Gibbs Sampling with Improper Prior Distributions

BU-1221-M December, 1993

James P. Hobert[1] and George Casella[2*]

Biometrics Unit

Cornell University

[1] Research supported by National Institute of Environmental Health Sciences Training Grant EHS-5-T32-ES07261-03 and National Science Foundation Grants DMS-9305547 and INT-9216784.

[2] Research supported by National Science Foundation Grant No. DMS-9305547

* James P. Hobert is a graduate student of statistics, Biometrics Unit, Cornell University. George Casella is Professor of Statistics, Biometrics Unit, Cornell University, Ithaca, NY 14853.

Key Words: Bayesian Hierarchical Models, Full Conditional Specification, Mixed Models, Variance Components, Exponential Family.

Abstract

It is not unusual, when considering conditionally independent hierarchical models (CIHM's) within a Bayesian framework, to assign improper prior distributions to the parameters of the second stage. Unfortunately, the technical difficulties which lead to the use of Markov chain Monte Carlo techniques (e.g. the Gibbs sampler) for sampling from the resulting posterior distributions also cause problems in establishing the integrability of these posterior distributions. In this work, we give conditions which guarantee the integrability of the resulting posterior distributions for some frequently used CIHM's. These CIHM's include the linear mixed model and those in which the first and second stage distributions can be written as a one-parameter exponential family and its two-parameter conjugate exponential family, respectively.

1. Introduction

Bayesian hierarchical models often lead to situations in which the posterior distribution is impossible to deal with analytically while the conditional distributions required for Gibbs sampling are of a simple form. A lack of prior information about the hyperparameters in these models often leads to the use of improper prior distributions. When using improper priors it is, of course, necessary to check that the posterior distribution is proper, i.e. that the marginal distribution of the data is not infinite. This is not always easily done, however, when using complicated hierarchical models. If one uses improper priors within a Bayesian hierarchical model and finds that the conditional distributions required for Gibbs sampling are of a simple form, it might be tempting to simply assume that the posterior is a proper distribution and implement the Gibbs sampler. Unfortunately, existence of the "Gibbs conditionals" in no way guarantees that the posterior distribution is proper. This is a dangerous situation since the Gibbs sampler can sometimes be employed when the posterior distribution is improper. This could clearly lead to seriously misleading conclusions.

For example, consider the common beta-binomial hierarchy

$$Y|p \sim Binomial(n,p) \qquad (1.1)$$
$$p \sim Beta(a,b)$$

where the posterior distribution from (1.1) is $\pi(p|Y,a,b)$, a beta density. With the added computational power of the Gibbs sampler, it is now possible to specify a prior distribution on the beta parameters, and fully marginalize. That is, if we add a prior $g(a,b)$ to (1.1), Gibbs sampling can allow us to calculate the posterior distribution by iteratively resampling from the three conditional posteriors $\pi(p|Y,a,b)$, $\pi(a|Y,b,p)$ and $\pi(b|Y,a,p)$. The full posterior can then seemingly be approximated (to any degree of accuracy) by

$$\pi(p|y) \approx \frac{1}{M}\sum_{i=1}^{M}\pi(p|y,a_i,b_i) \tag{1.2}$$

where M denotes the number of Gibbs iterations.

The problem arises when, as is often the case in practice, the prior $g(a,b)$ is improper, i.e. $\int g(a,b)\,da\,db=\infty$. It also often happens that, in such cases, the three conditional posteriors will all be proper densities, and can be used to generate random variables. Proceeding along, (1.2) can then be used to approximate $\pi(p|y)$ but, as we shall see, $\pi(p|y)$ may not exist. Thus, we could be in a case where the right-hand side of (1.2) gives a beautiful picture, but the picture is phony: there is nothing for it to approximate. This example is fully developed in Section 2.

The reason for this problem is that a set of conditional densities is not sufficient for determining a joint density. For random variables $X_1,\ldots,X_n$, consider the set of densities

$$f(x_i|x_{-i}) \quad i=1,2,\ldots,n \tag{1.3}$$

where x_{-i} represents the set $\{x_1,x_2,\ldots,x_{i-1},x_{i+1},\ldots,x_n\}$ and we use the notation $f(\cdot|\cdot)$ to stand for a generic density. Gelman and Speed (1993) call this set the "full conditional specification" and give conditions under which it uniquely determines a joint distribution $f(x_1,\ldots,x_n)$. This is the underlying workings of the Gibbs sampler, but existence of a full conditional specification (FCS) does not guarantee the existence of a unique joint distribution (see Example 2 in Casella and George (1992)). It is therefore quite important, when using improper prior distributions, to make sure that the posterior distribution is proper before using a Markov chain Monte Carlo (MCMC) technique such as the Gibbs sampler. In this paper we consider two fairly general models and give some conditions (on the prior distributions) under which the posteriors are guaranteed to be proper. This then allows the implementation of the Gibbs sampler with the assurance that it will converge to a legitimate posterior distribution.

An insidious feature of this problem is that it may be undetectable to a practicioner, as the right-hand side of (1.2) will often look pretty. Moreover, even monitoring raw histograms will not always reveal an underlying problem. A reason for this difficulty is that, if the posterior distribution is improper, we have a null Markov chain (Meyn and Tweedie 1993), which can be quite difficult to detect. In fact, this problem was brought to our attention by a colleague in Animal Science, who had been using a Gibbs sampler for almost two years until it "blew up" one day.

Thus, at this point, the only reliable mechanism for detecting improper posteriors is to eliminate them from consideration. In Section 2 we consider a hierarchical exponential family model, and show which improper priors will result in proper posteriors. We also look at a number of detailed examples. Section 3 considers the general linear mixed model, where in some cases we can give necessary and sufficient conditions for proper posteriors. Section 4 treats the one-way random effects model in detail, and illustrates the output from Gibbs chains with improper posterior distributions. Lastly, Section 5 is a short discussion.

2. Hierarchical Exponential Family Models

The first set of models we discuss are those in which the data are distributed according to a one-parameter exponential family, whose parameter is distributed according to the conjugate two-parameter exponential family (George, Makov and Smith 1992) and the hyperparameters are assigned improper priors. The model is

$$Y_{ij}|\theta_i \sim f(y_{ij}|\theta_i) = a(y_{ij})\exp(\theta_i y_{ij} - \psi(\theta_i)) \quad i=1,2,\ldots,K \ \ and \ \ j=1,2,\ldots,J$$

$$\theta_i|\alpha,\beta \sim f(\theta_i|\alpha,\beta) = \exp\{\beta(\alpha\theta_i - \psi(\theta_i)) - \phi(\alpha,\beta)\} \tag{2.1}$$

$$\alpha,\beta \sim \pi(\alpha,\beta)$$

where $\psi(\theta) = \ln \int a(y) e^{y\theta} dy$ and $\phi(\alpha,\beta) = \ln \int \exp\{\beta(\alpha\theta - \psi(\theta))\} d\theta$. We assume that all distributions in the first two stages of the above model are proper and that $\int \pi(\alpha,\beta)\, d\alpha\, d\beta = \infty$. Conditionally on the θ's, the Y's are assumed to be independent. In the same manner, the θ's are assumed independent given (α,β). The Y's are also assumed conditionally independent of (α,β) given the θ's.

The joint posterior distribution of $\theta = (\theta_1,\ldots,\theta_K)$, α and β is given by

$$\pi(\theta,\alpha,\beta|\mathbf{y}) = \frac{1}{m(\mathbf{y})} \prod_i \left(\prod_j f(y_{ij}|\theta_i) \right) f(\theta_i|\alpha,\beta)\pi(\alpha,\beta)$$

where $\mathbf{y} = (y_{11},\ldots,y_{JK})$ and

$$m(\mathbf{y}) = \int \cdots \int \left[\prod_i \left(\prod_j f(y_{ij}|\theta_i) \right) f(\theta_i|\alpha,\beta) \right] \pi(\alpha,\beta)\, d\theta_1 \cdots d\theta_K\, d\beta\, d\alpha. \qquad (2.2)$$

The joint posterior distribution, $\pi(\theta,\alpha,\beta|\mathbf{y})$ is a proper distribution if and only if m(**y**) is finite. Given the exponential family structure of our model, we can evaluate part of the integral in (2.2) by noting that

$$\int \left(\prod_j f(y_{ij}|\theta_i) \right) f(\theta_i|\alpha,\beta)\, d\theta_i = \left\{ \prod_j a(y_{ij}) \right\} \exp\left\{ \phi\left(\frac{y_{i.} + \alpha\beta}{\beta + J}, \beta + J \right) - \phi(\alpha,\beta) \right\} \qquad (2.3)$$

where $y_{i.} = \sum_j y_{ij}$. Substituting (2.3) into (2.2) gives the following proposition

Proposition 1. Given model (2.1), the posterior distribution, $\pi(\theta,\alpha,\beta|\mathbf{y})$, will be proper iff

$$\iint \exp\left\{\sum_i \phi\left(\frac{y_{i.}+\alpha\beta}{\beta+J},\beta+J\right)-K\phi(\alpha,\beta)\right\}\pi(\alpha,\beta)\,d\beta\,d\alpha < \infty \tag{2.4}$$

where $y_{i.}=\sum_j y_{ij}$.

We now discuss three special cases of model (2.1). All three models possess legitimate FCS's and therefore lend themselves to Gibbs sampling. On the other hand, using Proposition 1, we show that all three of these models can lead to improper posterior distributions.

