The Annals of Statistics
1981, Vol. 9, No. 2, 356–367

A BAYESIAN NONPARAMETRIC APPROACH TO RELIABILITY

BY R. L. DYKSTRA[1] AND PURUSHOTTAM LAUD

University of Missouri and University of Northern Illinois

It is suggested that problems in a reliability context may be handled by a Bayesian nonparametric approach. A stochastic process is defined whose sample paths may be assumed to be increasing hazard rates by properly choosing the parameter functions of the process. The posterior distribution of the hazard rates is derived for both exact and censored data. Bayes estimates of hazard rates and cdf's are found under squared error type loss functions. Some simulation is done and estimates graphed to better understand the estimators. Finally, estimates of the hazard rate from some data in a paper by Kaplan and Meier are constructed.

1. Introduction. Recently, the Bayesian nonparametric approach to statistical inference has received a good deal of attention. In this approach, a stochastic process is defined whose sample paths index distributions. Thus the distribution of the process serves as a prior over the indexed family. The goal of this approach is to obtain the scope and robustness of nonparametric procedures along with the mathematical elegance inherent in Bayesian methods.

The most common Bayesian nonparametric approach has been extensively discussed by Ferguson (1973), and consists of using a 'Dirichlet Process' prior: Ferguson defines a continuous time parameter stochastic process whose finite dimensional increments have a Dirichlet distribution. The sample paths of this process correspond to univariate probability measures. Ferguson shows that the posterior distribution of the process, given the complete observations, is also distributed as a Dirichlet stochastic process, and uses this posterior distribution for making statistical inferences. Doksum (1974) addresses his attention to prior stochastic processes that are 'tailfree' and/or 'neutral'. Susarla and Van Ryzin (1976) were able to obtain the posterior mean of censored data using a Dirichlet prior. Recently, Ferguson and Phadia (1976) were able to generalize these censored data results to more general "neutral to the right" processes.

This type of approach seems to have merit concerning statistical inference in a reliability context. What we propose, since the concept of hazard rate plays such a key role in statistical reliability, is to place the prior probability over the collection of hazard rates. This is done by defining an appropriate stochastic process whose sample paths are hazard rates. With this prior we derive the posterior distribution of the hazard rates for both right-censored and exact observations. This approach has the advantage of placing the prior probability on absolutely continuous rather than on discrete distributions, as is the case with the Dirichlet process prior. Moreover, Bayes estimators of the entire distribution under natural loss functions are absolutely continuous. We note that, since our prior random cdf's are not neutral to the right, the work of Doksum (1974) and Ferguson and Phadia (1979) does not apply.

Received March 1977; revised November 1979.

[1] The author's work was supported by the Office of Naval Research under Contract N00014-78-C-0655.

AMS 1970 *subject classifications.* Primary 62G99; secondary 62F15.

Key words and phrases. Hazard rates, increasing hazard rates, Bayes estimates, extended gamma process, posterior process, prior process.

2. The extended gamma process. Let $F(\cdot)$ be a left-continuous cdf such that $F(x) = 0$ for $x \le 0$. The survival function and cumulative hazard function corresponding to F are defined by $\bar{F}(x) = 1 - F(x)$ and $H(x) = -\ln \bar{F}(x)$. If for all x,

$$H(x) = \int_{[0,x)} r(t)\, dt,$$

then $r(x)$ is called the hazard rate of the distribution and in a reliability context indicates the propensity for failure of an item in the near future given that the item has survived till time x.

Let $G(\alpha, \beta)$ denote the gamma distribution with density

$$g(x \mid \alpha, \beta) = x^{\alpha-1} \exp(-x/\beta) I_{(0,\infty)}(x)/\Gamma(\alpha)\beta^{\alpha},$$

for $\alpha, \beta > 0$; $G(0, \beta)$ denotes a distribution degenerate at 0.

Let $\alpha(t)$, $t \ge 0$, be a nondecreasing left-continuous real-valued function such that $\alpha(0) = 0$, and let $\beta(t)$, $t \ge 0$, be a positive right-continuous real-valued function, bounded away from 0 with left-hand limits existing.

Let $Z(t)$, $t \ge 0$, defined on an appropriate probability space $(\Omega, \mathscr{F}, P)$, denote a gamma process with independent increments corresponding to $\alpha(t)$. That is, $Z(0) \equiv 0$, $Z(t)$ has independent increments, and for $t > s$, $Z(t) - Z)(s)$ is $G(\alpha(t) - \alpha(s), 1)$. It can be shown (see Ferguson (1973)) that such a process exists and that its distribution is uniquely determined. We assume WLOG that this process has nondecreasing left-continuous sample paths.

We now define a new stochastic process by

$$r(t) = \int_{[0,t)} \beta(s)\, dZ(s), \tag{2.1}$$

where the integration is with respect to the sample paths of the $Z(t)$ process. We say a process defined in this manner has an extended gamma distribution, and we denote such a process by saying $r(t)$ is $\Gamma(\alpha(\cdot), \beta(\cdot))$.

Of course if $r(t)$ is taken to be a random hazard rate, there will be a corresponding random cdf given by

$$F(x) = 1 - \exp\left[-\int_{[0,x)} r(t)\, dt\right].$$

From Doksum's work, $F(x)$ will be neutral to the right only if $H(x) = \int_{[0,x)} r(t)\, dt$ has independent increments. It is easily seen that even though $r(t)$ has independent increments, $H(x)$ will not, and hence the distributional results of Doksum will not apply.

The finite dimensional cdf's (or densities) of $r(t)$ appear to be rather intractable, although the distribution of the extended gamma process is "nice" in many ways.

