ILLUSTRATION OF BAYESIAN INFERENCE IN NORMAL DATA MODELS USING GIBBS SAMPLING

BY

ALAN E. GELFAND

SUSAN E. HILLS

AMY RACINE-POON

ADRIAN F. M. SMITH

TECHNICAL REPORT NO. 421

SEPTEMBER 6, 1989

PREPARED UNDER CONTRACT

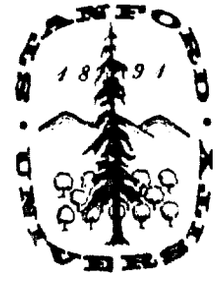
N00014-89-J-1627 (NR-042-267)

FOR THE OFFICE OF NAVAL RESEARCH

DEPARTMENT OF STATISTICS

STANFORD UNIVERSITY

STANFORD, CALIFORNIA



(4)

ILLUSTRATION OF BAYESIAN INFERENCE IN NORMAL DATA MODELS USING GIBBS SAMPLING

BY

ALAN E. GELFAND

SUSAN E. HILLS

AMY RACINE-POON

ADRIAN F. M. SMITH

TECHNICAL REPORT NO. 421

SEPTEMBER 6, 1989

Prepared Under Contract

N00014-89-J-1627 (NR-042-267)

For the Office of Naval Research

Herbert Solomon, Project Director

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

1. Introduction

Technical difficulties arising in the calculation of marginal posterior densities needed for Bayesian inference have long served as an impediment to the wider application of the Bayesian framework to real data. In the last few years there have been a number of advances in numerical and analytic approximation techniques for such calculations—see, for example, Naylor and Smith (1982, 1988), Smith *et al* (1985, 1987), Tierney and Kadane (1986), Shaw (1988), Geweke (1988)—but implementation of these approaches typically requires sophisticated numerical or analytic approximation expertise and possibly specialist software. In a recent paper, Gelfand and Smith (1988) described sampling based approaches for such calculations, which, by contrast, are essentially trivial to implement, even with limited computing resources. In this previous paper we entered caveats regarding the computational efficiency of such sampling based approaches, but our continuing investigations have shown that adaptive, iterative sampling achieved through the Gibbs sampler (Geman and Geman, 1984) is, in fact, surprisingly efficient, converging remarkably quickly for a wide range of problems.

Our objective in this paper is to provide illustrations of a range of applications of the Gibbs sampler in order to demonstrate its versatility and ease of implementation in practice. We begin by briefly reviewing the Gibbs sampler in Section 2. In Section 3, based upon computational experience with a variety of problems, we offer several suggestions on assessing the convergence of this iterative algorithm. In Section 4 we begin our illustrative analysis with a variance components model applied to a 'nasty' data set introduced in Box and Tiao (1973), whose Bayesian analysis therein involved elaborate exact and asymptotic methods. In addition, we illustrate the ease with which inferences for functions of parameters, such as ratios, can be made using Gibbs sampling. In Section 5, we take up the k-sample normal means problem in the general case of unbalanced data with unknown population variances. In particular, we show that the previously inaccessible case where the population means are ordered is straightforwardly handled through Gibbs sampling. Application is made to an unbalanced generated data set from normal populations with known ordered means and severely non-homogeneous variances. In Section 6, we look at a population linear growth curve model, as an illustration of the power of the Gibbs sampler in handling complex hierarchical models. We analyse data on the response over time of 30 rats to a treatment, with a total of 66

parameters in the hierarchical model specification. In Section 7, we analyse a two-period cross-over design involving the comparison of two drug formulations, in order to illustrate the ease with which the Gibbs sampler deals with complications arising from missing data in an originally balanced design. A summary discussion is provided in Section 8.

2. Gibbs sampling

In the sequel, densities will be denoted, generically, by square brackets so that joint, conditional and marginal forms appear, respectively, as $[X,Y]$, $[X|Y]$ and $[Y]$. The usual marginalisation by integration procedure will be denoted by forms such as

$$[X] = \int [X|Y]\cdot[Y].$$

Throughout, we shall be dealing with collections of random variables for which it is known (see, for example, Besag, 1974) that specification of all full conditional distributions uniquely determines the full joint density. More precisely, for such a collection of random variables $U_1, U_2, \ldots, U_k$, the joint density, $[U_1, U_2, \ldots, U_k]$, is uniquely determined by $[U_s|U_r, r \neq s]$, $s = 1,2,\ldots,k$. Our interest is in the marginal distributions, $[U_s]$, $s = 1,2,\ldots,k$.

An algorithm for extracting marginal distributions from the full conditional distribution was formally introduced as the Gibbs sampler in Geman and Geman (1984). The algorithm requires all the full conditional distributions to be 'available' for sampling, where 'available' is taken to mean that, for example, U_s can be generated straightforwardly and efficiently given specified values of the conditioning variables, U_r, $r \neq s$.

Gibbs sampling is a Markovian updating scheme which proceeds as follows. Given an arbitrary starting set of values $U_1^{(0)}, \ldots, U_k^{(0)}$, we draw $U_1^{(1)}$ from $[U_1|U_2^{(0)}, \ldots, U_k^{(0)}]$, then $U_2^{(1)}$ from $[U_2|U_1^{(1)}, U_3^{(0)}, \ldots, U_k^{(0)}]$ ... and so on up to $U_k^{(1)}$ from $[U_k|U_1^{(1)}, \ldots, U_{k-1}^{(1)}]$ to complete one iteration of the scheme. After t such iterations we would arrive at $(U_1^{(t)}, \ldots, U_k^{(t)})$. Geman and Geman show under mild conditions that $U_s^{(t)} \xrightarrow{d} U_s \sim [U_s]$ as $t \to \infty$. Thus, for t large enough we can regard $U_s^{(t)}$ as a simulated

observation from $[U_s]$.

Replicating this process m times produces m iid k-tuples $(U_{1j}^{(i)},\ldots,U_{kj}^{(i)})$, $j = 1,\ldots,m$. For any s, the collection $U_{s1}^{(i)},\ldots,U_{sm}^{(i)}$ can be viewed as a simulated sample from $[U_s]$. The marginal density is then estimated by the finite mixture density

$$\hat{[U_s]} = m^{-1} \sum_{j=1}^{m} [U_s | U_r = U_{rj}^{(i)}, r \neq s]. \tag{1}$$

See Gelfand and Smith, 1988, for further discussion.)

Since the expression (1) can be viewed as a 'Rao-Blackwellized' density estimator, relative to the more usual kernel density estimators based upon $U_{sj}^{(i)}$, $j = 1,\ldots,m$, estimation efficiency is high and we find $m = 100$ (at most 200) to be adequate in practice as a converged sample size on which to base the marginal density estimate.

Suppose interest centres on the marginal distribution for a variable V which is a function $g(U_1,\ldots,U_k)$ of $U_1,\ldots,U_k$. We note that evaluation of g at each of the $(U_{1j}^{(i)},\ldots,U_{kj}^{(i)})$ provides samples of V. In this case, a density estimate of the form (1) is not available, but an ordinary kernel density estimate can readily be calculated (see Section 4 for an illustration of this).

All applications we consider in this paper are within the Bayesian framework, where the U_s are unobservable, representing either parameters or missing data (and V can thus be a function of the parameters which we are interested in). All distributions will be viewed as conditional on the observed data, whence marginal distributions become the marginal posteriors needed for Bayesian inference or prediction.

3. Convergence diagnostics

Our experience thus far, in a variety of real and simulated data analyses, shows remarkably rapid convergence of the Gibbs sampler. In acquiring this computational experience, we have experimented with a variety of diagnostic tools to facilitate concluding whether or not the algorithm has 'converged'. Generally, the most natural and least sophisticated approach has been the most useful. Namely, for a fixed m

we increase t, overlay plots of the resulting estimated densities, and see if the estimates are equivalent under a 'thick felt tip pen' test. Similarly, we also increase m at fixed t to assess stability of the density estimate. In our experimentation we increase t in multiples of 10 never having required $t > 50$. We typically set $m = 50$ or $m = 100$. Requisite generation, even for larger t and m, is usually neither costly nor slow. Univariate plots are drawn by selecting 40 equally spaced points in the effective domain of the variable. We then evaluate the density estimate (of the form (1)) at these points and a spline-smoothed curve is drawn through these values. By effective domain, we mean the interval where, say, 99% of the mass lies. We occasionally require several passes to determine this domain and occasionally require more than 40 points to obtain a satisfying plot. Clearly, this plotting method could be refined, but such issues are not the main concern of this paper. In this regard, we also recommend a convenient check on calculations by using a simple trapezoidal integration on the collection of estimated density values to see how close the result is to 1.

Monitoring across iterations of summary statistics such as sample moments or quantiles has not proven effective. If we successively study differences or relative differences in such statistics, it is not easy to assess when these quantities are stable. Calculation of standard errors for such differences is difficult in part due to the unknown dependence structure between successive iterations (although comparison of iterations, say 10 apart, will mitigate the dependence issue). Sample reuse methods might be tried. However, rather than a comparison of a few summary statistics, we prefer an overall distributional comparison between iterates.

An attractive graphical tool, which we have found very useful, is the empirical quantile-quantile plot. With m constant across iterations, such plots are easily obtained. One only need order the generated samples at the iterations to be plotted. Using $m = 100$ (perhaps 200), under convergence the plotted points should generally be close to the 45° line. Creating such displays over increasing numbers of iterations enables us to distinguish inherent variation from lack of convergence. Such displays, with their inherent variation, accord with the aforementioned 'thick felt-tip pen' comparison. We offer illustrative displays in conjunction with the variance components example of Section 4.

We mention a final point which is pertinent to the assessment of convergence. The Markovian nature of the iterative process means that apparent convergence can be temporarily perturbed by an 'untypical' sample. This can upset our diagnostics for perhaps several subsequent iterations until stability is restored. An obvious way to 'robustify' the process is to work with pooled successive samples within some moving window. Again, we shall not focus on such refinements in this paper.

4. Variance component problems

Random effects models are very naturally modelled within the Bayesian framework. Nonetheless, calculation of the marginal posterior distributions of variance components and functions of variance components has proved a challenging technical problem. Box and Tiao (1973) report a substantial amount of detailed, sophisticated approximation work, both analytic and numerical. Skene (1983) considers purpose-built numerical techniques. The methods described by Smith *et al* (1985, 1987) require careful reparametrization dependent upon both the data and the choice of prior. In a similar spirit, Achcar and Smith (1988) discuss parameter transformations for successful implementation of Laplace's method (Tierney and Kadane, 1986). By comparison, the Gibbs sampling approach is remarkably simple.

