Z. Wahrscheinlichkeitstheorie verw. Gebiete
59, 55-66 (1982)

Zeitschrift für
Wahrscheinlichkeitstheorie
und verwandte Gebiete

Bayesian Nonparametric Statistical Inference for Poisson Point Processes*

Albert Y. Lo

Department of Statistics, Rutgers University, New Brunswick, N.J., USA
and Department of Mathematics and Statistics University of Pittsburgh, PA, USA

Summary. A random measure is said to be selected by a weighted gamma prior probability if the values it assigns to disjoint sets are independent gamma random variables with positive multipliers. If the intensity measure of a nonhomogeneous Poisson point process is selected by a weighted gamma prior probability and if a sample is drawn from the Poisson point process having this intensity measure, then the posterior random intensity measure given the observations is also selected by a weighted gamma prior probability. If the measure space is Euclidean and if the true intensity measure is continuous and finite, the centered posterior process, rescaled by the square root of the sample size, will converge weakly in Skorohod topology to a Wiener process subject to a change of time scale.

1. Introduction and Preliminaries

1.1. Introduction

The problem of statistical inference in Poisson point processes has a long history. A poisson point process is characterized by the intensity measure or the derivative of the intensity measure (with respect to certain σ-finite measure) called the intensity rate function. Thus, statistical inference problems for Poisson point processes concnetrate on methods of estimating and testing the intensity measure or the rate function. However, the methods employed are parametric in nature in that the intensity measure or rate function is assumed to be of some linear or parametric form (Brown (1972), Lewis (1972), Clevenson and Zidek (1977)). This situation has changed considerably with the appearance of Aalen (1978). In this article he proposes a class of nonparametric estimators for the intensity measures of the multiplicative point pro-

* This research was supported in part by the National Science Foundation Grants MCS 77-10376 and MCS 75-14194

0044-3719/82/0059/0055/$02.40

cesses and shows that they are well behaved asymptotically. However, his approach is from a sample theorist point of view.

Motivated by Ferguson (1973) and Doksum (1974), the object of this study is to extend the idea of the gamma conjugate prior when sampling from a Poisson distribution to the infinite dimensional case and demonstrate how this extension can be used to solve the statistical problems for the Poisson point processes from a nonparametric Bayesian point of view.

The type of prior process used here is the weighted gamma process. It was introduced by Dykstra and Laud (1981) as prior process for the hazard function and rate function in the reliability model. They show that the posterior processes are mixtures of gamma type processes. See also Lo (1978) for a similar result in a general point process model. In this article, we show that the weighted gamma processes are conjugate priors when sampling from the Poisson point process model.

In Sect. 2, we associate with each finite measure α on a measure space and each positive and α-integrable function β a probability on the space of finite set functions $\{v\}$. We say that v has a weighted gamma probability distribution with shape α and scale β, and denote this concisely by $v \sim \mathscr{P}_{\alpha,\beta}$. In Sect. 3 we demonstrate that if the prior distribution for the intensity measure v of a Poisson point process is $\mathscr{P}_{\alpha,\beta}$, then given a sample of size n; $N_1, N_2, \ldots, N_n$ from the Poisson point process, the posterior distribution for v is $\mathscr{P}_{\alpha+\Sigma N_i, \frac{\beta}{1+n\beta}}$. In Sect. 4, we further restrict our measure space to the unit q-cube. We show that the posterior process converges weakly in $D[0,1]^q$ to a Brownian process for almost all sample sequences, if the "true" intensity measure is continuous. In Sect. 5 we demonstrate how the results in the previous sections can be used to solve the nonparametric statistical inference problem connected with the Poisson point processes.

1.2. Some Preliminaries on Random Measures

Let X be a complete separable metric space, $\mathscr{A}$ the Borel σ-field and Ω the space of finite measures on $(X, \mathscr{A})$. This Ω topologized by weak convergence is a complete separable metric space (Prohorov (1956)). Let $\mathscr{M}$ be the σ-field on Ω generated by (weak) open sets. Thus, $(\Omega, \mathscr{M})$ is a measure space. Note that $\mathscr{M}$ is the smallest σ-field that makes all the projection maps $\mu \to \mu(A)$ $A \in \mathscr{A}$ measurable. See Matthes, Kerstan and Mecke (1978), Chap. 3. Let $\mathbb{N}$ be the space of finite counting measures on X. $\mathbb{N}$ is a closed subset of Ω. Let the σ-field $\mathscr{N}$ be $\{\mathbb{N} \cap C;\ C \in \mathscr{M}\}$, that is, the restriction of $\mathscr{M}$ to $\mathbb{N}$. Note that $(\mathbb{N}, \mathscr{N})$ is also a measure space.

Definition 1.1. A finitely additive random measure μ on the measure space $(X, \mathscr{A})$ is a family of random variables $\{\mu(A): A \in \mathscr{A}\}$, indexed by sets in $\mathscr{A}$, such that

(1) $\mu(A)$ is a positive real valued random variable for each $A \in \mathscr{A}$.
(2) $\mu(X) < \infty$ a.s.