THEOREM 2.1. *If $r(t)$ is distributed as $\Gamma(\alpha(\cdot), \beta(\cdot))$, then $r(t)$ has independent increments and for fixed t*

(2.2) *the characteristic function of $r(t)$ in some neighborhood of 0 is given by* $\psi_{r(t)}(\theta) = \exp[-\int_{[0,t)} \ln(1 - i\beta(s)\theta)\, d\alpha(s)]$,

$$\mathrm{E}r(t) = \int_{[0,t)} \beta(s)\, d\alpha(s), \tag{2.3}$$

and

$$\text{Var } r(t) = \int_{[0,t)} \beta^2(s)\, d\alpha(s). \tag{2.4}$$

The proof follows by first defining a sequence of partitions $0 = t_{n,1} < t_{n,2} < \cdots < t_{n,k(n)}$ whose norm goes to 0 and whose upper end point goes to ∞. Then the sequence of random functions

$$r_n(t) = \sum_{\{i>0; t_{n,i}<t\}} \beta(t_{n,i})[Z(t_{n,i}) - Z(t_{n,i-1})] \tag{2.5}$$

is defined. Appropriate limiting arguments then complete the proof.

Since the original gamma process $Z(t)$ is a pure jump process, the extended gamma process will also be a pure jump process.

3. Random hazard rates. Provided $\alpha(t)$ is not identically zero, we may assume that the sample paths of an extended gamma process $r(t)$ are well defined nondecreasing hazard rates corresponding to absolutely continuous distributions. Thus the conditional distribution of the observations $X_1, \cdots, X_n$ given $r(t)$ will be defined by

$$P(X_1 \geq x_1, \cdots, X_n \geq x_n \mid r(t)) = \prod_{i=1}^n \exp\left[-\int_{[0,x_i)} r(t)\, dt\right]. \tag{3.1}$$

Of course (3.1) and the distribution of $r(t)$ will determine the joint distribution of $X_1, \cdots, X_n, r(t)$ and will be used to derive the marginal distribution of $X_1, \cdots, X_n$ and the posterior distribution of $r(t)$ given the observed values of $X_1, \cdots, X_n$. Since the sample paths of the $r(t)$ process are nondecreasing functions a.s., we are placing our prior probability entirely within the class of distributions with nondecreasing hazard rates.

In assigning a prior probability measure by this method, one needs to input the functions $\alpha(t)$ and $\beta(t)$. One approach consists of defining nondecreasing mean and variance functions $\mu(t)$ and $\sigma^2(t)$. It would seem reasonable to assign as $\mu(t)$ the best "guess" of the hazard rate and use $\sigma^2(t)$ to measure the amount of uncertainty or variation in the hazard rate at the point t. Assuming $\mu(t)$, $\sigma^2(t)$ and $\alpha(t)$ are all differentiable, one can use (2.3) and (2.4) to set

$$\mu(t) = \int_{[0,t)} \beta(s)\alpha'(s)\, ds,$$

and

$$o^2(t) = \int_{[0,t)} \beta^2(s)\alpha'(s)\, ds.$$

Solving for $\alpha(t)$ and $\beta(t)$ yields

$$\beta(t) = \frac{d\sigma^2(t)}{dt} \bigg/ \frac{d\mu(t)}{dt}, \tag{3.2}$$

and

$$\frac{d\alpha(t)}{dt} = \left[\frac{d\mu(t)^2}{dt}\right] \bigg/ \frac{d\sigma^2(t)}{dt}, \tag{3.3}$$

which then determines the prior distribution. The form of the posterior distribution gives information on the effect of the prior and may help in choosing $\alpha(\cdot)$ and $\beta(\cdot)$.

The marginal distribution of an observation X can be found from (3.1) with the use of a limiting argument.

THEOREM 3.1. *If the prior over hazard rates is* $\Gamma(\alpha(\cdot), \beta(\cdot))$ *then the marginal survival function of an observation X is given by*

$$\bar{F}(t) = P(X \geq t) = \exp[-\int_{[0,t)} \ln(1 + \beta(s)(t - s))\, d\alpha(s)]. \tag{3.4}$$

The marginal survival function of the observations $X_1, \cdots, X_n$ can be found by methods similar to those used in Theorem 3.1 and is given in the following corollary.

COROLLARY 3.1. *If the prior over the hazard rates is* $\Gamma(\alpha(\cdot), \beta(\cdot))$, *then the joint marginal survival function of n observations* $X_1, \cdots, X_n$ *is*

$$(\bar{F}(t_1, \cdots, t_n) = P(X_1 \geq t_1, \cdots, X_n \geq t_n) = \exp[-\int_{[0,\infty)} \ln(1 + \beta(s) \textstyle\sum_{i=1}^{n} (s - t_i)^+)\, d\alpha(s)] \tag{3.5}$$

where $a^+ = \max\{a, 0\}$.

Thus the marginal survival function of $Y = \min(X_1, \cdots, X_n)$ is of the same form as the survival function of X_1 providing $\beta(s)$ is replaced by $n\beta(s)$.

The key problem in any Bayesian setting is to derive the posterior distribution. Moreover it is important to handle censored observations since reliability data are often of this type. If an extended gamma prior is used, the posterior distribution for right-censored observations is also an extended gamma process. The proof is given in Section 7.

THEOREM 3.2. *If the prior over the hazard rates is* $\Gamma(\alpha(\cdot), \hat{\beta}(\cdot))$, *then the posterior over the hazard rates, given m censored observations of the form* $X_1 \geq x_1, X_2 \geq x_2, \cdots, X_m \geq x_m$, *is* $\Gamma(\alpha(\cdot), \hat{\beta}(\cdot))$ *where*

$$\hat{\beta}(t) = \frac{\beta(t)}{1 + \beta(t) \sum_{i=1}^{m} (x_i - t)^+}. \tag{3.6}$$

The effect of censored observations is thus to decrease the slope of the sample paths to the left of the censoring points while leaving the slope of the values to the right unchanged.

We next address ourselves to the question of the posterior distribution of $r(t)$ given exact observations. The following states that the posterior can be expressed as a continuous mixture of extended gamma distributions. The dimension of the mixing measure increases with sample size. The proof is given in Section 6.