We shall illustrate the approach with a model involving only two variance components, but it will be clear that the development for more complicated models is no more difficult. Consider, then, the variance components model defined by

$$Y_{ij} = \theta_i + e_{ij}, \quad i = 1,\ldots,K, \quad j = 1,\ldots,J, \tag{2}$$

where, assuming conditional independence throughout, $[\theta_i|\mu,\sigma^2_\theta] = N(\mu,\sigma^2_\theta)$ and $[e_{ij}|\sigma^2_e] = N(0,\sigma^2_e)$. Let $\theta = (\theta_1,\ldots,\theta_K)$, $Y = (Y_{11},\ldots,Y_{KJ})$ and assume that $\mu,\sigma^2_\theta,\sigma^2_e$ are independent with priors $[\mu] = N(\mu_0,\sigma^2_0)$, $[\sigma^2_\theta] = IG(a_1,b_1)$, $[\sigma^2_e] = IG(a_2,b_2)$ (see, for example, Hill, 1965), where IG denotes the inverse gamma distribution and $\mu_0,\sigma^2_0,a_1,b_1,a_2,b_2$ are assumed known. It is then straightforward to verify that the Gibbs sampler is specified by:

$$\begin{aligned}[\sigma^2_\theta|Y,\mu,\theta,\sigma^2_e] &= IG(a_1+\tfrac{1}{2}K,\ b_1+\tfrac{1}{2}\Sigma(\theta_i-\mu)^2)\\ [\sigma^2_e|Y,\mu,\theta,\sigma^2_\theta] &= IG(a_2+\tfrac{1}{2}KJ,\ b_2+\tfrac{1}{2}\Sigma\Sigma(Y_{ij}-\theta_i)^2)\end{aligned} \tag{3}$$

$$[\mu|Y,\theta,\sigma_\theta^2,\sigma_\varepsilon^2] = N\left(\frac{\sigma_\theta^2\mu_0+\sigma_\varepsilon^2\Sigma\theta_i}{\sigma_\theta^2+K\sigma_\varepsilon^2},\frac{\sigma_\theta^2\sigma_\varepsilon^2}{\sigma_\theta^2+K\sigma_\varepsilon^2}\right)$$

$$[\theta|Y,\mu,\sigma_\theta^2,\sigma_\varepsilon^2] = N\left(\frac{J\sigma_\theta^2}{J\sigma_\theta^2+\sigma_\varepsilon^2}\bar{Y}+\frac{\sigma_\varepsilon^2}{J\sigma_\theta^2+\sigma_\varepsilon^2}\mu 1,\frac{\sigma_\theta^2\sigma_\varepsilon^2}{J\sigma_\theta^2+\sigma_\varepsilon^2}I\right),$$

where $\bar{Y} = (\bar{Y}_1,\ldots,\bar{Y}_K)$, $\bar{Y}_i = \Sigma Y_{ij}/J$, 1 is a $K\times 1$ column vector of 1's and I is a $K\times K$ identity matrix. In particular, in (3), we can allow the a_i and/or b_i equal to zero, representing a range of conventional improper priors for σ_θ^2 and σ_ε^2.

Box and Tiao (1973, Section 5.1.3) introduce two data sets for which the model (2) is appropriate. The second and more difficult set is generated from random normal deviates with $\mu = 5$, $\sigma_\theta^2 = 4$, $\sigma_\varepsilon^2 = 16$. The resultant data, summarised in Table 1, are badly behaved, in that the standard (ANOVA based) unbiased estimate of σ_θ^2 is negative, rendering inference about σ_θ^2 difficult. We shall use this example to provide a challenging low dimensional test of the Gibbs sampler.

Table 1: Generated Data (Box and Tiao, 1973, p 247)

$K = 6$ $J = 5$ $\bar{Y}_{..} = 5.6656$

Batch	1	2	3	4	5	6
$\bar{Y}$	6.2268	4.6560	7.5212	5.6848	6.0796	3.8252
S^2	8.8650	25.4900	25.6359	7.0935	14.3590	8.2691

Source	SS	df	MS
Between Batches	41.6816	5	8.3363
Within Batches	358.7014	24	14.9459
Total	400.3830	29	

$\hat{\sigma}_\varepsilon^2 = 14.9459$ $\hat{\sigma}_\theta^2 = -1.3219$

For illustrative purposes, we provide a Bayesian analysis based on the prior specification $[\sigma_\epsilon^2] = IG(0,0)$, $[\mu] = N(0, 10^{12})$, together with either

$$I\text{: } [\sigma_\theta^2] = IG(0,0) \qquad \text{or} \qquad II\text{: } [\sigma_\theta^2] = IG(\tfrac{1}{2}, b_1).$$

Under I, we have the improper prior for $(\sigma_\theta^2, \sigma_\epsilon^2)$ suggested by Hill (1965), which is a naive two dimensional extension of the familiar non-informative prior for a variance. Under II, we have a proper weak independent inverse chi square prior for σ_θ^2 which, depending upon b_1, 'supports' or 'differs from' the data (see Skene, 1976 for further detailed discussion). The two priors for σ_θ^2 differ considerably. Under I $[\sigma_\theta^2]$ is one-tailed, giving strong weight to the assertion that σ_θ^2 is near zero. As this is weakly confirmed by the data, the marginal posterior (Figure 1a) reflects this prior. Under II, $[\sigma_\theta^2]$ is two-tailed, having mode at $2b_1/3$. Interestingly, experimentation with b_1 varying up to 6 leads to an outcome similar to that under I. For all such b_1, the prior is virtually reproduced as the posterior (see Figure 2a for the case $b_1 = 1$). The data provide very little information about σ_θ^2.

Our experience with Gibbs sampling in this context is very encouraging. Under both I and II (with $b_1 = 1$), the iterative approach had no difficulty with the extreme skewness in the resultant posterior of σ_θ^2. Overall convergence was achieved under I within at most 20 iterations using $m = 100$, and under II within at most 10 iterations using $m = 100$. We demonstrate this in Figure 1, which, for case I, compares density estimates after 20 and 40 iterations for σ_ϵ^2 and σ_θ^2. Figure 2 presents the corresponding curves for case II after 10 and 20 iterations. In Figures 3 and 4, we show several empirical $Q-Q$ plots for σ_ϵ^2 and σ_θ^2, respectively, under II, again using $m = 100$ points. We compare the first iteration with the second, the second with the third, the third with the fourth and at 'convergence'—the ninth with the tenth. Note that for σ_ϵ^2 we essentially have convergence at the third iteration.

The variance ratio, $\sigma_\theta^2/\sigma_\epsilon^2$, or perhaps the intra-class correlation coefficient $\sigma_\theta^2/(\sigma_\theta^2+\sigma_\epsilon^2)$ are often quantities of interest. Remarks at the end of Section 2 show that obtaining the marginal posterior distribution for such variables is easily accomplished. Figure 5 shows the estimated density for the variance ratio under both I and II obtained after 20 iterations with $m = 1000$, the untypically large value of m arising from the awkward shape of the posterior. A density estimator with normal kernels was used with window

width suggested in Silverman (1986 p.48).

5. Normal means problems

The comparison of means presumed from normal populations is arguably the most ubiquitous model in statistical inference, but issues such as unbalanced sampling and heterogeneity of variances have typically forced compromises in frequentist and empirical Bayes approaches. Historically, this has also been somewhat true in the purely Bayesian setting. Frequently, with regard to variance parameters the proper Bayesian procedure of marginalisation by integration has been replaced by point estimation, in order to reduce the dimensionality of numerical integrations needed to obtain marginal posterior distributions for mean parameters. Gibbs sampling provides a means of performing such integrations without having to make approximations. The Gibbs sampler was introduced in the context of problems of very high dimension (such as image reconstruction, expert systems, neural networks) and has been spectacularly successful in such contexts. Its encouraging performance in our investigations is therefore not surprising since even a large multiparameter Bayesian problem is of small dimension compared to typical image processing problems.

In this section, we consider the comparison of I population means which, in conjunction with distinct unknown population variances and an exchangeable prior, results in a $2I+2$ parameter problem. We show that the implementation of the Gibbs sampler is straightforward. The more general case where the population means are represented as linear functions of a set of explanatory variables can be handled similarly using by now familiar distribution theory given in, for example, Lindley and Smith (1972). Such an example, involving 66 parameters, appears in Section 6.

Often there are implicit order restrictions on the means to be compared. For instance, it may be known that the means are increasing as we traverse the populations from $i = 1$ up to I. If we incorporate this information into our prior specification using order statistics, the integrations required for marginalisation are typically beyond the capacity of current numerical and analytic approximation methodology. However, as we show below, the Gibbs sampler is still straightforwardly implemented since normal full

conditionals be simply replaced by truncated normals.

The requisite distribution theory assuming no order restrictions on the means is as follows. Assuming conditional independence throughout, let $[Y_{ij}|\theta_i,\sigma_i^2] = N(\theta_i,\sigma_i^2)$ $i = 1,\ldots,I$, $j = 1,\ldots,n_i$, $[\theta_i|\mu,\tau^2] = N(\mu,\tau^2)$, $[\sigma_i^2] = IG(a_1,b_1)$, $[\mu] = N(\mu_0,\sigma_0^2)$ and $[\tau^2] = IG(a_2,b_2)$, where IG denotes the Inverse Gamma distribution and $a_1,a_2,b_1,b_2,\mu_0,\sigma_0^2$ are assumed known (often chosen to represent conventional improper prior forms; see Section 4). By sufficiency, we confine attention to $\bar{Y}_i = \Sigma Y_{ij}/n_i$ and $S_i^2 = \Sigma(Y_{ij}-\bar{Y}_i)^2/(n_i-1)$. Letting $\theta = (\theta_1,\ldots,\theta_I)$, $\sigma^2 = (\sigma_1^2,\ldots,\sigma_I^2)$ and $Y = (\bar{Y}_1,\ldots,\bar{Y}_I, S_1^2,\ldots,S_I^2)$, we have, for given data Y, the following full conditional distributions.

$$[\theta|Y,\sigma^2,\mu,\tau^2] = N(\theta^*,D^*), \tag{4}$$

where

$$\theta_i^* = \frac{n_i\bar{Y}_i\tau^2+\mu\sigma_i^2}{n_i\tau^2+\sigma_i^2}$$

$$D_{ii}^* = \frac{\sigma_i^2\tau^2}{n_i\tau+\sigma_i^2}, \qquad D_{ij}^* = 0, \qquad i \neq j,$$

$$[\sigma^2|Y,\theta,\mu,\tau^2] = \prod_{i=1}^{I}[\sigma_i^2|\bar{Y}_i,S_i^2,\theta_i],$$

where

$$[\sigma_i^2|\bar{Y}_i,S_i^2,\theta_i] = IG(a_1+\tfrac{1}{2}n_i, b_1+\tfrac{1}{2}\sum_{j=1}^{n_i}(Y_{ij}-\theta_i)^2),$$

$$[\mu|Y,\theta,\sigma^2,\tau^2] = N\left(\frac{\tau^2\mu_0+\sigma_0^2\Sigma\theta_i}{\tau^2+I\sigma_0^2}, \frac{\tau^2\sigma_0^2}{\tau^2+I\sigma_0^2}\right),$$

and

$$[\tau^2|Y,\theta,\sigma^2,\mu] = IG(a_2+\tfrac{1}{2}I, b_2+\tfrac{1}{2}\Sigma(\theta_i-\mu)^2).$$

Suppose now that the means are known to be ordered, say, $\theta_1 < \theta_2 < \ldots < \theta_I$. If we assume as our prior

that the θ_i arise as order statistics from a sample of size I from $N(\mu, \tau^2)$, then it is straightforward to show that $[\theta_i | Y, \theta_j, j \neq i, \sigma^2, \mu, \tau^2]$ is now precisely the marginal normal distribution in (4), but restricted to the interval $[\theta_{i-1}, \theta_{i+1}]$ (where we adopt the convention $\theta_0 = -\infty$, $\theta_{I+1} = +\infty$) and so again is straightforwardly available for sampling. The full conditional distributions for σ^2, τ^2 and μ remain exactly as above.