Example 1 Consider the following Bayesian hierarchical model which could, for instance, be used to analyze the famous "pump failure" data (Gaver and O'Muircheartaigh 1987).

$$\begin{aligned} Y_{ij}\mid\lambda_i &\sim Poisson(\lambda_i) \quad i=1,2,\ldots,K \quad j=1,2,\ldots,J \\ \lambda_i\mid r,s &\sim Gamma(r,s) \\ dF(r,s) &= s^a\,ds\,dr \quad (a>-1) \end{aligned} \tag{2.5}$$

We say $X \sim Gamma(r,s)$ if $f_X(t) \propto t^{r-1}\exp(-st)$. The FCS for this model is

$$\lambda_i\mid r,s,\mathbf{y} \sim Gamma(y_{i.}+r, J+s) \quad i=1,2,\ldots,K$$

$$f(r\mid\mathbf{y},\boldsymbol{\lambda},s) \propto s^{Kr}(\Gamma(r))^{-K}\left(\prod_i \lambda_i\right)^{r-1}$$

$$f(s\mid\mathbf{y},\boldsymbol{\lambda},r) \propto s^{Kr+a}\exp\left\{-s\sum_i \lambda_i\right\}.$$

George, Makov and Smith (1992) show that $f(r\mid\mathbf{y},\boldsymbol{\lambda},s)$ is an integrable function for $r\in\Re^+$ while $f(s\mid\mathbf{y},\boldsymbol{\lambda},r)$ is clearly integrable for $s\in\Re^+$ when $a>-1$. The FCS therefore consists of legitimate densities. We now use Proposition 1 to show that the posterior distribution, $\pi(\lambda, r, s|\mathbf{y})$, is improper. Substituting into (2.4) we have

$$\iint \exp\left\{\sum_i \phi\left(\frac{y_{i\cdot}+\alpha\beta}{\beta+J}, \beta+J\right) - K\phi(\alpha,\beta)\right\}\pi(\alpha,\beta)\, d\beta\, d\alpha$$

$$= \int_0^\infty \frac{\prod \Gamma(y_{i\cdot}+r)}{\Gamma^K(r)} \left\{\int_0^\infty \frac{s^{Kr+a}}{(s+J)^{y_{\cdot\cdot}+Kr}} ds\right\} dr. \tag{2.6}$$

If $y_{\cdot\cdot} \le a+1$, then the inside integral diverges. If $y_{\cdot\cdot} > a+1$ we have that (2.6) is equal to

$$= c\int_0^\infty \frac{\prod \Gamma(y_{i\cdot}+r)}{\Gamma^K(r)} \frac{\Gamma(Kr+a+1)}{\Gamma(y_{\cdot\cdot}+Kr)} dr. \tag{2.7}$$

If $a \ge 1$, then it is clear that

$$c\int_0^\infty \frac{\prod \Gamma(y_{i\cdot}+r)}{\Gamma^K(r)} \frac{\Gamma(Kr+a+1)}{\Gamma(y_{\cdot\cdot}+Kr)} dr \ge c\int_0^\infty \left(\frac{\Gamma(Kr+\underline{a}+1)}{\Gamma(y_{\cdot\cdot}+Kr)}\right) \prod_{i=1}^K \left\{\frac{\Gamma(y_{i\cdot}+r)}{\Gamma(r)}\right\} dr \tag{2.8}$$

where $\underline{a}$ denotes rounding a down to the nearest integer. Evaluating the gamma functions

$$c\int_0^\infty \left(\frac{\Gamma(Kr+\underline{a}+1)}{\Gamma(y_{\cdot\cdot}+Kr)}\right) \prod_{i=1}^K \left\{\frac{\Gamma(y_{i\cdot}+r)}{\Gamma(r)}\right\} dr \ge c\int_0^\infty \frac{r^{y_{\cdot\cdot}}}{(y_{\cdot\cdot}+Kr-1)^{y_{\cdot\cdot}-\underline{a}-1}} dr \tag{2.9}$$

But the integral on the right-hand side diverges as $\underline{a} \ge 1$. Finally, if $a \in (-1,1)$,

$$c\int_0^\infty \frac{\prod \Gamma(y_{i\cdot}+r)}{\Gamma^K(r)} \frac{\Gamma(Kr+a+1)}{\Gamma(y_{\cdot\cdot}+Kr)} dr \ge c\int_2^\infty \left(\frac{\Gamma(Kr+a+1)}{\Gamma(y_{\cdot\cdot}+Kr)}\right) \prod_{i=1}^K \left\{\frac{\Gamma(y_{i\cdot}+r)}{\Gamma(r)}\right\} dr$$

since the integrand is positive. For $r \in (2,\infty)$, we may replace a with -1 to get another lower bound

$$c\int_2^{\infty}\left(\frac{\Gamma(Kr+a+1)}{\Gamma(y_{\cdot}+Kr)}\right)\prod_{i=1}^{K}\left\{\frac{\Gamma(y_{i\cdot}+r)}{\Gamma(r)}\right\}dr \geq c\int_2^{\infty}\left(\frac{r}{(y_{\cdot}+Kr-1)}\right)^{y_{\cdot}}dr. \tag{2.10}$$

The integral on the right-hand side is divergent. Thus, by Proposition 1, the posterior distribution, $\pi(\lambda, r, s|\mathbf{y})$, is not proper for the hierarchical model (2.5). This example demonstrates a situation in which a person could use a perfectly reasonable FCS to construct a Gibbs chain in which the conditionals can be sampled but which <u>cannot</u> converge to a proper posterior distribution.

For the sake of concreteness, suppose K=J=2 and $a = 0$. The FCS in this case is

$$\lambda_1|r,s,\mathbf{y} \sim Gamma(y_{1\cdot}+r, s+2) \qquad \lambda_2|r,s,\mathbf{y} \sim Gamma(y_{2\cdot}+r, s+2)$$

$$f(r|\mathbf{y},\lambda,s) \propto s^{2r}(\Gamma(r))^{-2}(\lambda_1\lambda_2)^{r-1} \qquad f(s|\mathbf{y},\lambda,r) \propto s^{2r}\exp\{-s(\lambda_1+\lambda_2)\}.$$

Constructing a Gibbs chain in this situation would be quite straightforward. (Since, the posterior corresponding to this FCS is improper, a Gibbs chain generated using it is formally not a Gibbs chain. We henceforth use the term *pseudo-Gibbs* when referring to such chains.) The Metropolis algorithm (Smith and Roberts 1993) could be used to simulate from the conditionals corresponding to r and s. We believe this to be a dangerous situation since it is possible to construct a pseudo-Gibbs chain and then use the results of this chain to make inferences about a nonexistent posterior distribution.

<u>Example 2</u> Consider another special case of model (2.1), a full hierarchical specification for the beta-binomial model given in the introduction:

$$Y_{ij}|p_i \sim Binomial(n,p_i) \;\; i=1,2,\ldots,K \;\; j=1,2,\ldots,J$$

$$p_i|r,s \sim Beta(r,s) \tag{2.11}$$

$$dF(r,s) = ds\,dr.$$

The FCS for this hierarchical model is

$$p_i|r,s,y \sim Beta(y_{i\cdot}+r, Jn - y_{i\cdot}+s) \;\; i=1,2,\ldots,K$$

$$f(r|\mathbf{y},p,s) \propto \frac{\Gamma^K(r+s)}{\Gamma^K(r)}\left(\prod_i p_i\right)^{r-1}$$

$$f(s|\mathbf{y},p,r) \propto \frac{\Gamma^K(r+s)}{\Gamma^K(s)}\left(\prod_i (1-p_i)\right)^{s-1}.$$

George, Makov and Smith (1992) show that $f(r|\mathbf{y},p,s)$ and $f(s|\mathbf{y},p,r)$ are integrable functions whenever $K \geq 2$. We now use Proposition 1, as before, to show that the posterior distribution, $\pi(p,r,s|\mathbf{y})$, is improper. (We were unable to use the results of George, Makov and Smith (1993) to show that the FCS is legitimate for a more general prior like $dF(r,s) = s^a r^b ds\,dr$.) Substituting into (2.4) we have

$$\iint \exp\left\{\sum_i \phi\left(\frac{y_{i\cdot}+\alpha\beta}{\beta+J}, \beta+J\right) - K\phi(\alpha,\beta)\right\}\pi(\alpha,\beta)\,d\beta\,d\alpha$$

$$= \int_0^\infty\int_0^\infty \prod_i \frac{\Gamma(y_{i\cdot}+r)\Gamma(Jn-y_{i\cdot}+s)\Gamma(r+s)}{\Gamma(r)\Gamma(s)\Gamma(Jn+r+s)}\,ds\,dr.$$

Using the fact that

$$\frac{\Gamma(y_{i\cdot}+r)\Gamma(Jn-y_{i\cdot}+s)\Gamma(r+s)}{\Gamma(r)\Gamma(s)\Gamma(Jn+r+s)}$$

$$=\frac{(y_{i\cdot}+r-1)(y_{i\cdot}+r-2)\cdots r(Jn-y_{i\cdot}+s-1)(Jn-y_{i\cdot}+s-2)\cdots s}{(Jn+r+s-1)(Jn+r+s-2)\cdots(r+s)}.$$

we have

$$\int_0^\infty\int_0^\infty \prod_i \frac{\Gamma(y_{i\cdot}+r)\Gamma(Jn-y_{i\cdot}+s)\Gamma(r+s)}{\Gamma(r)\Gamma(s)\Gamma(Jn+r+s)}ds\,dr$$

$$\geq \int_0^\infty\int_0^\infty \prod_i \left\{\frac{r^{y_{i\cdot}}s^{Jn-y_{i\cdot}}}{(Jn-1+r+s)^{Jn}}\right\}ds\,dr = \int_0^\infty r^{\sum y_{i\cdot}}\left\{\int_0^\infty \frac{s^{KJn-\sum y_{i\cdot}}}{(Jn-1+r+s)^{KJn}}ds\right\}dr. \qquad (2.12)$$

The inside integral in (2.12) will be infinite if $\sum y_{i\cdot} \leq 1$. If $\sum y_{i\cdot} > 1$, we have

$$\int_0^\infty r^{\sum y_{i\cdot}}\left\{\int_0^\infty \frac{s^{KJn-\sum y_{i\cdot}}}{(Jn-1+r+s)^{KJn}}ds\right\}dr = c\int_0^\infty \frac{r^{\sum y_{i\cdot}}}{(Jn-1+r)^{\sum y_{i\cdot}-1}}dr$$

where c represents a constant which is independent of r. It follows that

$$\int_0^\infty \frac{r^{\sum y_{i\cdot}}}{(Jn-1+r)^{\sum y_{i\cdot}-1}}dr \geq \int_{Jn-1}^\infty \left\{\frac{r^{\sum y_{i\cdot}}}{(r+r)^{\sum y_{i\cdot}-1}}\right\}dr = 2^{-\left(\sum y_{i\cdot}-1\right)}\int_{Jn-1}^\infty r\,dr = \infty.$$

Therefore, the posterior distribution, $\pi(p,r,s|\mathbf{y})$, for the hierarchical model (2.11) is improper.

Example 3 The last special case of model (2.1) we consider is the normal model.

$$
\begin{aligned}
Y_{ij} \mid \xi_i &\sim Normal(\xi_i, 1) \quad i = 1, 2, \ldots, K \quad j = 1, 2, \ldots, J \\
\xi_i \mid \mu, \sigma^2 &\sim Normal(\mu, \sigma^2) \\
dF(\mu, \sigma^2) &= \left(\sigma^2\right)^{-(a+1)} d\mu \, d\sigma^2.
\end{aligned}
\tag{2.13}
$$

The posterior, $\pi(\xi, \mu, \sigma^2 | \mathbf{y})$, for model (2.13) exists iff $-K + 1 < 2a < 0$ (Berger and Robert 1990), while it is possible to sample from the FCS (and construct a pseudo-Gibbs chain) whenever $-K < 2a$. (If $-K < 2a$, σ^2 given ξ, μ and the data has a proper inverted gamma distribution.) Berger and Robert (1990) also show that, in this case, a proper posterior distribution implies existence of the posterior expectation and covariance matrix of the ξ's. Model (2.13) is a special case of the models in Section 3.

