HIERARCHICAL BAYESIAN ANALYSIS OF CHANGE POINT PROBLEMS

BY

BRADLEY P. CARLIN, ALAN E. GELFAND and ADRIAN F. M. SMITH

TECHNICAL REPORT NO. 437
OCTOBER 18, 1990

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

The literature on change point problems is, by now, enormous. Here we consider only the so–called nonsequential or fixed sample size version although an informal sequential procedure which follows from Smith (1975) is a routine consequence (see Section 6). Still the literature is substantial and we merely note several review articles which span both parametric and nonparametric approaches. These are Hinkley, Chapman and Runger (1980), Siegmund (1986), Wolfe and Schechtman (1984) and Zacks (1983).

Our focus is on a fully Bayesian parametric approach. Use of the Bayesian framework for inference with regard to the change point dates to work by Chernoff and Zacks (1964) and Shiryayev (1963). Smith (1975) presents the Bayesian formulation in the case of a finite sequence of independent observations. In particular he addresses three situations: (i) both the initial distribution and the changed distribution are known, (ii) only the former is known (iii) both are unknown. Details are given for binomial and normal models. Broemeling (1972) and Menzefricke (1981) also consider normal models. Bacon and Watts (1971), Ferreira (1975), Holbert and Broemeling (1977), Chin Choy and Broemeling (1980), Smith and Cook (1980), and Moen, Salazar and Broemeling (1985) look at change points in linear models. Booth and Smith (1982) and, in a slightly different fashion, West and Harrison (1986) consider time series models. Diaz (1982) and Hsu (1982) study sequences of gamma–type random variables. Raftery and Akman (1986) extend this to Poisson processes where the change point need not have occurred at an event time.

In our view wider use of the Bayesian framework has been impeded by the difficulties in computing required marginal posterior distributions of the change

point and of the model parameters. In many of the above papers unsatisfying compromises have been made with regard to model specification and/or assumed knowledge of at least some of the model parameters in order to reduce the dimensionality of numerical integrations required to obtain marginal posterior distributions. The objective of this paper is to demonstrate that, for a very broad range of hierarchical change point models, such compromise is not necessary. Rather, all desired marginal posterior densities can be obtained through a straightforward iterative Monte Carlo method. Needed random generation, whether or not conjugacy is assumed, can typically be carried out in a reasonably efficient manner. We concede that for any particular situation there may exist more efficient methods for obtaining desired marginal posteriors. By the same token, we can handle many situations which were previously inaccessible. Moreover the conceptual simplicity of our method may prove an attractive alternative to the analytic and/or numerical sophistication demanded by other methods. The subsequent development advances work reported in Gelfand and Smith (1990) and Gelfand et. al. (1990).

In particular in Section 2 we formulate the hierarchical Bayes change point model. We clarify what distributions are sought and what distributions can be readily sampled. In Section 3 we briefly review the Gibbs sampler which underlies our iterative Monte Carlo method. We also show how the distribution of and expectations of arbitrary functions of the model parameters can be obtained. In Section 4 through 7 we present a range of examples. As an elementary illustration in Section 4 we apply our methodology to the "Nile data" previously analyzed under simplifying assumptions by Cobb (1978) and Hinkley and Schechtman (1987). In Section 5 we examine the British coal mining accident data of Maguire, Pearson and

Wynn (1952) as extended and corrected by Jarrett (1979). This data has been discussed in Worsley (1986), Raftery and Akman (1986) and Siegmund (1988). Our approach routinely handles the complete data set as well as a reduced version where 20% is treated as missing. In Section 6 we show that independent observations are not required to implement our methodology as we consider a change in transition matrix for a sequence of observations from a Markov chain. As a final example in Section 7 we study the changing regressions model. Illustrating in the context of simple linear regression we add a further wrinkle by allowing for known order between the original and changed slope. We look at two data sets – a real one investigated by Bacon and Watts (1971) and a generated one. We offer some concluding remarks in Section 8.

2. The Bayesian Formulation

The simplest formulation of the change point problem assumes f and g known densities with $Y_i \sim f(y)$, $i=1,\cdots k$, $Y_i \sim g(Y)$, $i=k+1,\cdots n$, with k unknown taking values in $\mathcal{K}_n = \{1,2,\cdots n\}$. Thus k=n is interpreted as "no change". As a result with $Y \equiv (Y_1,\cdots Y_n)$ the likelihood $L(Y;k)$ becomes

$$L(Y;k) = \prod_{i=1}^{k} f(Y_i) \cdot \prod_{i=k+1}^{n} g(Y_i) \qquad (1)$$

from which, for instance, the MLE for k can be directly obtained. Distribution theory for the MLE is well discussed in the literature (see e.g., Hinkley, 1970). The Bayesian perspective is added by placing a prior density $\tau(k)$ on $\mathcal{K}_n$ whence the posterior density of $k \mid Y$ becomes

$$L(Y;k)\tau(k) \Big/ \sum_{k=1}^{n} L(Y;k)\tau(k) \qquad (2)$$

Features of (2) e.g., mode, quantiles, expectations are easy to obtain. The question of whether or not a change has occurred is addressed through the posterior

odds for no change, $P(k = n|Y)/\{1 - P(k = n|Y)\}$. Since the model assumes at most one change, to make inference about where it occurred we might use the mode and a highest posterior density credible interval of (2). It is easy to sample from (2) since it is a discrete distribution on $\mathcal{K}_n$. Note also that the independence of the Y_i's is not necessary provided that for each k we can write down the likelihood $L(Y;k)$. Such a situation arises in the Markov chain example of Section 6 and also for instance in changing time series applications and more general dynamic models. Nonetheless to simplify notation and presentation, for the remainder of this section we assume that the Y_i are independent.

Suppose, as is more realistic, that the original and changed densities are unknown but that we assume a parametric family $f(Y|\theta)$ for the former and a (possibly different) parametric family $g(Y|\eta)$ for the latter. Then the likelihood, $L(Y;k,\theta,\eta)$ becomes

$$\prod_{i=1}^{k} f(Y_i|\theta) \prod_{i=k+1}^{n} g(Y_i|\eta).$$

Assuming a prior $\pi(\theta,\eta,k)$ on θ,η and k the joint distribution of data and parameters is

$$L(Y;k,\theta,\eta) \cdot \pi(\theta,\eta,k) \tag{3}$$

Interest centers on the marginal posterior distribution of k as well as that of components of θ and of η.

For the moment take θ and η to be scalar. Suppose we seek e.g. the distribution $k|Y$. To obtain this modulo normalizing constant requires a double integration over θ and η; to obtain it exactly requires an additional triple integration over θ, η, and k. Such calculation will typically require numerical or analytic expertise and, of course, if θ and η are vectors the situation worsens.

Consider however the "full" conditional distributions, $k|Y,\theta,\eta$, $\theta|Y,k,\eta$ and $\eta|Y,k,\theta$. Treating the data Y as fixed, suppose that these full conditional distributions uniquely determine the joint (conditional) distribution $k,\theta,\eta|Y$, hence the marginal posteriors which we seek. Then the development in Section 3 shows that suitable sampling from the full conditional distributions enables marginal posterior density estimates. We now demonstrate that in the present case sampling from the full conditional distributions will, rather generally, be straightforward.

First of all, given θ and η, $k|Y,\theta,\eta$ is exactly of the form (2) and is, as noted above, straightforwardly sampled. Suppose θ,η and k are assumed independent so that $\pi(\theta,\eta,k) = \lambda(\theta)\gamma(\eta)\tau(k)$. If λ is conjugate with f and γ is conjugate with g then $\theta|Y,\eta,k$ does not depend upon η and is merely the prior λ updated by the data $Y_1 \cdots Y_k$ while $\eta|Y,\theta,k$ is γ updated by the data $Y_{k+1},\cdots Y_n$. Since λ and γ are thus standard parametric families, sampling from these full conditionals is routine. Conjugate priors can be made arbitrarily diffuse and have an attractive robustness property (see Morris 1983, p. 525).

However, we may wish to investigate nonconjugate possibly heavier–tailed priors. Suppose we drop the assumptions of conjugacy for θ and for η and also the assumption of independence for θ, η and k. Nonetheless it is still the case that, for any scalar parameter, its associated full conditional distribution is proportional to $L(Y;k,\theta,\eta)$. But then random generation methods such as the ratio of uniforms method or perhaps the rejection method (see Devroye, 1986 or Ripley, 1986) enable sampling from such a nonstandardized integrable density function. In the present paper we confine ourselves to conjugate examples. Our experience with nonconjugate models and the ratio of uniforms method will be reported in a future paper.

Extension to general hierarchical Bayes models is immediate. Suppose, for instance, that again θ and η are independent of each other and of k. Also suppose that the prior on θ takes the form $\lambda(\theta|\alpha)$ with conjugate prior $\rho(\alpha)$ on the hyperparameter α, the prior on η takes the form $\gamma(\eta|\beta)$ with conjugate prior $\varphi(\beta)$ on the hyperparameter β. Now the joint distribution of the data and all parameters is

$$L(Y;k,\theta,\eta) \cdot \tau(k) \cdot \lambda(\theta|\alpha) \cdot \rho(\alpha) \cdot \gamma(\eta|\beta) \cdot \varphi(\beta) \qquad (4)$$

Once again the full conditionals are easy to write down and to sample from. In particular:

$k|Y,\theta,\eta,\alpha,\beta$ does not depend upon α or β and is exactly of the form (2),

$\theta|Y,k,\eta,\alpha,\beta$ does not depend upon η or β and is the prior λ updated by the data $Y_1,\cdots Y_k$,

$\eta|Y,k,\theta,\alpha,\beta$ does not depend upon θ or α and is the prior γ updated by the data $Y_{k+1},\cdots Y_n$,

$\alpha|Y,k,\theta,\eta,\beta$ does not depend upon Y,k,η or β and is the hyperprior ρ updated by θ,

$\beta|Y,k,\theta,\alpha,\eta$ does not depend upon Y,k,θ or α and is the hyperprior φ updated by η.

Again if the assumed conjugacy and parameter independence are relaxed the aforementioned random generation methods can still be employed.

3. Review of the Gibbs Sampler

In the previous section we demonstrated that, for rather general hierarchical Bayes change point models, sampling from each of the full conditional distributions can be accomplished. Under mild conditions (see Besag, 1974) the specification of

all full conditional densities uniquely determines the full joint density hence all marginal densities. The Gibbs sampler, introduced formally in Geman and Geman (1984) and subsequently developed in detail for general Bayesian computations in Gelfand & Smith (1990) and Gelfand et. al. (1990), provides a mechanism for extracting marginal distribution from the full conditional distributions. Tanner and Wong (1987) develop the notion of substitution sampling which is closely related as discussed in Gelfand and Smith (1990).

For the remainder of this section densities will be denoted generically by square brackets so that joint, conditional and marginal forms appear respectively as $[U,V]$ $[U|V]$ and $[V]$. The usual marginalization by integration is denoted by forms such as $[U] = \int [U|V] \cdot [V]$. For a collection of random variables $U_1, U_2, \cdots U_p$ the full conditional densities are thus denoted by $[U_s | U_r, r \neq s]$, $s = 1,2,\cdots p$ and we seek the marginal densities $[U_s]$, $s = 1,2,\cdots p$.

Gibbs sampling is a Markovian updating scheme which proceeds as follows. Given an arbitrary starting set of values $U_1^{(0)},\cdots,U_p^{(0)}$, we draw $U_1^{(1)}$ from $[U_1 | U_2^{(0)},\cdots,U_p^{(0)}]$, then $U_2^{(1)}$ from $[U_2 | U_1^{(1)}, U_3^{(0)},\cdots,U_p^{(0)}]\cdots$ and so on up to $U_p^{(1)}$ from $[U_p^{(0)} | U_1^{(1)}\cdots,U_{p-1}^{(1)}]$ to complete one iteration of the scheme. After t such iterations we would arrive at $((U_1^{(t)},\cdots,U_p^{(t)})$. Geman and Geman show under mild conditions that $(U_1^{(t)},\cdots,U_p^{(t)}) \xrightarrow{d} (U_1,\cdots,U_p) \sim [U_1,U_2,\cdots,U_p]$ as $t \longrightarrow \infty$. Hence for t large enough, $U_s^{(t)}$ for example will be regarded as a simulated observation from $[U_s]$. Parallel replication of this process m times yields m iid p–tuples $(U_{1j}^{(t)},\cdots,U_{pj}^{(t)})$ $j = 1,\cdots,m$. Note that sample size at say the ℓ–th iteration may be increased from m to any specified size by randomly reusing the $U_{sj}^{(\ell-1)}$ with replacement.

