Shrinkage Priors for Nonparametric Bayesian Prediction of Nonhomogeneous Poisson Processes

Fumiyasu Komaki, *Member, IEEE*

Abstract—We consider nonparametric Bayesian estimation and prediction for nonhomogeneous Poisson process models with unknown intensity functions. We propose a class of improper priors for intensity functions. Nonparametric Bayesian inference with kernel mixture based on the class improper priors is shown to be useful, although improper priors have not been widely used for nonparametric Bayes problems. Several theorems corresponding to those for finite-dimensional independent Poisson models hold for nonhomogeneous Poisson process models with infinite-dimensional parameter spaces. Bayesian estimation and prediction based on the improper priors are shown to be admissible under the Kullback–Leibler loss. Numerical methods for Bayesian inference based on the priors are investigated.

Index Terms—Admissibility, Dirichlet process, gamma process, kernel mixture, predictive density.

I. PREDICTION BASED ON MODELS WITH FINITE DIMENSIONAL PARAMETER

STATISTICAL modeling and data analysis based on non-homogeneous Poisson point processes have various applications (e.g. [31], [33]). We consider nonparametric Bayesian inference with kernel mixture for nonhomogeneous Poisson processes from the viewpoint of predictive density theory. In the present paper, it is shown that Bayesian procedures based on a class of improper priors provide reasonable estimation and prediction. Several theorems concerning admissibility corresponding to those for finite-dimensional independent Poisson models are shown to hold for the nonhomogeneous Poisson process model.

In this section, we summarize the basic framework of predictive density theory mainly studied for finite-dimensional models. Suppose that we have an observation x from a probability density $p(x \mid \theta)$ that belongs to a parametric model $\{p(x \mid \theta) \mid \theta \in \Theta \in \mathbb{R}^d\}$. The objective is to predict an unobserved random variable y distributed according to $p(y \mid \theta)$ using a predictive density $q(y; x)$.

We adopt the Kullback–Leibler loss

$$D\{p(y \mid \theta), q(y; x)\} = \int p(y \mid \theta) \log \frac{p(y \mid \theta)}{q(y; x)} \mathrm{d}y.$$

from the true density $p(y \mid \theta)$ to a predictive density $q(y; x)$. A predictive density $q(y; x)$ is said to dominate another predictive density $q'(y; x)$ if the risk of $q(y; x)$ is not greater than that of $q'(y; x)$ for all θ and the strict inequality holds for at least one point θ in the parameter space. A predictive density $q(y; x)$ is said to be admissible if there does not exist a predictive density dominating $q(y; x)$. See [5] and [28] for basic frameworks of statistical decision theory concerning admissibility and its relationship with Bayes theory.

Many studies recommend Bayesian predictive densities

$$p_\pi(y \mid x) = \frac{\int p(y \mid \theta) p(x \mid \theta) \pi(\theta) \mathrm{d}\theta}{\int p(x \mid \overline{\theta}) \pi(\overline{\theta}) \mathrm{d}\overline{\theta}}$$

based on a prior $\pi(\theta)$ rather than plug-in densities $p(y \mid \widehat{\theta}(x))$ (e.g. [1], [11], [19]), where $\widehat{\theta}$ is an estimated value of θ. This is because there exist Bayesian predictive densities dominating plugin densities in many examples.

It is important to construct prior distributions for Bayesian inference when we do not have specific prior information about unknown parameters. Such a prior is called a noninformative prior or an objective prior. The Jeffreys prior is a theoretically important objective prior. Let $\theta = (\theta_1, \ldots, \theta_d)$ be an unknown parameter of a finite dimensional parametric statistical model and $I(\theta)$ be the corresponding Fisher information matrix. Then, the Jeffreys prior is defined by

$$\pi_{\mathrm{J}}(\theta) \mathrm{d}\theta_1 \cdots \mathrm{d}\theta_d = |I(\theta)|^{1/2} \mathrm{d}\theta_1 \cdots \mathrm{d}\theta_d,$$

where $|I(\theta)|$ is the determinant of $I(\theta)$. Noninformative priors such as the Jeffreys prior often becomes improper, i.e. the total mass of the prior is infinite. The Bayes theorem cannot be directly applied to improper priors because they are not probability distributions, whose total masses are 1. However, formal application of the Bayes theorem to improper priors sometimes gives useful prediction and estimation. Statistical decision theory concerning admissibility provides a basis of such generalized Bayesian procedure based on improper priors (e.g. [5]).

Bayesian predictive densities based on shrinkage priors often dominate the Bayesian predictive densities based on the Jeffreys prior when the dimension of the parameter space is

large. Shrinkage priors assign more weight to parameter values close to a subset in the parameter space than the Jeffreys prior. In particular, Bayesian prediction based on shrinkage priors for finite-dimensional models such as the multivariate Normal model and the multidimensional independent Poisson model have been investigated. See [10] for recent developments of parameter estimation theory based on shrinkage priors.

First, consider the d-dimensional Normal model. Suppose that x_i $(i = 1, \ldots, d)$ are independently distributed according to $\mathrm{N}(\mu_i, \sigma^2)$ and that y_i $(i = 1, \ldots, d)$ are independently distributed according to $\mathrm{N}(\mu_i, \tau^2)$, where $\mathrm{N}(\mu_i, \sigma^2)$ is the Normal distribution with mean μ_i and variance σ^2. Here, $\mu := (\mu_1, \ldots, \mu_d)$ is the unknown parameter and σ and τ are known positive constants. We consider prediction of $y = (y_1, y_2, \ldots, y_d)$ using $x = (x_1, x_2, \ldots, x_d)$ under the Kullback–Leibler loss. The Bayesian predictive density $p_{\mathrm{S}}(y \mid x)$ based on the prior $\pi_{\mathrm{S}}(\mu) = \|\mu\|^{-(d-2)} = (\sum_i \mu_i{}^2)^{-(d-2)/2}$ introduced by [32] dominates the Bayesian predictive density $p_{\mathrm{J}}(y \mid x)$ based on the Jeffreys prior $\pi_{\mathrm{J}}(\mu) = 1$ ([20]). This corresponds to the widely known result that the generalized Bayes estimator based on Stein's prior $\pi_{\mathrm{S}}(\mu)$ dominates the best invariant estimator $\widehat{\mu} = x$ when $d \geq 3$. See [12] for sufficient conditions for general priors other than the Stein prior and [7] for admissible predictive densities for Normal models. The asymptotics of minimax risk of predictive density estimation for non-parametric regression is studied by [34].

Next, consider the d-dimensional Poisson model, which is closely related to the nonhomogeneous Poisson process models considered herein. Intuitively speaking, nonhomogeneous Poisson process models are infinite-dimensional Poisson models. Suppose that x_i $(i = 1, \ldots, d)$ are independently distributed according to $\mathrm{Po}(s\lambda_i)$ and that y_i $(i = 1, \ldots, d)$ are independently distributed according to $\mathrm{Po}(t\lambda_i)$, where $\mathrm{Po}(s\lambda_i)$ is the Poisson distribution with mean $s\lambda_i$, $\lambda := (\lambda_1, \ldots, \lambda_d)$ is the unknown parameter, and s and t are known positive constants. We consider prediction of $y = (y_1, y_2, \ldots, y_d)$ using $x = (x_1, x_2, \ldots, x_d)$ under the Kullback–Leibler loss.

A natural class of priors including the Jeffreys prior $\pi_{\mathrm{J}}(\lambda)\mathrm{d}\lambda_1 \cdots \mathrm{d}\lambda_d = \lambda_1^{-\frac{1}{2}} \cdots \lambda_d^{-\frac{1}{2}} \mathrm{d}\lambda_1 \cdots \mathrm{d}\lambda_d$ is $\pi_\alpha(\lambda)\mathrm{d}\lambda_1 \cdots \mathrm{d}\lambda_d := \lambda_1^{\alpha_1 - 1} \cdots \lambda_d^{\alpha_d - 1} \mathrm{d}\lambda_1 \cdots \mathrm{d}\lambda_d$, where $\alpha_i > 0$ $(i = 1, \ldots, d)$. A class of improper prior densities

$$
\begin{aligned}
&\pi_{\alpha, \gamma}(\lambda)\mathrm{d}\lambda_1 \mathrm{d}\lambda_2 \cdots \mathrm{d}\lambda_d \\
&= \frac{\lambda_1^{\alpha_1 - 1} \lambda_2^{\alpha_2 - 1} \cdots \lambda_d^{\alpha_d - 1}}{(\lambda_1 + \lambda_2 + \cdots + \lambda_d)^\gamma} \mathrm{d}\lambda_1 \mathrm{d}\lambda_2 \cdots \mathrm{d}\lambda_d
\end{aligned}
\tag{1}
$$

with $\sum_i \alpha_i - \gamma > 0$ and $\alpha_i > 0$ $(i = 1, 2, \ldots, d)$ is investigated and Theorems 1 and 2 below are obtained ([21]).

Theorem 1 ([21]): When $\sum_i \alpha_i - \gamma > 1$ and $\alpha_i > 0$ $(i = 1, 2, \ldots, d)$, the Bayesian predictive density $p_{\alpha, \gamma}(y \mid x)$ based on $\pi_{\alpha, \gamma}(\lambda)$ is dominated by the Bayesian predictive density $p_{\tilde{\alpha}, \tilde{\gamma}}(y \mid x)$ based on $\pi_{\tilde{\alpha}, \tilde{\gamma}}(\lambda)$, where $\tilde{\gamma} := \sum_i \alpha_i - 1$ and $\tilde{\alpha} = (\tilde{\alpha}_1, \tilde{\alpha}_2, \ldots, \tilde{\alpha}_d) := (\alpha_1, \alpha_2, \ldots, \alpha_d)$.