3. Hierarchical Linear Mixed Models

The second set of models we discuss are Bayesian hierarchical versions of the linear mixed models (Searle, Casella and McCulloch 1992). The main interest in this section, as in the previous section, is to classify those improper priors that lead to legitimate FCS's into those for which the posterior distribution is proper and those for which it is not. We use a parametric improper prior distribution for the variance components which yields many standard improper forms, such as those discussed in Hill (1963) and Tiao and Tan (1965), as well as many not so standard forms, such as flat priors, as special cases. A nice property of this parametric improper prior is that its use leads to a manageable FCS which is identical in form to that given in Gelfand and Smith (1990) who describe the Gibbs sampler for the one-way random effects model using proper priors at all stages.

We define the model equation to be

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{u} + \varepsilon \tag{3.1}$$

where $\mathbf{Y}$ is an N x 1 vector of data, β is a p x 1 vector of fixed effects (parameters), $\mathbf{u}$ is a q x 1 vector of random effects (random variables), $\mathbf{X}$ and $\mathbf{Z}$ are known N x p and N x q design matrices, receptively, and ε is an N x 1 vector of residual errors. We assume that $\mathbf{u}$ can be partitioned into a series of r sub-vectors

$$\mathbf{u} = \left[\mathbf{u}_1' \, \mathbf{u}_2' \cdots \mathbf{u}_r'\right]'.$$

We also assume that $(\mathbf{Z}'\mathbf{Z})^{-1}$ and $(\mathbf{X}'\mathbf{X})^{-1}$ exist.

A typical Bayesian hierarchy used for mixed models such as this begins with the assumptions

$$(i) \;\; \mathbf{u} | \sigma_1^2, \ldots, \sigma_r^2 \sim N(\mathbf{0}, \mathbf{D})$$
$$(ii) \;\; \pi(\beta) = 1$$
$$(iii) \;\; \varepsilon | \sigma_\varepsilon^2 \sim N(\mathbf{0}, \mathbf{I}\sigma_\varepsilon^2)$$

where

$$\mathbf{D} = diag\left\{\mathbf{I}_{q_1}\sigma_1^2, \ldots, \mathbf{I}_{q_r}\sigma_r^2\right\} \quad \text{such that } \sum q_i = q.$$

It is clear that (iii) implies $\mathbf{Y} | \beta, \mathbf{u}, \sigma_\varepsilon^2 \sim N(\mathbf{X}\beta + \mathbf{Z}\mathbf{u}, \mathbf{I}\sigma_\varepsilon^2)$. Now consider the following (improper) priors for the above hyperparameters

$$\pi_i\left(\sigma_i^2 | a_i\right) = \left(\sigma_i^2\right)^{-(a_i+1)}$$

$$\pi_\varepsilon\left(\sigma_\varepsilon^2 | b\right) = \left(\sigma_\varepsilon^2\right)^{-(b+1)} \tag{3.2}$$

where the a_i's and b are known constants and each distribution is supported on the positive half-line. We use this form because it resembles that of the usual inverted gamma (proper) prior distributions. We write the Bayesian hierarchical model defined above as

$$\mathbf{Y}|\mathbf{u},\sigma_\varepsilon^2,\beta \sim N(\mathbf{X}\beta+\mathbf{Z}\mathbf{u},\mathbf{I}\sigma_\varepsilon^2)$$

$$\pi(\beta)=1 \qquad \mathbf{u}|\sigma_1^2,\ldots,\sigma_r^2 \sim N(\mathbf{0},\mathbf{D}) \qquad \pi_\varepsilon(\sigma_\varepsilon^2|b)=\left(\sigma_\varepsilon^2\right)^{-(b+1)} \tag{3.3}$$

$$\pi_i(\sigma_i^2|a_i)=\left(\sigma_i^2\right)^{-(a_i+1)}.$$

If $b > -\frac{N}{2}$ and $a_i > -\frac{q_i}{2}$ for all i, the FCS for model (3.3) is as follows.

$$f(\sigma_i^2|\sigma_1^2,\ldots,\sigma_{i-1}^2,\sigma_{i+1}^2,\ldots,\sigma_r^2,\mathbf{y},\mathbf{u},\sigma_\varepsilon^2,\beta)=IG\left(a_i+\frac{q_i}{2},2(\mathbf{u}'\mathbf{u})^{-1}\right)$$

$$f(\sigma_\varepsilon^2|\sigma_1^2,\ldots,\sigma_r^2,\mathbf{y},\mathbf{u},\beta)=IG\left(b+\frac{N}{2},2\left\{(\mathbf{y}-(\mathbf{X}\beta+\mathbf{Z}\mathbf{u}))'(\mathbf{y}-(\mathbf{X}\beta+\mathbf{Z}\mathbf{u}))\right\}^{-1}\right)$$

$$f(\mathbf{u}|\sigma_1^2,\ldots,\sigma_r^2,\mathbf{y},\sigma_\varepsilon^2,\beta)=N\left(\left(\mathbf{Z}'\mathbf{Z}+\sigma_\varepsilon^2\mathbf{D}^{-1}\right)^{-1}\mathbf{Z}'(\mathbf{y}-\mathbf{X}\beta),\sigma_\varepsilon^2\left(\mathbf{Z}'\mathbf{Z}+\sigma_\varepsilon^2\mathbf{D}^{-1}\right)^{-1}\right) \tag{3.4}$$

$$f(\beta|\sigma_1^2,\ldots,\sigma_r^2,\mathbf{y},\sigma_\varepsilon^2,\mathbf{u})=N\left((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{y}-\mathbf{Z}\mathbf{u}),\sigma_\varepsilon^2(\mathbf{X}'\mathbf{X})^{-1}\right).$$

Before stating Theorem 1, which gives necessary and sufficient conditions for the posterior distribution, $\pi(\sigma_\varepsilon^2,\sigma_1^2,\ldots,\sigma_r^2,\mathbf{u},\beta|\mathbf{y})$, to be proper, we develop an interesting connection between model (3.3) and the estimation technique called restricted maximum likelihood (REML). We can write the likelihood function of the variance components for model (3.3) as

$$L(\sigma_\varepsilon^2,\sigma_1^2,\ldots,\sigma_r^2|\mathbf{y}) \overset{def}{=} f\left(\mathbf{Y}|\sigma_\varepsilon^2,\sigma_1^2,\ldots,\sigma_r^2\right)$$
$$= \int f(\mathbf{Y}|\mathbf{u},\beta,\sigma_\varepsilon^2) f(\mathbf{u}|\sigma_1^2,\ldots,\sigma_r^2)\, d\mathbf{u}\, d\beta$$

where we have used the following three assumptions concerning the conditional independence structure of model (3.3): (i) given $\mathbf{u}$, $\mathbf{Y}$ is conditionally independent of $\sigma_1^2,\ldots,\sigma_r^2$, (ii) given $\sigma_1^2,\ldots,\sigma_r^2$, $\mathbf{u}$ is conditionally independent of β and σ_ε^2, and (iii) β, σ_ε^2 and $\sigma_1^2,\ldots,\sigma_r^2$ are mutually independent. This likelihood is evaluated in the Appendix, and is given by

$$L(\sigma_\varepsilon^2,\sigma_1^2,\ldots,\sigma_r^2|\mathbf{y}) =$$

$$\frac{\exp\left\{\frac{1}{2}\mathbf{y}'\left((\mathbf{ZDZ}'+\mathbf{I}\sigma_\varepsilon^2)^{-1}\mathbf{X}\left(\mathbf{X}'(\mathbf{ZDZ}'+\mathbf{I}\sigma_\varepsilon^2)^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'(\mathbf{ZDZ}'+\mathbf{I}\sigma_\varepsilon^2)^{-1}-(\mathbf{ZDZ}'+\mathbf{I}\sigma_\varepsilon^2)^{-1}\right)\mathbf{y}\right\}}{(2\pi)^{\frac{N-p}{2}}\left(\sigma_\varepsilon^2\right)^{\frac{N-q}{2}}\left|\mathbf{X}'(\mathbf{ZDZ}'+\mathbf{I}\sigma_\varepsilon^2)^{-1}\mathbf{X}\right|^{\frac{1}{2}}\left|\mathbf{Z}'\mathbf{Z}+\sigma_\varepsilon^2\mathbf{D}^{-1}\right|^{\frac{1}{2}}}. \quad (3.5)$$

The above likelihood is equivalent to the REML likelihood (Searle, Casella and McCulloch 1992 p.323) derived in the frequentist setting by considering the density function of a linear transformation of the data, $\mathbf{K}'\mathbf{y}$, given the variance components. The matrix $\mathbf{K}$ is any N x (N-p) matrix of rank N-p such that $\mathbf{K}'\mathbf{X}=\mathbf{0}$. This transformation is justified in a number of ways (Searle, Casella and McCulloch 1992 p.249) and leads to data independent of β. The REML likelihood can also be written in terms of the matrix $\mathbf{K}$ (Searle, Casella and McCulloch 1992 p.323). This correspondence further motivates our study of model (3.3).

We now state the theorem.

Theorem 1. Assume that $N-p>q$ and let $t=rank(\mathbf{Z}'\mathbf{P}_x\mathbf{Z})\le q$ where $\mathbf{P}_x=(\mathbf{I}-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')$. There are two cases :

Case 1: If $t = q$ or if $r = 1$ then conditions (iv), (v) and (vi) below are necessary and sufficient for the posterior distribution of model (3.3) to be proper.

Case 2: If $t < q$ and $r > 1$ then conditions (iv), (v) and (vi) are sufficient for the posterior of model (3.3) to be proper while necessary conditions result when (v) is replaced with $(v')\ q_i > -2a_i$.

$$\text{(iv)}\ \ a_i < 0$$
$$\text{(v)}\ \ q_i > q - t - 2a_i$$
$$\text{(vi)}\ \ N + 2\sum a_i + 2b - p > 0.$$

Proof. See the Appendix.

Before we explore some pseudo-Gibbs chains we note that since the priors on the variance components in model (3.3) are simple powers, Theorem 1 can be used to quickly ascertain which posterior moments of the variance components exist. For instance, assume we are in case 1 and we have already discovered that the posterior for our special case of (3.3) is proper. Suppose we would now like to know if the marginal posterior distribution of σ_i^2 has a second moment. We can answer this question quickly using Theorem 1 by exploiting the following fact. The existence of the second moment of the marginal posterior distribution of σ_i^2 is equivalent to the integrability of the posterior of a slightly different version of our model in which a_i is replaced by $a_i - 2$. Therefore the marginal posterior distribution of σ_i^2 necessarily has a second moment if conditions (v) and (vi) of Theorem 1 still hold when a_i is replaced by $a_i - 2$.

Since interest here is in estimation of variance components, study of the model (3.1) is sometimes simplified with the additional assumption that β, the fixed effects vector, is known. In such a case we get a stronger result than Theorem 1, whose proof is virtually the same.

If β is known, the FCS is still given by (3.4) without the density for β. The following theorem can be proved.

