*The Annals of Statistics*
1990, Vol. 18, No. 2, 603–616

# BAYES ESTIMATION FROM A MARKOV RENEWAL PROCESS

BY MICHAEL J. PHELAN

*Princeton University*

A procedure for Bayes nonparametric estimation from a Markov renewal process is developed. It is based on a conjugate class of a priori distributions on the parameter space of semi-Markov transition distributions. The class is characterized by a Dirichlet family of distributions for random Markov matrices and a Beta family of Lévy processes for random cumulative hazard functions. The main result is the derivation of the posterior law from an observation of the Markov renewal process over a period of time.

**1. Introduction.** This paper develops a procedure for Bayes estimation of the transition distributions from a Markov renewal process. The prior consists of a family of Dirichlet distributions on the space of Markov transition matrices and a Beta family of Lévy processes having sample paths in the space of cumulative hazard functions. The main result is the derivation of the posterior from an observation of the Markov renewal process over a period of time. The result shows that the chosen prior family is a conjugate family for this problem, and it extends the Bayes life-testing estimation procedure proposed by Hjort (1984, 1988) to the present context. Hjort (1984, 1988) conjectures that his procedure extends to the *subfamily* of hierarchical semi-Markov processes considered by Voelkel and Crowley (1984). Our results show this conjecture to be true, but we find the restriction to a hierarchical process unnecessary.

Our approach is based on the parametrization of the transition distributions in terms of transition probabilities of a Markov chain and cumulative hazard functions of life distributions. The estimation procedure is developed by concatenating a procedure proposed by Hjort (1984) for estimating Markov transition probabilities with that proposed for estimating cumulative hazards in life-testing. The result is a considerable generalization of the result in Brock (1973), who only considers Bayes estimation for the Markov transition probabilities. Moreover, our approach can incorporate a model for random right-censorship, so that the Bayes estimators of transition distributions provide alternatives to the nonparametric estimators proposed by Gill (1980a) and Phelan (1988).

The paper is organized as follows. The next section recalls some elementary facts about Markov renewal processes. In particular, we define the sample space and a probability measure for sample functions from the process ob-

Received April 1988; revised June 1989.

*AMS* 1980 *subject classifications*. Primary 62M02; secondary 62C99, 62G05.

*Key words and phrases*. Markov renewal process, Bayes nonparametric estimation, Beta processes, Dirichlet.

served over an interval of time, namely $[0, t]$. Most important is the identity given in equation (2.4) that specifies the measure in terms of a set of Markov transition probabilities and a set of cumulative hazard functions. This is followed by the specification of the a priori distribution over the parameter space of Markov matrices and cumulative hazards from the family of *Dirichlet times Beta process* priors using Definitions 2.1 and 2.2. In Section 3 we specify a model for Bayes nonparametric estimation from a Markov renewal process. This is followed immediately by the statement of the main result in Theorem 3.1 together with the rules for updating the a priori parameters. The remainder of the section is devoted to the proof of Theorem 3.1 and three propositions needed in that proof. Finally, in Section 4 we calculate the Bayes estimators under squared-error loss, make some closing remarks, and discuss generalizations of our results.

## 2. Markov renewal processes and the prior.

2.1. *Markov renewal processes.* Here we present some facts which follow from the constructive definition of a Markov renewal process given by Pyke (1961). Let $(J, S) = (J_n, S_n)$, $n \geq 0$, denote a Markov renewal process with state space $E \times R_+$ and transition distributions $\{Q_{ij}(t);\, i, j \in E, t \geq 0\}$, where $E = \{1, 2, \dots, m\}$ $m \geq 1$. For each $n$ we have

$$(2.1)\quad Q_{ij}(t) = P(J_{n+1} = j,\, S_{n+1} - S_n \leq t | J_n = i), \qquad i, j \in E, t \geq 0.$$

The process $J = (J_n)$, $n \geq 0$, is a Markov chain with state space $E$ and transition probabilities given by $p_{ij} = Q_{ij}(\infty)$, $i, j \in E$.

We assume that the process is stationary, so that the left-hand side of (2.1) is independent of $n$ and the duration the process has been active. As an example in medical clinical trials, let $E$ denote the collection of the patient's state during treatment, such as initial illness, in remission of symptoms, progression and relapse. The patient changes his status according to a Markov chain, the $n$th transition occurring at time $S_n$. The length of stay for a patient in a given state depends only on that state and the state to which he moves, and the patient sojourns between a pair of states $i$ and $j$ according to the distribution $Q_{ij}$.

According to Pyke and Schaufele (1964), we incur no loss of generality to assume that $Q_{ij} = p_{ij}\phi_i$, where $\phi_i$ denotes a distribution on $R_+$. For each $i \in E$, recall that the distribution $\phi_i$ is determined uniquely by its cumulative hazard function $b_i$ defined by

$$(2.2)\qquad b_i(t) = \int_{(0,t]} \frac{d\phi_i(s)}{1 - \phi_i(s-)}, \qquad t \geq 0,$$

where $\phi_i(s-) = \lim_{u \uparrow s} \phi_i(u)$. We will use an important representation of the transition distribution $Q_{ij}$ that is given by

$$(2.3)\qquad Q_{ij}(t) = p_{ij}\phi_i(t) = p_{ij}\Big(1 - \prod_{0 \leq s \leq t} (1 - db_i(s))\Big), \qquad t \geq 0,$$

where the above product-integral of $b_i$ is treated for example in Gill (1980b).

For each $n \geq 0$, let $\Pi_n = (E \times R_+)^{n+1}$ and let $\mathbf{\Pi}_n$ denote the Borel field on $\Pi_n$ generated by all subsets of $E$ and the Borel sets in $R_+$. Also let $\mathbb{N} = \{0, 1, 2, \ldots, \}$ and let $\mathbf{N}$ denote all subsets of $\mathbb{N}$. The sample space for a Markov renewal process observed over a period of time is given by the product space $(\mathbb{N} \times \Pi, \mathbf{N} \otimes \mathbf{\Pi})$, where $\Pi = \bigcup_{n=0}^{\infty} \Pi_n$ and $\mathbf{\Pi} = \bigvee_{n=0}^{\infty} \mathbf{\Pi}_n$. In particular, let $N = (N(t))$, $t \geq 0$, denote the Markov renewal counting process defined by

$$N(t) = \sum_{n=1}^{\infty} 1(S_n \leq t), \qquad t \geq 0, \tag{2.4}$$

and let $X_{n+1} = S_{n+1} - S_n$, $n \geq 0$, where $S_0 = 0$. Suppose the Markov renewal process is observed over $[0, t]$, $t > 0$. Then, according to Moore and Pyke (1968), almost all sample functions can be represented by a point in the sample space $(\mathbb{N} \times \Pi, \mathbf{N} \otimes \mathbf{\Pi})$ given by the finite tuple $(N(t), R(t))$, where $R(t) = (J_0, \ldots, J_{N(t)}, X_1, \ldots, X_{N(t)}, t - S_{N(t)})$. A probability measure $P$ is given on this space as follows. For any $n > 0$ and $G \in \mathbf{\Pi}_n$ we define

$$P(N(t) = n, R(t) \in G) = \int_G dP(n; \pi),$$

where, using (2.1), (2.2) and (2.3) we have

$$\begin{aligned} dP(n;\pi) &= p_{j_0}\big(1 - \phi_{j_n}(u_t)\big) \prod_{k=0}^{n-1} p_{j_k j_{k+1}}\, d\phi_{j_k}(x_{k+1}) \\ &= p_{j_0} \prod_{0 < s \leq u_t} \big(1 - db_{j_n}(s)\big) \prod_{k=0}^{n-1} p_{j_k j_{k+1}} \prod_{0 < s < x_{k+1}} \big(1 - db_{j_k}(s)\big) db_{j_k}(x_{k+1}) \end{aligned} \tag{2.5}$$

for $\pi = (j_0, j_1, \ldots, j_n, x_1, \ldots, x_n, u_t)$, where $1 \leq j_k \leq m$, $0 \leq k \leq n$, $x_k \geq 0$, $1 \leq k \leq n$, $u_t = t - x_1 - \cdots - x_n \geq 0$ and $p_{j_0} = P(J_0 = j_0)$. Of course, for $n = 0$ in (2.5) we replace $u_t$ with $t$ and the empty product over $k$ with a probability measure having unit mass at the point $(j_0, 0)$. Here (2.5) defines the projection of $P$ onto $(\Pi_n, \mathbf{\Pi}_n)$, $n \geq 0$, parametrized by the transition probabilities and the cumulative hazard functions. For a given set of these parameters, we take $(\mathbb{N} \times \Pi, \mathbf{N} \otimes \mathbf{\Pi}, P)$ as the underlying probability space for an observation from a Markov renewal process.

2.2. *The prior.* This section specifies the parameter space and an a priori distribution for our problem. Throughout, $(\Omega, \mathbf{F}, \mathbb{P})$ denotes a probability space upon which all random variables are defined. Let $(M_m, \mathbf{M}_m)$ denote the Borel space of $m \times m$ Markov matrices, where each element of $M_m$ is a transition probability matrix of a Markov chain on $E$. Let $(H, \mathbf{H})$ denote the Borel space of cumulative hazard functions, where each $b \in H$ is a nondecreasing, right-continuous function on $R_+$ satisfying $b(0) = 0$ and $\Delta b(t) \leq 1$, $t \geq 0$. Also let $(H^m, \mathbf{H}^m)$ denote the $m$-fold product space of $(H, \mathbf{H})$ with itself. The parameter space for our estimation problem is given by the product space $(\Theta, \mathbf{\Theta})$, where $\Theta = M_m \times H^m$ and $\mathbf{\Theta} = \mathbf{M}_m \otimes \mathbf{H}^m$.

To specify a random element of $(\Theta, \mathbf{\Theta})$, we proceed by specifying a random element of $(M_m, \mathbf{M}_m)$ and of $(H^m, \mathbf{H}^m)$ separately. Note that each matrix in

$M_m$ has rows that lie in the $m-1$-dimensional simplex in the unit cube. Let $\beta_1, \dots, \beta_m$ denote nonnegative quantities, and let $Z$ denote a random vector in the simplex. We use the notation $Z \sim \mathrm{Dir}(\beta_1, \dots, \beta_m)$ to denote that $Z$ is Dirichlet distributed with parameters $\beta_1, \dots, \beta_m$. Consider the following definition.

DEFINITION 2.1. Let $\boldsymbol{\beta} = (\beta_{ij})$, $i, j \in E$, denote a matrix of nonnegative quantities. Let $M = (M_{ij})$, $i, j \in E$, denote a random matrix defined on $(\Omega, \mathbf{F}, \mathbb{P})$. The rows of $M$ are independent random vectors such that for each $i \in E$, $(M_{i1}, \dots, M_{im}) \sim \mathrm{Dir}(\beta_{i1}, \dots, \beta_{im})$. We say $M$ is Dirichlet distributed with parameters $\boldsymbol{\beta}$ and denote this by $M \sim \mathrm{Dir}(\boldsymbol{\beta})$.

According to Definition 2.1, $M$ is a random matrix with values in $(M_m, \mathbf{M}_m)$ and distribution determined by the $m$-fold product measure of Dirichlet distributions over the simplex. Let $Q_0$ denote the probability measure on $(M_m, \mathbf{M}_m)$ induced by the mapping $M$. Here $Q_0$ serves as the a priori distribution of the transition probabilities of the process $J$.

Next we specify a random element of $(H, \mathbf{H})$. For this purpose, we introduce a family of Lévy processes developed in Hjort (1984, 1988). Let $b$ denote a fixed hazard function in $H$ with a finite number of discontinuities on the set $D$, and let $c = (c(t))$, $t \geq 0$, denote a positive, piecewise-continuous function. The process $A = (A_t)$, $t \geq 0$, is said to be a Beta process having parameters $c$ and $b$, which we denote $A \sim \mathrm{Beta}(c, b)$, provided $A$ is a process of independent increments having Lévy measure $\nu$ satisfying: For $t \geq 0$, $0 < x < 1$,

$$\nu(dt, dx) = \begin{cases} c(t) x^{-1} (1-x)^{c(t)-1} \, db(t) \, dx, & t \notin D, \\ f(t, x) \, dx, & t \in D, \end{cases} \tag{2.6}$$

where $f(t, \cdot)$ denotes the Beta density on $[0,1]$ having parameters $\beta_1 = c(t)\Delta b(t)$ and $\beta_2 = c(t)(1 - \Delta b(t))$, $t \in D$. The Beta family of Lévy processes is defined constructively in Theorem 3.1 of Hjort (1984, 1988), where it is shown that $\mathbb{P}(A \in H) = 1$. The Lévy measure characterizes the distribution of the increments of the process $A$ and appears in the exponent function of its Laplace transform; see, for example, Hjort (1988), equation (3.6). We note that $EA_t = \nu([0, t] \times [0, 1])$ for every $t$. The term Beta process derives from the fact that the increments of $A$ over short intervals are distributed *approximately* as Beta random variables, although these processes are almost surely discontinuous at those times in $D$, where the size of the jumps are Beta random variables whose densities appear in (2.6). Consider the following definition.

DEFINITION 2.2. For each $i \in E$, let $c_i$ denote a nonnegative, piecewise-continuous function and let $b_i \in H$ with a finite number of discontinuities on the set $D_i$, say. Let $\mathbf{A} = (A^1, \dots, A^m)$ denote a vector-valued process defined on $(\Omega, \mathbf{F}, \mathbb{P})$. The $A^i$ are independent and $A^i \sim \mathrm{Beta}(c_i, b_i)$, $i \in E$. We say $\mathbf{A}$ is a vector Beta process with parameter $(\mathbf{c}, \mathbf{b}) = ((c_1, \dots, c_m), (b_1, \dots, b_m))$ and denote this by $\mathbf{A} \sim \mathrm{Beta}(\mathbf{c}, \mathbf{b})$.

According to Definition 2.2, $\mathbf{A}$ is a random variable with values in $(H^m, \mathbf{H}^m)$. For each $i \in E$, let $Q_i$ denote the probability measure on $(H, \mathbf{H})$ induced by the mapping $A^i$. Then the probability measure on $(H^m, \mathbf{H}^m)$ induced by the mapping $\mathbf{A}$ is given by the product measure determined by the $Q_i$. This product measure serves as the a priori distribution of the cumulative hazard functions [see (2.2)] of the sojourn-time distributions of the process $(J, S)$.

We define a random element in $(\Theta, \boldsymbol{\Theta})$ as follows. Let $(M, \mathbf{A})$ denote a random variable defined on $(\Omega, \mathbf{F}, \mathbb{P})$ taking values in $(\Theta, \boldsymbol{\Theta})$. We assume that $(M, \mathbf{A})$ has distribution $Q$ over $(\Theta, \boldsymbol{\Theta})$, where $Q$ denotes the product measure given by Definitions 2.1 and 2.2 using the fixed parameters $\boldsymbol{\beta}$ and $(\mathbf{c}, \mathbf{b})$, so that $M$ and $\mathbf{A}$ are independent with $M \sim \mathrm{Dir}(\boldsymbol{\beta})$ and $\mathbf{A} \sim \mathrm{Beta}(\mathbf{c}, \mathbf{b})$. The distribution $Q$ is the a priori distribution for our problem where by varying the parameters $\boldsymbol{\beta}$ and $(\mathbf{c}, \mathbf{b})$ a family of a priori distributions is obtained. In order to interpret these parameters, we close this section with the following observations. We have

$$E(M_{ij}) = \beta_{ij} \Big/ \sum_{k \in E} \beta_{ik}, \qquad i, j \in E, \tag{2.7}$$

and

$$E(A_t^i) = b_i(t) \quad \text{and} \quad \mathrm{Var}(A_t^i) = \int_0^t \frac{(1 - \Delta b_i(s))}{c_i(s) + 1} \, db_i(s), \qquad i \in E, t \geq 0. \tag{2.8}$$

Thus the parameters $\boldsymbol{\beta}$ and $\mathbf{b}$ define an a priori guess on the transition probabilities and the cumulative hazards, respectively. Moreover, the function $\mathbf{c}$ parametrizes one's degree of prior belief in $\mathbf{b}$ because of the apparent way its size moderates the dispersion of $\mathbf{A}$ about $\mathbf{b}$.

**3. Bayes estimation.** The problem is Bayes estimation of the transition distributions from a Markov renewal process observed over a period of time. Our approach is to concatenate the procedure proposed by Hjort (1984) for estimating transition probabilities of a Markov chain with that for estimating cumulative hazards in life testing.

Let $(\Omega, \mathbf{F}, \mathbb{P})$ denote the probability space on which all random variables encountered below are defined. Let $(\mathbb{N} \times \Pi, \mathbb{N} \otimes \Pi)$ and $(\Theta, \boldsymbol{\Theta})$ denote the Borel spaces defined in Sections 2.1 and 2.2, respectively. In accordance with Section 2.2, let $(M, \mathbf{A})$ denote a random element in $(\Theta, \boldsymbol{\Theta}, Q)$, where $Q$ is the product measure defined above. Now consider a stochastic process $(J, S) = (J_n, S_n)$, $n \geq 0$, with state space $E \times R_+$. We assume that the conditional law of $(J, S)$ given $(M, \mathbf{A})$ is that of a Markov renewal process with transition probabilities $M = (M_{ij})$, $i, j \in E$, and cumulative rate functions $\mathbf{A} = (A^1, \dots, A^m)$. In particular, as the equivalent of (2.1) we have

$$\mathbb{P}(J_{n+1} = j, S_{n+1} - S_n \leq t | J_n = i, M, \mathbf{A}) = M_{ij}\Big(1 - \prod_{0 < s \leq t} (1 - dA_s^i)\Big), \qquad i, j \in E, t \geq 0. \tag{3.1}$$

Equation (3.1) specifies the regular conditional transition distributions of $(J, S)$ given $(M, \mathbf{A})$. Obviously, $J = (J_n)$, $n \geq 0$, is conditionally a Markov chain with state space $E$ and transition probabilities $M$. Moreover, the sojourn-times $\{X_n = S_n - S_{n-1},\ n \geq 1\}$ are conditionally independent given $(J, M, \mathbf{A})$, where for each $i \in E$, the sojourn-times in state $i$ have random cumulative hazard $A^i$.

Fix $t > 0$, and suppose the $(J, S)$ process is observed over $[0, t]$. We consider the random vector $(N(t), R(t))$ defined as in Section 2. Since $E$ is finite, it follows that $N(t)$ is almost surely finite so that $(N(t), R(t), M, \mathbf{A})$ lies almost surely in the sample space $(\mathbb{N} \times \Pi \times \Theta, \mathbb{N} \otimes \mathbf{\Pi} \otimes \mathbf{\Theta})$. A probability distribution over this sample space is defined as follows. We assume that $\mathbb{P}(J_0 = 1) = 1$. Then, using (3.1) we have

$$(3.2)\quad \mathbb{P}(N(t) = n,\ R(t) \in F,\ (M, \mathbf{A}) \in G) = \int_G \left[ \int_F dP(n; \pi | \theta) \right] dQ(\theta),$$

for $n \geq 0$, $G \in \mathbf{\Theta}$, $F \in \mathbf{\Pi}_n$, where $P(n; \cdot\, | \theta)$ is given by (2.4) for $\theta = ((p_{ij}), b_1, \dots, b_m) \in \Theta$ and $p_1 = 1$. According to Section 2.1, the latter defines a probability measure $P(\cdot\,|\theta)$, say, providing the regular conditional probability distribution of $(N(t), R(t))$ given $(M, \mathbf{A})$ at $\theta$.

We introduce the following random variables and processes: Fix $t > 0$ and define

$$(3.3)\qquad \tilde{N}_{ij} = \sum_{k=1}^{N(t)} 1(J_{k-1} = i,\ J_k = j), \qquad i, j \in E,$$

$$(3.4)\qquad N_i(s) = \sum_{k=1}^{N(t)} 1(J_{k-1} = i,\ X_k \leq s), \qquad s \geq 0,\ i \in E,$$

and

$$(3.5)\qquad Y_i(s) = 1(J_{N(t)} = i,\ t - S_{N(t)} \geq s) + \sum_{k=1}^{N(t)} 1(J_{k-1} = i,\ X_k \geq s), \qquad s \geq 0,\ i \in E.$$

Here the random variables $\{\tilde{N}_{ij},\ i, j \in E\}$ give the observed transition counts for the process $J$, while for $i \in E$, $N(i) = (N_i(s))$, $s \geq 0$ is a counting process defined over the completed sojourn-times in state $i$ and $Y(i) = (Y_i(s))$, $s \geq 0$, is the analog of a risk process in life testing. Notice that included in $Y(i)$ is the partially observed times $X_{N(t)+1}$ in the sense that its value is known only to exceed the backward recurrence time $t - S_{N(t)}$.

Let $\boldsymbol{\beta}$, $\mathbf{c}$ and $\mathbf{b}$ denote a fixed set of a priori parameters. For each $i \in E$, define the processes $B(i) = (B_i(s))$, $s \geq 0$, and $C(i) = (C_i(s))$, $s \geq 0$, by

$$(3.6)\qquad C_i(s) = c_i(s) + Y_i(s), \qquad s \geq 0,$$

and

$$(3.7)\qquad B_i(s) = \int_0^s \frac{c_i\, db_i + dN_i}{c_i + Y_i}, \qquad s \geq 0.$$

Here $B(i)$ is a weighted average of the prior guess $b_i$, and an analog of the Nelson estimator of cumulative hazard. The random variables $\boldsymbol{\beta} = (\beta_{ij} + \tilde{N}_{ij})$, $i, j \in E$, $\mathbf{C} = (C(1), \dots, C(m))$ and $\mathbf{B} = (B(1), \dots, B(m))$ play the role of the *updated* parameters in the posterior distribution.

The main result of the paper is the following theorem.

THEOREM 3.1. *For fixed* $\boldsymbol{\beta}$, $\mathbf{c}$ *and* $\mathbf{b}$, *suppose* $M$ *and* $\mathbf{A}$ *are independent with* $M \sim \mathrm{Dir}(\boldsymbol{\beta})$ *and* $\mathbf{A} \sim \mathrm{Beta}(\mathbf{c}, \mathbf{b})$ *defined by Definitions* 2.1 *and* 2.2, *respectively. Consider the probability model specified by* (3.1) *and* (3.2). *For* $t > 0$, *let* $(N(t), R(t))$ *denote an observation on the process* $(J, S)$ *over* $[0, t]$. *Then* $M$ *and* $\mathbf{A}$ *are conditionally independent given* $(N(t), R(t))$, *with conditional distributions given by* $\mathrm{Dir}(\boldsymbol{\beta})$ *and* $\mathrm{Beta}(\mathbf{C}, \mathbf{B})$, *respectively.*

The class of *Dirichlet times Beta process* priors is a conjugate family for Bayes estimation from a Markov renewal process. The naturalness of the problem and the product form of (2.4) and the prior lead one to this expectation, particularly when the first $n$ terms of the process are taken as data. Nevertheless, the observational scheme used here, which monitors the process over a period $[0, t]$, requires we establish two facts. Namely, notwithstanding the presence of semi-Markov dependence in the sample, the posterior is obtained by independently updating each factor in the prior, and the sojourn-times act on the posterior Beta process as do censored lifetimes in the framework of Hjort (1984, 1988). The care required in demonstrating the latter in Proposition 3.3 originates in the "censoring" of $X_{N(t)+1}$ to $t - S_{N(t)}$ being a function of the preceding observations.

The remainder of this section is devoted to proving Theorem 3.1. We proceed by first proving three needed propositions. Recall that $\Theta = M_m \times H^m$ and that for each $n$, $\Pi_n = (E \times R_+)^{n+1}$. Hence, for any $\theta \in \Theta$, we write $\theta = (\theta_0, \theta_1)$, where $\theta_0 \in M_m$ and $\theta_1 \in H^m$, and for any $\pi \in \Pi_n$, we write $\pi = (\pi_0, \pi_1)$, where $\pi_0 \in E^{n+1}$ and $\pi_1 \in R_+^{n+1}$. Now define

$$dP(n; \pi_0 | \theta_0) = \prod_{k=0}^{n-1} p_{j_k j_{k+1}}, \tag{3.8}$$

where $\pi_0 = (j_0, j_1, \dots, j_n)$ and $\theta_0 = (p_{ij})$, $i, j \in E$. Next, using equations (2.4), (3.2) and (3.8), define $dP(n; \pi_1 | \theta_1, \pi_0)$ so that

$$dP(n; \pi | \theta) = dP(n; \pi_1 | \theta_1, \pi_0)\, dP(n; \pi_0 | \theta_0), \tag{3.9}$$

where

$$\pi_1 = (x_1, x_2, \dots, x_n, u_t) \quad (u_t = t - x_1 - \cdots - x_n \ge 0),$$
$$\theta_1 = (b_1, b_2, \dots, b_m), \pi = (\pi_0, \pi_1)$$

and $\theta = (\theta_0, \theta_1)$. This obtains a convenient factorization of the kernel of the measure $P(n; \cdot\, | \theta)$ appearing in the bracketed term of (3.2).

We have the following proposition.

PROPOSITION 3.1. *Let* $(N(t), R(t), M, \mathbf{A})$ *denote the random element in* $(\mathbb{N} \times \Pi \times \Theta, \mathbf{N} \otimes \mathbf{\Pi} \otimes \mathbf{\Theta}, P)$, *where* $P$ *is determined by* (3.2). *Then*

$(X_1, \ldots, X_{N(t)}, \mathbf{A})$ *and* $M$ *are conditionally independent given* $(J_0, \ldots, J_{N(t)}, N(t))$.

PROOF. Since $N(t)$ is almost surely finite it suffices to prove the conditional independence on the set $\{N(t) = n\}$ for every $n \geq 0$. Fix $n$ and using (3.8) and (3.9), we begin by defining the $\mathbf{E}^{n+1}$-measurable function

$$f(n, \theta_0) = \int_{H^m \times R_+^{n+1}} dP(n; \pi_1 | \pi_0, \theta_1)\, dQ'(\theta_1), \qquad \pi_0 \in E^{n+1},$$

where $Q'$ is the distribution of $\mathbf{A}$, and $\mathbf{E}^{n+1}$ denotes all subsets of $E^{n+1}$. Next observe that according to (3.2) and using (3.8) we have

$$\mathbb{P}\big(N(t) = n, (J_0, \ldots, J_{N(t)}) \in F, M \in G\big)$$

$$= \int_{G \times F} f(n, \pi_0)\, dP(n; \pi_0 | \theta_0)\, dQ_0(\theta_0),$$

for $F \in \mathbf{E}^{n+1}$, and $G \in \mathbf{M}_m$. This defines a measure $\mu$, say, on $(E^{n+1} \times M_m, \mathbf{E}^{n+1} \otimes \mathbf{M}_m)$, which is the restriction of the distribution of $(N(t), J_0, \ldots, J_{N(t)}, M)$ to the set $\{N(t) = n\}$. Moreover, for all $G \in \mathbf{E}^{n+1} \otimes \mathbf{M}_m$, we have

$$\mathbb{P}\big(N(t) = n, (X_1, \ldots, X_{N(t)}) \in F, \mathbf{A} \in K, (J_0, \ldots, J_{N(t)}, M) \in G\big)$$

$$= \int_G \left[ \int_{K \times F} f^{\oplus}(n, \pi_0)\, dP(n; \pi_1 | \pi_0, \theta_1)\, dQ'(\theta_1) \right] d\mu(\pi_0, \theta_0),$$

for $F \in \mathbf{R}_+^{n+1}$ (the Borel sets in $R_+^{n+1}$), $K \in \mathbf{H}^m$ and where $f^{\oplus} = f^{-1}$ if $f > 0$ and $f^{\oplus} = 0$ if $f = 0$. Note that for each $n$ the term in brackets above is a measurable function of $\pi_0$. Moreover, for $\pi_0$ outside a set of $\mu$-measure zero, it defines a probability distribution over $(R_+^{n+1} \times H^m, \mathbf{R}_+^{n+1} \times \mathbf{H}^m)$. Therefore, by virtue of the equation above and the definition of $\mu$, it follows that the bracketed term determines a regular conditional distribution of $(X_1, \ldots, X_{N(t)}, \mathbf{A})$ given $(N(t), J_0, \ldots, J_{N(t)}, M)$ on $\{N(t) = n\}$. Since this is independent of $M$, the proposition is obtained. □

Proposition 3.1 implies that $\mathbf{A}$ and $M$ are conditionally independent given $(N(t), R(t))$. Hence, to obtain the posterior distribution of $(M, \mathbf{A})$, it suffices to individually derive the posterior of $M$ given $(N(t), J_0, J_1, \ldots, J_{N(t)})$ and the posterior of $\mathbf{A}$ given $(N(t), R(t))$.

The remaining two propositions are more technical, but they provide the thrust to the proof of Theorem 3.1. Fix $n \geq 0$ and consider $X_1, \ldots, X_{n+1}$ and $J_0, J_1, \ldots, J_n$. For each $i \in E$, define $N_i = \operatorname{card}\{k: J_k = i, k = 0, 1, \ldots, n-1\}$ and the random set $T_i = \{X_k: J_{k-1} = i, k = 1, \ldots, n\}$. Let $\tilde{Q}_i$ denote the random probability measure on $(H, \mathbf{H})$ defined by the following rule: For each $\omega \in \Omega$, if $N_i(\omega) = 0$, then set $\tilde{Q}_i(\omega) = Q_i$. Otherwise if $N_i(\omega) > 0$ and $T_i(\omega) = \{x_k, k = 1, \ldots, N_i(\omega)\}$, then set $\tilde{Q}_i(\omega)$ equal to the probability on $(H, \mathbf{H})$ obtained by formal application of Theorem 4.1 of Hjort (1984) with $Q_i$ as the starting measure and $T_i(\omega)$ as the observed sample. According to Corollary 4.1 of Hjort (1984), $\tilde{Q}_i(\omega)$ is the probability law of a Beta process with parameters obtained by updating $(c_i, b_i)$ according to the rules set by (ii) and (iii) of Theorem 4.1 of Hjort (1984).

We have the following proposition.

PROPOSITION 3.2. *Fix $n \geq 0$ and let $X = (X_0, X_1, \dots, X_n)$, $X_0 = 0$, $Y = (J_0, J_1, \dots, J_n)$ and $\pi = (1, j_1, \dots, j_n) \in E^{n+1}$. Consider the event $B(\pi) = \{Y = \pi\}$ and define $\mathbf{G} = \sigma(X, Y) \vee \sigma(B(\pi))$. Let $r > 0$ and let $G$ denote the rectangle $G_1 \times G_2 \times \cdots \times G_m \in \mathbf{H}^m$ and suppose $\mathbb{P}(B(\pi)) > 0$. Then on $B(\pi)$ we have*

$$\mathbb{P}(X_{n+1} > r, \mathbf{A} \in G | \mathbf{G}) = \prod_{i \neq j_n} \tilde{Q}_i(G_i) \tilde{E}_{j_n} \prod_{0 < s \leq r} \left(1 - dA_s^{j_n}\right) 1\left(A^{j_n} \in G_{j_n}\right),$$

*where $\tilde{E}_{j_n}$ is the expectation operator defined by the random measure $\tilde{Q}_{j_n}$. Hence the $\tilde{Q}_i$ give regular conditional distributions of the $A^i$ given $\mathbf{G}$.*

PROOF. Observe that almost surely we have

$$\begin{aligned}\mathbb{P}(X_{n+1} > r, \mathbf{A} \in G | \mathbf{G}) &= E\left(1(X_{n+1} > r) 1(\mathbf{A} \in G) | \mathbf{G}\right) \\ &= E\left(1(\mathbf{A} \in G) E\left(1(X_{n+1} > r) | \sigma(X, Y, \mathbf{A}) \vee \sigma(B(\pi))\right) | \mathbf{G}\right) \\ &= E\left(1(\mathbf{A} \in G) \prod_{0 < s \leq r} \left(1 - dA_s^{j_n}\right) | \mathbf{G}\right),\end{aligned}$$

where these equalities follow by definition, the inclusion $\mathbf{G} \subset \sigma(X, Y, \mathbf{A}) \vee \sigma(B(\pi))$, and assumption (3.1), respectively. Next recall that by hypothesis, the times $X_1, \dots, X_{n+1}$ are conditionally independent given $B(\pi)$ and $\mathbf{A}$, where the sojourn-times in state $i$, namely the set of times $T_i$, $i \neq j_n$, and $\{X_{n+1}\} \cup T_{j_n}$, $i = j_n$, have conditional distribution determined by the process $A^i$, for $i \in E$. Moreover, by virtue of the independence of $M$ and $\mathbf{A}$, and the independence of the $A^i$, it is easily shown that the random variables $(T_i, A^i)$, $i \neq j_n$, $(\{X_{n+1}\} \cup T_{j_n}, A^{j_n})$ are conditionally independent given $B(\pi)$. Therefore, on $B(\pi)$ we have

$$\begin{aligned}&E\left(1(\mathbf{A} \in G) \prod_{0 < s \leq r} \left(1 - dA_s^{j_n}\right) | \mathbf{G}\right) \\ &\quad = \prod_{i \neq j_n} E\left(1\left(A^i \in G_i\right) | \mathbf{G}\right) E\left(\prod_{0 < s \leq r} \left(1 - dA_s^{j_n}\right) 1\left(A^{j_n} \in G_{j_n}\right) | \mathbf{G}\right),\end{aligned}$$

where terms in the product on the right above depend only on those sojourns in $T_i$, $i \neq j_n$, and $T_{j_n}$, respectively.

Yet again by hypothesis and given $B(\pi)$, each of the subproblems posed by the conditional expectations above is equivalent to finding the posterior law of a Beta process $A^i$ given a sample of size $n_i = \operatorname{card}\{k\colon j_k = i,\ k = 0, 1, \dots, n-1\}$ lifetimes in $T_i$ from the model for Bayes estimation in life testing proposed by Hjort (1984). Hence, by Theorem 4.1 and Corollary 4.1 of Hjort (1984), it follows that on $B(\pi)$ we have

$$E\Big(1(\mathbf{A} \in G) \prod_{0<s\leq r} (1 - dA_s^{j_n}) \big| \mathbf{G}\Big)$$
$$= \prod_{i \neq j_n} \tilde{Q}_i(G_i) \tilde{E}_{j_n} \prod_{0<s\leq r} (1 - dA_s^{j_n}) 1\big(A^{j_n} \in G_{j_n}\big)$$

and the proposition is thus proved. □

It follows from Proposition 3.2 that on $B(\pi)$ we have

$$\mathbb{P}(X_{n+1} > r | \mathbf{G}) = \tilde{E}_{j_n} \prod_{0<s\leq r} (1 - dA_s^{j_n}), \qquad r \geq 0, \tag{3.10}$$

where $\tilde{E}_{j_n}$ denotes the expectation operator induced by $\tilde{Q}_{j_n}$. The expression on the right-hand side of (3.10) admits the following interpretation. For each $\omega \in B(\pi)$, let $\tilde{Q}_{j_n}(\omega)$ denote the probability measure identified by Proposition 3.2. Then, the right-hand side of (3.10) can be interpreted as the probability some positive random variable $Y$, say, exceeds $r$ in the context of a Bayes model for life testing with prior process $A^{j_n}$ distributed according to $\tilde{Q}_{j_n}(\omega)$. Using this interpretation, we see that in Proposition 3.2 the problem of further conditioning on the event $\{X_{n+1} > r\}$ is in essence equivalent to updating the distribution $\tilde{Q}_{j_n}(\omega)$ by conditioning on the event $\{Y > r\}$. The latter is formally handled by (i) of Theorem 4.1 of Hjort (1984). Hence, let $\tilde{Q}_{j_n}^r(\omega)$ denote the probability on $(H, \mathbf{H})$ obtained by formal application of (i) of Theorem 4.1 of Hjort (1984) to $\tilde{Q}_{j_n}(\omega)$ as described above. Consider the following proposition.

PROPOSITION 3.3. *Fix $n \geq 0$ and let $B(\pi)$, $\mathbf{G}$ and $G \in \mathbf{H}^m$ be defined as in Proposition 3.2. Let $r > 0$ and suppose $\mathbb{P}(B(\pi) \cap \{X_{n+1} > r\}) > 0$. Then on $B(\pi) \cap \{X_{n+1} > r\}$ we have*

$$\mathbb{P}\big(\mathbf{A} \in G \big| \mathbf{G} \vee \sigma(\{X_{n+1} > r\})\big) = \prod_{i \neq j_n} \tilde{Q}_i(G_i) \tilde{Q}_{j_n}^r(G_{j_n}).$$

*Second, define $\Omega_n = \{N(t) = n\}$ and suppose $\mathbb{P}(B(\pi) \cap \Omega_n) > 0$. Then on $B(\pi) \cap \Omega_n$ we have:*

$$\mathbb{P}\big(\mathbf{A} \in G \big| \mathbf{G} \vee \sigma(\Omega_n)\big) = \prod_{i \neq j_n} \tilde{Q}_i(G_i) \tilde{Q}_{j_n}^{u_t}(G_{j_n}),$$

*where $u_t = t - x_1 - \cdots x_n \geq 0$ for $X_1(\omega) = x_1, \ldots, X_n(\omega) = x_n$.*

PROOF. We define a probability $\mu$ as follows. For $G \in \mathbf{H}^m$, let $\mu(G) = \mathbb{P}(\mathbf{A} \in G | \{X_{n+1} > r\})$. The conditional probability in question can be computed from the measure $\mu$; namely on $\{X_{n+1} > r\}$ we have

$$\mathbb{P}\big(\mathbf{A} \in G \big| \mathbf{G} \vee \sigma(\{X_{n+1} > r\})\big) = \mu(G | \mathbf{G}).$$

Now it is easily shown that, outside a set of $\mu$-measure zero, we have

$$\mu(G|\mathbf{G}) = \mathbb{P}(\mathbf{A} \in G,\, X_{n+1} > r|\mathbf{G})/\mathbb{P}(X_{n+1} > r|\mathbf{G})$$

[see, for example, Billingsley (1979), page 404]. Hence, by virtue of Proposition 3.2 and equation (3.10), on $B(\pi)$ we have

$$\mu(G|\mathbf{G}) = \prod_{i \neq j_n} \tilde{Q}_i(G_i) U(r, j_n, G_{j_n}),$$

for $G = G_1 \times \cdots \times G_m$, and where

$$U(r, j_n, G_{j_n}) = \tilde{E}_{j_n} \prod_{0<s\le r} (1 - dA_s^{j_n}) 1(A^{j_n} \in G_{j_n}) / \tilde{E}_{j_n} \prod_{0<s\le r} (1 - dA_s^{j_n}).$$

But it follows from (i) of Theorem 4.1 of Hjort (1984), that for each $\omega \in B(\pi)$ we have

$$U(r, j_n, G_{j_n})(\omega) = \tilde{Q}_{j_n}^r(G_{j_n}, \omega),$$

so the first assertion is proved.

To prove the second assertion we define a probability $\nu$ as follows. For $G \in \mathbf{H}^m$, let $\nu(G) = \mathbb{P}(\mathbf{A} \in G|\Omega_n)$. Then, outside a set of $\nu$-measure zero and arguing as above, we have

$$\begin{aligned}\mathbb{P}(\mathbf{A}\in G|\mathbf{G} \vee \sigma(\Omega_n))1(N(t) = n) &= \nu(G|\mathbf{G}) \\ &= \mathbb{P}(\mathbf{A} \in G,\, N(t) = n|\mathbf{G})/\mathbb{P}(N(t) = n|\mathbf{G}).\end{aligned}$$

To evaluate this last expression, let $x_i \ge 0$, $i = 1, \dots, n$, such that $u_t = t - x_1 - \cdots - x_n \ge 0$. Observe that for any $\omega \in \Omega$, we have $X_1(\omega) = x_1, \dots, X_n(\omega) = x_n$, $N(t, \omega) = n$ if and only if $X_1(\omega) = x_1, \dots, X_n(\omega) = x_n$, $X_{n+1}(\omega) > u_t$. Therefore, on $B(\pi) \cap \{X_1 + \cdots + X_n \le t\}$ we have

$$\nu(G|\mathbf{G}) = \mathbb{P}(\mathbf{A} \in G,\, X_{n+1} > u_t|\mathbf{G})/\mathbb{P}(X_{n+1} > u_t|\mathbf{G}) = \prod_{i \neq j_n} \tilde{Q}_i(G_i) \tilde{Q}_i^{u_t}(G_{j_n}),$$

for $G = G_1 \times \cdots \times G_m$, where the last equality follows from the first assertion of this proposition with $r$ replaced by $u_t$. The second assertion is thus proved. □

We are now in a position to give a proof of Theorem 3.1.

Proof of Theorem 3.1. Proposition 3.1 implies that $M$ and $\mathbf{A}$ are conditionally independent given $(N(t), R(t))$. Therefore, it remains only to compute the conditional distribution of $M$ and of $\mathbf{A}$ given $(N(t), R(t))$.

Proposition 3.1 implies that, to compute the conditional law of $M$, it suffices to compute the conditional law of $M$ given $(N(t), J_0, J_1, \dots, J_{N(t)})$. Define $\Omega_n = \{N(t) = n\}$, for $n \ge 0$. Since $N(t)$ is almost surely finite and by an elementary calculation, we have

$$\mathbb{P}\big(M \in G|\sigma(N(t)\, J_0, J_1, \dots, J_{N(t)})\big) = \sum_{n \ge 0} \mathbb{P}(M \in G|\sigma(J_0, \dots, J_n))1_{\Omega_n},$$

for $G \in \mathbf{M}_m$. By hypothesis, for each $n \ge 0$, $J_0, J_1, \dots, J_n$ are observations over $[0, n]$ from a Markov chain having transition probabilities $M$ drawn from the distribution $\mathrm{Dir}(\boldsymbol{\beta})$. Hence, using the hypothesis of time homogeneity, the repeated application (i.e., over successive transitions) of (iii) of Theorem 2.2 of Hjort (1984) shows directly that $P(M \in G\ \sigma(J_0, J_1, \dots, J_n))1_{\Omega_n}$ is computed from the distribution $\mathrm{Dir}(\boldsymbol{\beta})$, where $N(T) = n$ in (3.3). This proves the desired result for $M$.

To compute the conditional law of $\mathbf{A}$ given $(N(t), R(t))$, define $\mathbf{G}_n = \sigma(J_0, J_1, \dots, J_n, X_1, \dots, X_n) \vee \sigma(\Omega_n)$, for $n \geq 0$. Since $N(t)$ is almost surely finite, we have

$$\mathbb{P}(\mathbf{A} \in G | \sigma(N(t), R(t))) = \sum_{n \geq 0} \mathbb{P}(\mathbf{A} \in G | \mathbf{G}_n) 1_{\Omega_n},$$

for $G \in \mathbf{H}^m$. For rectangles of the form $G = G_1 \times \cdots \times G_m$, Proposition 3.3 implies that

$$\mathbb{P}(\mathbf{A} \in G | \mathbf{G}_n) 1_{\Omega_n} = \prod_{i \neq J_n} \tilde{Q}_i(G_i) \tilde{Q}_{J_n}^{U_t}(G_{J_n}),$$

where $U_t = t - X_1 - \cdots - X_n \geq 0$, and where the $\tilde{Q}_i$ $i \neq J_n$, and $\tilde{Q}_{J_n}^{U_t}$ are determined as in Propositions 3.2 and 3.3. Hence, Corollary 4.1 of Hjort (1984) implies that $\mathbb{P}(\mathbf{A} \in G|\mathbf{G}_n)1_{\Omega_n}$ is computed from the distribution $\mathrm{Beta}(\mathbf{C}, \mathbf{B})$, where $N(t) = n$ in (3.4)–(3.7). This proves the desired result for $\mathbf{A}$ and completes the proof.

**4. Remarks and generalizations.** The present paper provides a procedure for Bayes estimation from a Markov renewal process. In particular, suppose the Bayes estimators of the Markov transition probabilities and the cumulative hazards of the sojourn-time distributions are desired. Then, under squared-error loss, these are given by the *posterior means* of the conditional distribution derived in Theorem 3.1. For the transition probabilities, we have

$$E(M_{ij} | N(t), R(t)) = \boldsymbol{\beta}_{ij} \Big/ \sum_{k \in E} \boldsymbol{\beta}_{ik}, \qquad i, j \in E, \tag{4.1}$$

where the $\boldsymbol{\beta}_{ij} = \tilde{N}_{ij} + \beta_{ij}$ are defined in Theorem 3.1. For the cumulative hazards of the sojourn-time distributions, we have

$$E(A_s^i | N(t), R(t)) = B_i(s), \qquad s \geq 0, i \in E, \tag{4.2}$$

where $B(i) = (B_i(s))$, $s \geq 0$, is defined by (3.7). By virtue of the conditional independence of $M$ and $\mathbf{A}$ given $(N(t), R(t))$, it follows from (3.1), (4.1) and (4.2) that the Bayes estimators of the transition distributions are given by

$$\begin{aligned} &E\left( M_{ij}\left(1 - \prod_{0 < u \leq s} (1 - dA_u^i)\right) \middle| N(t), R(t) \right) \\ &\qquad = \left( \boldsymbol{\beta}_{ij} \Big/ \sum_{k \in E} \boldsymbol{\beta}_{ik} \right) \left( 1 - \prod_{0 < u \leq s} (1 - dB_i(u)) \right), \end{aligned} \tag{4.3}$$

for $s \geq 0$ and $i, j \in E$.

The posterior analysis can proceed to compute other parameters of interest, such as mean sojourn-times, renewal functions, etc. Many of these are expressed as the posterior mean of a functional of $(M, \mathbf{A})$; the transition distributions computed in (4.3) provide a particular example. Alternatively, consider the problem of obtaining a Bayesian confidence set for the parameter, for which the posterior variances of the estimators are useful in constructing *approximate* Bayesian confidence intervals or bands. These are given by

$$\text{(4.4)} \quad \operatorname{Var}(M_{ij} | N(t), R(t)) = \beta_{ij}\Big(\sum_{k \neq j} \beta_{ik}\Big) \Big/ \left[\Big(\sum_{k} \beta_{ik}\Big)^2 \Big(\sum_{k} \beta_{ik} + 1\Big)\right],$$

and

$$\text{(4.5)} \qquad \operatorname{Var}(A_s^i | N(t), R(t)) = \int_0^s \frac{1 - \Delta B_i}{C_i + 1} dB_i.$$

In principle, the variance of a transition distribution is obtained from (4.4), (4.5) and numerous applications of the product rule for independent variables. Instead, we recommend working with $\operatorname{Var}(M_{ij} A_s^i | N(t), R(t))$ to construct approximate Bayesian confidence bands for the transition rates. This may be achieved using sample-path simulations from the posterior distribution.

The conjugate family of priors considered here have each member of the family specified by choice of the parameters $\boldsymbol{\beta}$, $\mathbf{c}$ and $\mathbf{b}$. The problem of estimating these parameters from an empirical record remains to be explored as part of the general methodology for approximate posterior analysis mentioned above. The estimators may be based on past observations of the $(J, S)$ process in leading to consistent empirical Bayes procedures.

The Bayes estimators of the transition distributions defined at (4.3) provide alternative nonparametric estimators to those proposed in Gill (1980a) and Phelan (1988). The estimators in Phelan (1988) are obtained by formally setting the $\beta_{ij}$ and $c_i$ equal to 0 in the updated versions of the parameters. On the other hand, those in Gill (1980a) are obtained similarly, provided one restricts Gill's problem and motivates the estimators accordingly. One way is to restrict the class of transition distributions to those that factor as in Phelan (1988), another is to restrict the censoring mechanism as below.

The results of this paper generalize in a number of ways. Obviously, we can accommodate additional observations provided these are of conditionally independent realizations of the Markov renewal process. In this case one simply aggregates over these realizations when computing the random variables defined in (3.3)–(3.5). The conclusion of Theorem 3.1 is the same. A more interesting generalization for applications allows for random censorship. Indeed a restricted version of random right-censorship of a Markov renewal process is accommodated without fundamental change to the proofs nor the conclusion of Theorem 3.1. This model allows for right-censorship of the sojourn-times without loss of the state information of the $J$ process. This is the essence of the restriction of Gill's censoring mechanism referred to above.

The proof of the main result rests on the simultaneous use of Theorems 2.2 and 4.1 of Hjort (1984); the former being concerned with Bayes inference from Markov chains, while the latter is concerned with Bayes inference from lifetimes. Individually these theorems cover a wider class of priors than those considered here, so that another immediate generalization of our result is obtained by using the wider class of priors, where a proof of the conjugate property would require similar steps. On the other hand, it has been suggested by a referee that such a program might flow more easily from the building of a time-discrete framework for renewal-type phenomena, taking the time-continuous case as a limit over a suitable sequence. This approach may prove particularly valuable in generalizing the whole framework to other observational schemes and to the Markov additive processes treated for example in Çinlar (1972).

**Acknowledgment.** The author acknowledges the valuable comments of an anonymous referee.

## REFERENCES

BILLINGSLEY, P. (1979). *Probability and Measure*. Wiley, New York.

BROCK, D. B. (1973). Bayesian analysis for Makov renewal processes. *Metron* **29** 185–198.

ÇINLAR, E. (1972). Markov additive processes. I. II. *Z. Wahrsch. Verw. Gebiete* **24** 85–93, 95–121.

GILL, R. D. (1980a). Nonparametric estimation based on censored observations of a Markov renewal process, *Z. Wahrsch. Verw. Gebiete* **35** 97–116.

GILL. R. D. (1980b). *Censoring and Stochastic Integrals. Math. Centre Tracts* **124** Amsterdam.

HJORT, N. L. (1984). Nonparametric Bayes estimators of cumulative intensities in models with censoring. Technical Report 762, Norwegian Computing Center.

HJORT, N. L. (1988). Nonparametric Bayes estimators based on Beta processes in models for life history data. SAND/24/1988, Norwegian Computing Center Report.

MOORE, E. H. and PYKE, R. (1968). Estimation of the transition distributions of a Markov renewal process. *Ann. Inst. Statist. Math.* **20** 411–424.

PHELAN, M. J. (1988). Nonparametric estimation from a censored Markov renewal process. Technical Report SOR-88-15, Princeton Univ.

PYKE, R. (1961). Markov renewal processes with finitely many states. *Ann. Math. Statist.* **32** 1243–1259.

PYKE, R. and SCHAUFELE, R. (1964). Limit theorems for Markov renewal processes. *Ann. Math. Statist.* **35** 1746–1764.

VOELKEL, J. G. and CROWLEY, J. (1984). Nonparametric inference for a class of semi-Markov processes with censored observations. *Ann. Statist.* **12** 142–160.

DEPARTMENT OF CIVIL ENGINEERING
AND OPERATIONS RESEARCH
PRINCETON UNIVERSITY
PRINCETON, NEW JERSEY 08544