Theorem 2 ([21]): For every $d \geq 1$, the Bayesian predictive densities based on the priors in the class $\{\pi_{\alpha, \gamma}(\lambda) : 0 <$

$\sum_{i=1}^d \alpha_i - \gamma \leq 1, \alpha_i > 0 \ (i = 1, 2, \ldots, d)\}$ defined by (1) are admissible under the Kullback–Leibler loss.

In particular, when $d \geq 3$, the Bayesian predictive density $p_{\pi_{\mathrm{S}}}(y \mid x)$ based on the shrinkage prior $\pi_{\mathrm{S}}(\lambda) := \pi_{\alpha=(\frac{1}{2}, \ldots, \frac{1}{2}), \gamma = \frac{d}{2} - 1}(\lambda)$ dominates the Bayesian predictive density $p_{\pi_{\mathrm{J}}}(y \mid x)$ based on the Jeffreys prior and is admissible under the Kullback–Leibler loss. Parameter estimation can be regarded as infinitesimal prediction under the Kullback–Leibler loss in the multivariate Poisson model ([22]). Intuitively, infinitesimal prediction means prediction for infinitely near future.

In the present paper, we generalize the results for finite-dimensional independent Poisson models to the results for nonhomogeneous Poisson models. The remainder of the present paper paper is organized as follows. In Section 2, a class of improper shrinkage priors for nonparametric Bayesian inference with kernel mixtures for nonhomogeneous Poisson models is introduced. Several theorems concerning admissibility of Bayesian predictive densities and Bayes estimators for nonhomogeneous Poisson models corresponding to those for finite-dimensional models are proved. In Section 3, numerical methods to evaluate Bayesian predictive densities and Bayes estimators are investigated. Finally, conclusions are presented in Section 4.

II. Nonparametric Bayesian Inference for Nonhomogeneous Poisson Processes

We consider nonhomogeneous Poisson processes on a region U in the Euclidean space $\mathbb{R}^d$. The results in the following can be generalized to those for nonhomogeneous Poisson processes on general spaces such as a Polish space.

Basic properties of nonparametric inference of nonhomogeneous Poisson processes using gamma process priors are given by [24] and [26]. Corresponding results for probability density estimation are given by [25]. Various estimation methods of intensity functions of nonhomogeneous Poisson process models have been studied (e.g. [4], [18]).

Conventional statistical decision theory mainly deals with problems with finite dimensional parameters. In the present paper, we show that the framework of statistical decision theory is effectively applied to prediction and estimation for nonhomogeneous Poisson process models with infinite dimensional parameter spaces. The results suggest that decision theoretic approach is also useful for other infinite dimensional problems.

A. Bayes Estimators and Bayesian Predictive Densities

Let $\lambda(u)$ be a positive function of $u \in U$ satisfying $\int_U \lambda(u)\mathrm{d}u < \infty$. We observe $\boldsymbol{x} = (N, x_1, x_2, \ldots, x_N)$ distributed according to the nonhomogeneous Poisson process $\mathscr{P}o(s\lambda)$ with intensity function $s\lambda$ $(s > 0)$. Here, N is the number of the observed points and $x_1, \ldots, x_N$ are observed points in U. We introduce a known constant $s > 0$ for later use. Let $X(B) := \#\{x_i \mid x_i \in B \ (1 \leq i \leq N)\}$ and $\lambda(B) := \int_B \lambda(u)\mathrm{d}u$ for $B \subset U$. If $X(B_i)$ $(i = 1, \ldots, k)$ are independently distributed according to the Poisson distribution with mean $s\lambda(B_i)$ for an arbitrary partition $(B_1, \ldots, B_k)$ of

U, $\boldsymbol{x}$ is said to be distributed according to the nonhomogeneous Poisson distribution with intensity function $s\lambda$. The function λ is the unknown parameter.

The likelihood for the nonhomogeneous Poisson process model is given by

$$\left\{\prod_{i=1}^{N} s\lambda(x_i)\right\} \exp\left\{-s \int \lambda(x)\mathrm{d}x\right\}$$

$$= \left\{\prod_{i=1}^{N} \overline{\lambda}(x_i)\right\}(sw)^N \exp(-sw) \tag{2}$$

$$\propto \left\{\prod_{i=1}^{N} \overline{\lambda}(x_i)\right\} \frac{(sw)^N}{N!} \exp(-sw) =: p_\lambda(\boldsymbol{x}), \tag{3}$$

where $w = |\lambda| := \int_U \lambda(u)\mathrm{d}u$ and $\overline{\lambda} := \lambda/w$ (see e.g. [8] p. 22). We identify λ with $(w, \overline{\lambda})$. The probability density (3) of the observed points $(x_1, \ldots, x_N)$ multiplied by $N!$ coincides with (2). These two representations do not make essential differences in the following discussions. We mainly use (3) in the following.

Let $\boldsymbol{y} = (M, y_1, y_2, \ldots, y_M)$ be a sample independent of $\boldsymbol{x}$ from the nonhomogeneous Poisson process $\mathscr{P}_o(t\lambda)$, where $t > 0$ is a known constant. We investigate estimation of λ and prediction of $\boldsymbol{y}$ using $\boldsymbol{x}$. In Subsection II-D, estimation is formulated as a limit of prediction.

First, we consider estimation of λ. From (3), the Kullback–Leibler divergence from the probability density $p_\lambda(\boldsymbol{y})$ corresponding to the intensity $t\lambda$ to another probability density $p_{\lambda'}(\boldsymbol{y})$ corresponding to the intensity $t\lambda'$ is

$$D(p_\lambda(\boldsymbol{y}), p_{\lambda'}(\boldsymbol{y}))$$

$$= \mathrm{E}_\lambda \log \frac{\left\{\prod_{i=1}^{M} \overline{\lambda}(y_i)\right\} \dfrac{(tw)^M}{M!} \exp(-tw)}{\left\{\prod_{i=1}^{M} \overline{\lambda'}(y_i)\right\} \dfrac{(tw')^M}{M!} \exp(-tw')}$$

$$= \mathrm{E}_\lambda \left[tw' - tw + M \log \frac{w}{w'} + \sum_{i=1}^{M} \log \frac{\overline{\lambda}(y_i)}{\overline{\lambda'}(y_i)}\right]$$

$$= tw\left(\frac{w'}{w} - 1 - \log \frac{w'}{w} + \int \overline{\lambda}(y) \log \frac{\overline{\lambda}(y)}{\overline{\lambda'}(y)}\mathrm{d}y\right) \tag{4}$$

$$= t \int \left(\lambda'(y) - \lambda(y) + \lambda(y) \log \frac{\lambda(y)}{\lambda'(y)}\right)\mathrm{d}y. \tag{5}$$

Suppose that a prior density $\pi(\mathrm{d}\lambda)$ is adopted and observation $\boldsymbol{x}$ is given. Let $p_\pi(\mathrm{d}\lambda \mid \boldsymbol{x})$ be the posterior distribution. If the posterior mean of λ has a density $\lambda_{\pi,\boldsymbol{x}}(u)$ with respect to the Lebesgue measure on U, the posterior mean of the Kullback–Leibler loss of an intensity estimator $\widehat{\lambda}$ is

$$\int D(p_\lambda(\boldsymbol{y}), p_{\lambda'}(\boldsymbol{y})) \, p_\pi(\mathrm{d}\lambda \mid \boldsymbol{x})$$

$$= t \iint \left(\widehat{\lambda}(y) - \lambda(y) + \lambda(y) \log \frac{\lambda(y)}{\widehat{\lambda}(y)}\right)\mathrm{d}y \, p_\pi(\mathrm{d}\lambda \mid \boldsymbol{x})$$

$$= t \iint \left\{\left(\widehat{\lambda}(y) - \lambda_{\pi,\boldsymbol{x}}(y) + \lambda_{\pi,\boldsymbol{x}}(y) \log \frac{\lambda_{\pi,\boldsymbol{x}}(y)}{\widehat{\lambda}(y)}\right)\right.$$

$$+ \left.\left(\lambda_{\pi,\boldsymbol{x}}(y) - \lambda(y) + \lambda(y) \log \frac{\lambda(y)}{\lambda_{\pi,\boldsymbol{x}}(y)}\right)\right\}$$

$$\times p_\pi(\mathrm{d}\lambda \mid \boldsymbol{x})\mathrm{d}y$$

and is minimized when $\widehat{\lambda} = \lambda_{\pi,\boldsymbol{x}}$. Here, we assume that the integral exists. Thus, the Bayes estimator of λ is the posterior mean $\lambda_{\pi,\boldsymbol{x}}$ given observation $\boldsymbol{x}$.

