting estimates of $\underset{\sim}{\theta}$, $\underset{\sim}{\Sigma}$ and $\underset{\sim}{\lambda}$ obtained by the EM algorithm. An E step is straightforward. The M step, however, requires a maximization over $\underset{\sim}{\mu}$, $\underset{\sim}{\Sigma}$ and $\underset{\sim}{\lambda}$ given $\underset{\sim}{\theta}$, and with more parameters than data this maximum is unbounded. If instead we 'integrate out the λ_i', again the E step is straightforward but now, without conjugacy, the M step cannot be achieved in closed form.

In the normal nonlinear model, taking $\hat{\underset{\sim}{\theta}}_i$ as say, the MLE or a nonlinear least squares estimator of $\underset{\sim}{\theta}_i$, the EM approximation replaces

$$\prod_{j=1}^{n_i} N(y_{ij} | \eta_{ij}(\underset{\sim}{\theta}_i),\ \tau_i^{-1}) \quad ,$$

with an approximate normal distribution for $\hat{\underset{\sim}{\theta}}_i$ based upon asymptotic theory. The remainder of the model specification is unchanged so that, with the resultant conjugacy, implementation of the EM algorithm is straightforward. However, since interesting applications tend to have small to moderate n_i the quality of the normal approximation is questionable.

With regard to the Laplace method approach, we first note that the joint posterior distribution of all the parameters is proportional to likelihood $\times$ prior. Hence any expectation, including any marginal distributions, can be expressed as a ratio of integrals. The Laplace method approximates both the numerator and denominator integrals. In particular, the version by Tierney and Kadane (1986) appears to offer the best such general approximation and has been tailored in Kass and Steffey (1989) for application to exchangeable hierarchical models.

When the model is conjugate, by which we mean that closed form integration over $\underset{\sim}{\theta}_i$ results, the Laplace approximation may be applied without difficulty. If not, as with the Student linear and normal nonlinear models, further techniques are needed. Kass and Steffey suggest that, if the dimensionality of $\underset{\sim}{\theta}$ is high, perhaps an approximate EM-type method might be

used! We note additionally that, in implementing the Laplace approximation, two function maximizations will always be required and at least one additional maximization will be required for each different expectation that is sought. Also, as demonstrated in Achcar and Smith (1990), the approximation is very sensitive to parameterization.

The Gibbs sampling approach is able to avoid many of the problems associated with the above alternative approximations, and appears to offer the most flexible and powerful method currently available for the routine analysis of challenging population model problems.

ACKNOWLEDGEMENTS

CIGA-GEIGY AG, Basle, Switzerland, provided financial support to the first author during part of this research. This paper formed part of the first annual Morris M DeGroot Memorial Lecture, given by the second author at Carnegie-Mellon University, September 20, 1991.

Table 1.

Measurements on 11 Girls and 16 Boys, at 4 Different Ages

	Girls - Age In Years			
Individual	**8**	**10**	**12**	**14**
1	21	20	21.5	23
2	21	21.5	24	25.5
3	20.5	24	24.5	26
4	23.5	24.5	25	26.5
5	21.5	23	22.5	23.5
6	20	21	21	22.5
7	21.5	22.5	23	25
8	23	23	23.5	24
9	20	21	22	21.5
10	16.5	19	19	19.5
11	24.5	25	28	28

	Boys - Age in Years			
Individual	**8**	**10**	**12**	**14**
12	26	25	29	31
13	21.5	22.5	23	26.5
14	23	22.5	24	27.5
15	25.5	27.5	26.5	27
16	20	23.5	22.5	26
17	24.5	25.5	27	28.5
18	22	22	24.5	26.5
19	24	21.5	24.5	25.5
20	23	20.5	31	26
21	27.5	28	31	31.5
22	23	23	23.5	25
23	21.5	23.5	24	28
24	17	24.5	26	29.5
25	22.5	25.5	25.5	26
26	23	24.5	26	30
27	22	21.5	23.5	25

Table 2.

Pharmacokinetic Data

	HOURS AFTER ADMINISTRATION							
PATIENT	2	4	6	8	10	24	28	32
1	1.09	0.75	0.53	0.34	0.23	0.02		
2	2.03	1.28	1.20	1.02	0.83	0.28		
3	1.44	1.30	0.95	0.68	0.52	0.06		
4	1.55	0.96	0.80	0.62	0.46	0.08		
5	1.35	0.78	0.50	0.33	0.18	0.02		
6	1.08	0.59	0.37	0.23	0.17	0.00		
7	1.32	0.74	0.46	0.28	0.27	0.03	0.02	0.00
8	1.63	1.01	0.73	0.55	0.41	0.01	0.06	0.02
9	1.26	0.73	0.40	0.30	0.21	0.00		
10	1.30	0.70	0.40	0.25	0.14	0.00		

Table 3.