A kernel density estimate for $[U_s]$ based upon the $U_{sj}^{(t)}$ can be readily obtained (see e.g., Silverman, 1986) and should be adequate if, at the last iteration the number of replications, m, is large enough. However using a Rao–Blackwell argument (see Gelfand and Smith, 1990) a density estimate of the form

$$[\hat{U}_s] = \sum_{j=1}^{m} [U_s | U_{rj}^{(t)}, r \neq s]/m \tag{5}$$

is better under a wide range of loss functions. This is not surprising since (5) takes advantage of the known structure in the model whereas the kernel density estimate does not. The form (5) is a discrete mixture distribution and is in fact a Monte Carlo integration to accomplish the desired marginalization.

In our context the U_i are the unknown parameters in the hierarchical change point model whence for instance in the situation (4) the density estimate for $k|Y$ becomes $m^{-1} \sum_{j=1}^{m} [k|Y, \boldsymbol{\theta}_j^{(t)}, \eta_j^{(t)}]$ where $[k|Y, \boldsymbol{\theta}, \eta]$ is of the form (2).

There may also be interest in a function of the parameters (see Section 5), i.e. a function of the U_i, say $W(U_1, \cdots U_p)$. Each p–tuple, $(U_{1j}^{(t)}, \cdots, U_{pj}^{(t)})$, provides an observed $W_j^{(t)} \equiv W(U_{1j}^{(t)}, \cdots, U_{pj}^{(t)})$ whose marginal distribution is approximately $[W]$ whence a kernel density estimate for $[W]$ using these $W_j^{(t)}$ can be developed. A "Rao–Blackwellized" density estimate analogous to (5) can also be obtained. If U_s actually appears as an argument of W the full conditional density $[W | U_r, r \neq s]$ can be obtained by univariate transformation from $[U_s | U_r, r \neq s]$.

In concluding this section we note that complete implementation of the Gibbs sampler requires that a determination of t be made and that across iterations, choice(s) of m be specified. In a challenging application some experimentation with different settings for t and m will likely be necessary. We do not view this as a deterrent since random generation is generally inexpensive and since there may be no feasible alternative. In the subsequent examples convergence was evaluated in a

univariate manner by plotting marginal posterior density estimates of the form (5) five iterations apart to see if they are equivalent under a "thick felt tip pen" test. Typically a somewhat small m is used until convergence is concluded at which point for a final iteration m is increased by an order of magnitude to develop the presented density estimate. We make no claims for this procedure. Assessment of convergence is a complex issue which we are currently investigating.

4. Normal Means Problem

As an elementary illustration we assume data from a normal distribution whose mean may have changed at some point during the period of observation. That is, $Y_i \sim N(\theta_1,\sigma_1^2)$ $i = 1,\cdots k$, $Y_i \sim N(\theta_2,\sigma_2^2)$, $i=k+1, \cdots n$. At the second stage we adopt the following prior distributions: k discrete uniform on $\mathcal{K}_n$; θ_1, $\theta_2 \sim N(\mu,\tau^2)$; σ_1^2, $\sigma_2^2 \sim IG(a,b)$ where IG denotes the inverse gamma distribution. At the third stage of the hierarchy, let $\mu \sim N(\mu_0,\sigma_0^2)$ and $\tau^2 \sim IG(c,d)$ where μ_0, σ_0^2, a, b, c, and d are known constants. This hierarchical structure is standard in the Bayesian literature and has already been used with the Gibbs sampler for comparing p independent normal means in Gelfand and Smith (1990). A simplified version for the change point problem was discussed in Smith (1975). Our primary interest focuses on the marginal posterior distributions of k, θ_1, and θ_2. All told we have a 7 parameter problem. The 7 corresponding full conditional distributions are:

$$\theta_1 | Y,\theta_2,\sigma_1^2,\sigma_2^2,\mu,\tau^2,k \sim N\left[\frac{\mu\sigma_1^2+\tau^2\sum_1^k Y_i}{\sigma_1^2+k\tau^2}, \frac{\sigma_1^2\tau^2}{\sigma_1^2+k\tau^2}\right]$$

$$\theta_2 | Y,\theta_1,\sigma_1^2,\sigma_2^2,\mu,\tau^2,k \sim N\left[\frac{\mu\sigma_2^2+\tau^2\sum_{k+1}^n Y_i}{\sigma_2^2+(n-k)\tau^2}, \frac{\sigma_2^2\tau^2}{\sigma_2^2+(n-k)\tau^2}\right]$$

$$\sigma_1^2 \mid \mathbf{Y},\theta_1,\theta_2,\sigma_2^2,\mu,\tau^2,k \sim \mathrm{IG}\left[a+\tfrac{k}{2}\,,\ \left\{\tfrac{1}{2}\sum_{1}^{k}(Y_i-\theta_1)^2+\tfrac{1}{b}\right\}^{-1}\right]$$

$$\sigma_2^2 \mid \mathbf{Y},\theta_1,\theta_2,\sigma_1^2,\mu,\tau^2,k \sim \mathrm{IG}\left[a+\tfrac{n-k}{2}\,,\ \left\{\tfrac{1}{2}\sum_{k+1}^{n}(Y_i-\theta_2)^2+\tfrac{1}{b}\right\}^{-1}\right]$$

$$\mu \mid \mathbf{Y},\theta_1,\theta_2,\sigma_1^2,\sigma_2^2,\tau^2,k \sim \mathrm{N}\left[\frac{\mu_0\tau^2+\sigma_0^2(\theta_1+\theta_2)}{\tau^2+2\sigma_0^2}\,,\ \frac{\tau^2\sigma_0^2}{\tau^2+2\sigma_0^2}\right]$$

$$\tau^2 \mid \mathbf{Y},\theta_1,\theta_2,\sigma_1^2,\sigma_2^2,\mu,k \sim \mathrm{IG}\left[c+1,\ \left\{\tfrac{1}{2}\sum_{l=1}^{2}(\theta_l-\mu)^2+\tfrac{1}{d}\right\}^{-1}\right]$$

with $$p(k \mid \mathbf{Y},\theta_1,\theta_2,\sigma_1^2,\sigma_2^2,\mu,\tau^2) = \frac{L(\mathbf{Y};k,\theta_1,\theta_2,\sigma_1^2,\sigma_2^2)}{\sum_{\mathcal{K}_n} L(\mathbf{Y};k,\theta_1,\theta_2,\sigma_1^2,\sigma_2^2)}$$

where $$L(\mathbf{Y};k,\theta_1,\theta_2,\sigma_1^2,\sigma_2^2) = \exp\left\{-\frac{1}{2\sigma_1^2}\sum_{1}^{k}(Y_i-\theta_1)^2-\frac{1}{2\sigma_2^2}\sum_{k+1}^{n}(Y_i-\theta_2)^2\right\}\Big/\sigma_1^k\,\sigma_2^{n-k}$$

The Gibbs sampler is thus easily implemented, the only slight inconvenience being that at each iteration the discrete full conditional for k must be restandardized, hence reevaluated at each of the n support points.

As an example we analyze the data set given in Table 1 which records annual volume of discharge from the Nile river at Aswan for the years 1871 to 1970. This data has been studied by Cobb (1978) and by Hinkley and Schechtman (1987). Both provide a conditional frequentist approach to estimating k but unrealistically

assume that all the parameters except k are known using data–based values $\mu_1 = 11.0$, $\mu_2 = 8.5$, and $\sigma_1 = \sigma_2 = 1.25$.

(Insert Table 1 here)

We chose relatively vague priors for our variances by taking a = b = c = d = 2, and chose the slightly more informative values of $\mu_0 = 10$ and $\sigma_0 = 2$ for the hyperprior on μ. Figure 1 presents the marginal posterior density estimate for the change point, Figure 2 for the mean level before and after the change. Convergence was achieved within 30 iterations and the densities were drawn using the 31^{st} with m = 100.

Clearly there is strong evidence that the change occurred at k = 28, or following the year 1898. Our estimate of the posterior probability of this event, .77, is slightly less than the Bayes estimate of .81 that Cobb obtained using his simplified model, but more than .64, the unconditional (frequentist) probability that the maximum likelihood estimator of k equals 28 using the asymptotic sampling distribution given in Hinkley (1970). The posterior modal values of 10.88 and 8.55 for θ_1 and θ_2 respectively, are close to the data–based values above. The marginal posterior for θ_1 is more spread than that for θ_2 since the change seems to have occurred early in the series.

5. A Poisson Process With Change Point

As a second illustration consider Bayesian analysis of a Poisson process with a change point. Such an analysis was recently given by Raftery and Akman (1986). Assuming conjugate priors they examine the collection of times between occurrences of the process allowing the time of change to be continuous. For chronologically ordered time intervals not necessarily of equal length we study the set of occurrence counts. We use a three stage hierarchical model but assume that the change point

occurs between intervals. Thus our model for the data is $Y_i \sim P_0(\theta t_i)$, $i = 1,\cdots k$, $Y_i \sim P_0(\lambda t_i)$, $i = k+1,\cdots n$. At the second stage we place independent priors over k, θ and λ : k discrete uniform $\mathcal{K}_n$, $\theta \sim G(a_1,b_1)$, and $\lambda \sim G(a_2,b_2)$ where G denotes the gamma distribution. At the third stage we take $b_1 \sim IG(c_1,d_1)$ independent of $b_2 \sim IG(c_2,d_2)$ and assume that a_1, a_2, c_1, c_2, d_1, and d_2 are known. Our interest again lies in the marginal posterior distributions of k, θ and λ. The collection of full conditional distributions are easily available as:

$$\theta|\,Y,\lambda,b_1,b_2,k \sim G\left[a_1 + \sum_1^k Y_i,\ \left[\sum_1^k t_i + b_1^{-1}\right]^{-1}\right]$$

$$\lambda|\,Y,\theta,b_1,b_2,k \sim G\left[a_2 + \sum_{k+1}^n Y_i,\ \left[\sum_{k+1}^n t_i + b_2^{-1}\right]^{-1}\right]$$

$$b_1|\,Y,\theta,\lambda,b_2,k \sim IG\left[a_1 + c_1,\ \left[\theta + d_1^{-1}\right]^{-1}\right]$$

$$b_2|\,Y,\theta,\lambda,b_1,k \sim IG\left[a_2 + c_2,\ \left[\lambda + d_2^{-1}\right]^{-1}\right]$$

and

$$p(k|\,Y,\theta,\lambda,b_1,b_2) = L(Y;k,\theta,\lambda) \Big/ \sum_k L(Y;k,\theta,\lambda)$$

where $$L(Y;k,\theta,\lambda) = \exp\{(\lambda-\theta)\sum_1^k t_i\} \cdot (\theta/\lambda)^{\sum_1^k Y_i}$$

A variable of interest might be the ratio $R = \theta/\lambda$. Following the discussion near the end of Section 3 we transform from θ to R to obtain as full conditional

distribution $R|Y,\lambda,b_1,b_2,k \sim G(a_1 + \sum_1^k Y_i,\ \lambda^{-1}(\sum_1^k t_i + b_1^{-1})^{-1})$. A density estimate as in (5) can now be obtained.

The Gibbs sampler, in general, straightforwardly handles missing data. Such points can be treated as additional model parameters whose full conditional distributions are immediate. However if the predictive distributions for such points are not of interest we would typically not perform the additional random generation required to incorporate them as model parameters. Rather, we would prefer to use the likelihood based solely upon the observed data. In the present case, for example, we need only set the associated Y and t for missing points to 0 in the likelihood and proceed as above.

A much–analyzed data set of intervals between British coal–mining disasters during the 112 year period 1851–1962 was gathered by Maguire et. al. (1952), extended and corrected by Jarrett (1979). Frequentist change point investigations appear in Worsley (1986) and in Siegmund (1988) while Raftery and Akman (1986) apply their Bayesian model. We apply ours using yearly intervals. The resulting observed annual counts appear in Table 2; $k \in \{1,2,\cdots,112\}$.

(Insert Table 2 here)

Raftery and Akman used a vague second stage prior by taking $a_1 = a_2 = .5$ and $b_1 = b_2 = 0$. We make our model comparable by taking their values for a_1 and a_2 and choosing vague third stage priors for b_1 and b_2, letting $c_1 = c_2 = 0$ and $d_1 = d_2 = 1$. Convergence of the algorithm was obtained after 15 iterations and again m = 100. Figure 3 shows the density estimates for $k|Y$ using the entire data set (solid lines) and using the data set with every fifth year deleted (dashed lines). Note that k = 41 is the posterior mode in both cases, and that the three largest spikes are k = 39, 40 and 41, meaning the change most probably occured sometime between late

1889 and early 1892. Raftery and Akman obtained similar results – a posterior mode of March 10, 1890 and a posterior median of August 27, 1890. Also note that in the missing data case, with no data at k = 5j the density estimates are the same for k = 5j and k = 5j-1, j = 1,$\cdots$,22.

