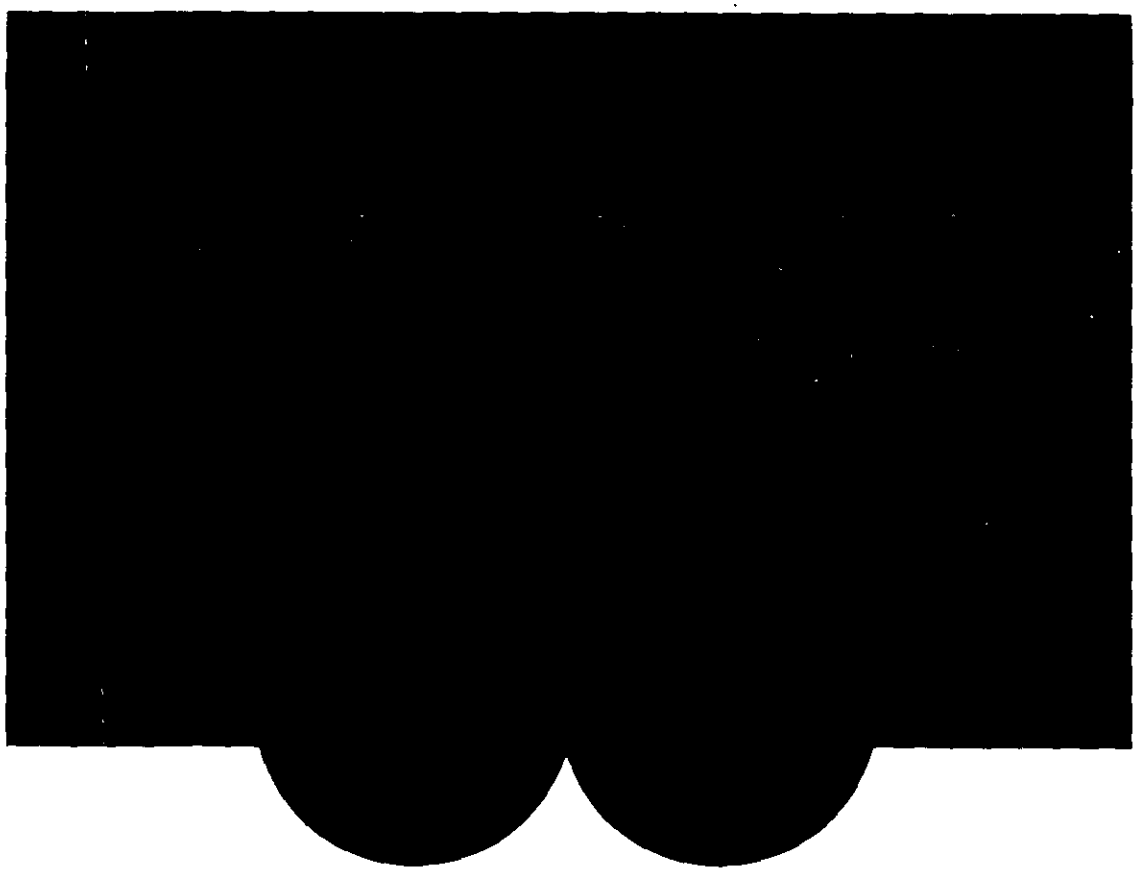


BAYESIAN ANALYSIS OF CHANGE-POINT HAZARD RATE MODELS

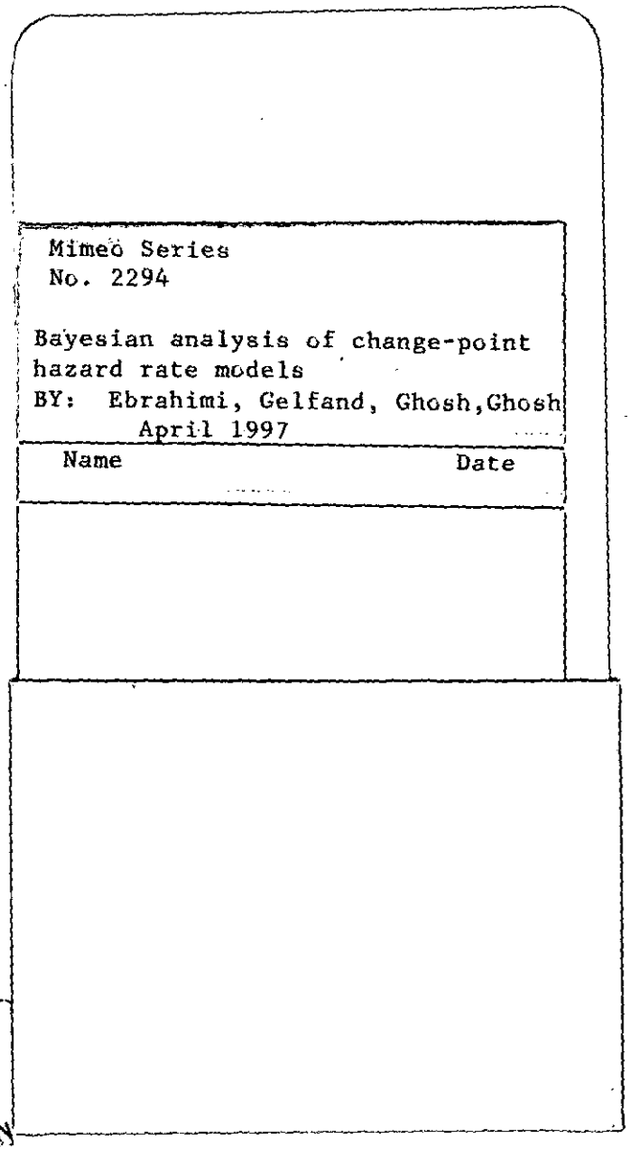
by

Nader Ebrahimi, Alan E. Gelfand, Malay Ghosh, and Sujit K. Ghosh

Institute of Statistics Mimeograph Series No. 2294

April 1997

NORTH CAROLINA STATE UNIVERSITY
Raleigh, North Carolina



BAYESIAN ANALYSIS OF CHANGE-POINT HAZARD RATE MODELS

by

Nader Ebrahimi
Division of Statistics
Northern Illinois University
DeKalb, IL 60115