Model	μ_1	μ_2	τ^{-1}	mean of Σ	
BOYS					
St_2	24.65	0.7781	2.65	1.7967	-0.0229
	(23.96, 25.94)	(0.5986, 0.9867)	(1.79, 3.87)		0.0742
NO	24.99	0.7978	2.90	2.5470	0.0063
	(24.30, 26.92)	(0.5952, 0.9931)	(1.86, 4.25)		0.0927
N1	24.67	0.7999	2.86	1.5290	-0.0021
	(23.84, 25.25)	(0.6161, 0.9799)	(1.92, 3.79)		0.0919
N2	24.66	0.7000	2.55	1.7041	0.0234
	(23.84, 25.28)	(0.4771, 0.9098)	(1.68, 3.56)		0.0719
GIRLS					
St_2	22.47	0.4322	0.49	3.1409	0.1070
	(21.39, 23.38)	(0.2175, 0.6085)	(0.28, 0.83)		0.0594

() denotes a 90% sample interval

APPENDIX The generalized ratio-of-uniforms method

Let $f(\theta)$ denote the unnormalized univariate density from which we wish to generate. The generalized ratio-of-uniforms method is as follows. If we generate bivariate points in the region C defined by

$$C = \left\{ (u, v) : 0 < u^{r+1} \leq f\left[\frac{v}{u^r}\right] \right\} , \qquad (A.1)$$

with $r > 0$, then the resulting 'ratio-of-uniforms' $\frac{v}{u^r}$ has distribution $f/\int f$. The efficiency of the method depends crucially upon the ease with which we can generate points within the region C. The strategy which has proved most successful is to contain C within a rectangle $R = [0, a] \times [b^-, b^+]$. We define a, b^- and b^+ shortly. It is shown in Wakefield, Gelfand and Smith (1991) that the probability of acceptance of a point generated in R is, in general, large if we generate instead from $\phi = \theta - \theta^*$, where θ^* is the mode of $f(\theta)$.

The aforementioned paper also recommends the use of $r = .5$. With this value and the 'mode-shift' we obtain the following strategy.

1 Determine $a = \left(\max_{\theta} f(\theta)\right)^{\frac{1}{r+1}}$,

2 Determine $b^- = \min_{\phi \leq 0} \phi[f(\phi + \theta^*)]^{r/r+1}$,

and $b^+ = \max_{\phi \geq 0} \phi[f(\phi + \theta^*)]^{r/r+1}$.

3 Generate $u \sim U(0, a)$ and $v \sim U(b^-, b^+)$. Let $\hat{\theta} = \frac{v}{u^r} + \theta^*$.

4 If $\hat{\theta}$ is not contained in the support of θ go to 3.

5 If $u^{r+1} \leq f(\hat{\theta})$ then accept $\hat{\theta}$, otherwise go to 3.

With this strategy, typical acceptance probabilities of around 0.8 have resulted for a range of models. As long as the maxima/minima defined in 1 and 2 exist the method can be applied. Apart from this restriction the method is completely general and does not, for example, need log-concavity of f, as per

the adaptive rejection sampling method described in Gilks and Wild (1992). The price of this generality is, of course, the need to carry out the maximisations/minimizations in order to find the bounding rectangle.

REFERENCES

ACHCAR, J. A., and SMITH, A. F. M. (1990). Aspects of reparametrization in approximate Bayesian inference. *Essays in Honour of George A. Barnard*, (J. Hodges, ed.). North-Holland, Amsterdam.

ANDREWS, D. F., and MALLOWS, C. L. (1974). Scale mixtures of normal distributions. *J. R. Statist. Soc., Series B*, **36**, 99-102.

BEAL, S. L., and SHEINER, L. B. (1980). The NONMEM system. *The American Statistician*, **34**, 118-119.

BERKEY, C. S. (1982). Bayesian approach for a nonlinear growth model. *Biometrics*, **38**, 953-961.

BESAG, J., YORK, J., and MOLLIÉ, A. (1991). Bayesian image restoration with two applications in spatial statistics. *Ann. Inst. Statist. Math.*, **43**, 1-59.