Figure 4 displays the density estimates for $\theta|Y$ (solid lines) and for $\lambda|Y$ (dashed lines) for the original and reduced data sets, and Figure 5 does the same for $R|Y$. For all three pairs of curves, as expected the reduced data set has the greater spread. We see that systematically deleting 20% of the data has no effect on the posterior modes of θ and λ (3.06 and 0.89, respectively), and shrinks the posterior mode of R only slightly from 3.25 to 3.18. Raftery and Akman obtained 3.41 as the posterior mean of R using their model. As a final remark, we note that, as in the last example, the posterior probability that k = n is essentially 0, indicating very strong evidence for a change.

6. Markov Chain Change Points

To our knowledge there is no previous literature examining the Markov chain change point problem from a Bayesian viewpoint. Nonetheless the problem is extremely important arising for instance in the analysis of spatial variations in base frequencies in DNA (Curnow & Kirkwood 1989 sec. 7) as well as in system user authentication contexts. Suppose then a sequence of a sequence of n observations $Y = (Y_1,\cdots,Y_n)$ from a process which is an p–state stationary Markov chain having either transition matrix A or precisely one change to a transition matrix B. The entries of A are $a_{ij} = P(Y_{t+1} = j|Y_t = i)$ whence $a_{ij} \geq 0$, $\sum_j a_{ij} = 1$; similarly for B with entries b_{ij}. We take independent Dirichlet priors on the rows of A yielding the prior $\lambda(A) = \prod_{i=1}^{p} D_{\lambda_i}(a_i)$ where

$$D_{\lambda_i}(a_i) = \frac{\Gamma(\sum_{j=1}^{p} \lambda_{ij})}{\prod_{j=1}^{p} \Gamma(\lambda_{ij})} \prod_{j=1}^{p} a_{ij}^{\lambda_{ij}}$$

We do similarly for B taking $\gamma(B) = \prod_{i=1}^{p} D\gamma_i(b_i)$. λ and γ offer a rich class of priors. In particular, special structure for A and B can be modeled through appropriate choices of the λ_i and γ_i. As before, we denote the prior on k by $\tau(k)$. Note that the multinomal change point problem occurs as a special case when the Y_i are independent and $a_i = a$, $b_i = b$.

We assume A,B,k independent whence the joint of distribution of the data and parameters is

$$f(Y|A,B,k) \cdot \tau(k) \cdot \lambda(A) \cdot \gamma(B) \tag{6}$$

where at $Y = y = (y_1, \cdots, y_n)$,

$$f(y|A,B,k) = \prod_{t=1}^{k-1} a_{y_t,y_{t+1}} \prod_{t=k}^{n-1} b_{y_t,y_{t+1}} \cdot \mu_0(y_1), \qquad 2 \le k \le n-1$$

$$= \prod_{t=1}^{n-1} a_{y_t,y_{t+1}} \cdot \mu_0(y_1), \qquad k = n \tag{7}$$

$$= \prod_{t=1}^{n-1} b_{y_t,y_{t+1}} \cdot \mu_0(y_1), \qquad k = 1$$

with μ_0 denoting the initial or starting state distribution. Using (6) and (7) we see that the full posteriors are as follows:

$$A|Y,B,k \sim \prod_{i=1}^{p} D_{\lambda_i + Z_i}(a_i) \tag{8}$$

$$B|Y,A,k \sim \prod_{i=1}^{p} D_{\gamma_i + Z_i'}(a_i) \tag{9}$$

where $Z_i = (Z_{i1},\cdots,Z_{ip})$ with Z_{ij} = # of transitions from i to j within observations $Y_1,\cdots,Y_k$, $Z_i' = (Z_{i1}',\cdots,Z_{ip}')$ with Z_{ij}' = # of transitions from i to j within observations $Y_k,\cdots,Y_n$, and

$$p(k|Y,A,B) = \frac{f(Y|A,B,k)\,\tau(k)}{\sum_{k=1}^{n} f(Y|A,B,k)\,\tau(k)} \tag{10}$$

Note that since we are conditioning on all the data, hence Y_1, μ_0 does not appear in (8) or (9) and cancels in (10). Thus we need not worry about its specification.

A three stage hierarchical model can be developed by placing priors on the λ_i and γ_i. Details are similar to the previous examples and are omitted. We note that work on hierarchical models for contingency tables as in e.g., Albert and Gupta (1982) is closely related.

We add the usual Bayesian caveat. The number of parameters involved in this analysis is $2r(r-1) + 1$. Hence the smaller n is relative to $2r(r-1) + 1$ the more the prior will drive the posterior densities. Other considerations may make it important to detect the change point as soon as possible rather than waiting to the end of the data sequence. Simple updating of the posterior given new data is not possible since with additional data the domain of k changes. If "change" versus "no change" is of primary concern, as an informal sequential procedure we might place a spike in the prior τ at $k = n$. For instance $\tau(n) = 1/2$ implies prior indifference regarding a change in the first n observations. Use of a uniform prior would imply odds of $n-1$ to 1 for a change. Monitoring the posterior odds for no change, $P(k=n|Y)/(1-P(K=n|Y))$, will reveal a downward trend to warn of change.

As an illustrative example we consider a three state stationary Markov chain where

$$A = \begin{bmatrix} 0.7000 & 0.1500 & 0.1500 \\ 0.3333 & 0.3333 & 0.3333 \\ 0.3333 & 0.3333 & 0.3333 \end{bmatrix}, \quad B = \begin{bmatrix} 0.3333 & 0.3333 & 0.3333 \\ 0.1500 & 0.7000 & 0.1500 \\ 0.3333 & 0.3333 & 0.3333 \end{bmatrix}$$

Table 3 gives a sequence of 50 observations generated with a change after k = 35.

(Insert Table 3 here)

$\tau(k)$ was taken to be uniform over $\{1,\cdots 50\}$. The Dirichlet priors are taken to be generalized uniforms, i.e. all λ_{ij}, $\gamma_{ij} = 1$. Turning to the marginal posteriors, in Figure 6 we show for $k|Y$ a comparison of the density at iteration 4 (solid line) and at iteration 5(dotted line) using m = 100. We would not conclude convergence. In Figure 7 we compare iterations 30 (solid line) and 31 (dotted line). Now a conclusion of convergence seems appropriate. In fact the density estimates have been roughly this stable since iteration 20. The marginal posterior for k has mode at 33 with roughly 60% of the mass on the set {33, 34, 35}. In Figure 8 we show the marginal posteriors for $a_{11}|Y$ and $b_{11}|Y$. In Figure 9 we show the marginal posteriors for $a_{22}|Y$ and $b_{22}|Y$. Note that for both a_{11} and b_{22} modes and concentration of mass agree with the "true" a_{11} and b_{22} respectively. Interestingly, for b_{11} the marginal posterior is one–tailed. This arises because for the observed sequence there are no transitions from state 1 to state 1 after k = 33.

7. Changing Linear Regression Models

Bayesian analysis of changing linear models includes the work of Bacon and Watts (1971), Ferreira (1975), Holbert and Broemeling (1977), Chin Choy and Broemeling (1980), Smith and Cook (1980) and Moen, Salazar and Broemeling (1985). Typically this work involves simplified models and/or considerable analytic effort. By contrast the Gibbs sampler enables analysis of complex models while

demanding little mathematical and computational expertise from the user. In what follows, we outline and exemplify the basic unconstrained model. Furthermore, in the spirit of robustness we investigate the impact of the choice of the prior on the results. We then add the constraint of ordered slope parameters, and show the surprising ease with which the Gibbs sampler provides a solution, illustrating with a second example.

7.1 Basic Unconstrained Model

Consider a three–stage hierarchical simple linear regression model, where at the first stage $Y_i \sim N(\alpha_1+\beta_1 x_i,\ \sigma_1^2)$, $i = 1,\cdots,k$, $Y_i \sim N(\alpha_2+\beta_2 x_i,\ \sigma_2^2)$, $i = k+1,\cdots, n$. At the second stage we take $\theta_1 = (\alpha_1,\beta_1)^T$, $\theta_2 = (\alpha_2,\beta_2)^T$ independent $N(\theta_0,\Sigma)$, and σ_1^2, σ_2^2 independent $IG(a_0,b_0)$. Again we assume k follows a discrete uniform distribution on $\mathcal{K}_n$.

At the third stage, we take the independent normal–Wishart form $\theta_0 \sim N(\mu,C)$, $\Sigma^{-1} \sim W((\rho V)^{-1},\rho)$. Thus we have a 13 parameter problem with known constants μ, C, V, ρ, a_0 and b_0.

For a given k define $B_i^{(k)} = (\sigma_i^{-2} X_i^{(k)T} X_i^{(k)} + \Sigma^{-1})^{-1}$, $b_i^{(k)} = \sigma_i^{-2} X_i^{(k)T} Y_i^{(k)} + \Sigma^{-1}$, $i = 1,2$ where $X_1^{(k)} = \begin{bmatrix} 1 & \cdots & 1 \\ X_1 & \cdots & X_k \end{bmatrix}^T$, $X_2^{(k)} = \begin{bmatrix} 1 & \cdots & 1 \\ X_{k+1} & \cdots & X_n \end{bmatrix}^T$, $Y_1^{(k)} = (Y_1,\cdots,Y_n)^T$, $Y_2^{(k)} = (Y_{k+1},\cdots Y_n)^T$. Let $\Delta = (2\,\Sigma^{-1} + C^{-1})^{-1}$. Then using standard distribution theory we obtain the following full conditional distributions: $\theta_i \sim N(B_i^{(k)} b_i^{(k)},\ B_i^{(k)})$ $i = 1,2$; $\sigma_1^2 \sim IG(a_0 + k/2,\ \{\frac{1}{2}(Y_1^{(k)} - X_1^{(k)}\theta_1)^T (Y_1^{(k)} - X_1^{(k)}\theta_1) + 1/b_0\}^{-1})$; $\sigma_2^2 \sim IG(a_0 + \frac{n-k}{2},\ \{\frac{1}{2}(Y_2^{(k)} - X_2^{(k)}\theta_2)^T (Y_2^{(k)} - X_2^{(k)}\theta_2) + 1/b_0\}^{-1})$; $\theta_0 \sim N(\Delta\{\Sigma^{-1}(\theta_1+\theta_2)+C^{-1}\mu\},\ \Delta)$; $\Sigma^{-1} \sim W(\{\sum_1^2 (\theta_i - \theta_0)(\theta_i - \theta_0)^T + \rho V\}^{-1},\ \rho+2)$ and $k \sim \rho(k \mid Y,\ \theta_1,\ \theta_2,\ \sigma_1^2,\ \sigma_2^2,\ \theta_0,\ \Sigma) = L(Y;\ k,\ \theta_1,\ \theta_2,\ \sigma_1^2,\ \sigma_2^2) / \sum_{\mathcal{K}_n} L(Y;\ k,\ \theta_1,\ \theta_2,\ \sigma_1^2,\ \sigma_2^2)$

where $L(Y;\ k,\ \theta_1,\ \theta_2,\ \sigma_1^2,\ \sigma_2^2) = \exp\{-\sum_{i=1}^{2} \frac{1}{2\sigma_i^2} (Y_i^{(k)} - X_i^{(k)}\theta_i)^T (Y_i^{(k)} - X_i^{(k)}\theta_i)\} / \sigma_1^k\, \sigma_2^{n-k}$.

Simulation from the Wishart distribution is easily done using the algorithm of Odell and Feiveson (1966) as outlined for the 2×2 case in Gelfand et. al. (1990).

We apply the Gibbs sampler thus defined to the data in Table 4, which come from Bacon and Watts (1971, p.530). In this data, X represents the log of the flow rate of water down an inclined channel (g./cm.sec.), and Y represents the log of the height of the stagnant surface layer (cm.) for different surfactants.

(Insert Table 4 here)

The data seem to indicate a decreasing linear trend that appears to become more steeply decreasing for $X > 0$. We apply our model using a vague prior for the σ_i^2, $a_0 = 0.1$ and $b_0 = 100$, along with a vague third–stage prior for θ_0, all entries of μ and C^{-1} equal to 0. For Σ^{-1} we compare two different Wishart priors, the first somewhat informative ($\rho = 4$, $V = \begin{bmatrix} 0.001 & 0 \\ 0 & 0.3 \end{bmatrix}$), and the second more vague ($\rho = 2$, $R = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$).