Alan E. Gelfand
Department of Statistics
University of Connecticut
Storrs, CT 06269

and

Malay Ghosh
Department of Statistics
University of Florida
Gainesville, FL 32611

Sujit K. Ghosh
Department of Statistics
North Carolina State University
Raleigh, NC 27695

Abstract

Change-point hazard rate models arise, for example, in applying "burn-in" techniques to screen defective items and in studying times until undesirable side effects occur in clinical trials. The classical approach develops estimates of model parameters, with particular interest in the threshold or change-point parameter, but exclusively in terms of asymptotic properties. Such asymptotics can be poor for the small to moderate sample sizes often encountered in practice. We propose a Bayesian approach, avoiding asymptotics, to provide more reliable inference conditional only upon the data actually observed. The Bayesian models can be fitted using simulation methods. We develop a very general formulation of the model but also look at special cases which offer particularly simple fitting. We illustrate with an application involving failure times of electrical insulation.

Key Words and Phrases: Burn-in; constant hazard rate, Gibbs sampling; order restrictions; posterior inference; Weibull model.

1 Introduction

In reliability theory a widely accepted procedure is to apply "burn-in" techniques to screen out defective items and improve the lifetimes of surviving items. One helpful tool for capturing "burn-in" is to model the aging process by the hazard function.

More specifically, suppose the lifetimes of items are independent, but that early failures appear to occur at one rate and late failures (after some threshold time) appear to occur at another rate. Let T denote the lifetime with density function $f(t)$ and survival function $\bar{F}(t) = P(T > t)$. Then a suitable form for the hazard function $h(t) = \frac{f(t)}{\bar{F}(t)}$ is

$$h(t) = h_0(t)I(0 \leq t \leq \tau) + h_1(t)I(t > \tau), \tag{1.1}$$

where $I(A) = \begin{cases} 1, & \text{if } x \epsilon A \\ 0, & \text{otherwise} \end{cases}$ and τ is the threshold parameter, $\tau \epsilon R^+$. From (1.1) the cumulative hazard function is

$$H(t) = \int_0^t h(x)dx = H_0(\min(t, \tau)) + [H_1(t) - H_1(\tau)]^+, \tag{1.2}$$

where $[a]^+ = a$ if $a > 0, = 0$ otherwise.

Because, survivors actually have greater resistance to fatal stresses than do newer ones, we assume that $h_0(t) \geq h_1(t)$ for all $t \geq 0$. In (1.1) if τ were known, to implement screening, items could be tested up to time τ and only survivors sold. However, in practice neither τ nor h_0 and h_1 will be known. The goal of this paper is to consider inference in this general case.

The above model is also applicable in analyzing biomedical data. For example, suppose patients in a clinical trial receive a treatment at time 0. The event times may represent the time until undesirable side effects occur, in which case we would have an initial hazard rate,

say $h_0(t)$ and expect a lower hazard rate say h_1 [illegible] some time.

The literature to date focuses primarily on e[illegible] perspective. For instance, Nguyen et al. (1984) c[illegible] $\lambda_1 < \lambda_0$. Basu et al. (1988) extend this to allow a[illegible] also Ebrahimi (1991) and Loader (1991) in this re[illegible] λ_0 are proposed and their asymptotic properties a[illegible] for the small to moderate sample sizes often enc[illegible] We adopt a fully Bayesian approach for this pro[illegible] reliable inference conditional only upon the data [illegible] of the model we obtain an entire posterior distri[illegible] the feature. In order to deflect customary criticis[illegible] in the prior specification, we adopt rather vague [illegible] of a likelihood analysis. We are unaware of any [illegible]

We remark that the hazard rate change poin[illegible] different from the usual change point set up. T[illegible] observations as $1, 2, \cdots$, and seeks to identify th[illegible] model occurred. Change can be considered in [illegible] after a fixed sample size has been collected. In [illegible] discrete valued variable.

The outline of the paper is as follows: In Sec[illegible] a sample of lifetimes collected under (1.1). We [illegible] using simulation methods. In Section 3 we speci[illegible] which yields particularly simple fitting. We al[illegible]

constant hazard rate change point model to a small data set of failure times of electrical insulation in Section 4. Finally we discuss some useful extensions in Section 5.

2 Bayesian formulation of the problem

Returning to (1.1), we adopt fully parametric modeling assuming $h_0(t) = h_0(t;\boldsymbol{\theta}_0)$ and $h_1(t) = h_1(t;\boldsymbol{\theta}_1)$. That is, h_0 and h_1 are two, possibly distinct, parametric families of hazard functions indexed by $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_1$. To capture the order restriction on h_0 and h_1 we let $S = \{(\boldsymbol{\theta}_0, \boldsymbol{\theta}_1) : h_0(t;\boldsymbol{\theta}_0) \geq h_1(t;\boldsymbol{\theta}_1) \text{ for all } t \geq 0\}$.