If the posterior is decomposed as $p_\pi(\mathrm{d}w \, \mathrm{d}\overline{\lambda} \mid \boldsymbol{x}) = p_\pi(\mathrm{d}w \mid \boldsymbol{x})p_\pi(\mathrm{d}\overline{\lambda} \mid \boldsymbol{x})$, then the Bayes estimators of w, $\overline{\lambda}$, and λ based on the prior π are given by

$$w_{\pi,\boldsymbol{x}} = \int w \, p_\pi(\mathrm{d}w \mid \boldsymbol{x}), \quad \overline{\lambda}_{\pi,\boldsymbol{x}} = \int \overline{\lambda} \, p_\pi(\mathrm{d}\overline{\lambda} \mid \boldsymbol{x}),$$

and

$$\lambda_{\pi,\boldsymbol{x}} = w_{\pi,\boldsymbol{x}} \overline{\lambda}_{\pi,\boldsymbol{x}},$$

respectively.

Next, we consider predictive densities of $\boldsymbol{y}$. The Kullback–Leibler divergence from the probability density $p_\lambda(\boldsymbol{y})$ corresponding to the intensity $t\lambda$ to another probability density $q(\boldsymbol{y}) = q(M, y_1, \ldots, y_M) = q(M)q(y_1, \ldots, y_M \mid M)$ is

$$D(p_\lambda(\boldsymbol{y}), q(\boldsymbol{y})) = \mathrm{E}_\lambda \left[\log \frac{p_{tw}^{\mathrm{Po}}(M) \prod_{i=1}^{M} \overline{\lambda}(y_i)}{q(M, y_1, \ldots, y_M)}\right], \tag{6}$$

where $p_{tw}^{\mathrm{Po}}(M) := \{(tw)^M/M!\} \exp(-tw)$ denotes the Poisson probability density with mean tw.

Since the posterior mean of the Kullback–Leibler divergence is

$$\int D(p_\lambda(\boldsymbol{y}), q(\boldsymbol{y}; \boldsymbol{x}))p_\pi(\mathrm{d}\lambda \mid \boldsymbol{x})$$

$$= \iint \sum_{M=0}^{\infty} \int \cdots \int p_{tw}^{\mathrm{Po}}(M)\left\{\prod_{i=1}^{M} \overline{\lambda}(y_i)\right\}$$

$$\times \log \frac{p_{tw}^{\mathrm{Po}}(M) \prod_{i=1}^{M} \overline{\lambda}(y_i)}{q(M, y_1, \ldots, y_M)}\mathrm{d}y_1 \cdots \mathrm{d}y_M p_\pi(\mathrm{d}w \, \mathrm{d}\overline{\lambda} \mid \boldsymbol{x}), \tag{7}$$

the Bayesian predictive density minimizing (7) is

$$p_\pi(M, y_1, \ldots, y_M \mid \boldsymbol{x})$$

$$= \iint p_{tw}^{\mathrm{Po}}(M)\left\{\prod_{i=1}^{M} \overline{\lambda}(y_i)\right\}p_\pi(\mathrm{d}w \, \mathrm{d}\overline{\lambda} \mid \boldsymbol{x}).$$

If the posterior is decomposed as $p_\pi(\mathrm{d}w \, \mathrm{d}\overline{\lambda} \mid \boldsymbol{x}) = p_\pi(\mathrm{d}w \mid \boldsymbol{x})p_\pi(\mathrm{d}\overline{\lambda} \mid \boldsymbol{x})$, then the Bayesian predictive density based on the prior π is given by

$$p_\pi(M, y_1, \ldots, y_M \mid \boldsymbol{x}) = p_\pi(M \mid \boldsymbol{x})p_\pi(y_1, \ldots, y_M \mid M, \boldsymbol{x}), \tag{8}$$

where

$$p_\pi(M \mid \boldsymbol{x}) = \int p_{tw}^{\mathrm{Po}}(M)p_\pi(\mathrm{d}w \mid \boldsymbol{x})$$

and

$$p_\pi(y_1, \ldots, y_M \mid M, \boldsymbol{x}) = \int \left\{\prod_{i=1}^{M} \overline{\lambda}(y_i)\right\}p_\pi(\mathrm{d}\overline{\lambda} \mid \boldsymbol{x}).$$

B. Kernel Mixture Models

We need to adopt a prior for λ in order to use Bayesian methods. First, we consider the gamma process prior $\mathscr{G}a(\alpha, \beta)$, where $\alpha(u)$ is a positive density function of u and β is a positive scalar that does not depend on u. Then, for an arbitrary partition $(B_1, \ldots, B_k)$ of U, $\mu_i := \mu(B_i)$ $(i = 1, \ldots, k)$ are independently distributed according to the gamma distributions $\mathrm{Ga}(\alpha_i, \beta)$ with densities $\{1/\Gamma(\alpha_i)\}(\mu_i^{\alpha_i-1}/\beta^{\alpha_i}) \exp(-\mu_i/\beta)$, where $\alpha_i := \alpha(B_i) = \int_{B_i} \alpha(u)\mathrm{d}u$. The mixture of the nonhomogeneous Poisson process $\mathscr{P}o(t\lambda)$ with respect to the prior $\mathscr{G}a(\alpha, \beta)$ for λ is the negative binomial process $\mathscr{N}e\mathscr{B}i(\alpha, t\beta/(1 + t\beta))$. For $\mathscr{N}e\mathscr{B}i(\alpha, t\beta/(1+t\beta))$ and an arbitrary partition $(B_1, \ldots, B_k)$ of U, the numbers N_i of points in B_i $(i = 1, \ldots, k)$ are independently distributed according to the negative binomial distribution with density $\binom{N_i+\alpha_i-1}{\alpha_i}\{t\beta/(1 + t\beta)\}^{N_i}\{1/(1+t\beta)\}^{\alpha_i}$. The posterior with respect to the prior $\mathscr{G}a(\alpha, \beta)$ and observation $\boldsymbol{x}$ is $\mathscr{G}a(\alpha+\sum_i \delta_{x_i}, 1/(s+1/\beta))$ (see [26]). Then, the posterior means of w, $\overline{\lambda}$, and λ are

$$w_{\alpha,\beta,\boldsymbol{x}} = \frac{\beta}{1+s\beta}(|\alpha| + N), \ \overline{\lambda}_{\alpha,\beta,\boldsymbol{x}} = \frac{\alpha + \sum_i \delta_{x_i}}{|\alpha| + N},$$

and

$$\lambda_{\alpha,\beta,\boldsymbol{x}} = \frac{\beta}{1+s\beta}\left(\alpha + \sum_i \delta_{x_i}\right),$$

respectively, where δ_{x_i} denotes the Dirac measure at x_i. Thus, the Bayesian predictive density based on the gamma process prior $\mathscr{G}a(\alpha, \beta)$ and observation $\boldsymbol{x}$ is

$$\mathscr{N}e\mathscr{B}i\left(\alpha + \sum_i \delta_{x_i}, \frac{(s+1/\beta)^{-1}t}{1 + (s+1/\beta)^{-1}t}\right)$$
$$= \mathscr{N}e\mathscr{B}i\left(\alpha + \sum_i \delta_{x_i}, \frac{t\beta}{1 + (s+t)\beta}\right).$$

Although the gamma process prior is a conjugate prior for the nonhomogeneous Poisson model, it is not natural to use this prior directly for λ because the measure $\alpha+\sum_i \delta_{x_i}$ is not absolutely continuous with respect to the intensity measure λ. In fact, the Kullback–Leibler loss of the posterior mean $\lambda_{\alpha,\beta,\boldsymbol{x}}$ with respect to observation $\boldsymbol{x}$ and the gamma prior $\mathscr{G}a(\alpha, \beta)$ is

$$D(t\lambda, t\lambda_{\alpha,\beta,\boldsymbol{x}})$$
$$= tw\left\{\frac{w_{\alpha,\beta}}{w} - 1 - \log\frac{w_{\alpha,\beta}}{w} + \int \overline{\lambda}(y) \log\frac{\overline{\lambda}(y)}{\overline{\lambda}_{\alpha,\beta,\boldsymbol{x}}(y)}\mathrm{d}y\right\}$$
$$= tw\left\{\frac{w_{\alpha,\beta}}{w} - 1 - \log\frac{w_{\alpha,\beta}}{w} + \int \overline{\lambda}(y) \log\frac{\overline{\lambda}(y)}{\overline{\alpha}(y)}\mathrm{d}y\right.$$
$$\left. + \log\left(1 + \frac{N}{|\alpha|}\right)\right\},$$

where $|\alpha| := \int_U \alpha(u)\mathrm{d}u$ and $\overline{\alpha} := \alpha/|\alpha|$. The amount of information concerning λ included in $\boldsymbol{x}$ increases as s increases. However, when s goes to infinity and t is fixed, N and the divergence $D(t\lambda, t\lambda_{\alpha,\beta})$ diverges to infinity. Here, $w_{\alpha,\beta}$ converges to w with probability 1. However, $\overline{\lambda}_{\alpha,\beta,\boldsymbol{x}}$ does not converge to $\overline{\lambda}$ in the Kullback–Leibler sense, although $\overline{\lambda}_{\alpha,\beta,\boldsymbol{x}}$ weakly converges to $\overline{\lambda}$.