Convergence was obtained within 50 iterations and densities are drawn at the 51^{st} using $m = 100$. Figure 10 plots the estimated marginal posterior for k using the informative (solid line) and vague (dashed line) priors on Σ^{-1}. Figure 11 does the same for β_1 and β_2, the slopes before and after the change (the β_2 estimates are on the left). Clearly with noninformative priors on θ_0, σ_1^2 and σ_2^2, changing the prior on Σ^{-1} can have a rather marked effect on the results. For example, k's posterior distribution is much more diffuse using the vague Wishart prior. Our intent here is not to claim one or the other solution is "correct," but rather to emphasize how simple it is using the Gibbs sampler for the data analyst to investigate prior robustness interactively.

The estimated posterior modes of β_1 for the informative and vague priors are -0.42 and -0.44 respectively, and for β_2 they are -1.01 and -0.98, respectively.

Though Bacon and Watts used a different parameterization, a somewhat less elaborate model, vague prior specification, and treated k as continuous, their results are quite comparable to ours: their joint posterior obtains its global maximum at $\beta_1 = -0.44$, $\beta_2 = -1.02$, with the change estimated to have occurred at $X = 0.054$. This corresponds to $k = 13$, which is the posterior median we obtain under either of our prior specifications.

7.2 Ordered Slopes Model

In certain applications we know that if there has been a change, the change must be in a certain direction. We would want to incorporate this prior knowledge into our analysis. This entails placing order constraint(s) on the parameters, which makes required integrations much more difficult, perhaps impossible. The Gibbs sampler can again be used to overcome this difficulty.

Consider the portion of the prior involving θ_1 and θ_2. Suppose we draw θ_i by first obtaining β_i and then α_i given β_i. Assume that β_1 and β_2 correspond to the larger and smaller respectively of two independent draws from the same marginal normal prior on the slopes as in the preceding section. Then the full posteriors for β_1 and β_2 are merely truncated versions of the previous full posteriors, truncated so that $\beta_2 < \beta_1$. The full conditional distributions for the α_i as well as for all other parameters are unchanged. One for one sampling from a truncated distribution can be effected using a suggestion in Devroye (1986, p. 38). More general order constraint can be modeled by assuming two independent but not identically distributed draws before ordering to obtain β_1 and β_2.

(Insert Table 5 about here)

To illustrate these ideas, consider the data set in Table 5 generated according to the model $Y_i \sim N(i, 5^2)$, $i = 1, \cdots 15$, $Y_i \sim N(30-i, 5^2)$, $i = 16 \cdots, 29$ so that the

true β_1 and β_2 are 1 and -1, respectively. The remaining prior structure agrees with that of the previous section except for Σ^{-1} where we take Wishart with $\rho = 3$ and V = I. We again obtained convergence within 50 iterations and our plots are based on m = 100. Figure 12 compares the resulting estimated marginal posteriors for β_1 and β_2 under the model of the previous section (solid line) and the present ordered slopes model (dashed line). The pair of β_1 curves are to the right in the figure. For β_1, the constrained model yields a more concentrated posterior. In particular, both of the tails are quite heavy in the unconstrained model ranging from roughly -8.0 to +13.0 while for the constrained model they are limited to roughly -1.2 to 5.7. Additional studies (not shown) show more and more dramatic differences between the two models as ρ is decreased; the tails become heavier and heavier for the basic model while remaining virtually unchanged in the constrained model.

8. Conclusion

We have shown how a broad range of hierarchical Bayes change point models can be straightforwardly analyzed using an iterative sampling based approach known as the Gibbs sampler. Interesting future applications lie in such as areas as time series and other dynamic models, generalized linear models and nonlinear models as well as in the employment of nonconjugate priors.

References

Albert, J.H. and Gupta, A.K. (1982). Mixtures of Dirichlet distributions and estimation in contingency tables. *Ann. Statist.* **10**, 1261–8.

Bacon, D.W. and Watts, D.G. (1971). Estimating the transition between two intersecting straight lines. *Biometrika* **58**, 525–34.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems (with discussion). *J.R. Statist. Soc.* B **36**, 192–326.

Booth, N.B. and Smith, A.F.M. (1982). A Bayesian approach to retrospective identification of change points. *J. Econ.* **19**, 7–22.

Broemeling, L. (1972). Bayesian procedures for detecting a change in a sequence of random variables. *Metron* **XXX**, 1–14.

Chernoff, H. and Zacks, S. (1964). Estimating the current mean of a normal distribution which is subjected to changes in time. *Ann. Math. Statist.* **35**, 999–1018.

Chin Choy, J. and Broemeling, L. (1980). Some Bayesian inferences for a changing linear model. *Technometrics* **22**, 71–8.

Cobb, G. (1978). The problem of the Nile: conditional solution to a changepoint problem. *Biometrika* **65**, 243–51.

Curnow, R.N. and Kirkwood, T.B.L. (1989). Statistical Analysis of Deoxyribonucleic Acid Sequence Data – A Review. *J.R. Statist. Soc.* A, **152**, 199–220.

Devroye, L. (1986). *Non–Uniform Random Variate Generation.* New York: Springer–Verlag.

Diaz, J. (1982). Bayesian detection of a change of scale parameter in sequences of independent gamma random variables. *J. Econ.* **19**, 23–9.

Ferreira, P.E. (1975). A Bayesian analysis of a switching regression model: a known number of regimes. *J. Am. Statist. Assoc.* **70**, 370–4.

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

Gelfand, A.E., Hills, S.E., Racine–Poon, A. and Smith, A.F.M. (1990). Illustration of Bayesian inference in normal data models using Gibbs sampling. Technical Report, Department of Mathematics, University of Nottingham.

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, **6**, 721–41.

Hinkley, D.V. (1970). Inference about the change–point in a sequence of random variables. *Biometrika* **57**, 1–17.

Hinkley, D. and Schechtman, (1987). Conditional bootstrap methods in the mean shift model. *Biometrika* **74**, 85–94.

Hinkley, D., Chapman, P. and Runger, G. (1980). Change point problems. University of Minnesota Technical Report 382.

Holbert, D. and Broemeling, L.D. (1977). . Bayesian inference related to shifting sequences and two–phase regression. *Commun. in Statist.* A **6**, 265–75

Hsu, D.A. (1982). A Bayesian robust detection of shift in the risk structure of stock market returns. *J. Am. Statist. Assoc.* **77**, 29–39.

Jarrett, R.G. (1979). A note on the intervals between coal–mining disasters. *Biometrika* **66**, 191–3.

Maguire, B.A., Pearson, E.S. and Wynn, A.H.A. (1952). The time intervals between industrial accidents. *Biometrika* **38**, 168–80.

Menzefricke, U. (1981). A Bayesian analysis of a change in the precision of a sequence of independent normal random variables at an unknown time point. *Appl. Statist.* **30**, 141–6.

Moen, D.H., Salazar, D. and Broemeling, L.D. (1985). Structural changes in multivariate regression models. *Commun. in Statist.* A **14**, 1757–68.

Morris, C.N. (1983). Natural exponential families with quadratic variance functions: statistical theory. *Ann. Statist.* 11, 515–29.

Odell, P.L. and Feiveson, A.H. (1966) A numerical procedure to generate a sample covariance matrix. *J. Am. Statist. Assoc.* 61, 198–203.

Raftery, A.E. and Akman, V.E. (1986). Bayesian analysis of a Poisson process with a change–point. *Biometrika* 73, 85–9.

Ripley, B. (1986). *Stochastic Simulation*, John Wiley & Sons, New York.

Shiryayev, A.N. (1963). On optimum methods in quickest detection problems. *Theory of Prob. and its Appl.* 8, 22–46.

Siegmund, D. (1986). Boundary crossing probabilities and statistical applications. *Ann. Statist.* 14, 361–404.

Siegmund, D. (1988). Confidence sets in change point problems. *Inter. Statist. Review J6* 1, 31–48.

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

Smith, A.F.M. (1975). A Bayesian approach to inference about a change–point in a sequence of random variables. *Biometrika* 62, 407–16.

Smith, A.F.M. and Cook, D.G. (1980). Straight lines with a change–point: a Bayesian analysis of some renal transplant data. *Appl. Statist.* 29, 180–9.

Tanner, M. and Wong, W.H. (1987). The calculation of posterior distributions by data augmentation (with discussion). *J. Am. Statist. Assoc.* 82, 528–50.

West, M. and Harrison, P.J. (1986). Monitoring and adaptation in Bayesian forecasting models. *J. Am. Statist. Assoc.* 81, 741–50.

Wolfe, D.A. and Schechtman, E. (1984). Nonparametric statistical procedures for the change point problem. *J. Statist. Planning and Inference* 9, 389–96.

Worsley, K.J. (1986). Confidence regions and tests for a change–point in a sequence of exponential family random variables. *Biometrika* 73, 91–104.

Zacks, S. (1983). Survey of classical and Bayesian approaches to the change point problem: fixed sample and sequential procedures of testing and estimation, In: *Recent Advances in Statistics*, Herman Chernoff Festschrift, p. 245–69, Academic Press, New York/London.

Table 1. Annual volume of the Nile River (discharge at Aswan, 10^{10} m^3) from 1871 to 1970, with apparent changepoint near 1898

Year	Vol.	Year	Vol.	Year	Vol.	Year	Vol.
1871	11.20	1896	12.20	1921	7.68	1946	10.40
1872	11.60	1897	10.30	1922	8.45	1947	8.60
1873	9.63	1898	11.00	1923	8.64	1948	8.74
1874	12.10	1899	7.74	1924	8.62	1949	8.48
1875	11.60	1900	8.40	1925	6.98	1950	8.90
1876	11.60	1901	8.74	1926	8.45	1951	7.44
1877	8.13	1902	6.94	1927	7.44	1952	7.49
1878	12.30	1903	9.40	1928	7.96	1953	8.38
1879	13.70	1904	8.33	1929	10.40	1954	10.50
1880	11.40	1905	7.01	1930	7.59	1955	9.18
1881	9.95	1906	9.16	1931	7.81	1956	9.86
1882	9.35	1907	6.92	1932	8.65	1957	7.97
1883	11.10	1908	10.20	1933	8.45	1958	9.23
1884	9.94	1909	10.50	1934	9.44	1959	9.75
1885	10.20	1910	9.69	1935	9.84	1960	8.15
1886	9.60	1911	8.31	1936	8.97	1961	10.20
1887	11.80	1912	7.26	1937	8.22	1962	9.06
1888	7.99	1913	4.56	1938	10.10	1963	9.01
1889	9.58	1914	8.24	1939	7.71	1964	11.70
1890	11.40	1915	7.02	1940	6.76	1965	9.12
1891	11.00	1916	11.20	1941	6.49	1966	7.46
1892	12.10	1917	11.00	1942	8.46	1967	9.19
1893	11.50	1918	8.32	1943	8.12	1968	7.18
1894	12.50	1919	7.64	1944	7.42	1969	7.14
1895	12.60	1920	8.21	1945	8.01	1970	7.40

Table 2: British Coalmining disaster data by year 1851–1962.
Data from Maguire et. al. (1952) as corrected by Jarrett (1979).