For an uncensored sample $T_1, T_2, \ldots T_n$ of lifetimes the likelihood takes the form

$$L(\boldsymbol{\theta}_0, \boldsymbol{\theta}_1, \tau; \boldsymbol{t}) = \prod_{j=1}^{n} h(t_j) \exp(-H(t_j)), \tag{2.1}$$

where $\boldsymbol{t} = (t_1, \ldots, t_n)$ denotes the observed values of the lifetimes. Let $v_j(\tau) = 1$ if $t_j \leq \tau, = 0$ if $t_j > \tau$. Then (2.1) becomes

$$\begin{aligned} L(\boldsymbol{\theta}_0, \boldsymbol{\theta}_1, \tau; \boldsymbol{t}) &= \prod_{j=1}^{n} h_0(t_j;\boldsymbol{\theta}_0)^{v_j(\tau)} h_1(t_j;\boldsymbol{\theta}_1)^{1-v_j(\tau)} \\ &\times \exp\{-\sum_{j=1}^{n}(H_0(\min(t_j,\tau);\boldsymbol{\theta}_0) + [H_1(t_j;\boldsymbol{\theta}_1) - H_1(\tau;\boldsymbol{\theta}_1)]^+)\} I(S). \end{aligned} \tag{2.2}$$

For our model we need to uniquely define the notion of "no change". That is, $\tau < t_j \ \forall j$ is not distinguishable from $\tau \geq t_j \ \forall j$; in either case there is no change-point during the period of observation. As a simple remedy, if we order the observation times, $t_{(1)} < t_{(2)} < \ldots < t_{(n)}$, we restrict the likelihood so that $\tau \geq t_{(1)}$. Certainty of a change-point during the period of observation would then add the further restriction, $\tau < t_{(n)}$.

To create a Bayesian model we require a prior specification for $\boldsymbol{\theta}_0, \boldsymbol{\theta}_1$ and τ. We assume that it takes the general form

$$f(\boldsymbol{\theta}_0, \boldsymbol{\theta}_1) \cdot I(S) \cdot f(\tau) \tag{2.3}$$

and that it is proper which assures that the posterior $f(\boldsymbol{\theta}_0, \boldsymbol{\theta}_1, \tau|\boldsymbol{t})$ is proper. Our prior information on τ places it on the interval $(0, b)$ with b possibly ∞. The actual support for τ is truncated according to the restrictions imposed by the likelihood. When τ is not bounded above we argue below that if $f(\tau)$ is improper the posterior must necessarily be improper. We could not assume that, e.g., $f(\tau) = 1$ or $f(\tau) = \tau^{-1}$.

Combining (2.2) and (2.3) provides the complete Bayesian specification and thus the posterior $f(\boldsymbol{\theta}_0, \boldsymbol{\theta}_1, \tau|t)$ which is proportional to

$$L(\boldsymbol{\theta}_0, \boldsymbol{\theta}_1, \tau; \boldsymbol{t}) \cdot f(\boldsymbol{\theta}_0, \boldsymbol{\theta}_1) I(S) \cdot f(\tau). \tag{2.4}$$

The posteriors $f(\boldsymbol{\theta}_0|\boldsymbol{t})$ and $f(\boldsymbol{\theta}_1|\boldsymbol{t})$ enable us to learn about the pre and post threshold hazards. In fact for each t, since $h_0(t; \boldsymbol{\theta}_0), H_0(t; \boldsymbol{\theta}_0), h_1(t; \boldsymbol{\theta}_1), H_1(t; \boldsymbol{\theta}_1), h(t; \boldsymbol{\theta}_0, \boldsymbol{\theta}_1, \tau)$ and $H(t; \boldsymbol{\theta}_0, \boldsymbol{\theta}_1, \tau)$ are all random variables, they all have posterior distributions which would be of interest as well. However, primary interest is in the posterior for $\tau, f(\tau|\boldsymbol{t})$ and when a change is not certain, $P(\tau > t_{(n)}|\boldsymbol{t})$.

The expression in (2.4) is not analytically tractable so we turn to simulation approaches for fitting such a model. That is, we seek to draw samples from the posterior in (2.4) in order to learn about its features. Customary iterative approaches using, e.g., a rejection method or a weighted bootstrap, see Smith and Gelfand (1992), are difficult here. A good importance sampling density is required, as it would be for Monte Carlo integration. But given the way that $\boldsymbol{\theta}_0, \boldsymbol{\theta}_1$ and τ are intertwined in (2.2), this will be difficult to develop. However, the form in (2.4) is nicely suited for Markov chain Monte Carol simulation using Gibbs sampling updating (Gelfand and Smith, 1990). To clarify we consider the full conditional distriubtions for $\boldsymbol{\theta}_0$, for $\boldsymbol{\theta}_1$ and for τ.

In particular, we immediately have $f(\boldsymbol{\theta}_0|\boldsymbol{\theta}_1, \tau, \boldsymbol{t})$ proportional to

$$\prod_{j=1}^{n} h_0(t_j;\boldsymbol{\theta}_0)^{v_j(\tau)} \cdot f(\boldsymbol{\theta}_0,\boldsymbol{\theta}_1) \cdot I(S) \cdot \exp\{-\sum_{j=1}^{n} H_0(\min(t_j,\tau);\boldsymbol{\theta}_0\} \tag{2.5}$$

and $f(\boldsymbol{\theta}_1|\boldsymbol{\theta}_0,\tau,\boldsymbol{t})$ proportional to

$$\prod_{j=1}^{n} h_1(t_j,\boldsymbol{\theta}_0)^{1-v_j(\tau)} f(\boldsymbol{\theta}_0,\boldsymbol{\theta}_1) I(S) \exp\{-\sum_{j=1}^{n} [H_1(t_j;\boldsymbol{\theta}_1), -H_1(\tau;\boldsymbol{\theta}_1)]^+\}. \tag{2.6}$$

In special cases (2.5) and (2.6) will be standard distributions so that sampling will be routine (see Section 3 for example). Generally, they will not be so that tailored strategies depending upon the respective forms will be required. Such tailoring is discussed for example in Gilks et al. (1996).

