The Annals of Statistics
1999, Vol. 27, No. 2, 562–588

NONPARAMETRIC BAYESIAN ESTIMATORS FOR COUNTING PROCESSES

BY YONGDAI KIM

Hankuk University of Foreign Study

This paper is concerned with nonparametric Bayesian inference of the Aalen's multiplicative counting process model. For a desired nonparametric prior distribution of the cumulative intensity function, a class of Lévy processes is considered, and it is shown that the class of Lévy processes is conjugate for the multiplicative counting process model, and formulas for obtaining a posterior process are derived. Finally, our results are applied to several practically important models such as one point processes for right-censored data, Poisson processes and Markov processes.

1. Introduction. For a given probability space $(\Omega, \mathscr{F}, P)$, a stochastic process $N(t)$ defined on the time interval $[0, \tau]$ is called a counting process if the sample paths of the process are right continuous step functions with value 0 at $t = 0$ and with a finite number of jumps, each positive and of size 1. In practice, $N(t)$ counts the number of certain events in time $[0, t]$, and so counting process models are widely used in a variety of fields where a sequence of times, each time corresponding to the occurrence of some event, constitutes the observable data.

One of the widely used counting process models in statistics is the Aalen's multiplicative intensity model [Aalen (1978)]. Let $\Lambda(t)$ be the cumulative intensity process of $N(t)$. That is, $\Lambda(t)$ is a predictable process such that $N(t) - \Lambda(t)$ is a martingale with respect to a given filtration $\mathscr{F}_t$ (increasing right-continuous, complete σ-field). The counting process $N(t)$ is called a multiplicative intensity model if the cumulative intensity process $\Lambda(t)$ can be written as

$$\Lambda(t) = \int_0^t Y(s)\, dA(s),$$

where $A(t)$ is a right-continuous nondecreasing function and $Y(t)$ is a nonnegative predictable process. Typically, Y is a censoring indicator and A is a parameter of interest. Then A can be a cumulative hazard function for survival data, a mean function for Poisson processes and a cumulative transition intensity function for Markov processes. Various developments of statistical theory for the multiplicative counting process model can be found in Karr (1986) and Anderson, Borgan, Gill and Keiding (1993).

In this paper, we are concerned with nonparametric Bayesian inference of the multiplicative intensity model. We will give a nonparametric prior distribution to the cumulative intensity function $A(t)$ and obtain the posterior

---

Received March 1997; revised October 1998.

AMS 1991 *subject classifications*. Primary 62C10; secondary 60G55.

Key words and phrases. Nonparametric Bayesian estimator, multiplicative counting process, Lévy process.

distribution: the conditional distribution of A given that N is observed continuously on $[0, \tau]$. For a prior distribution, Lévy processes without a Gaussian term and a deterministic term are used. Such processes have been widely used in Bayesian inference. Doksum (1974) noted the importance of Lévy processes as a class of conjugate prior distributions in survival analysis. For counting process models, Lo (1982, 1992) used the weighted gamma process as a prior distribution of the mean function when sampling a Poisson process model. Hjort (1990) used a beta process for a prior distribution of the cumulative hazard function and showed that the beta process is a conjugate prior for right-censored observations. Since all of the aforementioned models are special cases of the multiplicative intensity model, our approach provides a unified method for a wider class of nonparametric models. Along with the unification of the previously considered models, the multiplicative intensity model includes various complicated censoring–filtering problems. Examples are right-censored data with delayed entrance (left truncated) and the counter model. Details are provided in Section 4.

The outline of this paper is as follows. In Section 2, we present the semimartingale approach to Lévy processes since this approach has several advantages over the well-known Lévy formula. Section 3 contains the main results of this paper, the posterior distribution of the cumulative intensity function A given N. Finally in Section 4, examples are presented.

Throughout this paper, we will, if not otherwise stated, restrict all processes to the time interval $[0, 1]$. However, all of our results are also valid on $[0, \tau]$ for any $\tau > 0$.

2. Lévy processes and semimartingales. In this section, semimartingale approaches to Lévy process are reviewed, since they are extensively used in this paper. Before proceeding further, we explain why we use semimartingale approaches instead of the well-known Lévy formula. It is well known that a Lévy process A without a Gaussian term and a deterministic term is a pure jump process which can be rewritten as

$$A(t) = \sum_{s \le t:\, \Delta J(s)=1} \Delta A(s), \tag{2.1}$$

where $J(t) = \sum_{s \le t} I(\Delta A(s) \neq 0)$ and $\Delta A(t) = A(t) - A(t-)$. On the other hand, if the intensity function A is given by (2.1), we have

$$\begin{aligned} &\Pr\{N(1) = n \text{ and } \mathbf{S_n} = \mathbf{s_n} | A\} \\ &\quad = \prod_{t=0}^{1} [Y(t)\Delta A(t)]^{I(t \in \mathbf{s_n})} [1 - Y(t)\Delta A(t)]^{I(t \notin \mathbf{s_n})}, \end{aligned} \tag{2.2}$$

where $\mathbf{S_n} = (S_1, \dots, S_n)$, S_i is the time of the ith jump of N, and $\mathbf{s_n} \in [0, 1]^n$ such that $0 < s_1 < \cdots < s_n < 1$. Hence, roughly speaking, the posterior distribution of A given N can be obtained once the distributions of $J(t)$ and $\Delta A(t)$ are given a priori, which are easily derived from the semimartingale representation of Lévy processes as shown below.

Let $\mathscr{A}$ be the class of all right-continuous nondecreasing functions A defined on $[0, 1]$, zero at time zero, such that the size of each jump is less than or equal to 1, and let $\Sigma_{\mathscr{A}}$ be the σ-algebra generated by the Borel cylinder sets. Suppose that A is a Lévy process on $(\mathscr{A}, \Sigma_{\mathscr{A}})$ (without a deterministic term). Define a σ-finite measure μ on $[0, 1] \times [0, 1]$ such that

$$\mu([0, t] \times D) = \sum_{s \leq t} I(\Delta A(s) \in D)$$

for $t \in [0, 1]$ and $D \in \mathscr{B}[0, 1]$ where $\mathscr{B}[0, 1]$ denotes the σ-field on $[0, 1]$. Then it is easy to see that the random measure μ is a Poisson random measure [Jacod and Shiryaev (1987), page 70]. If we define a σ-finite measure ν on $[0, 1] \times [0, 1]$ by

$$\nu([0, t] \times D) = \mathrm{E}[\mu([0, t] \times D)]$$

for $t \in [0, 1]$ and $D \in \mathscr{B}[0, 1]$, then we have

$$\mathrm{E}[A(t)] = \int_0^t \int_0^1 x\nu(ds, dx) \tag{2.3}$$

and

$$\mathrm{Var}(A(t)) = \int_0^t \int_0^1 x^2\nu(ds, dx) - \left(\sum_{s \leq t} \int_0^1 x\nu(\{s\}, dx) \right)^2. \tag{2.4}$$

Here, the measure ν is called the compensator of the Poisson random measure μ. Conversely, suppose that a positive σ-finite measure ν on $[0, 1] \times [0, 1]$ satisfying

$$\nu(\{t\} \times [0, 1]) \leq 1 \tag{2.5}$$

is given. Then, there exists a unique Poisson random measure μ whose compensator is ν [Jacod (1979)], and a process A defined by

$$A(t) = \int_0^t \int_0^1 x\mu(ds, dx) \tag{2.6}$$

is a Lévy process defined on $(\mathscr{A}, \Sigma_{\mathscr{A}})$ provided that

$$\int_{[0, t] \times [0, 1]} x\nu(ds, dx) < \infty. \tag{2.7}$$

Consequently, we can construct a Lévy process on $(\mathscr{A}, \Sigma_{\mathscr{A}})$ simply by choosing a σ-finite measure ν on $[0, 1] \times [0, 1]$ satisfying (2.5) and (2.7). Furthermore, since the Lévy process A constructed by (2.6) satisfies the moment conditions (2.3) and (2.4), we can reflect prior information given by mean and variance to the prior process by choosing an appropriate σ-finite measure ν. More details on this issue are given in Section 4. In what follows, we call the measure ν simply the compensator of A.

REMARK. It is well known that any Lévy process $A(t)$ on $(\mathscr{A}, \Sigma_{\mathscr{A}})$ admits a Lévy representation

$$E[\exp(-\theta A(t))] = \left[\prod_{t_j \le t} E(-\theta U_j)\right] \exp\left[-\int_0^1 (1 - \exp(-\theta x))\, dL_t(x)\right],$$

where $\{L_t,\ t \ge 0\}$ is a continuous Lévy measure. Then the compensator ν of A becomes

$$\nu([0, t] \times D) = \int_D dL_t(x) + \sum_{t_j \le t} \int_D dH_j(x)$$

for $t \in [0, 1]$ and $D \in \mathscr{B}[0, 1]$, where $H_j(x)$ is a distribution function of a random variable U_j.

