Lecture Notes in Statistics 133

Edited by P. Bickel, P. Diggle, S. Fienberg, K. Krickeberg,
I. Olkin, N. Wermuth, S. Zeger

Dipak Dey
Peter Müller
Debajyoti Sinha (Editors)

Practical Nonparametric and Semiparametric Bayesian Statistics

Springer

Dipak Dey
Department of Statistics
University of Connecticut
Storrs, CT 06269

Peter Müller
ISDS
Duke University
Durham, NC 27708

Debajyoti Sinha
Department of Mathematics
University of New Hampshire
Durham, NH 03824

Library of Congress Cataloging-in-Publication Data

Practical nonparametric and semiparametric Bayesian statistics / Dipak Dey, Peter Müller, Debajyoti Sinha (editors).
p. cm. -- (Lecture notes in statistics ; 133)
Includes bibliographical references.
ISBN 0-387-98517-4 (softcover : alk. paper)
1. Bayesian statistical decision theory. 2. Nonparametric statistics. I. Dey, Dipak. II. Müller, Peter, 1963 Aug. 9- III. Sinha, Debajyoti. IV. Series: Lecture notes in statistics (Springer-Verlag) ; v. 133.
QA279.5.P73 1998
519.5'42--dc21
98-16460

Printed on acid-free paper.

Camera ready copy provided by the editors.
Printed and bound by Braun-Brumfield, Ann Arbor, MI.
Printed in the United States of America.

9 8 7 6 5 4 3 2 1

ISBN 0-387-98517-4 Springer-Verlag New York Berlin Heidelberg SPIN 10676439

19

A Bayesian Model for Fatigue Crack Growth

Liping Liu
Elja Arjas

ABSTRACT In this chapter we apply hierarchical Bayesian modeling to describe fatigue crack growth. Due to the special pattern of crack growth data, we propose a nonparametric model which is based on a minimum of probabilistic assumptions. We put a (prior) probability model on an unknown growth curve by considering a grid of time points and assuming that the increments between such grid points are independent exponential random variables, with expectations proportional to the length of the grid interval. The nonparametric structure enables us to arrive at a model which is both flexible enough to be fitted to different sets of growth data and, at the same time, enables us to interpret different parts of the model in an intuitively plausible manner. A Markov chain Monte Carlo method is adopted for the numerical computation. Numerical results show that our model behaves well in both being fitted to the data and in predicting future observables.

KEY WORDS: Exchangeability; Hierarchical modeling; Predictive distribution; Growth curve; Markov chain Monte Carlo.

19.1 Introduction

Physical devices or systems in use are typically subject to monotonically increasing wear or degradation. The failure or death of a system is then conveniently defined to be the event in which degradation reaches a certain threshold level. This point of view becomes particularly relevant in cases where the level of degradation can be measured to some degree of accuracy, and where the goal of the statistical modeling is to provide a systematic method for predicting the time to failure from such follow-up measurements. Utilization of system specific degradation readings can be expected to lead to a much smaller prediction error variance than methods which only use the failure time data (see Lu and Meeker, 1993).

The connection between the study of degradation or wear and that of growth curves is obvious. In the former "time" can be simply interpreted as mileage of driving, number of cycles in the trial, etc. There is extensive literature concerning statistical growth curve models. Most papers assume

linear or polynomial structure, and traditional and modern regression techniques were often adopted (see e.g. von Rosen, 1991, Fearn, 1975, Lindley and Smith, 1972, Tian, Shukla and Buncher, 1994, and Rao, 1987). For specific studies concerning cumulative damage, Bogdanoff and Kozin (1985) used the ideas of finite Markov chains, and formulated the so-called B-models. Lu and Meeker (1993) established a parametric nonlinear mixed-effects model to describe degradation measurements.

Due to the convenience of expressing uncertainties probabilistically in terms of posterior and predictive distributions, the Bayesian approach looks quite attractive to us. Combined with hierarchical structure and nonparametric modeling, we can arrive at a model which is flexible enough to be fitted to different sets of growth data. Furthermore, it is usually very easy to assign a "flat" prior probability to the unobservables, and to work with very weak structural and probabilistic assumptions.

The set-up considered in this paper is as follows. Suppose there are n physical devices under observation. For the k^{th} device, measurements $0 < Y_{k,1} < Y_{k,2} < \cdots < Y_{k,p_k}$ are made at times $0 < t_{k,1} < t_{k,2} < \cdots < t_{k,p_k} \leq T_{max}$, respectively, where T_{max} is the largest possible observational time (across all devices) and p_k is the number of observations available on device k.

In a designed experiment, the $t_{k,i}$'s above are often equally spaced. However, some observations might be missing, and some devices can fail before it has been possible to make as many measurements as were originally planned. In our model, we do not ask that the number of observations for different devices be the same, nor the times at which these observations were made. In this way, the data can be of a very general form and hence there is no need to consider separately a "missing data" or "incomplete observations" problem.

Our ultimate purpose is to predict the growth of device after a last recorded observation. This device can be a "new" one (previously unobserved), or one about which there are already some observations in the data set. We are especially interested in the time at which this device reaches a certain target level, Y^*. If Y^* is taken to be the threshold level described above, our task becomes naturally the prediction of the failure time, as we have to do in survival analysis.

In the next section we will present our model in detail. Section 3 describes a Metropolis-Hastings algorithm associated with the model. In section 4 an example is analyzed, and our results on prediction are compared with those of Lu and Meeker (1993). The paper is concluded with a brief discussion in section 5.