As for τ, suppose $t_{(r-1)} < \tau < t_{(r)}, r = 1, \cdots n+1$ (with $t_{(0)} = 0$ and $t_{(n+1)} = \infty$). Then, on this interval $f(\tau|\boldsymbol{\theta}_0,\boldsymbol{\theta}_1,\boldsymbol{t})$ is proportional to

$$k_r \exp\{-(n-r+1)(H_0(\tau,\boldsymbol{\theta}_0) - H_1(\tau,\boldsymbol{\theta}_1))\} \cdot f(\tau) \tag{2.7}$$

where $k_r = \prod_{j=1}^{r-1} h_0(t_{(j)};\boldsymbol{\theta}_0) \prod_{j=r}^{n} h_1(t_{(j)};\boldsymbol{\theta}_1) \cdot \exp\{-\sum_{j=1}^{r-1} H_0(t_{(j)};\boldsymbol{\theta}_0) - \sum_{j=r+1}^{n} H_1(t_{(j)};\boldsymbol{\theta}_1)\}$. (In defining k_r, products over an empty set of terms are set to 1, sums are set to 0. We also suppress the dependence on $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_1$). Hence, let $c_r = k_r \int_{t_{(r-1)}}^{t_{(r)}} f(\tau) \exp\{-(n-r+1)(H_0(\tau,\boldsymbol{\theta}_0) - H_1(\tau,\boldsymbol{\theta}_1))\} d\tau$, $c = \sum_{r=1}^{n+1} c_r$ and $p_r = c_r/c$. To sample τ we first randomly select an interval, i.e., we select $t_{(r-1)} < \tau < t_{(r)}$ with probability p_r. We then draw τ in this interval from the normalized version of (2.7). That is, from the density which is (2.7) devided by c_r. Note that k_r cancels out of this density. In special cases the integration for c_r can be done explicitly in which case the density over $t_{(r-1)} < \tau < t_{(r)}$ is routine to sample. (See Section 3 again for example.) Regardless, the univariate density in (2.7) will not be hard to sample. See again Gilks et al. (1996) or perhaps Devroye (1986).

We return to the comment earlier in this section where we asserted that if τ is not bounded above and $f(\tau)$ is improper so must the posterior be. More explicitly, suppose for each t, $\int_t^\infty f(\tau)d\tau = \infty$. Then

$$\int_0^\infty \int_{\Theta_0} \int_{\Theta_1} f(\boldsymbol{\theta}_0, \boldsymbol{\theta}_1, \tau|\boldsymbol{t}) d\boldsymbol{\theta}_1 d\boldsymbol{\theta}_0 d\tau$$

$$\geq \int_{t_{(n)}}^\infty \int_{\Theta_0} \int_{\Theta_1} f(\boldsymbol{\theta}_0, \boldsymbol{\theta}_1, \tau|\boldsymbol{t}) d\boldsymbol{\theta}_1 d\boldsymbol{\theta}_0 d\tau$$

$$= \int_{t_{(n)}}^\infty [\int_{\Theta_0} \int_{\Theta_1} k_{n+1}(\boldsymbol{\theta}_0, \boldsymbol{\theta}_1) f(\boldsymbol{\theta}_0, \boldsymbol{\theta}_1) I(S) d\boldsymbol{\theta}_1 d\boldsymbol{\theta}_0] \cdot f(\tau) d\tau.$$

Since the term in brackets is free of τ, the last integral does not exist.

We conclude this section by noting that h_0 and h_1 may have common parameters. In particular, this may be done to simplify the set S. For instance, if h_0 and h_1 are taken to be Weibull hazards and we require $h_0(t) = \lambda_0\gamma_0 t^{\gamma_0 - 1} > \lambda_1\gamma_1 t^{\gamma_1 - 1} = \lambda_1(t)$ for all $t \geq 0$ this necessitates $\gamma_0 = \gamma_1$, i.e., $h_0(t) = \lambda_0\gamma t^{\gamma-1}$ and $h_1(t) = \lambda_1\gamma t^{\gamma-1}$ with $\lambda_0 > \lambda_1$. The foregoing distribution theory can be directly extended to handle the general situation. The Weibull case is detailed in Section 3.

3 Two special cases

First we consider the case where $h_0(t) = \lambda_0, h_1(t) = \lambda_1, \lambda_0 > \lambda_1$. Classical analysis for this case using asymptotic results is presented in Nguyen et al. (1984) and Loader (1991). For an observed sample of values $t_1, \cdots, t_n$, (2.2) becomes

$$\lambda_0^{\sum_j I(t_j \leq \tau)} \lambda_1^{\sum_j I(t_j > \tau)} I(\lambda_0 > \lambda_1) \times \exp\{-\sum_j (\lambda_0 \min(t_j, \tau) + \lambda_1 [t_j - \tau]^+)\} \qquad (3.1)$$

For the prior on λ_0 and λ_1 we take conjugate independent gamma specifications restricted to $\{\lambda_0 > \lambda_1\}$, i.e., $f(\lambda_0, \lambda_1)$ is proportional to

$$\lambda_0^{a_0-1}\lambda_1^{a_1-1}\exp(-b_0\lambda_0 - b_1\lambda_1)I(\lambda_0 > \lambda_1). \tag{3.2}$$

