The Annals of Statistics
1979, Vol. 7, No. 1, 163–186

BAYESIAN NONPARAMETRIC ESTIMATION BASED ON CENSORED DATA

BY THOMAS S. FERGUSON[1] AND ESWAR G. PHADIA

University of California, Los Angeles and
The William Paterson College of New Jersey

Let $X_1, \cdots, X_n$ be a random sample from an unknown cdf F, let $y_1, \cdots, y_n$ be known real constants, and let $Z_i = \min(X_i, y_i)$, $i = 1, \cdots, n$. It is required to estimate F on the basis of the observations $Z_1, \cdots, Z_n$, when the loss is squared error. We find a Bayes estimate of F when the prior distribution of F is a process neutral to the right. This generalizes results of Susarla and Van Ryzin who use a Dirichlet process prior. Two types of censoring are introduced—the inclusive and exclusive types—and the class of maximum likelihood estimates which thus generalize the product limit estimate of Kaplan and Meier is exhibited. The modal estimate of F for a Dirichlet process prior is found and related to work of Ramsey. In closing, an example illustrating the techniques is given.

1. Introduction and summary. The problem of nonparametric estimation of a distribution function based on a sample partially censored on the right may be described as follows. We are given observations $Z_i = \min(y_i, X_i)$ $i = 1, \cdots, n$ where the $\{y_i\}$ are known numbers, and the $\{X_i\}$ are a sample from a population with unknown distribution function F. We are required to estimate F from the data $Z_1, \cdots, Z_n$.

This problem is encountered in many applied situations such as cancer research [10], and the study of survival data [7] and of baboons descending trees [11]. The basic paper of Kaplan and Meier [7] considers several nonparametric estimates of F and shows that one of them, the product limit estimate, is in fact a maximum likelihood estimate. These estimates have received considerable attention in recent years as is evidenced by the references in the paper of Breslow and Crowley [2] who treat large sample properties.

A recent paper of Susarla and Van Ryzin [10] contains a treatment of the problem from a Bayesian point of view. Using the Dirichlet process as a prior distribution for the unknown distribution function F, these authors obtain the mean of the posterior distribution of F given the data as an estimate of F. The main objective of the present paper is to extend their results to a more general class of prior distributions for F, namely, the processes neutral to the right introduced by Doksum [3]. The general theory of these processes and their use in the estimation of F given some right censored data is presented in Section 2. The formulas

Received September 1976; revised February 1978.
[1]Author's research was supported by NSF Grant No. MPS 72-04591
AMS 1970 *subject classifications.* Primary 62C10; Secondary 62G05, 60K99.
Key words and phrases. Bayesian nonparametric estimation, survival function, censored data, prior distribution, process neutral to the right, Dirichlet process, processes with independent increments, modal estimation.

expressing the posterior expectation of F given the data simplify for the subclass of homogeneous processes neutral to the right, and two such processes are considered in detail in Section 3, the gamma process, and a simple homogeneous process. In a recent announcement by Dykstra and Laud [4], a useful generalization of the gamma process is introduced for the problem of estimating the hazard function, $-\log(1 - F)$.

Since the Dirichlet process is a nonhomogeneous process neutral to the right, the estimates of Susarla and Van Ryzin may be derived from the general theory of Section 2. The details of this derivation are presented in Section 4.

In problems of this sort where the prior distribution of F may give positive probability to the event that F have a jump at a fixed point, it is useful to generalize earlier treatments of this problem to allow for two types of censoring information to be given to the statistician. In addition to information of the type $X_i \geqslant y_i$, as described in the introductory paragraph and called "inclusive censoring" in this paper, we also consider "exclusive censoring" where the information relayed to the statistician is of the type $X_i > y_i$. This latter type is the customary way of defining censoring and is the only one considered in Kaplan and Meier [7] and in Susarla and Van Ryzin [10].

There is a natural generalization of the product limit estimate to the two types of censoring. It arises in Sections 3 and 4 as a limit of our estimates as the prior sample size tends to zero. In Section 3, it is indicated that this estimate is still a maximum likelihood estimate when there are two types of censoring.

In an apparently unrelated paper by Ramsey [9] on Bayesian bioassay, problems are treated that are quite close to those considered here. Ramsey uses essentially the Dirichlet process as a prior distribution of the response curve, F, and obtains the modal estimate of F by maximizing the finite dimensional joint density (with respect to a suitable measure) of the posterior distribution. The bioassay problem may be considered as a censored sampling problem in which bioassay "successes" are observations censored on the left, and "failures" are observations censored on the right. Methods of this paper do not extend to problems in which observations may be either right or left censored. Yet the techniques Ramsey develops may be used when specialized to the case where all observations are failures to derive estimates for the problems considered here. We show in Section 6 that for the case when all observations are failures Ramsey's modal estimate has a simple closed form essentially given by the mean estimate of Susarla and Van Ryzin.

In the final section, the simple numerical example of Kaplan and Meier is reworked using different prior distributions in order to make a comparison of the estimates.

2. Bayesian estimation with priors neutral to the right. We review the use of processes neutral to the right as priors for nonparametric estimation problems. The posterior distribution of F given a censored sample is presented in Theorems 3 and 4, and the Bayes estimate is given in the corollary to Theorem 4.

2.1. *Basic properties of processes neutral to the right.* A random distribution function $F(t)$ on the real line, $\mathbb{R}$, is defined to be a stochastic process on $\mathbb{R}$ such that (a) $F(t)$ is nondecreasing a.s., (b) $F(t)$ is right-continuous a.s., (c) $\lim_{t\to\infty} F(t) = 0$ a.s., and (d) $\lim_{t\to\infty} F(t) = 1$ a.s. K. Doksum [3] has introduced a special class of random distribution functions, called processes neutral to the right, and has indicated the feasibility of their use in Bayesian nonparametric problems. Loosely speaking, a random distribution function $F(t)$ on the real line is neutral to the right if for every t_1 and t_2 with $t_1 < t_2$

$$(2.1) \qquad \frac{1 - F(t_2)}{1 - F(t_1)} \quad \text{is independent of} \quad \{F(t) : t \leqslant t_1\};$$

that is, if the proportion of mass $F(t)$ assigns to the subinterval (t_2, ∞) of the interval (t_1, ∞) is independent of what $F(t)$ does to the left of t_1.

If F is neutral to the right, then the process

$$(2.2) \qquad Y_t = -\log(1 - F(t))$$

is a nondecreasing process with independent increments. For the purposes of this paper, it is simpler to define a process neutral to the right through this property of the related process Y_t.

DEFINITION 1. A process $F(t)$ is said to be a random distribution function neutral to the right if it can be written in the form

$$(2.3) \qquad F(t) = 1 - e^{-Y_t}$$

where Y_t is a process with independent increments such that (a) Y_t is nondecreasing a.s., (b) Y_t is right continuous a.s., (c) $\lim_{t\to-\infty} Y_t = 0$ a.s., and (d) $\lim_{t\to+\infty} Y_t = \infty$ a.s.

We allow $Y_t = +\infty$ with positive probability for finite t. The usual rules (e.g. $\infty + c = \infty$ for finite c) apply.

Thus, in studying processes neutral to the right we may use the theory of Lévy [8] on a.s. nondecreasing processes with independent increments. Such a process Y_t, described in Definition 1, has at most countably many fixed points of discontinuity, call them $t_1, t_2, \cdots$ in some order. Let $S_1, S_2, \cdots$ represent the random heights of the jumps in Y_t at $t_1, t_2, \cdots$ respectively. Then, $S_1, S_2, \cdots$ are independent random variables. The $\{S_j\}$ are also independent of the rest of the process with the jumps removed.

$$(2.4) \qquad Z_t = Y_t - \Sigma_j S_j I_{[t_j, \infty)}(t),$$

where I_B represents the indicator function of the set B. The process Z_t has independent increments and is nondecreasing a.s. with $\lim_{t\to-\infty} Z_t = 0$ a.s. In addition, Z_t has no fixed points of discontinuity and so must have an infinitely divisible distribution with Lévy formula for the log of the moment generating function

$$(2.5) \qquad \log \mathcal{E} e^{-\theta Z_t} = -\theta b(t) + \textstyle\int_0^\infty (e^{-\theta z} - 1) dN_t(z)$$

where b is nondecreasing and continuous with $b(t) \to 0$ as $t \to -\infty$, and where N_t is a continuous Lévy measure; that is,

(i) for every Borel set B, $N_t(B)$ is continuous and nondecreasing,
(ii) for every real t, $N_t(\cdot)$ is a measure on the Borel subsets of $(0, \infty)$,
(iii) $\int_0^\infty z(1+z)^{-1} dN_t(z) \to 0 \quad$ as $\quad t \to -\infty$.

The main result of Doksum [3] for processes neutral to the right is that if $X_1, \cdots, X_n$ is a sample from F and if F is neutral to the right, then the posterior distribution of F given the sample is neutral to the right also. In the following theorem we describe the posterior distribution for a sample of size one. The general case of arbitrary sample size may then be treated by a repeated application of the theorem.

In the following theorems, Y_t is always related to F_t by (2.2) and (2.3). The process Y_t^- defined by

$$Y_t^- = \lim_{s \nearrow t} Y_s \tag{2.6}$$

also has independent increments, but is left-continuous a.s. The increment $Y_t - Y_t^-$ represents the jump at t. It is positive with positive probability if and only if t is a fixed point of discontinuity of the process.