THEOREM 3.3. *If the prior over the hazard rates is* $\Gamma(\alpha(\cdot), \beta(\cdot))$, *then the posterior over the hazard rates, given m observations of the form* $X_1 = x_1, \cdots, X_m = x_m$ *is a mixture of extended gamma processes. The distribution of the mixture is given by*

$$P(r(t) \in B \mid X_1, = x_1, \cdots, X_m = x_m) \tag{3.7}$$

$$= \frac{\int_{[0,x_m)} \cdots \int_{[0,x_1)} \prod_{i=1}^{m} \hat{\beta}(z_i) F(B; \Gamma(\alpha + \sum_{i=1}^{m} I_{(z_i,\infty)}, \hat{\beta})) \prod_{i=1}^{m} d[\alpha + \sum_{j=i+1}^{m} I_{(z_j,\infty)}](z_i)}{\int_{[0,x_m)} \cdots \int_{[0,x_1)} \prod_{i=1}^{m} \hat{\beta}(z_i) \prod_{i=1}^{m} d[\alpha + \sum_{j=i+1}^{m} I_{(z_j,\infty)}](z_i)}$$

Here $F(B; Q)$ *denotes the probability assigned to* $B \in \mathscr{B}_R$ *by a stochastic process which is distributed as* Q, $\hat{\beta}(\cdot)$ *is defined as in* (3.6), *and the iterated integrations are done with respect to* z_1 *through* z_m. *Of course* $\sum_{j=m+1}^{m} I_{(z_j,\infty)}(z_i) \equiv 0$.

The complexity of this distribution makes it difficult to see how a particular observation

$X_1 = x_1$ affects the posterior. A failure at time x_1 serves to increase the hazard rate prior to x_1. This increase is smaller for smaller t. This is evidenced by the weight function $\hat{\beta}(t) = \beta(t)[1 + \beta(t)(x_1 - t)^+]^{-1}$ in the mixing integral. The above effect is tempered by the rate at which $\alpha(t)$ increases so that $\hat{\beta}(t)$ and $\alpha(t)$ together determine where and how the increase in risk (the unit jump in the α function) occurs.

4. Bayes estimators. A natural loss function when estimating a hazard rate is the squared error type loss function used in Ferguson (1973) for distribution functions

$$L(r, \hat{r}) = \int_{[0,\infty)} (r(t) - \hat{r}(t))^2 \, dW(t) \tag{4.1}$$

where W is an arbitrary finite measure on $[0, \infty)$ such that

$$\int_{[0,\infty)} \int_{[0,t)} \beta^2(s) \, d\alpha(s) \, dW(t) < \infty.$$

The Bayes estimator $\hat{r}(t)$ which minimizes the expected loss is given by the posterior mean of $r(t)$.

With no censored observations, we may use the form of $Er(t)$ in (2.3) and the fact that the mean of a mixture of distributions is the mixture of the means (assuming existence) to express $\hat{r}(t)$ as

$$\hat{r}(t) = \frac{\int_{[0,x_m)} \cdots \int_{[0,x_1)} \int_{[0,t)} \prod_{i=0}^m \hat{\beta}(z_i) \prod_{i=0}^m d\left[\alpha(z_i) + \sum_{j=i+1}^m I_{(z_j,\infty)}(z_i)\right]}{\int_{[0,x_m)} \cdots \int_{[0,x_1)} \prod_{i=1}^m \hat{\beta}(z_i) \prod_{i=1}^m d\left[\alpha(z_i) + \sum_{j=i+1}^m I_{(z_j,\infty)}(z_i)\right]} \tag{4.2}$$

where integration is performed respectively with respect to $z_0, z_1, \cdots, z_m$ and $\hat{\beta}$ is given by (3.6).

Note that the denominator is the same form as the numerator, though the integral is of smaller dimension. Obviously $\hat{r}(t)$ is a nondecreasing function of t as expected. Censored observations can be incorporated in $\hat{r}(t)$ by defining $\hat{\beta}$ as in (3.6) where the x_i's include both censoring points and complete observations.

It would appear that the utility of this estimate is severely limited since it involves a multi-dimensional integral. We shall show in the next section, however, that $\hat{r}(t)$ is expressible in a manner that involves only one-dimensional integrals.

If the prime consideration is predictive in nature, the solution is different. Suppose

$$\bar{F}^*(t) = P(X_{n+1} \geq t \mid X_1 = x_1, \cdots, X_n = x_n) \tag{4.3}$$

denotes the conditional survival function of a future observation given n current observations. Then $\bar{F}^*(t)$ will be the function that minimizes $E \int_{[0,\infty)} (\bar{F}(t) - \hat{\bar{F}}(t))^2 \, dW(t)$ for any finite measure W. Thus $\bar{F}^*(t)$ is also the Bayes estimator of the survival function when the loss function is of the squared error type over cdf's. Diligent computation shows that

$$\bar{F}^*(t) = \exp\left[\int_{[0,\infty)} \ln(1 + \hat{\beta}(z_0)(t - z_0)^+) \, d\alpha(z_0)\right] \cdot J(\hat{\beta}^*)/J(\hat{\beta}) \tag{4.4}$$

where

$$J(\beta) = \int_{[0,x_n)} \cdots \int_{[0,x_1)} \prod_{i=1}^n \beta(z_i) \prod_{i=1}^n d[\alpha(z_i) + \textstyle\sum_{j=i+1}^n I_{(z_j,\infty)}(z_i)],$$

$\hat{\beta}$ is defined by (3.6) and $(1 + \hat{\beta}(z_i)(t - z_i)^+ \hat{\beta}^*(z_i) = \hat{\beta}(z_i)$. Similarly, Corollary 3.1 can be used to express the joint survival function of k future observations $X_{n+1}, \cdots, X_{n+k}$ conditional on the observed data. Thus

$$\bar{F}^*(t_{n+1}, \cdots, t_{n+k}) = P(X_{n+1} \geq t_{n+1}, \cdots, X_{n+k} \geq t_{n+k} \mid X_1 = x_1, \cdots, x_n = x_n) \tag{4.5}$$

is of the same form as (4.4) with $(t - z_i)^+$ replaced by $\sum_{j=n+1}^{n+k} (t_j - z_i)^+$. One of the consequences is that the minimum of k future observations has the conditional survival function given in (4.4) with $\hat{\beta}$ replaced by $k\hat{\beta}$.