In applications we choose the a_i and b_i so that the prior provides little information. For instance, if we were confident that $\tau < \tilde{\tau}$, we might choose a_0 and b_0 to yield mean $(\tilde{\tau}/4)^{-1}$ with very large variance (we used 10^3 for our illustration in Section 4) and a_1 and b_1 to yield mean $(3\tilde{\tau}/4)^{-1}$ again with very large variance. If the likelihood restricts $\tau < t_{(n)}$ we can set $\tilde{\tau} = t_{(n)}$. We prefer adopting proper specifications rather than improper forms such as $f(\lambda_0, \lambda_1) = (\lambda_0\lambda_1)^{-1}I(\lambda_0 > \lambda_1)$ or $f(\lambda_0, \lambda_1) = I(\lambda_0 > \lambda_1)$ in order to insure proper posteriors and encourage well behaved Markov chain Monte Carlo algorithms.

From (3.1) and (3.2) we see immediately that

$$f(\lambda_0|\lambda_1, \tau; \boldsymbol{t}) = Ga(a_0 + \sum_{j=1}^{n} I(t_j \le \tau), b_0 + \sum_{j=1}^{n} \min(t_j, \tau)) \cdot I(\lambda_0 > \lambda_1)$$

and

$$f(\lambda_1|\lambda_0, \tau; \boldsymbol{t}) = Ga(a_1 + \sum_{j=1}^{n} I(t_j > \tau), b_1 + \sum_{j=1}^{n} [t_j - \tau]^+) \cdot I(\lambda_0 > \lambda_1).$$

Both of these truncated Gamma densities are routine to sample.

For the prior on τ, we illustrate with the case of $f(\tau) = (t_{(n)} - t_{(1)})^{-1}, t_{(1)} < \tau < t_{(n)}$. Then (2.7) becomes $k_r \exp(-(\lambda_0 - \lambda_1)(n - r + 1)\tau), t_{(r-1)} < \tau < t_{(r)}$ with $k_r = (\lambda_0/\lambda_1)^r \exp\{-\lambda_0 \sum_{j=1}^{r-1} t_{(j)} - \lambda_1 \sum_{j=r}^{n} t_{(j)}\}$. Obviously, c_r hence p_r, can be computed explicitly so choosing an interval at random for τ is easily done. Then we can draw τ within this interval by simple c.d.f. inversion.

The Weibull case mentioned at the end of Section 2 adds an interesting wrinkle. Now we have $h_0(t) = \lambda_0\gamma t^{\gamma-1}$ and $h_1(t) = \lambda_1\gamma t^{\gamma-1}$ with $\lambda_0 > \lambda_1$. Again taking $f(\lambda_0, \lambda_1)$ as in (3.2), we assume γ is, a priori independent of λ_0 and λ_1, with $Ga(c, d)$. It is doubtful that we really believe γ to be independent of λ_0 and λ_1 but since it is difficult and arbitrary

to model dependence we prefer to let the data inform about this through the posterior. In application, we would choose c and d to give mean 1 (roughly expressing prior indifference as to whether the h_i are increasing or decreasing), again with very large variance.

Hence, after some manipulation, $f(\lambda_0, \lambda_1, \gamma, \tau|\boldsymbol{t})$ is proportional to

$$\lambda_0^{a_0+\sum I(t_j\le\tau)-1} \cdot \lambda_1^{a_1+\sum I(t_j>\tau)-1} \cdot \gamma^{c+n-1}$$
$$\cdot I(\lambda_0 > \lambda_1) \cdot \exp\{\gamma \sum_{j=1}^{n} \log t_j - b_0\lambda_0 - b_1\lambda_1 - d\gamma$$
$$-\lambda_0 \sum_{j=1}^{n} \min(t_j^\gamma, \tau^\gamma) - \lambda_1 \sum_{j=1}^{n} [t_j^\gamma - \tau^\gamma]^+\}. \tag{3.3}$$

From (3.3), analogous to the previous example we see that $f(\lambda_0|\lambda_1, \gamma, \tau, \boldsymbol{t}) = Ga(a_0 + \sum_{j=1}^{n} I(t_j \le \tau), b_0 + \sum_{j=1}^{n} \min(t_j^\gamma, \tau^\gamma))$ and $f(\lambda_1|\lambda_0, \gamma, \tau, \boldsymbol{t}) = Ga(a_1 + \sum_{j=1}^{n} I(t_j > \tau), b_1 + \sum_{j=1}^{n} [t_j^\gamma - \tau^\gamma]^+)$. As for γ, it is easiest to again pin down τ. If $t_{(r-1)} < \tau < t_{(r)}$ we obtain $f(\gamma|\lambda_0, \lambda_1, \tau, \boldsymbol{t})$ proportional to

$$\gamma^{c+n-1} \exp\{\gamma(\sum \log t_{(j)} - d) - \lambda_0 \sum_{j=1}^{r-1} e^{\gamma \log t_{(j)}} - \lambda_1 \sum_{j=r}^{n} e^{\gamma \log t_{(j)}} - (\lambda_0 - \lambda_1)(n - r + 1)\tau^\gamma\}. \tag{3.4}$$

The density in (3.4) is obviously nonstandard. We suggest sampling it using a Metropolis step within the Gibbs sampler as proposed in, e.g. Müller (1995). In fact we can create a good proposal density for (3.4) as follows.