THEOREM 1 (Doksum). *Let F be a random distribution function neutral to the right, and let X be a sample of size one from F. Then, the posterior distribution of F given $X = x$ is neutral to the right. The posterior distribution of an increment in Y_t to the right of x is the same as the prior distribution. The posterior distribution of an increment in Y_t to the left of x may be found by multiplying the prior density of the increment by e^{-y} and renormalizing. Thus, if an increment $Y_t - Y_s$ with $s < t < x$ has prior density $dG(y)$, the posterior density given $X = x$ is*

$$e^{-y} dG(y) / \int_0^\infty e^{-y} dG(y). \tag{2.7}$$

To complete the description of the posterior distribution, we need to know what happens to the increment $S = Y_x - Y_x^-$, the jump at x. Generally, Y_t will have a fixed point of discontinuity at x in the posterior whether or not there was one in the prior. However, there is no simple form for the posterior distribution, call it $H_x(s)$, of the jump there. A complete description of $H_x(s)$ may be found in Ferguson [6], page 624. In some special cases H_x may easily be evaluated. In the case where the process Y_t is homogeneous (to be treated in Section 3), H_x is easily written down in terms of the Lévy measure, N_t. The other simple case occurs when x is a prior fixed point of discontinuity.

THEOREM 2. *Under the hypotheses of Theorem 1, if x is a prior fixed point of discontinuity of F, then the posterior density of the jump in Y_t at x given $X = x$ may be found by multiplying the prior density of the jump by $(1 - e^{-s})$ and renormalizing. Thus,*

$$dH_x(s) = (1 - e^{-s}) dG_x(s) / \int_0^\infty (1 - e^{-s}) dG_x(s) \tag{2.8}$$

where G_x is the prior distribution of the jump at x, and H_x is the posterior distribution of the jump at x given $X = x$.

2.2. *Posterior distribution given a censored sample.* To complete the description of the posterior distribution of F for application to the censored sampling problem, we present a theorem for a censored sample of size one, and consider two cases, when the censoring information is exclusive, $X > x$, and when it is inclusive, $X \geqslant x$. The posterior distribution of F given a censored sample turns out to be simpler than that given an uncensored sample. In fact, the posterior distribution of F is the same as that of Theorem 1 except that the jump at the point x does not have to be treated differently, so there is no need for an analogue of Theorem 2. The increment at x is treated as if it were to the left of x for exclusive censoring, and to the right of x for inclusive censoring. The only difference between the two types of censoring occurs when x is a prior fixed point of discontinuity.

THEOREM 3. *Let F be a random distribution function neutral to the right, let X be a sample of size one from F, and let x be a real number.*

(a) *The posterior distribution of F given $X > x$ is neutral to the right; the posterior distribution of an increment to the right of x is the same as the prior distribution; the posterior distribution of an increment to the left of or including x is found by multiplying the prior density by e^{-y} and renormalizing as in* (2.7).

(b) *The posterior distribution of F given $X \geqslant x$ is neutral to the right; the posterior distribution of an increment to the right of or including x is the same as the prior distribution; the posterior distribution of an increment to the left of x is found by multiplying the prior density by e^{-y} and renormalizing as in* (2.7).

PROOF. Let $t_1 < t_2 < \cdots < t_n$ be arbitrary real numbers one of which is x, say $t_j = x$. Let $W_i = Y_{t_{i+1}} - Y_{t_i}$ for $i = 0, \cdots, j-2$ (where $t_0 = -\infty$ so that $Y_{t_0} \equiv 0$), $W_{j-1} = Y_{t_j}^- - Y_{t_{j-1}}$, $W_j = Y_{t_j} - Y_{t_j}^-$, and for $i = j+1, \cdots, n$, $W_i = Y_{t_i} - Y_{t_{i-1}}$. Under the prior distribution, $W_0, W_1, \cdots, W_n$ are independent random variables with joint density, say

$$f_{W_0, \cdots, W_n}(w_0, \cdots, w_n) = \Pi_{i=0}^n f_{W_i}(w_i) \tag{2.9}$$

with respect to some convenient product measure. Given F, the probability that $X > x$ is

$$1 - F(x) = e^{-Y_x} = e^{-\Sigma_0^j W_i}. \tag{2.10}$$

Since the posterior density of $W_0, \cdots, W_n$ given $X > x$ is proportional to the product of (2.9) and (2.10),

$$f_{W_0, \cdots, W_n}(w_0, \cdots, w_n | X > x) \alpha \left(\Pi_{i=0}^j e^{-w_i} f_{W_i}(w_i)\right)\left(\Pi_{i=j+1}^n f_{W_i}(w_i)\right), \tag{2.11}$$

we see that the W_i are independent for this posterior distribution as well. This shows that the posterior distribution of F is neutral to the right. Furthermore, the distribution of increments to the right of x are unchanged, while the distribution of increments to the left of and including x are changed by multiplying the prior densities by e^{-w} and renormalizing. Similarly, given F, the probability that $X \geqslant x$ is

$$1 - F^{-}(x) = e^{-Y_x^-} = e^{-\Sigma_0^{j-1} W_i}, \tag{2.12}$$

and the posterior distribution of $W_0, \cdots, W_n$, given $X \geqslant x$, is

$$f_{W_0, \ldots, W_n}(w_0, \cdots, w_n | X \geqslant x) \alpha \big(\Pi_{i=0}^{j-1} e^{-w_i} f_{W_i}(w_i)\big)\big(\Pi_{i=j}^{n} f_{W_i}(w_i)\big). \tag{2.13}$$

Part (b) of the theorem clearly follows from this.

2.3. *The general case.* We now combine Theorems 1, 2, and 3 for a sample of size n. We consider later the problem of Bayesian estimation of the distribution function under weighted squared error loss. The Bayes estimate is then the expected value of the distribution function, which, when the prior is neutral to the right, reduces to the problem of evaluating the moment generating function of Y_t at the point 1.

$$\mathcal{E}F(t) = 1 - \mathcal{E}e^{-Y_t} = 1 - M_t(1) \tag{2.14}$$

where the moment generating function (MGF) is defined as

$$M_t(\theta) = \mathcal{E}e^{-\theta Y_t}. \tag{2.15}$$

(Note the minus sign!) Therefore, we state the results in Theorem 4 below in terms of the posterior MGF.

We assume the observational data has three forms, m_1 "real" observations $X_1 = x_1, \cdots, X_{m_1} = x_{m_1}$, m_2 "exclusive" censorings $X_{m_1+1} > x_{m_1+1}, \cdots, X_{m_1+m_2} > x_{m_1+m_2}$, and m_3 "inclusive" censorings $X_{m_1+m_2+1} \geqslant x_{m_1+m_2+1}, \cdots, X_{m_1+m_2+m_3} \geqslant x_{m_1+m_2+m_3}$ where $m_1 + m_2 + m_3 = n$, the sample size. To compress the data even further, we introduce the following notation.

Let $u_1, \cdots, u_k$ be the distinct values among $x_1, \cdots, x_n$, ordered so that $u_1 < u_2 < \cdots < u_k$. Let $\delta_1, \cdots, \delta_k$ denote the number of "real" observations at $u_1, \cdots, u_k$ respectively, let $\lambda_1, \cdots, \lambda_k$ denote the number of exclusive censorings at $u_1, \cdots, u_k$ respectively, and let $\mu_1, \cdots, \mu_k$ denote the number of inclusive censorings at $u_1, \cdots, u_k$ respectively, so that $\Sigma_1^k \delta_i = m_1$, $\Sigma_1^k \lambda_i = m_2$, and $\Sigma_1^k \mu_i = m_3$. We shall refer to the vectors $\mathbf{u}$, $\boldsymbol{\delta}$, $\boldsymbol{\lambda}$, $\boldsymbol{\mu}$ as the *data*.

In the following theorem,

$$h_j = \Sigma_{i=j+1}^{k}(\delta_i + \lambda_i + \mu_i) \tag{2.16}$$

denotes the number of the x_i greater than u_j, and $j(t)$ denotes the number of u_i less than or equal to t. Also, we use $M_t^-(\theta)$ to denote the MGF of Y_t^-,

$$M_t^-(\theta) = \lim_{s \nearrow t} M_s(\theta). \tag{2.17}$$

$G_u(s)$ denotes the prior distribution of the jump in Y_t at u, and $H_u(s)$ denotes the posterior distribution of the jump in Y_t at u, given $X = u$ for one observation.

THEOREM 4. *Let F be a random distribution function neutral to the right, and let $X_1, \cdots, X_n$ be a sample of size n from F, yielding data $\mathbf{u}$, $\boldsymbol{\delta}$, $\boldsymbol{\lambda}$, $\boldsymbol{\mu}$. Then the posterior distribution of F given the data is neutral to the right, and Y_t has posterior* MGF

$$(2.18)\qquad M_t(\theta\,|\,data) = \frac{M_t(\theta + h_{j(t)})}{M_t(h_{j(t)})} \cdot \Pi_{i=1}^{j(t)} \left[\frac{M_{u_i}^-(\theta + h_{i-1})}{M_{u_i}^-(h_{i-1})} \cdot \frac{C_{u_i}(\theta + h_i + \lambda_i, \delta_i)}{C_{u_i}(h_i + \lambda_i, \delta_i)} \cdot \frac{M_{u_i}(h_i)}{M_{u_i}(\theta + h_i)} \right],$$

where, if u is a prior fixed point of discontinuity of Y_t,

$$(2.19)\qquad c_u(\alpha, \beta) = \int_0^\infty e^{-\alpha s}(1 - e^{-s})^\beta dG_u(s),$$

while, if u is not a prior fixed point of discontinuity of Y_t,

$$(2.20)\qquad \begin{aligned} C_u(\alpha, \beta) &= \int_0^\infty e^{-\alpha s}(1 - e^{-s})^{\beta-1} dH_u(s) \qquad &\text{if } \beta \geqslant 1 \\ &= 1 &\text{if } \beta = 0. \end{aligned}$$

PROOF. That the posterior distribution of F is neutral to the right given the data follows immediately from Theorems 1 and 3. Consider therefore a given value of t. The posterior MGF of Y_t is the product of the posterior MGF's of Y_{u_1}, $Y_{u_2} - Y_{u_1}, \cdots, Y_{u_{j(t)}} - Y_{u_{j(t)-1}}$, $Y_t - Y_{u_{j(t)}}$. The posterior MGF of $Y_{u_i} - Y_{u_{i-1}}$ is the product of the posterior MGF of $Y_{u_i}^- - Y_{u_{i-1}}$ and the posterior MGF of the jump at u_i, $Y_{u_i} - Y_{u_i}^-$. The increment $Y_{u_i}^- - Y_{u_{i-1}}$ has h_{i-1} observations to the right of it, so that the posterior distribution is obtained from the prior by multiping the prior density by $e^{-yh_{i-1}}$ and renormalizing. In terms of the prior MGF of $Y_{u_i}^- - Y_{u_{i-1}}$, which is $M_{u_i}^-(\theta)/M_{u_{i-1}}(\theta)$, this gives

$$(2.21)\qquad \frac{M_{u_{i-1}}(h_{i-1})}{M_{u_{i-1}}(\theta + h_{i-1})} \cdot \frac{M_{u_i}^-(\theta + h_{i-1})}{M_{u_i}^-(h_{i-1})}$$

as the posterior MGF. The point u_i has h_i observations to the right of it, λ_i exclusive censored observations, and δ_i real observations there. If there is a prior fixed point of discontinuity at u_i, the posterior distribution of the jump in Y_t at u_i is obtained by multiplying the prior density by $e^{-(h_i+\lambda_i)s}(1 - e^{-s})^{\delta_i}$ and renormalizing. In this case, the posterior MGF of the jump in Y_t at u_i is therefore

$$(2.22)\qquad \begin{aligned} &\int_0^\infty e^{-(\theta+h_i+\lambda_i)s}(1 - e^{-s})^{\delta_i} dG_{u_i}(s) \Big/ \int_0^\infty e^{-(h_i+\lambda_i)s}(1 - e^{-s})^{\delta_i} dG_{u_i}(s) \\ &\qquad = C_{u_i}(\theta + h_i + \lambda_i, \delta_i)/C_{u_i}(h_i + \lambda_i, \delta_i). \end{aligned}$$

On the other hand, if u_i is not a prior fixed point of discontinuity of Y_t, then it takes one "real" observation at u_i to generate a fixed point of discontinuity there, having distribution function $H_{u_i}(s)$. The remaining $\delta_i - 1$ "real" observations at u_i may be treated as in the previous case to give a posterior MGF of the jump in Y_t at

u_i as

$$(2.23)\qquad \int_0^\infty e^{-(\theta+h_i+\lambda_i)s}(1-e^{-s})^{\delta_i-1}dH_{u_i}(s)\Big/\int_0^\infty e^{-(h_i+\lambda_i)s}(1-e^{-s})^{\delta_i-1}dH_{u_i}(s)$$
$$= C_{u_i}(\theta+h_i+\lambda_i,\delta_i)/C_{u_i}(h_i+\lambda_i,\delta_i).$$

If $\delta_i = 0$, the posterior jump in Y_t at u_i is zero, so the MGF is identically one, which is also represented by $C_{u_i}(\theta + h_i + \lambda_i, \delta_i)/C_{u_i}(h_i + \lambda_i, \delta_i)$. Finally the MGF of $Y_t - Y_{u_{j(t)}}$ is, analogous to (2.21)

$$(2.24)\qquad \frac{M_{u_{j(t)}}(h_{j(t)})}{M_{u_{j(t)}}(\theta + h_j(t))}\cdot\frac{M_t(\theta + h_{j(t)})}{M_t(h_{j(t)})}.$$

Combining the terms (2.21), (2.22), (2.23), and (2.24) into a single product, and using $M_{-\infty} \equiv 1$, yields the product stated in the theorem.

We are ready now to exhibit the posterior expected value of $F(t)$ given the data. For this it is convenient to reduce the form of the expected value further with the introduction of the following notation. Let

$$(2.25)\qquad R_t(h) = M_t(h+1)/M_t(h)$$

and let

$$(2.26)\qquad r_u(\alpha, \beta) = C_u(\alpha + 1, \beta)/C_u(\alpha, \beta).$$

Finally, we let $S(t)$ denote the survival function

$$(2.27)\qquad S(t) = 1 - F(t)$$

and give the formula for the posterior expectation of $S(t)$ from which the posterior expectation of $F(t)$ may be computed if needed.

COROLLARY. *Under the assumptions of Theorem* 4,

$$(2.28)\qquad \mathcal{E}(S(t)|data) = M_t(1|data)$$
$$= R_t(h_{j(t)})\prod_{i=1}^{j(t)}\frac{R_{u_i}^-(h_{i-1})}{R_{u_i}(h_i)}r_{u_i}(h_i + \lambda_i, \delta_i).$$

The proof is by direct substitution.

3. The homogeneous process neutral to the right. In the applications of the results of the previous section to specific processes neutral to the right, difficulties are encountered in evaluating $H_u(s)$, the posterior distribution of a jump in Y_t at a point u at which a single observation fell. This function is needed in order to compute $C_u(\alpha, \beta)$ of (2.20). In one rather general case, $H_u(s)$ is easy to evaluate—the case where Y_t is a homogeneous process with independent increments.

DEFINITION 2. A random distribution function F neutral to the right is said to be homogeneous if the independent increment process $Y_t = -\log(1 - F(t))$ has Lévy function independent of t; that is, if the MGF of Y_t has the form

$$(3.1)\qquad M_t(\theta) = e^{\gamma(t)\int_0^\infty(e^{-\theta z}-1)dN(z)},$$

where $\gamma(t)$ is continuous nondecreasing, $\lim_{t\to-\infty}\gamma(t) = 0$, $\lim_{t\to+\infty}\gamma(t) = +\infty$, and where N is any measure on $(0, \infty)$ such that

$$\int_0^\infty z(1 + z)^{-1}dN(z) < \infty. \tag{3.2}$$

Thus, for the homogeneous process, there are no (prior) fixed points of discontinuity, there is no nonrandom part ($b \equiv 0$), and the Lévy measure has the simple form $N_t(\cdot) = \gamma(t)N(\cdot)$.

The general form of $H_u(s)$ for homogeneous processes neutral to the right is given in the following theorem without proof. It is stated explicitly in Ferguson [6] and may be derived from the computations of Example 4.1 of Doksum [3]. It is a striking fact that $H_u(s)$ is independent of u and γ and depends only on N.

THEOREM 5. *Let F be a random distribution function neutral to the right with* MGF *of Y_t of the form* (3.1) *with $\gamma(t)$ continuous. Let X be a sample of size one from F. The posterior distribution, given $X = x$, of the jump in Y_t at x is independent of x and γ, and is given by*

$$H_x(s) = \int_0^s(1 - e^{-z})dN(z)\Big/ \int_0^\infty(1 - e^{-z})dN(z). \tag{3.3}$$

That the denominator of (3.3) exists follows from (3.2). From this theorem we see that the posterior expected value of the survival function, $S(t) = 1 - F(t)$, can be written very simply. We shall evaluate this expectation in the next two sections for two particular homogeneous processes. For this we introduce the function

$$\begin{aligned} \varphi(\alpha, \beta, N) &= \int_0^\infty e^{-\alpha z}(1 - e^{-z})^\beta dN(z) \qquad &&\text{for } \beta \geqslant 1 \\ &= 1 &&\text{for } \beta = 0. \end{aligned} \tag{3.4}$$

All the functions used in (2.28) to form the posterior expected value, $\mathcal{E}(S(t)|\text{data})$, can be written in terms of φ as follows. From Theorem 5 and (2.20)

$$\begin{aligned} C_u(\alpha, \beta) &= \varphi(\alpha, \beta, N)/\varphi(0, 1, N) \qquad &&\text{for } \beta \geqslant 1 \\ &= 1 &&\text{for } \beta = 0; \end{aligned} \tag{3.5}$$

from (2.25) and (3.1),

$$R_t(h) = e^{-\gamma(t)\varphi(h, 1, N)}; \tag{3.6}$$

and from (2.26) and (3.5)

$$r_u(\alpha, \beta) = \varphi(\alpha + 1, \beta, N)/\varphi(\alpha, \beta, N). \tag{3.7}$$

3.1. *The gamma process.* When the independent increments of the process Y_t have gamma distributions, we say Y_t is a gamma process. In this case, the MGF of Y_t may be written in any of the forms

$$\begin{aligned} M_t(\theta) &= \frac{\tau^{\gamma(t)}}{\Gamma(\gamma(t))}\int_0^\infty e^{-\theta y}e^{-\tau y}y^{\gamma(t)-1}dy \\ &= \left(\frac{\tau}{\tau + \theta}\right)^{\gamma(t)} \\ &= e^{\gamma(t)\int_0^\infty(e^{-\theta z} - 1)e^{-\tau z}z^{-1}dz}. \end{aligned} \tag{3.8}$$

Thus, Y_t has a gamma distribution with shape parameter $\gamma(t)$ and reciprocal scale or intensity parameter τ, independent of t. We assume $\gamma(t)$ is continuous, though the formulas could be worked out with a little extra complexity in the discontinuous case. The Lévy measure has the form

$$dN(z) = e^{-\tau z} z^{-1} dz \tag{3.9}$$

so that the function $\varphi(\alpha, \beta, N)$ may be written

$$\varphi(\alpha, \beta, N) = \int_0^\infty e^{-\alpha z}(1 - e^{-z})^\beta e^{-\tau z} z^{-1} dz. \tag{3.10}$$

Here the function N is determined by the parameter τ, which may be combined with the parameter α. Thus for the gamma case we may write

$$\begin{aligned} \varphi_G(\alpha, \beta) &= \int_0^\infty e^{-\alpha z}(1 - e^{-z})^\beta z^{-1} dz \qquad \text{for} \quad \beta \geqslant 1 \\ &= 1 \qquad\qquad\qquad\qquad\qquad\qquad\quad \text{for} \quad \beta = 0. \end{aligned} \tag{3.11}$$

so that $\varphi(\alpha, \beta, N) = \varphi_G(\alpha + \tau, \beta)$. From (3.8) with $\theta = 1$, we see

$$\varphi_G(\alpha, 1) = \log\left(\frac{\alpha + 1}{\alpha}\right). \tag{3.12}$$

Furthermore, $\varphi_G(\alpha, \beta)$ may be found in terms of $\varphi_G(\alpha, 1)$ for integral $\beta > 1$ by expanding $(1 - e^{-z})^{\beta - 1}$ in the integrand of (3.11) in a binomial series,

$$\varphi_G(\alpha, \beta) = \Sigma_{i=0}^{\beta-1}\binom{\beta - 1}{i}(-1)^i \log\left(\frac{\alpha + i + 1}{\alpha + i}\right). \tag{3.13}$$

Finally, $R_t(h)$ of (3.6) can be simplified with the use of (3.12) as

$$R_t(h) = \left(\frac{h + \tau}{h + \tau + 1}\right)^{\gamma(t)}. \tag{3.14}$$

THEOREM 6. *In the sampling scheme of subsection* 2.3, *if the prior distribution of* Y_t *is the homogeneous gamma process with shape function* $\gamma(t)$, *and intensity* τ, *the posterior expectation of* $S(t)$ *is*

$$\begin{aligned} \mathcal{E}(S(t)|\text{data}) = &\left(\frac{h_{j(t)} + \tau}{h_{j(t)} + \tau + 1}\right)^{\gamma(t)} \\ &\cdot \Pi_{i=1}^{j(t)}\left[\left(\frac{(h_{i-1} + \tau)(h_i + \tau + 1)}{(h_{i-1} + \tau + 1)(h_i + \tau)}\right)^{\gamma(u_i)} \frac{\varphi_G(h_i + \lambda_i + \tau + 1, \delta_i)}{\varphi_G(h_i + \lambda_i + \tau, \delta_i)}\right]. \end{aligned} \tag{3.15}$$

The proof is by direct substitution.

REMARK 1. One of the useful features of the Dirichlet process $\mathcal{D}(\alpha)$ as a prior distribution is that there is a fairly reasonable interpretation of $\alpha(\mathbb{R})$ as the "prior sample size" and of $F_0(t) = \alpha(t)/\alpha(\mathbb{R})$ as the prior guess at the shape of $F(t)$. We would like to be able to find a similar interpretation for the processes neutral to the right. For the gamma process, the prior guess at the shape of $S(t)$ is

$$\mathcal{E}S(t) = M_t(1) = \left(\frac{\tau}{\tau + 1}\right)^{\gamma(t)}.$$

If our prior guess at the shape of $S(t)$ is given by $S_0(t)$, then we would choose γ and τ to satisfy

$$\left(\frac{\tau}{\tau+1}\right)^{\gamma(t)} = S_0(t) \tag{3.16}$$

for all t, which would determine $\gamma(t)$ for fixed τ to be

$$\gamma(t) = \log(S_0(t))/\log(\tau/(\tau+1)). \tag{3.17}$$

Hopefully, the remaining parameter τ reflects our prior "strength of belief" in some sense. This it seems to do, but not with the simple interpretation as "prior sample size" as we have with the Dirichlet process.

To see what effect the choice of τ has, consider a single uncensored observation and a gamma process prior with $\gamma(t)$ satisfying (3.17). From (3.15) we find

$$\begin{aligned} \mathcal{E}(S(t)|X=x) &= S_0(t)^{l(\tau)} && \text{for} \quad t < x \\ &= S_0(t)S_0(x)^{l(\tau)-1}l(\tau) && \text{for} \quad t \geqslant x \end{aligned}$$

where

$$l(\tau) = \log((\tau+2)/(\tau+1))/\log((\tau+1)/\tau).$$

The size of the jump at x is therefore

$$S_0(x)^{l(\tau)}(1 - l(\tau)). \tag{3.18}$$

The function l is monotone in τ with $l(\tau) \to 0$ as $\tau \to 0$, and $l(\tau) \to 1$ as $\tau \to \infty$; therefore, the size of the jump at x is monotone in τ and tends to one as $\tau \to 0$, and to zero as $\tau \to \infty$. For small values of τ, one is less sure of the prior shape and more willing to change the estimate of shape on the basis of a sample, than for large values of τ. So τ does measure prior "strength of belief" in some sense.

Unlike the Dirichlet process prior, for which the posterior expected size of the jump at an observation does not depend on where the observation occurs, the size of the jump at x depends on x through $S_0(x)$ as is seen in (3.18). This makes it difficult to determine how to relate τ to "prior sample size." One feature of the posterior jump that is independent of the position of the observation x is the proportion of the mass in the interval $[x, \infty)$ that goes into the jump at x, namely $1 - l(\tau)$. For survival type data, it seems preferable to have behavior of this sort.

REMARK 2. Another attractive property of the Dirichlet process prior is that as the prior sample size, $\alpha(\mathbb{R})$, tends to zero, the posterior estimate of $F(t)$ tends to the maximum likelihood estimate (MLE) no matter what the prior guess at the shape may be. This is well known when there are no censored observations and the MLE is the sample distribution function. We may hope that the corresponding property would hold for the gamma process—that as the "strength of belief" parameter τ tends to zero, the posterior expectation of $F(t)$ would converge to the MLE no matter what value is chosen for $F_0(t)$, provided it is fixed. After all, no strength of belief in one prior distribution should be the same as no strength of belief in another.

Unfortunately, this is not the case even when there is no censoring. As an example, let us suppose that we have n uncensored observations $X_1, \cdots, X_n$ taking on the distinct values $u_1 < u_2 < \cdots < u_n$ in some order. Then in (3.15) we have $\lambda_i = 0$, $\delta_i = 1$ and $h_i = n - i$. Now if $S_0(t) = 1 - F_0(t)$ is held fixed and positive for all t, if γ is chosen according to (3.17), and if τ tends to zero, we find that all terms not involving φ_G tend to one, so that

$$\text{(3.19)} \qquad \mathcal{E}(1 - F(t)|\text{data}) \to \prod_{i=1}^{j(t)} \frac{\varphi_G(n - i + 1, 1)}{\varphi_G(n - i, 1)}$$

$$= \frac{\varphi_G(n, 1)}{\varphi_G(n - j(t), 1)} = \frac{\log((n + 1)/n)}{\log((n - j(t) + 1)/(n - j(t)))}.$$

This is not the sample distribution function, but at least it is independent of F_0. Furthermore, the size of the jump at an observation depends only on the rank order of the observation and not otherwise on its value. For example, if there are two observations, $x_1 < x_2$, the larger observation is given weight $\log(\frac{3}{2})/\log 2 = .5850$, and the smaller observation is given weight .4150, so (3.19) reduces to the estimate

$$\text{(3.20)} \qquad \hat{F}(t) = .4150 I_{[x_1, \infty)}(t) + .5850 I_{[x_2, \infty)}(t).$$

The weights for a few other values of n are given in the accompanying table. It is seen that the weights given to the lowest $n - 1$ observations are approximately equal and that the largest observation gets weight between 40.9% to 44.3% greater than the others.

3.2. *A simple homogeneous process.* One is led to search for homogeneous processes that behave like the Dirichlet in regard to Remark 2, for which the posterior expectation of F converges to the sample distribution function as some "strength of belief" parameter tends to zero. Of the various possibilities, the following seems to be the simplest.

Let Y_t be the homogeneous process with MGF of the form

$$\text{(3.21)} \qquad M_t(\theta) = e^{\gamma(t)\int_0^\infty (e^{-\theta z} - 1)e^{-\tau z}(1 - e^{-z})^{-1} dz}$$

TABLE 1

Table of jumps in (3.19) *at the observations for various sample sizes n*

n										
2	0.4150	0.5850								
3	0.2905	0.2945	0.4150							
4	0.2243	0.2253	0.2284	0.3219						
5	0.1829	0.1833	0.1841	0.1866	0.2630					
6	0.1545	0.1547	0.1550	0.1557	0.1578	0.2224				
7	0.1338	0.1338	0.1340	0.1342	0.1348	0.1367	0.1926			
8	0.1179	0.1180	0.1181	0.1182	0.1184	0.1189	0.1206	0.1699		
9	0.1055	0.1055	0.1055	0.1056	0.1057	0.1059	0.1064	0.1078	0.1520	
10	0.0954	0.0954	0.0954	0.0955	0.0955	0.0956	0.0958	0.0962	0.0976	0.1375

where γ is continuous nondecreasing, and $\tau > 0$ is a parameter whose value, hopefully, will reflect "strength of belief." Since τ is no longer the reciprocal of the scale parameter for Y_t, the interpretation of τ is more nebulous. The Lévy measure has the form

$$dN(z) = e^{-\tau z}(1 - e^{-z})^{-1}dz \tag{3.22}$$

which is simply the Lévy measure for the gamma process with z^{-1} replaced by $(1 - e^{-z})^{-1}$.

From (3.4) we have for $\beta \geqslant 1$

$$\begin{aligned} \varphi(\alpha, \beta, N) &= \textstyle\int_0^\infty e^{-(\alpha+\tau)z}(1 - e^{-z})^{\beta-1}dz \\ &= \Gamma(\alpha + \tau)\Gamma(\beta)/\Gamma(\alpha + \beta + \tau) \end{aligned} \tag{3.23}$$

a formula for the beta function. In particular, for $\beta = 1$

$$\varphi(\alpha, 1, N) = \Gamma(\alpha + \tau)/\Gamma(\alpha + \tau + 1) = 1/(\alpha + \tau) \tag{3.24}$$

so that from (3.6)

$$R_t(h) = e^{-\gamma(t)/(h+\tau)}. \tag{3.25}$$

THEOREM 7. *Under the sampling scheme of subsection* 2.3, *if the prior distribution of Y_t is the homogeneous process with* MGF (3.21) *with γ continuous, the posterior expectation of $S(t) = 1 - F(t)$ is*

$$\begin{aligned} \mathcal{E}(S(t)|data) &= e^{-\gamma(t)/(h_{j(t)}+\tau)} \\ &\cdot \textstyle\prod_{i=1}^{j(t)}\left[e^{\gamma(u_i)(h_{i-1}-h_i)/((h_{i-1}+\tau)(h_i+\tau))}\left(\dfrac{h_i + \lambda_i + \tau}{h_i + \lambda_i + \delta_i + \tau}\right)\right]. \end{aligned} \tag{3.26}$$

The proof is by direct substitution of (3.23) and (3.25) into (2.28).

REMARK 3. If we fix the prior guess at S to be S_0 so that

$$\mathcal{E}(S(t)) = M_t(1) = e^{-\gamma(t)/\tau} = 1 - F_0(t) = S_0(t) \tag{3.27}$$

then $\gamma(t) = -\tau \log S_0(t)$. Therefore, we may put (3.26) in an alternate form

$$\begin{aligned} \mathcal{E}(S(t)|\text{data}) &= S_0(t)^{\tau/(h_{j(t)}+\tau)} \\ &\cdot \textstyle\prod_{i=1}^{j(t)}\left[S_0(u_i)^{-\tau(h_{i-1}-h_i)/((h_{i-1}+\tau)(h_i+\tau))}\left(\dfrac{h_i + \lambda_i + \tau}{h_i + \lambda_i + \delta_i + \tau}\right)\right]. \end{aligned} \tag{3.28}$$

If $S_0(t) > 0$ for all t, we have as $\tau \to 0$

$$\begin{aligned} \mathcal{E}(S(t)|\text{data}) &\to \textstyle\prod_{i=1}^{j(t)} \dfrac{h_i + \lambda_i}{h_i + \lambda_i + \delta_i} && \text{for} \quad t < u_k \\ &\to \dfrac{S_0(t)}{S_0(u_k)} \textstyle\prod_{i=1}^{k} \dfrac{h_i + \lambda_i}{h_i + \lambda_i + \delta_i} && \text{for} \quad t \geqslant u_k \end{aligned} \tag{3.29}$$

where $(h_k + \lambda_k)/(h_k + \lambda_k + \delta_k)$ is replaced by 1 if it is 0/0. This is a maximum likelihood estimate of $S(t)$ as is seen in the next remark. In particular, when there are no censored observations ($\lambda_i = 0$, $h_i + \delta_i = h_{i-1}$), this reduces to the sample distribution function.

To investigate the effect of the choice of τ, we specialize (3.28) to a single uncensored observation at x:

$$\mathcal{E}(S(t)|X = x) = S_0(t)^{\tau/(1+\tau)} \quad \text{for} \quad t < x$$
$$= S_0(t)S_0(x)^{-1/(1+\tau)} \cdot \frac{\tau}{1+\tau} \quad \text{for} \quad t \geqslant x. \tag{3.30}$$

For large values of τ, this is close to $S_0(t)$, while for small values of τ, it is close to the function with a single jump of size 1 at x. Thus, as in the gamma process case, τ measures prior "strength of belief"; but again it is hard to calibrate τ in terms of "prior sample size" because of the size of the jump

$$S_0(x)^{\tau/(1+\tau)} \cdot \frac{1}{1+\tau} \tag{3.31}$$

is still strongly affected by the value of $S_0(x)$. It is easily seen that for all homogeneous processes neutral to the right, the posterior jump size depends on $S_0(x)$ and tends to zero as $S_0(x)$ tends to zero.

REMARK 4. For the situation in which all censorings are exclusive censorings, Kaplan and Meier [7] have shown that the product limit estimate is a maximum likelihood estimate. By an application of the lemma in Section 5 one can extend their results to two types of censoring and obtain the maximum likelihood estimate of $S(t)$ which may be written as

$$\hat{S}(t) = \Pi_{i=1}^{j(t)} \frac{\lambda_i + h_i}{\delta_i + \lambda_i + h_i} \tag{3.32}$$

where, as before, $h_i = \Sigma_{j=i+1}^{k}(\delta_j + \lambda_j + \mu_j)$ denotes the number of observations to the right of u_i.

Two conventions are necessary for this to represent the class of all maximum likelihood estimates. First, if $\lambda_k = 0$ and $\delta_k = 0$, so that

$$\frac{\lambda_k + h_k}{\delta_k + \lambda_k + h_k} = \frac{0}{0}, \tag{3.33}$$

then this term is to be taken to represent an arbitrary number in [0, 1], the MLE being not unique in this case. Second, if the mass above u_k

$$m_k = \Pi_{i=1}^{k} \frac{\lambda_i + h_i}{\delta_i + \lambda_i + h_i} > 0, \tag{3.34}$$

then again the MLE is not unique, and the mass m_k can be distributed arbitrarily in (u_k, ∞) without changing the likelihood. Thus for $t > u_k$, if $\hat{S}(t)$ in (3.32) is positive ($= m_k$), we assume that it represents an arbitrary number in $[0, m_k]$, subject of course to the requirement that $\hat{S}$ be nonincreasing.

4. Application to a result of Susarla and Van Ryzin. As an example of the type of computations involved in the application of (2.24) to nonhomogeneous processes, we consider the Dirichlet process and derive the estimate of Susarla and Van Ryzin [10]. Let α be a nondecreasing right-continuous function on $\mathbb{R}$ such that $\lim_{t\to-\infty}\alpha(t) = 0$ and $\lim_{t\to+\infty}\alpha(t) = \alpha(\mathbb{R}) < \infty$. As pointed out in Ferguson [6], the Dirichlet process, $\mathfrak{D}(\alpha)$, can be defined as the random distribution function neutral to the right for which the MGF of $Y_t = -\log(1 - F(t))$ has the forms

$$(4.1)\qquad M_t(\theta) = \frac{\Gamma(\alpha(\mathbb{R}))}{\Gamma(\alpha(t))\Gamma(\alpha(\mathbb{R}) - \alpha(t))}\int_0^\infty e^{-(\alpha(\mathbb{R})-\alpha(t)+\theta)y}(1 - e^{-y})^{\alpha(t)-1}dy$$
$$= \frac{\Gamma(\alpha(\mathbb{R}))\Gamma(\alpha(\mathbb{R}) - \alpha(t) + \theta)}{\Gamma(\alpha(\mathbb{R}) - \alpha(t))\Gamma(\alpha(\mathbb{R}) + \theta)}$$
$$= e^{\int_0^\infty(e^{-\theta z}-1)dN_t(z)}$$

where the Lévy measure $N_t(z)$ is given by

$$(4.2)\qquad dN_t(z) = \frac{e^{-\alpha(\mathbb{R})z}(e^{\alpha(t)z} - 1)}{z(1 - e^{-z})}dz,$$

provided $0 < \alpha(t) < \alpha(\mathbb{R})$. Let us denote the distribution with MGF (4.1) as $\mathcal{H}(\alpha(t), \alpha(\mathbb{R}) - \alpha(t))$. Note that for $s < t$, $Y_t - Y_s \in \mathcal{H}(\alpha(t) - \alpha(s), \alpha(\mathbb{R}) - \alpha(t))$.

To be able to apply the corollary of Theorem 4, we need to evaluate the functions R and r. It is easy to compute $R_t(h)$ from (2.25) and (4.1)

$$(4.3)\qquad R_t(h) = \frac{\alpha(\mathbb{R}) - \alpha(t) + h}{\alpha(\mathbb{R}) + h}.$$

To evaluate $r_u(a, b)$, we consider two cases. First, suppose that u is a prior fixed point of discontinuity of Y_t, so that u is a point of discontinuity of α. If $\Delta(u)$ denotes the mass at u, i.e., $\Delta(u) = \alpha(u) - \alpha^-(u)$, then the prior distribution of the jump $Y_u - Y_u^-$ is $\mathcal{H}(\Delta(u), \alpha(\mathbb{R}) - \alpha(u))$. From (2.19) and (4.1)

$$(4.4)\qquad C_u(a, b) = \frac{\Gamma(\alpha(\mathbb{R}) - \alpha(u) + \Delta(u))\Gamma(\Delta(u) + b)\Gamma(\alpha(\mathbb{R}) - \alpha(u) + a)}{\Gamma(\Delta(u))\Gamma(\alpha(\mathbb{R}) - \alpha(u))\Gamma(\alpha(\mathbb{R}) - \alpha(u) + \Delta(u) + a + b)},$$

so that

$$(4.5)\qquad r_u(a, b) = \frac{\alpha(\mathbb{R}) - \alpha(u) + a}{\alpha(\mathbb{R}) - \alpha(u) + \Delta(u) + a + b}.$$

Next, consider the case when u is not a prior fixed point of discontinuity of Y_t. It is known from the general theory of Dirichlet processes that the posterior distribution of F given $X = u$ is a Dirichlet process with new parameter $\alpha_u(t) = \alpha(t) + I_{[u,\infty)}(t)$. Therefore, the posterior distribution of the jump in Y_t at u given $X = u$ is $\mathcal{H}(1, \alpha(\mathbb{R}) - \alpha(u))$. In evaluating (2.20), we arrive at formula (4.4) with $\Delta(u) = 0$, even in the case $b = 0$. Consequently, (4.5) is valid in all cases if $\Delta(u)$ is defined, as it should be, to be zero if u is a point of continuity of α.

The corollary to Theorem 4 then leads immediately to the following theorem.

THEOREM 8. *Under the sampling scheme of subsection* 2.3, *if the prior distribution of* F *is the Dirichlet process*, $\mathfrak{D}(\alpha)$, *then the posterior expectation of* $S(t)$ *is*

$$(4.6)\quad \mathcal{E}(S(t)|data) = \frac{\alpha(\mathbb{R}) - \alpha(t) + h_{j(t)}}{\alpha(\mathbb{R}) + n} \cdot \prod_{i=1}^{j(t)} \frac{(\alpha(\mathbb{R}) - \alpha^{-}(u_i) + h_{i-1})(\alpha(\mathbb{R}) - \alpha(u_i) + h_i + \lambda_i)}{(\alpha(\mathbb{R}) - \alpha(u_i) + h_i)(\alpha(\mathbb{R}) - \alpha^{-}(u_i) + h_i + \lambda_i + \delta_i)}.$$

A data point u_i is a censoring point if $\lambda_i > 0$ or $\mu_i > 0$. If u_i is not a censoring point, the corresponding factor in the term on the right of (4.6) is identically one, since $\lambda_i = 0$ and $h_{i-1} = h_i + \delta_i$. Therefore the product in (4.6) may be restricted to censoring points.

This is a version of a formula derived by Susarla and Van Ryzin [10]. The following differences should be noted.

1. In this paper it is assumed that the censoring points are given constants, whereas in [10] the censoring points are allowed to be random variables, chosen independent and identically distributed according to some known distribution, independently of the observations from F. Thus, their model fits some applications, such as the cancer study, mentioned in the introduction, very well. Here we take the point of view of the Bayesian who looks at the observations as information. This allows greater freedom. For example, this allows treatment of sequential problems in which future censoring points are chosen depending on past data.

2. In [10] all censoring was restricted to be exclusive censoring. Although this is a minor mathematical detail, generality is gained in our formulation without increased complexity of the formulas. Thus, the formula of Susarla and Van Ryzin is really (4.6) with $\boldsymbol{\mu} = \mathbf{0}$, which is really the same as $h_{j-1} = h_j + \delta_j + \lambda_j$ for $j = 1, \cdots, n$.

5. Modal estimation with Dirichlet process priors. In the problem of estimating a potency curve in a bioassay model, Ramsey [9] considers the modal estimates with Dirichlet process priors. The modal estimate of F for the problem of this paper can be reduced to Ramsey's estimate when the prior is a Dirichlet process because of the following considerations. All "real" observations, $X_i = x_i$, may be taken care of by "updating" the Dirichlet process prior to the posterior Dirichlet process. The remaining censored observations may be considered as bioassay "failures" with levels at the censoring points. Thus, the application of Ramsey's formulas when all observations are failures and the Dirichlet process prior is already updated by the real observations, gives the modal estimate of F for our problem.

However, in the bioassay problem Ramsey considers, there is no closed form solution for the modal estimates. Successive approximation is required in the general case. We see below that when all observations are failures, Ramsey's estimates can be written in a simple closed form. In particular, when applied to our problem, the modal estimate turns out to be the same as the mean estimate, (4.6).

Here, unlike the previous sections, the generality gained in allowing two types of censoring is paid for in greatly increased complexity of bothersome details in both statement and proof of the theorem. Therefore, for this theorem, we restrict consideration to exclusive censoring as is done in [7] and [10].

There are analytical difficulties in defining the mode of an infinite-dimensional distribution. We avoid these difficulties by restricting attention to finite-dimensional subsets of the variables. Let $t_1, \cdots, t_r$ be arbitrary points; we hope to compute the mode of the joint distribution of $(F(t_1), \cdots, F(t_r))$. This introduces a new difficulty, namely, that the modal value $\hat{F}(t_i)$ of $F(t_i)$ may depend on $t_1, \cdots, t_r$. After all, the vector mode of a multidimensional distribution need not be the same as the vector of modes of the marginal distributions. We shall see below that the modal value $\hat{F}(t_i)$ does not depend on $t_1, \cdots, t_r$ provided that all censoring points are included in $t_1, \cdots, t_r$ (more precisely, provided $t_1, \cdots, t_r$ contains all u_j for which $\lambda_j > 0$). This strong stability of the modal values permits us to *define* the modal estimate of $F(t)$ as the modal value of $F(t)$ in $(F(t), F(t_1), \cdots, F(t_r))$ when $t_1, \cdots, t_r$ contains all censoring points.

Therefore, let t be arbitrary, and let $t_1, \cdots, t_r$ be arbitrary distinct points containing t and all censoring points, and suppose that $t_1, \cdots, t_r$ are ordered in increasing order. We assume that F has a Dirichlet process prior with parameter α. In particular, the vector

$$(5.1) \qquad (p_1, p_2, \cdots, p_{r+1}) = (F(t_1), F(t_2) - F(t_1), \cdots, 1 - F(t_r))$$

has a finite dimensional Dirichlet distribution with parameters $(\beta_1, \beta_2, \cdots, \beta_{r+1})$, where $\beta_i = \alpha(t_i) - \alpha(t_{i-1})$ for $i = 1, \cdots, r+1$ with $\alpha(t_0) = 0$ and $\alpha(t_{r+1}) = \alpha(\mathbb{R})$. We assume for simplicity that $\beta_i > 0$ for $i = 1, \cdots, r+1$. This is equivalent to the assumption that α gives positive mass to every open interval.

Following Ramsey [9], we take, for convenience of analysis and interpretation, the density of the vector $(p_1, \cdots, p_r)$ with respect to the measure

$$(5.2) \qquad dv = \frac{\Pi_{i=1}^r dp_i}{\Pi_{i=1}^{r+1} p_i}$$

where $p_{r+1} \equiv 1 - \Sigma_1^r p_i$, over the simplex

$$S_r = \{(p_1, \cdots, p_r) : p_i \geqslant 0 \text{ for } i = 1, \cdots, r, \text{and } \Sigma_1^r p_i \leqslant 1\}.$$

The prior density of $(p_1, \cdots, p_r)$ over S_r with respect to dv is proportional to

$$(5.3) \qquad \Pi_{i=1}^{r+1} p_i^{\beta_i}.$$

We assume the data has the form of m_1 "real" observations, $X_1 = x_1, \cdots, X_{m_1} = x_{m_1}$, and m_2 exclusive censorings, $X_{m_1+1} > x_{m_1+1}, \cdots, X_n > x_n$, where $m_1 + m_2 = n$. As before, let $u_1, \cdots, u_k$ represent the distinct x_j's, let $\delta_1, \cdots, \delta_k$ represent the respective numbers of real observations at these points, and $\lambda_1, \cdots, \lambda_k$ represent the respective numbers of exclusive censored observations at these points, so that $\Sigma_1^k \delta_i = m_1$ and $\Sigma_1^k \lambda_i = m_2$. We refer to $\mathbf{u}, \boldsymbol{\delta}, \boldsymbol{\lambda}$ as the data.

As before, we let $h_j = \Sigma_{j+1}^k(\delta_i + \lambda_i)$, so that $h_j + \delta_j + \lambda_j = h_{j-1}$, and we let $j(t)$ denote the number of u_i less than or equal to t.

THEOREM 9. *Suppose $F \in \mathfrak{D}(\alpha)$, where α gives positive mass to every interval. Then the posterior modal (with respect to ν) estimate of $1 - F$ given the data is*

$$(5.4) \qquad 1 - \hat{F}(t) = \frac{\alpha(\mathbb{R}) - \alpha(t) + h_{j(t)}}{\alpha(\mathbb{R}) + n} \cdot \Pi_{i=1}^{j(t)} \frac{\alpha(\mathbb{R}) - \alpha(u_i) + h_i + \lambda_i}{\alpha(\mathbb{R}) - \alpha(u_i) + h_i}.$$

The proof is based on the following lemma.

LEMMA 1. *Let*

$$(5.5) \qquad f(\mathbf{P}) = \Pi_{j=1}^N\left[P_j^{\alpha_j}(P_j - P_{j+1})^{\beta_j}\right]$$

where $\alpha_i \geqslant 0$ and $\beta_i \geqslant 0$ for all i. The maximum of $f(\mathbf{P})$ subject to the constraints

$$(5.6) \qquad 1 = P_1 \geqslant P_2 \geqslant \cdots \geqslant P_N \geqslant P_{N+1} = 0$$

occurs at points $\hat{\mathbf{P}}$ such that

$$(5.7) \qquad \hat{P}_j = \Pi_{k=1}^{j-1} \frac{\Sigma_{k+1}^N(\alpha_i + \beta_i)}{\beta_k + \Sigma_{k+1}^N(\alpha_i + \beta_i)} \qquad \text{for} \quad j = 2, \cdots, N$$

with the convention that terms of the form $0/0$ in (5.7) represent arbitrary numbers in $[0, 1]$.

PROOF. Every point $(P_2, \cdots, P_N)$ in the constraint set (5.6) has a representation as

$$(5.8) \qquad P_j = \Pi_{k=1}^{j-1}x_k \qquad \text{for} \quad j = 2, \cdots, N$$

where $0 \leqslant x_j \leqslant 1$ for $j = 1, \cdots, N - 1$, and conversely every point of the form (5.8) is in the constraint set. In terms of the x_j, the function f becomes

$$(5.9) \qquad f = \Pi_{j=1}^{N-1}\left[(1 - x_j)^{\beta_j} x_j^{\Sigma_{j+1}^N(\alpha_i + \beta_i)}\right].$$

This may be maximized by maximizing the terms in square brackets separately for each $j = 1, \cdots, N - 1$. The point of maximum value is easily found to be

$$(5.10) \qquad x_j = \frac{\Sigma_{j+1}^N(\alpha_i + \beta_i)}{\beta_j + \Sigma_{j+1}^N(\alpha_i + \beta_i)} \qquad \text{for} \quad j = 1, \cdots, N - 1,$$

where $0/0$ denotes an arbitrary number in $[0, 1]$. Substitution of (5.10) into (5.8) yields (5.7) immediately.

PROOF OF THE THEOREM. Let $t_1, t_2, \cdots, t_r$ be arbitrary distinct points containing t and all censoring points, and suppose that $t_1, \cdots, t_r$ are ordered in increasing order. Let $n_1, n_2, \cdots, n_{r+1}$ denote the number of real observations that fall in the intervals $(-\infty, t_1]$, $(t_1, t_2]$, $\cdots$, (t_r, ∞) respectively and let l_j denote the number of censored observations at t_j, $j = 1, \cdots, r$ ($l_j = 0$ unless t_j is a censoring point).

The joint likelihood of observing a sample as described above is

$$(5.11)\qquad \Pi_{j=1}^{r+1}[F(t_j) - F(t_{j-1})]^{n_j}(1 - F(t_{j-1}))^{l_{j-1}} = \Pi_{j=1}^{r+1} p_j^{n_j}(\Sigma_{i=j}^{r+1} p_i)^{l_{j-1}}$$

where $F(t_0) = 0$ and $F(t_{r+1}) = 1$, and $l_0 = 0$. Since the prior density of $\mathbf{p}$ with respect to the measure $d\nu$ is proportional to (5.3), the posterior density of $\mathbf{p}$ with respect to $d\nu$ is proportional to

$$(5.12)\qquad \Pi_{j=1}^{r+1} p_j^{n_j+\beta_j}(\Sigma_{i=j}^{r+1} p_i)^{l_{j-1}}.$$

The mode of the posterior distribution is obtained by finding the point $\mathbf{p}$ that maximizes this expression subject to $p_i \geqslant 0$ for all i and $\Sigma_{i=1}^{r+1} p_i = 1$. Applying the results of Lemma 1 with $N = r + 1$, and $P_j = \Sigma_j^{r+1} p_i$ for $j = 1, \cdots, r + 1$, we obtain as the modal estimate at t_j

$$(5.13)\qquad 1 - \hat{F}(t_j) = \Sigma_{j+1}^{r+1}\hat{p}_i$$

$$= \Pi_{k=1}^{j}\frac{\Sigma_{k+1}^{r+1}(n_i + \beta_i + l_{i-1})}{n_k + \beta_k + \Sigma_{k+1}^{r+1}(n_i + \beta_i + l_{i-1})}.$$

Since $\Sigma_{k+1}^{r+1}\beta_i = \alpha(\mathbb{R}) - \alpha(t_k)$, and since $\Sigma_{k+1}^{r+1}(n_i + l_{i-1}) = h(t_k) + l_k$, where $h(t) = h_{j(t)}$ represents the number of observations, real or censored, greater than t, (5.13) reduces easily to

$$(5.14)\qquad 1 - \hat{F}(t_j) = \Pi_{k=1}^{j}\frac{\alpha(\mathbb{R}) - \alpha(t_k) + h(t_k) + l_k}{\alpha(\mathbb{R}) - \alpha(t_{k-1}) + h(t_{k-1})}$$

$$= \frac{\alpha(\mathbb{R}) - \alpha(t_j) + h(t_j)}{\alpha(\mathbb{R}) + n}\Pi_{k=1}^{j}\frac{\alpha(\mathbb{R}) - \alpha(t_k) + h(t_k) + l_k}{\alpha(\mathbb{R}) - \alpha(t_k) + h(t_k)}.$$

We are to show, for the t_j which is equal to t, that (5.14) reduces to (5.4). Note that the product in (5.14) may be taken over just those k for which $l_k \neq 0$, that is, over the t_k corresponding to the censoring points. This shows that $1 - \hat{F}(t)$ depends only on the censoring points among $t_1, \cdots, t_r$, and is thus independent of the choice of $t_1, \cdots, t_r$ provided all censoring points are included. Furthermore, if $t_k = u_i$, a censoring point, then $l_k = \lambda_i$ and $h(t_k) = h_i$. The estimate (5.4) follows easily from this.

REMARK 1. The following example shows that if not all censoring points are included in $t_1, \cdots, t_r$, then the mode of $F(t_j)$ may be different from that given in (5.4).

Suppose there is a sample of size one consisting of a censored observation at the point y, and let t be a point greater than y. The joint mode of $(F(y), F(t))$ is seen to be (from (5.4) with $h_1 = 0$ and $\lambda_1 = 1$)

$$\left(\frac{\beta_1}{\alpha(\mathbb{R}) + 1}, 1 - \frac{\beta_3}{\alpha(\mathbb{R}) + 1}\cdot\frac{\beta_2 + \beta_3 + 1}{\beta_2 + \beta_3}\right)$$

where $\beta_1 = \alpha(y)$, $\beta_2 = \alpha(t) - \alpha(y)$, and $\beta_3 = \alpha(\mathbb{R}) - \alpha(t)$. To compute the marginal mode of $F(t)$, we proceed as follows. Denote $F(y)$ and $1 - F(t)$ by p_1 and p_3

respectively. The posterior density of p_1 and p_3 with respect to the measure $dv = dp_1dp_3/(p_1(1 - p_1 - p_3)p_3)$ is $c(1 - p_1)p_1^{\beta_1}(1 - p_1 - p_3)^{\beta_2}p_3^{\beta_3}$, where c is a constant. The posterior marginal density of p_3 with respect to the measure $dv_3 = dp_3/(p_3(1 - p_3))$ is

$$f_3(p_3) = cp_3^{\beta_3}(1 - p_3)^{\beta_1+\beta_2}(\beta_1 p_3 + \beta_2).$$

Finding the value of p_3 that maximizes this expression, we obtain the modal estimate of $p_3 = 1 - F(t)$, as the unique root between zero and one of the quadratic equation

$$(\alpha + 1)\beta_1 p_3^2 + (\alpha\beta_2 - \beta_1 - \beta_1\beta_3)p_3 - \beta_2\beta_3 = 0,$$

where $\alpha = \beta_1 + \beta_2 + \beta_3$. That this does not yield the same value as in (5.4) may be verified by taking some numerical values for β_1, β_2 and β_3, e.g. $\beta_1 = 1$, $\beta_2 = 2$, $\beta_3 = 3$.

REMARK 2. In the bioassay problem, Ramsey [9] wondered (footnote on page 846) whether the modal estimate of the potency curve at an observational dose-level x_i is also the mean of the marginal posterior distribution of $F(x_i)$. If it were so, then Theorem 9 would follow directly from Theorem 8. We see that this is not the case by considering the following example.

Let z_1 and z_2 ($z_1 < z_2$) be observational dose levels. One animal at each of these dose levels is subjected to an experiment in which it is observed that the animal at z_1 has survived the dose while the animal at z_2 did not. The likelihood of this sample is $(1 - F(z_1))F(z_2)$. With the prior $\mathfrak{D}(\alpha)$, the modal estimate of F at z_1 and z_2 can be obtained by solving the equations

$$\frac{\beta_2}{p_2} + \frac{1}{1 - p_1} = \frac{\beta_1}{p_1}$$

and

$$\frac{\beta_2}{p_2} + \frac{1}{p_1 + p_2} = \frac{\beta_3}{1 - p_1 - p_2}$$

where $p_1 = F(z_1)$, $p_2 = F(z_2) - F(z_1)$, $\beta_1 = \alpha((-\infty, z_1])$, $\beta_2 = \alpha((z_1, z_2])$ and $\beta_3 = \alpha((z_2, \infty))$. The mean estimate of F at z_1 and z_2 is obtained from Antoniak [1] as

$$\beta_1\{\beta_2 + (\beta_2 + \beta_3)(\beta_1 + \beta_2 + 1)\}/(\alpha(\mathbb{R}) + 2)[\beta_2(\alpha(\mathbb{R}) + 1) + \beta_1\beta_3]$$

and

$$(\beta_1 + \beta_2 + 1)[(\alpha(\mathbb{R}) + 2)\beta_2 + \beta_1\beta_3]/(\alpha(\mathbb{R}) + 2)[\beta_2(\alpha(\mathbb{R}) + 1) + \beta_1\beta_3].$$

Again, the fact that these two sets of estimates are not equal may be verified by taking a particular set of numerical values of β_1, β_2 and β_3.

6. Numerical examples. We illustrate the application of the results obtained in earlier sections by reworking the example of Kaplan and Meier [7]. Their data is as follows:

(6.1) real observations at 0.8, 3.1, 5.4, and 9.2 months;
censored observations at 1.0, 2.7, 7.0, and 12.1 months.

All censored observations are of the exclusive type.

This example has been used by Susarla and Van Ryzin to compare the product limit estimate with their Bayes estimate using a Dirichlet process prior with parameter α on $(0, \infty)$ of the form $\alpha(t, \infty) = e^{-\theta t}$, where θ was chosen to be 0.1. A posteriori, this seems like a reasonable choice for θ since the maximum likelihood estimate of θ when the survival curve is known to be of the form $e^{-\theta t}$ is close to 0.1 (actually $0.097 \cdots$).

In the tables below, we compare the estimate obtained by Susarla and Van Ryzin with the Bayes estimates based on the homogeneous process priors neutral to the right discussed in Section 3. We take the prior guess at S to be the same as that chosen by Susarla and Van Ryzin.

$$S_0(t) = 1 - F_0(t) = e^{-0.1t} \qquad \text{for} \quad t > 0. \tag{6.2}$$

Corresponding to the choice by Susarla and Van Ryzin of a prior sample size $\alpha(\mathbb{R}) = 1$, we choose the intensity parameter $\tau = 1$ in formulas (3.15) and (3.28) because we feel that τ should correspond reasonably well to prior sample size, at least in the simple homogeneous process case. In these formulas, we have from (6.1) distinct observations at $u_1 = 0.8$, $u_2 = 1.0$, $u_3 = 2.7$, $u_4 = 3.1$, $u_5 = 5.4$, $u_6 = 7.0$, $u_7 = 9.2$, and $u_8 = 12.1$. Furthermore, we have $\delta_1 = \delta_4 = \delta_5 = \delta_7 = 1$, $\lambda_2 = \lambda_3 = \lambda_6 = \lambda_8 = 1$, and the rest of the δ_i and λ_i and all of the μ_i equal to zero so that $h_i = 8 - i$ for $i = 0, 1, \cdots, 8$.

(a) *The gamma process prior.* Substituting (6.2) and $\tau = 1$ into (3.16), we obtain $\gamma(t) = 0.1443t$. Therefore, from Theorem 6, we find our Bayes estimate to be

$$\hat{S}_G(t) = \left(\frac{9 - j(t)}{10 - j(t)}\right)^{0.1443t} \cdot \prod_{i=1}^{j(t)} \left[\left(\frac{(10-i)^2}{(11-i)(9-i)} \right)^{0.1443u_i} \left[\frac{\ln\left(\frac{11-i}{10-i}\right)}{\ln\left(\frac{10-i}{9-i}\right)} \right]^{\delta_i} \right]$$

where $j(t)$ is the number of observations less than or equal to t.

(b) *The simple homogeneous process prior.* From (3.29) with (6.2) for S_0 and $\tau = 1$, we obtain as the Bayes estimate

$$\hat{S}_H(t) = e^{-0.1t/(9-j(t))} \cdot \prod_{i=1}^{j(t)} \left[e^{+0.1u_i/((10-i)(9-i))} \left(\frac{10 - i - \delta_i}{10 - i} \right) \right].$$

(c) *The Dirichlet process prior.* As is found in [6] or as may be deduced from

TABLE 2
Comparison of the functional form of three estimates of $S(t)$

interval for t	$\hat{S}_G(t)$	$\hat{S}_H(t)$	$\hat{S}_D(t)$
[0.0, 0.8)	$(\frac{9}{10})^{0.1443t} \cdot 1.0000$	$e^{-t/90} \cdot 1.0000$	$(e^{-t/10}+8)\cdot 0.1111$
[0.8, 1.0)	$(\frac{8}{9})^{0.1443t} \cdot 0.8958$	$e^{-t/80} \cdot 0.8899$	$(e^{-t/10}+7)\cdot 0.1111$
[1.0, 2.7)	$(\frac{7}{8})^{0.1443t} \cdot 0.8979$	$e^{-t/70} \cdot 0.8915$	$(e^{-t/10}+6)\cdot 0.1272$
[2.7, 3.1)	$(\frac{6}{7})^{0.1443t} \cdot 0.9051$	$e^{-t/60} \cdot 0.8972$	$(e^{-t/10}+5)\cdot 0.1493$
[3.1, 5.4)	$(\frac{5}{6})^{0.1443t} \cdot 0.7749$	$e^{-t/50} \cdot 0.7554$	$(e^{-t/10}+4)\cdot 0.1493$
[5.4, 7.0)	$(\frac{4}{5})^{0.1443t} \cdot 0.6536$	$e^{-t/40} \cdot 0.6209$	$(e^{-t/10}+3)\cdot 0.1493$
[7.0, 9.2)	$(\frac{3}{4})^{0.1443t} \cdot 0.6977$	$e^{-t/30} \cdot 0.6582$	$(e^{-t/10}+2)\cdot 0.2091$
[9.2, 12.1)	$(\frac{2}{3})^{0.1443t} \cdot 0.5788$	$e^{-t/20} \cdot 0.5115$	$(e^{-t/10}+1)\cdot 0.2091$
[12.1, ∞)	$e^{-t/10} \cdot 0.9563$	$e^{-t/10} \cdot 0.9367$	$e^{-t/10} \cdot 0.9102$

(4.6) with $\alpha(t, \infty) = e^{-0.1t}$, the Bayes estimate with a $\mathcal{D}(\alpha)$ prior is given by

$$\hat{S}_D(t) = \frac{e^{-0.1t} + 8 - j(t)}{9} \cdot \prod_{i=1}^{j(t)} \frac{e^{-0.1u_i} + 8 - i + \lambda_i}{e^{-0.1u_i} + 8 - i}.$$

These estimates are compared in Tables 2 and 3. Table 2 is arranged to bring out the behavior of the estimates between the observations. In Table 3, the actual values of the estimates are computed for all the observational points. When there is only one number for an estimate at an observational point, the estimated survival curve is continuous at that point. When there are two numbers, the upper number represents the left limit of the function at the point, and the lower number represents the right limit. The difference represents therefore the mass assigned to the point by the estimate.

The product limit estimate, or MLE,

$$\hat{S}_{PL}(t) = \prod_{i=1}^{j(t)} \left(\frac{9 - i - \delta_i}{9 - i} \right)$$

is included in Table 3 for comparison. It is seen that all the estimates are quite close. However, the estimate based on the Dirichlet process prior comes closer to the product limit estimate than the other two; in particular, it assigns more mass to the real observations and less mass between the observations and at the tails than the other two. This may be because an intensity parameter value of $\tau = 1$ represents a prior sample size of somewhat more than one (and a somewhat larger prior sample size for the gamma prior than for the simple homogeneous prior). But the situation is not clear because some of the information obtained from a real observation is used to change the relative weight of points to the left of it, unlike for the Dirichlet process where the only change is at the observation itself.

It should be recalled from the discussion of Remark 2 in Section 3, that as $\tau \to 0$ the estimate $\hat{S}_G(t)$ does not converge to the product limit estimate. For the example

TABLE 3
Numerical comparison of five estimates of $S(t)$

t	0.8	1.0	2.7	3.1	5.4	7.0	9.2	12.1
	.9879			.8448	.6723		.4762	
$\hat{S}_G(t)$		.8807	.8523			.5217		.2851
	.8837			.7142	.5493		.3379	
	.9912			.8521	.6781		.4844	
$\hat{S}_H(t)$		.8789	.8577			.5212		.2793
	.8810			.7100	.5425		.3229	
	.9915			.8559	.6841		.5015	
$\hat{S}_D(t)$		.8783	.8605			.5220		.2714
	.8803			.7066	.5348		.2924	
	1.0000			.8750	.7000		.5250	
$\hat{S}_{PL}(t)$		.8750	.8750			.5250		.2625
	.8750			.7000	.5250		.2625	
	1.0000			.8821	.7207		.5590	
$\hat{S}_{G,0}(t)$		.8821	.8821			.5590		.3270
	.8821			.7207	.5590		.3270	

discussed here, it converges instead to

$$\hat{S}_{G,0}(t) = \Pi_{i=1}^{j(t)} \left[\frac{\log\left(\frac{10-i}{9-i}\right)}{\log\left(\frac{9-i}{8-i}\right)} \right]^{\delta_i} \quad \text{for} \quad t \leqslant 12.1.$$

This estimate gives more weight to the right tails than the product limit estimate. However, since all estimates of Table 2 give greater weight to the right tail, perhaps this estimate will turn out better than the product limit estimate. Upon computing the values of this estimate at the observational points, we see in Table 3 that it gives quite a bit more weight to the right tail than any of the other estimates.

REFERENCES

[1] ANTONIAK, C. (1974). Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *Ann. Statist.* **2** 1152–1174.

[2] BRESLOW, N. and CROWLEY, J. (1974). A large sample study of the life table and product limit estimates under random censorship. *Ann. Statist.* **2** 437–453.

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

[4] DYKSTRA, RICHARD and LAUD, P. (1976). A nonparametric Bayesian approach toward reliability (preliminary report)-Abstract. *Bull. Inst. Math. Statist.* **5** 171.

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

[6] FERGUSON, T. S. (1974). Prior distributions on spaces of probability measures. *Ann. Statist.* **2** 615–629.

[7] KAPLAN, E. L. and MEIER, P. (1958). Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.* **53** 457–481.
[8] LÉVY, P. (1937). *Théorie de l'Addition des Variables Aléatoire*. Gauthier-Villars, Paris.
[9] RAMSEY, F. L. (1972). A Bayesian approach to bioassay. *Biometrics* **28** 841–858.
[10] SUSARLA, V. and VAN RYZIN, J. (1976). Nonparametric Bayesian estimation of survival curves from incomplete observations. *J. Amer. Statist. Assoc.* **71** 897–902.
[11] WAGNER, STEPHEN S. and ALTMANN, STUART A. (1973). What time do the baboons come down from the trees? *Biometrics* **29** 623–635.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF CALIFORNIA
LOS ANGELES, CALIFORNIA 90024