Now, we present how to derive the distributions of $J(t)$ and $\Delta A(t)$ for a given Lévy process on $(\mathscr{A}, \Sigma_{\mathscr{A}})$. Let A be a Lévy process on $(\mathscr{A}, \Sigma_{\mathscr{A}})$ with a compensator ν. Define a set $\mathbf{U}$ by setting $\mathbf{U} = \{t | \nu(\{t\} \times [0, 1]) > 0\}$. Then it is easy to see that $\mathbf{U}$ is the set of times of fixed discontinuity of A. We will always assume that the number of elements in $\mathbf{U}$ is finite and denote it by $\mathbf{U} = \{u_1, \dots, u_l\}$. Further, we assume that ν can be decomposed into

$$\nu([0, t] \times D) = \int_0^t \int_D dF_s(x)\, ds + \sum_{u_j \le t} \int_D dH_j(x) \tag{2.8}$$

for $t \in [0, 1]$ and $D \in \mathscr{B}[0, 1]$, where F_s are σ-finite measures on $[0, 1]$ such that $\int_0^1 x\, dF_s(x) < \infty$ and H_j are distribution functions on $[0, 1]$. Suppose that $\nu([0, 1] \times [0, 1]) < \infty$ and $\mathbf{U} = \varnothing$. Then A becomes a compound Poisson process such that $J(t)$ is a Poisson process with intensity function $\lambda(t)$ where $\lambda(t) = \int_0^1 dF_t(x)$ and the distribution function of $\Delta A(t)$ conditional on $\Delta A(t) > 0$ is $F_t(x)/\lambda(t)$. If $\mathbf{U} \neq \varnothing$, J becomes a compound Poisson process with fixed discontinuity at $u_1, \dots, u_l$ and H_j is the distribution function of $\Delta A(u_j)$, in which case A is called an *extended compound Poisson process*. If ν is not a finite measure, no nice description of the processes $J(t)$ and $\Delta A(t)$ exists. However, we can always approximate any Lévy process A on $(\mathscr{A}, \Sigma_{\mathscr{A}})$ with a compensator ν given by (2.8) by a sequence of extended compound Poisson processes as follows. Define

$$F_t^n(x) = \int_{1/n}^{x \vee 1/n} dF_t(u)$$

and construct a compound Poisson process $A_n(t)$ with the compensator ν_n,

$$\nu_n([0, t] \times D) = \int_0^t \int_D dF_s^n(x)\, ds + \sum_{u_j \le t} \int_D dH_j(x) \tag{2.9}$$

for $t \in [0, 1]$ and $D \in \mathscr{B}[0, 1]$. The next theorem is a simple corollary of Theorem 3.13 in Jacod and Shiryaev (1987).

THEOREM 2.1. *As $n \to \infty$, A_n converges to A weakly on $\mathscr{D}$, the space of right-continuous functions with left limits existing on* $[0, 1]$ *equipped with Skorohod topology*.

When $\nu([0,1]\times[0,1]) < \infty$, the posterior distribution of A can be calculated directly since the distributions of $J(t)$ and $\Delta A(t)$ are available, while when $\nu([0,1]\times[0,1]) = \infty$ the posterior distribution of A can be obtained as the limit of the posterior distributions of A_n, which are the main contents of Section 3.

3. Posterior distribution. We begin by stating the model to be studied. For a given intensity function A, let N be a counting process such that

$$\left\{N(t) - \int_0^t Y(s)\,dA(s), \mathscr{F}_t^N\right\} \tag{3.1}$$

is a martingale where $\mathscr{F}_t^N$ is a filtration generated by N and $Y(s)$ is a predictable process with respect to $\mathscr{F}_t^N$. For a prior distribution, we assume that A is a Lévy process defined on $(\mathscr{A}, \Sigma_{\mathscr{A}})$ with a compensator ν. The problem of nonparametric Bayesian inference consists in construction of the posterior distribution, the conditional distribution of A given a realization of N on $[0, 1]$.

REMARK. Note that we use $\mathscr{F}_t^N$ as a filtration of N, which means that we can only observe a realization of N. In practice, we may have more information than N, in which case the complete probabilistic model of the additional information is needed for obtaining the posterior distribution. For example, in survival analysis we have censoring information. One simple model of the censoring information is that censoring times are independent of survival times, and our results can be applied to this, since the process A is independent of the censoring information.

In the model, we assume that $Y(t)$ is predictable with respect to $\mathscr{F}_t^N$, which means that $Y(t)$ can be reconstructed from the observation of N. In fact, for any predictable process Y with respect to $\mathscr{F}_t^N$, there always exists a sequence of measurable functions $\{\phi_n\colon [0,1]^{n+1} \to [0,1]\}$ such that $Y(t) = \phi(t\colon N)$, where

$$\phi(t\colon N) = \phi_0(t) + \sum_{n=1}^{\infty} \phi_n(t\colon \mathbf{S_n}) I(S_n < t \le S_{n+1}), \tag{3.2}$$

$\mathbf{S_n} = (S_1, \dots, S_n)$, and each S_i represents the time of the i-jump of N. For proof, see Lemma III.1.29 in Jacod and Shiryaev (1987).

In this section, we derive the posterior distribution of A given N when A is a Lévy process on $(\mathscr{A}, \Sigma_{\mathscr{A}})$ with the compensator ν in (2.8). First, in Theorem 3.1 we derive the posterior distribution of A when the prior process is an extended compound Poisson process by direct calculation, and it is proved in Theorem 3.2 that the result obtained in Theorem 3.1 still holds for any Lévy process priors defined on $(\mathscr{A}, \Sigma_{\mathscr{A}})$.

In this section, we will make extensive use of the distribution of $\mathbf{S_n}$, which are defective. That is, $\Pr\{S_i \in [0,1]\} < 1$. To avoid unnecessary notational complication, we assume that S_i is defined on $[0,1] \cup \{\infty\}$, where $\Pr\{S_i = \infty\} = 1 - \Pr\{S_i \in [0,1]\}$.

Since we use $\mathscr{F}_t^N$ for the filtration of N, without loss of generality we assume that the probability space $(\Omega, \mathscr{F}, P)$ is the same as $(\mathscr{A} \times \mathscr{N}, \Sigma_{\mathscr{A}} \otimes \Sigma_{\mathscr{N}}, P)$ where $\mathscr{N}$ is the space of right-continuous step functions defined on $[0,1]$ with value 0 at $t = 0$ and with a finite number of jumps, each positive and of size 1, and $\Sigma_{\mathscr{N}}$ is the σ-field generated by the Borel cylinder sets. We denote the marginal distributions of A and N by P_A and P_N, respectively.

THEOREM 3.1. *For a given A, let N be a counting process defined in* (3.1). *Suppose that a priori A is a Lévy process on $(\mathscr{A}, \Sigma_{\mathscr{A}})$ with a compensator ν in* (2.8) *and we observe $N(1) = n$ and $\mathbf{S_n} = \mathbf{s_n}$. If $\nu([0,1] \times [0,1]) < \infty$, then the posterior process A given N is again a Lévy process on $(\mathscr{A}, \Sigma_{\mathscr{A}})$ with the compensator*

$$
\begin{aligned}
\nu^p([0,t] \times D) = {} & \int_0^t \int_D (1 - Y(s)x)\, dF_s(x)\, ds \\
& + \sum_{u_i \in \mathbf{s_n}} c_1(u_i)^{-1} \int_D Y(u_i)x\, dH_i(x) I(u_i \le t) \\
& + \sum_{u_i \notin \mathbf{s_n}} c_2(u_i)^{-1} \int_D (1 - Y(u_i)x)\, dH_i(x) I(u_i \le t) \\
& + \sum_{s_i \notin \mathbf{U}} c_3(s_i)^{-1} \int_D Y(s_i)x\, dF_{s_i}(x) I(s_i \le t)
\end{aligned}
\tag{3.3}
$$

for $t \in [0,1]$ and $D \in \mathscr{B}[0,1]$, where

$$
c_1(u_i) = \int_0^1 Y(u_i)x\, dH_i(x),
$$

$$
c_2(u_i) = \int_0^1 (1 - Y(u_i)x)\, dH_i(x)
$$

and

$$
c_3(s) = \int_0^1 Y(s)x\, dF_s(x).
$$

PROOF. Without loss of generality, we assume that $\int_0^1 dF_t(x) < \infty$ for all $t \in [0,1]$. Let $\lambda(t) = \int_0^1 dF_t(x)$ and define distribution functions $G_t(x)$ on $[0,1]$ by $G_t(x) = F_t(x)/\lambda(t)$. Consider a sequence of nondecreasing positive random variables $\{T_n\}$ such that $J(t) = \sum I(T_n \le t)$ is a Poisson process with the intensity function $\lambda(t)$. Let $\{\xi_n\}$ be a sequence of random variables on $[0,1]$ such that for a given $\{T_n\}$, ξ_n are independent random variables with distribution G_{T_n}. With the fixed set of times of discontinuity $\mathbf{U} = \{u_1, \ldots, u_l\}$, define a process $A(t)$ by

$$A(t) = \sum_{n=1}^{\infty} \xi_n I(T_n \le t) + \sum_{i=1}^{l} \eta_i I(u_i < t), \tag{3.4}$$

where η_i are independent random variables with distributions H_i. Then the process $A(t)$ is a Lévy process with the compensator ν in (2.8).

Let $\mathbf{T_m} = (T_1, \dots, T_m)$ and let $\mathbf{Z_m} = (\xi_1, \dots, \xi_m, \eta_1, \dots, \eta_l)$. For a given N, let $Q(\cdot : N)$ be a probability measure on $(\mathscr{A}, \Sigma_{\mathscr{A}})$ induced from a Lévy process whose compensator is given by (3.3). To prove the theorem, it suffices to show that

$$P(B \cap C) = \int_C Q(B : N)\, dP_N$$

for any $B \in \Sigma_{\mathscr{A}}$ and $C \in \Sigma_{\mathscr{N}}$. However, by Lemma III.1.29 in Jacod and Shiryaev (1988), it suffices to consider only the sets B and C of the forms

$$B = \{J(1) = m, \mathbf{T_m} \le \mathbf{t_m},\ \mathbf{Z_m} \le \mathbf{z_m}\}$$

and

$$C = \{N(1) = n,\ \mathbf{S_n} \le \mathbf{s_n}\}$$

for any given vectors $\mathbf{t_m} \in [0,1]^m$, $\mathbf{z_m} \in [0,1]^{m+l}$ and $\mathbf{s_n} \in [0,1]^n$, where $\mathbf{T_m} \le \mathbf{t_m}$ is defined to be $T_i \le t_i$ for all $i = 1, \dots, m$. We define $\mathbf{Z_m} \le \mathbf{z_m}$ and $\mathbf{S_n} \le \mathbf{s_n}$ similarly.