First note that the function $\varphi(s) = (1+\epsilon)s - \epsilon e^s, \epsilon > 0$ is concave in s with maximum at $s_0 = \log(\epsilon^{-1}(1+\epsilon)) > 0$. Hence $s - \epsilon e^s < -\epsilon s + \varphi(s_0)$. Identifying s_j as $\gamma \log t_{(j)}$ and taking $\epsilon = \lambda_0$ if $j \le r-1$ and $\epsilon = \lambda_1$ if $j \ge r$ yields

$$\gamma \sum \log t_{(j)} - \lambda_0 \sum_{j=1}^{r-1} e^{\gamma \log t_{(j)}} - \lambda_1 \sum_{j=r}^{n} e^{\gamma \log t_{(j)}}$$
$$-\gamma(\lambda_0 \sum_{j=1}^{r-1} \log t_{(j)} + \lambda_1 \sum_{j=r}^{n} \log t_{(j)}) + const.$$

Hence, as a function of γ, the exponential term in (3.4) is bounded above

$$\exp\{-\gamma(\lambda_0 \sum_{j=1}^{r-1} \log t_{(j)} + \lambda_1 \sum_{j=r}^{n} \log t_{(j)} - d)\}.$$

A proposal density for γ becomes $Ga(c+n, d+\lambda_0 \sum_{j=1}^{r-1} \log t_{(j)} + \lambda_1 \sum_{j=r}^{n} \log t_{(j)})$. In fact, any $t_{(j)} < 1$ can be deleted from the summations since it provides a negative contribution to the exponential term in (3.4).

Lastly, for the full conditional distribution for τ, again if $f(\tau) = (t_{(n)} - t_{(1)})^{-1}$, $t_{(1)} < \tau < t_{(n)}$ then, analogous to the previous example, $f(\tau|\lambda_0, \lambda_1, \gamma, \boldsymbol{t})$ is proportional to $k_r' \exp\{-(\lambda_0 - \lambda_1)(n-r+1)\tau^\gamma\}, t_{(r-1)} < \tau < t_{(r)}$ with $k_r' = (\lambda_0/\lambda_1)^r \exp\{-\lambda_0 \sum_{j=r}^{r} t_{(j)}^\gamma - \lambda_1 \sum_{j=r}^{n} t_{(j)}^\gamma\}$. We can sample τ^γ as above, hence τ.

4 A numerical illustration

In Table 1, we have data which represent failure times in hours for electrical insulation in which the insulation was subjected to a continuously increasing voltage stresses.

Table 1: Failure times for electrical insulation

219.3	79.4	86	150.2	21.7	18.5
121.9	40.5	147.1	35.1	42.3	48.7

Lawless (1982) fits a simple exponential model to these data. We consider the constant hazard rate change point model of the previous section.

The output of the Gibbs sampler is a collection of vectors $(\lambda_{0\ell}^*, \lambda_{1\ell}^*, \tau_\ell^*)$ $\ell = 1, \cdots m$ (= 5000 in this case) assumed to be essentially a random sample from $f(\lambda_0, \lambda_1, \tau|\boldsymbol{t})$. Using the components of the sample vector enables us to provide inference about λ_0, λ_1 and τ respectively. In table 2 we present the posterior median and 95% equal tailed intervals estimates for λ_0, λ_1 and τ.

Table 2: Inference summary for constant hazard change-point model.

	Posterior Median	Posterior Interval Estimate
λ_0	.0175	(.0077,.0792)
λ_1	.0103	(.0039,.0206)
τ	55.34	(19.18,188.53)

In Figure 1a) we graph the posteriors for λ_0/λ_1 and in 1b) for τ with a flat prior superimposed. Despite the small sample size, we do find evidence that λ_1 is roughly 60% of λ_0 and that the threshold is roughly 50 hours.

Lastly it may be of interest to examine the survival function, $\bar{F}(t) = \exp\{-\lambda_0 \min(t,\tau) - \lambda_1[t-\tau]^+\}$. As a function of λ_0, λ_1 and τ, at a given t, it too has a posterior distribution $f(\bar{F}(t)|\boldsymbol{\tau})$. The realizations $(\lambda_{0\ell}^*, \lambda_{1\ell}^*, \tau_\ell^*)$ provide realizations $\bar{F}_\ell^*(t)$ from this posterior. From these realizations we can obtain posterior means and 95% individual confidence bands. These are graphed in Figure 2.

5 Concluding remarks

We have described how to handle a general change-point hazard rate problem using a Bayesian approach in the absence of censoring. The case where some of the lifetimes are censored can also be conveniently handled. Suppose, for instance, that for item i, the actual lifetime is not observed. Rather, the item is removed from test after time w_i at which point it has not failed so $T_i > w_i$. All we need to do is introduce T_i as a latent variable into our setup in (2.1) and (2.2). The form in (2.2) is unchanged except that T_i is now unknown and restricted to be greater than w_i. Multiplying (2.2) by (2.3) now yields a posterior for $\boldsymbol{\theta}_0, \boldsymbol{\theta}_1, \tau$ and T_i given $t_1, t_2, \cdots, t_{i-1}, w_i, t_{i+1}, \cdots t_n$. Again adopting Gibbs sampling to fit the model, the full conditional distributions for $\boldsymbol{\theta}_0$, and for $\boldsymbol{\theta}_1$ now given $T_i = t_i$ are exactly as in (2.5) and (2.6). Also given $T_i = t_i$ the full conditional distribution for τ is as in (2.7). Lastly, we need to sample T_i given $\boldsymbol{\theta}_0, \boldsymbol{\theta}_1, \tau$ with $T_i > w_i$, i.e., we need to sample T_i from

the distribution with hazard (1.1) subject to the restriction. If we draw U_i which is an exponential(1) random variable and invert $U_i = H_0(\min(t_i, \tau); \boldsymbol{\theta}_0) + [H_1(t_i; \boldsymbol{\theta}_1) - H_1(\tau; \boldsymbol{\theta}_1)]^+$ to solve for t_i, then t_i is a realization from the unconstrained distribution. Hence, we retain t_i if $t_i > w_i$. If not we draw a new U_i.