The fact that $\hat{\beta}^*$ in (4.4) is a nonincreasing function which is equal to $\hat{\beta}$ at 0 guarantees that $\bar{F}^*(t)$ is a bonafide survival function. The first factor of $\bar{F}^*(t)$ in (4.4) would be the survival function of a future observation were the observations censored at $x_1, \cdots, x_n$ rather than observed. Thus, the second factor gives the information gained by observing "deaths" rather than "losses" using the terminology of Kaplan and Meier (1958).

5. Computation and simulation. The presence of multi-dimensional integrals in our estimates would appear to make computation extremely difficult. The following theorem enables us to work with integrals of only one dimension. The integrands are powers of the $\hat{\beta}$ function and the integration is with respect to the α measure.

THEOREM 5.1. *If $\alpha(\cdot)$ and $\beta(\cdot)$ are defined as in Section 2, then*

$$\int_{[0,x_n)} \cdots \int_{[0,x_1)} \prod_{i=1}^n \hat{\beta}(z_i) \prod_{i=1}^n d[\alpha(z_i) + \textstyle\sum_{j=i+1}^n I_{(z_j,\infty)}(z_i)] \tag{5.1}$$

$$= \textstyle\sum_{\mathbf{e}} k(\mathbf{e}) \left[\prod_{I(\mathbf{e})} \int_{[0,x_i)} \hat{\beta}(t)^{e_i}\, d\alpha(t) \right]$$

where $0 < x_n \leq x_{n-1} \leq \cdots \leq x_1 < \infty$ and the sum is over all vectors $\mathbf{e} = (e_1, \cdots, e_n)$ of nonnegative integers such that $\sum_{i=1}^j e_i \leq j$, $j = 1, \cdots, n-1$ and $\sum_{i=1}^n e_i = n$. Here $I(\mathbf{e}) = \{j : e_j \geq 1\}$ and

$$k(\mathbf{e}) = \textstyle\prod_{I(\mathbf{e})} P(j - 1 - \sum_1^{j-1} e_i, e_j - 1) = \prod_{I(\mathbf{e})} [(j-1) - \sum_1^{j-1} e_i]!/[j - \sum_1^j e_j]! \tag{5.2}$$

where $P(n, r)$ denotes the number of permutations of n things taken r at a time.

The proof follows from integrating with respect to the indicator functions and applying combinatorial techniques.

EXAMPLE. Consider the very specialized case where $\alpha(\cdot)$ is flat except for a jump at zero. That is, $\alpha(0) = 0$ and $\alpha(x) = \alpha$, $x > 0$. In this case, $r(t)$ is a constant function whose value is a $G(\alpha, \beta(0))$ random variable. The only value of $\beta(t)$ that matters is $\beta(0) = \beta$. Since the parameter in an exponential distribution is just its constant failure rate, this is equivalent to putting a $G(\alpha, \beta)$ prior over the parameter θ of an exponential density.

If we have complete observations at $x_1, \cdots, x_n$ and censored observations at $x_{n+1}, \cdots, x_{n+m}$, we can specify the posterior distribution of $r(t)$ from Theorems 3.2 and 3.3. Since the posterior of an exponential distribution with a gamma prior is again gamma, the distribution of $r(t_0)$, $t_0 > 0$, specified by the mixture in Theorem 3.3 must also be a gamma distribution.

In this example, the Bayes estimate of $r(t)$, $t > 0$, is a constant (free of t) and may be expressed in terms of Theorem 5.1.

Let $\#\mathbf{e}$ denote the number of nonzero components of $\mathbf{e}$. Then from Theorem 5.1, the numerator of $\hat{r}(t)$ in (4.2) equals

$$\hat{\beta}(0)^{n+1} \textstyle\sum_{\mathbf{e}} k(\mathbf{e}) \alpha^{\#\mathbf{e}} = \hat{\beta}(0)^{n+1} \sum_{i=1}^{n+1} \alpha^i \sum_{\{\mathbf{e};\#\mathbf{e}=i\}} k(\mathbf{e}).$$

However, it can be shown that $\sum k(\mathbf{e})$, whose sum is over $\{e; \#e = i\}$, is the coefficient of Z^i in $Z(Z+1)(Z+2)\cdots(Z+n)$ (the modulus of Sterling numbers of the first kind). Thus, the numerator of $\hat{r}(t)$ equals

$$\beta^{n+1}(1+\beta \textstyle\sum_1^{n+m} x_i)^{-(n+1)}\alpha(\alpha+1)\cdots(\alpha+n).$$

By similar treatment, the denominator of $\hat{r}(t)$ equals

$$\beta^{n}(1+\beta \textstyle\sum_1^{n+m} x_i)^{-n}\alpha(\alpha+1)\cdots(\alpha+n-1).$$

Thus

$$\hat{r}(t) = \beta(\alpha+n)(1+\beta \textstyle\sum_1^{n+m} x_i)^{-1} = (\alpha+n)(\beta^{-1}+\sum_1^{n+m} x_i)^{-1}, \qquad t>0.$$

This agrees with the posterior mean for uncensored data given by Mann, Schafer, and Singpurwalla (1974, page 414). To conclude the example, note that as $n \to \infty$, $\hat{r}(t) \sim$ [total number of failures]/[total time on test].

In order to observe the performance of our Bayes estimators, samples from Weibull and exponential distributions were taken and the corresponding Bayes estimators computed. In all cases the sample size was 11 and the prior parameter functions $\alpha(t) = t$, and $\beta(t) \equiv 2$ were used. Thus the expected value of the prior hazard rate would be $\int_{[0,t)} \beta(s)\, d\alpha(s) = 2t$. This is the hazard rate of a Weibull distribution with mean .8862. All observations were complete (not censored). If one decreases $\beta(\cdot)$ and increases $\alpha(\cdot)$ in such a way that the mean of the prior $\int_{[0,t)} \beta(s)\, d\alpha(s)$ is unchanged, the variance of the prior $\int_{[0,t)} \beta^2(s)\, d\alpha(s)$ will be decreased. This specifies a more precise prior distribution and hence the prior will have more influence in posterior estimates.