Year	Count	Year	Count	Year	Count	Year	Count
1851	4	1881	2	1911	0	1941	4
1852	5	1882	5	1912	1	1942	2
1853	4	1883	2	1913	1	1943	0
1854	1	1884	2	1914	1	1944	0
1855	0	1885	3	1915	0	1945	0
1856	4	1886	4	1916	1	1946	1
1857	3	1887	2	1917	0	1947	4
1858	4	1888	1	1918	1	1948	0
1859	0	1889	3	1919	0	1949	0
1860	6	1890	2	1920	0	1950	0
1861	3	1891	2	1921	0	1951	1
1862	3	1892	1	1922	2	1952	0
1863	4	1893	1	1923	1	1953	0
1864	0	1894	1	1924	0	1954	0
1865	2	1895	1	1925	0	1955	0
1866	6	1896	3	1926	0	1956	0
1867	3	1897	0	1927	1	1957	1
1868	3	1898	0	1928	1	1958	0
1869	5	1899	1	1929	0	1959	0
1870	4	1900	0	1930	2	1960	1
1871	5	1901	1	1931	3	1961	0
1872	3	1902	1	1932	3	1962	1
1873	1	1903	0	1933	1		
1874	4	1904	0	1934	1		
1875	4	1905	3	1935	2		
1876	1	1906	1	1936	1		
1877	5	1907	0	1937	1		
1878	5	1908	3	1938	1		
1879	3	1909	2	1939	1		
1880	4	1910	2	1940	2		

Table 3: A sequence of 50 observations for the 3–state Markov chain example

1	1	2	2	1	1	1	1	1	1	2	3	3	2	3	2	1	1	2	1
1	3	3	1	1	1	3	2	1	1	1	1	1	3	2	2	3	1	2	2
2	2	2	2	2	3	2	3	2	2										

Table 4. Stagnant band height data; x = log (flow rate in g./cm.sec.), y = log (band height in cm.), from Bacon & Watts (1971)

x	y	x	y	x	y	x	y
0.11	0.44	0.11	0.43	0.34	0.25	−1.08	0.99
−0.80	0.90	0.11	0.43	1.19	−0.65	−1.08	1.03
0.01	0.51	−0.63	0.81	0.59	−0.01	0.44	0.13
−0.25	0.65	−0.63	0.83	0.85	−0.30	0.34	0.24
−0.25	0.67	−1.39	1.12	0.85	−0.33	0.25	0.30
−0.12	0.60	−1.39	1.12	0.99	−0.46		
−0.12	0.59	0.70	−0.13	0.99	−0.43		
−0.94	0.92	0.70	−0.14	0.25	0.33		

Table 5: Generated data for ordered slopes example

x_i	y_i	x_i	y_i	x_i	y_i	x_i	y_i
1.00	−6.71	9.00	12.06	17.00	10.70	25.00	−1.28
2.00	−6.42	10.00	9.90	18.00	9.07	26.00	3.70
3.00	3.85	11.00	15.03	19.00	12.69	27.00	−2.04
4.00	11.25	12.00	9.69	20.00	8.12	28.00	2.00
5.00	11.96	13.00	21.99	21.00	11.18	29.00	−8.35
6.00	−4.58	14.00	14.20	22.00	3.58		
7.00	3.90	15.00	6.04	23.00	10.80		
8.00	13.63	16.00	16.43	24.00	12.76		

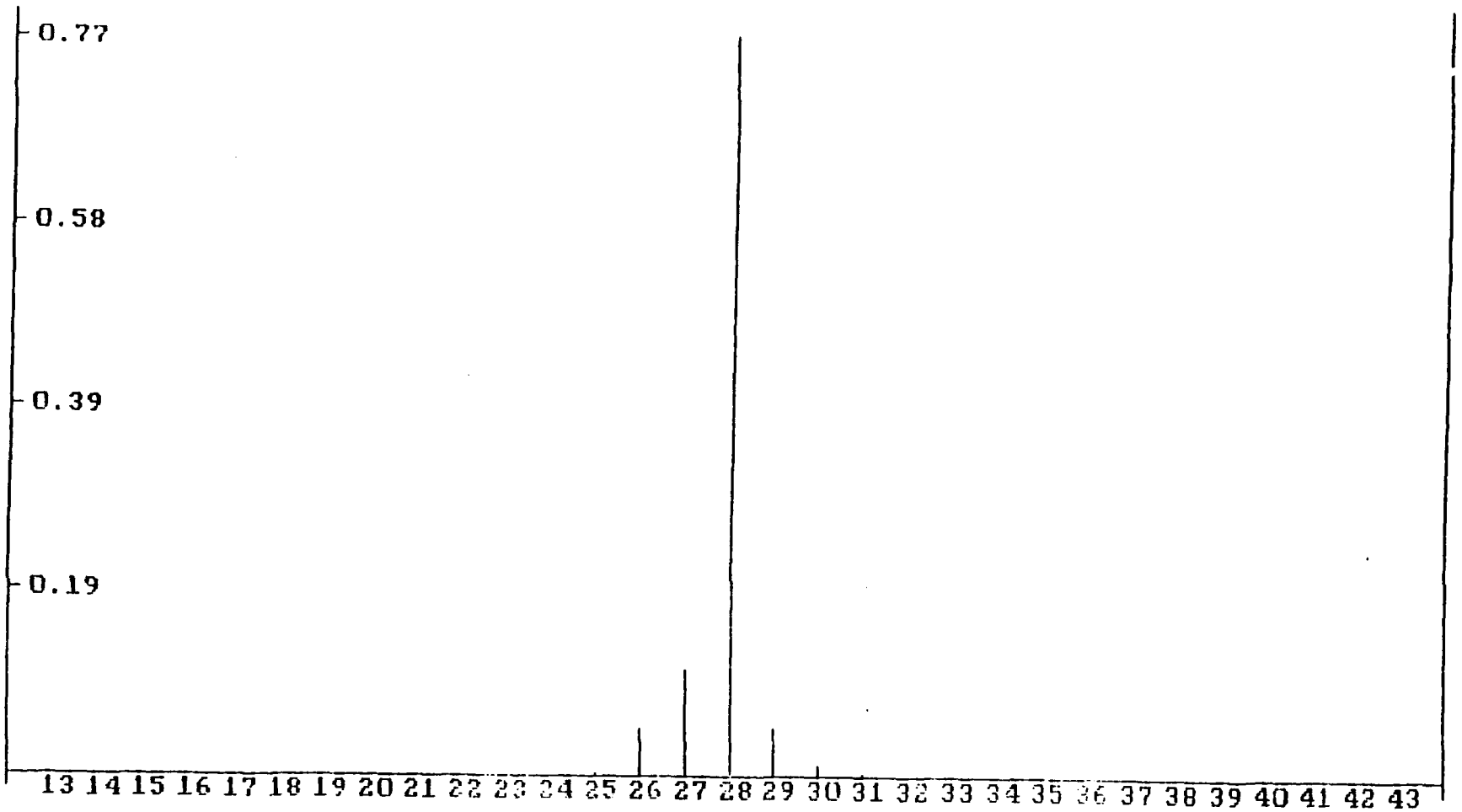

Figure 1: Estimate of the density for $k|Y$ for the 'Nile' data

$a = b = c = d = 2$, $\mu_0 = 10$, $\sigma_0 = 2$

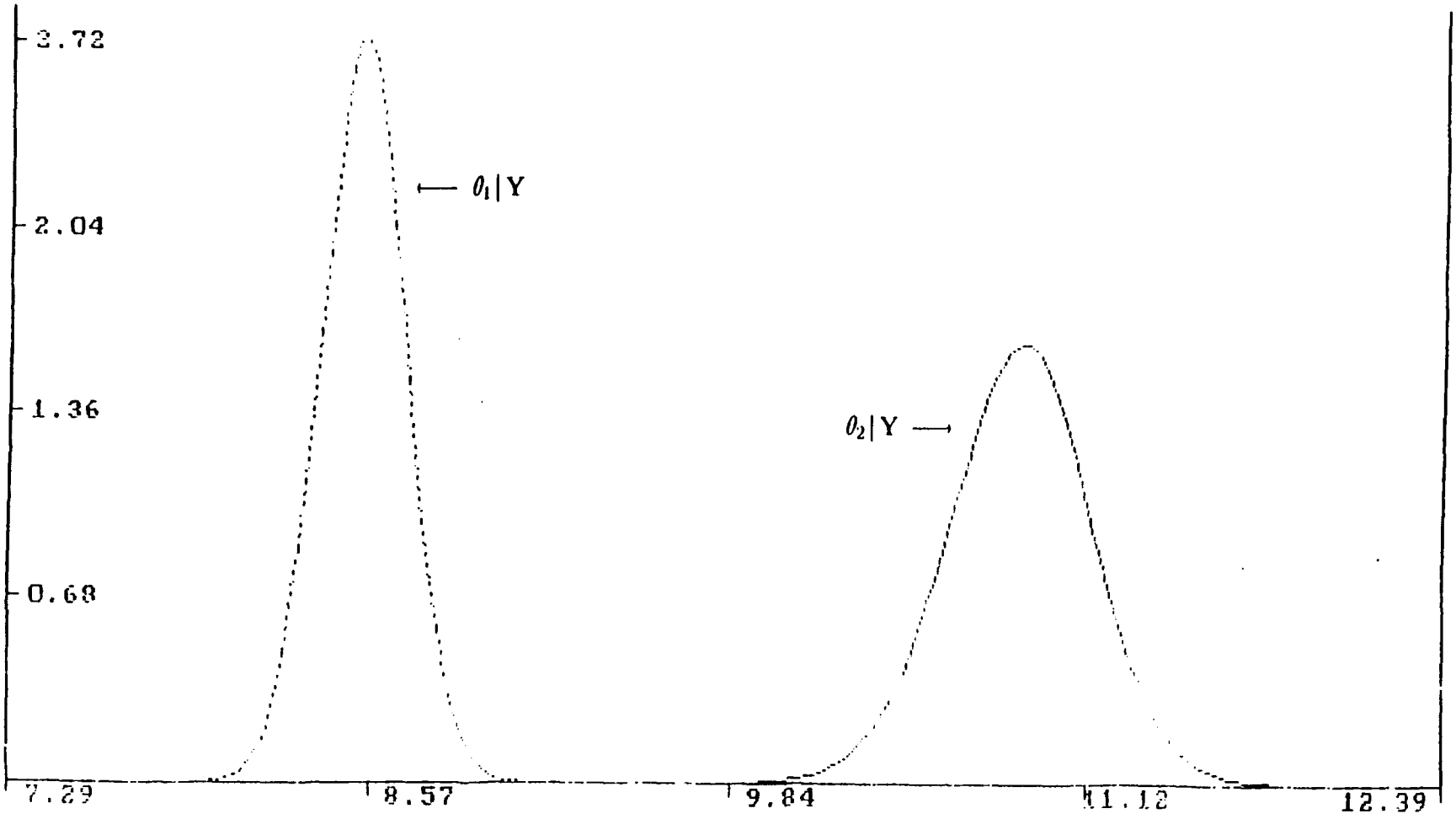

Figure 2: Estimates of the densities for $\theta_1|Y$ and $\theta_2|Y$ for the 'Nile' data

$a = b = c = d = 2$, $\mu_0 = 10$, $\sigma_0 = 2$

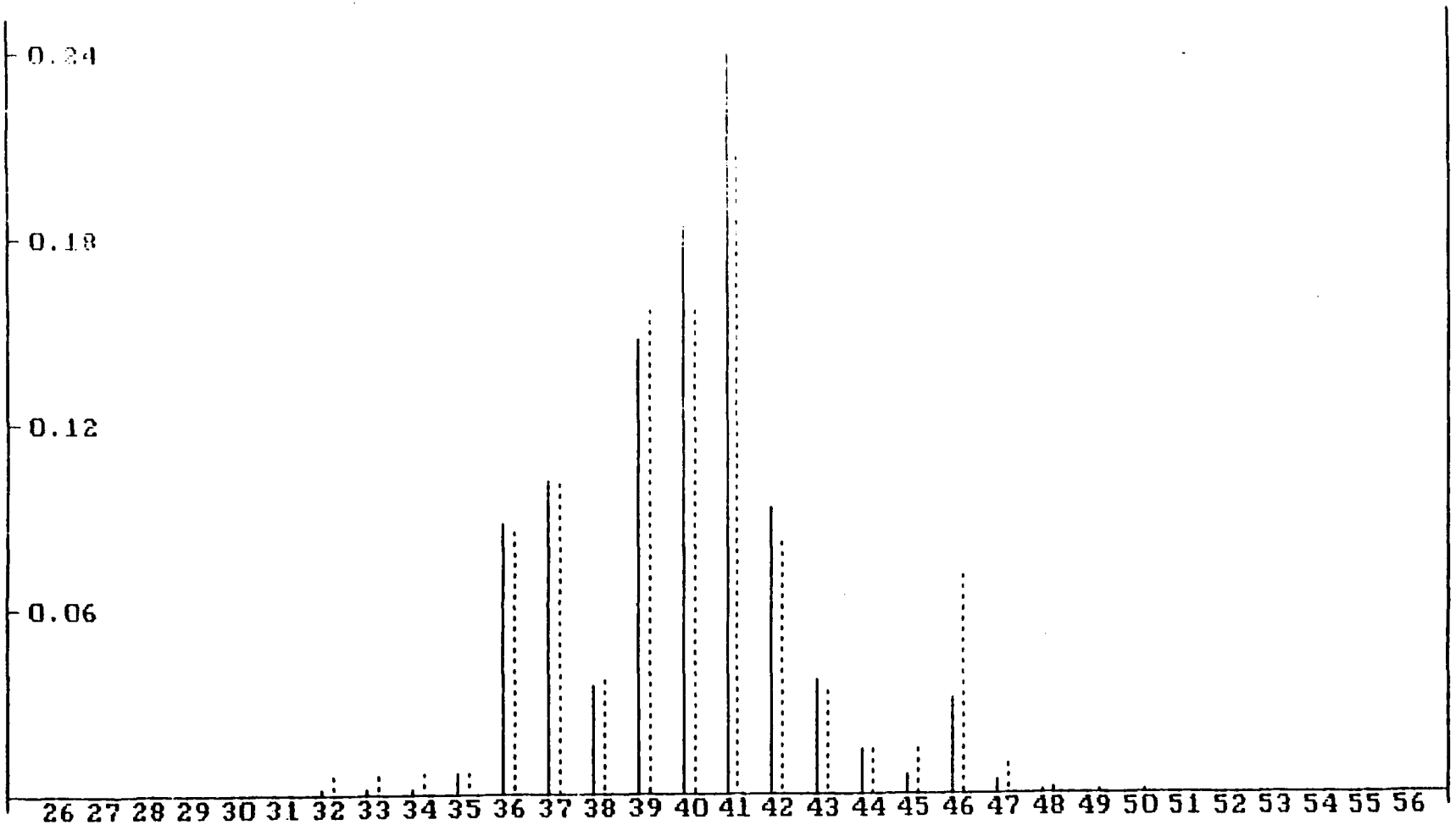

Figure 3: Estimates of the density for k|Y for the 'coal mine disaster' data

$a_1 = a_2 = .5$, $c_1 = c_2 = 0$, $d_1 = d_2 = 1$

full data = _____, 20% missing (reduced) data = ...