The case where $h(t)$ is a truncated "bath-tub" model, i.e., $h_1(t) = \lambda_1$ with $h_0(t) > \lambda_1$ may be of interest as in, e.g., Basu et al. (1988). We might seek to treat this as a semiparametric model by not specifying a parametric form for h_0. The sampling-based fitting machinery could accommodate the various Bayesian nonparametric hazard approaches which have appeared in the literature, e.g., Dirchlet processes (Ferguson, 1974), Gamma processes (Kalbfleisch, 1978), extended Gamma processes (Dykstra and Laud, 1981), Beta process (Hjort, 1990) and mixtures of Betas (Gelfand and Mallick, 1995). We have not examined this approach but a larger data set than that of Table 1 would be required to justify considering it.

Lastly, we can handle the case of more than one threshold if it were appropriate. For instance, in screening defectives it might be sensible to model two stages of burn-in. In a clinical trial, there might be an initial hazard rate for a side effect which, after a period of time, changes to an intermediate lazard rate before settling into a long term hazard rate. An illustrative $h(t)$ would take the form

$$h(t) = \lambda_0 I(0 < t \leq \tau_1) + \lambda_1 I(\tau_1 < t \leq \tau_2) + \lambda_2 I(t > \tau_2)$$

with appropriate order restrictions on λ_0, λ_1 and λ_2.

REFERENCES

Basu, A.P., Ghosh, J.K. and Joshi, S.N. (1988). On estimating change point in a failure rate. *Statistics and Decision Theory, 4*, 239-252.

Devroye, L. (1986). *Non-uniform random variate generation.* Springer, New York.

Dykstra, R. L. and Laud, P. (1981). "A Bayesian nonparametric approach to reliability", *Annals of Statistics, 9*, 356-367.

Ebrahimi, N. (1991). On estimating change point in a mean residual life function. *Sankhya A, 53*, 209-219.

Ferguson, T.S. (1973). "A Bayesian analysis of some nonparametric problems", *Annals of statistics 1*, 209-230.

Gelfand, A.E. and Mallick, B. (1995). "Bayesian analysis of proportional hazards models built form monotone functions", *Biometrics, 51*, 843-852.

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

Gilks, W.R., Richardson, S. and Spiegelhalter, D.J. (1996). *Markov Chain Monte Carlo in Practice*, Chapman and Hall, London.

Hjort, N.L. (1990). "Nonparametric Bayes estimation based on beta processes in models for life history data". *Annals of Statistics, 18*, 1259-1294.

Kalbfleisch, J.D. (1978). Nonparametric Bayesian analysis of survival time data, *J.R.S.S.-B 40*, 214-221.

Lawless, J.F. (1982). *Statistical models and methods for lifetime data.* John Wiley, New York.

Loader, C.R. (1991). Inference for a hazard rate change point. *Biometrika, 78*, 749-757.

Müller, P. (1996). Metropolis posterior integration schemes, Tech Report, ISDS, Duke University.

Nguyen, H.T., Rogers, G.S. and Walker, E.A. (1984). Estimation in change point hazard rate models. *Biometrika, 71*, 229-304.

Smith, A.F.M. and Gelfand, A.E. (1992). Bayesian statistics without tears. *Amer. Statist. 46*, 84-88.