Bayes estimates are computed under both loss functions, i.e., integral squared error loss on hazard rates and cdf's. The hazard rate corresponding to the Bayes estimate of the cdf is graphed along with the estimated hazard rate for the purpose of comparison. Thus on Figures 1 and 2, the posterior Bayes estimate of the hazard rate is denoted by a solid line, while the hazard rate which corresponds to the posterior Bayes estimate of the cdf is denoted by the line made up of alternate dashes and plusses. Figure 1 depicts the Bayes estimates of the hazard rate when the random sample comes from a Weibull distribution with failure rate $3t$ while Figure 2 depicts the Bayes estimates of the hazard rate when the sample comes from an exponential distribution with the same mean. The estimated failure rates in Figure 1 conform fairly well to the true failure rate of $3t$ even though the mean of the prior process was different from the true rate. The estimates are quite responsive to the data and yet reasonably smooth. When the sample comes from an exponential distribution rather than a Weibull, the estimates are more nearly constant, as one would hope. Finally, Figure 3 depicts estimates of the hazard rate for the data given by Kaplan and Meier (1958). The prior was arbitrarily taken to be $\alpha(t) = t$, $\beta(t) \equiv 4$ and the starred lines indicate censored values. A slight peaking occurs in estimates of the hazard rate at complete observations. In comparing our estimates with those of the cdf given by Susarla and Van Ryzin (1979) and Furguson and Phadia (1979), ours are somewhat closer to the Kaplan-Meier product limit estimate. Also, our estimates have the advantage of being continuous.

6. Proofs of theorems. In this section we consider stochastic processes defined on an appropriate probability space $(\Omega, \mathscr{F}, P)$. We use $\mathscr{R}^{\mathscr{R}}$ to denote the set of all nonnegative functions on the nonnegative real line $\mathscr{R}$ and $B_{\mathscr{R}}$ to denote the usual smallest α-algebra generated by finite dimensional cylinder sets. The distribution of a stochastic process r is the measure induced on $(\mathscr{R}^{\mathscr{R}}, B_{\mathscr{R}})$ by the measurable function $r: \Omega \to \mathscr{R}^{\mathscr{R}}$. Since, with probability one, the sample paths $r(t, \omega)$ of our stochastic process are failure rates we can define a probability measure $\tilde{P}$ on the product space $(\mathscr{R}^{\mathscr{R}} \times \mathscr{R}, \mathscr{B}_{\mathscr{R}} \times \mathscr{B})$ by extending $\tilde{P}(B \times C) = \int_A F_\omega(C)\, dP(\omega)$ to the usual product σ-algebra of $\mathscr{B}_{\mathscr{R}}$ and $\mathscr{B}$. Here $A = r^{-1}(B)$, $B \in \mathscr{B}_{\mathscr{R}}$ and $F_\omega(C)$ is the probability assigned to $C \in \mathscr{B}$ by the distribution

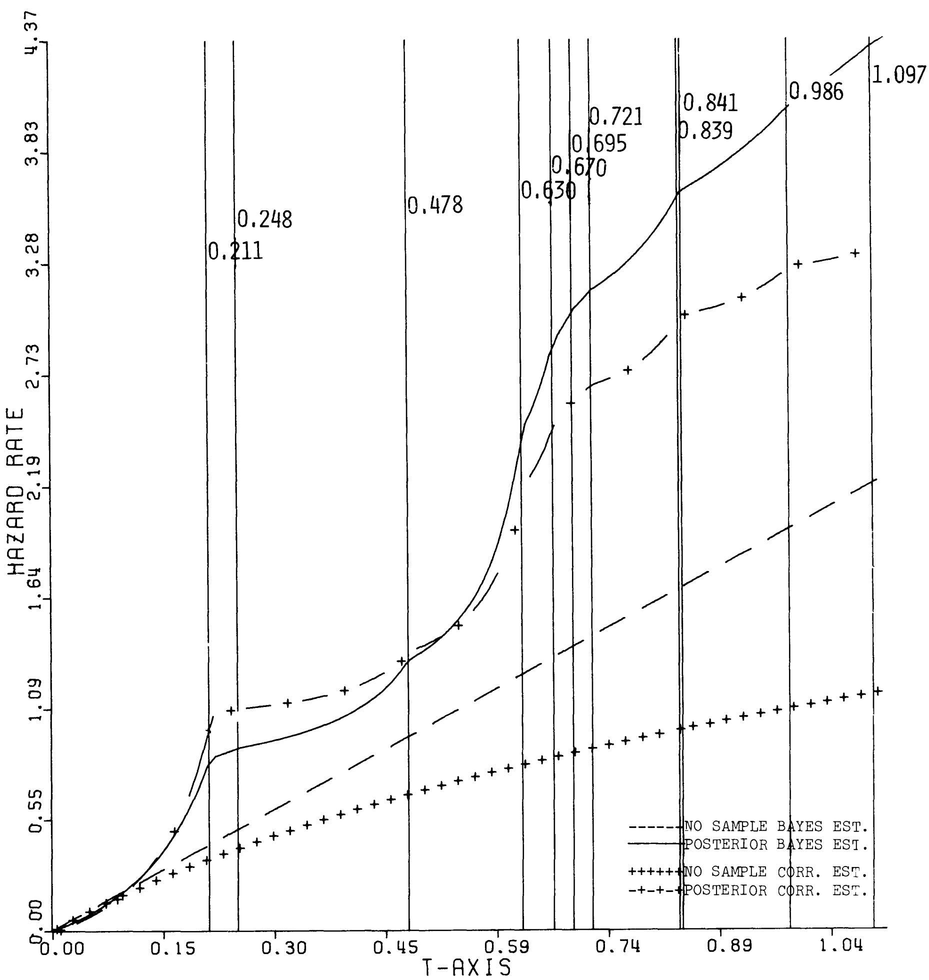

FIG. 1. Sample from Weibull distribution with hazard rate $r(t) = 3t$ and mean $\mu = .7236$.

Prior: $\alpha(t) = t$, $\beta(t) = 2$

function corresponding to $r(\cdot, \omega)$. A marginal probability measure on $(\mathscr{R}, \mathscr{B})$ is determined by $\hat{P}(C) = \tilde{P}(\mathscr{R}^{\#} \times C) = \int_\Omega F_\omega(C)\, dP(\omega)$, $C \in \mathscr{B}$. The posterior distribution of the process for a single observation is a function $\phi(\cdot, \cdot): \mathscr{B}_{\#} \times \mathscr{R} \to [0, 1]$ such that (i) ϕ is Borel-measurable in the second argument, (ii) for each fixed $X \in \mathscr{R}$, $\phi(\cdot, x)$ is a probability measure on $(\mathscr{R}^{\#}, \mathscr{B}_{\#})$ and (iii) $\int_C \phi(B, x)\, d\hat{P}'(x) = \tilde{P}(B \times C)$ for all $B \in \mathscr{B}_{\#}$ and $C \in \mathscr{B}$. The extension for several observations is straightforward.

For convenience we adopt the following notation:

(i) $g(x; \alpha, \beta) = x^{\alpha-1}\exp(-x/\beta)I_{[0,\infty)}(x)/\alpha^{\beta}\Gamma(\alpha)$; $g(x; \alpha) = g(x; \alpha, 1)$.

(ii) $\Delta\alpha_i = \alpha(t_i) - \alpha(t_{i-1})$, t_i abbreviates $t_i^{(n)}$.

(iii) $\beta_i = \beta(t_i)$.

(iv) $\sum = \sum_{i=1}^{k(n)}$, $\prod = \prod_{i=1}^{k(n)}$.

(v) $B_n(\mathbf{u}, \beta, \tau, \mathbf{y}) = \{(u_1, \cdots, u_{k(n)}) \in \mathscr{R}^{k(n)}: \sum \beta_i u_i I_{[0,\tau_1)}(t_i) > y_1, \sum \beta_i u_i I_{[\tau_1,\tau_2)}(t_i) > y_2, \cdots, \sum \beta_i u_i I_{[\tau_{k-1},\tau_k)}(t_i) > y_k\}$. Often $B_n(\mathbf{u}, \beta, \tau, \mathbf{y})$ is abbreviated $B_n(\mathbf{u}, \beta)$.

(vi) $F(B; Q)$ = probability assigned to $B \in \mathscr{B}_{\#}$ by a stochastic process with distribution Q.

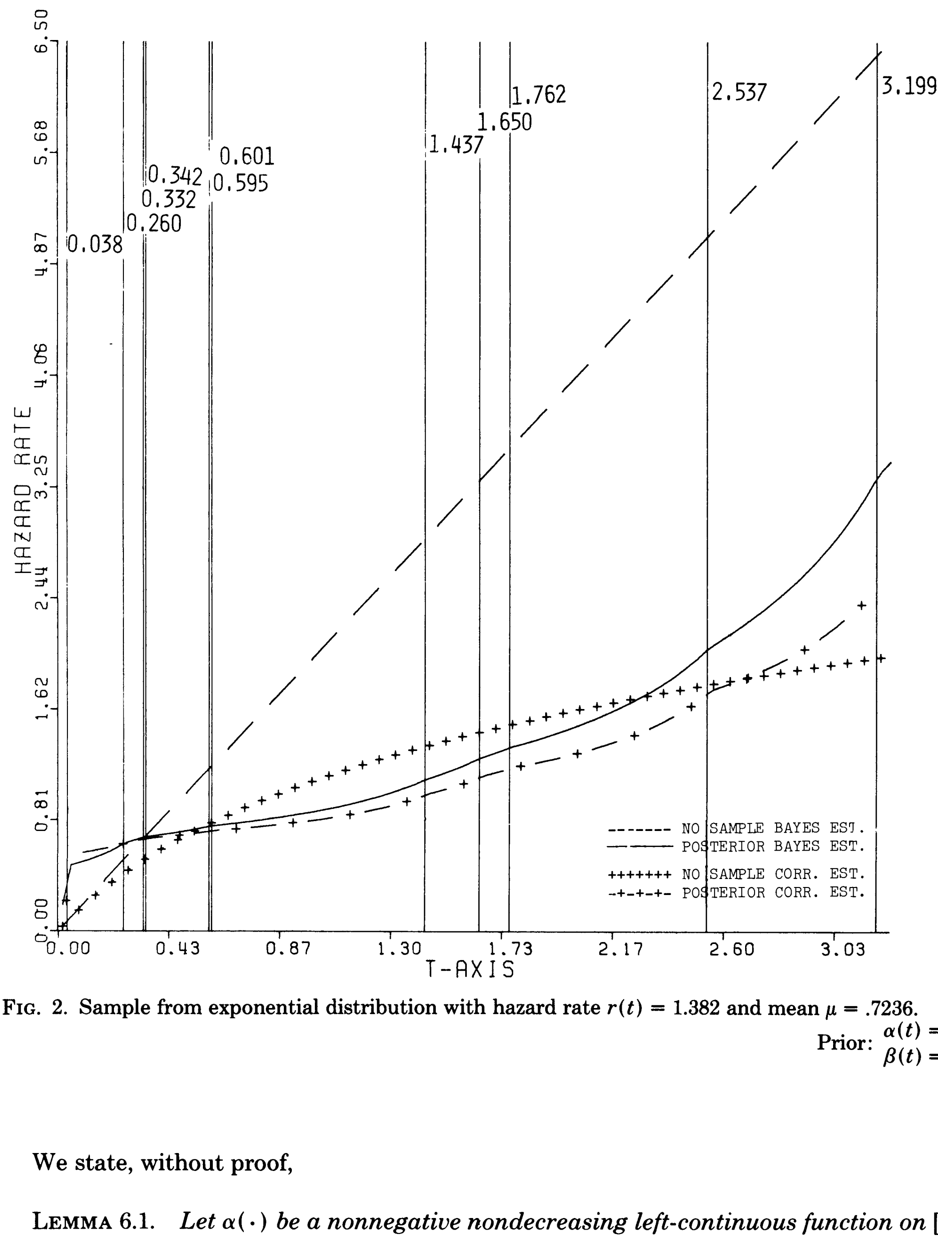

FIG. 2. Sample from exponential distribution with hazard rate $r(t) = 1.382$ and mean $\mu = .7236$.