Theorem 2. The posterior distribution for model (3.3), with β known, will be a proper distribution if and only if the following three conditions are satisfied for i = 1,...,r.

$$\text{(i)}\ a_i < 0$$
$$\text{(ii)}\ q_i > -2a_i$$
$$\text{(iii)}\ N + 2\sum a_i + 2b > 0.$$

4. A Detailed Example: One-Way Random Effects

In this section we consider the one-way random effects model in detail. First it is shown that all of the densities in the FCS are proper under minimal restrictions. The conditions required for a proper posterior distribution are then derived and compared with these minimal restrictions. Finally, a simulated data set is used to demonstrate what happens to the Gibbs sampler in two situations where the FCS consists of proper densities, but the posterior distribution is improper.

Consider the standard one-way random effects model

$$Y_{ij} = \mu + \theta_i + \varepsilon_{ij} \qquad i = 1,2,\ldots,k \quad j = 1,2,\ldots,J \tag{4.1}$$

where θ_i *iid* $N(0,\sigma_\theta^2)$, ε_{ij} *iid* $N(0,\sigma_\varepsilon^2)$, and the θ's and ε's are mutually independent. If we write this model as a Bayesian hierarchical model (using the priors discussed above) we have

$$Y_{ij}|\mu,\theta_i,\sigma_\varepsilon^2 \ iid\ N(\mu+\theta_i,\sigma_\varepsilon^2)$$

$$\pi(\mu) = 1 \qquad \theta_i|\sigma_\theta^2 \ iid\, N(0,\sigma_\theta^2) \qquad \pi(\sigma_\varepsilon^2) = \left(\sigma_\varepsilon^2\right)^{-(b+1)} \tag{4.2}$$

$$\pi(\sigma_\theta^2) = \left(\sigma_\theta^2\right)^{-(a+1)}.$$

One further assumption, which is not obvious from the model, is that of *conditional independence* which states that, conditional on the θ's, the Y_{ij}'s are independent of σ_θ^2. This assumption makes sense since a knowledge of σ_θ^2 should not effect the distribution of the Y_{ij}'s if we already know the θ's.

Model (4.2) is a fairly complicated Bayesian hierarchical model. No closed form solution exists for the posterior distribution $\pi(\sigma_\theta^2, \sigma_\varepsilon^2, \theta, \mu|\mathbf{y})$. This fact might lead one to consider using the Gibbs sampler in order to estimate some features of this posterior. However, after a bit of algebra, the FCS which is necessary for Gibbs sampling is given by

$$f(\sigma_\theta^2|\theta) = IG\left(\frac{k}{2}+a, 2\left(\sum_{i=1}^{k}\theta_i^2\right)^{-1}\right)$$

$$f(\sigma_\varepsilon^2|\theta,\mu,\mathbf{y}) = IG\left(\frac{N}{2}+b, 2\left(\sum_{i=1}^{k}\sum_{j=1}^{J}(y_{ij}-(\mu+\theta_i))^2\right)^{-1}\right)$$

$$f(\theta_i|\sigma_\theta^2,\sigma_\varepsilon^2,\mu,\mathbf{y}) = N\left(\frac{\sigma_\theta^2}{J\sigma_\theta^2+\sigma_\varepsilon^2}(y_{i\cdot}-J\mu), \frac{\sigma_\varepsilon^2\sigma_\theta^2}{J\sigma_\theta^2+\sigma_\varepsilon^2}\right) \tag{4.3}$$

$$f(\mu|\sigma_\varepsilon^2,\theta,\mathbf{y}) = N\left(\bar{y}_{\cdot\cdot}-\bar{\theta}_{\cdot}, \frac{\sigma_\varepsilon^2}{N}\right)$$

where

$$y_{i\cdot} = \frac{1}{J}\sum_{j=1}^{J}y_{ij} \qquad \bar{y}_{\cdot\cdot} = \frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{J}y_{ij} \qquad \bar{\theta}_{\cdot} = \frac{1}{k}\sum_{i=1}^{k}\theta_i\,.$$

The inverted gamma densities have the parameterization described in Berger (1985 p.561). These four conditional densities will be legitimate densities whenever $k+2a>0$ and $N+2b>0$, so the Gibbs sampler may be implemented whenever these two conditions hold.

Although the posterior distribution may not be calculated in closed form, we can, through the use of inequalities, establish which a's and b's lead to proper (integrable) posteriors. Since this model is a special case of the one considered in Section 3, we can apply Theorem 1 to derive conditions which guarantee a proper posterior.

If we regard the model (4.2) in matrix form, and reconcile notation with the general mixed model notation (3.1), we find that the $\mathbf{X}$ matrix is a vector of ones and the $\mathbf{Z}$ matrix is an incidence matrix. With some matrix algebra we then find that the rank of $\mathbf{Z}'\mathbf{P}_{\mathbf{x}}\mathbf{Z}$ is equal to $k-1$. We now apply Theorem 1, Case 1, with $t=k-1$, $p=1$, $q_i=q$ and $a_i=a$ to get the following Corollary.

Corollary 1. The one-way random effects model (4.1), with prior specification (4.2), will have a proper posterior distribution if and only if

$$\text{(i)}\ \ k-1>-2a>0$$
$$\text{(ii)}\ \ N+2a+2b-1>0$$

Note that the (a,b) pairs that satisfy Corollary 1 are a subset of those that satisfy $k+2a>0$ and $N+2b>0$, the conditions that make all of the densities in (4.3) proper. Thus, a proper posterior necessarily yields proper conditionals, but it is possible to have a legitimate set of densities in (4.3) that correspond to no proper posterior. In particular, choosing $a=0$ results in an improper posterior.

If the FCS consists of legitimate densities, but the posterior distribution is not proper, what happens to the pseudo-Gibbs chain? We now attempt to shed some light on this question with some simulations. The following data were simulated using model (4.1) with $\sigma_\theta^2 = 5, \sigma_\varepsilon^2 = 2$ and $\mu = 10$.

<u>Data y_{ij}</u>

Class	y_{i1}	y_{i2}	y_{i3}	y_{i4}	y_{i5}	$\bar{y}_{i\cdot}$
i = 1	8.60	8.98	9.32	7.07	8.23	8.44
i = 2	11.32	11.36	9.18	9.02	10.43	10.26
i = 3	7.74	5.05	7.38	6.11	6.31	6.52
i = 4	10.64	9.84	11.36	11.33	10.01	10.64
i = 5	14.08	10.74	14.13	14.84	14.84	13.73
i = 6	9.44	8.29	8.18	8.59	8.91	8.68
i = 7	11.35	10.98	11.71	8.73	11.05	10.76

In our experience with pseudo-Gibbs chains based on model (4.1) we have seen two different phenomena. In order to demonstrate these, we form two pseudo-Gibbs chains based on the two models shown in the table below.

Priors	Model 1	Model 2
$\pi(\sigma_\varepsilon^2)$	$(\sigma_\varepsilon^2)^{16}$	$\frac{1}{\sqrt{\sigma_\varepsilon^2}}$
$\pi(\sigma_\theta^2)$	$\frac{1}{\sqrt{\sigma_\theta^2}}$	$\frac{1}{(\sigma_\theta^2)^{\frac{3}{2}}}$
$\pi(\mu)$	1	1
$m(Y) < \infty$?	NO	NO

Model 1 corresponds to the (a, b) pair (-1/2, -17) and model 2 to (1/2, -1/2). The prior distribution on σ_ε^2 in Model 1 is admittedly ridiculous. However, any choice of b which violates the condition $N + 2b + 2a - 1 > 0$ will probably lead to a ridiculous prior if the magnitude of a is much smaller than N.

These two models were chosen for the following reasons.

a. They both yield legitimate FCS's, so a pseudo-Gibbs chain can be constructed for each.
b. They both have improper posterior distributions, so the output from the pseudo-Gibbs chain is nonsense.
c. Model 1 is "extremely bad" and can be detected as such. The impropriety of Model 2 is virtually undetectable by monitoring the output of the chain.

The same arbitrary starting values were used for both chains. For each chain a "burn-in" of 2 million was used (attempted) and then a final Gibbs chain of length 2000 was created by taking the next 2000 observations after the 2,000,000th. Write the chain as

$$^{(1)}\sigma_\varepsilon^2, {}^{(1)}\sigma_\theta^2, {}^{(1)}\theta_1, {}^{(1)}\theta_2, \ldots, {}^{(1)}\theta_7, {}^{(1)}\mu, {}^{(2)}\sigma_\varepsilon^2, {}^{(2)}\sigma_\theta^2, {}^{(2)}\theta_1, {}^{(2)}\theta_2, \ldots, {}^{(2)}\theta_7, {}^{(2)}\mu, {}^{(3)}\sigma_\varepsilon^2, \ldots$$

where the superscripts on the left refer to iteration number (by one iteration we mean one full cycle through all ten variables) and the variables with superscripts equal to 1 correspond to the starting values. We now describe the results.

The pseudo-Gibbs chain for model 1 could not be constructed because the chain "blew up" after about 1500 iterations. To be specific, each time we attempted to burn the chain in, the GAUSS program would terminate with an overflow message and at the time of termination, each of the ten variables had a value exceeding 10^{300}. Figure 1 is a histogram of the observations $^{(j)}\sigma_\theta^2$,

$j = 1,2,\ldots,1200$. (The minimum, maximum and mean of these numbers were $(0.80, 1.8\text{x}10^{60}, 5.7\text{x}10^{57})$.) Figure 2 is the analogous plot for σ_ε^2.

The pseudo-Gibbs chain for model 2, on the other hand, was quite well behaved. Figure 3 is a histogram of the observations ${}^{(i+2,000,000)}\sigma_\theta^2$, $i = 1,2,\ldots,2000$ along with a density estimate for $\pi\left(\sigma_\theta^2 | \mathbf{y}\right)$ calculated using the formula

$$\hat{\pi}_{\sigma_\theta^2}(t\,|\,\mathbf{y}) \approx \frac{1}{2000}\sum_{i=1}^{2000}\frac{1}{\Gamma(\alpha)\beta_i^\alpha t^{\alpha+1}}\exp\left\{-\frac{1}{\beta_i t}\right\} \tag{4.4}$$

where $\alpha = (k/2) + a$ and $\beta_i = \left(2 / \sum_{j=1}^{7} {}^{(i+2\cdot10^6)}\theta_j^2\right)$ as in (4.3). Figure 4 is the analogous plot for σ_ε^2. Both pictures appear quite reasonable and give no obvious indication that no joint posterior distribution, $\pi(\sigma_\theta^2, \sigma_\varepsilon^2, \theta, \mu | \mathbf{y})$, exists.

The conclusion from this example is that although some pseudo-Gibbs chains may "blow up" and warn the user that a problem exists, others can appear quite reasonable. This fact brings up an interesting question. "What are the probabilistic properties of these pseudo-Gibbs Markov chains and what do the usual Monte Carlo sums converge to if they converge at all?"

