BAYESIAN MODELS FOR A CHANGE-POINT IN FAILURE RATE

by

Nancy L. Campbell

A DISSERTATION

Presented to the Faculty of

The Graduate College at the University of Nebraska

In Partial Fulfillment of Requirements

For the Degree of Doctor of Philosophy

Major: Mathematics and Statistics

Under the Supervision of Professor K.M. Lal Saxena

Lincoln, Nebraska

August, 1995

DISSERTATION TITLE

Bayesian Models for a Change-Point in Failure Rate

BY

Nancy L. Campbell

SUPERVISORY COMMITTEE:

APPROVED	DATE
K.M. Lal Saxena Signature	7/21/95
K.M. Lal Saxena Typed Name	
P. Lahiri Signature	7/21/95
Partha Lahiri Typed Name	
Kun Liang Lu Signature	7/21/95
Kun-Liang Lu Typed Name	
Dong Ho Park Signature	7/21/95
Dong Ho Park Typed Name	
F. Choobineh Signature	7/21/95
Fred Choobineh Typed Name	
Linda J. Young Signature	7/21/95
Linda Young Typed Name	

GRADUATE COLLEGE
UNIVERSITY OF NEBRASKA

BAYESIAN MODELS FOR A CHANGE-POINT IN FAILURE RATE

Nancy L. Campbell, Ph.D.

University of Nebraska, 1995

Advisor: K.M. Lal Saxena

The failure rate function $r(x)$ provides a way to study the aging of a unit in a reliability study or in the analysis of survival data. A change in trend in the aging of the unit may be modelled by a change-point, call it θ, in the failure rate. Such a model can be expressed by writing

$$r(x) = \begin{cases} a(x) & 0 \leq x \leq \theta \\ b(x) & x > \theta. \end{cases}$$

For example, components may go through a "burn-in" phase in which the failure rate is high, after which the failure rate levels off at a lower rate.

Estimation of the change-point may be of interest in certain situations. In the example mentioned above, a manufacturer may want to sell only those components which have survived the burn-in phase, that is, have survived to time θ. We consider a Bayesian approach to the estimation for two change-point models, the constant-to-constant case and the increasing-to-constant case, also referred to as the truncated inverted bathtub shaped model. For the former model, marginal and posterior densities are obtained for both informative and noninformative priors. When noninfor-

mative priors are used, simulation results lead us to recommmend some choices over others. The simulation results also lead us to make recommendations as to which estimator to use in various cases, the posterior mean or the posterior mode. Empirical Bayes and hierarchical Bayes estimation procedures are also considered. For the truncated inverted bathtub shaped model, both parametric and semiparametric approaches are considered. Simulation results again lead us to make recommendations for implementation. A brief discussion of possible approaches to Bayesian estimation of a change-point in the mean residual life function is included in the final section.

Acknowledgements

I would like to express sincere gratitude to my advisor, Professor Lal Saxena, for his guidance and patience as I prepared this dissertation. I appreciate his constant effort to achieve a balance between encouraging me to work independently, and providing valuable advice and direction. I would also like to thank Professors Partha Lahiri, Dong Ho Park, Kun-Liang Lu, Fred Choobineh, and Linda Young for serving on my supervisory committee, and for providing helpful input.

I would like to thank the faculty and staff of the Department of Mathematics and Statistics, and especially Professor Jim Lewis, who has been a great source of support and encouragement throughout my graduate career. I am grateful to the many professors and fellow graduate students who have helped and encouraged me along the way.

Finally, I want to thank all of my friends and family who have been so important to my success in graduate school. I especially want to thank my parents for their ceaseless support of and interest in my work. Last, but certainly not least, I want to thank my husband Tom for weathering bad times with undying patience, and for making good times even better, throughout this endeavor.

Contents

1 Introduction

Properties of life distributions play an important role in reliability theory and analysis of lifetime data. In particular one may wish to examine the way in which a unit ages. The failure rate is one way to study aging. The failure rate function $r(x)$ of a life distribution F, with density f and support $[0, \infty)$, is defined as

$$r(x) = \frac{f(x)}{\overline{F}(x)} \quad x \geq 0,$$

where $\overline{F} = 1 - F$ is the survival function.

A life distribution may be classified according to the behavior of its failure rate function. For example, if $r(x)$ is nondecreasing in x the distribution is in the "Increasing Failure Rate" class. More simply we say that F is an IFR distribution. The "Decreasing Failure Rate" (DFR) class is defined similarly. Not all distributions have monotone failure rate. If there exists a point θ such that $r(x)$ is nonincreasing $\forall x \leq \theta$ and nondecreasing $\forall x > \theta$, the distribution belongs to the "Decreasing Increasing Failure Rate" (DIFR) class. The IDFR class is defined similarly. The failure rate functions in the latter two classes are also referred to as "bathtub-shaped" and "inverted bathtub-shaped" failure rates. In these cases, θ is known as a change-point of $r(x)$.

More generally, a change-point of $r(x)$ indicates some change in trend in the function, and determination of this point may be of interest in an analysis. For example, a component in a reliability study may have a high failure rate when new, which decreases as the product ages until a time θ when the failure rate begins to

increase due to negative effects of aging.

A model involving a change-point in the failure rate function can be expressed as follows. Write

$$r(x) = \begin{cases} a(x) & 0 \le x \le \theta \\ b(x) & x > \theta. \end{cases} \tag{1}$$

Here $a(\theta)$ may be equal to $b(\theta)$. A distribution with "smooth" failure rate function in the DIFR (IDFR) class can be included in model (1), if $a(x)$ and $b(x)$ represent the same function, a function which is nonincreasing (nondecreasing) up to time θ and nondecreasing (nonincreasing) after time θ.

Classical (i.e., non-Bayesian) estimation of the change-point θ has been considered by several authors. Nguyen, Rogers, and Walker (1984) and Yao (1986) proposed estimators for θ when $a(x) \equiv a$, $b(x) \equiv b$, a and b being two unknown constants. The former authors showed that the maximum likelihood estimator (MLE) of θ does not exist in certain cases, and instead proposed a consistent estimator of θ based on a stochastic process $X_n(t)$ which converges to 0 at $t = \theta$. The latter author proposed a restricted MLE which avoids the existence problems of the standard MLE. The constant-to-constant model was also considered by Loader (1991), Matthews and Farewell (1982), and Matthews, Farewell, and Pike (1985), all of whom focused on likelihood ratio tests for occurrence of change. Achcar and Bolferine (1989) considered the same model but included the case of randomly right-censored data. They also focused on determination of whether a change had occurred, rather than on estimation of the change-point.

The constant-to-constant model described above is fully parametric, as the distribution can be written explicitly. More general models have also been considered by other authors. Basu, Joshi, and Ghosh (1988) developed two estimators for θ under a truncated bathtub model, assuming that $a(x)$ in (1) is nonincreasing while $b(x) \equiv b$. Both estimators are based on test statistics designed to determine whether the failure rate is constant, and both are shown to be consistent estimators. The authors also determined the asymptotic distribution of the estimator of Nguyen, Rogers and Walker (1984) discussed in the constant-to-constant case.

Kulasekera (1988) considered a model in which both the left derivative of $a(x)$ at θ and the right derivative of $b(x)$ at θ exist but are not equal. The estimator for θ in this case maximizes an expression which estimates the difference in the two derivatives. The same author also considered a model in which $a(x)$ and $b(x)$ are both nondecreasing (nonincreasing) but have different forms. Arunkumar (1969) and Kulasekera and Saxena (1991) proposed estimators for θ under the DIFR (IDFR) model, based on minimization (maximization) of a nonparametric estimator of $r(x)$. The latter authors established the consistency of their estimator as well.

The models mentioned thus far do not assume that knowledge concerning θ is available prior to the analysis, with the exception of upper and lower bounds in some cases. When prior information is available, a Bayesian approach incorporates such knowledge into the estimation procedure. Even when prior information is vague, one may argue that a Bayesian analysis is desirable in order to satisfy the Likelihood Principle. See Berger and Wolpert (1984) for such an argument. A brief summary

of both parametric and nonparametric Bayesian approaches is included at the end of this section.

A Bayesian approach to the change-point problem has been considered by Ghosh, Joshi, and Mukhopadhyay (1993) in the constant-to-constant case, using certain diffuse priors. They established consistency of the posterior mode as an estimator of θ for that particular prior, using an asymptotic expansion of the marginal posterior density of θ. Achcar and Bolferine (1989) also considered this model; however as in the non-Bayesian case they focused on estimation of the ratio b/a in order to determine whether a change occurred, rather than on estimation of the change-point itself. They assumed the prior distribution for θ to be the discrete uniform distribution over the sample points $x_1, \ldots, x_n$. Raftery and Akman (1986) considered the Bayesian estimation of a change-point in a Poisson process, with the rate of the process changing from a constant a to a constant b. Gamma priors were considered for a and b, and a uniform prior for the change-point.

It should be noted that there is a distinction between the problem dealing with a change-point in a failure rate and the problem of a change-point in a sequence of random variables. There is a large amount of literature on the latter problem. A Bayesian approach has been considered for inference in such problems by many authors, including Chernoff and Zacks (1965) and Smith (1975) and more recently Carlin, Gelfand, and Smith (1992) and Barry and Hartigan (1992).

In this dissertation, a Bayesian approach to estimation of a change-point in the failure rate function is considered for increasingly complex models of the form (1).

The constant-to-constant model is considered first, in Section 2. Section 2.1 examines models assuming a and b to be known constants; the case when they are unknown is discussed in Section 2.2. In both situations it is first assumed that a prior density $\pi(\theta)$ can be specified by the analyst, followed by a discussion of choices for noninformative priors when this assumption does not hold. The truncated inverted bathtub model is presented in Section 3, first for $a(x)$ of a particular form, then for $a(x)$ unknown. In the former case, a parametric approach similar to that for the constant-to-constant case is proposed. In the latter case, a semiparametric Bayesian approach is used.

For all of the models considered, the likelihood function, prior density, marginal density, and posterior density are examined. Following the examination of the form of these functions, observations are made on the basis of simulations done for the particular model. Because we do not consider a particular loss function here, there is no clear choice for the estimator of the change-point. All of the information concerning θ is contained in the posterior distribution, and we would like to select an estimator which performs well in repeated experiments. Candidates for an estimator include the posterior mean and mode, among others. The simulations consider both of these quantities, providing a comparison for varying model parameters, prior parameters, and data sets.

We point out that, besides the failure rate function, the mean residual life (MRL) function of a life distribution is also used in the study of aging of units. As in the case of a failure rate function, a change point in the MRL function is also indicative of a change in trend in aging of units. Non-Bayesian estimation of a change-point in

the MRL function is considered by Ebrahimi (1991). In Section 4 we discuss briefly a Bayesian approach to the estimation of change-point in the MRL function.

The remainder of this section gives a brief summary of general Bayesian approaches to estimation. In a parametric Bayesian estimation problem, unknown quantities may be denoted θ, which could be a scalar or a vector. In our problem θ will represent the change-point mentioned earlier, and possibly some nuisance parameters which we do not wish to estimate, but must be included as unknowns. A classical statistical approach utilizes information from a sample to make inferences about θ, which is considered a *fixed* quantity. Properties of estimators of θ are based on the sampling distribution of certain statistics. Hence the theory involves not only the observed data points, but also considers all possible data sets for the study. A Bayesian analysis, on the other hand, considers only the observed values from the data set, and incorporates prior information about θ in the study. Here θ is considered an unknown *random* quantity which varies according to a specified prior distribution $\pi(\theta)$. If the observations on X have a distribution $f(x \mid \theta)$, then the marginal distribution of X and the posterior distribution of θ are, respectively,

$$m(x \mid \pi) = \int_{\Theta} f(x \mid \theta)\pi(\theta)d\theta, \quad \pi(\theta \mid x) = \frac{f(x \mid \theta)\pi(\theta)}{m(x \mid \pi)}.$$

The former represents the distribution of X, independent of θ but taking the prior distribution into account, while the latter represents the distribution of θ conditioned on the observed data **X**. Inference concerning θ is then based on the posterior distribution, since this distribution contains all of the information we have about θ. For

example, one may estimate θ through the posterior mean or mode. When θ involves nuisance parameters, $\pi(\theta)$ denotes the joint prior distribution of all of the unknown parameters and $\pi(\theta \mid x)$ denotes the joint posterior. The marginal posterior density of the parameter of interest, in our case the change-point, is obtained by integrating out the nuisance parameters.

One can see, from the expressions above for the marginal and posterior densities, that Bayesian estimation procedures rely heavily on evaluation of integrals. Unless the functional forms of the likelihood and posterior are such that their product is analytically tractable, numerical integration procedures are necessary for an exact Bayesian analysis. There has been much discussion in the literature as to computational methods applicable for Bayesian problems. Monte Carlo and quadrature methods for integration can be useful but computationally intensive tools. Alternatively, several approximation techniques have been proposed. For example, Walker (1967) examines asymptotic normality of the posterior density, Lindley (1980) proposes approximation of posterior moments based on first-order error terms of the normal approximation, and Tierney and Kadane (1986) suggest an approximation to the marginal density and posterior moments based on LaPlace's method. These approximations require regularity conditions which in many cases are not restrictive. However, we will see that in the constant-to-constant failure rate model considered in Section 2, such regularity conditions do not hold. For this reason, in our simulations we have used numerical integration based on Gaussian and Laguerre quadrature methods as well as Sinc quadrature. We have found that, because the expressions to be

integrated often involve exponential terms, Sinc quadrature performs quite well using very few nodes. See Lund and Bowers (1992) for a description of the Sinc method. As the dimensionality of integrals does not exceed three in the models considered here, implementation of these numerical methods is feasible.

A nonparametric Bayesian analysis, though based on the same principles as the parametric described above, proceeds differently. The unknown quantity of interest is not a scalar or a vector, but a function. Typically the function is the cumulative distribution function (cdf) or the probability density function (pdf) of the distribution of $\mathbf{X}$. Ferguson (1964) first considered such a problem, and introduced the concept of a process prior placed on the unknown function. He placed a gamma process prior on the cdf $F(x)$ of the distribution, and obtained an expression for the posterior mean of $F(x)$ as a function of x. In the semiparametric model mentioned earlier, the increasing portion of the failure rate function, $a(x)$, is the unknown function. The change-point θ is still the unknown parameter of interest, hence an approach combining the parametric and nonparametric methods is used. This technique, which we refer to as semiparametric, will be described in detail in Section 3.2.

2 Constant-to-Constant Model

Consider the special case of (1) when $a(x)$ and $b(x)$ are constants. Then model (1) becomes

$$r(x) = \begin{cases} a & 0 \leq x \leq \theta \\ b & x > \theta, \end{cases}$$

which yields the density function

$$f(x \mid \theta) = \begin{cases} ae^{-ax} & 0 \leq x \leq \theta \\ be^{-bx-(a-b)\theta} & x > \theta. \end{cases} \tag{2}$$

Let $x_1 \leq x_2 \leq \ldots \leq x_n$ be an ordered random sample of n lifetimes from the distribution (2), and let $\mathbf{x}$ denote the vector $(x_1, \ldots, x_n)$. Let

$$k = k(\theta) = number\ of\ x_i{}'s \leq \theta.$$

The likelihood function for the sample is

$$\prod_{i=1}^{n} f(x_i \mid \theta) = a^k b^{n-k} \exp[-a \sum_{i=1}^{k} x_i - b \sum_{k+1}^{n} x_i - (n-k)(a-b)\theta]. \tag{3}$$

Ghosh, Joshi, and Mukhopadhyay (1993) examine properties of the likelihood function (3). They observe that the function is discontinuous at each data point, and has an upward jump at each of these points. In addition, the likelihood is linear with negative slope in each continuous strip, and is constant for $\theta > x_n$. The authors also note that with a fairly high probability, the likelihood function has a local maxima much larger than θ. These properties are of particular interest when a and b are known and a uniform prior is specified for θ. In this situation, the posterior density of θ has all of the characteristics of the likelihood function.

2.1 a, b Known Constants

We begin with the model involving a single unknown parameter, the change-point θ, assuming the constants a and b in (2) to be known.

2.1.1 Gamma Prior for θ

Consider a gamma prior density for θ, with shape parameter $r(>0)$ and scale parameter $\lambda(>0)$. Note that the analysis can be done with any distributional form of a specified prior density $\pi(\theta)$. We choose a gamma prior density here because its support $[0,\infty)$ coincides with the parameter space Θ, and because various combinations of r and λ provide a wide range of distributional shapes to represent one's prior knowledge regarding θ. Note that the gamma distribution includes the exponential distribution as a special case, when $r = 1$. We have included this case in many of the simulations. Restricting the prior to be exponential does not provide as rich a class of priors as does the gamma distribution, with just one parameter. However it does provide some flexibility, and proves to be a much more manageable class of priors when we consider a hierarchical Bayes approach in Section 2.4.

We first obtain an expression for the marginal density of the data. This involves integrating $\pi(\theta)\prod_{i=1}^{n} f(x_i \mid \theta)$ over the parameter space Θ. Note that $k = k(\theta)$ increases as θ increases, for fixed $x_1 \leq \ldots \leq x_n$. Thus we integrate piecewise, with $k = 0$ over the interval $[0, x_1)$, with $k = 1$ over the interval $[x_1, x_2)$, $\ldots$, and with $k = n$ over the interval $[x_n, \infty)$.

Letting $x_0 = 0$ and $x_{n+1} = \infty$, we write the marginal density as

$$\begin{aligned} m(\mathbf{x}) &= \sum_{k=0}^{n} \int_{x_k}^{x_{k+1}} a^k b^{n-k} \exp[-a \sum_{i=1}^{k} x_i - b \sum_{k+1}^{n} x_i - (n-k)(a-b)\theta] \pi(\theta) d\theta \\ &\propto \sum_{k=0}^{n} a^k b^{n-k} \exp[-a \sum_{i=1}^{k} x_i - b \sum_{k+1}^{n} x_i] \\ &\quad \int_{x_k}^{x_{k+1}} \theta^{r-1} \exp[-\theta[\lambda + (n-k)(a-b)]] d\theta \\ &= m^*(\mathbf{x}), \ \ say, \end{aligned} \tag{4}$$

where $m^*(\mathbf{x})$ throughout this paper will denote a function proportional to $m(\mathbf{x})$. Constants are omitted as they cancel in the numerator and denominator of the expression for the posterior density. The posterior density, from (2) and (4), can be written as

$$\pi(\theta \mid \mathbf{x}) = \frac{a^k b^{n-k} \exp[-a \sum_{i=1}^{k} x_i - b \sum_{k+1}^{n} x_i - \theta(\lambda + (n-k)(a-b))] \theta^{r-1}}{m^*(\mathbf{x})}. \tag{5}$$

Once the posterior distribution has been determined, the decision-maker can select an estimator for θ such as the posterior mean or the posterior mode, both of which are readily computed. Choice of estimator is discussed in the following subsection, where the behavior of both mean and mode are observed empirically through simulations.

Simulation Results

Tables 1 and 2 display results of simulations for the constant-to-constant model using a gamma prior for θ. In each case, 100 samples were generated from the appropriate distribution with fixed change-point θ. Samples of size 20 were used in the simulations of Table 1, with an increase to size 50 in Table 2. Posterior estimators for each sample were computed for different values of the prior parameters. The values reported are

the average values across the 100 samples. The posterior mean $\hat{\theta}_{MN}$ and posterior mode $\hat{\theta}_{MD}$ are displayed, with the corresponding mean squared errors displayed in parentheses below each estimator. The Sinc quadrature method was used to evaluate the integrals.

Several combinations of (a, b) have been considered, to include both the DFR and IFR models. In the first two cases, the first column represents a choice of prior parameters for which the data and the prior distribution are in agreement, with respect to the mean. For example, in the first row $\theta = 0.5$, and the prior mean $r/\lambda = 0.5$ as well. For all such cases, the posterior mean overestimates θ while the posterior mode underestimates. The posterior mean is closer to the selected value of θ than is the posterior mode for the smaller samples, while the performance is comparable for larger sample size. The other columns display results for differing values of r and λ, to represent varying degrees of discord between the data and prior. In these cases there is not a clear choice for an estimator, in some cases $\hat{\theta}_{MN}$ performs better, and in other cases $\hat{\theta}_{MD}$ does. When the sample size is increased in Table 2, both estimators are influenced more by the data than the prior, hence both come closer to θ.

In the third case displayed in each table, the values of θ, a, and b are chosen to represent observations on a larger scale. Here the change-point $\theta = 5$, but a and b are adjusted so that the change-point occurs at roughly the 40th percentile of the distribution. With gamma priors, choice of scale has an effect on the estimators in that specifying priors to this scale results in greater prior variance, and hence in

greater posterior variance. Thus the magnitude of the error in estimation is greater. This is especially true of the first two columns. We note from the results in this row that small changes in prior parameter specification have just a slight effect on the posterior estimators. This observation indicates some robustness in the Bayesian estimators, where robustness is measured in terms of changes in prior specification.

We point out that the last row in each of the two tables represents parameter choices which may not be realistic. Here $a = 2$, $b = 1$, and $\theta = 5$, in which case an extremely small percentage of the population survive beyond time θ, as $\overline{F}(\theta) < 0.00001$. With virtually no observations falling above θ, we cannot expect to estimate the value accurately, and indeed a change-point occuring at this point in the life distribution would most likely not be of interest. However, we have included this case in the simulations as a matter of curiosity. Not surprisingly, the estimators do not appear to be reliable in this situation. Though the estimators came close for a few of the parameter choices, overall the results were quite erratic. We do not consider such cases in further sections.

a	b	θ		(r,λ)	(r,λ)	(r,λ)	(r,λ)	(r,λ)
2	1	.5		(1, 2)	(2, 5)	(3, 5)	(1, 4)	(3, 4)
			$\hat{\theta}_{MN}$	0.5770 (0.0655)	0.4691 (0.0177)	0.5963 (0.0300)	0.3777 (0.0344)	0.6939 (0.0726)
			$\hat{\theta}_{MD}$	0.3516 (0.0621)	0.3665 (0.0436)	0.4563 (0.0268)	0.2401 (0.0986)	0.4970 (0.0303)
1	2	.5		(1, 2)	(2, 5)	(3, 5)	(1, 4)	(3, 4)
			$\hat{\theta}_{MN}$	0.5199 (0.0627)	0.4539 (0.0246)	0.5517 (0.0293)	0.3780 (0.0442)	0.6242 (0.0562)
			$\hat{\theta}_{MD}$	0.4238 (0.0772)	0.4282 (0.0397)	0.4895 (0.0347)	0.3542 (0.0745)	0.5288 (0.0399)
.2	.1	5		(1, .2)	(1.1,.25)	(10, 2)	(10.1, 1.8)	(12, 2)
			$\hat{\theta}_{MN}$	5.7702 (6.5509)	5.2886 (4.2856)	4.9612 (0.2633)	5.4197 (0.4868)	5.5289 (0.5274)
			$\hat{\theta}_{MD}$	3.4806 (6.3563)	3.4325 (6.2214)	4.6911 (0.7733)	5.0223 (0.6289)	5.2449 (0.7249)
.2	.1	5		(1, .2)	(1.1, .25)	(10, 2)	(10.1, 1.8)	(12, 2)
			$\hat{\theta}_{MN}$	5.5574 (1.0569)	4.9798 (0.0241)	3.9195 (1.2055)	2.5170 (6.1794)	5.9254 (0.8934)
			$\hat{\theta}_{MD}$	0.9059 (16.9522)	4.4860 (0.2865)	4.2425 (0.5810)	2.2017 (7.8720)	5.4827 (0.2662)

Table 1: Constant-to-constant Model, a and b Known, Gamma Prior, n = 20