We consider only the case when $N(1) = 1$. Extension of the results to $N(1) > 1$ is straightforward. Suppose $J(1) = m$ with $\mathbf{T_m} = \mathbf{t_m}$ and $\mathbf{Z_m} = \mathbf{z_m}$. Then we can write

$$\begin{aligned} \Pr\{N(1) = 1,\ S_1 \le s_1 \mid A\} &= \sum_{k=1}^{m} \Pr\{N(1) = 1,\ S_1 = t_k \mid A\} I(t_k \le s_1) \\ &\quad + \sum_{k=1}^{l} \Pr\{N(1) = 1,\ S_1 = u_k \mid A\} I(u_k \le s_1) \end{aligned} \tag{3.5}$$

and in view of (2.2) we have

$$\Pr\{N(1) = 1,\ S_1 = t_k \mid A\} = \Phi(\mathbf{z_m}, \mathbf{t_m}, t_k)$$

and

$$\Pr\{N(1) = 1,\ S_1 = u_k \mid A\} = \Phi(\mathbf{z_m}, \mathbf{t_m}, u_k),$$

where

$$\begin{aligned} \Phi(\mathbf{z_m}, \mathbf{t_m}, s) = {} & \prod_{i=1}^{m} (1 - \phi_0(t_i) z_i)^{I(t_i < s)} (1 - \phi_1(t_i : s) z_i)^{I(t_i > s)} (\phi_0(t_i) z_i)^{I(t_i = s)} \\ & \times \prod_{i=1}^{l} (1 - \phi_0(u_i) z_{m+i})^{I(u_i < s)} (1 - \phi_1(u_i : s) z_{m+i})^{I(u_i > s)} \\ & \times (\phi_0(u_i) z_{m+i})^{I(u_i = s)}. \end{aligned}$$

For the marginal distribution of A, we can write

$$\begin{aligned}
&\Pr\{J(1)=m,\ (\mathbf{T_m},\mathbf{Z_m})\le(\mathbf{t_m},\mathbf{z_m})\}\\
&\quad=\Pr\{J(1)=m\}\Pr\{\mathbf{T_m}\le\mathbf{t_m}|J(1)=m\}\\
&\qquad\times\Pr\{\mathbf{Z_m}\le\mathbf{z_m}|\mathbf{T_m}=\mathbf{t_m},\ J(1)=m\}.
\end{aligned}\tag{3.6}$$

By the definition of the process J, we have

$$\Pr\{J(1)=m\}=\frac{\Lambda(1)^m}{m!}e^{-\Lambda(1)},\tag{3.7}$$

where $\Lambda(t)=\int_0^t\lambda(v)\,dv$. Also, it is well known that conditional on $J(1)=m$

$$\Pr\{\mathbf{T_m}\le\mathbf{t_m}|J(1)=m\}=m!\Lambda(1)^{-m}\int_0^{t_1}\int_{v_1}^{t_2}\cdots\int_{v_{m-1}}^{t_m}\prod_{i=1}^m\lambda(v_i)\,dv_i.\tag{3.8}$$

Also, it is clear from the definition of the process A that

$$\begin{aligned}
&\Pr\{\mathbf{Z_m}\le\mathbf{z_m}|\mathbf{T_m}=\mathbf{t_m},\ J(1)=m\}\\
&\quad=\int_0^{z_1}\cdots\int_0^{z_m}\int_0^{z_{m+1}}\cdots\int_0^{z_{m+l}}\prod_{i=1}^m dG_{t_i}(w_i)\prod_{i=1}^l dH_i(w_{m+i}).
\end{aligned}\tag{3.9}$$

Combining (3.7), (3.8) and (3.9), we can get the marginal distribution of A, and in turn by integrating out (3.5) with the marginal distribution of A, we have

$$\begin{aligned}
&\Pr\{A\in B,\ N\in C\}\\
&\quad=\int_B\Pr\{N(1)=1,\ S_1\le s_1\mid A\}\,dP_A\\
&\quad=e^{-\Lambda(1)}\sum_{k=1}^m\bigg[\int_{\mathbf{O(t_m)}\times[\mathbf{0},\mathbf{z_m}]}\Phi(\mathbf{w_m},\mathbf{v_m},v_k)\\
&\qquad\qquad\times I(v_k\le s_1)\,\mathbf{dG_{v_m}}(\mathbf{w_m})\,\mathbf{d\Lambda}(\mathbf{v_m})\bigg]\\
&\qquad+e^{-\Lambda(1)}\sum_{k=1}^l\bigg[\int_{\mathbf{O(t_m)}\times[\mathbf{0},\mathbf{z_m}]}\Phi(\mathbf{w_m},\mathbf{v_m},u_k)\\
&\qquad\qquad\times I(u_k\le s_1)\,\mathbf{dG_{v_m}}(\mathbf{w_m})\,\mathbf{d\Lambda}(\mathbf{v_m})\bigg],
\end{aligned}\tag{3.10}$$

where

$$\begin{aligned}
\mathbf{dG_{v_m}}(\mathbf{w_m})&=\prod_{i=1}^m dG_{v_i}(w_i)\prod_{i=1}^l dH_i(w_{m+i}),\\
\mathbf{d\Lambda}(\mathbf{v_m})&=\prod_{i=1}^m\lambda(v_i)\,dv_i,\\
\mathbf{O(t_m)}&=\{\mathbf{v_m}\in[0,1]^m\colon\ \mathbf{v_m}\le\mathbf{t_m}\text{ and }v_1\le\cdots\le v_m\}
\end{aligned}$$

and

$$[\mathbf{0}, \mathbf{z_m}] = \{\mathbf{w_m} \in [0,1]^{m+l} \colon \mathbf{w_m} \le \mathbf{z_m}\}.$$

The proof would be done if we show that $\int_c Q(B:N)\,dP_N$ is the same as (3.10). Suppose that the distribution of A conditional on $N(1) = 1$ and $S_1 = s_1$ is $Q(\cdot : N)$. Then

$$Q(B:N) = \Pr\{J(1) = m,\ (\mathbf{T_m}, \mathbf{Z_m}) \le (\mathbf{t_m}, \mathbf{z_m}) \mid N(1) = 1,\ S_1 = s_1\}.$$

First, consider the case of $s_1 = u_k$. Then in view of $Q(\cdot : N)$, we have

$$\Pr\{J(1) = m | N = 1,\ S_1 = u_k\} = \frac{B(1:u_k)^m}{m!} \exp(-B(1:u_k)), \tag{3.11}$$

where

$$B(t:u) = B_0(t:u) + B_1(t:u),$$
$$B_0(t:u) = \int_0^{t\wedge u} \lambda(v)(1 - \phi_0(v)\,d(v))\,dv,$$
$$B_1(t:u) = \int_u^{t\vee u} \lambda(v)(1 - \phi_1(v:u)\,d(v))\,dv$$

and $d(s) = \int_0^1 x\,dG_s(x)$. Also, we have

$$\begin{aligned} &\Pr\{\mathbf{T_m} \le \mathbf{t_m} | J(1) = m,\ N(1) = 1,\ S_1 = u_k\} \\ &\quad = \frac{B(1:u_k)^{-m}}{m!} \int_{\mathbf{O(t_m)}} \prod_{i=1}^m dB(v_i : u_k) \end{aligned} \tag{3.12}$$

and

$$\begin{aligned} &\Pr\{\mathbf{Z_m} \le \mathbf{z_m} | \mathbf{T_m} = \mathbf{t_m},\ J(1) = m,\ N(1) = 1,\ S_1 = u_k\} \\ &\quad = \int_{[\mathbf{0}, \mathbf{z_m}]} \Phi(\mathbf{w_m}, \mathbf{t_m}, u_k)\,\mathbf{dG_{t_m}}(\mathbf{w_m}) \\ &\qquad \times \left(\int_{[\mathbf{0}, \mathbf{1}]} \Phi(\mathbf{w_m}, \mathbf{t_m}, u_k)\,\mathbf{dG_{t_m}}(\mathbf{w_m}) \right)^{-1}, \end{aligned} \tag{3.13}$$

where $\mathbf{1}$ is a vector of length $m+l$ whose elements are all 1. Combining (3.11), (3.12) and (3.13), we have

$$\begin{aligned} &\Pr\{J(1) = m, (\mathbf{T_m}, \mathbf{Z_m}) \le (\mathbf{t_m}, \mathbf{z_m}) \mid N(1) = 1,\ S_1 = u_k\} \\ &\quad = \exp(-B(1:s_1)) \int_{\mathbf{O(t_m)} \times [\mathbf{0}, \mathbf{z_m}]} \Phi(\mathbf{w_m}, \mathbf{v_m}, u_k)\,\mathbf{dG_{v_m}}(\mathbf{w_m})\,\mathbf{d\Lambda}(\mathbf{v_m}) \\ &\qquad \times \left[\phi_0(u_k) e_k \prod_{i=1}^{k-1} (1 - \phi_0(u_i) e_i) \prod_{i=k+1}^{l} (1 - \phi_1(u_i : u_k) e_i) \right]^{-1}, \end{aligned} \tag{3.14}$$

where $e_i = \int_0^1 x\,dH_i(x)$.

Next, consider the case of $s_1 \notin \mathbf{U}$. Then conditional on $S_1 = s_1$ and $J(1) = m$, with probability 1 there exists a $k \in \{1, \ldots, m\}$ such that $T_k = s_1$. So we can write

$$
\begin{aligned}
&\Pr\{J(1) = m, (\mathbf{T_m}, \mathbf{Z_m}) \le (\mathbf{t_m}, \mathbf{z_m}) \mid N(1) = 1,\ S_1 = s_1\} \\
&\quad = \sum_{k=1}^{m} \bigl[\Pr\{J(1) = m, T_k = s_1 | N(1) = 1,\ S_1 = s_1\} \\
&\qquad \times \Pr\bigl\{(\mathbf{T_m}, \mathbf{Z_m}) \le (\mathbf{t_m}, \mathbf{z_m}) | J(1) = m, \\
&\qquad\qquad N(1) = 1,\ T_k = S_1 = s_1\bigr\}\bigr].
\end{aligned} \tag{3.15}
$$

Here, it is easy to show that

$$
\begin{aligned}
&\Pr\{J(1) = m,\ T_k = s_1 \mid N(1) = 1,\ S_1 = s_1\} \\
&\quad = \frac{B_0(s_1 : s_1)^{k-1}}{(k-1)!} \exp(-B_0(s_1 : s_1)) \frac{B_1(1 : s_1)^{m-k}}{(m-k)!} \exp(-B_1(1 : s_1))
\end{aligned} \tag{3.16}
$$

and

$$
\begin{aligned}
&\Pr\{\mathbf{T_m} \le \mathbf{t_m} | J(1) = m,\ N(1) = 1,\ T_k = S_1 = s_1\} \\
&\quad = \frac{(k-1)!(m-k)!}{B_0(s_1 : s_1)^{k-1} B_1(1 : s_1)^{m-k}} \\
&\qquad \times \int_{\mathbf{O}_k(\mathbf{t_m} : s_1)} \prod_{i=1}^{k-1} dB_0(v_i : s_1) \prod_{i=k+1}^{m} dB_1(v_i : s_1),
\end{aligned} \tag{3.17}
$$

where

$$
\begin{aligned}
\mathbf{O}_k(\mathbf{t_m} : s_1) = \bigl\{\mathbf{v_m} \in [0,1]^m \colon \mathbf{v_m} \le \mathbf{t_m} \quad &\text{and} \\
&v_1 \le \cdots \le v_{k-1} \le v_k = s_1 \le v_{k+1} \le \cdots \le v_m\bigr\}.
\end{aligned}
$$

Also it is true that

$$
\begin{aligned}
&\Pr\{\mathbf{Z_m} \le \mathbf{z_m} | \mathbf{T_m} = \mathbf{t_m},\ J(1) = m,\ N(1) = 1,\ T_k = S_1 = s_1\} \\
&\quad = \int_{[\mathbf{0}, \mathbf{z_m}]} \Phi(\mathbf{w_m}, \mathbf{t_m}, t_k)\, \mathbf{dG_{t_m}}(\mathbf{w_m}) \\
&\qquad \times \left(\int_{[\mathbf{0}, \mathbf{1}]} \Phi(\mathbf{w_m}, \mathbf{t_m}, t_k)\, \mathbf{dG_{t_m}}(\mathbf{w_m}) \right)^{-1}.
\end{aligned} \tag{3.18}
$$

Now combining (3.15), (3.16), (3.17) and (3.18) with (3.14), we have

$$\begin{aligned}
&\Pr\{J(1)=m, (\mathbf{T_m},\mathbf{Z_m}) \le (\mathbf{t_m},\mathbf{z_m}) | N(1)=1,\ S_1=s_1\}\\
&\quad = I(s_1 \notin \mathbf{U}) \exp(-B(1:s_1))\\
&\qquad \times \sum_{i=1}^{m}\Bigg[\int_{\mathbf{O}_k(\mathbf{t_m}: s_1)\times[\mathbf{0},\,\mathbf{z_m}]} \Phi(\mathbf{w_m},\mathbf{v_m}, v_i)\, \mathbf{dG_{v_m}}(\mathbf{w_m})\, \mathbf{d\Lambda}_{(i)}(\mathbf{v_m})\\
&\qquad\qquad \times \Bigg\{ v(s_1)\phi_0(s_1)\prod_{i=1}^{l}(1-\phi_0(u_i)e_i)^{I(u_i<s_1)}\\
&\qquad\qquad\qquad \times (1-\phi_1(u_i:s_1)e_i)^{I(u_i>s_1)}\Bigg\}^{-1}\Bigg] \qquad (3.19)\\
&\qquad + e^{-B(1:s_1)}\sum_{i=1}^{l} I(u_i=s_1)\\
&\qquad \times \Bigg[\int_{\mathbf{O(t_m)}\times[\mathbf{0},\mathbf{z_m}]} \Phi(\mathbf{w_m},\mathbf{v_m},u_i)\,\mathbf{dG_{v_m}}(\mathbf{w_m})\,\mathbf{d\Lambda}(\mathbf{dv_m})\\
&\qquad\quad \times \Bigg\{\phi_0(u_i)e_i\prod_{k=1}^{i-1}(1-\phi_0(u_k)e_k)\prod_{k=i+1}^{l}(1-\phi_1(u_k:u_i)e_i)\Bigg\}^{-1}\Bigg],
\end{aligned}$$

where

$$\mathbf{d\Lambda}_{(i)}(\mathbf{v_m}) = \prod_{j=1,\ j\ne i}^{n} \lambda(v_j)\,dv_j.$$

For the marginal distribution of N, first consider $\Pr\{S_1 > s_1\}$. By the definition of the process, we can write

$$\begin{aligned}
\Pr\{S_1 > s\} &= \mathrm{E}\bigg\{\prod_{t\in[0,\,s]}(1-\phi_0(t)\,dA(t))\bigg\}\\
&= \prod_{t\in[0,\,s]}(1-\phi_0(t)\,dA^0(t)),
\end{aligned}$$

where $A^0(t)=\mathrm{E}(A(t))$. Therefore

$$\Pr\{S_1 > s\} = \bigg[\prod_{u_i\le s}(1-\phi_0(u_i)e_i)\bigg]\exp\bigg\{-\int_0^s \lambda(v)\phi_0(v)\,d(v)\,dv\bigg\}.$$

Since $A(t)$ has independent increments,

$$\begin{aligned}
&\Pr\{S_2 > s_2 \mid S_1 = s_1\}\\
&\quad = \bigg[\prod_{s_1<u_i\le s_2}(1-\phi_1(u_i:s_1)e_i)\bigg]\exp\bigg\{-\int_{s_1}^{s_2}\lambda(v)\phi_1(v:s_1)\,d(v)\,dv\bigg\}.
\end{aligned}$$

Hence

$$\begin{aligned}
&\Pr\{N(1)=1,\ S_1 \le s_1\}\\
&\quad = \Pr\{S_2 > 1,\ S_1 \le s_1\}\\
&\quad = \int_0^{s_1}\Bigg[\prod_{v<u_i\le 1}(1-\phi_1(u_i:v)e_i)\\
&\qquad\qquad \times \exp\bigg\{-\int_v^1 \lambda(w)\phi_1(w:v)\,d(w)\,dw\bigg\}\Bigg]\,dP_1(v),
\end{aligned}$$

where $P_1(v) = \Pr\{S_1 \le v\}$. Consequently,

$$\begin{aligned}
&\Pr\{N(1)=1,\ S_1 \le s_1,\ S_1 \notin \mathbf{U}\}\\
&\quad = \int_0^{s_1}\lambda(s)\Bigg[\prod_{u_i<s}(1-\phi_0(u_i)e_i)\prod_{u_i>s}(1-\phi_1(u_i:s)e_i)\phi_0(s)\,d(s)\\
&\qquad\qquad \times \exp\bigg\{-\int_0^1 \lambda(v)(\phi_0(v)^{I(v\le s)}\phi_1(v:s)^{I(v>s)})\,d(v)\,dv\bigg\}\Bigg]\,ds.
\end{aligned} \tag{3.20}$$

In a similar manner we can show that

$$\begin{aligned}
&\Pr\{N(1)=1, S_1 \le s_1, S_1 \in \mathbf{U}\}\\
&\quad = \sum_{k=1}^{m}\Bigg[I(u_k \le s_1)\phi_0(u_k)e_k\prod_{j=1}^{k-1}(1-\phi_0(u_i)e_i)\prod_{j=k+1}^{m}(1-\phi_1(u_i:u_k)e_i)\\
&\qquad\qquad \times \exp\bigg\{-\int_0^1 \lambda(v)(\phi_0(v)^{I(v<u_k)}\phi_1(v:u_k)^{I(v>u_k)})\,d(v)\,dv\bigg\}\Bigg]
\end{aligned} \tag{3.21}$$

Integrating out (3.19) with (3.20) and (3.21) yields the same distribution as in (3.10), which completes the proof. □

Now, we proceed to drop the condition $\nu([0,1]\times[0,1]) < \infty$ in Theorem 3.1. If $\nu([0,1]\times[0,1]) = \infty$, F_s may not be a finite measure, in which case the process A jumps infinitely many times on a finite time interval. Before going further, we explain why we need a process with infinitely many jumps. Consider a random variable X on $[0,1]$ whose cumulative hazard function is A. In many situations a priori we want that the support of X covers $[0,1]$. Equivalently, we want

$$\prod_{i=1}^{n}\Pr\{X \in [t_{i-1}, t_i)|A\} > 0 \tag{3.22}$$

for any sequence $0 = t_0 < t_1 < \cdots < t_n = 1$. If A is random, we want that (3.22) holds with probability 1, which is impossible if A has only finitely many jumps.

The main idea of obtaining the posterior distribution is to approximate the Lévy process $A(t)$ with the compensator (2.8) by a sequence of extended compound Poisson processes $A_n(t)$ with the compensator ν^n given by (2.9). To complete the proof, we shall show that the sequence of the posterior distributions of A_n converges weakly to the posterior distribution of A in some sense. Unfortunately, the convergence of joint distributions does not always imply the convergence of the conditional distributions. Some sufficient conditions for the convergence of conditional distributions are available in Sweeting (1989). In our situation, however, those conditions in Sweeting (1989) are not applicable since the posterior distribution is not continuous on $\mathscr{N}$ due to the fixed times of discontinuity. To circumvent this difficulty, we will directly prove the convergence of the posterior distributions.

Throughout this section, we always assume that A_n and A are Lévy processes on $(\mathscr{A}, \Sigma_{\mathscr{A}})$ with the compensators (2.9) and (2.8), and given A_n and A, N_n and N are multiplicative intensity counting processes with cumulative intensity functions A_n and A and the predictable processes

$$Y_n(t) = \phi(t : N_n)$$

and

$$Y(t) = \phi(t : N),$$

where ϕ is given by (3.2).

LEMMA 1. *For all* $f \in B(\mathscr{N})$,

$$\mathrm{E}[f(N_n)] \to \mathrm{E}[f(N)], \tag{3.23}$$

where $B(\mathscr{N})$ *is a class of all bounded measurable functions on* $\mathscr{N}$.

PROOF. Since $N_n \to_{\mathscr{L}} N$ by Theorem A.1 in the Appendix and f is bounded, for a given $\varepsilon > 0$, we can choose some constants k and M such that

$$\mathrm{E}[f(N_n)I(N_n(1) \geq k)] < \varepsilon$$

and

$$\mathrm{E}[f(N)I(N(1) \geq k)] < \varepsilon$$

for all $n > M$. So it suffices to show that

$$\mathrm{E}[f(N_n)I(N_n(1) < k)] \to \mathrm{E}[f(N)I(N(1) < k)] \tag{3.24}$$

for all k. Since $f(N_n)I(N_n(1) < k)$ and $f(N)I(N(1) < k)$ are bounded measurable functions for $\sigma(S_1^n, \dots, S_k^n)$ and $\sigma(S_1, \dots, S_k)$, respectively, where S_i^n and S_i are the times of the ith-jump of N_n and N, respectively, (3.24) would be true if

$$\Pr\{(S_1^n, \dots, S_k^n) \in B\} \to \Pr\{(S_1, \dots, S_k) \in B\} \tag{3.25}$$

for all $B \in \mathscr{B}^k[0, 1]$ as $n \to \infty$.

We shall use mathematical induction to prove (3.25). Since the processes A_n and A have independent increments, it is easy to see that the cumulative hazard functions G_1^n and G_1 of S_1^n and S_1 are

$$G_1^n(t) = \int_0^t \phi_0(s)\, dA_n^0(s)$$

and

$$G_1(t) = \int_0^t \phi_0(s)\, dA^0(s),$$

where $A_n^0 = \mathrm{E}(A_n)$ and $A^0 = \mathrm{E}(A)$. Here,

$$A_n^0(t) = \int_0^t a_n(s)\, ds + \sum_{i=1}^{l} a_i I(u_i \le t)$$

and

$$A^0(t) = \int_0^t a(s)\, ds + \sum_{i=1}^{l} a_i I(u_i \le t),$$

where $a_n(s) = \int_{1/n}^1 u\, dF_s(u)$, $a(s) = \int_0^1 u\, dF_s(u)$ and $a_i = \int_0^1 u\, dH_i(u)$. Since $a_n \to a$, the bounded convergence theorem implies that

$$\sup_{B \in \mathscr{B}[0,1]} \left| \Pr\{S_1^n \in B\} - \Pr\{S_1 \in B\} \right| \to 0.$$

Suppose the assertion (3.25) is true for $k = m$. Define a sequence of probability measures $Q_n(\cdot : \mathbf{s_m})$ and $Q(\cdot : \mathbf{s_m})$ on $\mathscr{B}[0,1]$ indexed by $\mathbf{s_m} \in [0,1]^m$ by

$$Q_n(B : \mathbf{s_m}) = \Pr\{S_{m+1}^n \in B \mid \mathbf{S_m^n} = \mathbf{s_m}\}$$

and

$$Q(B : \mathbf{s_m}) = \Pr\{S_{m+1} \in B \mid \mathbf{S_m} = \mathbf{s_m}\},$$

where $\mathbf{S_m^n} = (S_1^n, \dots, S_m^n)$ and $\mathbf{S_m} = (S_1, \dots, S_m)$. Then it is easy to see that the cumulative hazard functions corresponding to $Q_n(\cdot : \mathbf{s_m})$ and $Q(\cdot : \mathbf{s_m})$ are G_{m+1}^n and G_{m+1}, where

$$G_{m+1}^n(t : \mathbf{s_m}) = \int_0^t \phi_{m+1}(s : \mathbf{s_m}) I(s_m \le s)\, dA_n^0(s)$$

and

$$G_{m+1}(t : \mathbf{s_m}) = \int_0^t \phi_{m+1}(s : \mathbf{s_m}) I(s_m \le s)\, dA^0(s).$$

Again the bounded convergence theorem implies

$$\sup_{B \in \mathscr{B}[0,1],\, \mathbf{s_m} \in [0,1]^m} \left| Q_n(B : \mathbf{s_m}) - Q(B : \mathbf{s_m}) \right| \to 0. \tag{3.26}$$

For $B^{m+1} \in \mathscr{B}^{m+1}[0,1]$, let $B(\mathbf{s_m}) = \{s\colon (s_1, \ldots, s_m, s) \in B^{m+1}\}$. Then, we have

$$\begin{aligned} &\left|\Pr\{\mathbf{S^n_{m+1}} \in B\} - \Pr\{\mathbf{S_{m+1}} \in B\}\right| \\ &\quad = \left|E_{P_n}(Q_n(B(\mathbf{S^n_m}):\mathbf{S^n_m})) - E_P(Q(B(\mathbf{S_m}):\mathbf{S_m}))\right| \\ &\quad \le \left|E_{P_n}(Q_n(B(\mathbf{S^n_m}):\mathbf{S^n_m}) - Q(B(\mathbf{S^n_m}):\mathbf{S^n_m}))\right| \\ &\qquad + \left|E_{P_n}(Q(B(\mathbf{S^n_m}):\mathbf{S^n_m})) - E_P(Q(B(\mathbf{S_m}):\mathbf{S_m}))\right|. \end{aligned} \tag{3.27}$$

The first term of the last inequality in (3.27) converges to 0 due to (3.26), and the second term also converges to 0 since $Q(B(\mathbf{S_m}):\mathbf{S_m})$ is a measurable function for $\sigma(S_1, \ldots, S_m)$. □

THEOREM 3.2. *Theorem* 3.1 *still holds when* $\nu([0,1]\times[0,1]) = \infty$.

PROOF. Let A_n be the extended compound Poisson process with the compensator ν^n defined by (2.9). By Theorem A.1 in the Appendix, we have

$$\begin{aligned} &\mathrm{E}\left[\exp\left\{-\int_0^1 g(t)\,dA_n(t) - f(t)\,dN_n(t)\right\}\right] \\ &\qquad \to \mathrm{E}\left[\exp\left\{-\int_0^1 g(t)\,dA(t) - f(t)\,dN(t)\right\}\right] \end{aligned} \tag{3.28}$$

for all $f, g \in C^+[0,1]$ where $C^+[0,1]$ is the class of all nonnegative continuous functions on $[0,1]$.

Define a sequence of probability measures $Q_n(\cdot : y)$ and $Q(\cdot : y)$ on $\mathscr{A}$ indexed by $y \in \mathscr{N}$ by

$$Q_n(G : N_n) = \Pr\{A_n \in G | N_n\}$$

and

$$Q(G : N) = \Pr\{\tilde{A} \in G | N\},$$

where $\tilde{A}$ is a Lévy process on $(\mathscr{A}, \Sigma_{\mathscr{A}})$ with the compensator ν in (3.3) for given N.

To prove the theorem, we shall show that

$$\begin{aligned} &\int_{\mathscr{N}}\int_{\mathscr{A}} \exp\left\{-\int_0^1 g(t)\,da(t) - f(t)\,dy(t)\right\} Q(da : y) P_N(dy) \\ &\qquad = \mathrm{E}\left[\exp\left\{-\int_0^1 g(t)\,dA(t) - f(t)\,dN(t)\right\}\right]. \end{aligned} \tag{3.29}$$

for all $f, g \in C^+[0,1]$. Let

$$\Phi_n(y) = \int_{\mathscr{A}} \exp\left\{-\int_0^1 g(t)\,da(t)\right\} Q_n(da : y)$$

and

$$\Phi(y) = \int_{\mathscr{A}} \exp\left\{ -\int_0^1 g(t)\, da(t) \right\} Q(da : y).$$

Because of (3.28), (3.29) is equivalent to

$$\tag{3.30} \begin{aligned} &\int_{\mathscr{N}} \exp\left\{ -\int_0^1 f(t)\, dy(t) \right\} \Phi_n(y) P_{N_n}(dy) \\ &\quad \to \int_{\mathscr{N}} \exp\left\{ -\int_0^1 f(t)\, dy(t) \right\} \Phi(y) P_N(dy). \end{aligned}$$

Now it can be shown that

$$\tag{3.31} \int_{\mathscr{N}} \exp\left\{ -\int_0^1 f(t)\, dy(t) \right\} (\Phi_n(y) - \Phi(y)) P_{N_n}(dy) \to 0.$$

The proof is in the Appendix. Also, by Lemma 1, we have

$$\tag{3.32} \int_{\mathscr{N}} \exp\left\{ -\int_0^1 f(t)\, dy(t) \right\} \Phi(y)(P_{N_n}(dy) - P_N(dy)) \to 0.$$

Combining (3.31) and (3.32) we conclude (3.30), which completes the proof. □

Let us next consider multiple sample paths. Let $N_1, \dots, N_n$ be i.i.d. multiplicative counting processes with a common cumulative intensity function A. Define a new process $N(t)$ by

$$N(t) = \sum_{k=1}^{n} N_k(t),$$

let q be the number of the distinct jumps of N and let $v_1, \dots, v_q$ be the distinct points at which the process N makes jumps. The following theorem can be proved by repeated application of Theorem 3.2, conditioning first on N_1, then on N_2 and so on.

THEOREM 3.3. *Let A be a Lévy process on $(\mathscr{A}, \Sigma_{\mathscr{A}})$ with a compensator ν,*

$$\nu([0,t] \times D) = \int_0^t \int_D dF_s(x)\, ds$$

for $t \in [0,1]$ and $D \in \mathscr{B}[0,1]$. Then the posterior distribution of A given $N_1, \dots, N_n$ is also a Lévy process on $(\mathscr{A}, \Sigma_{\mathscr{A}})$ with the compensator

$$\begin{aligned} \nu^p([0,t] \times D) = &\int_0^t \int_D \prod_{i=1}^{n} (1 - Y_i(s)x)\, dF_s(x)\, ds \\ &+ \sum_{k=1}^{q} c_k^{-1} \int_D \prod_{i=1}^{n} \eta(Y_i(v_k)x, \Delta N_i(v_k))\, dF_{v_k}(x) I(v_k \le t), \end{aligned}$$

for $t \in [0,1]$ and $D \in \mathscr{B}[0,1]$, where

$$\eta(x, y) = x^{I(y>0)} (1 - x)^{I(y=0)}$$

and

$$c_k = \int_0^1 \prod_{i=1}^{n} \eta(Y_i(v_k)x, \Delta N_i(v_k))\, dF_{v_k}(x).$$

REMARK. In Theorem 3.3, we assume that $\mathbf{U} = \varnothing$, which is needed only for the notational convenience and can be easily dropped out.

For multivariate counting processes, we can also obtain a similar result as in Theorem 3.3. Let $\mathbf{N}_1, \ldots, \mathbf{N}_n$ be i.i.d. k-variate counting processes with intensity processes $\mathbf{\Lambda}_1, \ldots, \mathbf{\Lambda}_n$, where $\mathbf{N}_i = (N_{i1}, \ldots, \mathbf{N}_{ik})$, $\mathbf{\Lambda}_i = (\Lambda_{i1}, \ldots, \Lambda_{ik})$ and Λ_{ij} can be represented by

$$\Lambda_{ij} = \int_0^t Y_{ij}(s)\, dA_j(s)$$

for some predictable processes Y_{ij} with respect to $\mathscr{F}^{\mathbf{N}_i}$. Suppose that the common cumulative intensity functions $\mathbf{A} = (A_1, \ldots, A_k)$ are independent Lévy processes on $(\mathscr{A}, \Sigma_{\mathscr{A}})$ with the compensators $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_k)$. Define a new process $\mathbf{N}_{\bullet} = (N_{\bullet 1}, \ldots, N_{\bullet k})$ by

$$N_{\bullet j}(t) = \sum_{i=1}^{n} N_{ij}(t)$$

and let q_j be the number of distinct jumps of $N_{\bullet j}$ and $v_{1j}, \ldots, v_{q_j j}$ be the distinct points at which the process $N_{\bullet j}$ makes a jump. We state the main result without proof. The proof can be done by modifying the proof of the univariate case.

THEOREM 3.4. *Suppose that for* $j = 1, \ldots, k$,

$$\nu_j([0, t] \times D) = \int_0^t \int_D dF_s^j(x)\, ds$$

for $t \in [0, 1]$ *and* $D \in \mathscr{B}[0, 1]$. *Then the posterior distributions of* A_j *given* $\mathbf{N}_1, \ldots, \mathbf{N}_n$, *are mutually independent Lévy processes on* $(\mathscr{A}, \Sigma_{\mathscr{A}})$ *with the compensator* ν_j^p,

$$\begin{aligned}\nu_j^p([0, t] \times D) = &\int_0^t \int_D \prod_{i=1}^{n} (1 - Y_{ij}(s)x)\, dF_s^j(x)\, ds \\ &+ \sum_{i=1}^{q_j} c_{ij}^{-1} \int_D \prod_{m=1}^{n} \eta(Y_{mj}(v_{ij})x, \Delta N_{mj}(v_{ij}))\, dF_{v_{ij}}^j(x) I(v_{ij} \le t),\end{aligned}$$

for $t \in [0, 1]$ *and* $D \in \mathscr{B}[0, 1]$, *where*

$$\eta(x, y) = x^{I(y>0)}(1 - x)^{I(y=0)}$$

and

$$c_{ij} = \int_0^1 \prod_{m=1}^{n} \eta(Y_{mj}(v_{ij})x, \Delta N_{mj}(v_{ij}))\, dF_{v_{ij}}^j(x).$$

4. Examples. Let A be a Lévy process on $(\mathscr{A}, \Sigma_{\mathscr{A}})$ with the compensator ν in (2.8). Then the prior mean of A becomes

$$\mathrm{E}[A(t)] = \int_0^t \int_0^1 x\, dF_s(x)\, ds + \sum_{k=1}^{l} I(u_k \le t) \int_0^1 x\, dH_k(x).$$

If we define distribution functions $G_t(x)$ on $[0, 1]$ by $G_t(x) = \int_0^x y\, dF_t(y)/\lambda(t)$, where $\lambda(t) = \int_0^1 y\, dF_t(y)$, then the prior mean becomes

$$\mathrm{E}[A(t)] = \int_0^t \lambda(s)\, ds + \sum_{k=1}^{l} I(u_k \le t) e_k,$$

where $e_k = \int_0^1 x\, dH_k(x)$. Also the prior variance of the process A is

$$\mathrm{Var}[A(t)] = \int_0^t \int_0^1 x\, dG_s(x) \lambda(s)\, ds + \sum_{k=1}^{l} I(u_k \le t) \biggl(\int_0^1 x^2\, dH_k(x) - e_k^2 \biggr).$$

So to choose a Lévy process corresponding to prior information given by mean and variance, we simply choose a mean function $\lambda(s)$ and distribution functions G_t and H_k appropriately. Note that the distributions G_t do not affect the prior mean given $\lambda(t)$, and they control the variance of the process through their means. For example, let us consider the beta process defined by Hjort (1990). We can redefine the beta process with parameters $A(t)$ and $c(t)$ by a Lévy process with $\lambda(t) = dA(t)/dt$ and $G_t(x)$ being a beta distribution function with parameters 1 and $c(t)$, that is,

$$G_t(x) = \int_0^x c(t)(1-y)^{c(t)-1}\, dy.$$

In this view, the name "beta process" fits well.

Now we present several examples of practical importance. For simplicity, we always assume that $\mathbf{U} = \varnothing$.

EXAMPLE 1. *Right-censored survival data.* Let $X_1, \dots, X_n$ be i.i.d. random survival times given a cumulative hazard function A. Assume that $(T_1, \delta_1), \dots, (T_n, \delta_n)$ is observed where $T_i = \min(X_i, C_i)$, $\delta_i = I(X_i \le C_i)$ and $C_1, \dots, C_n$ are the censoring times. If we let

$$N_i(t) = I(T_i \le t \text{ and } \delta_i = 1),$$

then the cumulative intensity process $\Lambda_i(t)$ of $N_i(t)$ with respect to $\mathscr{F}_t$ becomes

$$\Lambda_i(t) = \int_0^t Y_i(s)\, dA(s), \tag{4.1}$$

where $Y_i(s) = I(T_i \ge s)$ and

$$\mathscr{F}_t = \mathscr{F}_t^{N_i} \vee \mathscr{F}_t^{C_i},$$

where $\mathscr{F}_t^{C_i} = \sigma(I(C_i \le s),\ s \le t)$.

Under the assumption of independent censoring (i.e., X_i and C_i are independent), we can consider (4.1) as the intensity process of N_i with respect to $\mathscr{F}_t^{N_i}$ conditional on $\mathscr{F}_t^{C_i}$. Hence all necessary calculations can be done conditionally. If the prior distribution of A is a beta process with parameters $A^0(t)$ and $c(t)$, the posterior distribution is also a beta process with parameters

$$A^p(t) = \int_0^t \frac{c(s)}{c(s)+Y_\bullet(s)}\,dA^0(s) + \frac{1}{c(s)+Y_\bullet(s)}\,dN_\bullet(s) \tag{4.2}$$

and

$$c^p(t) = c(t) + Y_\bullet(t), \tag{4.3}$$

where $N_\bullet(t) = \sum_{i=1}^n N_i(t)$ and $Y_\bullet(t) = \sum_{i=1}^n Y_i(t)$, which was already obtained by Hjort (1990).

REMARK. In usual models of survival data, A is defined on $[0,\infty)$ instead of on the compact set $[0,t]$, in which case we need to put a prior mass on the space of intensity functions defined on $[0,\infty)$. However, this can be done by assigning the prior mass on the compact sets $[0,t]$ for all $t>0$ in a certain consistent way. For example, we can define a beta process A on $[0,\infty)$ with parameters A^0 and c by a stochastic process such that on any compact set $[0,t]$, A is a beta process with parameters A^0 and c. The process A is well defined on $[0,\infty)$ because the beta process has independent increments and so all finite dimensional distributions are uniquely determined which satisfy the consistency condition of Kolmogorov's existence theorem [see, for example, Shiryaev (1991)]. With this definition, we can conclude that the posterior distribution of A on $[0,\infty)$ is also a beta process with parameters A^p and c^p in (4.2) and (4.3), since the posterior distribution of A on $[0,t]$ is a beta process with the same parameters for all $t>0$.

EXAMPLE 2. *Left-truncated, right-censored survival data.* Left truncation is very common in fields like demography and epidemiology where observations consist of an entrance time (left truncation), an exit time and a type of exit (either death or right censoring). See Anderson, Borgan, Gill and Keiding (1993) and references therein. Let $X_1,\dots,X_n$ be i.i.d. survival times with cumulative hazard function A. Let $E_1,\dots,E_n$ and $C_1,\dots,C_n$ be entrance times and censoring times, respectively, which are independent of X_i's. Now, define $M_i(t) = I(X_i \le t)$ and

$$N_i(t) = \int_0^t I(E_i < s \le C_i)\,dM_i(s).$$

Then conditional on $E_1,\dots,E_n$ and $C_1,\dots,C_n$, N_i is a multiplicative counting process with the cumulative intensity process

$$\Lambda(t) = \int_0^t Y_i(s)\,dA(s),$$

where $Y_i(s) = I(E_i < s \le C_i)$. Therefore, if the prior distribution of A is a beta process with parameters $A^0(t)$ and $c(t)$, the posterior distribution is also a beta process with parameters

$$A^p(t) = \int_0^t \frac{c(s)}{c(s) + Y_\bullet(s)}\, dA^0(s) + \frac{1}{c(s) + Y_\bullet(s)}\, dN_\bullet(s)$$

and

$$c^p(t) = c(t) + Y_\bullet(t),$$

where $N_\bullet(t) = \sum_{i=1}^n N_i(t)$ and $Y_\bullet(t) = \sum_{i=1}^n Y_i(t)$.

EXAMPLE 3. *Poisson process.* Let $N_1, \dots, N_n$ be i.i.d. Poisson processes with the mean function A. In this case, $Y_i(t) \equiv 1$ for $i = 1, \dots, n$. If the prior distribution of A is a beta process with parameters $A^0(t)$ and $c(t)$, then the posterior distribution is again a beta process with parameters

$$A^p(t) = \int_0^t \frac{c(s)}{c(s) + n}\, dA^0(s) + \frac{1}{c(s) + n}\, dN_\bullet(s)$$

and

$$c^p(t) = c(t) + n.$$

EXAMPLE 4. *Censored Poisson process.* A nonparametric Bayesian problem for a censored Poisson process model was considered by Lo (1992) as follows. Let $M_1, \dots, M_n$ be i.i.d. Poisson processes with the common mean function A. Define risk functions Y_i such that

$$\begin{aligned} Y_i(t) &= 1 \text{ if the process } M_i \text{ is observed at time } t, \\ &= 0 \text{ if the process } M_i \text{ is not observed at time } t \end{aligned}$$

and Y_i is assumed to be left continuous and independent of M_i. A censored Poisson process N_i can be defined by

$$N_i = \int_0^t Y_i(s)\, dM_i(s).$$

In this case, for a given Y_i, the cumulative intensity process Λ_i of N_i is

$$\Lambda_i(t) = \int_0^t Y_i(s)\, dA(s).$$

Hence if the prior distributions of A is a beta process with parameters $A^0(t)$ and $c(t)$, then the posterior distribution is a beta process with parameters

$$A^p(t) = \int_0^t \frac{c(s)}{c(s) + Y_\bullet(s)}\, dA^0(s) + \frac{1}{c(s) + Y_\bullet(s)}\, dN_\bullet(s)$$

and

$$c^p(t) = c(t) + Y_\bullet(t).$$

EXAMPLE 5. *Counter model.* A counter is a device for detecting and registering instantaneous pulse-type signals. A familiar example is the electron multiplier. All physically realizable counters are imperfect, incapable of detecting all signals that enter their detection chambers. After a particle or signal is registered, a counter must recuperate or renew in preparation for the next arrival. This readjustment period is called *locked time*. See Pyke (1958), Karlin and Taylor (1975) and Kao and Smith (1993) for details. Suppose that the arrival process M follows a Poisson process with mean function A and locked times $C_1, C_2, \dots$ are i.i.d. positive random variables which are independent of M. Then the observable process N is

$$N(t) = \int_0^t Y(s)\, dM(s),$$

where $Y(s) = I(s \geq S_{N(s-)} + C_{N(s-)})$ and S_i is the time of the ith jump of N. The counter model is also used for disease event history data where subjects under observation experience a certain disease repeatedly, and once the disease occurs it lasts for a certain period of time in which no new disease can occur.

Let $M_1, \dots, M_n$ be i.i.d. Poisson processes with mean function A. Let $C_{i,1}, C_{i,2}, \dots$ be i.i.d. positive random variables which are independent of M_i's. Now, define a process $N_i(t)$ by

$$N_i(t) = \int_0^t Y_i(s)\, dM_i(s),$$

where

$$Y_i(s) = I(s \geq S_{i,\, N_i(s-)} + C_{i,\, N_i(s-)}).$$

Then conditional on C_{ij}'s, N_i is a multiplicative counting process with the cumulative intensity process

$$\Lambda_i(t) = \int_0^t Y_i(s)\, dA(s).$$

So if the prior distribution of A is a beta process with parameters $A^0(t)$ and $c(t)$, then the posterior distribution is a beta process with parameters

$$A^p(t) = \int_0^t \frac{c(s)}{c(s) + Y_\bullet(s)}\, dA^0(s) + \frac{1}{c(s) + Y_\bullet(s)}\, dN_\bullet(s)$$

and

$$c^p(t) = c(t) + Y_\bullet(t).$$

REMARK. The counter model and the censored Poisson process are similar in the sense that the two processes are thinned Poisson processes. However, the thinning procedure in the counter model depends on the underlying process while it is independent of the underlying process in the censored Poisson process. It is worth noting that Bayesian estimators of any thinned Poisson processes can be obtained by applying our results as long as the thinning procedure is predictable.

EXAMPLE 6. *Finite state space Markov chain.* Suppose that $X_1(t), \dots, X_n(t)$ are i.i.d. Markov processes moving around in the state space $\{1, \dots, k\}$ with cumulative transition intensity functions A_{jl}. Introduce

$$N_{ijl}(t) = \text{number of observed transitions } j \to l \text{ of } X_i \text{ during [0,t]}$$

and

$$Y_{ij}(t) = I(X_i(t-) = j).$$

Then the cumulative intensity process Λ_{ijl} of N_{ijl} becomes

$$\Lambda_{ijl}(t) = \int_0^t Y_{ij}(s)\, dA_{jl}(s).$$

Therefore, if the prior distributions of A_{jl} are independent beta processes with parameters A^0_{jl} and c_{jl}, the posterior distributions are also independent beta processes with parameters

$$A^p_{jl}(t) = \int_0^t \frac{c_{jl}(s)}{c_{jl}(s) + Y_{\bullet j}(s)}\, dA^0_{jl}(s) + \frac{1}{c_{jl}(s) + Y_{\bullet j}(s)}\, dN_{\bullet jl}(s)$$

and

$$c^p_{jl}(t) = c_{jl} + Y_{\bullet j}(t),$$

where $N_{\bullet jl}(t) = \sum_{i=1}^n N_{ijl}(t)$ and $Y_{\bullet j}(t) = \sum_{i=1}^n Y_{ij}(t)$.

APPENDIX

A.1. Weak convergence of multiplicative counting processes.

LEMMA 2. *Let A_n and A be a sequence of functions on $\mathscr{A}$ such that $A_n \to A$ with respect to the total variation norm. Suppose that N_n and N are counting processes defined on $(\mathscr{N}, \Sigma_{\mathscr{N}})$, whose cumulative intensity processes Λ_n and Λ are given by*

$$\Lambda_n(t) = \int_0^t \phi(s : N_n)\, dA_n(s)$$

and

$$\Lambda(t) = \int_0^t \phi(s : N)\, dA(s),$$

where ϕ is defined in (3.2). *Then as $n \to \infty$, $N_n \to_{\mathscr{L}} N$ on $\mathscr{D}$.*

PROOF. First, we will show that the sequence of N_n is tight. For this, it suffices to show that for a given $\varepsilon > 0$, there exists some constants M and k such that $\Pr\{N_n(1) > k\} < \varepsilon$ for all $n > M$. By the Chebyshev inequality,

$$\begin{aligned}\Pr\{N_n(1) > k\} &\le \mathrm{E}_{P_n}[N_n(1)]/k\\ &= \mathrm{E}_{P_n}[\Lambda_n(1)]/k\\ &\le A_n(1)/k,\end{aligned}$$

where the last inequality holds because of $\phi \le 1$. Since $A_n(1) \to A(1)$, we can choose k and M sufficiently large so that $A_n(1)/k < \varepsilon$ for $n > M$, which completes the proof of tightness.

Now, we can apply Theorem IX.2.11 in Jacod and Shiryaev (1987) to conclude the theorem. □

THEOREM A.1. *For given A_n and A, let N_n and N be multiplicative counting processes defined on $(\mathscr{N}, \Sigma_{\mathscr{N}})$ with the cumulative intensity processes Λ_n and Λ given by*

$$\Lambda_n(t) = \int_0^t \phi(s : N_n)\, dA_n(s)$$

and

$$\Lambda(t) = \int_0^t \phi(s : N)\, dA(s),$$

where ϕ is given by (3.2). *Suppose that A is a Levy process on $(\mathscr{A}, \Sigma_{\mathscr{A}})$ with the compensator*

$$\nu([0, t] \times D) = \int_0^t \int_D dF_s(x)\, ds + \sum_{u_j \le t} \int_D dH_j(x)$$

for $t \in [0, 1]$ and $D \in \mathscr{B}[0, 1]$ and a given set $\mathbf{U} = \{u_1, \ldots, u_l\}$, and that A_n are Levy processes on $(\mathscr{A}, \Sigma_{\mathscr{A}})$ with the compensators

$$\nu_n([0, t] \times D) = \int_0^t \int_{D\cap[1/n, 1]} dF_s(x)\, ds + \sum_{u_j \le t} \int_D dH_j(x)$$

for $t \in [0, 1]$, $D \in \mathscr{B}[0, 1]$ and $n = 1, 2, \ldots$. Then

$$(N_n, A_n) \to_{\mathscr{L}} (N, A)$$

on $\mathscr{D} \times \mathscr{D}$.

PROOF. It suffices to show that for all f and g in $C^+[0, 1]$,

$$\begin{aligned}&\mathrm{E}\left[\exp\left\{-\int_0^t g(s)\, dA_n(s) - f(s)\, dN_n(s)\right\}\right]\\ &\qquad \to \mathrm{E}\left[\exp\left\{-\int_0^t g(s)\, dA(s) - f(s)\, dN(s)\right\}\right].\end{aligned} \tag{A.1}$$

Define a function $\Phi(a)$ on $\mathscr{A}$ by

$$\Phi(A) = \mathrm{E}\left[\exp\left\{-\int_0^1 f(t)\,dN(t)\right\}\bigg|A\right]\exp\left(-\int_0^t g(s)\,dA(s)\right),$$

where $\mathrm{E}(\cdot|A)$ denotes a conditional expectation for a given A. Then (A.1) is equivalent to

$$\mathrm{E}(\Phi(A_n)) \to \mathrm{E}(\Phi(A)).$$

Define a new process $\tilde{A}_n$ by

$$\tilde{A}_n(t) = \sum_{s\le t} \Delta A(s) I(\Delta A(s) \ge 1/n).$$

Then it is easy to see that $\tilde{A}_n$ is a Levy processes on $(\mathscr{A}, \Sigma_{\mathscr{A}})$ with the compensator ν_n, and hence $\tilde{A}_n =_d A_n$. Since

$$\sup_{t\in[0,1]} \left|\tilde{A}_n(t) - A(t)\right| \to 0$$

with probability 1, Lemma 2 implies

$$\Phi(\tilde{A}_n) - \Phi(A) \to 0 \tag{A.2}$$

with probability one. Since Φ is a bounded function, the proof is done. □

A.2. Proof of (3.31).

LEMMA 3. *Let $z_1, \dots, z_n$ and $w_1, \dots, w_n$ be real numbers whose absolute values are less than or equal to* 1. *Then*

$$\left|\prod_{i=1}^n z_i - \prod_{i=1}^n w_i\right| \le \sum_{i=1}^n |z_i - w_i|.$$

The proof can be found in any standard book [see, e.g., Durrett (1991)].

LEMMA 4. *For any given $C \in \mathscr{B}[0,1]$, let $\tilde{C} = C\backslash\mathbf{U}$. Then*

$$\Pr\{N_n \textit{ makes no jump on } \tilde{C}\} \ge \exp\left(-A^0(1)\mu(C)\right),$$

where μ is the Lebesgue measure.

PROOF. Since we are only interested in the behavior of N_n on $C\backslash\mathbf{U}$, without loss of generality, we can assume that $A^0(s)$ is continuous. Since $A_n^0 \le A^0$, for any open interval $C = (a, b)$ we have

$$\begin{aligned}
&\Pr\{N_n \text{ makes no jump on } C\}\\
&\quad = \mathrm{E}\left[\mathrm{E}\left\{\exp\left(-\int_a^b \phi_k(s:\mathbf{T_k})\, dA_n^0(s)\right) \,\middle|\, N_n(a) = k\right\}\right]\\
&\quad \geq \exp\left\{-\int_a^b dA_n^0(s)\right\}\\
&\quad \geq \exp\left\{-\int_a^b dA^0(s)\right\}\\
&\quad \geq \exp\{-A^0(1)(b-a)\}.
\end{aligned}$$

The same idea can be employed to show that for any set C of a finite union of disjoint open intervals,

$$\Pr\{N_n \text{ makes no jump on } C\} \geq \exp\{-A^0(t)\mu(C)\}.$$

For any open set $C \in [0, 1]$, there exist countably many disjoint open intervals $O_1, O_2, \dots$ such that

$$C = \bigcup_{i=1}^{\infty} O_i.$$

Let $C_m = \bigcup_{i=1}^{m} O_i$. Then since

$$\{N_n \text{ makes no jump on } C_m\} \downarrow \{N_n \text{ makes no jump on } C\}$$

as $m \to \infty$ and for each m,

$$\Pr\{N_n \text{ makes no jump on } C_m\} \geq \exp\{-A^0(1)\mu(C_m)\},$$

we can conclude that

$$\Pr\{N_n \text{ makes no jump on } C\} \geq \exp\{-A^0(1)\mu(C)\}.$$

For any set $C \in \mathscr{B}[0, 1]$, there exists an open set C^o such that $C \subset C^o$ and $\mu(C^o) - \mu(C) < \varepsilon$ for a given $\varepsilon > 0$. Then

$$\begin{aligned}
\Pr\{N_n \text{ makes no jump on } C\} &\geq \Pr\{N_n \text{ makes no jump on } C^o\}\\
&\geq \exp\{-A^0(1)\mu(C^o)\}\\
&\geq \exp\{-A^0(1)(\mu(C) + \varepsilon)\}.
\end{aligned}$$

Since ε is arbitrary, we complete the proof. □

PROOF OF (40). For a given $\varepsilon > 0$, define a sequence of sets $C_n(\varepsilon)$ by

$$C_n(\varepsilon) = \left\{s \in [0, 1] \,\middle|\, \int_0^{1/n} u\, dF_s(u) > \varepsilon\right\}.$$

Since $\int_0^{1/n} uF_s(du) \to 0$ for all $s \in [0, 1]$,

$$\mu(C_n(\varepsilon)) \to 0 \tag{A.3}$$

as $n \to \infty$. By Lemma 4, we have

$$\text{(A.4)} \qquad \Pr\{N_n \text{ makes no jump on } C_n(\varepsilon)\} \geq \exp(-A^0(1)\mu(C_n(\varepsilon))),$$

and so we can choose a constant M_1 sufficiently large such that for all $n > M_1$,

$$1 - \Pr\{N_n \text{ makes no jump on } C_n(\varepsilon)\} < \varepsilon.$$

Also, since $N_n \to_{\mathscr{L}} N$, there exist constants K and M_2 such that

$$\text{(A.5)} \qquad \Pr\{N_n(1) > K\} < \varepsilon$$

for all $n > M_2$.

Let $B(C_n(\varepsilon)) = \{y \in \mathscr{N} \mid y \text{ makes no jump on } C_n(\varepsilon)\}$ and $B = \{y \in \mathscr{N} \mid y(1) \leq K\}$. Then $\mathscr{N} = B^c \cup (B \cap B(C_n(\varepsilon))) \cup (B \cap B^c(C_n(\varepsilon)))$, $P_n(B^c) \leq \varepsilon$ and $P_n(B \cap B^c(C_n(\varepsilon)) \leq \varepsilon$. Therefore, we have

$$\text{(A.6)} \qquad \begin{aligned} &\left| \int_{\mathscr{N}} \exp\left(-\int_0^1 f(t)\, dy(t) \right) (\Psi_n(y) - \Psi(y)) P_{N_n}(dy) \right| \\ &\qquad \leq \int_{B \cap B(C_n(\varepsilon))} \left| \Psi_n(y) - \Psi(y) \right| P_{N_n}(dy) + 4\varepsilon. \end{aligned}$$

Since $N_n(1) \leq K$ and

$$\int_0^{1/n} x\, dF_t(x) \leq \varepsilon$$

whenever $\Delta N_n(t) = 1$ and $N_n \in B \cap B(C_n(\varepsilon))$, by applying Lemma 3 to the Lévy's formulas of Ψ_n and Ψ, we can show that for $n > \max\{M_1, M_2\}$,

$$\begin{aligned} &\int_{B \cap B(C_n(\varepsilon))} |\Psi_n(y) - \Psi(y)| P_{N_n}(dy) \\ &\quad \leq \int_{B \cap B(C_n(\varepsilon))} \sum_{s:\, \Delta y(s) = 1,\, s \notin \mathbf{U}} \left| c_3^n(s)^{-1} \int_0^1 \exp(-g(s)x) Y(s) x\, dF_s^n(x) \right. \\ &\qquad\qquad \left. - c_3(s)^{-1} \int_0^1 \exp(-g(s)x) Y(s) x\, dF_s(x) \right| \\ &\qquad + \left| \exp\left(\int_0^1 \int_0^1 (1 - e^{g(s)x})(1 - Y(s)x)\, dF_s^n(x) \right) ds \right. \\ &\qquad\quad \left. - \exp\left(\int_0^1 \int_0^1 (1 - e^{g(s)x})(1 - Y(s)x)\, dF_s(x) \right) ds \right| P_{N_n}(dy) \\ &\quad \leq M\varepsilon, \end{aligned}$$

where the constant M depends on neither n nor ε. Since ε is arbitrary, the proof is done. $\square$

Acknowledgments. The author thanks the referees for valuable suggestions which significantly improved the earlier version of the paper. Thanks are also due to the late advisor Robert Bartoszyński for his continual support and encouragement and Ramani Pilla for proofreading the paper thoroughly.

REFERENCES

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

ANDERSON, P. K., BORGAN, Ø., GILL, R. D. and KEIDING, N. (1993). *Statistical Models Based on Counting Processes*. Springer, New York.

DOKSUM, K. A. (1974). Tailfree and neutral random probabilities and their posterior distributions. *Ann. Probab.* **2** 183–201.

DURRETT, R. (1991). *Probability: Theory and Examples*. Wadsworth and Brooks/Cole, Belmont, CA.

HJORT, N. L. (1990). Nonparametric Bayes estimators based on beta processes in models for life history data. *Ann. Statist.* **18** 1259–1294.

JACOD, J. (1979). Calcul stochastique et problèmes de martingales. *Lecture Notes in Math.* **714**. Springer, Berlin.

JACOD, J. and SHIRYAEV, A. N. (1987). *Limit Theorems for Stochastic Processes*. Springer, New York.

KARLIN, S. and TAYLOR, H. M. (1975). *A First Course in Stochastic Processes*, 2nd ed. Academic Press, New York.

KAO, E. and SMITH, M. S. (1993). On renewal processes relating to counter models: the case of phase-type interarrival times. *J. Appl. Probab.* **30** 175–183.

KARR, A. F. (1986). *Point Processes and Their Statistical Inference*. Dekker, New York.

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

LO, A. Y. (1992). Bayesian inference for Poisson process models with censored data. *J. Nonparametr. Statist.* **2** 71–80.

PYKE, R. (1958). On renewal processes related to type I and type II counter models. *Ann. Math. Statist.* **29** 737–754.

SHIRYAEV, A. N. (1991). *Probability*. Springer, New York.

SWEETING, T. J. (1989). On conditional weak convergence. *J. Theoret. Probab.* **2** 461–474.

DEPARTMENT OF STATISTICS
HANKUK UNIVERSITY OF FOREIGN STUDY
YONGIN, KYUNGKI-DO
449-791 KOREA
E-MAIL: kimy@stat.hufs.ac.kr