5. Conclusions

It is often the case that use of improper prior distributions within hierarchical exponential family and hierarchical linear mixed models lead to full conditional specifications with simple forms. Given these simple forms, it is easy, in fact trivial in some cases, to construct a Gibbs chain converging to a stationary distribution which is the posterior distribution. The existence of an FCS does not, unfortunately, guarantee that the corresponding posterior distribution is proper. It is therefore important to ascertain which improper priors result in proper posterior distributions for these models. The fact that it is possible to construct reasonable looking pseudo-Gibbs chains magnifies this importance.

In this paper, we have somewhat addressed this problem for hierarchical exponential family models and hierarchical linear mixed models. It should be noted that these two models are special cases of the conditionally independent hierarchical models (CIHM's) defined in Kass and Steffey (1989). It may be possible to find more general conditions for the existence of proper posterior distributions in terms of CIHM's.

Monte Carlo sums of observations from a true Gibbs chain possess many nice properties such as asymptotic normality (Tierney 1991). Many of these properties obtain because the Gibbs chains are positive Harris recurrent (Meyn and Tweedie 1993). The authors conjecture that this may not be the case when dealing with pseudo-Gibbs chains. An interesting problem would be to explore the transience/recurrence properties of these chains and their ramifications on Monte Carlo sums.

References

Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*, 2nd edition. New York: Springer-Verlag.

Berger, J. O. and Robert, C. (1990). Subjective hierarchical Bayes estimation of a multivariate normal mean: on the frequentist interface. *The Annals of Statistics*. **18** 617-651.

Casella, G. and George, E. I. (1992). Explaining the Gibbs sampler. *The American Statistician*. **46** 167-174.

Gaver, D. P. and O'Muircheartaigh, I. G. (1987). Robust empirical Bayes analysis of event rates. *Technometrics* **29** 1-15.

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

Gelman, A. and Speed, T. P. (1993). Characterizing a joint probability distribution by conditionals. *Journal of the Royal Statistical Society, Series B.*

George, E. I., Makov, U. E. and Smith, A. F. M. (1992). Conjugate likelihood distributions. CSS Technical Report #92-04, Center for Statistical Sciences, Univ. of Texas at Austin.

Hill, B. M. (1965). Inference about variance components in the one-way model. *J. Am. Statist. Assoc.* **60** 806-825.

Kass, R. E. and Steffey, D. (1989). Approximate Bayesian inference in conditionally independent hierarchical models (parametric empirical Bayes models). *J. Am. Statist. Assoc.* **84** 717-726.

Meyn, S. P. and Tweedie, R. L. (1993). *Markov Chains and Stochastic Stability.* London: Springer-Verlag.

Marshall, A. W. and Olkin, I. (1979). *Inequalities: Theory of Majorization and its Applications.* New York: Academic Press.

Smith, A. F. M. and Roberts, G. O. (1993). Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society, Series B.* **55** 3-23.

Searle, S. R. (1982). *Matrix Algebra Useful for Statistics.* New York: Wiley.

Searle, S. R., Casella, G. and McCulloch, C. E. (1992). *Variance Components.* New York: Wiley.

Tiao, G. C. and Tan, W. Y. (1965). Bayesian analysis of random-effect models in the analysis of variance. I. Posterior distribution of variance-components. *Biometrika* **52** 37-53.

Tierney, L. (1991). Markov chains for exploring posterior distributions. *The Annals of Statistics.*

Appendix

Before developing conditions under which the posterior distribution is proper, and hence so is the Gibbs chain, we define some notation and give a Lemma which is needed in the sequel. Let the (real-valued) eigenvalues of a v x v symmetric matrix $\mathbf{S}$ be written as

$$\lambda_{\max}(\mathbf{S}) = \lambda_1(\mathbf{S}) \geq \lambda_2(\mathbf{S}) \geq \cdots \geq \lambda_{v-1}(\mathbf{S}) \geq \lambda_v(\mathbf{S}) = \lambda_{\min}(\mathbf{S}).$$

Further, let $\lambda_{sp}(\mathbf{S})$ represent the smallest non-zero eigenvalue of $\mathbf{S}$. Then we have the following result.

Lemma (Marshall and Olkin 1979): If two symmetric matrices $\mathbf{S}_1$ and $\mathbf{S}_2$ are both n.n.d. then

$$|\mathbf{S}_1 + \mathbf{S}_2| \geq \prod_{i=1}^{v}\left[\lambda_i(\mathbf{S}_1) + \lambda_i(\mathbf{S}_2)\right] \qquad \text{and} \qquad |\mathbf{S}_1 + \mathbf{S}_2| \leq \prod_{i=1}^{v}\left[\lambda_i(\mathbf{S}_1) + \lambda_{n-i+1}(\mathbf{S}_2)\right].$$

Proof of Theorem 1.

Using the conditional independence structure of model (3.3), we can write

$\pi\left(\sigma_\varepsilon^2, \sigma_1^2, \ldots, \sigma_r^2, \mathbf{u}, \beta | \mathbf{y}\right)$

$$= \frac{f\left(\mathbf{Y}|\mathbf{u}, \sigma_\varepsilon^2\right) f\left(\mathbf{u}|\sigma_1^2, \ldots, \sigma_r^2\right) \pi_\varepsilon\left(\sigma_\varepsilon^2 | b\right) \pi_1\left(\sigma_1^2 | a_1\right) \cdots \pi_r\left(\sigma_r^2 | a_r\right) \pi(\beta)}{\int f\left(\mathbf{Y}|\mathbf{u}, \sigma_\varepsilon^2\right) f\left(\mathbf{u}|\sigma_1^2, \ldots, \sigma_r^2\right) \pi_\varepsilon\left(\sigma_\varepsilon^2 | b\right) \pi_1\left(\sigma_1^2 | a_1\right) \cdots \pi_r\left(\sigma_r^2 | a_r\right) \pi(\beta)\, d\mathbf{u}\, d\beta\, d\sigma_1^2 \cdots d\sigma_r^2\, d\sigma_\varepsilon^2} \qquad \text{(A.1)}$$

Let $m(\mathbf{y})$ represent the denominator of (A.1). The likelihood is

$$L\left(\sigma_\varepsilon^2, \sigma_1^2, \ldots, \sigma_r^2 | \mathbf{y}\right) = \int f(\mathbf{Y}|\mathbf{u}, \beta, \sigma_\varepsilon^2) f(\mathbf{u}|\sigma_1^2, \ldots, \sigma_r^2)\, d\mathbf{u}\, d\beta \qquad \text{(A.2)}$$

The integral on the right-hand side of (A.2) can be evaluated as follows. The inner integral (w.r.t. $\mathbf{u}$) can be calculated by "completing the quadratic form" and recognizing the multivariate normal density kernel. The outer integral (w.r.t. β) is done the same way using the fact that

$$\left(\mathbf{ZDZ}'+\mathbf{I}\sigma_\varepsilon^2\right)^{-1}=\frac{\mathbf{I}}{\sigma_\varepsilon^2}-\frac{1}{\sigma_\varepsilon^2}\mathbf{Z}\left(\sigma_\varepsilon^2\mathbf{D}^{-1}+\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}' \tag{A.3}$$

which follows from the Schur compliment (Searle 1982 p.261). Evaluation of this integral gives

$$L\left(\sigma_\varepsilon^2,\sigma_1^2,\ldots,\sigma_r^2|\mathbf{y}\right) =$$

$$\frac{\exp\left\{\frac{1}{2}\mathbf{y}'\left((\mathbf{ZDZ}'+\mathbf{I}\sigma_\varepsilon^2)^{-1}\mathbf{X}\left(\mathbf{X}'(\mathbf{ZDZ}'+\mathbf{I}\sigma_\varepsilon^2)^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'(\mathbf{ZDZ}'+\mathbf{I}\sigma_\varepsilon^2)^{-1}-(\mathbf{ZDZ}'+\mathbf{I}\sigma_\varepsilon^2)^{-1}\right)\mathbf{y}\right\}}{(2\pi)^{\frac{N-p}{2}}\left(\sigma_\varepsilon^2\right)^{\frac{N-q}{2}}\left|\mathbf{X}'(\mathbf{ZDZ}'+\mathbf{I}\sigma_\varepsilon^2)^{-1}\mathbf{X}\right|^{\frac{1}{2}}\left|\mathbf{Z}'\mathbf{Z}+\sigma_\varepsilon^2\mathbf{D}^{-1}\right|^{\frac{1}{2}}}. \tag{A.4}$$

Using (A.3) we can rewrite the determinant

$$\begin{aligned}\left|\mathbf{X}'(\mathbf{ZDZ}'+\mathbf{I}\sigma_\varepsilon^2)^{-1}\mathbf{X}\right|^{\frac{1}{2}} &= \left(\sigma_\varepsilon^2\right)^{-\frac{p}{2}}\left|\mathbf{X}'\mathbf{X}\right|^{\frac{1}{2}}\left|\mathbf{I}-(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}+\sigma_\varepsilon^2\mathbf{D}^{-1}\right)^{-1}\mathbf{Z}'\mathbf{X}\right|^{\frac{1}{2}}\\ &= \left(\sigma_\varepsilon^2\right)^{-\frac{p}{2}}\left|\mathbf{X}'\mathbf{X}\right|^{\frac{1}{2}}\left|\mathbf{I}-\mathbf{Z}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}+\sigma_\varepsilon^2\mathbf{D}^{-1}\right)^{-1}\right|^{\frac{1}{2}}\\ &= \left(\sigma_\varepsilon^2\right)^{-\frac{p}{2}}\left|\mathbf{X}'\mathbf{X}\right|^{\frac{1}{2}}\left|\sigma_\varepsilon^2\mathbf{D}^{-1}+\mathbf{Z}'\left(\mathbf{I}-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right)\mathbf{Z}\right|^{\frac{1}{2}}\left|\mathbf{Z}'\mathbf{Z}+\sigma_\varepsilon^2\mathbf{D}^{-1}\right|^{-\frac{1}{2}}.\end{aligned}$$

Substituting this into (A.4) yields

$$L\left(\sigma_\varepsilon^2,\sigma_1^2,\ldots,\sigma_r^2|\mathbf{y}\right) =$$

$$\frac{\exp\left\{\frac{1}{2}\mathbf{y}'\left((\mathbf{ZDZ}'+\mathbf{I}\sigma_\varepsilon^2)^{-1}\mathbf{X}\left(\mathbf{X}'(\mathbf{ZDZ}'+\mathbf{I}\sigma_\varepsilon^2)^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'(\mathbf{ZDZ}'+\mathbf{I}\sigma_\varepsilon^2)^{-1}-(\mathbf{ZDZ}'+\mathbf{I}\sigma_\varepsilon^2)^{-1}\right)\mathbf{y}\right\}}{(2\pi)^{\frac{N-p}{2}}\left(\sigma_\varepsilon^2\right)^{\frac{N-q-p}{2}}|\mathbf{D}|^{\frac{1}{2}}|\mathbf{X}'\mathbf{X}|^{\frac{1}{2}}|\sigma_\varepsilon^2\mathbf{D}^{-1}+\mathbf{Z}'\mathbf{P}_\mathbf{x}\mathbf{Z}|^{\frac{1}{2}}} \tag{A.5}$$

where $\mathbf{P}_\mathbf{x}$ is defined in the statement of Theorem 1. We use the following notation throughout the remainder of this proof.