CARLIN, B. P. and POLSON, N. G. (1992). Nonconjugate Bayesian models using the Gibbs sampler. To appear in *Canadian J. of Statist.*

FEARN, T. (1975). A Bayesian approach to growth curves. *Biometrika*, **62**, 89-100.

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. Amer. Statist. Assoc.*, **85**, 972-985.

GELFAND, A. E., and SMITH, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *J. Amer. Statist. Assoc.*, **85**, 398-409.

GELFAND, A. E., and SMITH, A. F. M. (1991). Gibbs sampling for marginal posterior expectations. *Commun. Statist. Th. and Math.*, **20**, 1747-1766.

GEMAN, S., and GEMAN, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *I.E.E.E. Transactions on Pattern Analysis and Machine Intelligence*, 6, 721-741.

GILKS, W. R., and WILD, P. (1992). Adaptive rejection sampling for Gibbs sampling. To appear in *Applied Statistics*.

HASTINGS, W. K. (1970). Monte Carlo sampling methods using Markov Chain and their applications. *Biometrika*, 57, 97-109.

HILLS, S. E. and SMITH A. F.M. (1992). Parameterization issues in Bayesian inference. To appear in *Bayesian Statistics 4* (J. M. Bernardo, *et al*, eds.). Oxford University Press.

JOHNSON, N. L., and KOTZ, S. (1972). *Distributions in Statistics: Continuous Multivariate*. John Wiley and Son, New York.

KASS, R. E., and STEFFEY, D. (1989). Approximate Bayesian inference in conditionally independent hierarchical models (Parametric Empirical Bayes Models). *J. Amer. Statist. Assoc.*, 84, 717-726.

LANGE, N. ,CARLIN, B. P. and GELFAND, A. E. (1992). Hierarchical Bayes models for the progression of HIV infection using longitudinal CD4+ counts. To appear in *J. Amer. Statist. Assoc.*

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

LINDSTROM, M. J., and BATES, D. M. (1990). Nonlinear mixed effects models for repeated measures data. *Biometrics*, 46, 673-687.

NOVICK, M. R., JACKSON, P. H., THAYER, D. T., and COLE, N. S. (1972). Estimating multiple regression in *m*-groups; a cross-validation study. *Brit. J. Math. Statist. Psychology*, 25, 33-50.

ODELL, P. L., and FEIVESON, A. H. (1966). A numerical procedure to generate a sample covariance matrix. *J. Amer. Statist. Assoc.*, 61, 198-203.

O'HAGAN, A. (1987). Modelling with heavy tails. *Bayesian Statistics 3*, (J. M. Bernardo, *et al*, eds.). Oxford University Press.

POTTHOFF, R. F., and ROY, S. N. (1964). A generalized multivariate analysis of variance useful especially for growth curve problems. *Biometrika*, 51, 313-326.

RACINE, A., GRIEVE, A. P., FLÜHLER, H., and SMITH, A. F. M. (1986). Bayesian methods in practice: experiences in the pharmaceutical industry (with discussion). *Applied Statistics*, 35, 93-150.

RACINE-POON, A. (1985). A Bayesian approach to non-linear random effects models. *Biometrics*, 41, 1015-1024.

RACINE-POON, A. (1992). SAGA: Sample Assisted Graphical Analysis. To appear in *Bayesian Statistics 4* (J. M. Bernardo, *et al*, eds.). Oxford University Press.

RACINE-POON, A., and SMITH, A. F. M. (1990). Population models. *Statistical Methodology in the Pharmaceutical Sciences* (D. Berry, e.d). Marcel Dekker, New York.

RAFTERY, A. E., and LEWIS, S. (1992). How many iterations in the Gibbs sampler? To appear in *Bayesian Statistics 4* (J. M. Bernardo, *et al*, eds.). Oxford University Press.

RIPLEY, B. D. (1987). Stochastic simulation. John Wiley and Son, New York.

SKENE, A. M., and WAKEFIELD, J. C. (1990). Hierarchical models for multi-centre binary response studies. *Statistics in Medicine*, 9, 919-929.

SMITH, A. F. M. (1983). Bayesian approaches to outliers and robustness. *Specifying Statistical Models* (J. P. Florens, *et al*, eds.), 13-35. Springer-Verlag, New York.

SWAMY, P. A. V. B. (1970). Efficient inference in a random coefficient regression model. *Econometrica*, 38, 311-323.