In order to overcome the difficulty caused by the fact that $\alpha+\sum_i \delta_i$ is not absolutely continuous with respect to λ, kernel mixture models are widely used. A nonnegative function $k : U \times U \to \mathbb{R}_{\geq 0}$ satisfying $\int k(y, u)\mathrm{d}y = 1$ is called a kernel function. Suppose that an intensity function λ is represented by

$$\lambda(y) = k(y, \mu) := \int k(y, u)\mu(\mathrm{d}u),$$

where $\mu(\mathrm{d}u)$ is a finite measure on U. Then,

$$w = |\lambda| := \int \lambda(y)\mathrm{d}y = \iint k(y, u)\mu(\mathrm{d}u)\mathrm{d}y$$
$$= \int \mu(\mathrm{d}y) = |\mu|.$$

We identify μ with $(w, \overline{\mu})$, where $\overline{\mu} := \mu/|\mu|$. Then, we have

$$\overline{\lambda}(y) = \frac{\lambda(y)}{|\lambda|} = k(y, \overline{\mu}) = \int k(y, u)\overline{\mu}(\mathrm{d}u).$$

We hereinafter consider kernel mixture models because these models are reasonable under the Kullback–Leibler loss. The results in the following can be generalized to the setting in which the kernel function has an unknown parameter by introducing a prior for the unknown parameter.

Example: The Gaussian kernel

$$k(y, u) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(y - u)^2\right\} \qquad (y, u \in \mathbb{R})$$

is frequently used in applications. For simplicity, we assume that $\sigma > 0$ is known. □

Assume a gamma process prior $\mathscr{G}a(\alpha, \beta)$ for μ. Then, $\overline{\mu} := \mu/w$ is distributed according to the Dirichlet process $\mathscr{D}i(\alpha)$ (see e.g. [13] p. 96). For every partition $(B_1, \ldots, B_k)$ of U, $(\overline{\mu}(B_1), \ldots, \overline{\mu}(B_k))$ is distributed according to the k-dimensional Dirichlet distribution $\mathrm{Di}(\alpha(B_1), \ldots, \alpha(B_k))$. The weight parameter w is distributed according to the gamma distribution $\mathrm{Ga}(|\alpha|, \beta)$ independently of $\overline{\mu}$. Thus, the simultaneous distribution of w and $\overline{\mu}$ is given by

$$p_\alpha^{\mathscr{D}i}(\mathrm{d}\overline{\mu})\, p_{|\alpha|,\beta}^{\mathrm{Ga}}(w)\, \mathrm{d}w, \tag{9}$$

where $p_\alpha^{\mathscr{D}i}(\mathrm{d}\overline{\mu})$ denotes the Dirichlet process measure and $p_{|\alpha|,\beta}^{\mathrm{Ga}}(w) = \{1/\Gamma(|\alpha|)\}(w^{|\alpha|-1}/\beta^{|\alpha|}) \exp(-w/\beta)$ is the gamma probability density.

C. Improper Priors

It is difficult to determine the scale parameter β from the viewpoint of objective Bayes. Let c be an arbitrary positive constant for time scale change. Then, $\mathscr{P}o(s\lambda) = \mathscr{P}o(\tilde{s}\tilde{\lambda})$, where $\tilde{s} := cs$ and $\tilde{\lambda} := \lambda/c$. Inference for λ is equivalent to inference for $\tilde{\lambda}$ because the time scale change does not affect the essence of the problem. Thus, the objective prior should be (relatively) invariant with respect to the time scale change. However, the gamma process prior $\mathscr{G}a(\alpha, \beta)$ is not relatively invariant if β is finite. One method by which to construct an invariant posterior is to adopt the improper prior

$$\pi_\alpha(\mu)\mathrm{d}\mu = p_\alpha^{\mathscr{D}i}(\mathrm{d}\overline{\mu})\, p_{|\alpha|,\infty}^{\mathrm{Ga}}(w)\, \mathrm{d}w, \tag{10}$$

which could be intuitively denoted by $\mu^{\alpha-1}\mathrm{d}\mu$, that is obtained by taking the limit $\beta \to \infty$. The posterior based on the prior π_α is invariant with respect to the time scale change. The prior (10) is a natural generalization of the improper prior $\prod_{i=1}^{d}(\mu_i^{\alpha_i-1}\mathrm{d}\mu_i)$ discussed by [21] for the finite-dimensional independent Poisson model.

Here, we consider a generalization of the gamma process prior (9). Let

$$\pi_{\alpha,\beta,\gamma}(\mathrm{d}\mu) := p_\alpha^{\mathscr{D}}(\mathrm{d}\overline{\mu})\, p_{|\alpha|-\gamma,\beta}^{\mathrm{Ga}}(w)\, \mathrm{d}w, \qquad (11)$$

where $\gamma < |\alpha|$. We consider

$$\pi_{\alpha,\gamma}(\mathrm{d}\mu) := \pi_{\alpha,\beta=\infty,\gamma}(\mathrm{d}\mu) = p_\alpha^{\mathscr{D}}(\mathrm{d}\overline{\mu})\, p_{|\alpha|-\gamma,\infty}^{\mathrm{Ga}}(w)\, \mathrm{d}w, \qquad (12)$$

which is a generalization of (10), by taking the limit $\beta \to \infty$. We denote the distributions of $\lambda(u) = \int k(u,v)\mu(\mathrm{d}v)$ corresponding to (11) and (12) by $\pi_{\alpha,\beta,\gamma}(\mathrm{d}\lambda)$ and $\pi_{\alpha,\gamma}(\mathrm{d}\lambda)$, respectively, by abuse of notation without confusion.

We investigate Bayesian inference based on the prior $\pi_{\alpha,\beta,\gamma}$. From (3) and (11), the posterior distribution of μ with respect to prior $\pi_{\alpha,\beta,\gamma}$ and observed data $\boldsymbol{x}$ is proportional to

$$\left\{\prod_{i=1}^{N} k(x_i,\overline{\mu})\right\} p_\alpha^{\mathscr{D}}(\mathrm{d}\overline{\mu}) \frac{(sw)^N}{N!} \exp(-sw) p_{|\alpha|-\gamma,\beta}^{\mathrm{Ga}}(w)\, \mathrm{d}w$$

$$\propto \left\{\prod_{i=1}^{N} k(x_i,\overline{\mu})\right\} p_\alpha^{\mathscr{D}}(\mathrm{d}\overline{\mu}) p_{|\alpha|-\gamma+N,\beta/(s\beta+1)}^{\mathrm{Ga}}(w)\, \mathrm{d}w. \qquad (13)$$

Thus, the posterior and the Bayesian predictive density for the kernel mixture models have more complex forms than those for the simple Gamma–Poisson processes.

The posterior mean of $\overline{\lambda}$ given $\boldsymbol{x}$ is

$$\overline{\lambda}_{\alpha,\boldsymbol{x}}(y) = \frac{\lambda_{\alpha,\beta,\gamma,\boldsymbol{x}}(y)}{|\lambda_{\alpha,\beta,\gamma,\boldsymbol{x}}|} = \frac{\int k(y,\overline{\mu})\{\prod_{i=1}^{N} k(x_i,\overline{\mu})\} p_\alpha^{\mathscr{D}}(\mathrm{d}\overline{\mu})}{\int \{\prod_{i=1}^{N} k(x_i,\overline{\mu})\} p_\alpha^{\mathscr{D}}(\mathrm{d}\overline{\mu})}, \qquad (14)$$

not depending on β, γ, or s. For precise treatment of quantities related to posterior distributions based on disintegration, see e.g. [17].

The posterior means of w and λ are

$$w_{|\alpha|-\gamma,\beta,N} = |\lambda_{\alpha,\beta,\gamma,\boldsymbol{x}}| = \frac{1}{s+1/\beta}(|\alpha|-\gamma+N) \qquad (15)$$

and

$$\lambda_{\alpha,\beta,\gamma,\boldsymbol{x}} = w_{|\alpha|-\gamma,\beta,N}\overline{\lambda}_{\alpha,\boldsymbol{x}} = \frac{1}{s+1/\beta}(|\alpha|-\gamma+N)\overline{\lambda}_{\alpha,\boldsymbol{x}}, \qquad (16)$$

respectively. Since the posterior density of w depends on α and γ only through $|\alpha|-\gamma$ and on $\boldsymbol{x} = (N, x_1, \ldots, x_N)$ only through N, we denote the posterior mean of w by $w_{|\alpha|-\gamma,\beta,N} := |\lambda_{\alpha,\beta,\gamma,\boldsymbol{x}}|$.

From (13), the Bayesian predictive density is given by

$$p_{t/(t+s+1/\beta),|\alpha|-\gamma+N}^{\mathrm{NB}}(M) p_\alpha(y_1, \ldots, y_M \mid M, \boldsymbol{x}),$$

where

$$p_\alpha(y_1, \ldots, y_M \mid M, \boldsymbol{x})$$
$$= \frac{\int \{\prod_{j=1}^{M} k(y_j,\overline{\mu})\}\{\prod_{i=1}^{N} k(x_i,\overline{\mu})\} p_\alpha^{\mathscr{D}}(\mathrm{d}\overline{\mu})}{\int \{\prod_{i=1}^{N} k(x_i,\overline{\mu})\} p_\alpha^{\mathscr{D}}(\mathrm{d}\overline{\mu})}. \qquad (17)$$

and

$$p_{t/(s+t+1/\beta),|\alpha|-\gamma+N}^{\mathrm{NB}}(M)$$
$$= \frac{\Gamma(M+N+|\alpha|-\gamma)}{M!\,\Gamma(N+|\alpha|-\gamma)}$$
$$\times \left(\frac{t}{t+s+1/\beta}\right)^M \left(\frac{s+1/\beta}{t+s+1/\beta}\right)^{|\alpha|-\gamma+N}.$$

Here, $p_{t/(s+t+1/\beta),|\alpha|-\gamma+N}^{\mathrm{NB}}(M)$ is the negative binomial distribution with success probability $t/(t+s+1/\beta)$ and number of failures $|\alpha|-\gamma+N$. We can evaluate the posterior mean (16) of λ and the predictive density (17) using Markov chain Monte Carlo (MCMC) methods as discussed in Section III.