$$\mathbf{M}_1 = (\mathbf{ZDZ}'+\mathbf{I}\sigma_\varepsilon^2)^{-1} \quad \text{and} \quad \mathbf{M}_2 = \left(\mathbf{X}'(\mathbf{ZDZ}'+\mathbf{I}\sigma_\varepsilon^2)^{-1}\mathbf{X}\right)^{-1}.$$

The next step is to find conditions for which the integral

$$\int L\left(\sigma_\varepsilon^2,\sigma_1^2,\ldots,\sigma_r^2\,|\,\mathbf{y}\right)\pi_1\left(\sigma_1^2|a_1\right)\cdots\pi_r\left(\sigma_r^2|a_r\right)\pi_\varepsilon\left(\sigma_\varepsilon^2|b\right)d\sigma_1^2\cdots d\sigma_r^2 =$$

$$\frac{\pi_\varepsilon\left(\sigma_\varepsilon^2|b\right)}{(2\pi)^{\frac{N-p}{2}}\left(\sigma_\varepsilon^2\right)^{\frac{N-q-p}{2}}|\mathbf{X}'\mathbf{X}|^{\frac{1}{2}}}\int\frac{\exp\left\{\frac{1}{2}\mathbf{y}'(\mathbf{M}_1\mathbf{X}\mathbf{M}_2\mathbf{X}'\mathbf{M}_1-\mathbf{M}_1)\mathbf{y}\right\}\pi_1\left(\sigma_1^2|a_1\right)\cdots\pi_r\left(\sigma_r^2|a_r\right)}{|\mathbf{D}|^{\frac{1}{2}}|\sigma_\varepsilon^2\mathbf{D}^{-1}+\mathbf{Z}'\mathbf{P}_\mathbf{x}\mathbf{Z}|^{\frac{1}{2}}}d\sigma_1^2\cdots d\sigma_r^2 \tag{A.6}$$

is finite. We first examine the exponential function in (A.6). Write it as

$$f(\mathbf{t}) = \exp\left\{\frac{1}{2}\mathbf{y}'(\mathbf{M}_1\mathbf{X}\mathbf{M}_2\mathbf{X}'\mathbf{M}_1-\mathbf{M}_1)\mathbf{y}\right\}$$

where $\mathbf{t}=\left(\sigma_1^2,\ldots,\sigma_r^2\right)$. A straightforward, but lengthy, differentiation argument will verify that $f(\mathbf{t})$ is non-decreasing in all of its arguments. Since

$$f(\mathbf{0}) = \exp\left\{-\frac{1}{2\sigma_\varepsilon^2}\mathbf{y}'\mathbf{P}_\mathbf{X}\mathbf{y}\right\}$$

and (A.7)

$$\lim_{\mathbf{t}\to\infty} f(\mathbf{t}) = \exp\left\{-\frac{1}{2\sigma_\varepsilon^2}\mathbf{y}'\left(\mathbf{P}_\mathbf{z}-\mathbf{P}_\mathbf{z}\mathbf{X}(\mathbf{X}'\mathbf{P}_\mathbf{z}\mathbf{X})^{-1}\mathbf{X}'\mathbf{P}_\mathbf{z}\right)\mathbf{y}\right\}$$

where $\mathbf{P_Z} = \left(\mathbf{I} - \mathbf{Z}(\mathbf{Z'Z})^{-1}\mathbf{Z'}\right)$, the exponential term is bounded and the integral in (A.6) will be finite if and only if

$$\int \frac{\pi_1(\sigma_1^2|a_1)\cdots\pi_r(\sigma_r^2|a_r)}{|\mathbf{D}|^{\frac{1}{2}}|\sigma_\varepsilon^2\mathbf{D}^{-1} + \mathbf{Z'P_xZ}|^{\frac{1}{2}}} d\sigma_1^2 \cdots d\sigma_r^2 < \infty. \qquad \text{(A.8)}$$

Note that we have assumed that $\mathbf{X'P_zX}$ is invertible. If this is not the case, a similar argument using the generalized inverse $(\mathbf{X'P_zX})^-$ can be used to establish condition (A.6).

This is the point where the proofs of cases 1 and 2 become different.

Case 1: $t = q$ ($\mathbf{Z'P_XZ}$ is of full rank) or $r = 1$.

We first consider $t = q$. From the Lemma we have

$$\prod_{i=1}^{r}\left(\frac{\sigma_\varepsilon^2}{\sigma_i^2} + \lambda_{\min}(\mathbf{Z'P_XZ})\right)^{q_i} \le \left|\mathbf{Z'P_XZ} + \sigma_\varepsilon^2\mathbf{D}^{-1}\right| \le \prod_{i=1}^{r}\left(\frac{\sigma_\varepsilon^2}{\sigma_i^2} + \lambda_{\max}(\mathbf{Z'P_XZ})\right)^{q_i}.$$

Substituting into (A.8) shows that the integral is finite if

$$\prod_{i=1}^{r}\int_0^\infty \frac{\left(\sigma_i^2\right)^{-(a_i+1)}}{\left(\sigma_\varepsilon^2 + \sigma_i^2\cdot\lambda_{\min}(\mathbf{Z'P_XZ})\right)^{\frac{q_i}{2}}} d\sigma_i^2 < \infty. \qquad \text{(A.9)}$$

Analogously, replacing $\lambda_{\min}(\mathbf{Z'P_XZ})$ in (A.9) with $\lambda_{\max}(\mathbf{Z'P_XZ})$ will result in a necessary condition for finiteness of (A.8). However, if (A.9) is true with $\lambda_{\min}(\mathbf{Z'P_XZ})$, it is also true with $\lambda_{\max}(\mathbf{Z'P_XZ})$, which results in the condition that (A.8) is true if and only if

$$\int_0^\infty \frac{\left(\sigma_i^2\right)^{-(a_i+1)}}{\left(\sigma_\varepsilon^2 + \lambda\sigma_i^2\right)^{\frac{q_i}{2}}} d\sigma_i^2 < \infty \text{ for } i = 1,\ldots,r \qquad \text{(A.10)}$$

where $\lambda > 0$ is constant. If $a_i < 0$ and $q_i > -2a_i$ then

$$\int_0^\infty \frac{\left(\sigma_i^2\right)^{-(a_i+1)}}{\left(\sigma_\varepsilon^2 + \lambda\sigma_i^2\right)^{\frac{q_i}{2}}} d\sigma_i^2 = c_i\left(\sigma_\varepsilon^2\right)^{-\left(a_i+\frac{q_i}{2}\right)} \tag{A.11}$$

where c_i is constant w.r.t. σ_ε^2. If either one of these conditions is not satisfied, the integral will diverge. Therefore, the product of integrals in (A.9) will be finite if and only if conditions (iv) and (v) of Theorem 1 are satisfied. When they are indeed satisfied, we have

$$\prod_{i=1}^{r}\int_0^\infty \frac{\left(\sigma_i^2\right)^{-(a_i+1)}}{\left(\sigma_\varepsilon^2 + \lambda\sigma_i^2\right)^{\frac{q_i}{2}}} d\sigma_i^2 = c\left(\sigma_\varepsilon^2\right)^{-\left(\frac{q}{2}+\sum a_i\right)} \tag{A.12}$$

where c is constant w.r.t. σ_ε^2 and we use the fact that $\sum q_i = q$.

Combining expression (A.12) with (A.6), and adding the integral over σ_ε^2, we find that $m(\mathbf{y})$ will be finite if

$$\int_0^\infty \frac{\exp\left\{-\frac{1}{2\sigma_\varepsilon^2}\mathbf{y}'\left(\mathbf{P_z} - \mathbf{P_z X}\left(\mathbf{X'P_z X}\right)^{-1}\mathbf{X'P_z}\right)\mathbf{y}\right\}}{\left(\sigma_\varepsilon^2\right)^{\frac{N+2\sum a_i+2b-p+2}{2}}} d\sigma_\varepsilon^2 < \infty \tag{A.13}$$

where we have replaced the exponential term in (A.6) with the upper bound in (A.7). However, (A.10) is proportional to an inverted gamma density, and therefore the integral in (A.13) will be finite if condition (vi) of Theorem 1 is satisfied.

Thus, we have proved that conditions (iv), (v) and (vi) are sufficient. It is easy to demonstrate that they are necessary using only what has already been estimated. Replacing the exponential function in (A.6) with the lower bound in (A.7) shows that conditions (iv) and (v) are necessary. That condition (vi) is necessary given that (iv) and (v) hold follows from the fact that

$$m(\mathbf{y}) \geq \int_0^\infty \frac{\exp\left\{-\frac{1}{2\sigma_\varepsilon^2}\mathbf{y}'\mathbf{P_X}\mathbf{y}\right\}}{\left(\sigma_\varepsilon^2\right)^{\frac{N+2\sum a_i+2b-p+2}{2}}} d\sigma_\varepsilon^2 \tag{A.14}$$

since, as before, the integrand in (A.14) is the kernel of an inverted gamma density function.

We have thus proved the Theorem when $t = q$, that is, if $\mathbf{Z'P_XZ}$ is of full rank. The other part of Case 1 is where $r = 1$, but $\mathbf{Z'P_XZ}$ is not of full rank. In that case write

$$\left|\sigma_\varepsilon^2\mathbf{D}^{-1} + \mathbf{Z'P_XZ}\right|^{\frac{1}{2}} = \left|\mathbf{I}\frac{\sigma_\varepsilon^2}{\sigma_1^2} + \mathbf{Z'P_XZ}\right|^{\frac{1}{2}} = \left|\mathbf{I}\frac{\sigma_\varepsilon^2}{\sigma_1^2} + \mathbf{H'\Lambda H}\right|^{\frac{1}{2}}$$

where $\mathbf{H}$ is orthogonal and Λ is a diagonal matrix of the eigenvalues of $\mathbf{Z'P_XZ}=\mathbf{Z'P_X'P_XZ}$. Since $\mathbf{Z'P_XZ}$ is a positive semi-definite (p.s.d.) matrix it has t positive eigenvalues and $q-t$ zero eigenvalues. Therefore

$$\left|\mathbf{I}\frac{\sigma_\varepsilon^2}{\sigma_1^2} + \mathbf{H'\Lambda H}\right|^{\frac{1}{2}} = \left|\mathbf{I}\frac{\sigma_\varepsilon^2}{\sigma_1^2} + \Lambda\right|^{\frac{1}{2}}$$

$$\geq \left(\frac{\sigma_\varepsilon^2}{\sigma_1^2} + \lambda_{sp}\right)^{\frac{t}{2}}\left(\frac{\sigma_\varepsilon^2}{\sigma_1^2}\right)^{\frac{q-t}{2}}.$$

The arguments above can then be used to establish the theorem.

Case 2: $t < q$ and $r > 1$.

In order to integrate over $\Re^{r+}$ in (A.8), we will integrate over the mutually exclusive sets

$$\left\{\left(\sigma_{i_1}^2,\ldots,\sigma_{i_r}^2\right)\in\Re^{r+}\ s.t.\sigma_{i_1}^2<\cdots<\sigma_{i_r}^2\right\} \tag{A.15}$$

where $(i_1,\ldots,i_r)$ is one of the $r!$ permutations of $(1,\ldots,r)$ and then add the results. Let S denote the set in (A.15) with $(i_1,\ldots,i_r)=(1,\ldots,r)$ and consider

$$\int_S \frac{\pi_1\left(\sigma_1^2|a_1\right)\cdots\pi_r\left(\sigma_r^2|a_r\right)}{|\mathbf{D}|^{\frac{1}{2}}\left|\sigma_\varepsilon^2\mathbf{D}^{-1}+\mathbf{Z}'\mathbf{P}_\mathbf{x}\mathbf{Z}\right|^{\frac{1}{2}}}d\sigma_1^2\cdots d\sigma_r^2. \tag{A.16}$$

Using the Lemma we have

$$\begin{aligned}\left|\sigma_\varepsilon^2\mathbf{D}^{-1}+\mathbf{Z}'\mathbf{P}_\mathbf{x}\mathbf{Z}\right| &\geq \prod_{i=1}^{q}\left[\lambda_i\left(\sigma_\varepsilon^2\mathbf{D}^{-1}\right)+\lambda_i(\mathbf{Z}'\mathbf{P}_\mathbf{x}\mathbf{Z})\right]\\ &\geq \prod_{i=1}^{t}\left[\lambda_i\left(\sigma_\varepsilon^2\mathbf{D}^{-1}\right)+\lambda_{sp}(\mathbf{Z}'\mathbf{P}_\mathbf{x}\mathbf{Z})\right]\prod_{i=t+1}^{q}\left[\lambda_i\left(\sigma_\varepsilon^2\mathbf{D}^{-1}\right)\right]\end{aligned}$$

and on the set S

$$\prod_{i=1}^{t}\left[\lambda_i\left(\sigma_\varepsilon^2\mathbf{D}^{-1}\right)+\lambda_{sp}(\mathbf{Z}'\mathbf{P}_\mathbf{x}\mathbf{Z})\right]\prod_{i=t+1}^{q}\left[\lambda_i\left(\sigma_\varepsilon^2\mathbf{D}^{-1}\right)\right]=$$

$$\left(\frac{\sigma_\varepsilon^2}{\sigma_1^2}+\lambda_{sp}(\mathbf{Z}'\mathbf{P}_\mathbf{X}\mathbf{Z})\right)^{q_1}\left(\frac{\sigma_\varepsilon^2}{\sigma_2^2}+\lambda_{sp}(\mathbf{Z}'\mathbf{P}_\mathbf{X}\mathbf{Z})\right)^{q_2}\cdots\left(\frac{\sigma_\varepsilon^2}{\sigma_r^2}+\lambda_{sp}(\mathbf{Z}'\mathbf{P}_\mathbf{X}\mathbf{Z})\right)^{t-\sum_{i=1}^{r-1}q_i}\left(\frac{\sigma_\varepsilon^2}{\sigma_r^2}\right)^{q-t}.$$

Substituting into (A.16) gives

$$\int_S \frac{\pi_1\left(\sigma_1^2|a_1\right)\cdots\pi_r\left(\sigma_r^2|a_r\right)}{|\mathbf{D}|^{\frac{1}{2}}\left|\sigma_\varepsilon^2\mathbf{D}^{-1}+\mathbf{Z}'\mathbf{P}_\mathbf{x}\mathbf{Z}\right|^{\frac{1}{2}}}d\sigma_1^2\cdots d\sigma_r^2$$

$$\leq \int_0^{\infty}\cdots\int_0^{\infty} \frac{\left(\sigma_\varepsilon^2\right)^{\frac{t-q}{2}} \pi_1\left(\sigma_1^2 | a_1\right)\cdots\pi_r\left(\sigma_r^2 | a_r\right)}{\left(\sigma_\varepsilon^2 + \sigma_r^2 \lambda_{sp}(\mathbf{Z}'\mathbf{P}_x\mathbf{Z})\right)^{\frac{t}{2} - \sum_{i=1}^{r-1}\frac{q_i}{2}} \prod_{i=1}^{r-1}\left(\sigma_\varepsilon^2 + \sigma_i^2 \lambda_{sp}(\mathbf{Z}'\mathbf{P}_x\mathbf{Z})\right)^{\frac{q_i}{2}}} d\sigma_1^2 \cdots d\sigma_r^2 \qquad \text{(A.17)}$$

As in the previous argument, under conditions (iv) and (v) the integral in (A.17) equals

$$c\left(\sigma_\varepsilon^2\right)^{-\left(\frac{q}{2}+\sum a_i\right)}$$

where c is a constant which does not depend on σ_ε^2. This result depends on the permutation $(i_1,\ldots,i_r) = (1,\ldots,r)$ only through the constant c. Therefore

$$\int \frac{\pi_1\left(\sigma_1^2 | a_1\right)\cdots\pi_r\left(\sigma_r^2 | a_r\right)}{|\mathbf{D}|^{\frac{1}{2}}\left|\sigma_\varepsilon^2\mathbf{D}^{-1} + \mathbf{Z}'\mathbf{P}_x\mathbf{Z}\right|^{\frac{1}{2}}} d\sigma_1^2 \cdots d\sigma_r^2 \leq c'\left(\sigma_\varepsilon^2\right)^{-\left(\frac{q}{2}+\sum a_i\right)}$$

where, again, c' does not vary with σ_ε^2. The remainder of the proof of this case closely resembles the corresponding part of the proof of Case 1.

□

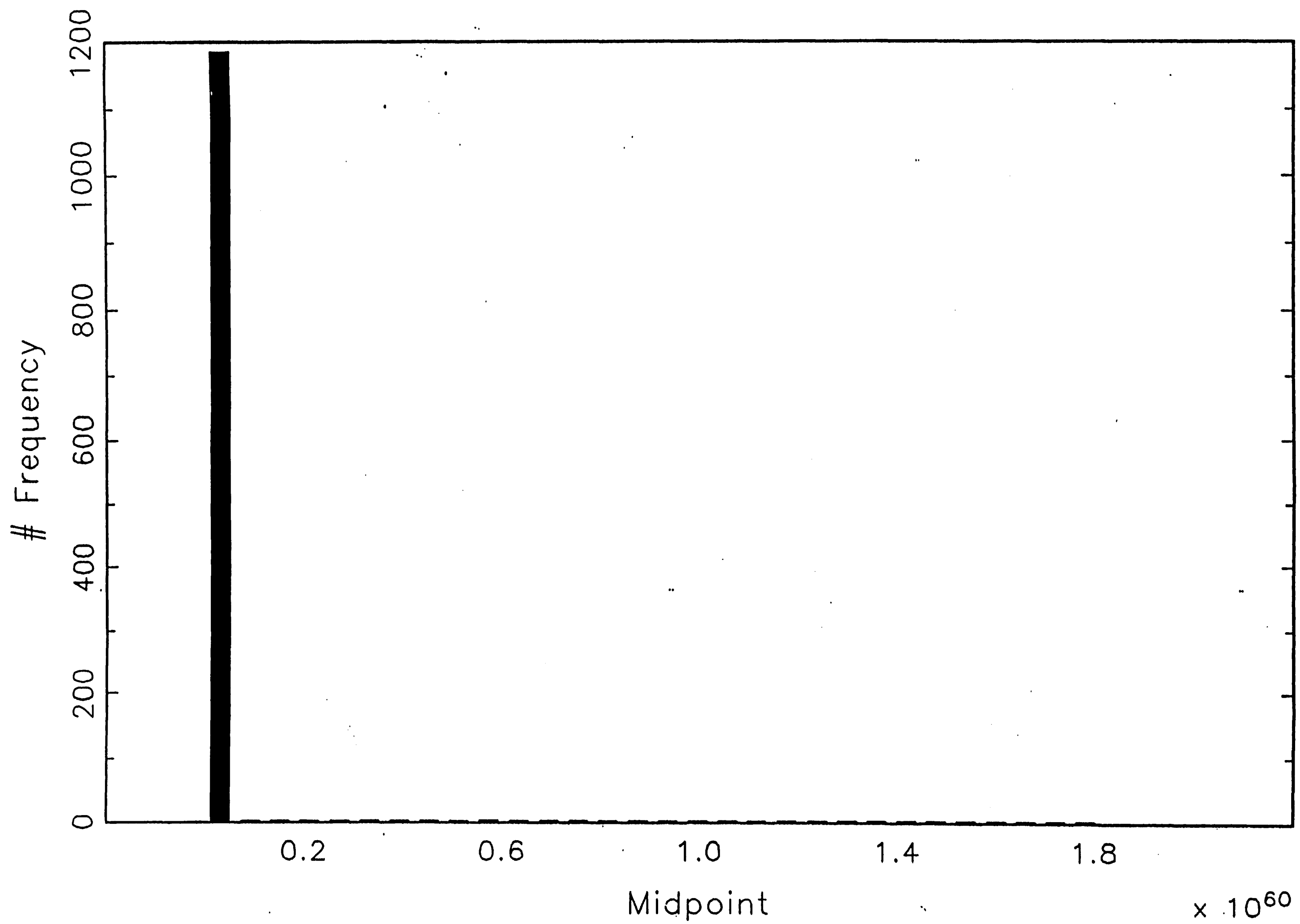

Figure 1. Histogram of ${}^{(i)}\sigma_\theta^2$, $i = 1,\dots,1200$, for model 1.