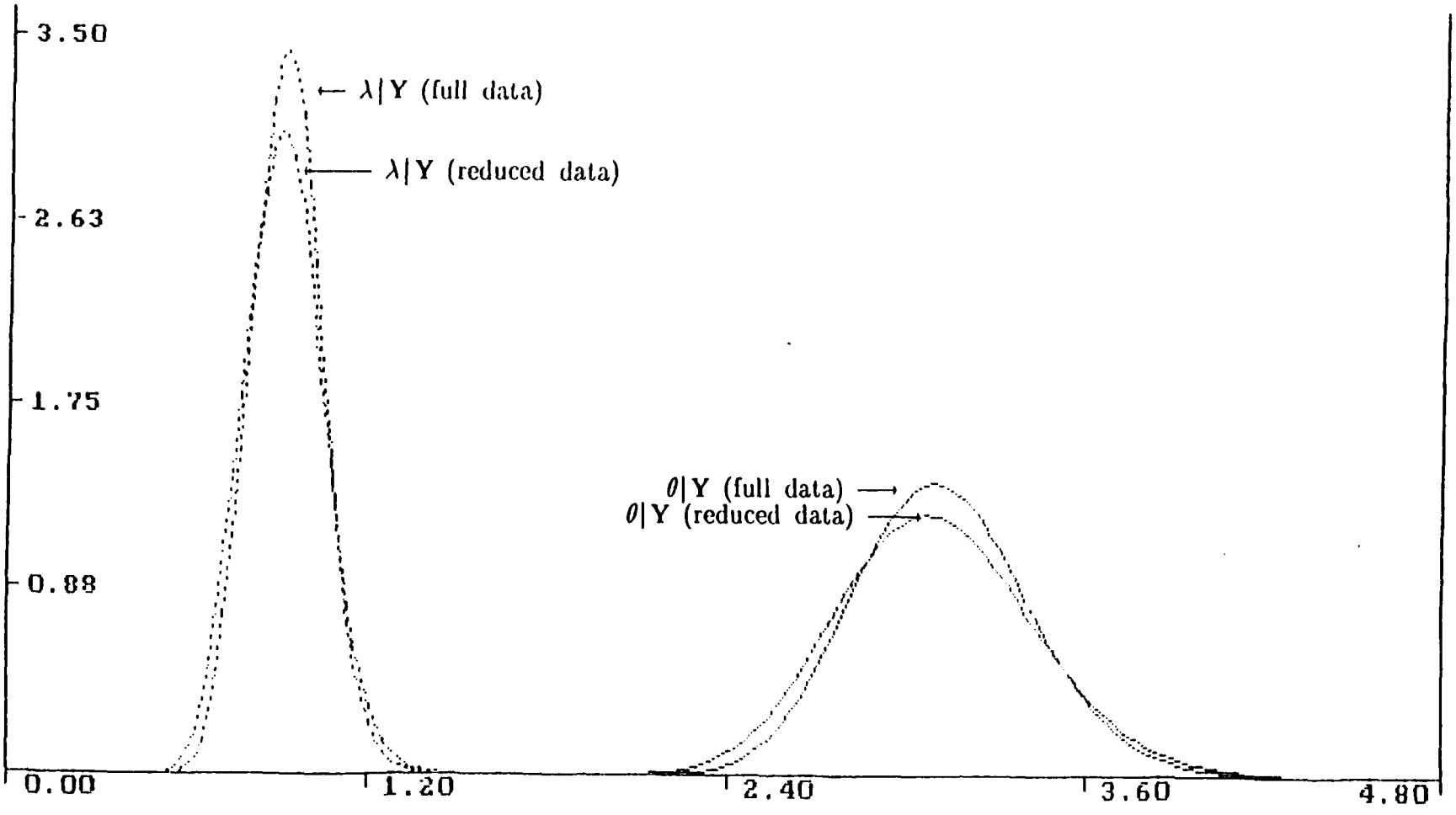

Figure 4: Estimates of the densities for $\theta|Y$ and $\lambda|Y$ for the 'coal mine disaster' data

$a_1 = a_2 = .5$, $c_1 = c_2 = 0$, $d_1 = d_2 = 1$

using full data and 20% missing (reduced) data

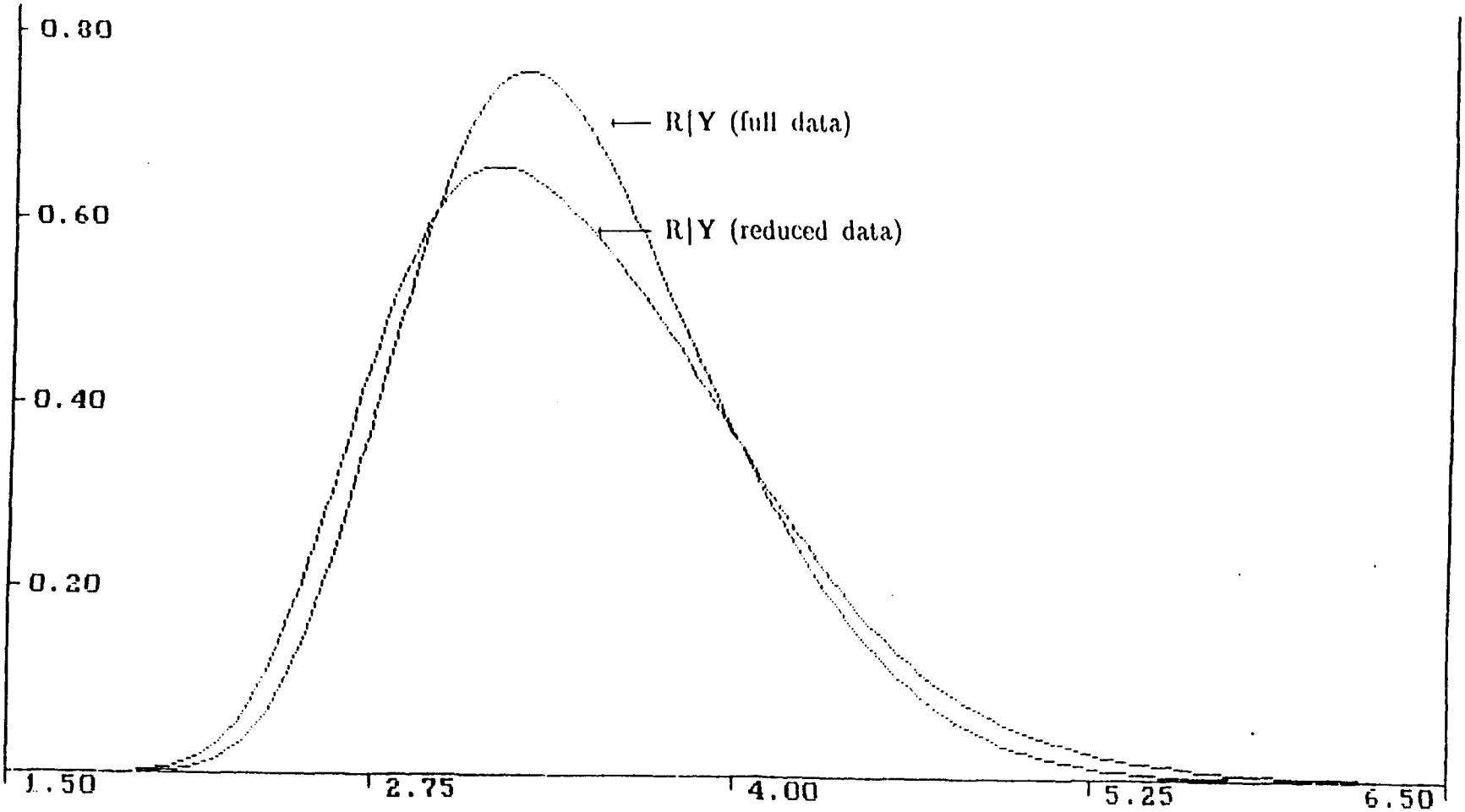

Figure 5: Estimates of the density for R|Y for the 'coal mine disaster' data

$a_1 = a_2 = .5$, $c_1 = c_2 = 0$, $d_1 = d_2 = 1$

using full data and 20% missing (reduced) data

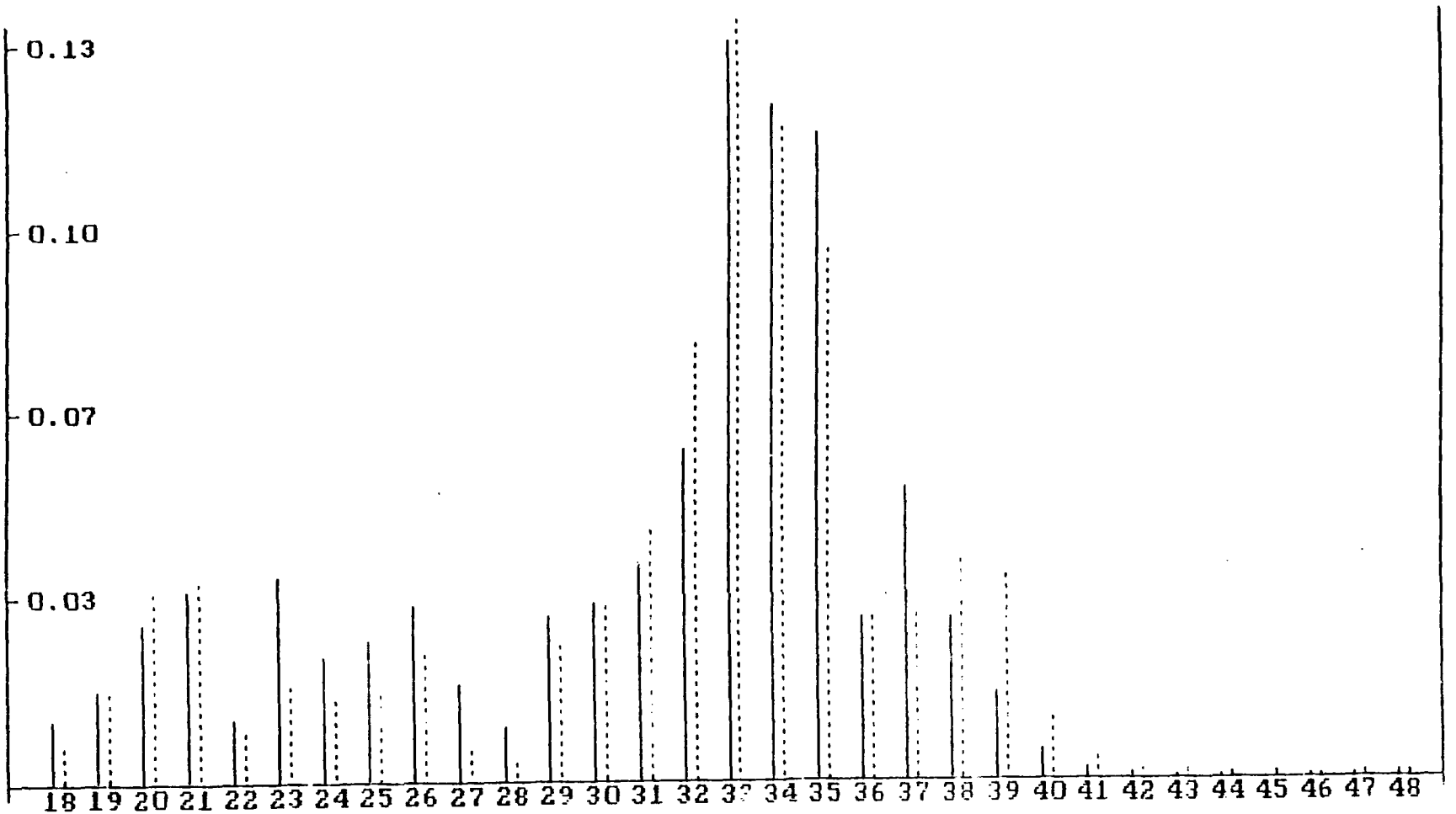

Figure 6: Estimates of the density for k|Y for the Markov chain data, iterations 4 and 5.