By taking the limit $\beta \to \infty$, we obtain the Bayes estimate and the Bayesian predictive density based on the improper prior $\pi_{\alpha,\gamma}$. From (13), the posterior with respect to the improper prior $\pi_{\alpha,\gamma}$ and observation $\boldsymbol{x}$ is proportional to

$$\left\{\prod_{i=1}^{N} k(x_i,\overline{\mu})\right\} p_\alpha^{\mathscr{D}}(\mathrm{d}\overline{\mu}) \frac{(sw)^N}{N!} \exp(-sw) p_{|\alpha|-\gamma,\infty}^{\mathrm{Ga}}(w)\, \mathrm{d}w$$

$$\propto \left\{\prod_{i=1}^{N} k(x_i,\overline{\mu})\right\} p_\alpha^{\mathscr{D}}(\mathrm{d}\overline{\mu}) p_{|\alpha|-\gamma+N,1/s}^{\mathrm{Ga}}(w)\, \mathrm{d}w.$$

The posterior means of $\overline{\lambda}$, w and λ with the improper prior $\pi_{\alpha,\gamma}$ are

$$\overline{\lambda}_{\alpha,\boldsymbol{x}}(y) = \frac{\int k(y,\overline{\mu})\{\prod_{i=1}^{N} k(x_i,\overline{\mu})\} p_\alpha^{\mathscr{D}}(\mathrm{d}\overline{\mu})}{\int \{\prod_{i=1}^{N} k(x_i,\overline{\mu})\} p_\alpha^{\mathscr{D}}(\mathrm{d}\overline{\mu})}, \qquad (18)$$

$$w_{|\alpha|-\gamma,N} = |\lambda_{\alpha,\gamma,\boldsymbol{x}}| = \frac{1}{s}(|\alpha|-\gamma+N), \qquad (19)$$

and

$$\lambda_{\alpha,\gamma,\boldsymbol{x}} = w_{|\alpha|-\gamma,N}\overline{\lambda}_{\alpha,\boldsymbol{x}} = \frac{1}{s}(|\alpha|-\gamma+N)\overline{\lambda}_{\alpha,\boldsymbol{x}}, \qquad (20)$$

respectively.

The Bayesian predictive density with the improper prior $\pi_{\alpha,\gamma}$ is given by

$$p_{t/(t+s),|\alpha|-\gamma+N}^{\mathrm{NB}}(M) p_\alpha(y_1, \ldots, y_M \mid M, \boldsymbol{x}),$$

where

$$p_{t/(t+s),|\alpha|-\gamma+N}^{\mathrm{NB}}(M)$$
$$= \frac{\Gamma(M+N+|\alpha|-\gamma)}{M!\,\Gamma(N+|\alpha|-\gamma)} \left(\frac{t}{t+s}\right)^M \left(\frac{s}{t+s}\right)^{|\alpha|-\gamma+N}.$$

D. Estimation as Infinitesimal Prediction

For the finite-dimensional independent Poisson model, Bayesian parameter estimation under the Kullback–Leibler loss is formulated as infinitesimal Bayesian prediction ([22], [23]). We derive the corresponding results for nonhomogeneous Poisson processes. Using this formulation, an integral representation of the Kullback–Leibler risk of a predictive density is obtained. This representation provides a unified framework for prediction and estimation and is a basis for later discussions. Intuitively, for the nonhomogeneous Poisson model, estimation is infinitesimal prediction and prediction is cumulative estimation.

Let $\boldsymbol{z} = (N + M, x_1, \ldots, x_N, y_1, \ldots, y_M)$. Then, the density of $\boldsymbol{z}$ is

$$p_\lambda(\boldsymbol{z}) = \left\{ \prod_{i=1}^{N+M} \overline{\lambda}(z_i) \right\} \frac{\{(s+t)w\}^{N+M}}{(N+M)!} \exp\{-(s+t)w\}. \tag{21}$$

Since the conditional density

$$p_\lambda(\boldsymbol{x} \mid \boldsymbol{z}) = \frac{p_\lambda(\boldsymbol{x}, \boldsymbol{z})}{p_\lambda(\boldsymbol{z})} = \frac{p_\lambda(\boldsymbol{x}) p_\lambda(\boldsymbol{y})}{p_\lambda(\boldsymbol{z})}$$
$$= \binom{N+M}{N} \left(\frac{s}{s+t} \right)^N \left(\frac{t}{s+t} \right)^M \tag{22}$$

does not depend on λ, $\boldsymbol{z}$ is a sufficient statistic when $\boldsymbol{x}$ and $\boldsymbol{y}$ are observed. We denote $p_\lambda(\boldsymbol{x} \mid \boldsymbol{z})$ by $p(\boldsymbol{x} \mid \boldsymbol{z})$. Let $p_\pi(\boldsymbol{x}, \boldsymbol{z}) := \int p_\lambda(\boldsymbol{x}, \boldsymbol{z}) \pi(\mathrm{d}\lambda)$ and $p_\pi(\boldsymbol{z}) := \int p_\lambda(\boldsymbol{z}) \pi(\mathrm{d}\lambda)$, where $\pi(\mathrm{d}\lambda)$ is a prior. Then, we have

$$p_\pi(\boldsymbol{y} \mid \boldsymbol{x}) = p_\pi(\boldsymbol{z} \mid \boldsymbol{x}) = \frac{p(\boldsymbol{x} \mid \boldsymbol{z}) p_\pi(\boldsymbol{z})}{p_\pi(\boldsymbol{x})}. \tag{23}$$

We consider a nonhomogeneous Poisson process $\boldsymbol{z}_\tau = (N_\tau, z_1, \ldots, z_{N_\tau})$ distributed according to $\mathscr{P}_\mathscr{O}(\tau\lambda)$ on U with time $\tau \geq 0$. When we need to explicitly specify time τ, z_i is denoted by $z_{\tau,i}$. For $a, b > 0$, the difference between $\boldsymbol{z}_{a+b} = (N_{a+b}, z_1, \ldots, z_{N_{a+b}})$ and $\boldsymbol{z}_a = (N_a, z_1, \ldots, z_{N_a})$ is defined by $\boldsymbol{z}_{a+b} - \boldsymbol{z}_a := (N_{a+b} - N_a, z_{N_a+1}, \ldots, z_{N_{a+b}})$. For all $a, b > 0$, $\boldsymbol{z}_a$ and $\boldsymbol{z}_{a+b} - \boldsymbol{z}_a$ are independently distributed according to $\mathscr{P}_\mathscr{O}(a\lambda)$ and $\mathscr{P}_\mathscr{O}(b\lambda)$, respectively. This spatial point process $\boldsymbol{z}_\tau$ on U is called a pure immigration process. Then, the simultaneous distribution of $\boldsymbol{x}$ and $\boldsymbol{y}$ is identical to that of $\boldsymbol{z}_s$ and $\boldsymbol{z}_{s+t} - \boldsymbol{z}_s$. The probability density of $\boldsymbol{z}_\tau$ is denoted by $p_{\lambda,\tau}(\boldsymbol{z}_\tau)$. From (23), the Kullback–Leibler risk is represented by

$$\mathrm{E}_\lambda \left[\log \frac{p_\lambda(\boldsymbol{y} \mid \boldsymbol{x})}{p_\pi(\boldsymbol{y} \mid \boldsymbol{x})} \right]$$
$$= \mathrm{E}_\lambda \left[\log \frac{p_{\lambda,s+t}(\boldsymbol{z}_{s+t})}{p_{\pi,s+t}(\boldsymbol{z}_{s+t})} \right] - \mathrm{E}_\lambda \left[\log \frac{p_{\lambda,s}(\boldsymbol{z}_s)}{p_{\pi,s}(\boldsymbol{z}_s)} \right]$$
$$= \int_s^{s+t} \frac{\partial}{\partial \tau} \mathrm{E}_\lambda \left[\log \frac{p_{\lambda,\tau}(\boldsymbol{z}_\tau)}{p_{\pi,\tau}(\boldsymbol{z}_\tau)} \right] \mathrm{d}\tau, \tag{24}$$

where $p_{\pi,\tau}(\boldsymbol{z}_\tau) = \int p_{\lambda,\tau}(\boldsymbol{z}_\tau) \pi(\mathrm{d}\lambda)$. Here, we assume that the expectation $\mathrm{E}_\lambda \left[\log\{p_{\lambda,\tau}(\boldsymbol{z}_\tau)/p_{\pi,\tau}(\boldsymbol{z}_\tau)\} \right]$ exists and is differentiable with respect to $\tau \in [s, s+t]$.

In order to evaluate the integrand of (24), we prepare Lemma 2, which is a generalization of Lemma 1 below, which is essentially used in [23].

Lemma 1: Suppose that N_τ $(\tau \geq 0)$ is distributed according to a Poisson distribution with mean τw, where w is a fixed positive real number. Let h be a function from the nonnegative integers to the real numbers. Assume that $\sum_{n=0}^\infty |h(n)|(\theta^n/n!) \exp(-\theta) < \infty$ for all $\theta > 0$. Then,

$$\frac{\mathrm{d}}{\mathrm{d}\tau} \mathrm{E}\left[h(N_\tau) \right] = \mathrm{E}\left[\left(\frac{N_\tau}{\tau} - w \right) h(N_\tau) \right]$$
$$= w \, \mathrm{E}\left[h(N_\tau + 1) - h(N_\tau) \right].$$

The proof is straightforward.

Lemma 2: Suppose that $\boldsymbol{z}_\tau = (N_\tau, z_1, \ldots, z_{N_\tau})$ is distributed according to the nonhomogeneous Poisson process $\mathscr{P}_\mathscr{O}(\tau\lambda)$ $(\tau \geq 0)$.

Let h_n $(n = 0, 1, 2, \ldots)$ be functions from $U^n \times \mathbb{R}_{\geq 0}$ to $\mathbb{R}$, where $U^0 := \emptyset$ and h_0 is a constant. Every function $h_n(z_1, \ldots, z_n, \tau)$ is symmetric with respect to $z_1, \ldots, z_n$ and is differentiable with respect to $\tau \geq 0$ for every fixed $z_1, \ldots, z_n$. Let h be a function from $(\cup_{n=0}^\infty U^n) \times \mathbb{R}_{\geq 0}$ to $\mathbb{R}$ defined by $h(\boldsymbol{z}, \tau) = h(n, z_1, \ldots, z_n, \tau) := h_n(z_1, \ldots, z_n, \tau)$, where $\boldsymbol{z} = (n, z_1, \ldots, z_n)$. Assume that $f(r, s) := \mathrm{E}[h(\boldsymbol{z}_r, s)]$ exists and is differentiable with respect to $(r, s) \in \mathbb{R}_{>0}^2$ in a neighborhood of every $r = s > 0$ and that $\frac{\partial}{\partial s} \mathrm{E}[h(\boldsymbol{z}_r, s)] = \mathrm{E}[\frac{\partial}{\partial s} h(\boldsymbol{z}_r, s)]$.

Then,

$$\frac{\mathrm{d}}{\mathrm{d}\tau} \mathrm{E}\left[h(\boldsymbol{z}_\tau, \tau) \right] = \int_U \lambda(y) \mathrm{E}\left[h(\boldsymbol{z}_\tau + \delta_y) - h(\boldsymbol{z}_\tau) \right] \mathrm{d}y$$
$$+ \mathrm{E}\left[\left. \frac{\partial h}{\partial r}(\boldsymbol{z}_\tau, r) \right|_{r=\tau} \right], \tag{25}$$

where $\boldsymbol{z}_\tau + \delta_y := (N_\tau + 1, z_1, \ldots, z_{N_\tau}, y)$.

The proof is given in Appendix A.

Now, we have an integral representation of the risk of a Bayesian predictive density.

Theorem 3: Suppose that $\boldsymbol{z}_\tau = (N_\tau, z_1, \ldots, z_{N_\tau})$ is distributed according to the nonhomogeneous Poisson process $\mathscr{P}_\mathscr{O}(\tau\lambda)$ $(\tau \geq 0)$.

Assume that

$$f(r, s) := \mathrm{E}\left[\log \frac{p_{s,\lambda}(\boldsymbol{z}_r)}{p_{s,\pi}(\boldsymbol{z}_r)} \right]$$

exists and is differentiable with respect to $(r, s) \in \mathbb{R}_{>0}^2$ in a neighborhood of every $r = s > 0$ and that

$$\frac{\partial}{\partial s} \mathrm{E}\left[\log \frac{p_{s,\lambda}(\boldsymbol{z}_r)}{p_{s,\pi}(\boldsymbol{z}_r)} \right] = \mathrm{E}\left[\frac{\partial}{\partial s} \log \frac{p_{s,\lambda}(\boldsymbol{z}_r)}{p_{s,\pi}(\boldsymbol{z}_r)} \right].$$

Then,

$$\mathrm{E}_\lambda \left[\log \frac{p_\lambda(\boldsymbol{y} \mid \boldsymbol{x})}{p_\pi(\boldsymbol{y} \mid \boldsymbol{x})} \right]$$
$$= \int_s^{s+t} \mathrm{E}_\lambda \left[w_{\pi,\boldsymbol{z}_\tau,\tau} - w - w \log \frac{w_{\pi,\boldsymbol{z}_\tau,\tau}}{w} \right.$$
$$\left. + w \int_U \overline{\lambda}(y) \, \log \frac{\overline{\lambda}(y)}{\overline{\lambda}_{\pi,\boldsymbol{z}_\tau,\tau}(y)} \mathrm{d}y \right] \mathrm{d}\tau$$

$$= \int_s^{s+t} \mathrm{E}_\lambda \left[\int_U \left\{ \lambda_{\pi, \mathbf{z}_\tau, \tau}(y) - \lambda(y) \right. \right.$$
$$\left. \left. + \lambda(y) \log \frac{\lambda(y)}{\lambda_{\pi, \mathbf{z}_\tau, \tau}(y)} \right\} \mathrm{d}y \right] \mathrm{d}\tau. \quad (26)$$

The proof is given in Appendix A.

Since the integrand (26) multiplied by t coincides with the Kullback–Leibler loss (5) for the intensity estimator $\lambda_{\pi, \mathbf{z}, \tau}$, Bayes estimation under the Kullback–Leibler loss for the non-homogeneous Poisson model can be regarded as infinitesimal Bayesian prediction as in the finite-dimensional independent Poisson model.

When the prior is decomposable as $\pi(\mathrm{d}\lambda) = \pi(\mathrm{d}w)\pi(\mathrm{d}\overline{\lambda})$, the assumption in Theorem 3 and Theorem 4 in the next subsection can be easily verified in many problems. If $\pi(\mathrm{d}\lambda) = \pi(\mathrm{d}w)\pi(\mathrm{d}\overline{\lambda})$, then

$$f(r, s) = \mathrm{E}_\lambda \left[\log \frac{\prod_{i=1}^{N_r} \lambda(z_{r,i}) \exp(-sw)}{\int \prod_{j=1}^{N_r} \lambda'(z_{r,j}) \exp(-sw')\pi(\mathrm{d}\lambda')} \right]$$

$$= \mathrm{E}_\lambda \left[\log \frac{\prod_{i=1}^{N_r} w^{N_r} \exp(-sw)}{\int \prod_{j=1}^{N_r} w'^{N_r} \exp(-sw')\pi(\mathrm{d}w')} \right]$$

$$+ \mathrm{E}_\lambda \left[\log \frac{\prod_{i=1}^{N_r} \overline{\lambda}(z_{r,i})}{\int \prod_{j=1}^{N_r} \overline{\lambda}'(z_{r,j})\pi(\mathrm{d}\overline{\lambda}')} \right]. \quad (27)$$

The first term in (27) can be explicitly evaluated for our priors. The second term in (27) is represented by

$$h(r) := \mathrm{E}_\lambda \left[\log \left\{ \frac{\prod_{j=1}^n \overline{\lambda}(z_j)}{\int \prod_{k=1}^n \overline{\lambda}_{\overline{\mu}}(z_k)\pi(\mathrm{d}\overline{\mu})} \right\} \right]$$

$$= \exp(-rw) \sum_{n=0}^\infty \frac{(rw)^n}{n!} c(n), \quad (28)$$

where

$$c(n) := \int \cdots \int \left\{ \prod_{i=1}^n \overline{\lambda}(z_i) \right\} \log \left\{ \frac{\prod_{j=1}^n \overline{\lambda}(z_j)}{\int \prod_{k=1}^n \overline{\lambda}_{\overline{\mu}}(z_k)\pi(\mathrm{d}\overline{\mu})} \right\}$$
$$\mathrm{d}z_1 \cdots \mathrm{d}z_n.$$

Here, $c(n)$ is the cumulative Kullback–Leibler risk when we predict $z_1, \ldots, z_n$ independently distributed according to a probability density $\overline{\lambda}$ using the simultaneous Bayesian predictive density $\int \prod_{j=1}^{N_r} \overline{\lambda}'(z_{r,j})\pi(\mathrm{d}\overline{\lambda}')$.

Therefore, in order to verify the assumption in Theorem 3, it is sufficient to show that (28) converges for every $0 \leq r < \infty$.

In many problems, we can verify that (28) converges for every $0 \leq r < \infty$ using the inequality

$$0 \leq c(n) \leq n \left\{ \sup_{\overline{\mu} \in B} D(\overline{\lambda}, \overline{\lambda}_{\overline{\mu}}) + \int_B \pi(\mathrm{d}\overline{\mu}) \right\}, \quad (29)$$

where B is an arbitrary measurable set in the space of intensity functions. The inequality (29) is used to define the index of resolvability by [3] (see also [2]). Thus, a sufficient condition is that there exists B such that $\sup_{\overline{\mu} \in B} D(\overline{\lambda}, \overline{\lambda}_{\overline{\mu}}) < \infty$ and $\pi(B) > 0$. For example, the nonhomogeneous Poisson model with the gamma prior and without a kernel discussed in Subsection II-B does not satisfy this condition.