(3) If $\{A_1^i, \ldots, A_{k_i}^i\}$; $i=1, \ldots, n$, are n collections of measurable disjoint sets in $\mathscr{A}$, then

$$\left(\mu\left(\sum_{j=1}^{k_i} A_j^i\right); i=1, \ldots, n\right) \stackrel{\mathscr{L}}{=} \left(\sum_{j=1}^{k_i} \mu(A_j^i); i=1, \ldots, n\right)$$

where $\stackrel{\mathscr{L}}{=}$ denotes equality in law.

We call a finitely additive random measure μ a sequentially continuous random measure if it is sequentially continuous[1] from above in the sense that if $A_n \in \mathscr{A}$, all n and $A_n \downarrow \emptyset$ when $n \to \infty$, then $\mathscr{P}\{\mu: \mu(A_n) \downarrow 0\} = 1$. According to Harris (1968), any sequentially continuous random measure μ on $(X, \mathscr{A})$ has a version which is σ-additive, i.e., $\mathscr{P}\{\mu: \mu \text{ is } \sigma\text{-additive}\} = 1$,[2] and any sequentially continuous random counting measure N on $(X, \mathscr{A})$ has a version which is σ-additive. To summarize, we have the following:

Proposition 1.1. *Let μ be a sequentially continuous random measure on a complete separable metric space X with preassigned finite dimensional distributions, then there exists a unique probability measure on $(\Omega, \mathscr{M})$ with identical finite dimensional distributions.*

Let μ be a random measure on X. The Laplace transform of μ is defined by

$$L_\mu(f) = E e^{-\mu(f)}, \quad f \in K^+$$

where K^+ is the family of positive functions on X and

$$\mu(f) = \int f(x)\, \mu(dx).$$

The following result is well known (see for example Kallenberg (1976), page 16).

Proposition 1.2. *Probability measures on $(\Omega, \mathscr{M})$ or $(\mathbb{N}, \mathscr{N})$ are uniquely determined by their Laplace transforms.*

2. The Gamma and Weighted Gamma Priors

It is well known that the family of gamma distributions is closed under convolutions for the same scale parameter. This property allows us to construct a prior probability on the space of finite measures. Fortunately, this prior probability turns out to be also tractable from a Bayesian point of view in the Poisson case. We start with

Definition 2.1. Let α be a finite σ-additive measure on $(X, \mathscr{A})$. The random measure μ on $(X, \mathscr{A})$ is said to be selected by a gamma prior probability if for all disjoint measurable sets A_j; $j=1, \ldots, k$, $\{\mu(A_j); j=1, \ldots, k\}$ is a family of

[1] This term is suggested by Lucien Le Cam

[2] $\mathscr{P}^*\{\mu: \mu \text{ is } \sigma\text{-additive}\} = 1$ where $\mathscr{P}^*$ is the outer probability measure if the set is not measurable

independent gamma random variables with means $\alpha(A_j)$; $j=1, \ldots, k$, respectively and scale parameter unity.

We say that μ is a gamma random measure with shape α and scale unity. Note that conditions of Proposition 1.1 hold, so there exists a unique probability $\mathscr{P}_{\alpha,1}$ on $(\Omega, \mathscr{M})$ with these finite dimensional distributions. We denote this by $\mu \sim \mathscr{P}_{\alpha,1}$. The following two results are clear,

(a) $\int_\Omega \mu(A) \mathscr{P}_{\alpha,1}(d\mu) = \alpha(A)$, all $A \in \mathscr{A}$

(b) $\int_\Omega \{\mu(A)-\alpha(A)\}\{\mu(B)-\alpha(B)\} \mathscr{P}_{\alpha,1}(d\mu) = \alpha(A \cap B)$, all $A, B \in \mathscr{A}$.

For each positive valued α-integrable β, the map $\mu \to \beta\mu = \nu$, where $\beta\mu$ is defined by $\beta\mu(A) = \int_A \beta(x)\mu(dx)$, $A \in \mathscr{A}$, is a measurable map from $(\Omega, \mathscr{M})$ into $(\Omega, \mathscr{M})$. This can be verified easily when β is an indicator and the general conclusion follows from a limiting argument. If $\mu \sim \mathscr{P}_{\alpha,1}$, we denote the induced probability measure on the range space by $\mathscr{P}_{\alpha,\beta}$. Thus, $\nu \sim \mathscr{P}_{\alpha,\beta}$ and ν is a weighted gamma random measure. We have the following:

Proposition 2.1.

(a′) $\int_\Omega \nu(A) \mathscr{P}_{\alpha,\beta}(d\nu) = \beta\alpha(A)$, *all* $A \in \mathscr{A}$.

(b′) $\int_\Omega \{\nu(A)-\beta\alpha(A)\}\{\nu(B)-\beta\alpha(B)\} \mathscr{P}_{\alpha,\beta}(d\nu) = \beta^2\alpha(A \cap B)$, *all* $A, B \in \mathscr{A}$, *provided* β^2 *is* α*-integrable.*

(c′) *If* A_j; $j=1, \ldots, k$ *are measurable disjoint subsets of* X, *then* $\nu(A_j)$; $j=1, \ldots, k$ *are independent under* $\mathscr{P}_{\alpha,\beta}$.

d′) $\int_\Omega e^{-\nu(f)} \mathscr{P}_{\alpha,\beta}(d\nu) = e^{-\alpha(\log[1+\beta f])}$, *all* $f \in K^+$.

Proof. (a′) is a consequence of the change of variable technique.

To prove (b′), a change of variable puts the left hand side of (b′) into

$$\int_\Omega \int_A \beta(x)\{\mu(dx)-\alpha(dx)\} \int_B \beta(x)\{\mu(dx)-\alpha(dx)\} \mathscr{P}_\alpha(d\mu).$$

Now this is equal to $\beta^2\alpha(A \cap B)$ if $\beta(x) = a \cdot I_{\{x \in C\}}$, when $a > 0$ and $C \in \mathscr{A}$. We can conclude (b′) with a limiting argument.

(c′) is true because if $A \cap B = \emptyset$, $\beta\mu(A)$ is measurable with respect to the σ-field generated by the family $\{\mu(A \cap C); C \in \mathscr{A}\}$ while $\beta\mu(B)$ is measurable with respect to the σ-field generated by the family $\{\mu(B \cap C); C \in \mathscr{A}\}$. The assertion follows because these two families are independent.

(d′) One first verifies this equality when f is a simple function, then a limiting argument via the dominated convergence theorem concludes the proof. □

Next, we show that the weighted gamma random measure also preserves the convolution property in the following way. We denote $\beta\mu$ where $\mu \sim \mathscr{P}_{\alpha,1}$ by $\beta\mu_\alpha$. Let $\{\alpha_j; j \in Z^+\}$ be a family of finite measures on $(X, \mathscr{A})$ such that $\sum_j \alpha_j(X) < \infty$. We define

$$\alpha(A) = \sum_j \alpha_j(A), \quad \text{all } A \in \mathscr{A}.$$

It is clear that α is also a finite measure on $(X, \mathscr{A})$.

Proposition 2.2. *For each positive valued α-integrable β, $\beta\mu_\alpha \overset{\mathscr{L}}{=} \sum \beta\mu_{\alpha_j}$, where the summands are independent.*

Proof. It suffices to show that, for $\beta \in K^+$,

$$L_{\beta\mu_\alpha}(f) = \prod_j L_{\beta\mu_{\alpha_j}}(f); \quad \text{all } f \in K^+.$$

Note that by a change of variable, the above is equivalent to

$$L_{\mu_\alpha}(f\beta) = \prod_j L_{\mu_{\alpha_j}}(f\beta); \quad \text{all } \beta, f \in K^+.$$

Thus, we only need to show

$$L_{\mu_\alpha}(g) = \prod_j L_{\mu_{\alpha_j}}(g); \quad \text{all } g \in K^+.$$

But then this is equivalent to showing that $\mu_\alpha \overset{\mathscr{L}}{=} \sum_j \mu_{\alpha_j}$, where the summands are independent. That this is indeed the case follows from the convolution property of gamma random variables. □

3. The Posterior Random Measures

Associate with each $v \in \Omega$, there is a Poisson point process $N = \{N(A);\ A \in \mathscr{A}\}$ with distribution denoted by $\mathscr{P}_v$. Note that sets of the form $\{N : N(A_l) = k_l;\ l = 1, \ldots, s\}$ where $k_l \in Z^+$; $l = 1, \ldots s$ and $\{A_l;\ l = 1, \ldots, s\}$ is any measurable partition of X form a determining class for $\mathscr{P}_v$ and

$$\mathscr{P}_v\{N : N(A_l) = k;\ l = 1, \ldots, s\} = \prod_{l=1}^{s} \frac{v^{k_l}(A_l)}{k_l!}\, e^{-v(A_l)}.$$

We call v the intensity measure of the Poisson point process. Thus, for each $v \in \Omega$, $\mathscr{P}_v$ is a probability on $(\mathbb{N}, \mathscr{N})$. Note that $\int_{\mathbb{N}} e^{-N(h)} \mathscr{P}_v(dN) = e^{-v(1-e^{-h})}$, all $h \in K^+$. The proof is the same as the proof for Proposition 2.1 (d'). Let $\mathscr{P}_v^n$ be the product probability on $(\mathbb{N} \times \mathbb{N} \times \ldots \times \mathbb{N},\ \mathscr{N} \otimes \mathscr{N} \otimes \ldots \otimes \mathscr{N}) = (\mathbb{N}^n, \mathscr{N}^n)$ and denote $\mathbf{N} = (N_1, \ldots, N_n)$. Let $\mathscr{P}$ be a probability on $(\Omega, \mathscr{M})$, then

(i) for each $A = \{\mathbf{N} : N_i(A_l^i) = k^i;\ l = 1, \ldots, s_i,\ i = 1, \ldots, n\} \in \mathscr{N}^n$, where for each i, $\{A_l^i;\ l = 1, \ldots, s_i\}$ is a measurable partition of X, we have $\mathscr{P}_v^n(A)$ is a measurable function in v;

(ii) there exists a $\mathscr{P}$-null subset of Ω such that for each v outside of this $\mathscr{P}$-null set, $\mathscr{P}_v^n$ is a probability on $(\mathbb{N}^n, \mathscr{N}^n)$.