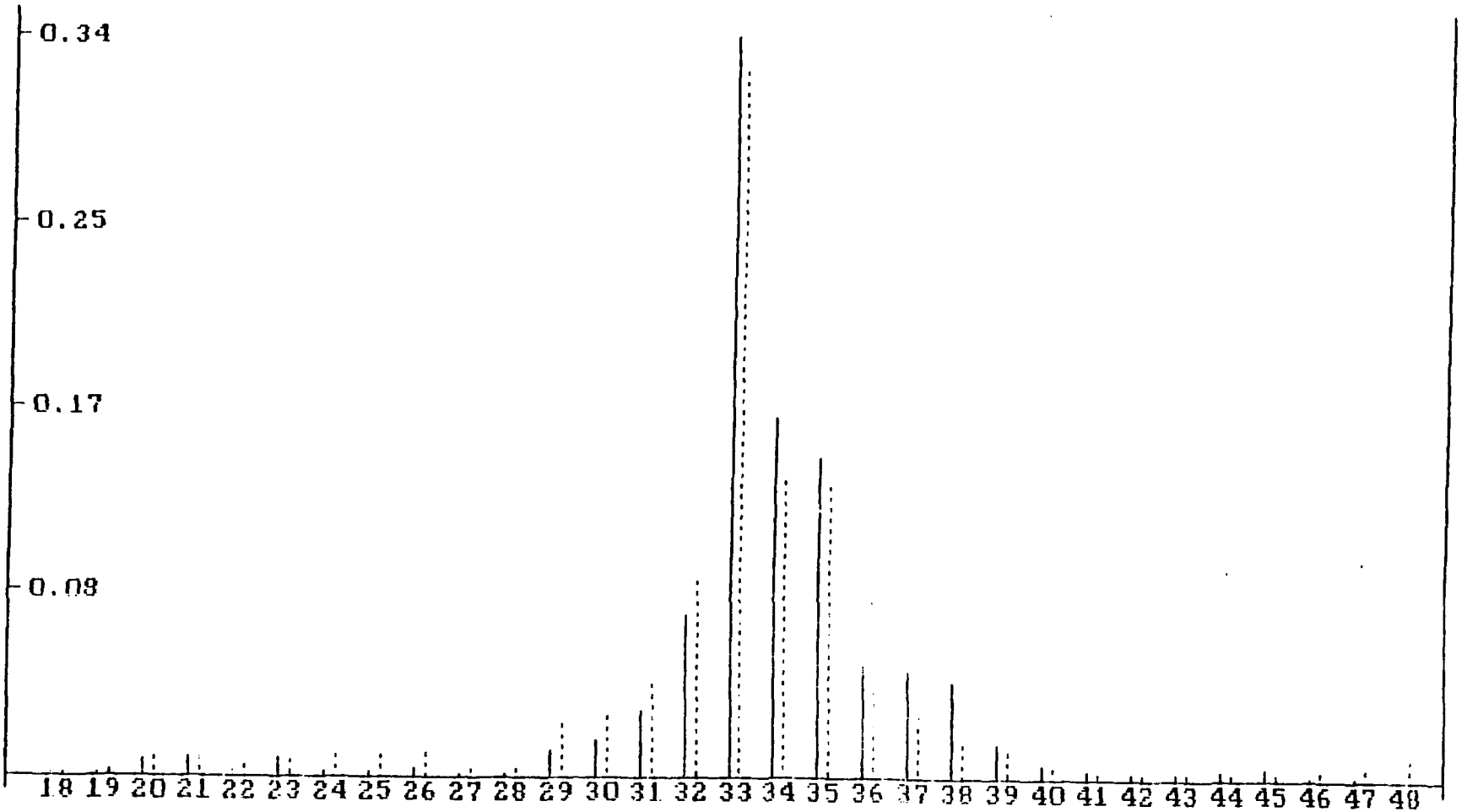

Figure 7: Estimates of the density for k|Y for the Markov chain data, iterations 30 and 31.

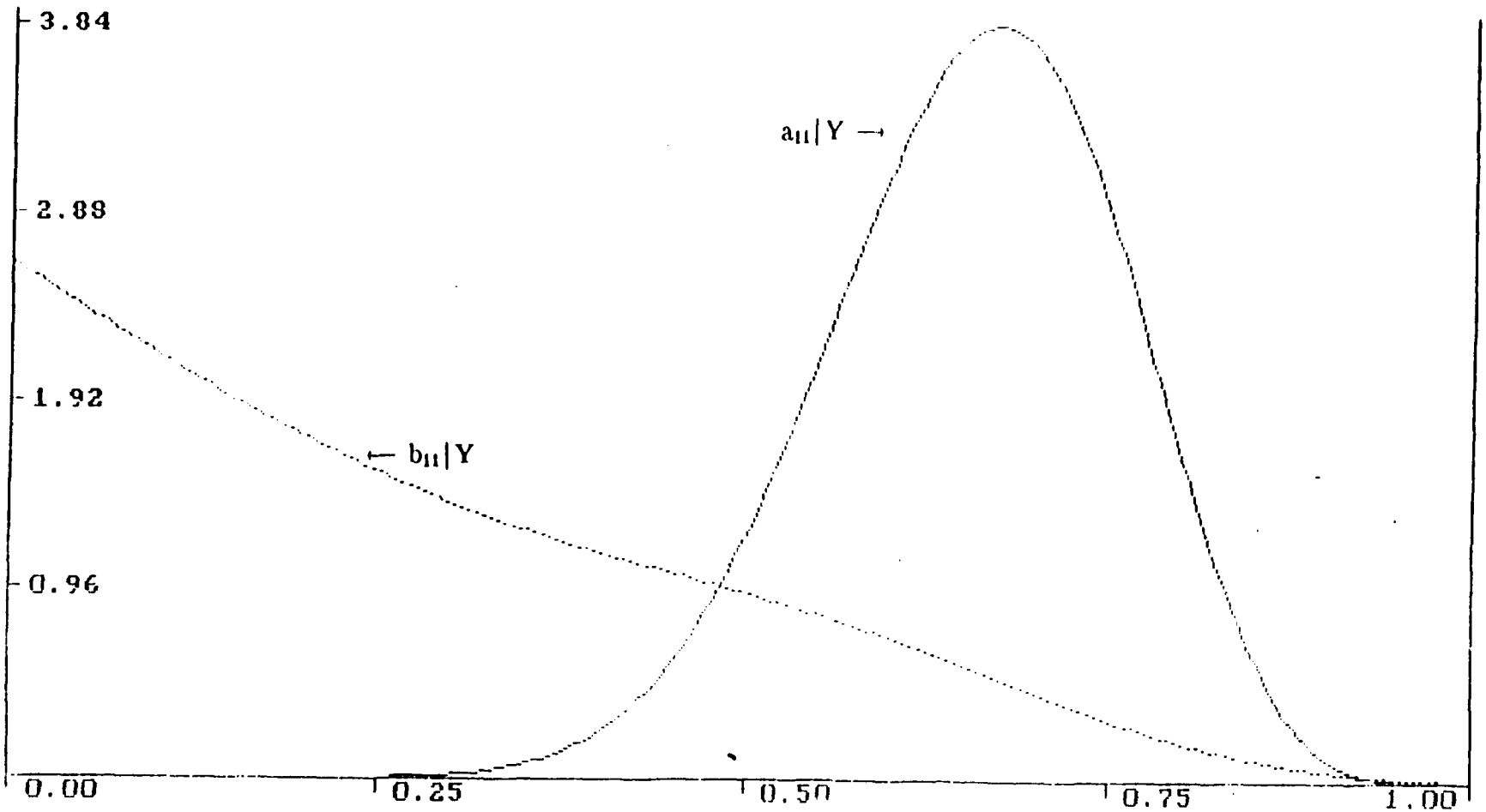

Figure 8: Estimates of the densities for $a_{11}|Y$ and $b_{11}|Y$ for the Markov chain data

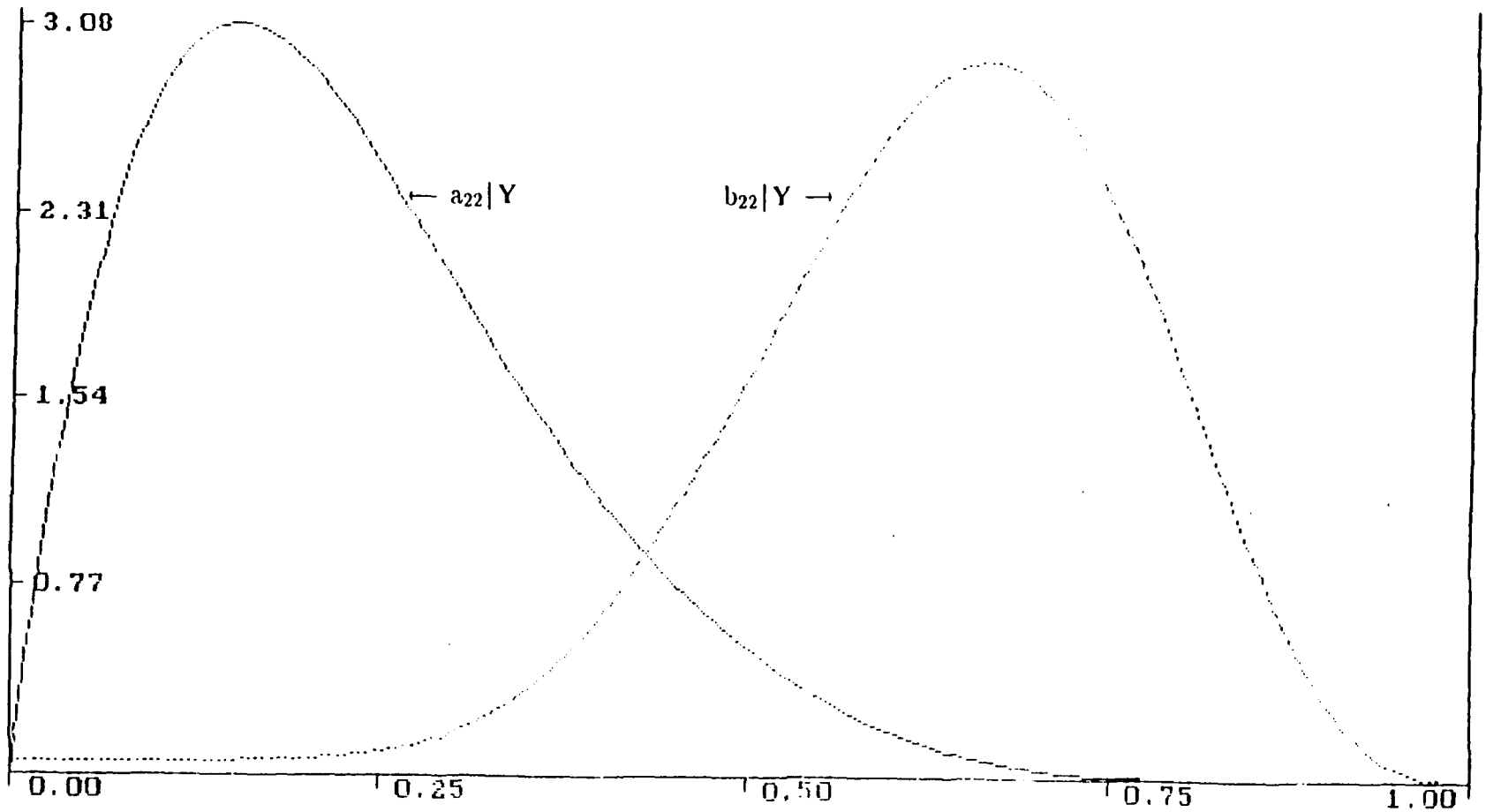

Figure 9: Estimates of the densities for $a_{22}|Y$ and $b_{22}|Y$ for the Markov chain data