If the power series of r in (28) converges for all $0 \leq r < \infty$, we have

$$\frac{\mathrm{d}}{\mathrm{d}r} h(r) = -w \exp(-rw) \sum_{n=0}^\infty \frac{(rw)^n}{n!} c(n)$$

$$+ \exp(-rw) \sum_{n=1}^\infty \frac{w(rw)^{n-1}}{(n-1)!} c(n)$$

$$= w \exp(-rw) \sum_{n=0}^\infty \frac{(rw)^{n-1}}{n!} \{c(n+1) - c(n)\} \quad (30)$$

$$= w \mathrm{E}_\lambda \left[\int \overline{\lambda}(y) \log \frac{\overline{\lambda}(y)}{\overline{\lambda}_{\pi, \mathbf{z}_r}(y)} \mathrm{d}y \right],$$

because

$$c(n+1) - c(n)$$

$$= \int \overline{\lambda}(y) \int \cdots \int \left\{ \prod_{i=1}^n \overline{\lambda}(z_i) \right\}$$

$$\times \log \left\{ \frac{\overline{\lambda}(y) \prod_{i=1}^n \overline{\lambda}(z_i)}{\int \overline{\lambda}_{\overline{\mu}}(y) \prod_{i=1}^n \overline{\lambda}_{\overline{\mu}}(z_i)\pi(\mathrm{d}\overline{\mu})} \right\} \mathrm{d}z_1 \cdots \mathrm{d}z_n \mathrm{d}y$$

$$- \int \cdots \int \left\{ \prod_{i=1}^n \overline{\lambda}(z_i) \right\}$$

$$\times \log \left\{ \frac{\prod_{i=1}^n \overline{\lambda}(z_i)}{\prod_{i=1}^n \int \overline{\lambda}_{\overline{\mu}}(z_i)\pi(\mathrm{d}\overline{\mu})} \right\} \mathrm{d}z_1 \cdots \mathrm{d}z_n$$

$$= \int \cdots \int \left\{ \prod_{i=1}^n \overline{\lambda}(z_i) \right\} \int \overline{\lambda}(y) \log \frac{\overline{\lambda}(y)}{\overline{\lambda}_{\pi, z^n}(y)} \mathrm{d}y \mathrm{d}z_1 \cdots \mathrm{d}z_n, \quad (31)$$

where

$$\overline{\lambda}_{\pi, z^n}(y) = \frac{\int \prod_{i=1}^n \overline{\lambda}_{\overline{\mu}}(z_i)\overline{\lambda}_{\overline{\mu}}(y) \pi(\mathrm{d}\overline{\mu})}{\int \prod_{j=1}^n \overline{\lambda}_{\overline{\mu}}(z_j)\pi(\mathrm{d}\overline{\mu})}.$$

Thus, the difference (31) between $c(n+1)$ and $c(n)$ is the individual Kullback–Leibler risk of the Bayesian predictive density for predicting z_{n+1} using $z_1, \ldots, z_n$, where $z_1, \ldots, z_{n+1}$ are independently distributed according to probability density $\overline{\lambda}$. The power series of r in (28) converges for $0 \leq r < \infty$ if and only if the power series in (30) converges for $0 \leq r < \infty$.

E. Shrinkage Priors Dominating the Prior π_α and Their Admissibility

We propose shrinkage priors dominating $\pi_\alpha(\mathrm{d}\mu) = \mu^{\alpha-1}\mathrm{d}\mu$ when $|\alpha| > 1$. We prove admissibility of Bayes estimators and Bayesian predictive densities based on the shrinkage priors.

Although few studies on admissibility concerning infinite-dimensional models have been carried out, we show that Blyth's method with a convex loss is useful for our infinite-dimensional problem because the method works even when the support of the prior is a small subset of the whole parameter space. The key idea is to decompose the problem into two sub-problems: a one-dimensional problem with an improper prior, and an infinite-dimensional problem with a proper prior.

Theorem 4: Suppose that $\boldsymbol{z}_\tau = (N_\tau, z_1, \ldots, z_{N_\tau})$ is distributed according to the nonhomogeneous Poisson process $\mathscr{P}_\mathfrak{o}(\tau\lambda)$ ($\tau \geq 0$).

Assume that

$$f(r, s) := \mathrm{E}\left[\log \frac{p_{s,\lambda}(\boldsymbol{z}_r)}{p_{s,\pi}(\boldsymbol{z}_r)}\right]$$

exists and is differentiable with respect to $(r, s) \in \mathbb{R}^2_{>0}$ in a neighborhood of every $r = s > 0$ and that

$$\frac{\partial}{\partial s}\mathrm{E}\left[\log \frac{p_{s,\lambda}(\boldsymbol{z}_r)}{p_{s,\pi}(\boldsymbol{z}_r)}\right] = \mathrm{E}\left[\frac{\partial}{\partial s}\log \frac{p_{s,\lambda}(\boldsymbol{z}_r)}{p_{s,\pi}(\boldsymbol{z}_r)}\right].$$

1) When $|\alpha| - \gamma > 1$, the Bayesian predictive density $p_{\alpha,\gamma}(\boldsymbol{y} \mid \boldsymbol{x})$ based on $\pi_{\alpha,\gamma}(\mathrm{d}\mu)$ is dominated by the Bayesian predictive density $p_{\alpha,\tilde\gamma}(\boldsymbol{y} \mid \boldsymbol{x})$ based on $\pi_{\alpha,\tilde\gamma}(\mathrm{d}\mu)$, where $\tilde\gamma := |\alpha| - 1$.
2) When $|\alpha| - \gamma > 1$, the generalized Bayes estimator of λ based on $\pi_{\alpha,\gamma}(\mathrm{d}\mu)$ is dominated by the generalized Bayes estimator of λ based on $\pi_{\alpha,\tilde\gamma}(\mathrm{d}\mu)$, where $\tilde\gamma := |\alpha| - 1$.

The proof is given in Appendix A.

In particular, if $|\alpha| > 1$, the Bayesian predictive density based on $\pi_\alpha(\mathrm{d}\mu) = \pi_{\alpha,\gamma=0}(\mathrm{d}\mu) = \mu^{\alpha-1}\mathrm{d}\mu$ is dominated by the Bayesian predictive density based on $\pi_{\alpha,\tilde\gamma=|\alpha|-1}(\mathrm{d}\mu)$.

Lemma 3: Let X be a Poisson random variable with mean $\theta \geq 0$, and let c be an arbitrary positive constant. Then,

$$\theta\,\mathrm{E}_\theta\left[\log(X + 1 + c) - \log(X + 1)\right] \leq c - c\exp(-\theta) < c.$$

The proof of Lemma 3, which is used in [21], is given in Appendix A for self-containedness.

Next, we prove admissibility of Bayes estimators and Bayesian predictive densities based on our shrinkage priors.

Suppose that the parameter space Λ is a set of finite measures that are mutually absolutely continuous with respect to the Lebesgue measure on $U \subset \mathbb{R}^d$. We assume that if $\lambda = (w, \overline{\lambda}) \in \Lambda$, then $(w', \overline{\lambda}) \in \Lambda$ for all $w' > 0$. Thus, $\Lambda = \mathbb{R}^+ \times \overline{\Lambda}$, where $\mathbb{R}^+$ is the set of positive real numbers and $\overline{\Lambda}$ is a set of probability densities that are mutually absolutely continuous with respect to the Lebesgue measure.

Note that the support of a prior for the kernel mixture model only covers a small subset of the whole parameter space Λ if Λ is a large set as in ordinary settings for nonparametric inference.

Let $\mathscr{A}$ be the space of all finite measures on U such that the measure is mutually absolutely continuous with respect to the Lebesgue measure. For estimation of λ, we choose an intensity estimate from $\mathscr{A}$. Let $\mathscr{P}$ be the space of all probability measures on $\cup_{m=0}^\infty U^m$ such that the marginal probability $P(m)$ of $P \in \mathscr{P}$ is positive (i.e. not equal to 0) for every nonnegative integer m and the conditional probability $P(\cdot \mid m)$ on U^m has density with respect to the Lebesgue measure on $U^m \subset \mathbb{R}^m$ for every m. For prediction problem, we choose a probability density from $\mathscr{P}$.

We assume that $\int k(y, u)\overline{\mu}(\mathrm{d}u) \in \overline{\Lambda}$ with probability 1 if $\overline{\mu}$ is distributed according to $\mathscr{D}i(\alpha)$. This condition is naturally satisfied in ordinary settings for nonparametric intensity estimation.

Theorem 5:
1) The Bayesian predictive density based on $\pi_{\alpha,\gamma}(\mathrm{d}\mu) := p_\alpha^{\mathscr{D}i}(\mathrm{d}\overline{\mu})\,p_{|\alpha|-\gamma,\infty}^{\mathrm{Ga}}(w)\mathrm{d}w$ is admissible under the Kullback–Leibler loss if $|\alpha| - 1 \leq \gamma < |\alpha|$.
2) The generalized Bayes estimator $\lambda_{\alpha,\gamma,\boldsymbol{x}}$ based on $\pi_{\alpha,\gamma}(\mathrm{d}\mu) := p_\alpha^{\mathscr{D}i}(\mathrm{d}\overline{\mu})\,p_{|\alpha|-\gamma,\infty}^{\mathrm{Ga}}(w)\mathrm{d}w$ is admissible under the Kullback–Leibler loss if $|\alpha| - 1 \leq \gamma < |\alpha|$.

The proof is given in Appendix A.

In particular, if $|\alpha| > 1$, then the Bayesian predictive density based on $\pi_{\alpha,\gamma=|\alpha|-1}$ dominating that based on $\pi_\alpha = \pi_{\alpha,\gamma=0}$ is admissible.