In sampling from the truncated normal distribution, the rejection method (discarding ineligible observations sampled from the non-truncated distribution) will tend to be wasteful and slow, particularly if $\theta_{i+1} - \theta_{i-1}$ is small. To draw an observation from $N(c, d^2)$ restricted to (a, b) a convenient 'one-for-one' sampling method is the following (Devroye, 1986). Generate U, a random uniform $(0, 1)$ variate and calculate $Y = c + d\Phi^{-1}(p(U; a, b, c, d))$, where

$$p(U; a, b, c, d) = \Phi\left(\frac{a-c}{d}\right) + U\left(\Phi\left(\frac{b-c}{d}\right) - \Phi\left(\frac{a-c}{d}\right)\right)$$

with Φ denoting the standard normal cdf. It is straightforward to show that Y has the desired distribution. These ideas are easily extended to give a general account of Bayesian analysis for order restricted parameters, but details will be deferred to a subsequent paper.

In order to study the performance of Gibbs sampling in the above setting, we analysed generated data so as to be able to calibrate the results against the known situation. For the purpose of illustration, we created a rather unbalanced, extremely non-homogeneous, data set by setting $I = 5$ and for the ith population, $i = 1, \ldots, 5$ drawing $n_i = 2i + 4$ independent observations from $N(i, i^2)$. The simulated data is summarised in Table 2, and we note, in particular, the inversion of order of the sample means, $\bar{Y}_4$ and $\bar{Y}_5$.

Table 2: Summary of simulated data for Normal Means problem

Sample	1	2	3	4	5
n_i	6	8	10	12	14
$\bar{Y}_i$	0.3191	2.034	3.539	6.398	4.811
S_i^2	0.2356	2.471	5.761	8.758	19.670

For illustration, we specified priors $[\mu] = N(0, 10^5)$, $[\sigma_\epsilon^2] = IG(\frac{1}{2}, 1)$ and $[\tau^2] = IG(\frac{1}{2}, 1)$. For the Gibbs sampler, convergence was achieved within ten iterations for the unordered case using $m = 100$. The ordered case required at most twenty iterations, again using $m = 100$, except for $[\theta_4|Y]$ and $[\theta_5|Y]$ which both required $m = 1000$. Rather than graphically documenting the convergence in this case, we compare the unordered and ordered marginal posteriors. Let $[\theta_i|\hat{Y}]_u$ and $[\theta_i|\hat{Y}]_0$ denote, respectively, the unordered and ordered density estimates. In Figure 6a we consider, for example, θ_2 and see that $[\theta_2|\hat{Y}]_u$ and $[\theta_2|\hat{Y}]_0$ have roughly the same mode but that $[\theta_2|\hat{Y}]_0$ is less dispersed. Utilizing the order information results in a sharper inference. In Figure 6b, we consider both θ_4 and θ_5. As would be expected, given the sufficient statistics, $[\theta_5|\hat{Y}]_u$ lies to the left of $[\theta_4|\hat{Y}]_u$ and is very dispersed. Utilizing the order information places $[\theta_4|\hat{Y}]_0$ and $[\theta_5|\hat{Y}]_0$ in the proper stochastic order, pulls the modes in the correct direction and reduces dispersion.

6. An hierarchical model

Applications of hierarchical models of the kind introduced by Lindley and Smith (1972) abound in fields as diverse as educational testing (Rubin, 1981), cancer studies (DuMouchel and Harris, 1983) and biological growth curves (Strenio, Weisberg and Bryk, 1983). However, both Bayesian and empirical Bayesian methodologies for such models are typically forced to invoke a number of approximations, whose consequences are often unclear under the multiparameter likelihoods induced by the modelling. See, for example, Morris (1983), Racine-Poon (1985) and Racine-Poon and Smith (1989) for details of some approaches to implementing hierarchical model analysis. By contrast, a full implementation of the Bayesian approach is easily achieved using the Gibbs sampler, at least for the widely used normal linear hierarchical model structure.

For illustration, we focus on the following population growth problem. In a study conducted by the CIBA-GEIGY company, the weights of thirty young rats were measured weekly, for five weeks. The data are given in Table 3, with weight measurements available for all five weeks.

Table 3: Rat population growth data

Rat	x_{i1}	x_{i2}	x_{i3}	x_{i4}	x_{i5}	Rat	x_{i1}	x_{i2}	x_{i3}	x_{i4}	x_{i5}
1	151	199	246	283	320	16	160	207	248	288	324
2	145	199	249	293	354	17	142	187	234	280	316
3	147	214	263	312	328	18	156	203	243	283	317
4	155	200	237	272	297	19	157	212	259	307	336
5	135	188	230	280	323	20	152	203	246	286	321
6	159	210	252	298	331	21	154	205	253	298	334
7	141	189	231	275	305	22	139	190	225	267	302
8	159	201	248	297	338	23	146	191	229	272	302
9	177	236	285	340	376	24	157	211	250	285	323
10	134	182	220	260	296	25	132	185	237	286	331
11	160	208	261	313	352	26	160	207	257	303	345
12	143	188	220	273	314	27	169	216	261	295	333
13	154	200	244	289	325	28	157	205	248	289	316
14	171	221	270	326	358	29	137	180	219	258	291
15	163	216	242	281	312	30	153	200	244	286	324

$x_{i1} = 8$, $x_{i2} = 15$, $x_{i3} = 22$, $x_{i4} = 29$, $x_{i5} = 36$ days, $i = 1,\ldots,30$.

For the time period considered, it is reasonable to assume individual straight-line growth curves so that, under homoscedastic normal measurement errors,

$$Y_{ij} \sim N(\alpha_i + \beta_i x_{ij}, \sigma^2) \quad (i = 1,\ldots,k;\ j = 1,\ldots,n_i)$$

provides the full measurement model (with $k = 30$, $n_i = 5$, and x_{ij} denoting the age in days of the ith rat when measurement j was taken). The population structure is modelled as

$$\begin{pmatrix}\alpha_i \\ \beta_i\end{pmatrix} \sim N\left\{\begin{pmatrix}\alpha_0 \\ \beta_0\end{pmatrix}, \Sigma\right\}, \quad (i = 1,\ldots,k)$$

assuming conditional independence throughout. A full Bayesian analysis now requires the specification of a prior for σ^2, $\mu = (\alpha_0, \beta_0)^\top$ and Σ. Typical inferences of interest in such studies include marginal posteriors for the population prarameters α_0, β_0 and predictive intervals for individual future growth given the first week measurement. We shall see that these are easily obtained using the Gibbs sampler.

For the prior specification, we take

$$[\mu, \Sigma^{-1}, \sigma^2] = [\mu][\Sigma^{-1}][\sigma^2]$$

to have a normal-Wishart-inverse-gamma form,

$$[\mu] = N(\eta, C)$$

$$[\Sigma^{-1}] = W((\rho R)^{-1}, \rho)$$

$$[\sigma^2] = IG\left(\frac{\nu_0}{2}, \frac{\nu_0 \tau_0^2}{2}\right).$$

Rewriting the measurement model for the ith individual as $Y_i \sim N(X_i \theta_i, \sigma^2 I_{n_i})$ with $\theta_i = (\alpha_i, \beta_i)^\top$ and X_i denoting the appropriate design matrix, and defining

$$Y = (Y_1, \ldots, Y_k)^\top, \qquad \bar{\theta} = k^{-1} \sum_{i=1}^{k} \theta_i, \qquad n = \sum_{i=1}^{k} n_i,$$

$$D_i = \sigma^{-2} X_i^\top X_i + \Sigma^{-1}$$

$$V = (k\Sigma^{-1} + C^{-1})^{-1},$$

the Gibbs sampler for $\theta = (\theta_1, \ldots, \theta_k)$, Σ, σ^2 (a total of 66 parameters in the above example) is straightforwardly seen to be specified by the following conditional distributions:

$$[\theta_i | Y, \mu, \Sigma^{-1}, \sigma^2] = N\{D_i(\sigma^{-2} X_i^\top Y_i + \Sigma^{-1}\mu), D_i\} \quad (i = 1, \ldots, k) \tag{5}$$

$$[\mu | Y, \{\theta\}, \Sigma^{-1}, \sigma^2] = N\{V(k\Sigma^{-1}\bar{\theta} + C^{-1}\eta), V\}$$

$$[\Sigma^{-1} | Y, \{\theta\}, \mu, \sigma^2] = W\{[\Sigma(\theta_i - \mu)(\theta_i - \mu)^\top + \rho R]^{-1}, k + \rho\}$$

$$[\sigma^2|Y,\{\theta\},\mu,\Sigma^{-1}] = IG\left(\frac{n+\nu_0}{2}, \frac{1}{2}\{\sum_i (Y_i - X_i\theta_i)^\top (Y_i - X_i\theta_i) - \nu_0\tau_0\}\right).$$

For the analysis of the rat growth data given above, the prior specification was defined by:

$$C^{-1} = 0, \quad \nu_0 = 0, \quad \rho = 2, \quad R = \begin{pmatrix} 100 & 0 \\ 0 & 0.1 \end{pmatrix},$$

reflecting rather vague initial information relative to that to be provided by the data. Simulation from the Wishart distribution for the 2×2 matrix Σ^{-1} is easily accomplished using the algorithm of Odell and Feiveson (1966): with $G(.,.)$ denoting gamma distributions, draw independently from

$$[U_1] = G\left(\frac{\nu}{2}, \frac{1}{2}\right), \quad [U_2] = G\left(\frac{\nu-1}{2}, \frac{1}{2}\right), \quad [N] = N(0,1);$$

set

$$W = \begin{bmatrix} U_1 & N\sqrt{U_1} \\ N\sqrt{U_1} & U_2+N^2 \end{bmatrix};$$

then if $S^{-1} = (H^{\frac{1}{2}})^\top (H^{\frac{1}{2}})$,

$$\Sigma^{-1} = (H^{\frac{1}{2}})^\top W (H^{\frac{1}{2}}) \sim W(S^{-1}, \nu).$$

The iterative process can be conveniently monitored by observing $Q-Q$ plots for $\alpha_0, \beta_0, \sigma^2$ and the eigenvalues of Σ^{-1}. For the data set summarized in Table 3, convergence is achieved with about 35 cycles of $m = 50$ drawings.

As we remarked earlier, full Bayesian analysis of structured hierarchical models involving covariates has hitherto presented difficulties and a number of Bayes/empirical Bayes approximation methods have been proposed. Racine-Poon and Smith (1989) review a number of these and demonstrate, with a range of real and simulated data analyses, that the EM-type algorithm given by Racine-Poon (1985) seems to be the best of these proposed approximations. However, it can be seen from Figure 7, where we present the estimated posterior marginals for the population parameters, that, even with this fairly substantial data set of 30×5 observations, the EM-type approximation is not really an adequate substitute for the more refined numerical

approximation provided by the Gibbs sampler. (Here, the EM-based 'posterior density' is the normal conditional form (5) with the converged estimates from the Racine-Poon algorithm substituted for the conditioning parameters.)

To further underline the effectiveness of the Gibbs sampler, and the danger of point-estimation based approximations in hierarchical models, we reanalysed two subsets of the complete data set of 150 observations given in Table 3, chosen to present an increasing challenge to the algorithms. One subset consisted of 90 observations, obtained by omitting the final data point from rats 6–10, the final two data points from rats 11–20, the final three from rats 21–25 and the final four from rats 26–30. The other subset consisted of 75 observations, obtained from the 90 by retaining only one of the observations for each of the rats 16–30. Convergence for the first subset required about 50 iterations of $m = 50$; convergence for the second about 65 iterations of $m = 50$.