Let $\mathscr{P}_2$ be defined as follows:

$$\mathscr{P}_2\{A \times B\} = \int_B \mathscr{P}_v^n(A)\, \mathscr{P}(dv), \quad \text{all } A \in \mathscr{N}^n,\ B \in \mathscr{M}.$$

Let Q be the marginal probability distribution of $N_1, \ldots, N_n$, that is

$$Q(A) = \mathscr{P}_2(A \times \Omega), \qquad A \in \mathscr{N}^n.$$

Let $\mathscr{P} = \mathscr{P}_{\alpha,\beta}$ i.e., the weighted gamma distribution and denote the marginal distribution Q in this case by $Q_{\alpha,\beta}$.

Definition 3.1. We call a nonnegative real valued function P on $\mathbb{N}^n \times \mathscr{M}$ a posterior random measure given $N_1, \ldots, N_n$ if

(i) for each $\mathbf{N} = (N_1, \ldots, N_n) \in \mathbb{N}^n$, $P(\mathbf{N}, \cdot)$ is a probability on $\mathscr{M}$;

(ii) for each $B \in \mathscr{M}$ $P(\cdot, B)$ is a measurable function of $\mathbf{N} = (N_1, N_2, \ldots, N_n)$; and

(iii) for all $A \in \mathscr{N}^n$, $B \in \mathscr{M}$,

$$\mathscr{P}_2(A \times B) = \int_A P(\mathbf{N}, B)\, Q_{\alpha,\beta}(d\mathbf{N}).$$

The $\mathscr{P}_3$ probability measure disintegrates into the prior and the model distribution on one hand, the marginal and the posterior distribution on the other. Specializing to the Gamma-Poisson family, we have the following disintegration lemma in the case of $n=1$.

Lemma 3.1. *Let g be a positive valued measurable or quasi-integrable (with respect to $\mathscr{P}_2$) function defined on $(\Omega \times \mathbb{N}, \mathscr{M} \otimes \mathscr{N})$, then*

$$\int_\Omega \int_{\mathbb{N}} g(v, N)\, \mathscr{P}_v(dN)\, \mathscr{P}_{\alpha,\beta}(dv) = \int_{\mathbb{N}} \int_\Omega g(v, N)\, \mathscr{P}_{\alpha+N, \frac{\beta}{1+\beta}}(dv)\, Q_{\alpha,\beta}(dN). \tag{2.1}$$

Proof. It suffices to show that the equality holds for $g(v, N) = I_{\{(v, N) \in B \times A\}}$, where $B \in \mathscr{M}$ and $A \in \mathscr{N}$. But then because of Proposition 1.2., this is equivalent to showing, for all $f, h \in K^+$,

$$E E_v\, e^{-v(f) - N(h)} = E E_N\, e^{-v(f) - N(h)} \tag{2.2}$$

where E_v is the conditional expectation given v and E_N is the conditional expectation given N. To show (2.2), we note that $Ee^{-v(f)} = e^{-\alpha(\log[1+\beta f])}$ when $v \sim \mathscr{P}_{\alpha,\beta}$, and $E_v\, e^{-N(h)} = e^{-v(1-e^{-h})}$ because given v, N is a Poisson point process with intensity measure v. Thus, $EE_v\, e^{-v(f)-N(h)} = e^{-\alpha(\log[1+\beta(f+1-e^{-h})])}$. On the other hand, $Ee^{-N(h)} = e^{-\alpha(\log[1+\beta(1-e^{-h})])}$ is the Laplace transform of the marginal point process. If given N, $E_N\, e^{-v(f)} = e^{-(\alpha+N)\left(\log\left[1+\frac{\beta}{1+\beta} f\right]\right)}$, then

$$\begin{aligned}
EE_N\, e^{-v(f)-N(h)} &= Ee^{-N(h)}\, E_N\, e^{-v(f)} \\
&= Ee^{-N(h)}\, e^{-(\alpha+N)\left(\log\left[1+\frac{\beta}{1+\beta} f\right]\right)} \\
&= e^{-\alpha\left(\log\left[1+\frac{\beta}{1+\beta} f\right]\right)}\, Ee^{-N\left(h+\log\left[1+\frac{\beta}{1+\beta} f\right]\right)} \\
&= e^{-\alpha\left(\log\left[1+\frac{\beta}{1+\beta} f\right]\right)} \cdot e^{-\alpha\left(\log\left[1+\beta\left\{1-e^{-\left(h+\log\left[1+\frac{\beta}{1+\beta} f\right]\right)}\right\}\right]\right)} \\
&= e^{-\alpha\left(\log\left[\left(1+\frac{\beta}{1+\beta} f\right)\left(1+\beta\left\{1-e^{-h-\log\left[1+\frac{\beta}{1+\beta} f\right]}\right\}\right)\right]\right)} \\
&= e^{-\alpha\left(\log\left[1+\frac{\beta}{1+\beta} f+\beta\left\{1+\frac{\beta}{1+\beta} f-e^{-h}\right\}\right]\right)} \\
&= e^{-\alpha(\log[1+\beta\{1+f-e^{-h}\}])}
\end{aligned}$$

which we recognize as the joint Laplace transform, so that (2.2) is true and the lemma is proved. □

Remark. This proof via Laplace transform was communicated to me by Lucien Le Cam.

The following theorem says that weighted gamma priors are conjugate priors when sampling from a Poisson point process.

Theorem 3.1. *Let $v \sim \mathscr{P}_{\alpha,\beta}$, and given v, $N_1, \ldots, N_n$ is a sample of size n from a Poisson point process with intensity measure v, then the posterior distribution of v given $N_1, N_2, \ldots, N_n$ is*

$$\mathscr{P}_{\alpha+\sum_1^n N_i,\ \frac{\beta}{1+n\beta}}.$$

Proof. Since the posterior distribution of v given N_1 is $\mathscr{P}_{\alpha+N_1,\ \frac{\beta}{1+\beta}}$ by Lemma 3.1, we can use $\mathscr{P}_{\alpha+N_1,\ \frac{\beta}{1+\beta}}$ as the new prior and observe N_2. The posterior distribution given N_1 and N_2 will be

$$\mathscr{P}_{\alpha+N_1+N_2,\ \frac{\frac{\beta}{1+\beta}}{1+\frac{\beta}{1+\beta}}} = \mathscr{P}_{\alpha+N_1+N_2,\ \frac{\beta}{1+2\beta}}.$$

This procedure generalizes to n, and the proof is completed. □

We use the following example to illustrate the idea.

Example 3.1. If $\beta(x) = \frac{1}{\theta}$, where $\theta > 0$, we have a gamma random measure with shape parameter α and constant scale $1/\theta$. It is easy to see that the marginal point process is defined by

$$Q\{N: N(A_l) = k_l;\ l=1, \ldots, s\} = \prod_{l=1}^{s} \frac{\Gamma(\alpha(A_l)+k_l)}{\Gamma(\alpha(A_l))\, k_l!} \left(\frac{1}{1+\frac{1}{\theta}}\right)^{\alpha(A_l)} \left(\frac{1/\theta}{1+\frac{1}{\theta}}\right)^{k_l}$$

where $k_l \in Z^+$; $l=1, \ldots, s$ and $\{A_l;\ l=1, \ldots, s\}$ is a measurable partition of X. This is an independent negative binomial point process. Its Laplace transform is given by

$$E e^{-N(h)} = e^{-\alpha\left(\log\left[1+\frac{1}{\theta}-e^{-h}\right]\right)}.$$

The posterior random measure after n observations is again a gamma random measure with shape parameter $\alpha + N_1 + \ldots + N_n$ and scale $\frac{1}{n+\theta}$.

4. The Limiting Posterior Distributions

Investigation into the consistency of posterior distribution has a rather long history. A general proof of this phenomenon is given by Doob (1949) who uses martingale methods to obtain the results that the posterior distributions converge to the distribution degenerate at the true parameter under which the sample is drawn, for almost all true parameters. Stronger results replacing the almost surely parameter set by the support of the prior probability were obtained by LeCam (1953), Freedman (1963), Fabius (1964) and Berk (1970) under various assumptions.

If the problems on the consistency of the posterior distributions are more or less settled, the search for the limiting posterior distribution has been successful only in the parametric case, typically, the parameter space is a Borel subset of the Euclidean space. See Le Cam (1953) and (1958) or more recently, Walker (1969) and Dawid (1970). Johnson (1970) derives an asymptotic expansion for the posterior distribution, having the normal distribution as the leading term. Weak convergence, while more or less set for the classical statistics based on the empirical distribution, remains incompleted in the nonparametric Bayesian case. Nevertheless, the weak convergence of the posterior Dirichlet processes, centered and rescaled is recently proved in Lo (1978). The limiting Gaussian process is the Brownian bridge as expected. In that same article, the $D[0,1]$ weak convergence of the posterior weighted gamma processes, centered and rescaled, is also established under the condition that the true intensity measure is continuous and finite. This latter result is now being extended to the following:

Proposition 4.1. *Let α be continuous and with support equal to $[0,1]^q$. If the true intensity measure μ_0 is continuous, then for almost all sample sequence, the posterior processes, centered and rescaled by the square root of the sample size converge weakly in $D[0,1]^q$ to the Wiener process W with zero means and second moments $EW(A)\,W(B)=\mu_0(A\cap B)$, for all measurable $A, B\subset[0,1]^q$.*

Proof. Because the posterior processes are independent "increment", to show finite dimensional distributions converge is equivalent to show that for each measurable $A\subset[0,1]^q$

$$\text{a.s. } [\mathscr{P}_{\nu_0}], \qquad X_n(A)\overset{\mathscr{L}}{\Rightarrow} N(0,\mu_0(A))$$

where

$$\begin{aligned} X_n(A)&=\sqrt{n}\left\{\nu(A)-\sum_{i=0}^{n}\frac{\beta}{1+n\beta}N_i(A)\right\}, \qquad \text{with } N_0(A)=\alpha(A)\\ &\overset{\mathscr{L}}{=}\sqrt{n}\sum_{i=0}^{n}\left\{\frac{\beta}{1+n\beta}\mu_{N_i}(A)-\frac{\beta}{1+n\beta}N_i(A)\right\}, \end{aligned}$$

where the equility in law holds because of Proposition 2.2 which also implies the summands are independent random variables. Thus the central limit theorem for triangular arrays (Loéve (1963)) applies and we conclude that $X_n(A)\overset{\mathscr{L}}{\Rightarrow} N(0,\sigma^2)$ a.s. $[\mathscr{P}_{\mu_0}]$, where σ^2 is the limiting posterior variance $X_n(A)$. To evaluate σ^2, note that the posterior variance of $X_n(A)$ is

$$n\sum_{i=0}^{n}\left(\frac{\beta}{1+n\beta}\right)^2 N_i(A)=\left(\frac{n\beta}{1+n\beta}\right)^2\frac{1}{n}\sum_{i=0}^{n}N_i(A)$$

which tends to $\mu_0(A)$ a.s. $[\mathscr{P}_{\mu_0}]$ as $n\to\infty$ via an argument using Egorov's theorem, see Loéve (1963), page 595.

To check tightness, we use Theorem 3 of Bickel and Wichura (1971). Using their notations, it suffices to check conditions of Theorem 3 there with $(X_n, \mu)\in\mathscr{C}(\beta_1, \gamma)$ replaced by $(X_n, \mu_n)\in\mathscr{C}(\beta_1, \gamma)$ and μ_n converges weakly to a continuous measure μ on $[0, 1]^q$. See the remarks following Theorem 3 of Bickel and Wichura (1971). But then a necessary condition for $(X_n, \mu_n)\in\mathscr{C}(\beta_1, \gamma)$ is

$$E_N\{X_n^2(B)\,X_n^2(C)\}\leqq\mu_n(B)\,\mu_n(C), \qquad \text{with } \beta=2,\ \gamma=4$$

where B, C are neighboring blocks.

Because

$$\begin{aligned}E_N\{X_n^2(B)\,X_n^2(C)\}&=E_N\{X_n^2(B)\}\,E_N\{X_n^2(C)\}\\&=\left\{\left(\frac{n\beta}{1+n\beta}\right)^2\frac{1}{n}\sum_{i=0}^{n}N_i(B)\right\}\left\{\left(\frac{n\beta}{1+n\beta}\right)^2\frac{1}{n}\sum_{i=0}^{n}N_i(C)\right\}\quad \text{a.s. } [\mathscr{P}_{\mu_0}]\end{aligned}$$

we only need to set $\mu_n(A)=\left(\frac{n\beta}{1+n\beta}\right)^2\frac{1}{n}\sum_{i=0}^{n}N_i(A)$ which converges weakly to a continuous $\mu_0(A)$, the true intensity measure and thus tightness is in force. □

When $q=1$, the above result specializes to those in Lo (1978) and we have the following:

Proposition 4.2. *If the true intensity measure is continuous, then a.s.* $[\mathscr{P}_{\mu_0}]$,

$$P\{\sup_{0\leqq t\leqq 1}|X_n(t)|\leqq\lambda\}\to\frac{1}{\sqrt{2\pi\mu_0(1)}}\int_{-\infty}^{\infty}\sum_{k=-\infty}^{\infty}(-1)^k\exp\left\{-\frac{(u-2k\lambda)^2}{2\mu_0(1)}\right\}du,$$

$$P\{\sup_{0\leqq t\leqq 1}X_n(t)>\lambda\}\to\frac{1}{\sqrt{2\pi\,\mu_0(1)}}\int_{\lambda}^{\infty}\exp\left\{-\frac{u^2}{2\mu_0(1)}\right\}du, \qquad \lambda>0,$$

where $\mu_0(1)=\mu_0([0, 1])$.

Proof. These are standard consequences of weak convergence. □

5. Applications

In this section, unless otherwise specified, we assume $(X, \mathscr{A})=(R, \mathscr{B})$ where R is the real line and $\mathscr{B}$ is the Borel σ-field. Generalizations to R^q amount to a change of notation. The statistical decision problems we consider are those connected with the intensity measure μ of a Poisson point process. Let $\mu\sim\mathscr{P}_{\alpha,\beta}$ and given $N_1, \ldots, N_n$, a sample of size n from $\mathscr{P}_\mu$, $\mu\sim\mathscr{P}_{\alpha+\sum N_i,\ \frac{\beta}{1+n\beta}}$. We shall first find the Bayes rule for the no sample problem and update the prior random measure to the posterior random measure. Then we give the Bayes rule for the n sample problem.

(a) Estimation of the Intensity Measure. Let the space of actions of the statistician be the space of all finite measures on $(R, \mathscr{B})$. Let the loss function be

$$L(\mu, \hat{\mu}) = \iint_{\{s \leq t\}} \{\mu(t) - \mu(s) - \hat{\mu}(t) + \hat{\mu}(s)\}^2 \, W(dt, ds).$$

Where $W(dt, ds)$ is a given finite measure on $(R \times R, \mathscr{B} \otimes \mathscr{B})$. The Bayes rule for the no sample problem will be the rule which minimizes the quantity

$$EL(\mu, \hat{\mu}) = \iint_{\{s \leq t\}} E\{\mu(t) - \mu(s) - \hat{\mu}(t) + \hat{\mu}(s)\}^2 \, W(dt, ds)$$

which is

$$\hat{\mu}(t) - \hat{\mu}(s) = E\{\mu(t) - \mu(s)\} = \int_s^t \beta(x) \, \alpha(dx).$$

Thus the Bayes rule for the n sample problem will be the rule which minimizes

$$\begin{aligned} EL(\mu, \hat{\mu}) &= \iint_{\{s \leq t\}} E\{\mu(t) - \mu(s) - \hat{\mu}(t) + \hat{\mu}(s)\}^2 \, W(dt, ds) \\ &= \iint_{\{s \leq t\}} EE_N\{\mu(t) - \mu(s) - \hat{\mu}(t) + \hat{\mu}(s)\}^2 \, W(dt, ds). \end{aligned}$$

Thus

$$\begin{aligned} \hat{\mu}(t) - \hat{\mu}(s) &= E_N\{\mu(t) - \mu(s)\} \\ &= \int_s^t \frac{\beta(x)}{1 + n\beta(x)} \sum_{i=0}^n N_i(dx), \qquad \text{where } N_0 = \alpha \\ &= \int_s^t \frac{\beta(x)}{1 + n\beta(x)} \alpha(dx) + \sum_{i=1}^n \int_s^t \frac{\beta(x)}{1 + n\beta(x)} N_i(dx). \end{aligned}$$

In particular, if $\beta(x) \equiv 1/\theta$, $\theta > 0$ a constant, then the Bayes estimate for the n sample problem reduces to

$$\frac{\theta}{\theta + n} \frac{\alpha\{(s, t]\}}{\theta} + \frac{n}{\theta + n} \frac{1}{n} \sum_{i=1}^n N_i\{(s, t]\}$$

a convex linear combination of the prior guess and the sample empirical intensity. In the case that $s \equiv 0$, we have the estimate of the intensity function,

$$\frac{\theta}{\theta + n} \frac{\alpha(t)}{\theta} + \frac{n}{\theta + n} \bar{N}(t), \text{ where } \bar{N}(t) = \frac{1}{n} \sum_{i=1}^n N_i(t).$$

In the sequel, we assume $\beta(x) \equiv 1/\theta$, $\theta > 0$ for simplicity.

(b) Estimation of the Moments of the Intensity Measure. Suppose the statistician is interested in the r^{th} moment of the intensity measure.

$$m_r = E\{\int x^r N(dx) \mid \mu\} = \int x^r \mu(dx).$$

Using a quadratic loss function of the form

$$L(m_r, \hat{m}_r) = (m_r - \hat{m}_r)^2,$$

the Bayes rule for the n sample problem is

$$\frac{1}{\theta} \int x^r \sum_{i=0}^{n} N_i(dx) = \frac{\theta}{\theta+n} \frac{1}{\theta} \int x^r \alpha(dx) + \frac{n}{\theta+n} \frac{1}{n} \sum_{i=1}^{n} \int x^r N_i(dx),$$

which again is a convex linear combination of the prior guess of the r^{th} moment and the r^{th} sample moment.

(c) Estimation of the Factorial Moment Measure. Let

$$M_{[r]}(A) = N(A)(N(A)-1)(N(A)-2)\dots(N(A)-(r-1)).$$

Then

$$E\{M_{[r]}(A) \mid \mu\} = \mu^r(A)$$

is called the factorial moment measure for the point process N. Let the loss function be $\int_{-\infty}^{\infty} (\mu^r(t) - \hat{\mu}^r(t))^2 \, W(dt)$ where $W(dt)$ is a finite measure on $(R, \mathscr{B})$. Then the Bayes rule for the no sample problem is

$$E\mu^r(t) = \left(\frac{1}{\theta}\right)^r \frac{\Gamma(\alpha(t)+r)}{\Gamma(\alpha(t))}.$$

Therefore the Bayes rule for n sample problem is

$$\left(\frac{1}{\theta+n}\right)^r \frac{\Gamma\left(\alpha(t) + \sum_{i=1}^{n} N_i(t) + r\right)}{\Gamma\left(\alpha(t) + \sum_{i=1}^{n} N_i(t)\right)}.$$

(d) Estimation of the Convolution Intensity. Suppose there are two independently distributed Poisson point processes. $\tilde{N}$, N having intensity measures $\tilde{\mu}$, μ respectively. The statistician is interested in the convolution intensity measure, i.e.

$$\mu(t) = E\{\int \tilde{N}(t-x) \, N(dx) \mid \tilde{\mu}, \mu\} = \int \tilde{\mu}(t-x) \, \mu(dx).$$

We assume $\tilde{\mu} \sim \mathscr{P}_{\tilde{\alpha},\tilde{\theta}}$ and $\mu \sim \mathscr{P}_{\alpha,\theta}$. Moreover, $\tilde{\mu}$ and μ are independently distributed. Let the loss function be of a quadratic nature as before. Under this set up, the Bayes rule for the no sample problem will be

$$E \int \tilde{\mu}(t-x) \, \mu(dx) = \frac{1}{\tilde{\theta}\theta} \int \tilde{\alpha}(t-x) \, \alpha(dx)$$

and the Bayes rule for the (n, m) sample problem, where $\tilde{N}_1, \dots, \tilde{N}_n$ is a sample of size n from $\tilde{N}$, and $N_1, \dots, N_m$ is a sample of size m from N, is then

$$\frac{1}{(\tilde{\theta}+n)(\theta+m)}\int\sum_{i=0}^{n}\tilde{N}_i(t-x)\sum_{j=0}^{m}N_j(dx)=\frac{1}{(\tilde{\theta}+n)(\theta+m)}\int\sum_{i=0}^{n}\sum_{j=0}^{m}\int\tilde{N}_i(t-x)\,N_j(dx)$$

where $\tilde{N}_0=\tilde{\alpha}$, $N_0=\alpha$ as usual.

Acknowledgements. This is a revision of part of the author's Ph.D. thesis written under the supervision of Kjell Doksum. Thanks are also due to Richard Barlow for his continual support and encouragement, David Blackwell for introducing the author to Bayesian Statistics and Lucien Le Cam for numerous penetrating discussions.

References

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

Berk, R.H.: Consistency a posteriori. Ann. Math. Statist. **41**, 894–906 (1970)

Bickel, P.J., Wichura, M.J.: Convergence criteria for multiparameters stochastic processes and some applications. Ann. Math. Statist. **42**, 1656–1670 (1971)

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

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

Dawid, A.P.: On the limiting normality of posterior distributions. Proc. Cambridge Phil. Soc. **67**, 625–633 (1970)

Doksum, K.: Tailfree and neutral random probabilities and their posterior distributions. Ann. Probability **2**, 183–201 (1974)

Doob, J.: Application of the theory of martingales. Coll. Int. du CNRS. Paris, 22–28 (1949)

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

Fabius, J.: Asymptotic behavior of Bayes' estimates. Ann. Math. Statist. **35**, 846–856 (1964)

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

Freedman, D.A.: On the asymptotic behavior of Bayes estimates in the discrete case. Ann. Math. Statist. **34**, 1386–1403 (1963)

Harris, T.: Counting measures monotone random set funcitons. Z. Wahrscheinlichkeitstheorie verw. Gebiete **10**, 102–119 (1968)

Johnson, R.A.: Asymptotic expansions associated with posterior distributions. Amer. Math. Soc. **41**, 851–864 (1970)

Kallenberg, O.: Random Measures. Berlin: Akademie-Verlag and Academic Press 1976

Le Cam, L.: Les propriétés asymptotiques des solutions des Bayes. Publ. Inst. Statist. Univ. Paris **7**, 17–35 (1958)

Le Cam, L.: On some asymptotic properties of maximum likelihood estimates and related Bayes estimates. University of California Public Statistics **1**, 277–330 (1953)

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

Lo, A.Y.: Some contributions to Bayesian nonparametric statistical inference. Ph.D. dissertation, University of California, Berkeley (1978)

Lo, A.Y.: On a class of Bayesian nonparametric estimates: II. Rate function estimates. [To appear (1978)]

Loéve, M.: Probability Theory, 3rd ed. New York: D. Van Nostrand. 1963

Matthes, K., Kerstan, J., Mecke, J.: Infinitely Divisible Point Process. New York: Wiley 1978

Prohorov, Yu.V.: Convergence of random processes and limit theorems in probability theory. Theor. Probability Appl. **1**, 157–214 (1956)

Walker, A.M.: On the asymptotic behavior of posterior distributions. J. Roy. Statist. Soc. B **31**, 80–88 (1969)

Received July 1, 1980