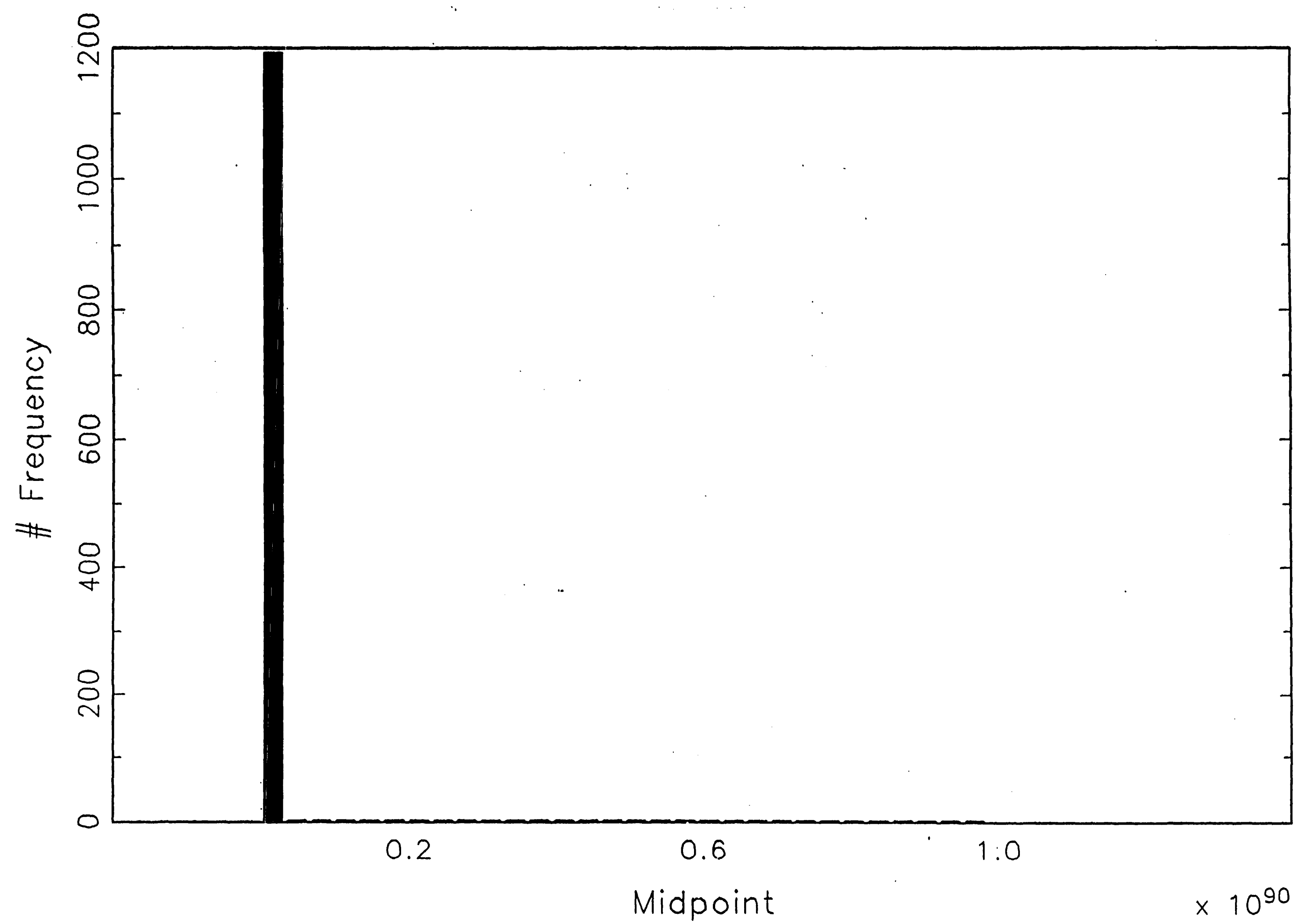

Figure 2. Histogram of ${}^{(i)}\sigma_{\varepsilon}^{2}$, $i = 1, \ldots, 1200$, for model 1.

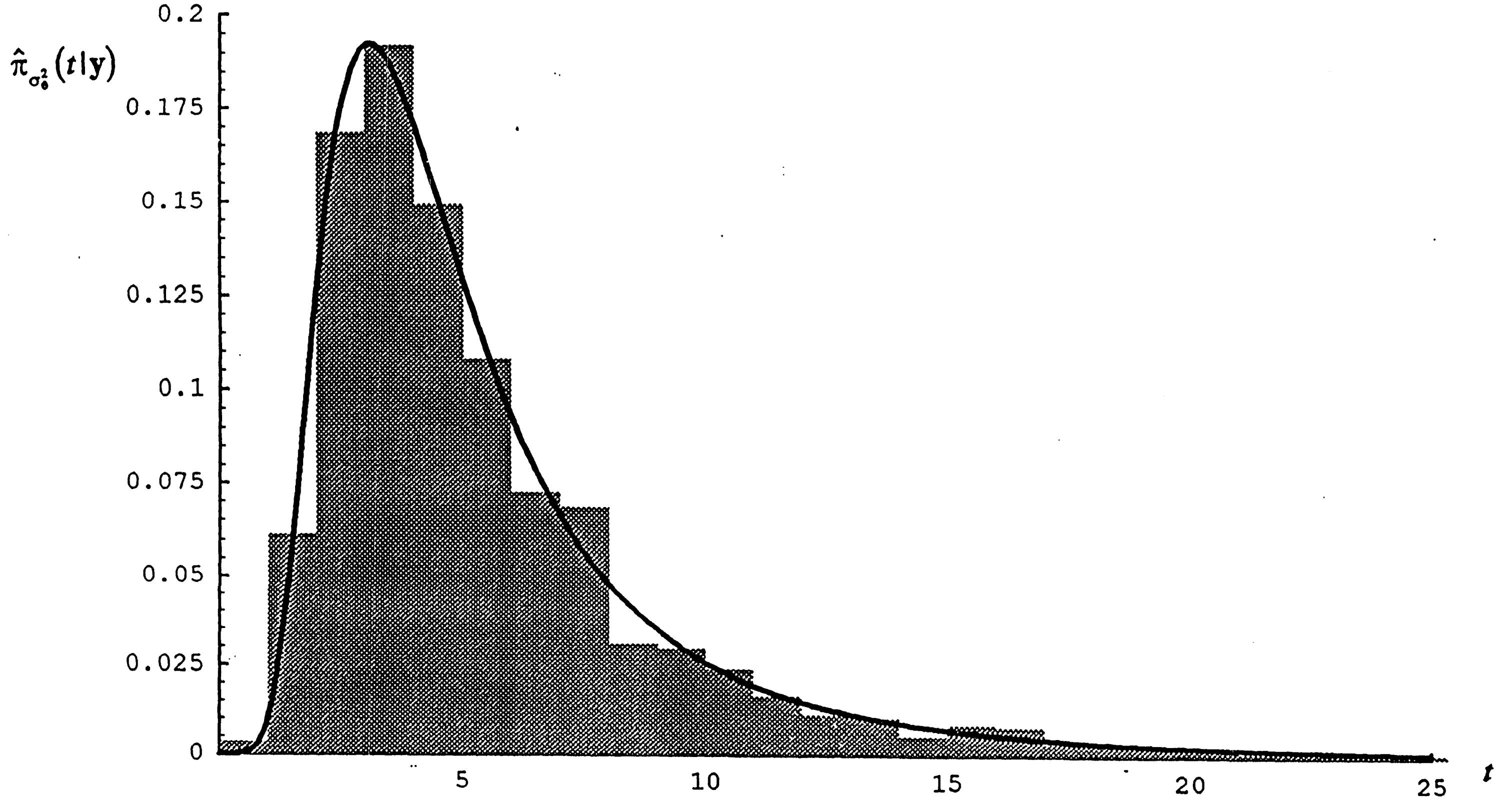

Figure 3. Density estimate of $\pi(\sigma_\theta^2|\mathbf{y})$ and histogram of ${}^{(i+2\cdot10^6)}\sigma_\theta^2$, $i=1,\ldots,2000$, for model 2.

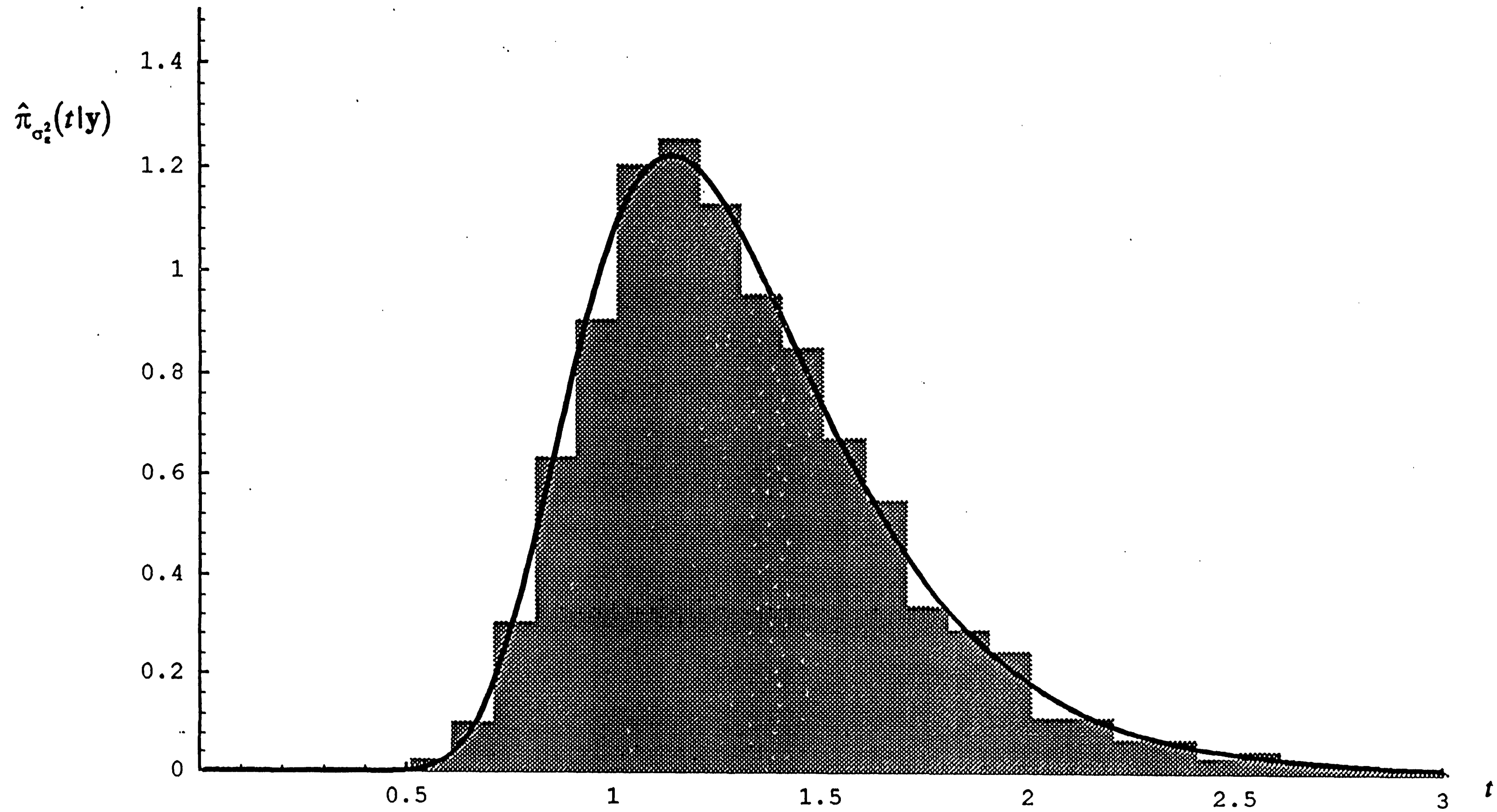

Figure 4. Density estimate of $\pi(\sigma_\varepsilon^2|\mathbf{y})$ and histogram of ${}^{(i+2\cdot10^6)}\sigma_\varepsilon^2$, $i=1,\ldots,2000$, for model 2.