Figure 8 summarizes the marginal posteriors for the growth rate parameter obtained for the two data subsets from the Gibbs and EM-type algorithms, respectively. It can be seen that while the EM approximation is perhaps tolerable for the full data set (Figure 7), it is very poor for the smaller data sets.

Consider now the data set of 90 observations and suppose that the problem of interest is the prediction of the future growth pattern for one of the rats for which there is currently just the first observation available ($i = 26,\ldots,30$). Specifically, suppose we consider predicting Y_{ij}, $j = 2,3,4,5$, corresponding to $x_{i2} = 15$, $x_{i3} = 22$, $x_{i4} = 29$, $x_{i5} = 36$ days. Then, formally,

$$[Y_{ij}|Y] = \int [Y_{ij}|\theta_i,\sigma^2]*[\theta_i,\sigma^2|Y],$$

where

$$[Y_{ij}|\theta_i,\sigma^2] = N(\alpha_i+\beta_i x_{ij},\sigma^2). \tag{6}$$

An estimate of $[Y_{ij}|Y]$ of the form (1) is thus easily obtained by averaging $[Y_{ij}|\theta_i,\sigma^2]$ over pairs of (θ_i,σ^2) obtained at the final cycle of the Gibbs sampler. Figure 9 shows, for $i = 26$, bands drawn through the individual 95% predictive interval limits calculated at days 15, 22, 29 and 36, together with the subsequently observed values at those points.

Alternatively, we could view the omitted or, in general, as yet unobserved data points as missing data. The Gibbs sampler could then be implemented treating such Y_{ij} as unobservable (in addition to the model parameters) since the required full conditional distributions have the form (6).

7. Missing data in a cross-over trial

The balanced two-period cross-over design is widely used; for example, in the pharmaceutical industry for bioequivalence studies involving a standard and a new drug formulation, A and B, say (Racine *et al*, 1986; Racine-Poon *et al*, 1987). Assuming n subjects, the standard random effects model for a two-period cross-over is given by:

$$Y_{i(jk)} = \mu + (-1)^{(j-1)}\left(\frac{\varphi}{2}\right) + (-1)^{(k-1)}\left(\frac{\pi}{2}\right) + \delta_i + \varepsilon_{i(jk)},$$

where

$Y_{i(jk)}$ = response to the ith subject ($i = 1,\ldots,n$) receiving the jth formulation ($j = 1,2$) in the kth period ($k = 1,2$);

μ = overall mean level of response;

φ = difference in formulation effects;

π = difference in period effects;

δ_i = random effect of ith subject;

$\varepsilon_{i(jk)}$ = measurement error.

The δ_i, $\varepsilon_{i(jk)}$ are assuming independent, for all i, j, k, with $\varepsilon_{i(jk)} \sim N(0, \sigma_1^2)$, $\delta_i \sim N(0, \sigma_2^2)$.

Suppose now that subjects $i = 1,\ldots,M$ have data missing from one of the two periods; subjects $i = M+1,\ldots,n$ have complete data. We shall write

$$Y_i = \begin{pmatrix} U_i \\ V_i \end{pmatrix}, \quad X_i = \begin{pmatrix} X_{iu} \\ X_{iv} \end{pmatrix} \quad (i = 1,\ldots,M),$$

where the 'observations' within subject i have been labelled such that V_i is the observed data, U_i is missing, and X_i defines the corresponding design matrix. For subjects $i = M+1,\ldots,n$, $Y_i = V_i$ simply denotes all the observed data. We write $U = (U_1,\ldots,U_M)^\top$, $V = (V_1,\ldots,V_n)^\top$. Then, conditional on

$$\theta = (\mu, \phi, \pi)^\top, \qquad \Sigma = \begin{pmatrix} \sigma_{12}^2 & \sigma_2^2 \\ \sigma_2^2 & \sigma_{12}^2 \end{pmatrix},$$

where $\sigma_{12}^2 = \sigma_1^2 + \sigma_2^2$, we have

$$Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} \sim N\left\{ \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix} \theta, S \right\} \equiv N(X\theta, S),$$

where

$$S = \begin{bmatrix} \Sigma & & 0 \\ & \Sigma & \\ & \ddots & \\ 0 & & \Sigma \end{bmatrix}.$$

Here, Y is $2n\times 1$, X is $2n\times 3$ and S is $2n\times 2n$.

It is convenient to work with $\theta, \sigma_1^2, \sigma_3^2$, where $\sigma_3^2 = \sigma_1^2 + 2\sigma_2^2$, and to note that, if we define

$Y_{i.}^{+}$ = average response of the ith subject

$$= \begin{cases} \frac{1}{2}(Y_{i(11)} + Y_{i(22)}) \\ \frac{1}{2}(Y_{i(21)} + Y_{i(12)}) \end{cases} \text{ for } \begin{matrix} AB \\ BA \end{matrix} \text{ sequence}$$

$Y_{i.}^{-}$ = difference of the two responses for the ith subject

$$= \begin{cases} \frac{1}{2}(Y_{i(11)} - Y_{i(22)}) \\ \frac{1}{2}(Y_{i(21)} - Y_{i(12)}) \end{cases} \text{ for } \begin{matrix} AB \\ BA \end{matrix} \text{ sequence,}$$

then:

for the AB sequence,

$$\begin{pmatrix} Y_{i\cdot}^{-} \\ Y_{i\cdot}^{+} \end{pmatrix} \sim N\left\{ \begin{pmatrix} \frac{\phi+\pi}{2} \\ \mu \end{pmatrix}, \frac{1}{2}\begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_3^2 \end{pmatrix} \right\};$$

for the BA sequence,

$$\begin{pmatrix} Y_{i\cdot}^{-} \\ Y_{i\cdot}^{+} \end{pmatrix} \sim N\left\{ \begin{pmatrix} \frac{\pi-\phi}{2} \\ \mu \end{pmatrix}, \frac{1}{2}\begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_3^2 \end{pmatrix} \right\}.$$

If we now make the prior specification

$$[\theta, \sigma_1^2, \sigma_3^2] = N(\eta, C) IG\left(\frac{\nu_1}{2}, \frac{\nu_1\tau_1}{2}\right) IG\left(\frac{\nu_3}{2}, \frac{\nu_3\tau_3}{2}\right) I_{(\sigma_1^2 \le \sigma_3^2)},$$

it can be seen that

$$[\sigma_1^2, \sigma_3^2 | U, V, \theta] \propto (\sigma_1^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma_1^2} SS_1\right\} (\sigma_1^2)^{-\left(\frac{\nu_1}{2}+1\right)} \exp\left\{-\frac{\nu_1\tau_1}{2\sigma_1^2}\right\}$$

$$\times (\sigma_3^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma_3^2} SS_3\right\} (\sigma_3^2)^{-\left(\frac{\nu_3}{2}+1\right)} \exp\left\{-\frac{\nu_3\tau_3}{2\sigma_3^2}\right\} I_{(\sigma_1^2 \le \sigma_3^2)},$$

where

$$SS_1 = 2 \sum_{AB\ seq.} \left[Y_{i\cdot}^{-} - \left(\frac{\phi+\pi}{2}\right)\right]^2 + 2 \sum_{BA\ seq.} \left[Y_{i\cdot}^{-} - \left(\frac{\pi-\phi}{2}\right)\right]^2,$$

$$SS_3 = 2 \sum_{i=1}^{n} (Y_{i\cdot}^{+} - \mu)^2.$$

It follows that a Gibbs sampler for σ_1^2, σ_3^2, θ and U is specified by:

$$[\sigma_1^2 | U, V, \theta, \sigma_3^2] = IG\left(\frac{n+\nu_1}{2}, \frac{SS_1 + \nu_1\tau_1}{2}\right) I_{(\sigma_1^2 \le \sigma_3^2)}$$

$$[\sigma_3^2 | U, V, \theta, \sigma_1^2] = IG\left(\frac{n+\nu_3}{2}, \frac{SS_3 + \nu_3\tau_3}{2}\right) I_{(\sigma_1^2 \le \sigma_3^2)}$$

$$[\theta|U,V,\sigma_1^2,\sigma_2^2] = N\{D(X^TS^{-1}Y+C^{-1}\eta),D\},$$

with

$$X^TS^{-1}Y = \sum_{i=1}^{n} X_i^T\Sigma^{-1}Y_i,$$

$$X^TS^{-1}X = \sum_{i=1}^{n} X_i^T\Sigma^{-1}X_i,$$

$$D = X^TS^{-1}X+C^{-1},$$

$$[U|V,\theta,\sigma_1^2,\sigma_3^2] = N\left\{X_u\theta+\frac{\sigma_2^2}{\sigma_{12}^2}(V-X_w\theta),\ \sigma_{12}^2\left[1-\left(\frac{\sigma_2^2}{\sigma_{12}^2}\right)^2\right]I_M\right\},$$

with

$$X_u = \begin{pmatrix} X_{1u} \\ \vdots \\ X_{Mu} \end{pmatrix}, \qquad U = \begin{pmatrix} U_1 \\ \vdots \\ U_M \end{pmatrix},$$

$$X_w = \begin{pmatrix} X_{1v} \\ \vdots \\ X_{Mv} \end{pmatrix}, \qquad W = \begin{pmatrix} V_1 \\ \vdots \\ V_M \end{pmatrix}.$$

Table 4 summarizes data from a (complete data) trial conducted with $n = 10$ subjects, in which responses are treated as missing from subject 1 in period 1, subject 3 in period 2 and subject 6 in period 2.

Table 4: Data from a two-period cross-over trial

Subject	*Sequence*	*Period 1*	*Period 2*
1	AB	1.40	1.65
2	AB	1.64	1.57
3	BA	1.44	1.58
4	BA	1.36	1.68
5	BA	1.65	1.69
6	AB	1.08	1.31
7	AB	1.09	1.43
8	AB	1.25	1.44
9	BA	1.25	1.39
10	BA	1.30	1.52

A = new tablet, B = standard tablet; formulations of Carbamazepine.

Data are observations of the logarithms of 1.52 maxima of concentration-time curves: see Maas *et al* (1987) for background and further details.

Convergence was achieved within 30 iterations of $m = 50$. Figure 10a shows the marginal posterior density for σ_1^2; Figure 10b shows the marginal posterior density for ϕ, the treatment effect. The Gibbs sampler also automatically provides 'predictive' densities for the missing responses; these are shown in Figure 11 and their locations may be compared with the actual missing values. Finally, the ease with which the Gibbs sampler permits analysis of this cross-over model enables informative sensitivity studies to be performed. In particular, we can easily study the difference between the treatment posterior density based on the complete data, the data omitting the assumed missing values, and the data based on just the seven subjects for whom full data was assumed. The resulting posteriors are shown in Figure 12 and reveal a typical finding in such trials. Namely: that if there is missing (at random) data from a subject we might just

as well ignore the subject altogether; also, that the loss of 30% of subjects in a small trial results in substantially increased inferential uncertainty.

8. Summary discussion

The range of normal data problems considered above as illustrations of the ease with which numerical Bayesian inferences can be obtained via Gibbs sampling, include the following aspects:

awkward posterior distributions, otherwise requiring subtle and sophisticated numerical or analytic approximation techniques (Sections 4 and 5);

further distributional complexity introduced by order constraints on model parameters (Section 5);

dimensionality problems, typically putting out of reach the implementation of other sophisticated approximation techniques (Section 6);

messy and intractable distribution theory arising from missing data in designed experiments (Section 7);

general functions of model parameters, including so-called Fieller-Creasy problems (Section 4);

awkward predictive inference (Section 6).

In all these situations, we have seen that the Gibbs sampler approach is straightforward to specify distributionally, trivial to implement computationally and with output readily translated into required inference summaries.

The potential of the methodology is enormous, rendering straightforward the analysis of a number of problems hitherto regarded as intractable from a Bayesian perspective. Work is in progress in extending the range of implementation. First, by developing, where necessary, purpose-built efficient random variate generators for conditional distribution forms arising in particular classes of applications; secondly, by facilitating the reporting of bivariate and conditional inference summaries, in addition to univariate marginal

curves. We plan to report shortly on various of these extensions.

Acknowledgements

This research was partly supported by the UK Science and Engineering Research Council, and the US Office of Naval Research.

References

Achcar J A and Smith A F M (1989). Aspects of reparametrization in approximate Bayesian inference. In *Essays in Honor of George A Barnard*, J Hodges editor; North-Holland.

Besag J (1974). Spatial interaction and the statistical analysis of lattice systems (with discussion). *Journal of the Royal Statistical Society B*, 36, 192–326.

Box G E P and Tiao G C (1973). *Bayesian inference in statistical analysis*. Addison-Wesley, Reading, MA.

Devroye L (1986). *Non-uniform random variate generation*. Springer-Verlag, NY.

DuMouchel W H and Harris J E (1983). Bayes methods for combining the results of cancer studies in humans and other species (with Discussion). *Journal of the American Statistical Association*, 78, 293–315.

Gelfand A E and Smith A F M, *Sampling based approaches to calculating marginal densities*, Department of Mathematics Technical Report, University of Nottingham.

Geman S and Geman D (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, 721–741.

Geweke J (1988). Antithetic acceleration of Monte Carlo integration in Bayesian inference. *Journal of Econometrics*, 38, 73–90.

Hill B M (1965). Inference about variance components in the one-way model. *Journal of the American Statistical Association*, 60, 806–825.

Lindley D V and Smith A F M (1972). Bayes estimates for the linear model (with Discussion). *Journal of the Royal Statistical Society, Series B*, 34, 1–41.

Maas B, Garnett W R, Pollock I M and Carnstock T J (1987). A comparative bioavailability study of Carbamazepine tablets and the chewable formulation. *Therapeutic Drug Monitoring*, Vol 9, 28–33.

Morris C (1983). Parametric empirical Bayes inference: theory and applications. *Journal of the American Statistical Association*, 78, 47–59.

Naylor J C and Smith A F M (1982). Applications of a method for the efficient computation of posterior distributions. *Applied Statistics*, 31, 214–225.

Naylor J C and Smith A F M (1988). Econometric illustrations of novel numerical integration strategies for Bayesian inference. *Journal of Econometrics*, 38, 103–126.

Odell P L and Feiveson A H (1966). A numerical procedure to generate a sample covariance matrix. *Journal of the American Statistical Association*, 61, 198–203.

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

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, Grieve A P, Flühler H and Smith A F M (1987). A two-stage procedure for bioequivalence studies. *Biometrics*, 43, 842–856.

Racine-Poon A and Smith A F M (1989). Population Models. To appear in *Statistical methodology in the Pharmaceutical Sciences* (Ed D Berry). Dekker.

Rubin D B (1981). Estimation in parallel randomized experiments. *Journal of Educational Statistics*, 6, 377–401.

Shaw J E H (1988). A quasirandom approach to integration in Bayesian statistics. *Annals of Statistics*, 16, 895–914.

Silverman B W (1986). *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, London.

Skene A M (1976). *Analytical methods for a balanced linear model*. Unpublished PhD thesis, Department of Statistics, University College, London.

Skene A M (1983). Computing marginal distributions for the dispersion parameters of analysis of variance models. *The Statistician*, 32, 99–108.

Smith A F M, Skene A M, Shaw J E H, Naylor J C and Dransfield M (1985). The implementation of the Bayesian paradigm. *Communications in Statistics, Theory and Methods*, 14, 1079–1102.

Smith A F M, Skene A M, Shaw J E H, Naylor J C (1987). Progress with numerical and graphical methods for Bayesian statistics. *Statistician*, 36, 75–82.

Strenio J F, Weisberg H I and Bryk A S (1983). Empirical Bayes estimation of individual growth-curve parameters and their relationship to covariates. *Biometrics*, 39, 71–86.

Tierney L and Kadane J B (1986). Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association*, 81, 82–86.

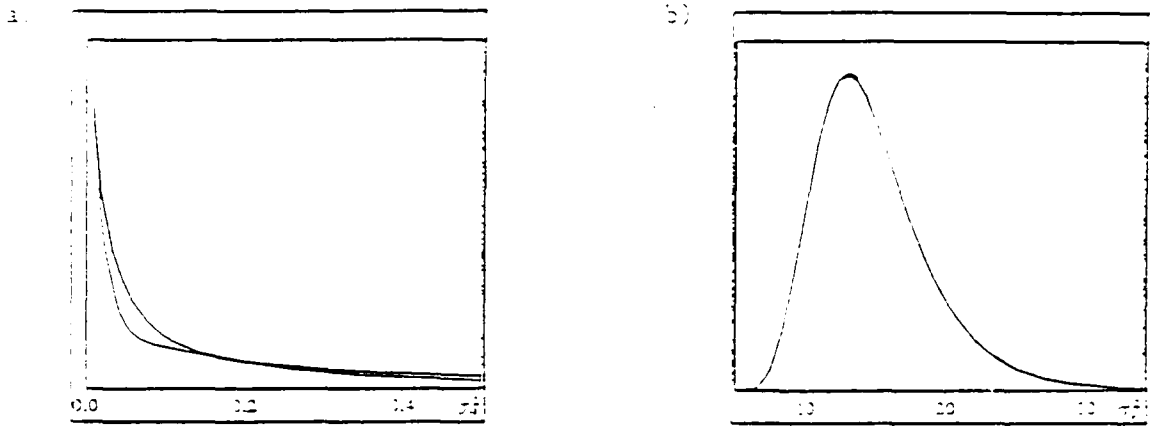


Figure 1: Convergence of Estimated Densities of Variance Components Under Prior Specification I

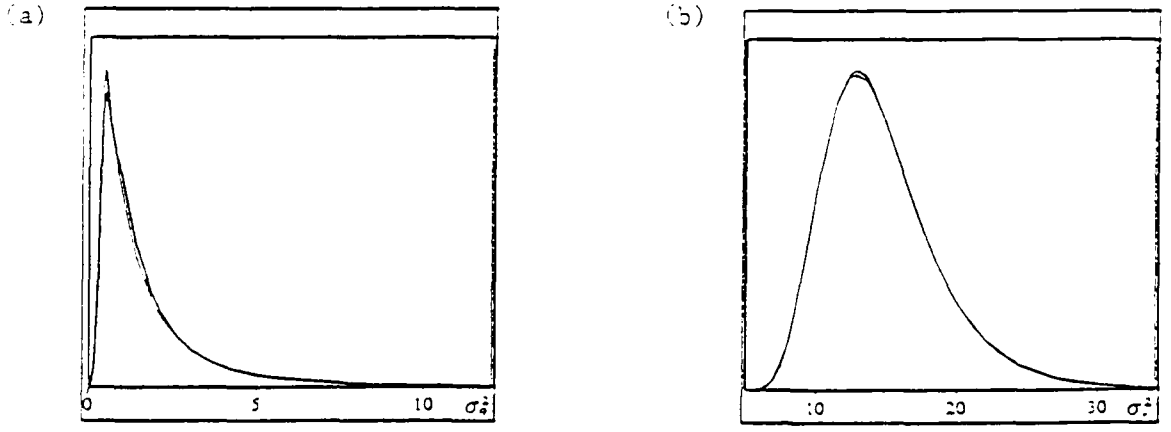


Figure 2: Convergence of Estimated Densities of Variance Components Under Prior Specification II

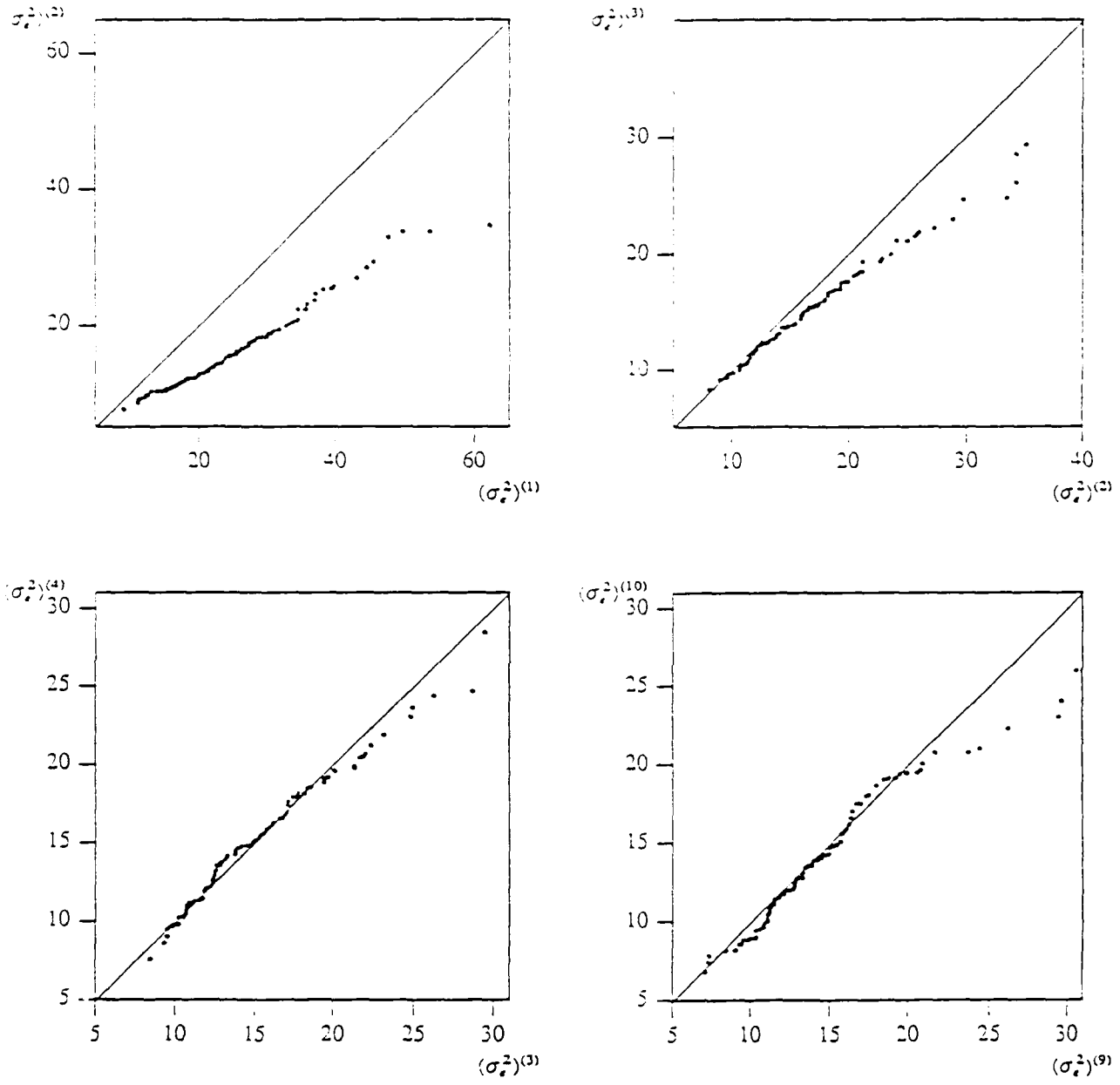


Figure 3: Empirical Q-Q plots for σ_e^2 (Prior Specification II)

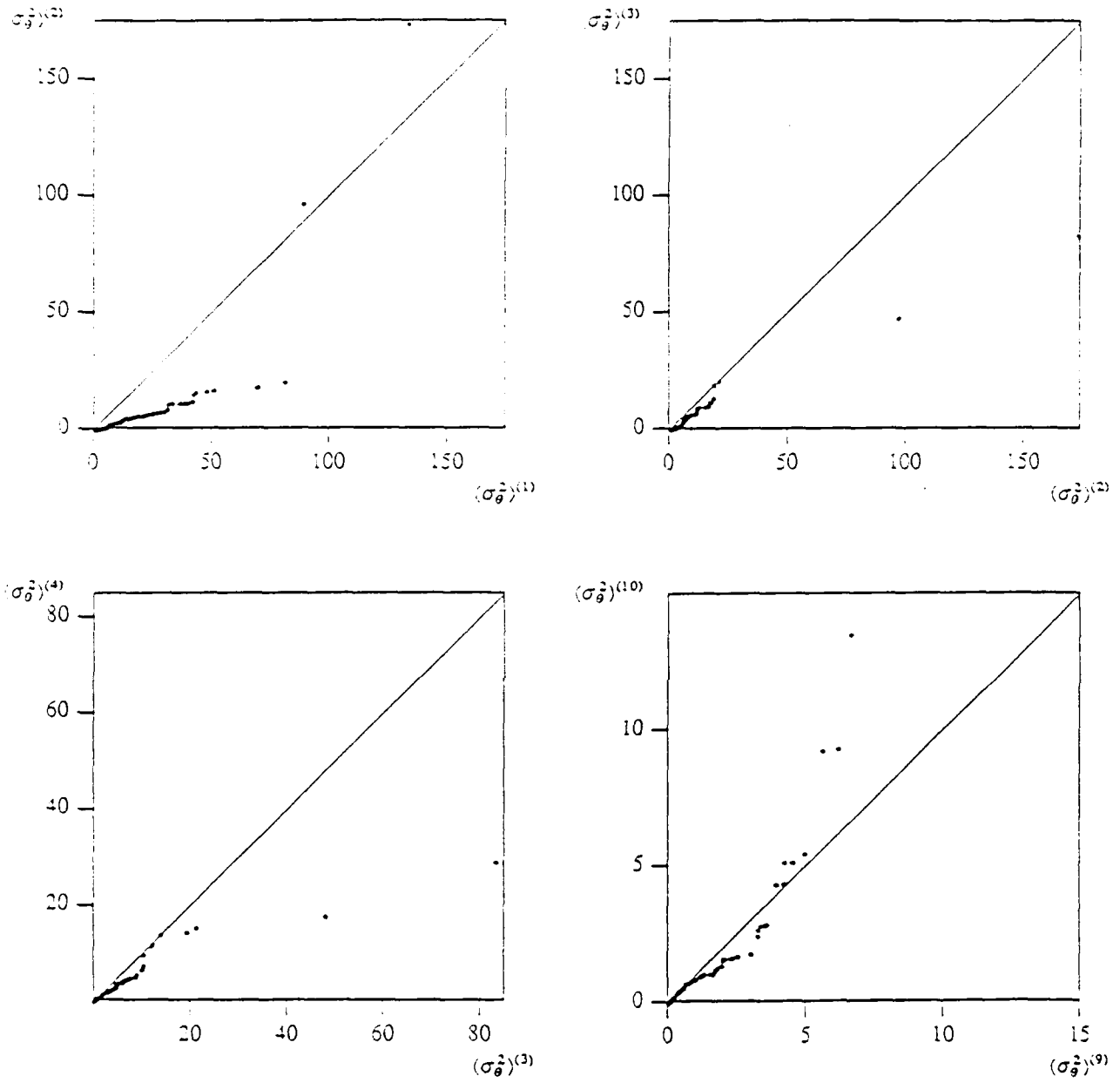


Figure 4: Empirical Q-Q plots for σ_θ^2 (prior specification II)

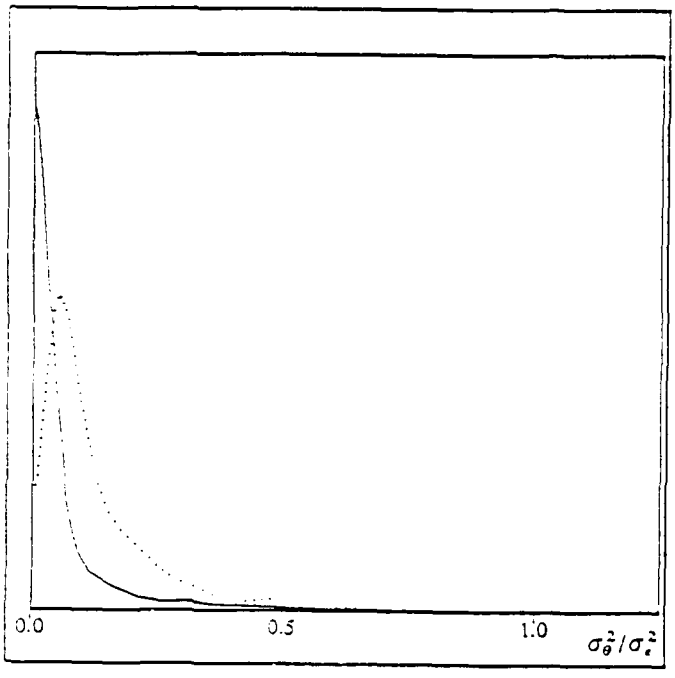
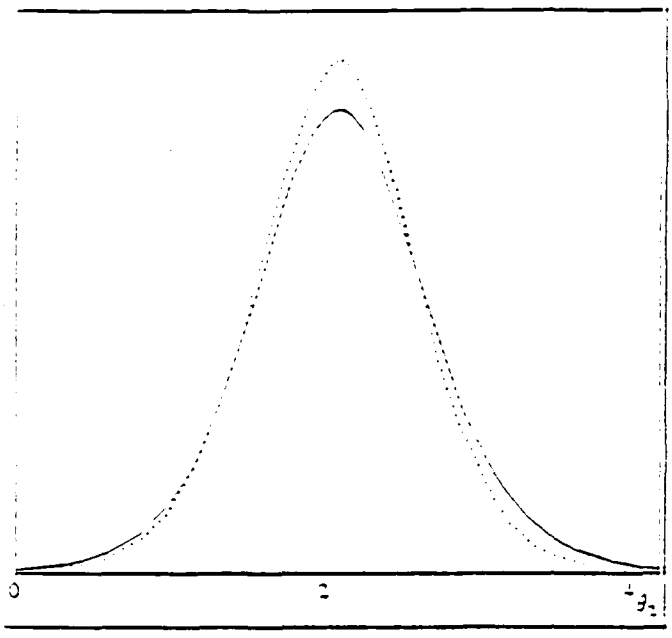


Figure 5: Estimated Density of the variance ratio $\sigma_\theta^2/\sigma_e^2$ for the variance components problem

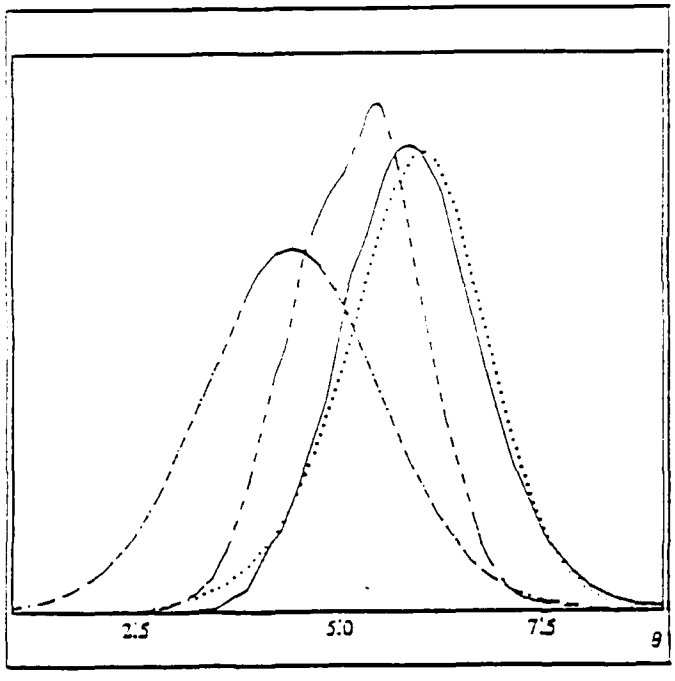
—— prior specification I, ···· prior specification II

(a)



—— is $[\theta_2|Y]_u$ ···· is $[\theta_2|Y]_o$

(b)



···· is $[\theta_4|Y]_u$ ·-·- is $[\theta_5|Y]_u$

--- is $[\theta_4|Y]_o$ —— is $[\theta_5|Y]_o$

Figure 6: Comparison of estimated densities of means, unordered and ordered cases