Prior: $\alpha(t) = t$, $\beta(t) = 2$

We state, without proof,

LEMMA 6.1. *Let $\alpha(\cdot)$ be a nonnegative nondecreasing left-continuous function on* $[0, \infty]$ *with* $\alpha(0) = 0$. *For a sequence of partitions* $0 = t_0 < t_1 < \cdots < t_{k(n)}$ *whose norm goes to zero and upper end point goes to infinity, define* $\alpha_n(0) = 0$ *and* $\alpha_n(t) = \sum \alpha(t_i) I_{(t_{i-1}, t_i]}(t) + \alpha(t_{k(n)}) I_{(t_{k(n)}, \infty)}(t)$, $t \in (0, \infty)$. *Let* $B = \{r(\cdot) \in \mathscr{R}^{\#} : r(\tau_1) > y_1, r(\tau_2) - r(\tau_1) > y_2, \cdots, r(\tau_k) - r(\tau_{k-1}) > y_k\}$ *where k is any positive integer and* $\tau_1 < \cdots < \tau_k$, $y_1, \cdots, y_k$ *are nonnegative real numbers. Define* $r_n(t, \omega)$ *as in* (2.5), $A = r^{-1}(B)$ *and* $A_n = r_n^{-1}(B)$. *Then*

$$\int_\Omega I_{A_n}(\omega)\, dP(\omega) \to \int_\Omega I_A(\omega)\, dP(\omega), \quad \text{i.e., } F(B; \Gamma(\alpha_n, \beta)) \to F(B; \Gamma(\alpha, \beta)) \text{ and} \tag{6.1}$$

$$\lim_{n\to\infty} \int_{B_n(\mathbf{u},\beta)} \prod g(u_i; \Delta\alpha_i)\, du_i = F(B; \Gamma(\alpha, \beta)). \tag{6.2}$$

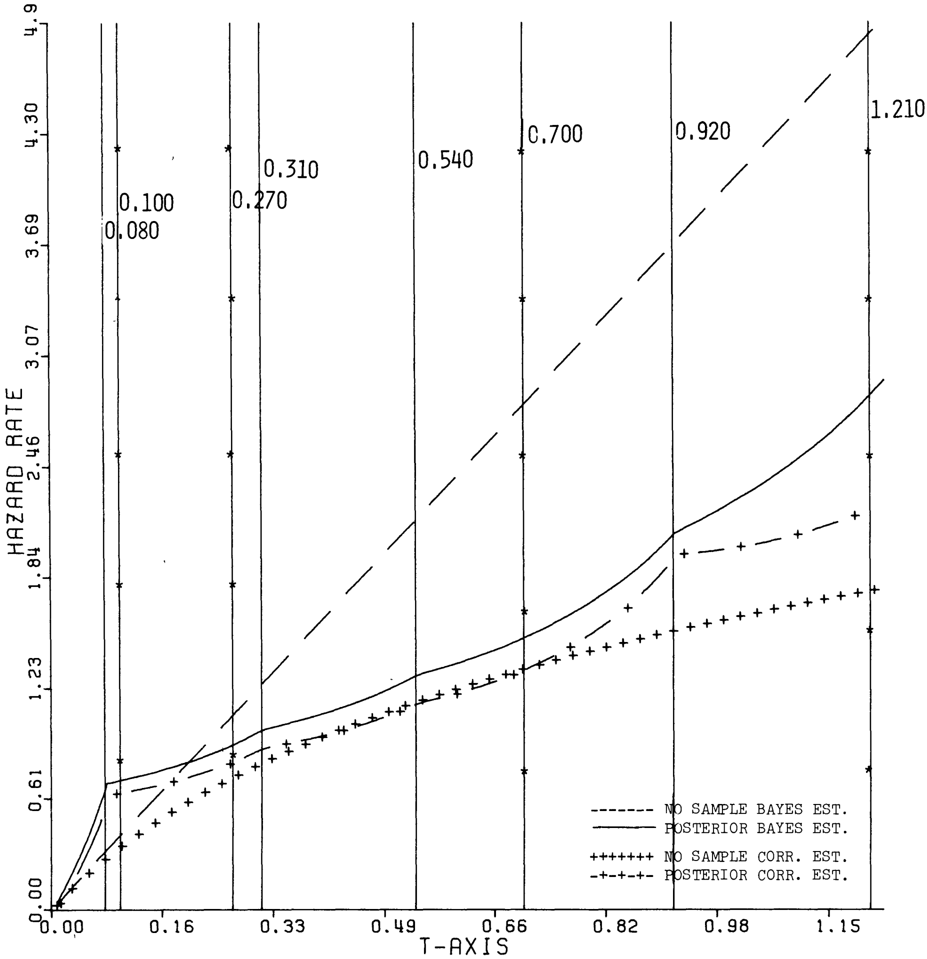

FIG. 3. Data from Kaplan-Meier paper. (Starred lines indicate censored data.)