19.2 The model

The model is based on the assumption that the growth curves of all devices have the same shape. This is, in fact, our only structural assumption. The

differences between devices are thought to arise from the fact that they grow at different but constant speeds. In other words, their "operational ages" are linear transformations of calendar time. More precisely, we assume that a measurement of the k^{th} device at time t, $Y_k(t)$, can be expressed as

$$Y_k(t) = g_0(\theta_k \cdot t) + \varepsilon_{k,t}, \tag{1}$$

where θ_k is a positive random variable representing the relative speed of growth of device k, g_0 is an unknown function which is continuous and strictly increasing, and $\varepsilon_{k,t}$ is a measurement error.

As is apparent, the function g_0 can be interpreted as the "common" growth curve of the population under investigation. The coefficient θ_k is the relative speed of growth of device k with respect to calendar time t. In the present Bayesian model both g_0 and θ_k are random. The random function

$$g_k(t) = g_0(\theta_k \cdot t) \tag{2}$$

can be simply interpreted as the growth curve associated with the device indexed by k. Notice that model (1) is not identifiable, since for any positive constant c,

$$g_0(\theta_k \cdot t) = g_0(c\theta_k \cdot \frac{t}{c}) = \tilde{g}_0((c\theta_k) \cdot t),$$

where $\tilde{g}_0(u) = g_0(u/c)$. But assuming that the functions $g_1, \ldots, g_n$, as well as the parameters $\theta_1, \ldots, \theta_n$, are exchangeable (see e.g. Bernardo and Smith, 1994), we now somewhat arbitrarily use the growth curve of device 1 as a baseline, writing then (2) in the form

$$g_k(t) = g_1(\xi_k \cdot t) = g(\xi_k \cdot t), \tag{3}$$

where $\xi_k = \theta_k/\theta_1$ and $g = g_1$. Now, since $\xi_1 \equiv 1$, model (3) is identifiable.

Suppose now that $\theta_1, \theta_2, \ldots, \theta_n$ are i.i.d. and have prior distribution Gamma(α_θ, β_θ), where α_θ is the shape and β_θ the scale parameter. Then, it is easy to verify that given θ_1, the derived parameters ξ_k will be i.i.d. and will follow the (prior) distribution Gamma(α_θ, $\theta_1\beta_\theta$). This confirms the fact that the particular choice of device 1 as a baseline is irrelevant since it only influences the value of the scale parameter. A reasonable assumption for the prior distribution of ξ_k would be therefore: for $k = 2, 3, \ldots, n$, ξ_k follows a Gamma distribution Gamma(α, β), where α and β are two independent random parameters with prior distributions Exp(λ) and Exp(ρ), respectively, with λ and ρ being given hyperparameters. We will generally choose the hyperparameters λ and ρ to be the same, which corresponds to the exchangeability idea that we don't in our prior assume that devices indexed from 2 to n grow faster or slower than device 1. Intuitively, a large value of α indicates that all devices will grow with a speed close to that of device 1, whereas a small value of α reflects the fact that the speeds can be very different. For this reason we also let α be random, so that the model can be updated to correspond to the heterogeneity between devices

in the data. Apparently a large value of ξ_k indicates that device k grows faster than device 1, and if most devices in the data actually grow faster than device 1, we will have a large posterior probability that β is smaller than α.

In principle, g is thought of as a nonparametrically defined function. In practice, we assume here for simplicity that g is piecewise linear. The points at which the slope of g changes its value belong to a fixed grid of points (see Arjas and Gasbarra, 1994). So what we need to specify for the model is a positive, strictly increasing sequence $g(t_0) < g(t_1) < g(t_2) < \cdots < g(t_N)$, where $0 = t_0 < t_1 < t_2 < \cdots < t_N = CT_{max}$ are equally-spaced points on $[0, CT_{max}]$. The constant C must be greater than 1, so that in case $\xi_k > 1$, $g(\xi_k t)$ is still well-defined. The prior distribution of g is here specified by assuming that, for $i = 1, 2, \ldots, N$, each increment $\Delta g(t_i) = g(t_i) - g(t_{i-1})$ of g follows an exponential distribution $\mathrm{Exp}((t_i - t_{i-1})^{-1}\mu)$. Here μ is a given hyperparameter. Note that this is somewhat different from considering the increments of a gamma process on the grid intervals.

This prior of g guarantees that g is increasing. The prior mean of $\Delta g(t_i)$ is $(t_i - t_{i-1})/\mu$, which is proportional to the length $\Delta t_i = t_i - t_{i-1}$. This means that the speed of growth has no trend according to the prior, and the posterior distribution of g will acquire additional information from the data so as to be able to describe the behaviour of the "true" growth curve of device 1. If one knows more about g, one can add more specifications or restrictions to the prior of g. Later in this paper, we will consider a convexity condition on g.

It is quite common to assume that the measurement errors follow a normal distribution with mean 0. Here the variables $\varepsilon_{k,t}$ are assumed to be mutually independent and follow a normal distribution $\mathcal{N}(0, \sigma^2)$, with σ^2 having an inverse Gamma distribution $IG(a, b)$, where a and b are hyperparameters. Therefore conjugacy can be realized here (see Roberts and Smith, 1993).

Now the hierarchical structure can be summarized in Figure 1 below. Hyperparameters and data are surrounded by squares, and random parameters are surrounded by circles. Arrows are used to indicate their dependence. The function $g(\cdot)$ in the figure corresponds to a set of parameters as described above.

19.3 A Markov Chain Monte Carlo method

A Markov chain Monte Carlo (MCMC) method is used for the numerical computation. As evidenced by recent literature, MCMC is a very powerful tool in dealing with hierarchical Bayesian models. For the general theory of MCMC, the reader is referred to, for example, Besag et al. (1995). In the following we present the main steps of the algorithm handling our crack growth model.

Denote by $x = (\alpha, \xi_2, \xi_3, \ldots, \xi_n, g(\cdot), \sigma^2)$ the vector of all the model

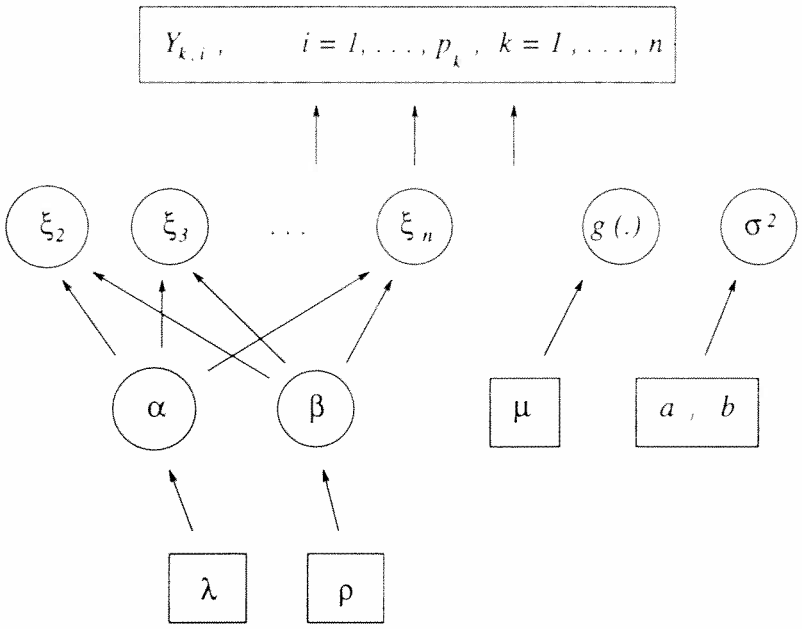

FIGURE 1. Hierarchical structure of the model.

parameters, and by $\pi(dx)$ and $\pi(dx|Y_{k,i} \quad i = 1, \ldots, p_k, \ k = 1, \ldots, n) = \pi(dx|data)$ the prior and posterior distributions of x, respectively. The goal is to construct an ergodic Markov chain $x^{(j)} = (\alpha^{(j)}, \xi_2^{(j)}, \xi_3^{(j)}, \ldots, \xi_n^{(j)}, g^{(j)}(\cdot), (\sigma^2)^{(j)})$, which has the unique stationary distribution given by $\pi(dx|data)$ (see Roberts and Smith, 1994). We first generate an initial value $x^{(0)}$, for example, from the prior distribution $\pi(dx)$. This is then updated from $x^{(j)}$ to $x^{(j+1)}$, always one coordinate at a time, and keeping all other coordinates fixed.

Given all other coordinates of x, the conditional density of α is proportional (in α) to

$$\frac{\beta^{(n-1)\alpha}}{(\Gamma(\alpha))^{n-1}}\Big(\prod_{k=2}^{n}\xi_k\Big)^{\alpha-1}e^{-\lambda\alpha},$$

so the Metropolis-Hastings algorithm can be used here. We generate a proposal α' of $\alpha^{(j+1)}$ from $\mathrm{Exp}(\lambda)$, then calculate the Hastings ratio

$$r_\alpha = \left(\frac{\Gamma(\alpha^{(j)})}{\Gamma(\alpha')}\right)^{n-1} \cdot \beta^{(n-1)(\alpha'-\alpha^{(j)})}\Big(\prod_{k=2}^{n}\xi_k^{(j)}\Big)^{\alpha'-\alpha^{(j)}},$$

and let $\alpha^{(j+1)} = \alpha'$ with probability $\min(1, r_\alpha)$, and $\alpha^{(j+1)} = \alpha^{(j)}$ otherwise.

It is easy to see that the conditional distribution of β, given all other coordinates, is the Gamma distribution $\mathrm{Gamma}((n-1)\alpha+1, \ (\sum_{k=2}^{n}\xi_k+\rho))$. Therefore β can be updated by draws directly from the complete conditional posterior distribution.

The conditional distribution of ξ_k is proportional to

$$\xi_k^{\alpha-1} \cdot e^{-\beta\xi_k} \prod_{i=1}^{p_k} [\frac{1}{\sqrt{2\pi}\sigma} \exp\{-\frac{1}{2\sigma^2}(Y_{k,i} - g(\xi_k \cdot t_{k,i}))^2\}].$$

Again we can use the Metropolis-Hastings algorithm. We first generate a proposal ξ' of $\xi_k^{(j+1)}$ from Gamma$(\alpha^{(j+1)}, \beta^{(j+1)})$, then calculate the corresponding Hastings ratio

$$r_\xi = \exp\{-\frac{1}{2(\sigma^2)^{(j)}} \sum_{i=1}^{p_k} [(g^{(j)}(\xi' \cdot t_{k,i}))^2 - (g^{(j)}(\xi_k^{(j)} \cdot t_{k,i}))^2$$

$$-2Y_{k,j}(g^{(j)}(\xi' \cdot t_{k,i}) - g^{(j)}(\xi_k^{(j)} \cdot t_{k,i}))]\},$$

and let $\xi_k^{(j+1)} = \xi'$ with probability $\min(1, r_\xi)$.

For the piecewise linear function $g(\cdot)$, we use the increments $\Delta g(t_i) = g(t_i) - g(t_{i-1})$ as our model parameters $(i = 1, 2, \ldots, N)$. When $\Delta g(t_i)$ is updated, all $g(t_j)$'s $(j \geq i)$ will move upwards or downwards by the same amount, according to whether $\Delta g(t_i)$ is made larger or smaller. The updating is realized in a similar way as in the previous step, i.e., by generating a proposal from the prior distribution, then calculating the corresponding Hastings ratio from the likelihood, and finally deciding whether to accept it or not.

The updating of σ^2 is equally simple. Given all other parameters and the data, the conditional distribution of σ^2 is inverse Gamma

$$IG(a + \frac{1}{2}\sum_{k=1}^{n} p_k \ , \quad b + \frac{1}{2}\sum_{k=1}^{n}\sum_{i=1}^{p_k}(Y_{k,i} - g(\xi_k \cdot t_{k,i}))^2).$$

A new value $(\sigma^2)^{(j+1)}$ can be drawn from this distribution directly to replace $(\sigma^2)^{(j)}$.

After the j^{th} generic step of the algorithm is finished, we can simulate any quantity of interest. From (3), we have

$$g_k^{(j)}(t) = g^{(j)}(\xi_k^{(j)} \cdot t),$$

and a corresponding measurement can be simulated by using the expression

$$Y_k^{(j)}(t) = g_k^{(j)}(t) + \varepsilon_{k,t}^{(j)},$$

where $\varepsilon_{k,t}^{(j)}$ is generated independently from $\mathcal{N}(0, (\sigma^2)^{(j)})$.

If the k^{th} device is still below the target level Y^* at the end of the trial, since the simulated curve $g_k^{(j)}(t)$ is strictly increasing and piecewise linear, we can easily get the simulated crossing time $(t_k^*)^{(j)}$ which satisfies

$$g_k^{(j)}((t_k^*)^{(j)}) = Y^*.$$

It is not difficult to check that the Markov chain we constructed is irreducible and aperiodic; therefore by the ergodic theorem, the predictive distribution of $g_k(t)$ is approximated by

$$P(g_k(t) \in B | Y_{k,i} \;\; i = 1, \ldots, p_k \;, k = 1, \ldots, n)$$

$$= \int_{\mathcal{X}} P(g_k(t) \in B|x)\pi(dx|data) \approx \frac{1}{J} \sum_{j=1}^{J} I(g_k^{(j)}(t) \in B),$$

where B is any Borel set on R^+, $\mathcal{X}$ is the parameter space, and J is a large enough positive integer. A similar formula for t_k^* is also true.

From a computational point of view, it may be more attractive to sample from the algorithm only at times j which are $\tilde{J} > 1$ units apart from each other. Choosing two positive integers, say J_0 and $\tilde{J}$, if J_0 and $\tilde{J}$ are not too small and if the Markov chain is constructed properly, then $\{g_k^{(J_0+j\cdot\tilde{J})}(t) \;;\;\; j = 1, 2, \ldots\}$ can be regarded approximately as a random sample from the predictive distribution of $g_k(t)$. A similar statement holds again for t_k^*.

19.4 An example: growth of crack lengths

This study was largely motivated by the paper from Lu and Meeker (1993), who analyzed the same crack growth data that we are using. This data set was first used by Hudak et al. (1978), and later by Bogdanoff and Kozin (1985). The measurements are the crack lengths in a compact tension steel test operated in different laboratories. Lu and Meeker (1993) got the data visually from Figure 4.5.2 of Bogdanoff and Kozin (1985), and presented the data set in Table 1 of their own paper. Only the measurements up to the "time" of 0.12 million cycles were reported, and a target level corresponding to failure was defined to be 1.6 inches. The data are shown in a graphical form in Figure 2. It should be mentioned that in the original data set there were also measurements beyond these cut-off points. There were altogether 21 test trials, and in 12 of these trials the length of the crack had reached the target level before the end of observation. In all 21 cases, the initial crack length was chosen to be 0.9 inches. Our goal is to reanalyze this data set, by applying the growth curve model and the inferential method introduced above. From Figure 2, it appears that our structural assumption regarding a common shape for the growth curves is quite reasonable.

From the data it is natural to set the value of T_{max} at 0.12. The constant C was here taken to be 2, which turned out to be large enough in the actual computations performed. Each observational interval (of length 0.01) was further divided into 4 subintervals of equal length, i.e., we let $t_i = 0.01 \times i/4$ (see section 2). Unless stated otherwise, we expressed "vague" prior knowledge in specifying the prior distribution, choosing the values of

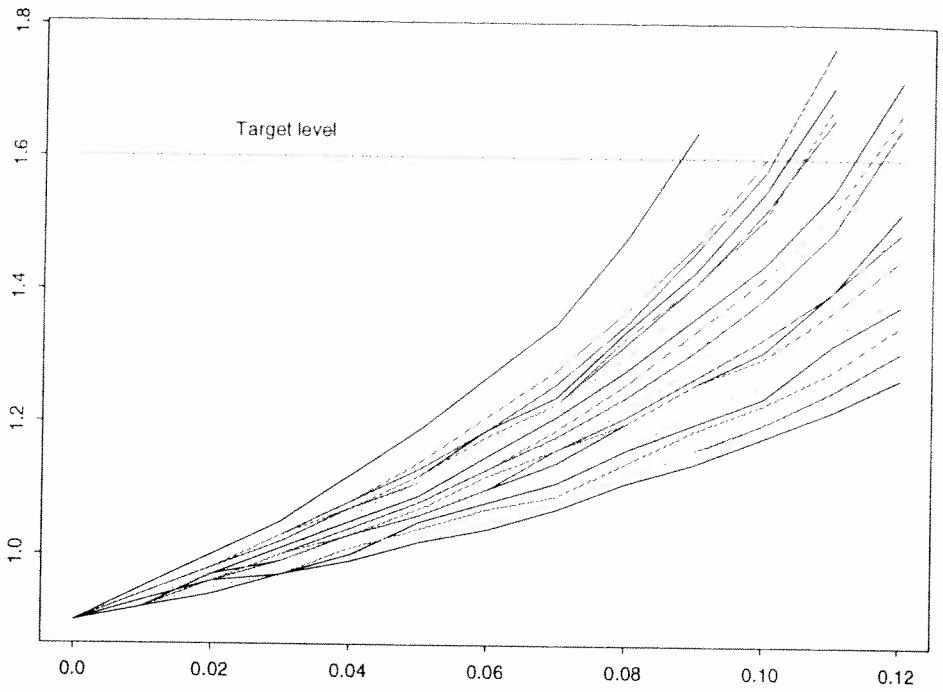

FIGURE 2. Fatigue crack growth data of 21 trials. The figure plots Y_{ki} on the ordinate against t on the abscissa.

the hyperparameters as follows: $\lambda = \rho = 0.05$, $\mu = 20$, $a = 3$ and $b = 0.02$. In this way the prior expectations of α and β are the same (see section 2).

The program was coded in C and run in a Sun Ultra 1 workstation. It ran quite fast, but the burn-in times were sometimes long. In the experiments, we discarded the first $J_0 = 6000$ iterations, picked a sample from every $\tilde{J} = 50$ iterations (see section 3), and did altogether 156000 iterations to get an MCMC sample of 3000 observations. This appeared to be sufficient to lead to stable Monte Carlo estimates.

An obvious concern in a study like this, where there are no strong a priori arguments either in favor or against some particular model, is how to assess the performance of the model we have decided to use. Considering that the primary goal here is prediction, an obvious suggestion is to pretend that one or several readings $Y_{k,i}$ in the data set were actually unknown, then use our model to determine their predictive distributions using the method described in section 3, and finally compare these with the true readings. If $Y_{k,i}$ can really be regarded as a random variable drawn from its (continuous) predictive distribution $F_{Y_{k,i}}$, then $F_{Y_{k,i}}(Y_{k,i})$ is uniformly distributed in (0, 1). So, by varying the readings pretended to be unknown, we should get a "pseudo" (0, 1)-uniform random sample. 21 such simulation trials were performed, one for each growth curve. In the k^{th} trial, we pretended to know all other length readings in the data, except for device k for which only the values up to 0.07 million cycles were assumed known. In this way we obtained 94 "pseudo" (0, 1)-uniform random variables (2 from trial 1, 3 from trial 2, 4 from each trial indexed from 3 to 8, and 5 from each of the remaining trials). The corresponding cumulative empirical distribution is

plotted in Figure 3. Note, however, that since we have repeated observations from the same device, these 94 variables are not independent and therefore straightforward goodness-of-fit testing based on this empirical distribution would not be correct. The main message in this figure appears to be that the predictive distributions tend to be somewhat more dispersed than the actual future length readings. At least in part this must be a consequence of having used a "flat" prior for σ^2 with prior mean $E_{prior}(\sigma^2) = 0.01$. Calculating the Kolmogorov-Smirnov statistic gave $d = 1.374$, which for an independent sample would be close to the 5 percent critical value 1.36. Changing the value of hyperparameter b from 0.02 to 0.0006 to correspond to a smaller measurement error variance σ^2 (with prior mean $E_{prior}(\sigma^2) = 0.0003$) improving the fit of the model somewhat. The value of d was then 0.919, which for an independent sample would be below the 5 percent critical value. The prior and posterior means and standard deviations of σ^2 in the above two cases are listed in Table 1.

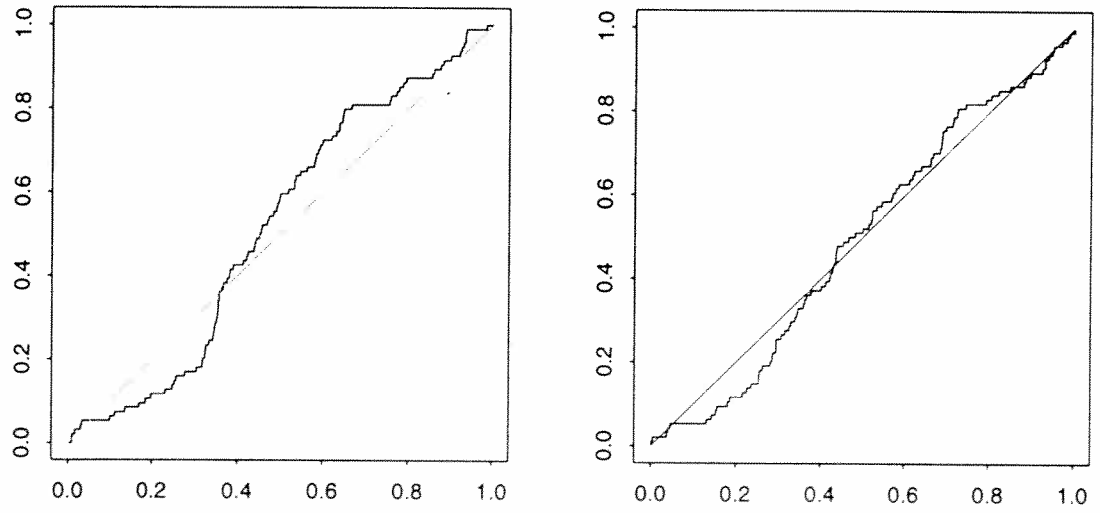

FIGURE 3. Cumulative empirical distributions of 94 "pseudo" (0, 1) uniform random variables. In the left the hyperparameter b was taken to be 0.02, and in the right $b = 0.0006$.

As stated earlier, our main goal was to make prediction. There are two obvious kinds of prediction: one is to predict the true value of $g_k(t)$ at a certain time t, and the other is to predict the crossing (failure) time of an device which has not yet reached the target level Y^* at the end of the trial. Since the first 12 devices in the data had reached the target level of

	Prior		Posterior	
	Mean	Standard error	Mean	Standard error
Experiment 1	0.01	0.01	0.0000285	0.00000245
Experiment2	0.0003	0.0003	0.0000109	0.00000105

TABLE 1. Prior and posterior means and standard deviations of σ^2 in two goodness-of-fit experiments.

$Y^* = 1.6$ inches before 0.12 million test cycles, in the following, predictions are made only for devices indexed from 13 to 21. Figure 4 shows the predictive distributions of the times at which these 9 crack lengths would reach the target level. As this data set is a part of a larger one (see Bogdanoff and Kozin, 1985), the true crossing times are actually known. The dotted straight lines show their positions, and the corresponding quantiles of their predictive distributions.

As an alternative method of assessment, we can use the mean or median of the predictive distribution as a point estimate, and compare this with the observed true value. This enables us to also compare our results with those obtained by Lu and Meeker (1993). Let us denote by T_k the true crossing time of the device indexed by k, by $\hat{T}_k$ Lu and Meeker's estimate of T_k (calculated from the estimated crack growth curves in their article), and by $\tilde{T}_k$ (resp. $\tilde{T}'_k$) our estimate of T_k by using the mean (resp. median) of the predictive distribution. Define the statistics

$$\Delta\hat{T}_k = \sum_{k=13}^{21} (\hat{T}_k - T_k), \qquad \hat{s} = \sqrt{\frac{1}{8}\sum_{k=13}^{21} (\hat{T}_k - T_k)^2};$$

$$\Delta\tilde{T}_k = \sum_{k=13}^{21} (\tilde{T}_k - T_k), \qquad \tilde{s} = \sqrt{\frac{1}{8}\sum_{k=13}^{21} (\tilde{T}_k - T_k)^2}$$

and

$$\Delta\tilde{T}'_k = \sum_{k=13}^{21} (\tilde{T}'_k - T_k), \qquad \tilde{s}' = \sqrt{\frac{1}{8}\sum_{k=13}^{21} (\tilde{T}'_k - T_k)^2}.$$

Then the numerical values obtained here were $\Delta\hat{T}_k = -0.0090$, $\hat{s} = 4.57 \times 10^{-3}$; $\Delta\tilde{T}_k = -0.0041$, $\tilde{s} = 3.51 \times 10^{-3}$ and $\Delta\tilde{T}'_k = -0.0029$, $\tilde{s}' = 3.89 \times 10^{-3}$. From this it seems that our point estimates are slightly more accurate than those provided by Lu and Meeker. Notice that our model assumptions

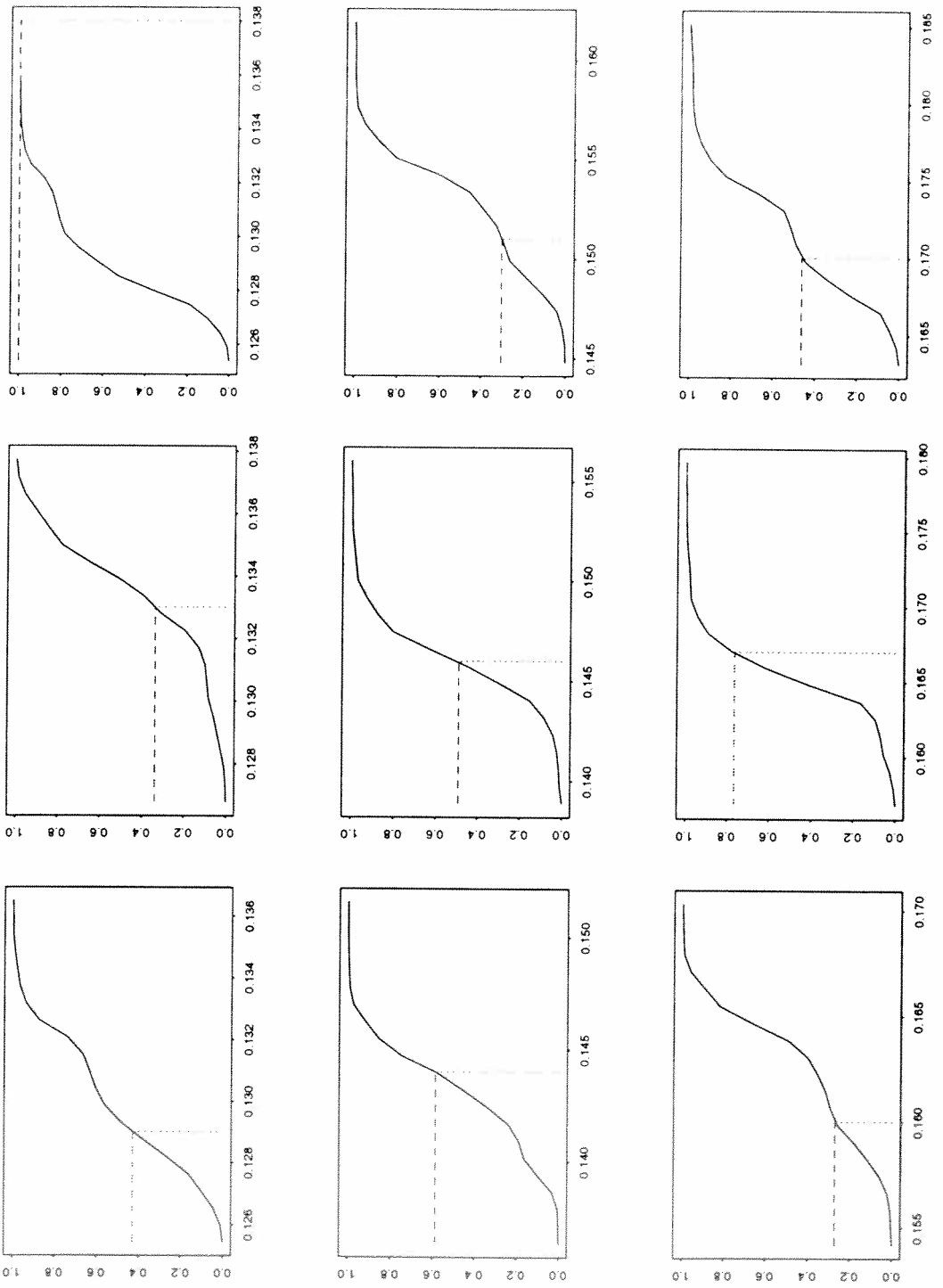

FIGURE 4. Predictive distributions of the crossing times to the target level $Y^* = 1.6$ inches, for trials indexed from 13 to 21. The dotted lines are the true crossing times, and the dashed lines indicate their corresponding quantiles.

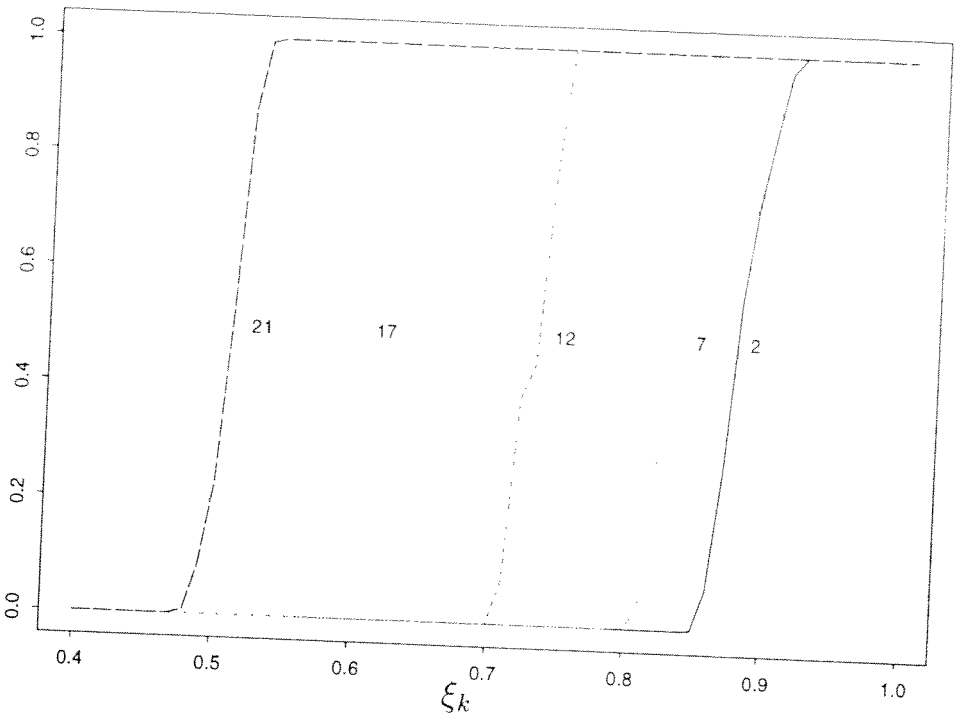

FIGURE 5. Posterior distribution functions for the parameter ξ_k, when $k = 2, 7, 12, 17$ and 21.

were much weaker than theirs. There is one bad prediction, viz. for the crossing time in trial 15. This same value was apparently underestimated also by Lu and Meeker (1993), so that the crossing time of this trial seems to be an "outlier". Some trials with other hyperparameter values were also made, and it seems that different choices of the hyperparameters have little influence on the corresponding results of both kinds of predictions.

Further study about some particular parameters can also be carried out conveniently by using the simulated sample $x^{(j)}$. In Figure 5, we plotted the posterior cumulative distribution functions of the "relative speed" parameters ξ_k for $k = 2$, 7, 12, 17 and 21. They are all less than 1, which was the assumed value of ξ_1. It is easy to see that a curve on the left side corresponds to slow growth, and on the right to fast growth. The posterior variances of the ξ_k's are small, which partly validates our common shape assumption in all growth curves.

19.5 Discussion

A model's ability to predict future growth beyond the point T_{max} at which all observation is terminated is often doubtful. Here, since the baseline $g(\cdot)$ is estimated when the condition $g(t) \leq Y^*$ is true, we can reasonably use our model to predict future growth for those slowly growing devices which were still below Y^* at time T_{max}. In other words, the information in the data seems to be sufficient for reasonable prediction.

Another advantage of the present approach is that different random

quantities in this hierarchical model can be viewed as corresponding to different sources of variation in a crack growth curve, and hence they can be given an intuitive interpretation. The common growth curve g_0 in (1), and later the baseline g in (3), are parameters describing population characteristics, whereas the speed θ_k in (1), and later ξ_k in (3), are individual characteristics representing the (relative) differences between different devices.

We did not estimate what might be described as the time-to-failure distribution of a generic device from the same population as the data, which was a primary goal of the analysis of Lu and Meeker (1993). In our Bayesian setting, this would have corresponded to introducing a hypothetical "new" device about which there was no crack growth data available, and then considering the predictive distribution of its failure time. This could have been done routinely by adding corresponding "dummy" growth curves to the simulation scheme.

We do not claim that our method improves the accuracy of predictions very much. Rather, the emphasis of this paper is on how to establish a hierarchical model under weak assumptions, how to assess the adequacy of the model, and how to make reasonable predictions. However, if some additional knowledge besides the data is available, it could be integrated into the prior distribution, and better predictions would be obtained.

If we add the further structural assumption that the growth curves are all convex, which is indeed strongly suggested by Figure 2, the only thing we need to change is the prior distribution of g, which is then constrained to be in the domain

$$0 < \Delta g(t_1) \leq \Delta g(t_2) \leq \cdots \leq \Delta g(t_N).$$

The corresponding posterior density is simply the original density restricted to the domain above and divided by a normalizing constant. In Figure 6, we have plotted 1000 simulated growth curves for the 13^{th} trial, by using the original model (3), and the model based on this further convexity assumption, respectively. It is not surprising that, from the figure as well as from the numerical results, it makes almost no difference whether the convexity condition is added to the prior or not, since the growth curves sampled from the posterior exhibit an apparent convex shape in any case. The convexity message convoyed by the data seems to be so strong that it is unnecessary for the prior to do the same. The situation becomes of course different if we are actually aiming at predicting beyond the target level Y^*, where there is no data.

In Bogdanoff and Kozin (1985), there is a second crack growth data set, which likewise appears to be consistent with our assumption of a common shape for the growth curves. Although this assumption is weaker than in most growth models considered elsewhere, it still puts a tight restriction on the "operational ages" $\theta_k \cdot t$. Arjas et al. (1996) used stochastic processes to characterize these "operational ages", so that their model is more flexible than the present one in fitting different kinds of empirical data. Here,

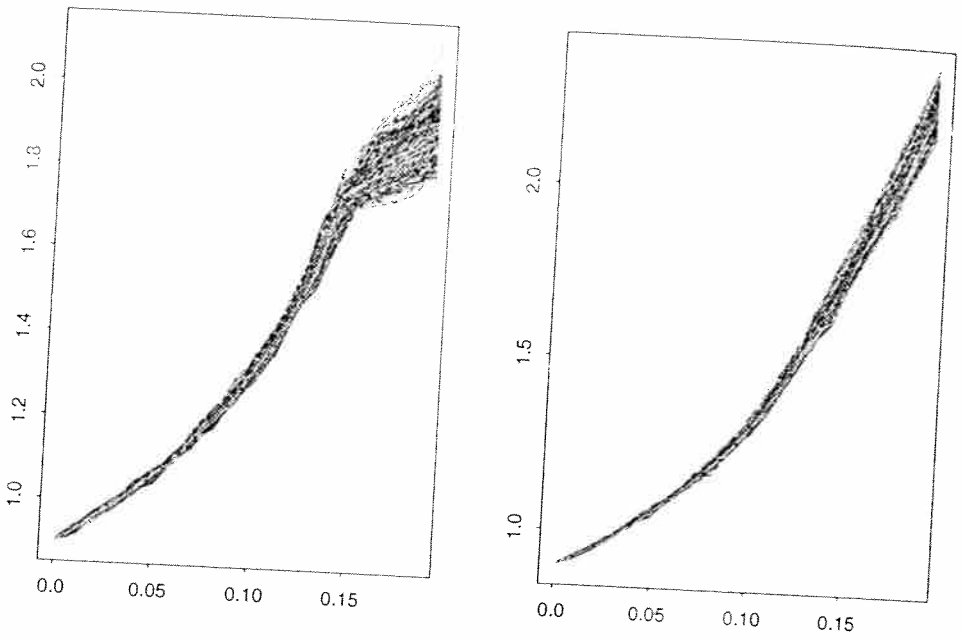

FIGURE 6. Simulated fatigue crack growth curves of size 1000 for trial 13, when the original model and the model with further convexity assumption are used respectively. The figure plots Y_{ki} on the y-axis against t on the abscissa.

however, the much simpler model works quite well, especially in prediction. We believe that, if the data set were less regular, a similar technique as in Arjas et al. (1996) could be used to tackle the problem successfully.

19.6 Acknowledgements

This research was undertaken while the first author was visiting the Department of Mathematical Sciences, University of Oulu, Finland, and was supported by grants from the Academy of Finland.

References

Arjas, E., and Gasbarra, D. (1994), "Nonparametric Bayesian Inference from Right Censored Survival Data, Using the Gibbs Sampler," *Statistica Sinica*, 4, 505-524.

Arjas, E., Liu, L., and Maglaperidze, N. (1996), "Prediction of growth: a Hierarchical Bayesian Approach," *to appear in Biometrical Journal.*

Bernardo, J.M., and Smith, A.F.M. (1994), *Bayesian Theory*, West Sussex: John Wiley & Sons Ltd.

Besag, J., Green, P., Higdon, D., and Mengerson, K. (1995), "Bayesian Computation and Stochastic Systems," *Statistical Science*, 10, 3-66.

Bogdanoff, J.L., and Kozin, F. (1985), *Probabilistic Models of Cumulative Damage*, New York: John Wiley.

Fearn, T. (1975), "A Bayesian Approach to Growth Curves," *Biometrika*, 62, 89-100.

Hudak, S.J., Jr., Saxena, A., Bucci, R.J., and Malcolm, R.C. (1978), "Development of Standard Methods of Testing and Analyzing Fatigue Crack Growth Rate Data," Technical report AFML-TR-78-40, Westinghouse R & D center, Westing house electric corporation, ittsburgh, PA 15235.

Lindley, D.V., and Smith, A.F.M. (1972), "Bayes Estimates for the Linear Model," *Journal of the Royal Statistical Society, Series B*, 34, 1-42.

Lu, C.J., and Meeker, W.Q. (1993), "Using Degradation Measures to Estimate a Time-to-Failure Distribution," *Technometrics*, 35, 161-174.

Rao, C.R. (1987), "Prediction of Future Observations in Growth Curve Models," *Statistical Science*, 2, 434-471.

Roberts, G. O., and Smith, A. F. M. (1994), "Simple Conditions for the Convergence of the Gibbs Sampler and Metropolis-Hastings Algorithm," *Stochastic Processes and their Applications*, 49, 207-216.

Smith, A.F.M., and Roberts, G.O. (1993), "Bayesian Computation via the Gibbs Sampler and Related Markov Chain Monte Carlo Methods," *Journal of the Royal Statistical Society, Series B*, 55, 3-23.

Tian, J.J., Shukla, R., and Buncher, C.R. (1994), "On Prediction of Future Observation in Growth Curve Model," *Statistics in Medicine*, 13, 2205-2217.

von Rosen, D. (1991), "The Growth Curve Model: a Review," *Communications in Statistics, Theory and Methods*, 20, 2791-2822.