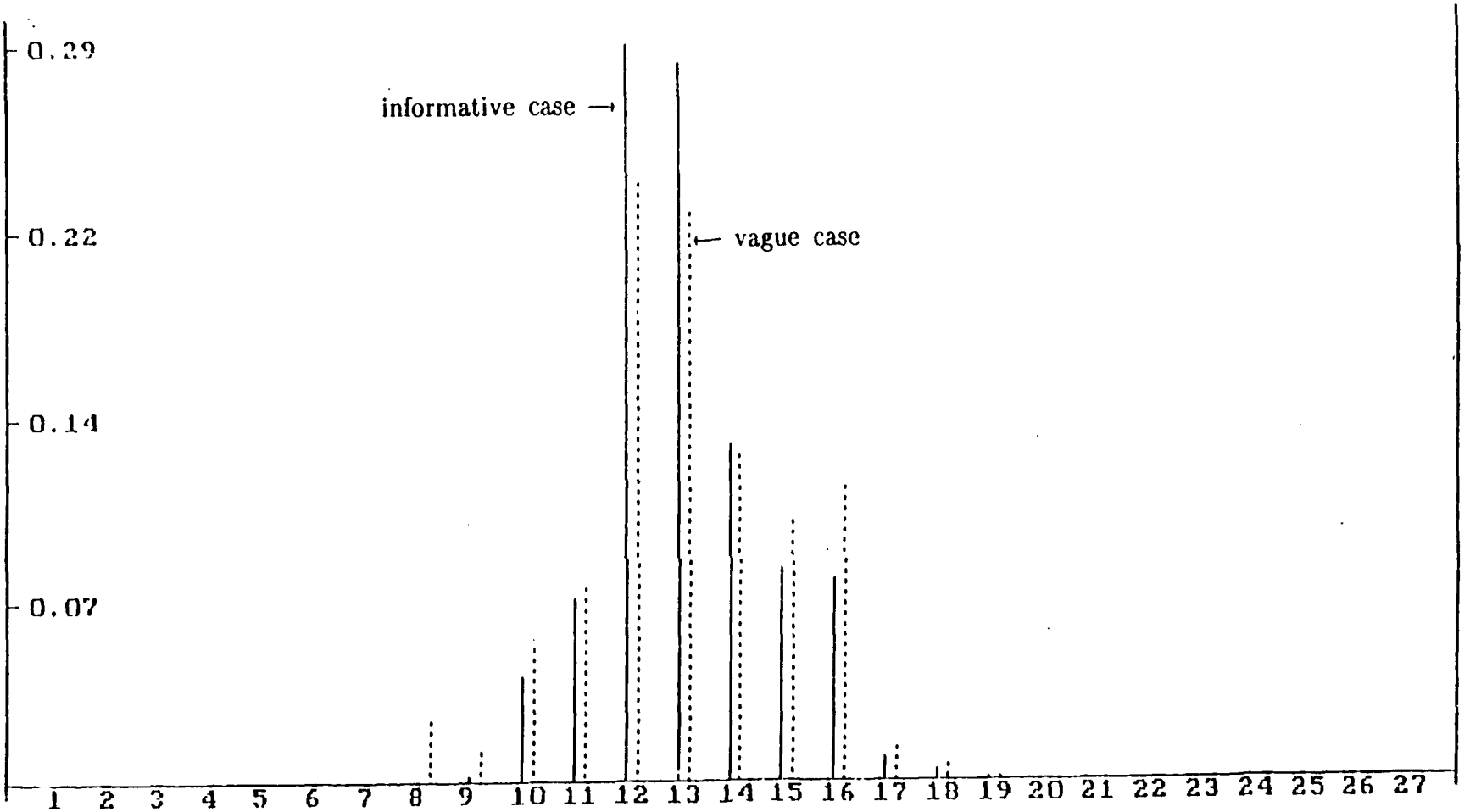

Figure 10: Estimate of the density for k|Y for the Bacon & Watts (1971) data with m = 100 $a_0 = .1$, $b_0 = 100$, informative case $\rho = 4$, $V = \begin{bmatrix} 0.001 & 0 \\ 0 & 0.3 \end{bmatrix}$, vague case $\rho = 2$, $V = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$.

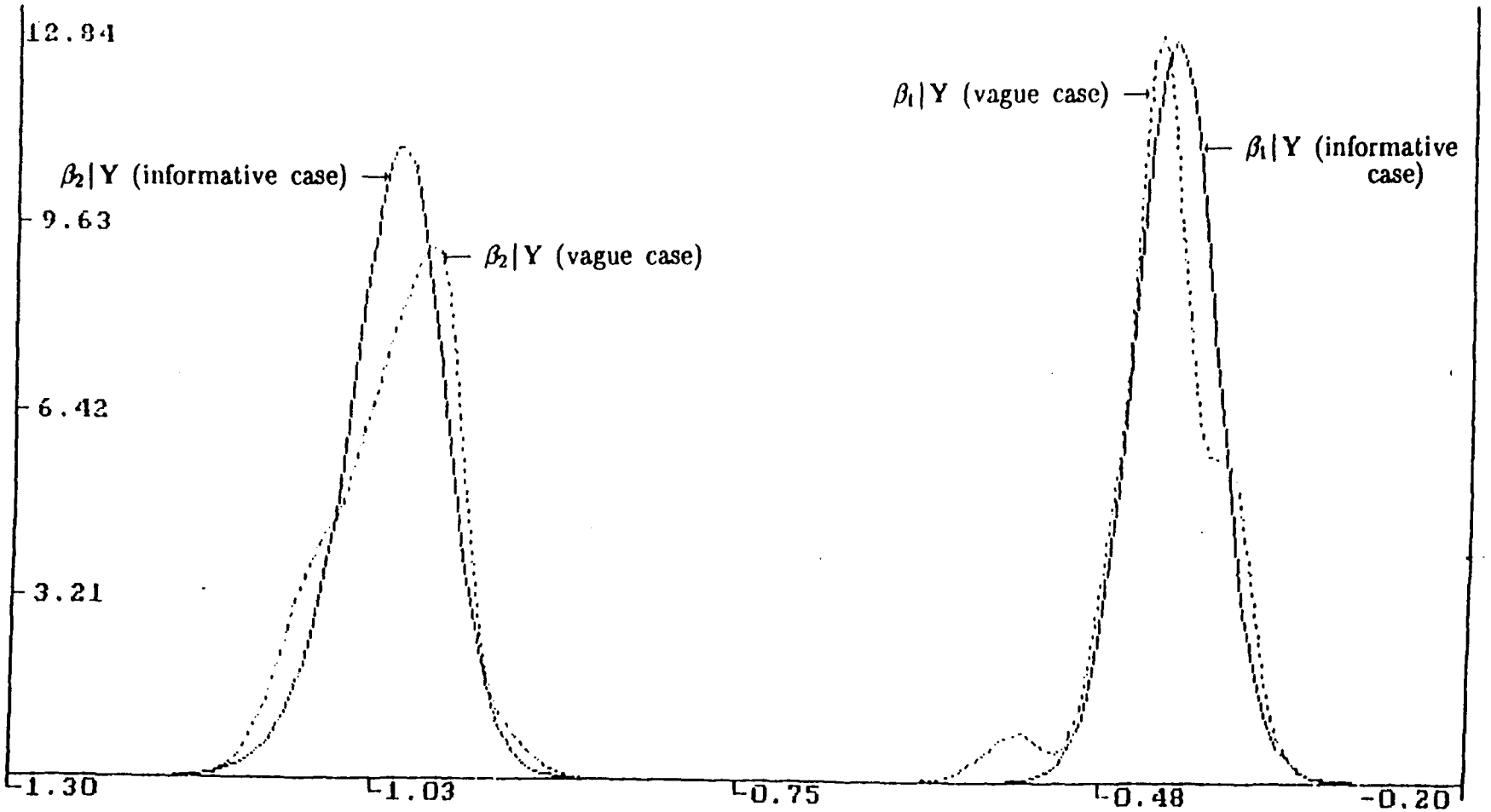

Figure 11: Estimates of the densities for $\beta_1|Y$ and $\beta_2|Y$ for the Bacon & Watts (1971) data with m = 100, $a_0 = .1$, $b_0 = 100$, informative case $\rho = 4$, $V = \begin{bmatrix} 0.001 & 0 \\ 0 & 0.3 \end{bmatrix}$, vague case $\rho = 2$, $V = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$.

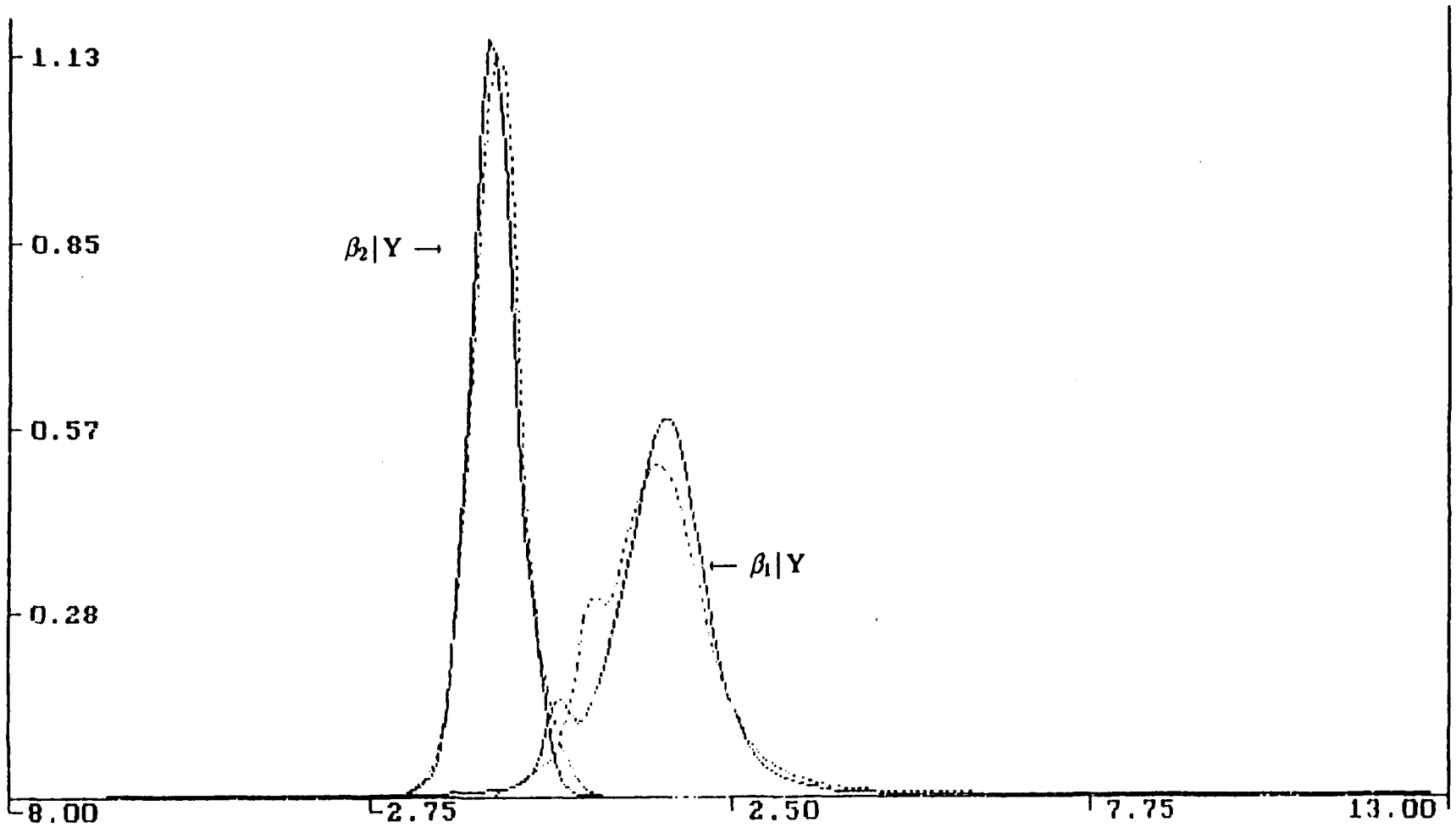

Figure 12: Estimates of the densities for $\beta_1|Y$ and $\beta_2|Y$ for the data in Table 5, unordered = _____, ordered = ...

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 437	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Hierarchical Bayesian Analysis Of Change Point Problems		5. TYPE OF REPORT & PERIOD COVERED TECHNICAL REPORT 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Bradley P. Carlin, Alan E. Gelfand 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 October 18, 1990 13. NUMBER OF PAGES 42
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)

Change points, changing Markov chains, changing Poisson processes, changing regressions, Gibbs sampler, hierarchical Bayes models.

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

A general approach to hierarchical Bayes change point models is presented. In particular desired marginal posterior densities are obtained utilizing the Gibbs sampler, an iterative Monte Carlo method. This approach avoids sophisticated analytic and numerical high dimensional integration procedures. We include application to changing regressions, changing Poisson processes, and changing Markov chains. Within these contexts we handle several previously inaccessible problems.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED