Ann. Inst. Statist. Math.
Vol. 41, No. 2, 227–245 (1989)

ON A CLASS OF BAYESIAN NONPARAMETRIC ESTIMATES: II. HAZARD RATE ESTIMATES

ALBERT Y. LO* AND CHUNG-SING WENG

Department of Statistics, State University of New York at Buffalo, Buffalo, NY 14261, U.S.A.

(Received October 1, 1985; revised October 12, 1988)

Abstract. The Bayes estimation of hazard rates for a family of multiplicative point processes is considered. We study the model for which a hazard rate can be linearly parametrized by a freely varied measure. The weighted gamma process is assumed to be the prior distribution of this measure; the posterior distributions and the posterior means are given in explicit forms. Examples of the evaluation of posterior means are given.

Key words and phrases: Bayesian method, hazard rates, gamma process.

1. Introduction

An important problem in statistics is the estimation of the hazard rate of a point process. Estimation procedures appropriate for parametric models can be found in the literature of life table analysis and competing risks models (Chiang (1968)), Poisson point process models (Brown (1972), Grandell (1972) and Clevenson and Zidek (1977)) and general point process models (Lewis (1972)). Aalen (1978) presented a unified theory of nonparametric inference for the cumulative hazards of a multiplicative counting process from a frequentist viewpoint; he showed that the above models as well as models in reliability theory, Markov chains with censoring, and birth and death process are special cases of the multiplicative counting process model. Bayesian inference for the Poisson point process was considered by Lo (1982); the parameters of interest are also the cumulative hazards (also called cumulative intensities).

In this paper, we consider the Bayes estimation of the derivatives of the cumulative hazards (called hazard rates) for a multiplicative counting process model. The idea of our approach is that estimating a density and estimating a hazard rate are analogous affairs, and a successful attempt of one generally leads to a feasible approach for the other. The approach

*The research of this author is supported in part by NSF Grant MCS 81-02523-01.

taken here is then based on the mixture method developed in Lo ((1984), referred to as Part I hereafter) for the Bayes estimation of a density; the difference is in the choice of a weighted gamma process as the prior distribution on the mixing measure. Dykstra and Laud's (1981) results in reliability theory are shown to be special cases.

In Section 2, the multiplicative counting process (Aalen (1978)) and its associated likelihood function are introduced and illustrated using simple examples. Section 3 discusses weighted gamma priors and presents a key tool of this paper: a Fubini-type lemma for the weighted gamma random measure. Applications of this lemma to the study of the path property of a weighted gamma random measure and the Bayes estimation of the cumulative intensity of a nonhomogeneous Poisson process are given; the connection of this lemma and a similar result on the Dirichlet process are also discussed. Section 4 shows that if the prior distribution of the mixing measure is a weighted gamma distribution, the posterior distribution is a mixture of weighted gamma distributions, and the posterior mean can be derived in an explicit form. Section 5 discusses the choice of priors. Section 6 discusses a Monte Carlo method for evaluating some posterior quantities using the Chinese restaurant process of sampling partitions (Aldous (1985) and Kuo (1986); this method can also be used to approximate posterior quantities appearing in Lo (1984)); numerical examples are given.

Some of the results in this paper appeared in Lo (1978).

2. The likelihood function for the hazard rates in mixture models

2.1 *The likelihood function*

Let $N(t) = (N_1(t),\ldots, N_q(t))$, $t \in [0, 1]$ be a vector counting process, such that

(i) each $N_j(t)$ is a right continuous point process with jump size one,

(ii) at any instant, at most one jump can occur in any of the component processes,

(iii) $EN_j(1) < \infty$ for each j, and

(iv) the jump times are totally inaccessible; i.e., for any sequence of stopping times $S_1,\ldots, S_n,\ldots$, $P\left\{\lim_{k\to\infty} S_k = T_n \text{ and } S_k < T_n \text{ for all } k\right\} = 0$ (Meyer (1966)), where $T_1, T_2,\ldots, T_{N^*(1)}$ are the jump times, and $N^*(1) = \sum_{1\le j\le q} N_j(1)$.

Then it is well known (Aalen (1978)) that there exists a unique continuous increasing k-variate process $A(t) = (A_1(t),\ldots, A_q(t))$ which is adapted to the σ-fields $\mathscr{F}_t = \sigma\{N(s), s \le t\}$, such that

(2.1) for each j, $M_j(t) = N_j(t) - A_j(t)$ is a square-integrable martingale,

(2.2) $\langle M_j, M_j\rangle = A_j$, and

$$(2.3) \qquad \langle M_i, M_j \rangle = 0 \quad \text{for} \quad i \neq j\,,$$

where $\langle \cdot\,, \cdot \rangle$ is the inner product defined in Kunita and Watanabe (1967).

In this paper, we are interested in estimating the derivatives of the $A_j(\cdot)$'s, and we assume

(v) each path $A_j(\cdot)$ is absolutely continuous with respect to Lebesgue measure on $[0, 1]$.

Assumptions (i) to (v) imply that there exists a k-variate left-continuous process $\Lambda(t) = (\Lambda_1(t), \ldots, \Lambda_q(t))$, having right-hand limits, such that for each j and each $t \in [0, 1]$

$$(2.4) \qquad A_j(t) = \int I_{\{0 \le s \le t\}} \Lambda_j(s)\, ds.$$

The vector process $\Lambda(t)$ is called the hazard (intensity) rate process of $N(t)$. Next is the definition of a multiplicative counting process.

DEFINITION 2.1. A vector counting process $N(t)$ is called a multiplicative counting process if it satisfies (i) to (v), and its hazard rate process $\Lambda(t)$ satisfies the following multiplicative property

(vi) $\Lambda_j(t) = Y_j(s) r_j(s)$,

where both $Y_j(s)$'s and $r_j(s)$'s are nonnegative, left-continuous functions with right-hand limits, $Y_j(s)$'s are observables and $r_j(s)$'s are deterministic functions called hazard (intensity) rates.

Next, we discuss the likelihood function of a multiplicative counting process. Jacod (1975) showed that, under some weak conditions, the distribution of a multiplicative counting process is absolutely continuous with respect to the joint distribution of q i.i.d. homogeneous Poisson processes, and the likelihood function of a multiplicative counting process is proportional to

$$(2.5) \qquad \left[\prod_{1 \le i \le N^*(1)} r_{J_i}(T_i) \right] \exp\left\{ -\int I_{\{0 \le s \le 1\}} \sum_{1 \le j \le q} Y_j(s) \times r_j(s)\, ds \right\},$$

where $J_n = j$ if jump number n occurs in the j-th component N_j.

Denote the total number of jumps of the j-th component by $n(j)$ and its jump times by $T_{j1}, \ldots, T_{jn(j)}, j = 1, \ldots, q$. Equation (2.5) can be written as

$$(2.6) \qquad \prod_{1 \le j \le q} \left[\prod_{1 \le i \le n(j)} r_j(T_{ji}) \right] \exp\left\{ -\int I_{\{0 \le s \le 1\}} Y_j(s) \times r_j(s)\, ds \right\}.$$

Several examples of (2.6) are given in Section 4 of Aalen (1978). In the following, we describe three simple cases.

Example 2.1. (The life testing model) Let q units with independent and identically distributed life times be the subjects of life testing and let $r(t) = f(t)/[1 - F(t)]$ be the failure rate of the units. Suppose we start observation when the units are new, and terminate observation after the n-th failure has occurred (i.e., q-n censored observations). Without loss of generality we denote the n failure times by $T_1, \ldots, T_n$. Let $Y(t)$ = number of units on test just before age t. Note that $Y(t)$ is a left continuous integer valued step function which vanishes outside some bounded interval. The likelihood function is proportional to

$$(2.7) \qquad \left[\prod_{1 \le i \le n} r(T_i) \right] \exp \left\{ - \int I_{\{0 \le s \le 1\}} Y(s) \times r(s) ds \right\},$$

which is (2.6) with $q = 1$. See the discussion by Lindley in Cox (1972).

Example 2.2. (The multiple decrement model) Let n independent units be observed over the time interval $[0, 1]$. Assume that the observation of each unit is a Markov process $X(t)$, $t \in [0, 1]$, with one transient state 0 and absorbing states $\{1, \ldots, q\}$, and that all units start at the initial state 0. Denote the number of processes that jump to state j by $n(j)$. Let the jump times to the j-th state be $T_{j1}, \ldots, T_{jn(j)}$, $j = 1, \ldots, q$ and $Y(t)$ be the number of processes that stay at state 0 just before time t. The likelihood function (Hoem (1971) or Aalen (1976)) is proportional to

$$(2.8) \qquad \prod_{1 \le j \le q} \left[\prod_{1 \le i \le n(j)} r_j(T_{ji}) \right] \exp \left\{ - \int I_{\{0 \le s \le 1\}} Y(s) \times r_j(s) ds \right\}.$$

This is (2.6) with $Y_j(s) = Y(s)$ for $j = 1, \ldots, q$.

Example 2.3. (The nonhomogeneous Poisson process model) Suppose we observe a Poisson point process with cumulative intensity υ on the unit interval. Suppose the derivative of the υ exists, i.e., $\upsilon(t) = \int I_{\{0 \le s \le t\}} r(s) ds$, and we are interested in the estimation of the intensity rate $r(t)$ instead of υ. Let the jump times be $T_1, \ldots, T_n$. The likelihood function is given by

$$(2.9) \qquad \left[\prod_{1 \le i \le n} r(T_i) \right] \exp \left\{ - \int I_{\{0 \le s \le 1\}} r(s) ds \right\}.$$

This is equivalent to (2.6) with $Y(s) = 1$ and $q = 1$.

2.2 *A mixture model for the hazard rates*

This paper is concerned with the Bayes estimation of the hazard rates $r_j(t)$, $j = 1, \ldots, q$ for the model specified by (2.6). In particular, we consider the case that each hazard rate $r(t)$ can be represented as a mixture of a known kernel by a freely varied measure μ. That is, there is a given nonnegative kernel $k(t,v)$ defined on $([0,1]\times R, \mathscr{F}\otimes\mathscr{B})$ where R is a Euclidean space and $\mathscr{F}$ and $\mathscr{B}$ are Borel σ-fields, such that

$$r(t|\mu)=\int_R k(t,v)\mu(dv) \qquad \text{for} \qquad t\in[0,1]\,. \tag{2.10}$$

We also assume that for each $\mu\in\Theta$ the space of finite measures on $(R,\mathscr{B})$, $\int r(t|\mu)dt$ is finite. In this case, the likelihood function (2.6) of $\boldsymbol{\mu}=(\mu_1,\ldots,\mu_q)$ can be written as

$$\begin{aligned} L(\boldsymbol{\mu}) &= \prod_{1\le j\le q}\left[\prod_{1\le i\le n(j)} r_j(T_{ji})\right]\exp\left\{-\int I_{\{0\le s\le 1\}}Y_j(s)\times r_j(s)ds\right\} \\ &= \prod_{1\le j\le q}\left[\prod_{1\le i\le n(j)}\int k_j(T_{ji},v)\mu_j(dv)\right] \\ &\quad\times\exp\left\{-\iint I_{\{0\le s\le 1\}}Y_j(s)\times k_j(s,v)ds\mu_j(dv)\right\} \\ &= \prod_{1\le j\le q} L_j(\mu_j)\,, \end{aligned} \tag{2.11}$$

where $r_j(t)=\int_R k_j(t,v)\mu_j(dv)$, k_j is given and μ_j varies in Θ for $j=1,\ldots,q$. The mixing measures μ_j are the parameters to be estimated.

An inspection of (2.11) and a knowledge of conjugate priors indicate that gamma type priors on the μ's will facilitate the computations; the outer product also suggests independent gamma priors. We describe such gamma priors in the next section.

3. The weighted gamma prior distributions

The theory developed in this section supplements that of Lo (1982), and we first briefly describe the settings in that article. For each α in Θ, there is a gamma random measure υ on $(R,\mathscr{B})$ with shape (mean) measure α. The finite dimensional distributions of υ are determined by: for each (measurable) partition $B_1,\ldots,B_m$ of R, $\upsilon(B_j)$, $j=1,\ldots,m$ are independent gamma $(\alpha(B_j),1)$, $j=1,\ldots,m$ random variables.

Let $\beta(v)$ be a given nonnegative α-integrable function defined on R, and define

$$\mu(A)=\int I_{\{v\in A\}}\beta(v)\upsilon(dv)\,. \tag{3.1}$$

Since μ is a weighted υ which is a gamma random measure, μ is called a weighted gamma random measure with shape α and scale β; we sometimes write μ as $\beta\upsilon$ to stress the dependence of μ on β and υ. Likewise, we define the deterministic measure $\beta\alpha$. The distribution of μ is denoted by $P(d\mu|\alpha,\beta)$, i.e., $P(d\mu|\alpha,\beta)$ is a probability on the measure space $(\Theta,\mathscr{M})$ where $\mathscr{M}$ is the σ-field carried by Θ and generated by the weak convergence of measures. The mean and the Laplace transform of μ are given by Lo (1982) as

$$\int_\Theta \int_R f(v)\mu(dv)P(d\mu|\alpha,\beta) = \int_R f(v)\beta(v)\alpha(dv) \tag{3.2}$$

and

$$\begin{aligned}\int_\Theta \exp\left\{-\int_R f(v)\mu(dv)\right\} P(d\mu|\alpha,\beta) \\ = \exp\left\{-\int_R \log[1+\beta(v)f(v)]\alpha(dv)\right\},\end{aligned} \tag{3.3}$$

where f is any nonnegative function.

First, we generalize (3.3) to an "updating" equality for the Laplace transform of a weighted gamma random measure.

PROPOSITION 3.1. *For nonnegative functions f on $(R,\mathscr{B})$ and g on $(\Theta,\mathscr{M})$,*

$$\begin{aligned}\int_\Theta g(\mu)\exp\left\{-\int_R f(y)\mu(dy)\right\} P(d\mu|\alpha,\beta) \\ = \exp\left\{-\int_R \log[1+\beta(y)f(y)]\alpha(dy)\right\} \\ \cdot\int_\Theta g(\mu)P(d\mu|\alpha,\beta/[1+\beta f]) .\end{aligned} \tag{3.4}$$

PROOF. A change of variable reduces (3.4) to

$$\begin{aligned}\int_\Theta g(\beta\upsilon)\exp\left\{-\int_R f(y)\beta(y)\upsilon(dy)\right\} P(d\upsilon|\alpha,1) \\ = \exp\left\{-\int_R \log[1+\beta(y)f(y)]\alpha(dy)\right\} \\ \cdot\int_\Theta g(\beta\upsilon)P(d\upsilon|\alpha,1/[1+\beta f]) .\end{aligned}$$

Therefore, it suffices to show that for any nonnegative function h on $(R,\mathscr{B})$,

$$(3.5)\quad \int_\Theta g(\upsilon)\exp\left\{-\int_R h(y)\upsilon(dy)\right\}P(d\upsilon\,|\,\alpha,1)$$
$$=\exp\left\{-\int_R \log[1+h(y)]\alpha(dy)\right\}\int_\Theta g(\upsilon)P(d\upsilon\,|\,\alpha,1/[1+h])\,.$$

We only need to prove (3.5) for $g(\upsilon)=\exp\left\{-\int_R k(x)\upsilon(dx)\right\}$ where k is any nonnegative function. In this case, the left side of (3.5) reduces to $\exp\left\{-\int_R \log[1+h(y)+k(y)]\alpha(dy)\right\}$ by (3.3). By the same (3.3), the right side of (3.5) becomes

$$\exp\left\{-\int_R \log[1+h(y)]\alpha(dy)\right\}$$
$$\cdot\exp\left\{-\int_R \log[1+k(y)[1+h(y)]^{-1}]\alpha(dy)\right\},$$

which also reduces to $\exp\left\{-\int_R \log[1+h(y)+k(y)]\alpha(dy)\right\}$. $\square$

The key tool of this paper is the following Fubini-type theorem. Denote a point mass at x by δ_x.

LEMMA 3.1. *Let g be any nonnegative function defined on $(R\times\Theta, \mathscr{B}\otimes\mathscr{M})$, then*

$$(3.6)\quad \int_\Theta\int_R g(\nu,\mu)\mu(d\nu)P(d\mu\,|\,\alpha,\beta)=\int_R\int_\Theta g(\nu,\mu)P(d\mu\,|\,\alpha+\delta_\nu,\beta)\beta\alpha(d\nu)\,.$$

PROOF. A proof of this lemma can be obtained by letting g be an indicator function, and then evaluating both sides of (3.6); see Lo (1978). However, the Laplace transform argument in Lo (1982) suggests the following streamlined proof. It suffices to prove (3.6) for the g's of the form

$$g(\nu,\mu)=\exp\{-\lambda\nu\}\exp\left\{-\int_R f(y)\mu(dy)\right\},$$

where λ is any nonnegative real number and f is any nonnegative function.

For this g, we apply Proposition 3.1, and then (3.2) to the left side of (3.6) to obtain

$$\int_\Theta\int_R \exp\{-\lambda\nu\}\mu(d\nu)\exp\left\{-\int_R f(y)\mu(dy)\right\}P(d\mu\,|\,\alpha,\beta)$$

$$= \exp\left\{-\int_R \log[1+\beta(y)f(y)]\alpha(dy)\right\}$$

$$\cdot \int_\Theta \int_R \exp\{-\lambda v\}\mu(dv)P(d\mu|\alpha, \beta/[1+\beta f])$$

$$= \exp\left\{-\int_R \log[1+\beta(y)f(y)]\alpha(dy)\right\}$$

$$\cdot \int_R \exp\{-\lambda v\}\beta(v)[1+\beta(v)f(v)]^{-1}\alpha(dv) .$$

On the other hand, an application of (3.3) to the inner integral of the right side of (3.6) yields

$$\int_R \exp\{-\lambda v\}\int_\Theta \exp\left\{-\int_R f(y)\mu(dy)\right\}P(d\mu|\alpha+\delta_v, \beta)\beta(v)\alpha(dv)$$

$$= \int_R \exp\{-\lambda v\}\exp\left\{-\int_R \log[1+\beta(y)f(y)](\alpha+\delta v)(dy)\right\}\beta(v)\alpha(dv)$$

$$= \exp\left\{-\int_R \log[1+\beta(y)f(y)]\right\}\alpha(dy)$$

$$\cdot \int_R \exp\{-\lambda v\}[1+\beta(v)f(v)]^{-1}\beta(v)\alpha(dv) .$$

Hence, the two sides of (3.6) are equal for these g's. □

The rest of this section provides applications of Lemma 3.1 in different contexts, and is independent of the rest of the paper.

3.1 *Discreteness of weighted gamma random measures*

The first application of the lemma is to show that the weighted gamma random measure is discrete with probability one.

COROLLARY 3.1. *$P(\mu\colon \mu$ is discrete$|\alpha, \beta) = 1$.*

PROOF. It suffices to prove the result for $\beta = 1$. A measure μ is discrete if and only if $\mu\{v\colon \mu\{v\} = 0\} = 0$ (or equivalently, $\mu\{v\colon \mu\{v\} > 0\} = \mu(R) < \infty$); hence, it is enough to show that $\int_\Theta \mu\{v; \mu\{v\} = 0\}P(d\mu|\alpha, 1) = 0$. Notice that

$$\int_\Theta \mu\{v\colon \mu\{v\} = 0\}P(d\mu|\alpha, 1) = \int_\Theta \int_R I_{\{\mu\{v\}=0\}}\mu(dv)P(d\mu|\alpha, 1)$$

$$= \int_R \int_\Theta I_{\{\mu\{v\}=0\}}P(d\mu|\alpha+\delta_v, 1)\alpha(dv) ,$$

by Lemma 3.1, and the fact that the integrand is a measurable function of μ and v. Consider the inner integral which equals $P(\mu: \mu\{v\} = 0 \mid \alpha + \delta_v, 1)$. For a fixed v, this probability is zero, since $\mu\{v\}$ is a gamma $(\alpha\{v\} + 1, 1)$ random variable according to $P(d\mu \mid \alpha + \delta_v, 1)$. □

Remark 3.1. The above argument is essentially due to Berk and Savage (1979). Alternatively, Kingman ((1975), p. 15) pointed out that one can modify the arguments of Blackwell (1973) to prove Corollary 3.1.

3.2 *On Ferguson's theorem*

The next application of Lemma 3.1 concerns a theorem of Ferguson ((1973), Theorem 1 in Section 3). Ferguson's theorem is known to be equivalent to Lemma 1 in Part I (Lo (1984)). The following corollary then essentially provides a proof of Ferguson's theorem using the Laplace transform.

COROLLARY 3.2. *Lemma* 3.1 *implies Lemma* 1 *of Part* I.

PROOF. First note that a Dirichlet random probability G is equal in distribution to $\mu/\mu(R)$ where μ is a gamma random measure (i.e., $\beta = 1$). Hence, Lemma 1 in Part I (Lo (1984)) is equivalent to

$$\begin{aligned}(3.7)\qquad &\int_\Theta \int_R g(v, \mu/\mu(R))[\mu(dv)/\mu(R)]P(d\mu \mid \alpha, 1)\\ &= \int_R \int_\Theta g(v, \mu/\mu(R))P(d\mu \mid \alpha + \delta_v, 1)\{\alpha(dv)/\alpha(R)\}\,.\end{aligned}$$

An application of Lemma 3.1 to the left side of (3.7) yields

$$(3.8)\qquad \int_R \int_\Theta g(v, \mu/\mu(R))[1/\mu(R)]P(d\mu \mid \alpha + \delta_v, 1)\alpha(dv)\,.$$

Since μ is a gamma random measure, $\mu(R)$ and $\mu/\mu(R)$ are independent (this follows from a simple extension of Theorem 1.2.3 in Bickel and Doksum (1977)). Hence, (3.8) reduces to

$$\int_R \int_\Theta g(v, \mu/\mu(R))P(d\mu \mid \alpha + \delta_v, 1)\{\alpha(dv)/\alpha(R)\}\,,$$

which is the right side of (3.7). □

3.3 *Bayes estimation of a Poisson cumulative intensity*

We conclude this section with an application of Lemma 3.1 to the problem of the Bayes estimation of a Poisson cumulative intensity μ (Lo

(1982)). Suppose $\mu \sim P(d\mu | \alpha, \beta)$ and given μ, $N_1, \ldots, N_n$ is an i.i.d. sample from a Poisson process with cumulative intensity μ. Then the posterior distribution of μ (Lo (1982)) given $N_1, \ldots, N_n$ is

$$P\left(d\mu \,\middle|\, \alpha + \sum_{1 \le i \le n} N_i, \beta/[1+n\beta] \right). \tag{3.9}$$

Suppose we like to choose a μ^* to minimize a risk $E[L(\mu,\mu^*)]$, where $L(\mu,\mu^*)$ is an integrable loss function, the Bayes theorem states that for a given sample $N_1, \ldots, N_n$, we should choose μ^* to minimize

$$\int_\Theta L(\mu,\mu^*) P\left(d\mu \,\middle|\, \alpha + \sum_{1 \le i \le n} N_i, \beta/[1+n\beta] \right). \tag{3.10}$$

Lo (1982) considered the loss function $L(\mu,\mu^*) = \int_R [\mu(y) - \mu^*(y)]^2 W(dy)$, where W is a given weight function, and obtained

$$\begin{aligned} \mu^*(t) &= \int_\Theta \mu(t) P\left(d\mu \,\middle|\, \alpha + \sum_{1 \le i \le n} N_i, \beta/[1+n\beta] \right) \\ &= \int I_{\{s \le t\}} \beta(s)[1+n\beta(s)]^{-1} \alpha(ds) \\ &\quad + \sum_{1 \le i \le n} \int I_{\{s \le t\}} \beta(s)[1+n\beta(s)]^{-1} N_i(ds). \end{aligned} \tag{3.11}$$

Perhaps a better choice of the loss is $L(\mu,\mu^*) = \int_R [\mu(y) - \mu^*(y)]^2 \mu(dy)$, which does not depend on an extraneous weight function. For this loss, the problem becomes choosing μ^* to minimize

$$\int_\Theta \int_R [\mu(t) - \mu^*(t)]^2 \mu(dt) P\left(d\mu \,\middle|\, \alpha + \sum_{1 \le i \le n} N_i, \beta/[1+n\beta] \right). \tag{3.12}$$

At first sight, this seems to be a formidable problem. However, an application of Lemma 3.1 reduces (3.12) to

$$\int_R \int_\Theta [\mu(t) - \mu^*(t)]^2 P\left(d\mu \,\middle|\, \alpha + \sum_{1 \le i \le n} N_i + \delta_t, \beta/[1+n\beta] \right) \alpha(dt). \tag{3.13}$$

Hence, the minimization is achieved by, for each t,

$$\begin{aligned} \mu^*(t) &= \int_\Theta \mu(t) P\left(d\mu \,\middle|\, \alpha + \sum_{1 \le i \le n} N_i + \delta_t, \beta/[1+n\beta] \right) \\ &= \int I_{\{s \le t\}} \beta(s)[1+n\beta(s)]^{-1} \alpha(ds) + \beta(t)/[1+n\beta(t)] \end{aligned} \tag{3.14}$$

$$+ \sum_{1 \le i \le n} \int I_{\{s \le t\}} \beta(s)[1 + n\beta(s)]^{-1} N_i(ds) ,$$

i.e., the new loss function results in an additional term $\beta(t)/[1 + n\beta(t)]$.

4. Posterior distributions

For the priors on the μ_j's, we choose $\alpha_j \in \Theta$, β_j to be nonnegative and α_j-integrable, and let $\mu_j \sim P(d\mu_j | \alpha_j, \beta_j)$, $j = 1,\ldots,q$. The μ_j's are assumed to be independent. The posterior distribution P_N of $\boldsymbol{\mu} = (\mu_1,\ldots,\mu_k)$ is defined by

$$\int_{\Theta^k} \left[\prod_j g_j(\mu_j) \right] P_N(d\mu) = \prod_j \frac{\int_\Theta g_j(\mu_j) L_j(\mu_j) P(d\mu_j | \alpha_j, \beta_j)}{\int_\Theta L_j(\mu_j) P(d\mu_j | \alpha_j, \beta_j)} , \tag{4.1}$$

where g_j's are nonnegative functions and j runs from 1 to q. To evaluate the posterior distribution $P_N(d\mu)$, we need only to evaluate (4.1) in the case $q = 1$ because the q factors are independent and can be handled separately. Suppressing the subscript j, (4.1) becomes

$$\int_\Theta g(\mu) P_N(d\mu) = \frac{\int_\Theta g(\mu) L(\mu) P(d\mu | \alpha, \beta)}{\int_\Theta L(\mu) P(d\mu | \alpha, \beta)} , \tag{4.2}$$

where g is a nonnegative function and

$$L(\mu) = \left[\prod_{1 \le i \le n} \int_R k(T_i, v) \mu(dv) \right] \exp \left\{ - \int_R \int I_{\{0 \le s \le 1\}} Y(s) k(s, v) ds \mu(dv) \right\} . \tag{4.3}$$

The following theorem characterizes the posterior distribution of μ. First, some notation. Let

$$\begin{aligned} \boldsymbol{v} &= (v_1,\ldots,v_n) , \\ \beta^*(v) &= \beta(v) \Big/ \left[1 + \beta(v) \int I_{\{0 \le s \le 1\}} Y(s) k(s, v) ds \right] , \\ k^*(t, v) &= \beta^*(v) k(t, v) , \end{aligned} \tag{4.4}$$

and

$$\mu_{n, k^*, \alpha}(C) = \int_C \left[\prod_{1 \le i \le n} k^*(T_i, v_i) \right] \prod_{1 \le i \le n} \alpha_{i-1}(dv_i) ,$$

where $C \in B^n$, $\alpha_0 = \alpha$, and $\alpha_{i-1}(dv_i) = \left(\alpha + \sum_{1\le j\le i-1} \delta_{v_j}\right)(dv_i)$ for $i \ge 2$.

THEOREM 4.1. *Assume that the likelihood function of a point process is proportional to* (4.3) *and the prior distribution of μ is $P(d\mu\,|\,\alpha,\beta)$. The posterior distribution* (4.2) *is given by*

$$\int_\Theta g(\mu)P_N(d\mu) = \frac{\int_{R^n}\int_\Theta g(\mu)P(d\mu\,|\,\alpha_n,\beta^*)\mu_{n,k^*,\alpha}(dv)}{\mu_{n,k^*,\alpha}(R^n)}. \tag{4.5}$$

PROOF. Apply Proposition 3.1 once and Lemma 3.1 n times to the numerator and denominator of (4.2) and then simplify. □

Theorem 4.1 defines the posterior distribution which summarizes the posterior information about the parameter μ and can be used to compute the Bayes estimate of μ. However, the mixture form of the posterior distribution (4.5) is quite complicated; hence, descriptions of some posterior quantities are necessary. In the next theorem, we illustrate the application of (4.5) to find the posterior mean of the hazard rate $r(t\,|\,\mu)$; the evaluation of the posterior k-th moment of $r(t\,|\,\mu)$ is similar and will not be given.

THEOREM 4.2. *For each t,*

$$E_N[r(t\,|\,\mu)] = \int_R k^*(t,v)\alpha(dv) + \sum_{\boldsymbol p} W(\boldsymbol p) \sum_{1\le i\le N(\boldsymbol p)} e_i r_i(t\,|\,\boldsymbol p), \tag{4.6}$$

where $W(\boldsymbol p) = \varphi(\boldsymbol p)\Big/\sum_{\boldsymbol p}\varphi(\boldsymbol p)$, $\varphi(\boldsymbol p) = \prod_{1\le i\le N(\boldsymbol p)}\left[(e_i-1)!\int_R \prod_{j\in C(i)} k^*(T_j,v)\alpha(dv)\right]$,

$$r_i(t\,|\,\boldsymbol p) = \frac{\int_R k^*(t,v)\prod_{j\in C(i)} k^*(T_j,v)\alpha(dv)}{\int_R \prod_{j\in C(i)} k^*(T_j,v)\alpha(dv)}; \tag{4.7}$$

$\boldsymbol p$ is a partition of $\{1,\dots,n\}$, $\{C(i): i = 1,\dots,N(\boldsymbol p)\}$ are the cells (subsamples) of the partition $\boldsymbol p$, and e_i is the number of elements in $C(i)$.

PROOF. Put $g(\mu) = r(t\,|\,\mu) = \int_R k(t,y)\mu(dy)$ in Theorem 4.1 to obtain

$$E_N[r(t\,|\,\mu)] = \frac{\int_{R^n}\int_\Theta\int k(t,y)\mu(dy)P(d\mu\,|\,\alpha_n,\beta^*)\mu_{n,k^*,\alpha}(dv)}{\int_{R^n}\mu_{n,k^*,\alpha}(dv)} \tag{4.8}$$

$$= \frac{\int_{R^n}\int_\Theta \int k^*(t,y)\upsilon(dy)P(d\upsilon|\alpha_n,1)\mu_{n,k^*,\alpha}(dv)}{\int_{R^n}\mu_{n,k^*,\alpha}(dv)},$$

where the last equality follows from a change of variables. Apply Lemma 3.1 to the numerator once and write t as T_{n+1} to obtain

$$E_N[r(t|\mu)] = \frac{\int_{R^{n+1}}\mu_{n+1,k^*,\alpha}(dv)}{\int_{R^n}\mu_{n,k^*,\alpha}(dv)}. \tag{4.9}$$

This is precisely the last equation on p. 353 in Part I (Lo (1984)) without the coefficient $[\alpha(R)+n]^{-1}$. The rest of the derivation follows that of Theorem 2 in Part I and will not be reproduced. □

We conclude this section by relating these results to those obtained by Dykstra and Laud (1981). Let $k(t,v) = I_{\{v\le t\}}$. Theorem 4.1 specializes to their Theorem 3.3 (Lemma 2 in Part I also specializes to their Theorem 5.1; see Lemma 2.1 in Brunner and Lo (1989)). Furthermore, expression (4.6) reduces to

$$E_N[r(t|\mu)] = \int I_{\{0\le v\le t\}}\beta^*(v)\alpha(dv) + \sum_{\boldsymbol{m}} W(\boldsymbol{m}) \sum_{1\le i\le n} m_i \times \frac{\int I_{\{0\le v\le T(i)\}}\beta^*(v)^{m_i+1}I_{\{0\le v\le t\}}\alpha(dv)}{\int I_{\{0\le v\le T(i)\}}\beta^*(v)^{m_i}\alpha(dv)}; \tag{4.10}$$

where $W(\boldsymbol{m}) = \phi(\boldsymbol{m})\Big/\Big[\sum_{\boldsymbol{m}}\phi(\boldsymbol{m})\Big]$, $\phi(\boldsymbol{m}) = k(\boldsymbol{m})\prod_i \int I_{\{0\le v\le T(i)\}}\beta^*(v)^{m_i}\alpha(dv)$; $\boldsymbol{m} = (m_1,\dots,m_n)$, m_i are nonnegative integers such that $s_j = \sum_{1\le i\le j} m_i \le j$ for each $j = 1,\dots,n-1$, and $s_n = n$; $k(\boldsymbol{m}) = \prod_{i^*}(i-1-s_{i-1})!/(i-s_i)!$; the product $\prod_{i^*}$ is over the set of i such that $m_i \ge 1$; $0 < T(n) < \cdots < T(1) < \infty$ are the ordered statistics of the T's.

Expression (4.10) is slightly simpler than (4.6). However, we must point out that such reduction is possible only because the kernel k takes on the simplest form. In general, (4.6) should be used.

5. The choice of k, α and β

In this section, we discuss the choices of the kernel k and prior parameters α and β. It suffices to consider the case that $q = 1$. As the first example for the choice of the kernel k, we let

(5.1) $$k(t, v) = \sum_{1 \le j \le q} I_{\{t \in \Delta_j, v \in \Delta_j\}} ,$$

where $\{\Delta_j : j = 1, \ldots, q\}$ is a partition of the time interval $[0, 1]$. This gives us a piecewise constant hazard rate model. In fact, $r(t|\mu) = \mu(\Delta_j)$ if $t \in \Delta_j$, and expression (4.6) reduces to, for each j and $t \in \Delta_j$,

(5.2) $$E_N[r(t|\mu)] = \frac{[\alpha(\Delta_j) + n_j]\beta(t)}{1 + \beta(t)\int I_{\{s \in \Delta_j\}} Y(s)ds} ,$$

where n_j is the number of observations in Δ_j.

Let $\alpha(\Delta_j) \to 0$ and $\beta(t) \to 0$, (5.2) becomes, for each j and $t \in \Delta_j$,

(5.3) $$E_N[r(t|\mu)] = n_j \Big/ \int I_{\{s \in \Delta_j\}} Y(s)ds ,$$

which is the maximum likelihood estimate of the constant hazard rate (see p. 54 in Cox and Oakes (1984)).

The other determinations of k are more subtle. In general, the family $\{k(\cdot, v) : v \in R\}$ is the collection of extreme points of $L_k = \{r(\cdot|\mu) : \mu \in \Theta\}$ in the sense that any $r \in L_k$ can be represented as a mixture of $k(\cdot, v)$, $v \in R$ for some mixing measure $\mu \in \Theta$. In the following, we list some cases where such integral representations exist.

(5.4) $$k(t, v) = I_{\{t \le v\}}, \qquad v \in R \quad (= [0, \infty))$$

implies L_k is the family of nondecreasing hazard rates on $[0, 1]$. This is due to a form of the Khintchine-Shepp theorem (p. 158 in Feller (1971)) for a finite measure instead of a probability (similarly $k(t, v) = I_{\{v \le t\}}$ gives a family of nonincreasing hazard rates).

(5.5) $$k(t, v) = e^{-tv}, \qquad v \in R ,$$

here L_k is the family of completely monotone hazard rates. It is also a very smooth subfamily of decreasing hazard rates.

(5.6) $$k(t, v) = I_{\{|t-a| \ge v\}}, \qquad v \in [0, \infty) ;$$

the rates obtained are increasing to the right of a and decreasing to the left of a and hence, L_k is the family of symmetric U-shaped hazard rates with the minimum at a (reversing the inequality sign in (5.6) gives the family of unimodal and symmetric rates).

In case no determination of k is available, one may use

$$r(t|\mu) = \iint I_{\{0<\tau<\infty, 0<v<\infty\}} \tau k(\tau(t-v))\mu(dv, d\tau), \qquad t \in [0,1], \tag{5.7}$$

for a prescribed kernel k such that $k(t) = 0$ for $t \leq 0$ and for each finite constant C, $\int I_{\{-\infty<t\leq C\}} k(t)\,dt < \infty$. Denote the collection of hazard rates that can be represented as (5.7) for some μ supported by $[0,\infty) \times [0,\infty)$ by L_k. It can be shown that the L^1-closure of L_k contains all the $r(t)$'s, such that $\int I_{\{0<s<1\}} r(s)\,ds < \infty$.

The choice of β and α depends on k. An examination of (4.6) and (4.7) indicates that an idea analogous to conjugate priors also prevails. In particular, it is convenient to regard $\beta(v)k(\cdot, v)$ in (4.7) as densities with v being the "parameter" and choose $\alpha(dv)$ to be a conjugate "prior" when sampling from the parametric family $\{\beta(v)k(\cdot, v), v \in R\}$.

6. The numerical evaluation of some posterior quantities

The number of summands for the dominating term $\sum_{p}$ in (4.6) is equal to the Bells number B_n where n is the sample size. The Bells number increases roughly as $n!$, and the exact evaluation of the estimates (4.6) is formidable for sample sizes larger than twelve (the evaluation of the monotone rate estimate (4.10) is only slightly simpler; for a detailed discussion, see Brunner and Lo (1989)). This section discusses Monte Carlo approximations to the posterior mean (4.6).

A careful inspection of $\sum_{p}$ in (4.6) indicates that it is actually

$$\sum_{p} q(p)N(t|p) \Big/ \sum_{p} q(p)D(p), \tag{6.1}$$

where

$$q(p) = \alpha(R)^{N(p)} \left[\prod_i (e_i - 1)! \right] \Gamma(\alpha(R))/\Gamma(\alpha(R)+n), \tag{6.2}$$

$$N(t|p) = \left\{ \prod_i \left[\int \prod_{j \in C(i)} k^*(T_j, v)\,\alpha(dv)/\alpha(R) \right] \right\} \sum_i r_i(t|p), \tag{6.3}$$

$$D(p) = \prod_i \left[\int \prod_{j \in C(i)} k^*(T_j, v)\,\alpha(dv)/\alpha(R) \right]. \tag{6.4}$$

The weights $\{q(p)\}$ are probability weights on the collection of partitions of the set $\{1,\ldots,n\}$ (for a proof, put $g_i = 1$ in Lemma 2 in Lo (1984)), suggesting a more efficient method by simulating a partition having distribution $q(\cdot)$.

A random partition having distribution $q(\cdot)$ can be simulated based on the Chinese restaurant process with parameter $\alpha(R)$ (Aldous (1985), p. 92, see also Kuo (1986)). Imagine n persons arriving sequentially at an initially empty restaurant with a large number of unoccupied tables. Person j either sits at an empty table with probability $\alpha(R)/(\alpha(R)+j-1)$, or else sits at an occupied table with probability proportional to the number of occupants at that table. The resulting configuration of the occupied tables is a random partition with distribution $q(\cdot)$.

Repeat a Chinese restaurant process with parameter $\alpha(R)$ a total of M times to obtain random partitions $\boldsymbol{p}_1, \boldsymbol{p}_2, \ldots, \boldsymbol{p}_M$. The Monte Carlo approximation to (6.1) is then given by the ratio

$$\sum_{1\le k\le M} N(t\,|\,\boldsymbol{p}_k) \Big/ \sum_{1\le k\le M} D(\boldsymbol{p}_k)\;; \tag{6.5}$$

it is well known that the standard error of this approximation is $O_P(M^{-1/2})$ (Rubenstein (1981)).

Remark 6.1. This Monte Carlo method provides an approximation of quantities of the form $\sum_{\boldsymbol{p}'} g_0(\boldsymbol{p}') \Big/ \sum_{\boldsymbol{p}} g(\boldsymbol{p})$ where $\boldsymbol{p}'$ is a partition of $\{1,\ldots,n,n+1,\ldots,n+m\}$ and g_0 and g are given functions of partitions, since one can approximate the numerator and the denominator separately. Note that the proofs in Theorems 4.1 and 4.2 indicate that the dominating term of the posterior $(m+1)$-th moment of any linear function of μ is of the form $\sum_{\boldsymbol{p}'} g_0(\boldsymbol{p}') \Big/ \sum_{\boldsymbol{p}} g(\boldsymbol{p})$. In particular, the posterior variance of $r(t\,|\,\mu)$ can be approximated, since it corresponds to $m = 1$. It is clear that this method can also be used to approximate the posterior moments of (any linear function of) the mixing distribution G appearing in Part I (Lo (1984)), since the dominating terms for these are also of the forms $\sum_{\boldsymbol{p}'} g_0(\boldsymbol{p}') \Big/ \sum_{\boldsymbol{p}} g(\boldsymbol{p})$.

Example 6.1 The following Table 1 (Aalen (1978)) shows times to copulation of *Drosophila*. The model likelihood function is given by (2.6) with $k = 1$ and $Y(t) = M(t) \times F(t)$, where $M(t)$ and $F(t)$ are the numbers of male and female flies that have not been involved in any copulation up to

Table 1. Times in seconds at initiations of mating (Aalen (1978)).

Ebony flies:	143, 180, 184, 303, 380, 431, 455, 475, 500, 514, 521, 552, 558, 606, 650, 667, 683, 782, 799, 849, 901, 995, 1131, 1216, 1591, 1702, 2212.
Oregon flies:	555, 742, 746, 795, 934, 967, 982, 1043, 1055, 1067, 1081, 1296, 1353, 1361, 1462, 1731, 1985, 2051, 2292, 2335, 2514, 2570, 2970.

time t (each fly only mates once). Two experiments are carried out: one for Ebony flies [$M(0) = 40$, $F(0) = 30$], and one for Oregon flies [$M(0) = 39$, $F(0) = 29$].

A very smooth decreasing hazard rate mixture model is assumed for the data; this is achieved by choosing an exponential kernel $k(t, v) = v \times \exp\{-vt\}$. Gamma (a, b) (with mean a/b) distributions are employed for the shape probability $\alpha(\cdot)/\alpha(R)$. Figure 1 displays the posterior mean (4.6) where the dominating term $\sum_p$ is approximated by (6.5); a gamma (1, 5) distribution is assumed for $\alpha(\cdot)/\alpha(R)$. Figure 2 displays the change of posterior means (based on the Ebony flies data) for different gamma (a, b) shape probabilities.

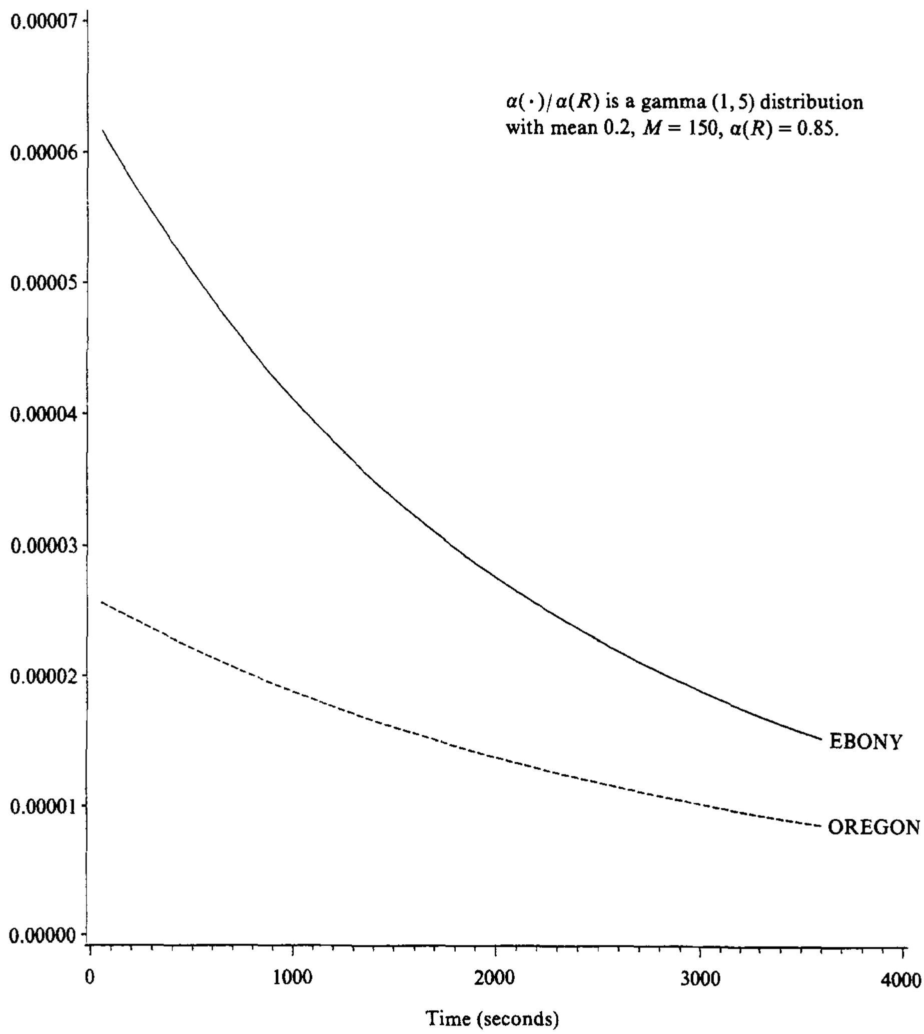

Fig. 1. Posterior means of hazard rates.

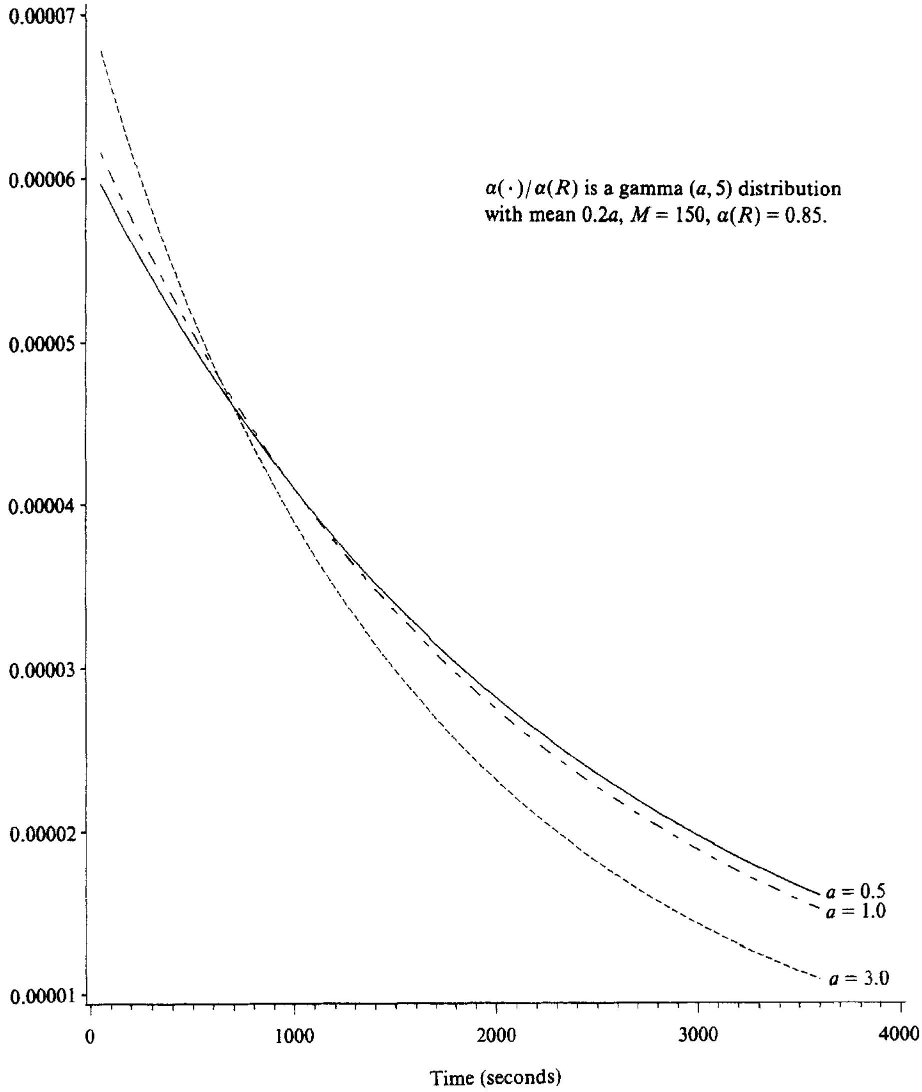

Fig. 2. Posterior means of hazard rates.

Acknowledgement

The authors are grateful to a referee for a meticulous proofreading as well as numerous helpful suggestions.

REFERENCES

Aalen, O. (1976). Nonparametric inference in connection with multiple decrement models, *Scand. J. Statist.*, 3, 15–27.

Aalen, O. (1978). Nonparametric inference for a family of counting processes, *Ann. Statist.*, **6**, 701–726.

Aldous, D. J. (1985). *Exchangeability and Related Topics*, Lecture Notes in Mathematics, Vol. 1117, Springer, Berlin-New York.

Berk, R. and Savage, I. R. (1979). Dirichlet processes produce discrete measure: an elementary proof, *Contributions to Statistics, Jaraslar Hajek Memorial Volume*, Academia, Prague.

Bickel, P. and Doksum, K. A. (1977). *Mathematical Statistics*, Holden-Day, Inc.

Blackwell, D. (1973). Discreteness of Ferguson selection, *Ann. Statist.*, **1**, 356–358.

Brown, M. (1972). Statistical analysis of non-homogeneous Poisson point processes, *Stochastic Point Processes*, (ed. P. A. W. Lewis), Wiley, New York.

Brunner, L. J. and Lo, A. Y. (1989). Bayes methods for a symmetric and unimodal density and its mode, *Ann. Statist.*, **17**, No. 4.

Chiang, C. L. (1968). *Introduction to Stochastic Processes in Biostatistics*, Wiley, New York.

Clevenson, M. L. and Zidek, J. W. (1977). Bayes linear estimators of the intensity function of the nonstationary Poisson process, *J. Amer. Statist. Assoc.*, **72**, 112–120.

Cox, D. R. (1972). Regression models and life tables, *J. Roy. Statist. Soc. Ser. B*, **34**, 187–220.

Cox, D. R. and Oakes, D. (1984). *Analysis of Survival Data*, Chapman & Hall, London-New York.

Dykstra, R. L. and Laud, P. (1981). A Bayesian nonparametric approach to reliability, *Ann. Statist.*, **9**, 356–367.

Feller, W. (1971). *An Introduction to Probability Theory and Its Applications*, Vol. II, Wiley, New York.

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

Grandell, J. (1972). Statistical inference for doubly stochastic Poisson point process, *Stochastic Point Processes*, (ed. P. A. W. Lewis), Wiley, New York.

Hoem, J. M. (1971). Point estimation of forces of transition in demographic models, *J. Roy. Statist. Soc. Ser. B*, **3**, 275–289.

Jacod, J. (1975). Multivariate point process: Predictable projection, Radon-Nikodym derivatives, representation of martingales, *Z. Wahrsch. Verw. Gebiete*, **31**, 235–253.

Kingman, J. F. K. (1975). Random discrete distributions, *J. Roy. Statist. Soc. Ser. B*, **37**, 1–22.

Kunita, H. and Watanabe, S. (1967). On square integrable martingales, *Nagoya Math. J.*, **30**, 209–245.

Kuo, L. (1986). Computations of mixtures of Dirichlet processes, *SIAM J. Sci. Statist. Comput.*, **7**, 60–71.

Lewis, P. A. W. (1972). Recent results in the statistical analysis of univariate point processes, *Stochastic Point Processes*, (ed. P. A. W. Lewis), Wiley, New York.

Lo, A. Y. (1978). Bayesian nonparametric rate function methods, Tech. Report, Dept. of Math. & Statist., Univ. of Pittsburgh, Pennsylvania.

Lo, A. Y. (1982). Bayesian nonparametric statistical inference for Poisson point process, *Z. Wahrsch. Verw. Gebiete*, **59**, 55–66.

Lo, A. Y. (1984). On a class of Bayesian nonparametric estimates: I. Density estimates, *Ann. Statist.*, **12**, 351–357.

Meyer, P. A. (1966). *Probability and Potentials*, Blaisdell, Waltham, Massachusetts.

Rubenstein, R. Y. (1981). *Simulation and the Monte Carlo Method*, Wiley, New York.