TIERNEY, L., and KADANE, J. (1986). Accurate approximations for posterior moments and marginal densities. *J. Amer. Statist. Assoc.*, **81**, 82-86.

WAKEFIELD, J. C., GELFAND, A. E., and SMITH, A. F. M. (1991). Efficient generation of random variates via the ratio-of-uniforms method. To appear in *Statistics and Computing*.

WAKEFIELD, J. C. (1992). Comment to: Parameterization issues in Bayesian inference, by A. F. M. Smith and S. E. Hills. To appear in *Bayesian Statistics 4* (J. M. Bernardo, *et al*, eds.). Oxford University Press.

ZEGER, S. L., and KARIM, M. R. (1991). Generalized linear models with random effects; a Gibbs sampling approach. *J. Amer. Statist. Assoc.*, **86**, 79-86.

Figure 1: least squares estimates of intercepts and slopes, plotting symbol is individual number

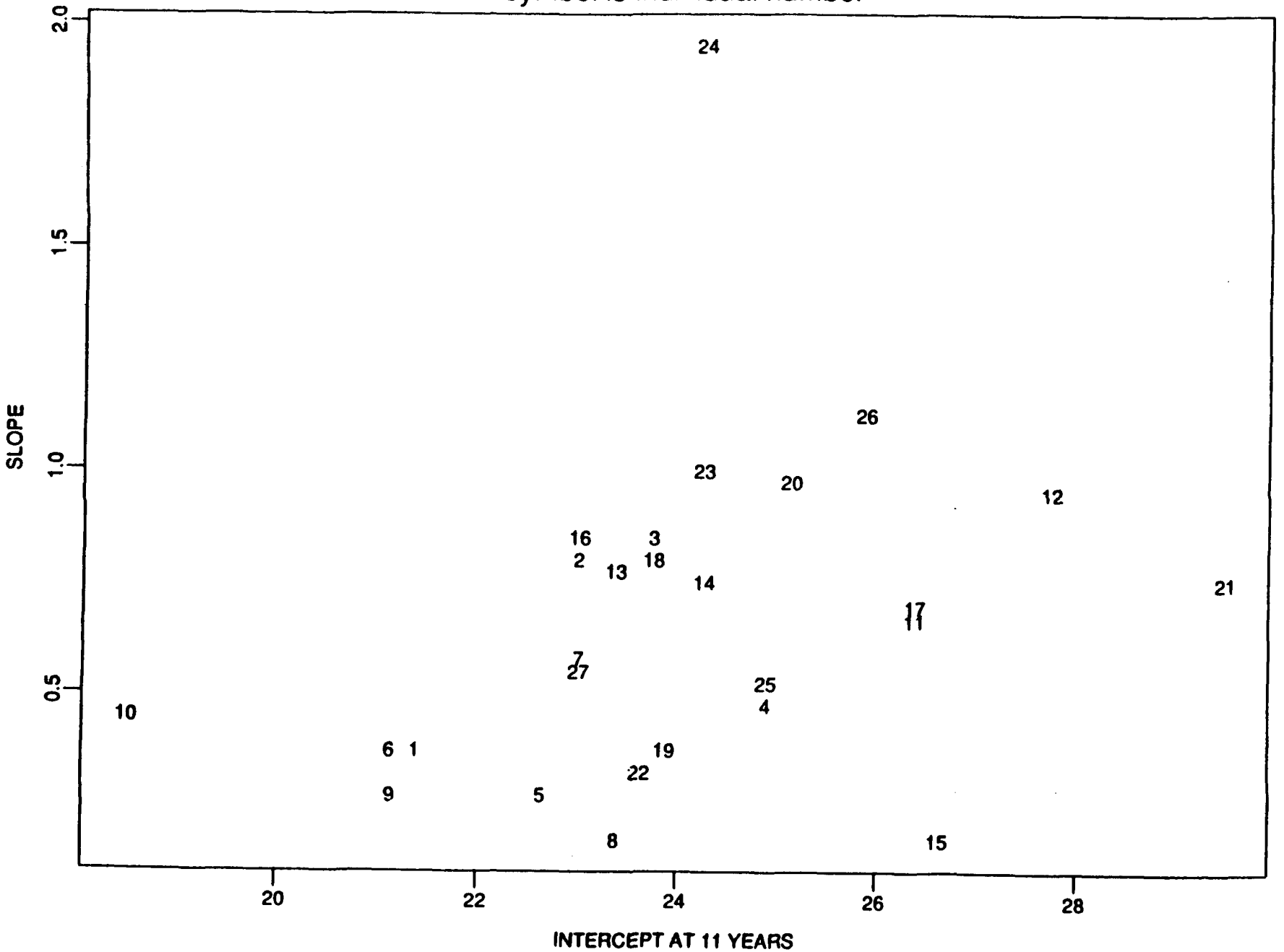


Figure 2: boxplots of lambda's for the boys. Boxplots are drawn containing 5%, 10%, 50%, 90% and 95% of samples