Prior: $\alpha(t) = t$, $\beta(t) = 4$

PROOF OF THEOREM 3.2. First consider the case $m = 1$. It suffices to show that the posterior probability of sets of the form B defined in Lemma 6.1 equals that assigned by $\Gamma(\alpha(\cdot), \hat{\beta}(\cdot))$. Recall $r_n(t)$ from (2.5), and let A_n and A be defined as in Lemma 6.1. Then $r_n(t) \to r(t)$ a.s. and $I_{A_n} \to I_A$ a.s. Thus

$$P(r(\cdot) \in B \mid X \geq x)$$

$$= \int_A \exp\left[-\int_{[0,x)} r(t)\,dt\right] dP(\omega) \Big/ \int_\Omega \exp\left[-\int_{[0,x)} r(t)\,dt\right] dP(\omega)$$

$$= \lim_{n\to\infty} \int_\Omega \exp\left[-\int_{[0,x)} r_n(t)\,dt\right] I_{A_n}(\omega)\,dP(\omega) \Big/ \lim_{n\to\infty} \int_\Omega \exp\left[-\int_{[0,x)} r_n(t)\,dt\right] dP(\omega)$$

$$= \lim_{n\to\infty} \int_{B_n(\mathbf{u},\beta)} \exp\left[-\sum \beta_i (x-t_i)^+ u_i\right] \prod g(u_i; \Delta\alpha_i)\,du_i \Big/ \lim_{n\to\infty} \left[1 + \beta_i (x - t_i)^+\right]^{-\Delta\alpha_i}$$

$$= \lim_{n\to\infty} \int_{B_n(\mathbf{u},\beta)} \prod g(u_i; \Delta\alpha_i, 1 + \beta_i(x - t_i)^+)\, du,$$

$$= \lim_{n\to\infty} \int_{B_n(\mathbf{v},\hat\beta)} \prod g(v_i; \Delta\alpha_i)\, dv_i \qquad \text{where} \qquad \beta_i u_i = \hat\beta_i v_i$$

$$= F(B; \Gamma(\alpha, \hat\beta)) \qquad \text{by (6.2) of Lemma 6.1.}$$

This proves the theorem for $m = 1$. Using a parenthesized subscript to emphasize the dependence of $\hat\beta$ on the sample size, we have $\hat\beta_{(j)}(t)/[1 + \hat\beta_{(j)}(t)(x_{j+1} - t)^+] = \hat\beta_{(j+1)}(t)$. The theorem follows by induction.

PROOF OF THEOREM 3.3. First consider the case of $m = 1$. It suffices to show that for sets of the form B defined in Lemma 6.1

$$\int_{[x,\infty)} \phi(B, s) f(s)\, ds = P(r(\cdot) \in B, X \geq x) \tag{6.3}$$

where $f(x) = -d/dx \int_\Omega \exp[-\int_{[0,x)} r(s)\, ds]\, dP(\omega)$ is the marginal density of X and $\phi(B, x)$ denotes the family of distributions (for $m = 1$) given in (3.7). Let $r_n(t)$ be defined as in (2.5), $A_n = r_n^{-1}(\beta)$,

$$f_n(x) = \frac{-d}{dx} \int_\Omega \exp\left[-\int_{[0,x)} r_n(s)\, ds\right] dP(\omega),$$

and

$$\phi_n(B, x) = \frac{-d}{dx} \int_\Omega \exp\left[-\int_{[0,x)} r_n(s)\, ds\right] I_{A_n}(\omega)\, dP(\omega)/f_n(x).$$

Thus,

$$\int_{[x,\infty)} \phi_n(B, s) f_n(s)\, ds = \int_\Omega \exp\left[-\int_{[0,x)} r_n(s)\, ds\right] I_{A_n}(\omega)\, dP(\omega). \tag{6.4}$$

The right-hand side of (6.4) converges to $\int_\Omega \exp[-\int_{[0,x)} r(s)\, ds] I_A(\omega)\, dP(\omega) = P(r(\cdot) \in B, X \geq x)$. It can be shown that $f_n(x) \to f(x)$ and $\phi_n(B, x) f_n(x) \to \phi(B, x) f(x)$. Hence by a generalization of LDCT (see theorem in Royden (1968), page 89) the left-hand side of (6.4) converges to $\int_{[x,\infty)} \phi(B, s) f(s)\, ds$ upon observing that $0 \leq \phi_n(B, x) f_n(x) \leq f_n(x)$ and $\int_{[x,\infty)} f_n(s)\, ds \to \int_{[x,\infty)} f(s)\, ds$. This concludes the proof for $m = 1$.

For $m = 2$ a similar proof can be given by taking the posterior distribution after the first observation as the prior for the second. $\phi_n(B, x)$ and $f_n(x)$ are similarly defined except that $r(t, \omega)$ is distributed as a mixture of extended gamma processes. The detailed computations are more cumbersome and one needs a generalization of an unsymmetric Fubini theorem given by Cameron and Martin (1941) to interchange the order of certain integrals that are encountered. Using the LDCT and the result proved for $m = 1$, one arrives at the result for $m = 2$. The proof for arbitrary m follows by induction.

REFERENCES

[1] CAMERON, R. H. and MARTIN, W. T. (1941). An unsymmetric Fubini theorem. *Bull. Amer. Math. Soc., Ser.* 2. **47** 121-125.

[2] DOKSUM, K. A. (1974). Tailfree and neutral random probabilities and their posterior distributions. *Ann. Probability* **2** 183-201.

[3] FERGUSON, T. S. (1973). A Bayesian analysis of some nonparametric problems. *Ann. Statist.* **1** 209-230.

[4] FERGUSON, T. S. and PHADIA, E. G. (1979). Bayesian nonparametric estimation based on censored data. *Ann. Statist.* **7** 163–186.

[5] KAPLAN, E. L. and MEIER, PAUL (1958). Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.* **53** 457–481.

[6] LOÉVE, M. (1963). *Probability Theory.* Van Nostrand, New York and Princeton.

[7] MANN, N. R., SCHAFER, R. E., and SINGPURWALLA, N. D. (1974). *Methods for Statistical Analysis of Reliability and Life Data.* Wiley, New York.

[8] ROYDEN, H. L. (1968). *Real Analysis*, 2nd Ed. MacMillan, New York.

[9] SUSARLA, V. and VAN RYZIN, J. (1976). Nonparametric Bayesian estimation of survival curves from incomplete observations. *J. Amer. Statist. Assoc.* **71** 897–902.

DEPARTMENT OF STATISTICS
UNIVERSITY OF IOWA
IOWA CITY, IOWA 52240

DEPARTMENT OF MATHEMATICAL SCIENCE
UNIVERSITY OF NORTHERN ILLINOIS
DE KALB, ILLINOIS 60115