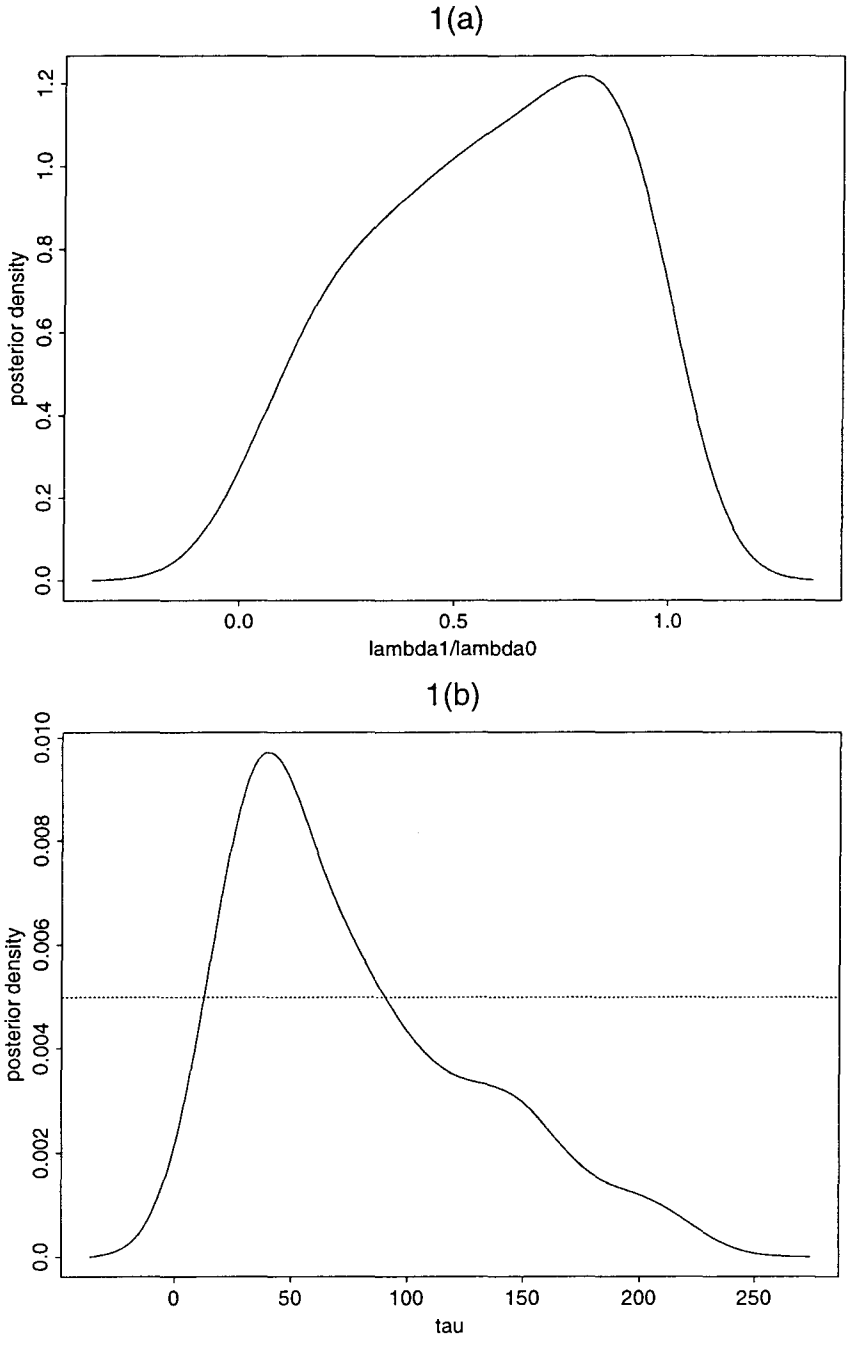


Figure 1: Posterior densities for lambda1/lambda0 (1a) and for tau (1b).

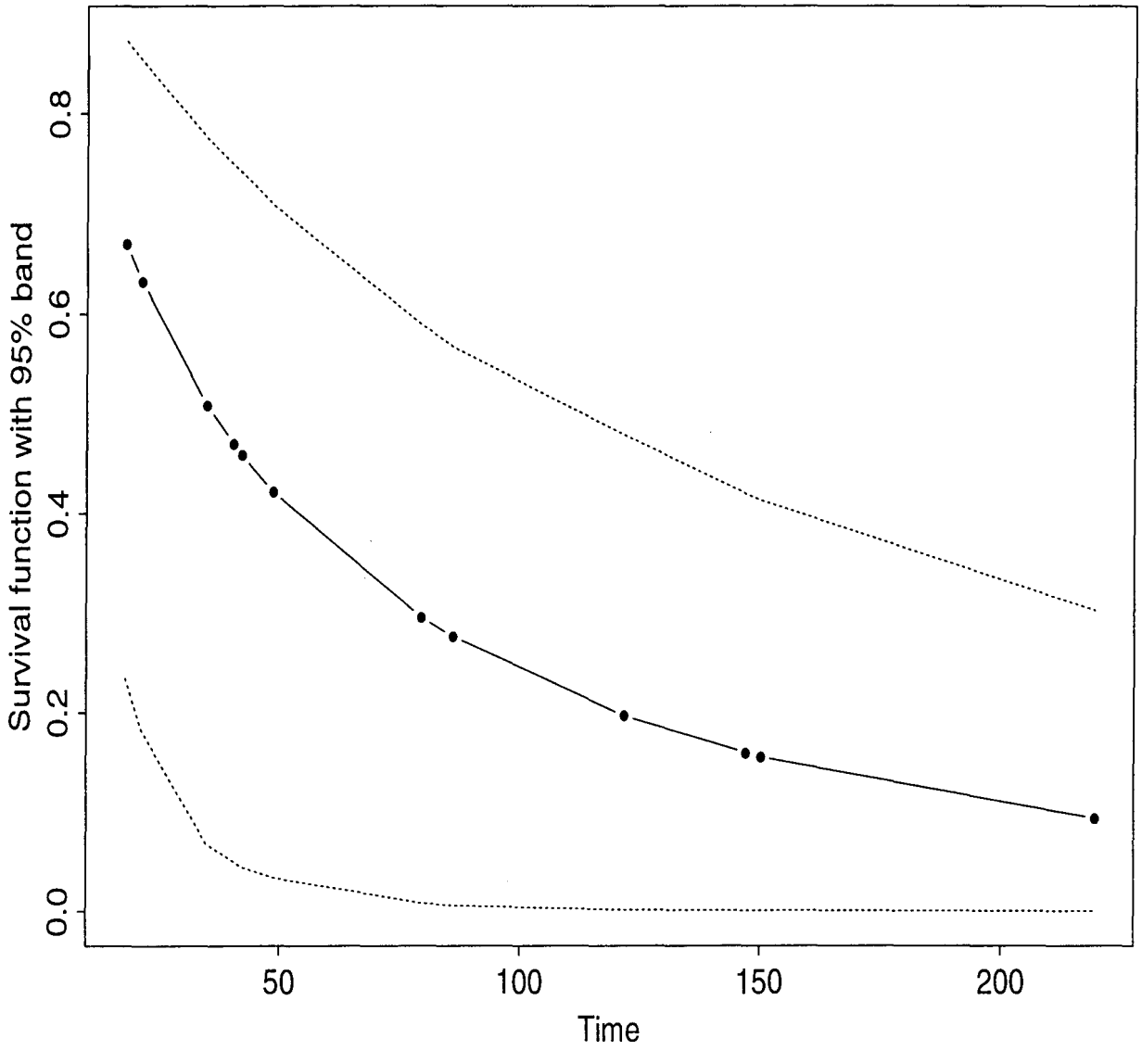


Figure 2: Estimated survival curve with .95 individual confidence bands.