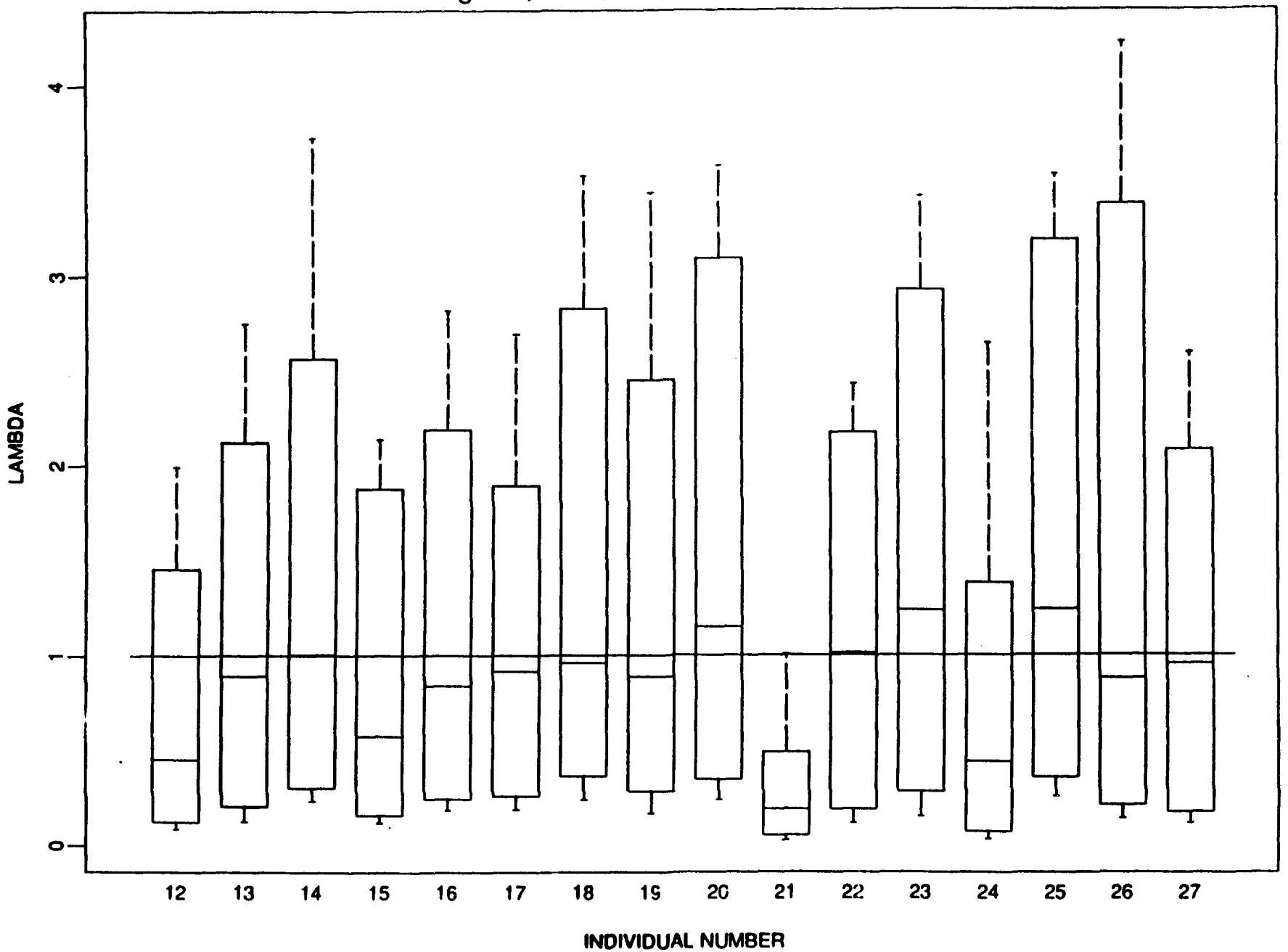


Figure 3: Difference between growth of boys and girls versus age

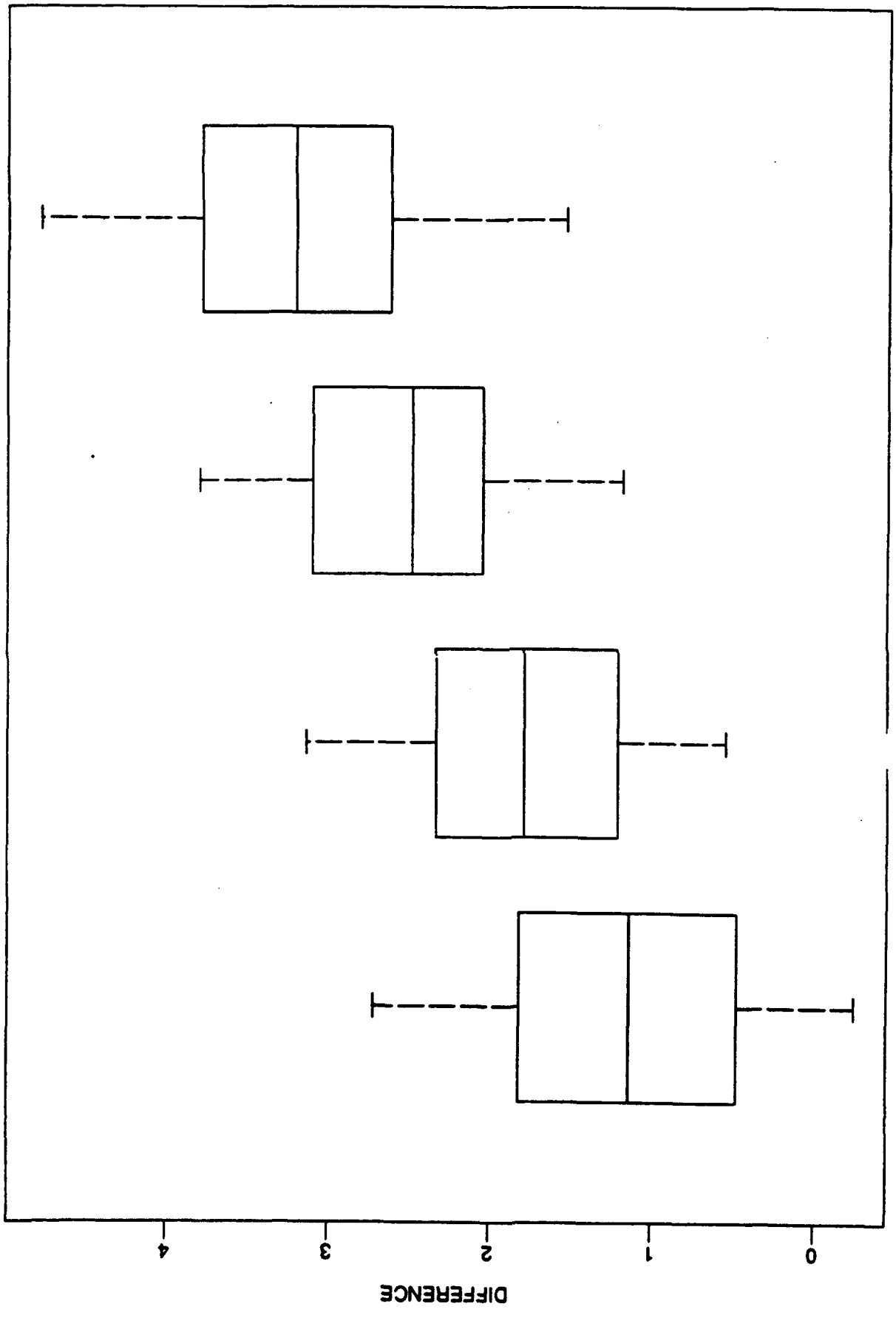
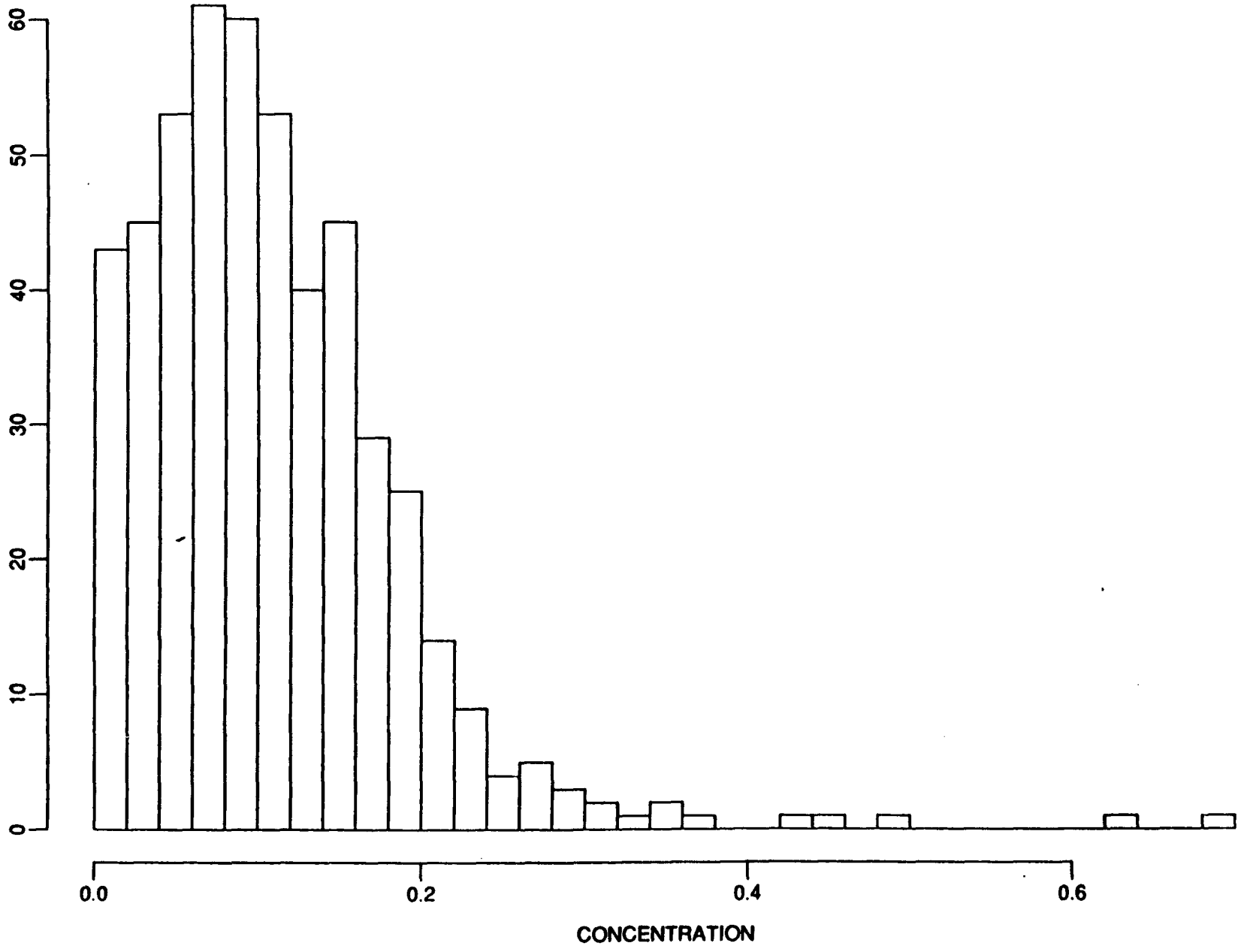


Figure 4: predictive distribution for concentration for individual 2 at 32 hours



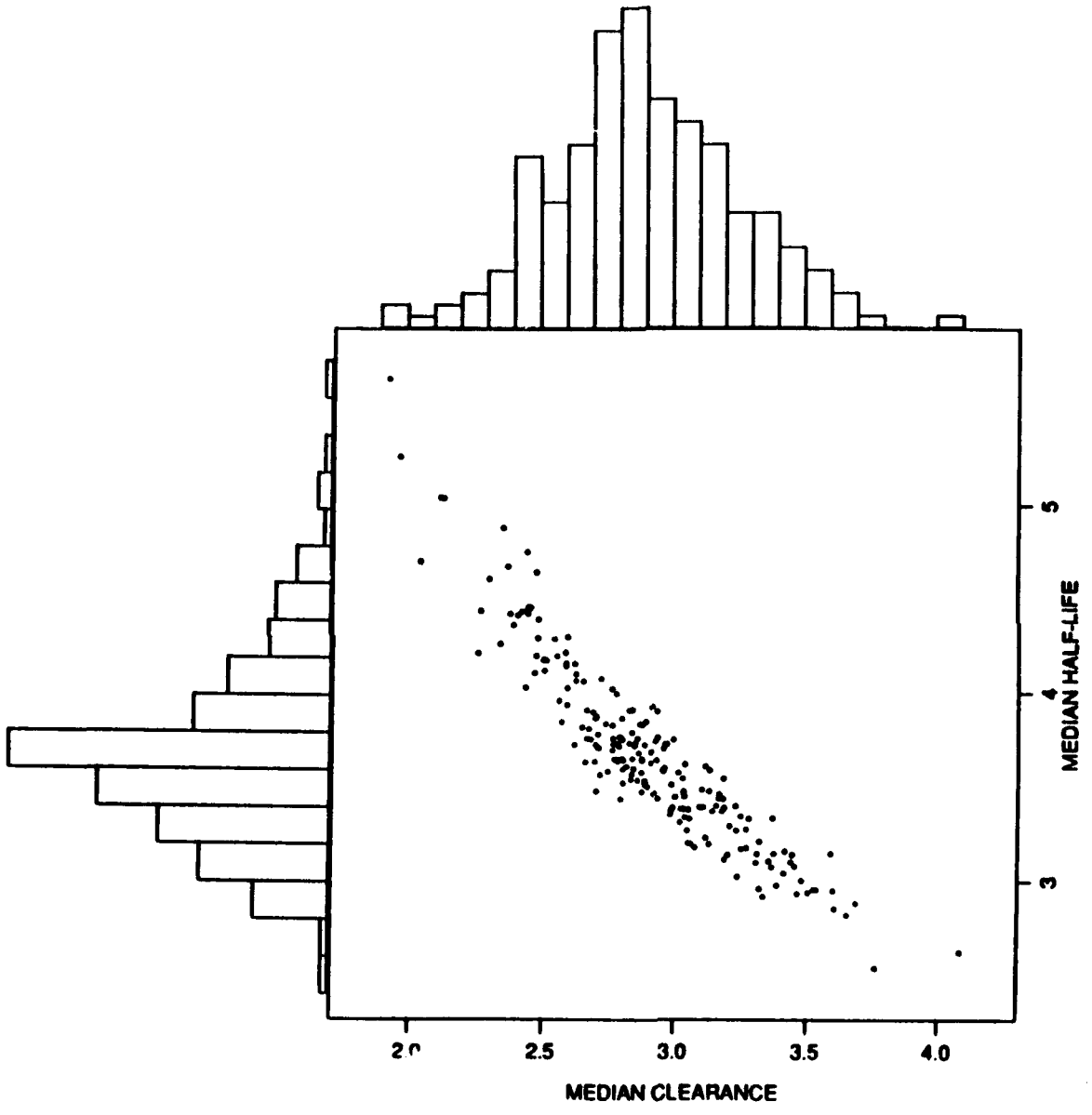


Figure 5: sample from the bivariate posterior of the median of the clearance and the median of the half-life

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 456	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Bayesian Analysis of Linear and Nonlinear Population Models Using the Gibbs Sampler		5. TYPE OF REPORT & PERIOD COVERED Technical Report 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) J.C. Wakefield, A.F.M. Smith, A. Racine-Poon, A.E. Gelfand		8. CONTRACT OR GRANT NUMBER(s) N0025-92-J-1264
9. PERFORMING ORGANIZATION NAME AND ADDRESS Office of Naval Research Statistics & Probability Program Code 111		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR-042-267
11. CONTROLLING OFFICE NAME AND ADDRESS		12. REPORT DATE July 21, 1992 13. NUMBER OF PAGES 40
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19.

Keywords: Population models; hierarchical models; Bayesian analysis; Gibbs sampler; population outliers; nonlinear models; mean-variance relationships.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

See Reverse Side

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

SUMMARY

A fully Bayesian analysis of linear and nonlinear population models has previously been unavailable, as a consequence of the seeming impossibility of performing the necessary numerical integrations in the complex multi-parameter structures typically arising in such models. It is demonstrated that, for a variety of linear and nonlinear population models, a fully Bayesian analysis can be implemented in a straightforward manner using the Gibbs sampler. The approach is illustrated with examples involving challenging problems of outliers and mean-variance relationships in population modelling.