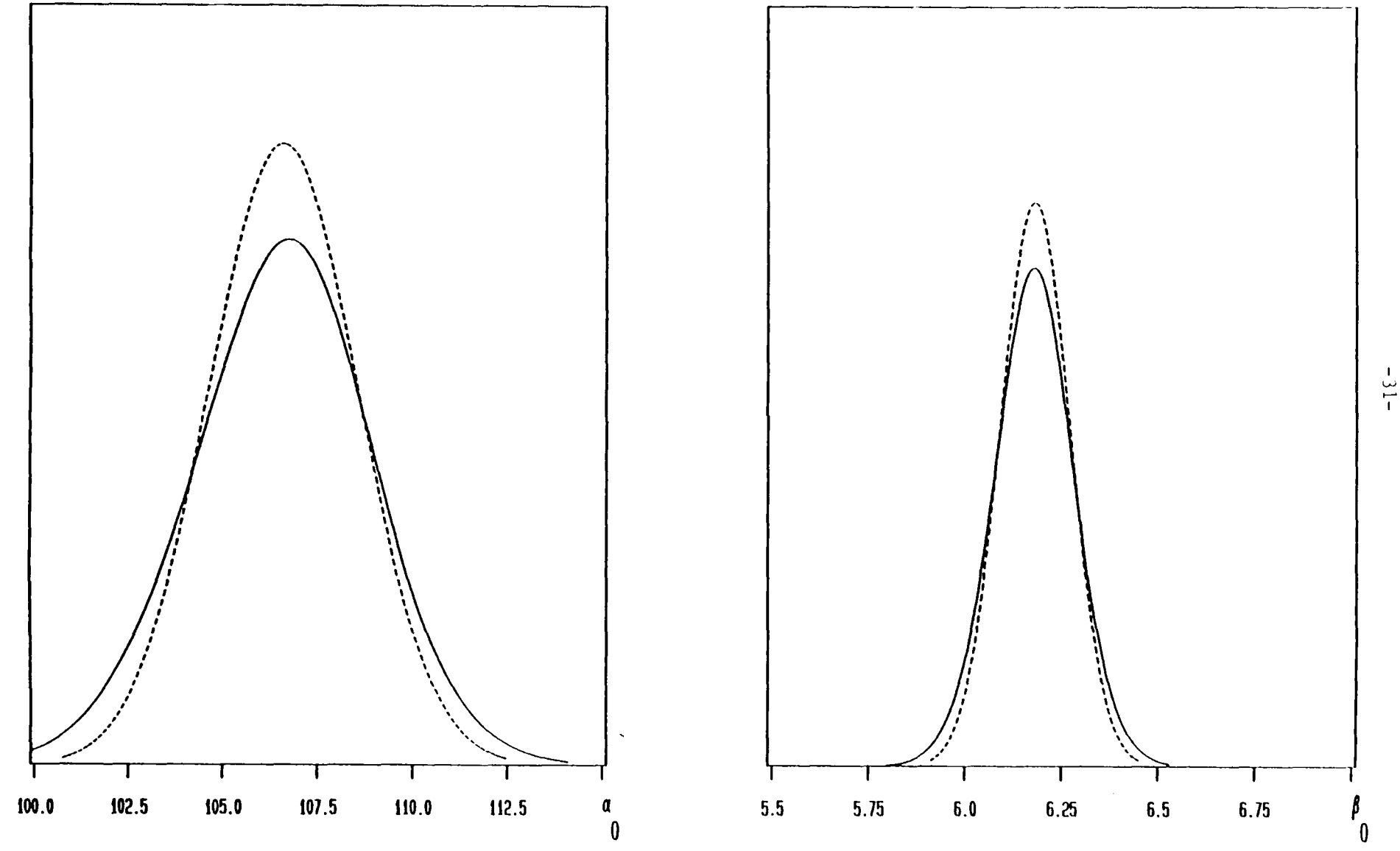


Figure 7: Estimated densities for population initial weight and growth rate for 150 observation case (__ is Gibbs sampler, – – – is EM)

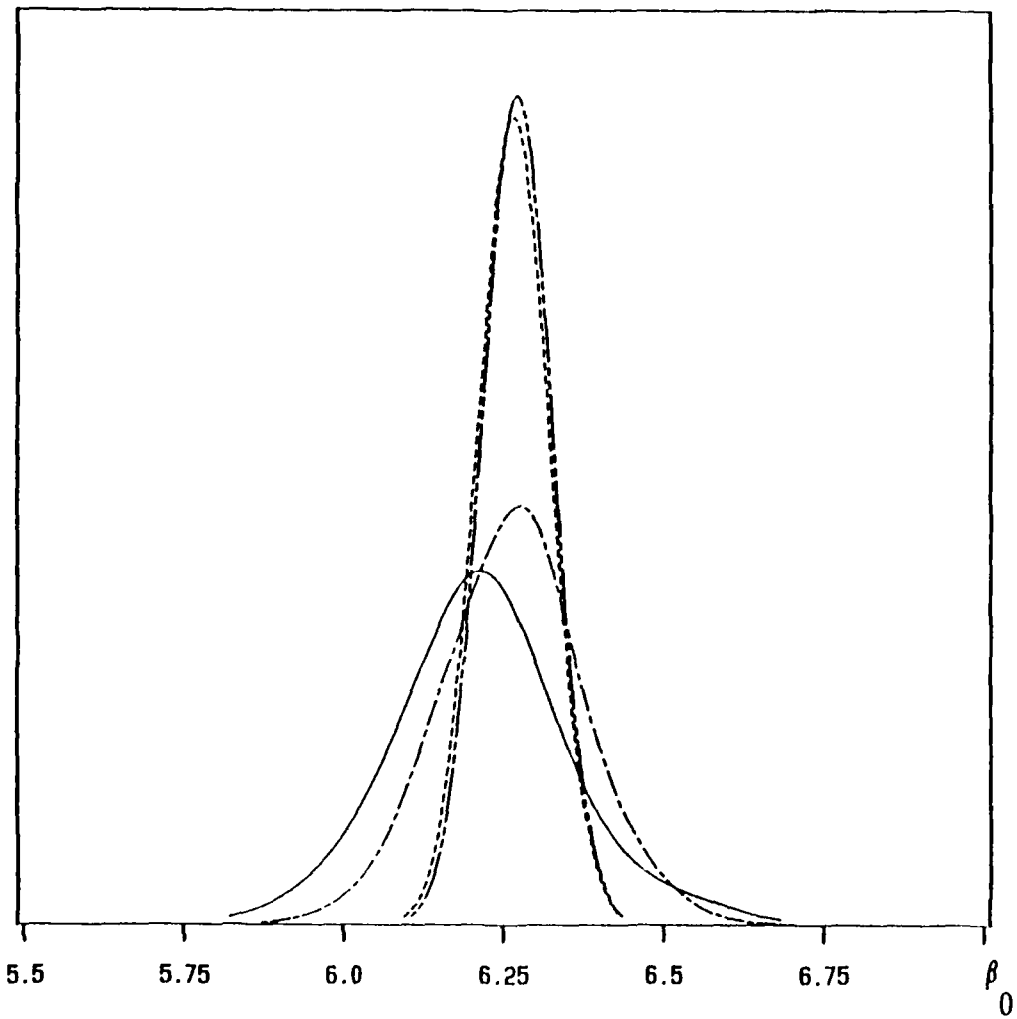


Figure 8: Estimated densities for population growth rate
for 90 observation case (_.._ is Gibbs sampler, _.. is EM)
for 75 observation case (is Gibbs sampler, – – – is EM)

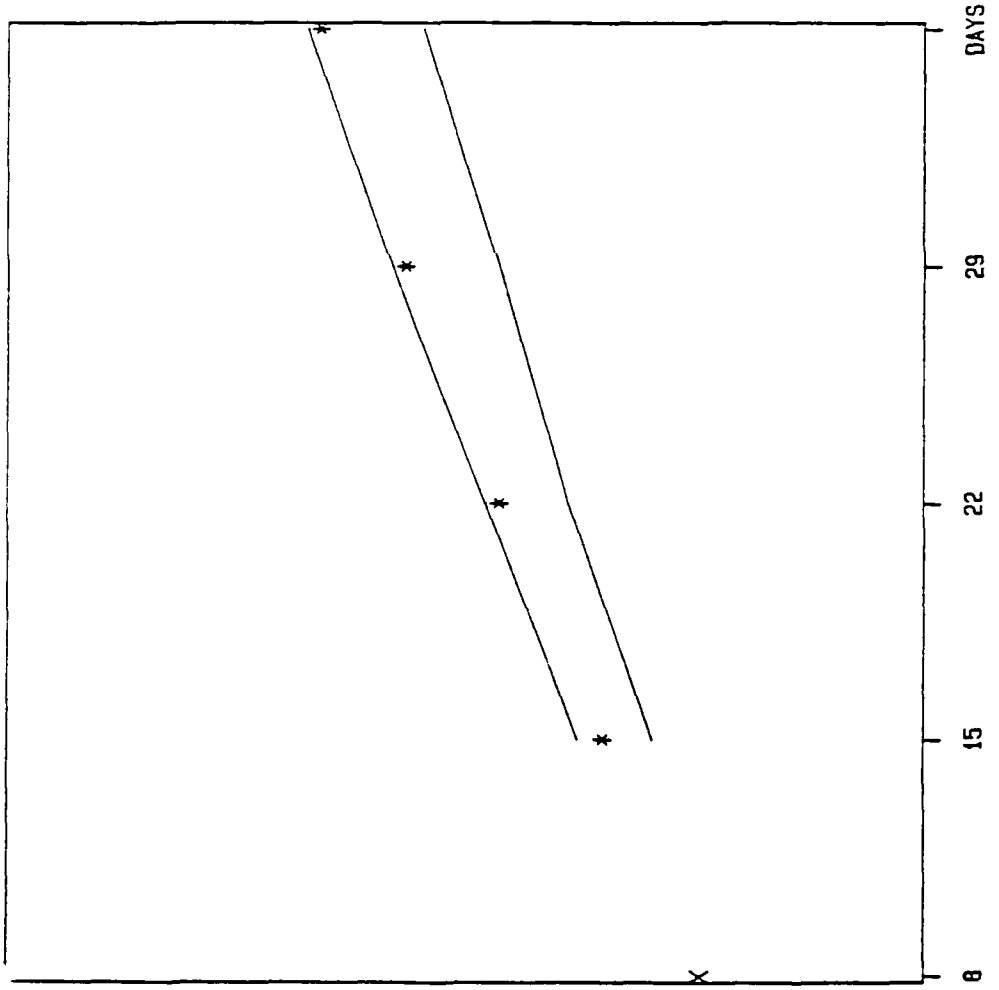
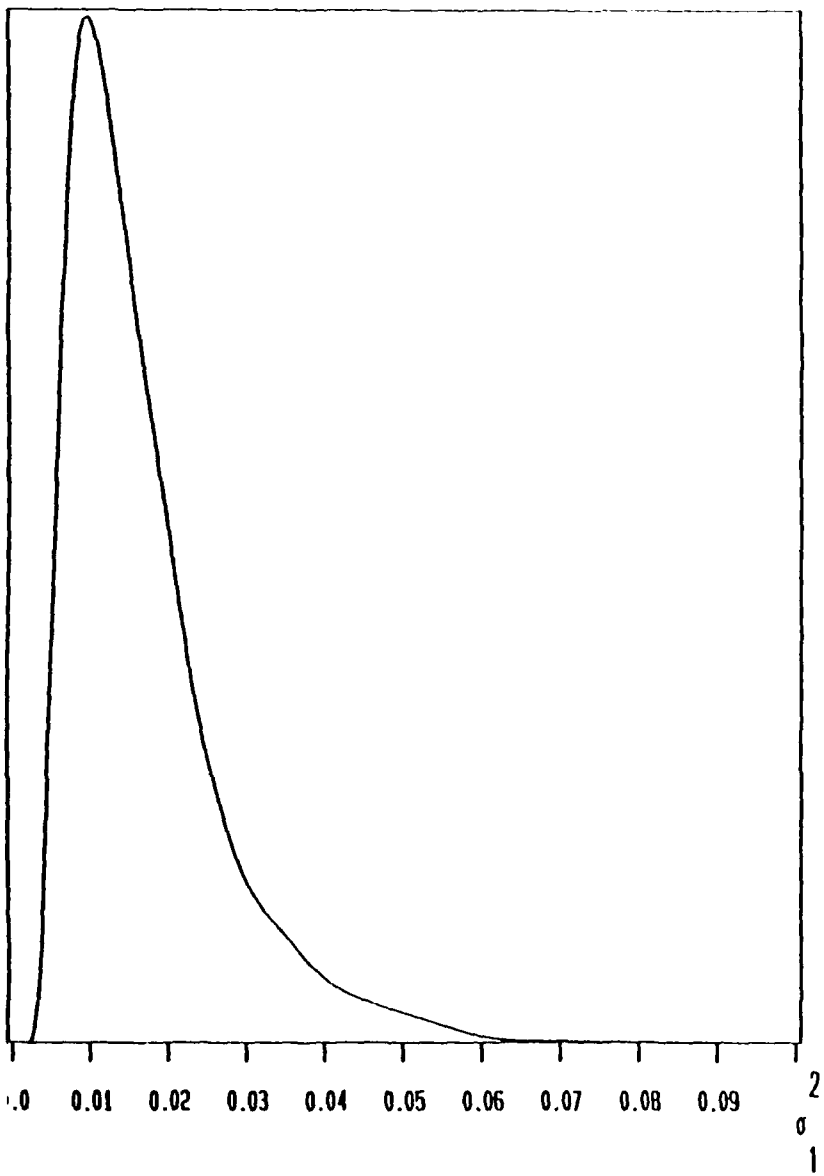


Figure 9: Estimated 95% predictive intervals for furture observations (*) given the first observation (x) of rat 26 (90 observation case).



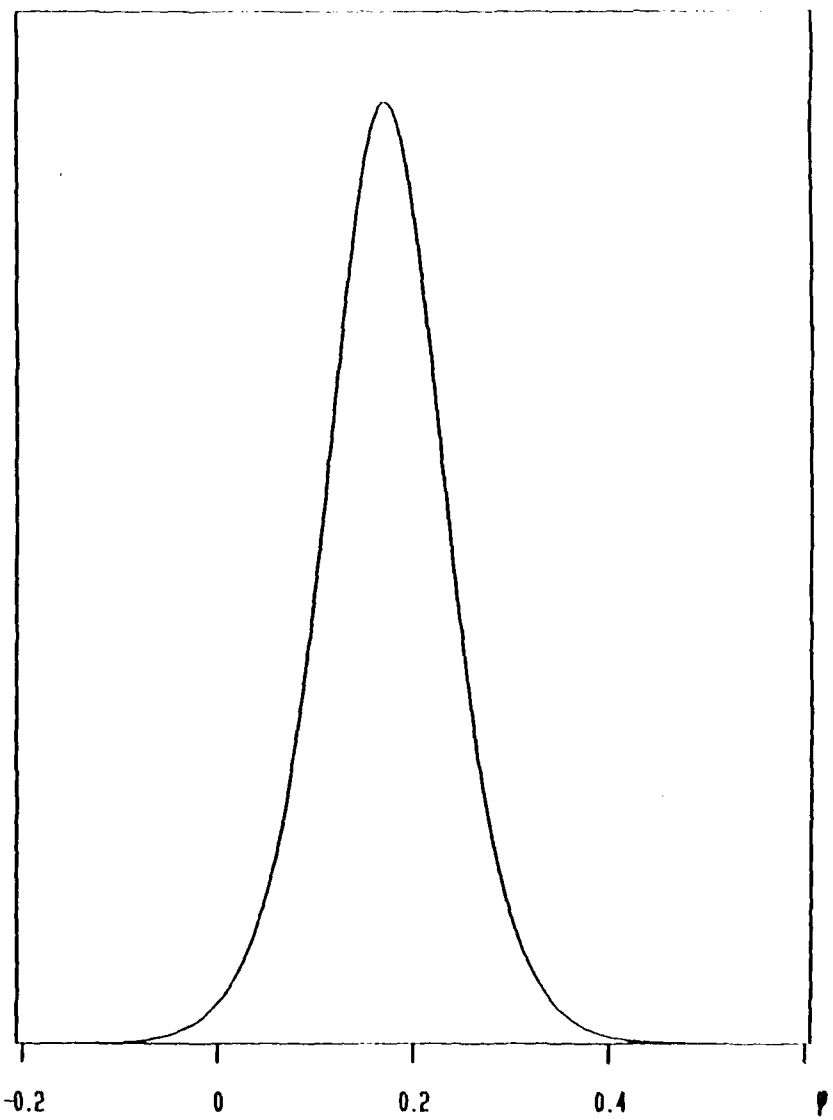


Figure 10: Estimated densities for σ_1^2 and φ in the cross-over trial.

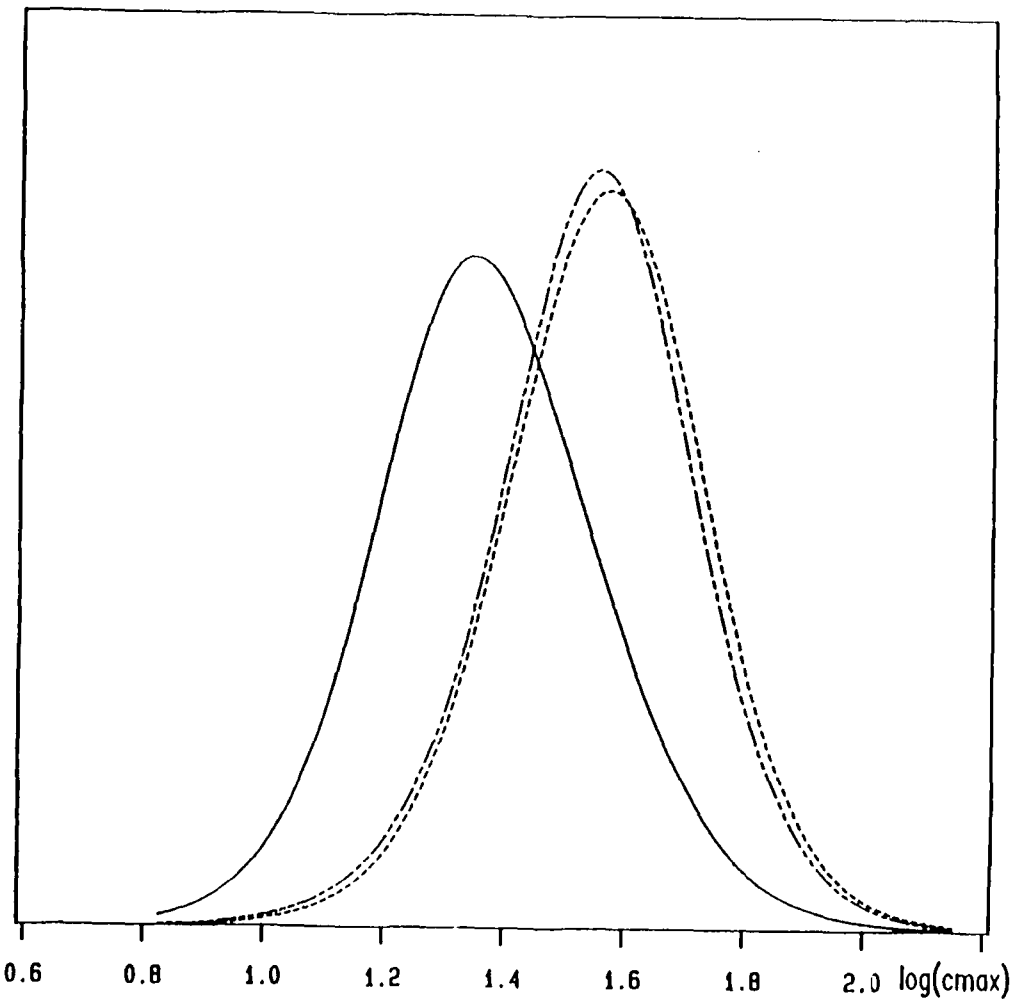


Figure 11: Estimated predictive densitites for the missing data values in cross-over trial
___= subject 1, actual value=1.40; ...= subject 3, actual value=1.44; _._._ = subject 6, actual value=1.31

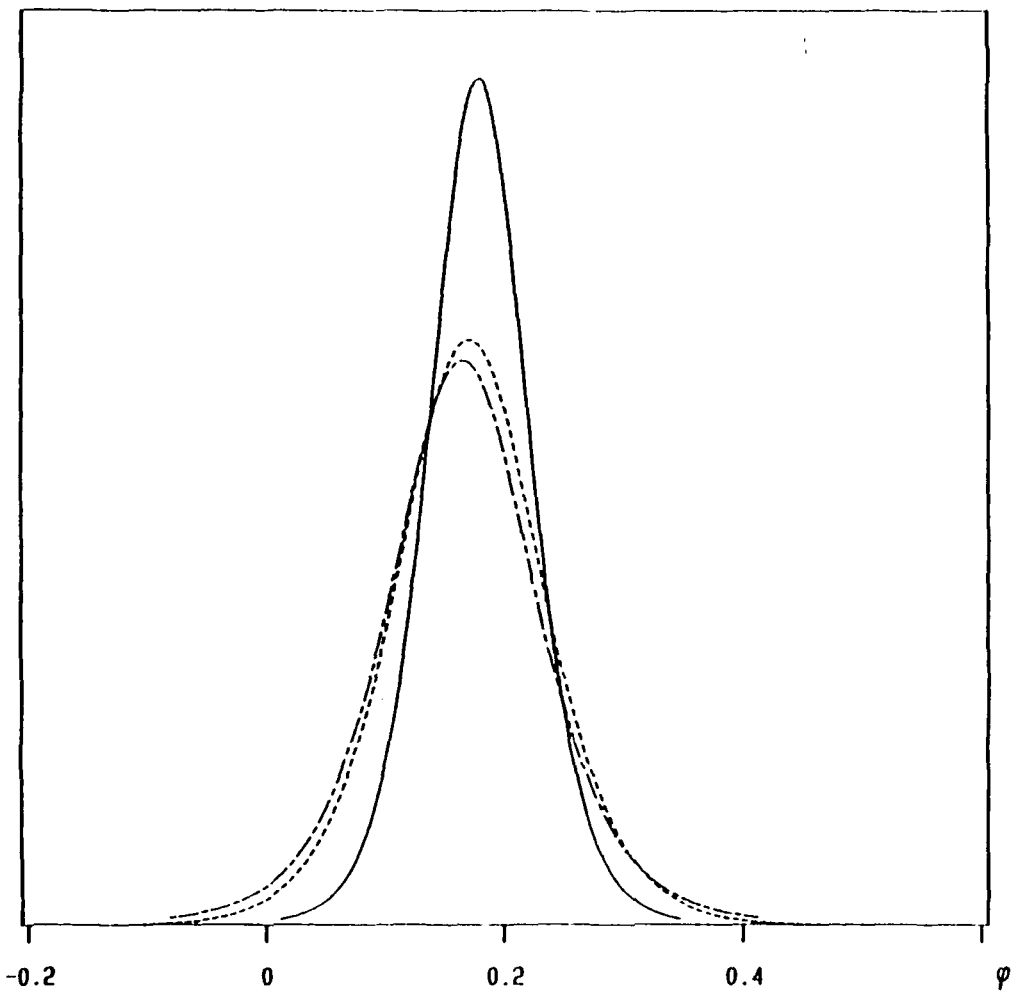


Figure 12 : Posterior densities of φ in the cross–over trial
_=complete data; – – – – = omitting the assumed missing value; . . . = based on 7 subjects

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER -21	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Illustration Of Bayesian Inference In Normal Data Models Using Gibbs Sampling		5. TYPE OF REPORT & PERIOD COVERED TECHNICAL REPORT 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Alan E. Gelfand, Susan E. Hills, Amy Racine-Poon, and Adrian F. M. Smith		8. CONTRACT OR GRANT NUMBER(s) N00014-89-J-1627
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics Stanford University Stanford, CA 94305		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR-042-267
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Statistics & Probability Program Code 1111		12. REPORT DATE September 6, 1989 13. NUMBER OF PAGES 37
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. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Bayesian inference; marginalisation; Gibbs sampler; variance components; order-restricted inference; hierarchical models; missing data; non-linear parameters; density estimation.

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

Use of the Gibbs sampler as a method for calculating Bayesian marginal posterior and predictive densities is reviewed and illustrated with a range of normal data models, including: variance components; unordered and ordered means; hierarchical growth curves, and missing data in a cross-over trial. In all cases the approach is straightforward to specify distributionally, trivial to implement computationally, with output readily adapted for required inference summaries.

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)