III. NUMERICAL EVALUATION OF PREDICTIVE DENSITIES AND ESTIMATORS

In this section, we explore numerical methods by which to evaluate Bayesian predictive densities and Bayes estimates.

In the posterior (13) with respect to the prior $\pi_{\alpha,\beta,\gamma}(\mathrm{d}\mu) := p_\alpha^{\mathscr{D}i}(\mathrm{d}\overline{\mu})\,p_{|\alpha|-\gamma,\beta}^{\mathrm{Ga}}(w)\mathrm{d}w$, w and $\overline{\mu}$ are independently distributed. The posterior distribution of w is $p_{|\alpha|-\gamma+N,\beta/(s\beta+1)}^{\mathrm{Ga}}(w)\,\mathrm{d}w$. The posterior distribution of $\overline{\mu}$ is proportional to a Dirichlet process mixture

$$\int\left\{\prod_{l=1}^N \int k(x_l, u_l)\overline{\mu}(\mathrm{d}u_l)\right\}p_\alpha^{\mathscr{D}i}(\mathrm{d}\overline{\mu})$$
$$= \int \cdots \int \left\{\prod_{l=1}^N k(x_l, u_l)\right\}p_\alpha^{\mathrm{CR}}(\mathrm{d}u_1, \ldots, \mathrm{d}u_N), \quad (32)$$

where

$$p_\alpha^{\mathrm{CR}}(\mathrm{d}u_1, \ldots, \mathrm{d}u_N) = \int\left\{\prod_{l=1}^N \overline{\mu}(\mathrm{d}u_l)\right\}p_\alpha^{\mathscr{D}i}(\mathrm{d}\overline{\mu})$$

is the distribution of the Chinese restaurant process with measure α (e.g. [27] p. 193). If random variables $u_1, u_2, \ldots, u_N$ are distributed according to a Chinese restaurant process $P_\alpha^{\mathrm{CR}}(\mathrm{d}u_1, \ldots, \mathrm{d}u_N)$, then they are sequentially distributed according to

$$P_\alpha^{\mathrm{CR}}(\mathrm{d}u_1) = \frac{1}{|\alpha|}\alpha(\mathrm{d}u_1)$$

and

$$P_\alpha^{\mathrm{CR}}(\mathrm{d}u_{k+1} \mid u_1, \ldots, u_k)$$
$$= \frac{1}{|\alpha| + k}\left\{\alpha(\mathrm{d}u_{k+1}) + \sum_{i=1}^k \delta_{u_i}(\mathrm{d}u_{k+1})\right\}$$

($k = 1, 2, \ldots, N - 1$), see e.g. [27].

We can numerically evaluate quantities concerning the posterior density of λ such as the Bayes estimate of λ and Bayesian predictive density of $\boldsymbol{y}$ given $\boldsymbol{x}$ using MCMC based on the representation (32). Various MCMC methods for density estimation (e.g. [27]) based on nonparametric Bayes with Dirichlet priors can be applied to our problem.
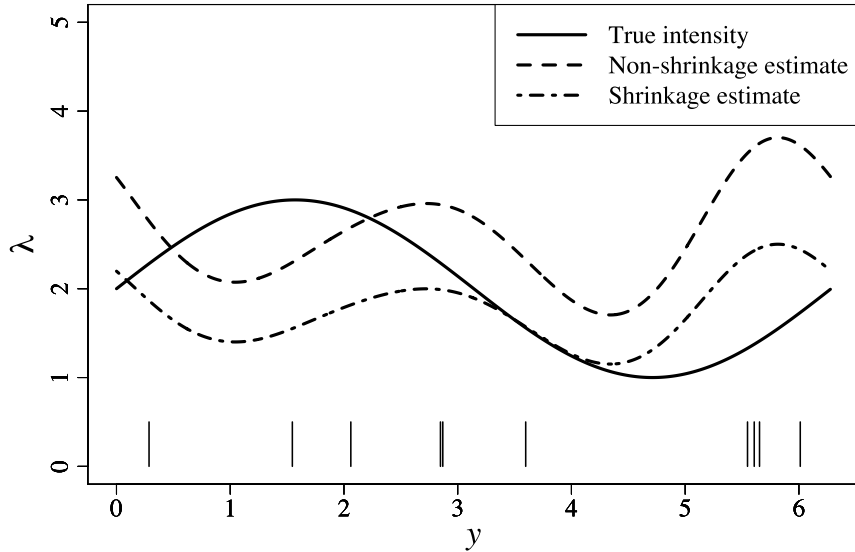
Fig. 1.   Bayesian estimates of λ based on non-shrinkage prior $\pi_{\alpha=1,\beta=\infty,\gamma=0}$ and shrinkage prior $\pi_{\alpha=1,\beta=\infty,\gamma=|\alpha|-1}$. The (sorted) observed points indicated by the vertical ticks at the bottom are located at $0.29, 1.55, 2.06, 2.85, 2.87, 3.60, 5.55, 5.61, 5.65,$ and 6.01.

From (18), the posterior mean $\overline{\lambda}$ with respect to $\pi_{\alpha,\beta,\gamma}$ and observation $\boldsymbol{x}$ is

$$\overline{\lambda}_{\alpha,\boldsymbol{x}} = \frac{\int k(y,\overline{\mu})\{\prod_{l=1}^{N} k(x_l,\overline{\mu})\} p_{\alpha}^{\mathscr{D}i}(\mathrm{d}\overline{\mu})}{\int \{\prod_{l=1}^{N} k(x_l,\overline{\mu})\} p_{\alpha}^{\mathscr{D}i}(\mathrm{d}\overline{\mu})}$$

$$= \left[\int \cdots \int k(y,u_{N+1})\{\prod_{l=1}^{N} k(x_l,u_l)\}\right.$$

$$\left. \times \, p_{\alpha}^{\mathrm{CR}}(\mathrm{d}u_1,\ldots,\mathrm{d}u_N,\mathrm{d}u_{N+1})\right]$$

$$\Big/ \left[\int \cdots \int \{\prod_{l=1}^{N} k(x_l,u_l)\} p_{\alpha}^{\mathrm{CR}}(\mathrm{d}u_1,\ldots,\mathrm{d}u_N)\right]$$

$$= \left[\int \cdots \int k(y,u_{N+1}) p_{\alpha}^{\mathrm{CR}}(\mathrm{d}u_{N+1} \mid u_1,\ldots,u_N)\right.$$

$$\left. \times \, \{\prod_{l=1}^{N} k(x_l,u_l)\} p_{\alpha}^{\mathrm{CR}}(\mathrm{d}u_1,\ldots,\mathrm{d}u_N)\right]$$

$$\Big/ \left[\int \cdots \int \{\prod_{l=1}^{N} k(x_l,u_l)\} p_{\alpha}^{\mathrm{CR}}(\mathrm{d}u_1,\ldots,\mathrm{d}u_N)\right], \quad (33)$$

which does not depend on β or γ. Thus, the posterior mean of $\overline{\lambda}$ with respect to the prior $\pi_{\alpha,\beta,\gamma}$ is $\overline{\lambda}_{\alpha,\boldsymbol{x}}$, which does not depend on β or γ.

We can obtain the Bayes estimates of λ based on $\pi_{\alpha,\beta,\gamma}$ and $\pi_{\alpha,\gamma} := \pi_{\alpha,\beta=\infty,\gamma}$ using (15), (16), (19), (20), and (33).

Example: We consider intensity functions on a unit circle $[0, 2\pi)$ for simplicity. Let the true intensity function be

$$\lambda(u) = \sin(u) + 2.$$

We obtain Bayes estimates of λ based on priors $\pi_{\alpha=1,\beta=\infty,\gamma=0}$ and $\pi_{\alpha=1,\beta=\infty,\gamma=|\alpha|-1}$. Then, $|\alpha| := \int_0^{2\pi} \alpha(u)\mathrm{d}u = 2\pi > 1$. We adopt the von Mises kernel

$$k_{\kappa}(x;\mu) = \frac{\exp\{\kappa \cos(x-\mu)\}}{2\pi I_0(\kappa)}$$

with $\kappa = 5$. The observation time length s is set to be 1.

In Figure 1, Bayesian estimates of λ based on non-shrinkage prior $\pi_{\alpha=1,\beta=\infty,\gamma=0}$ and shrinkage prior $\pi_{\alpha=1,\beta=\infty,\gamma=|\alpha|-1}$ with respect to observation $(0.29, 1.55, 2.06, 2.85, 2.87, 3.60, 5.55, 5.61, 5.65, 6.01)$ are shown. The Bayesian estimates based on the non-shrinkage prior and the shrinkage prior are $\lambda_{\alpha=1,\gamma=0,\boldsymbol{x}} = (N+2\pi)\overline{\lambda}_{\alpha,\boldsymbol{x}}$, and $\lambda_{\alpha=1,\gamma=|\alpha|-1,\boldsymbol{x}} = (N+1)\overline{\lambda}_{\alpha,\boldsymbol{x}}$, respectively, where $\overline{\lambda}_{\alpha,\boldsymbol{x}}$ is numerically evaluated by MCMC.

From Theorem 4, the estimator based on the shrinkage prior dominates that based on the non-shrinkage prior. Specifically, the Kullback–Leibler risk of the estimator based on the shrinkage prior is smaller than that of the estimator based on the non-shrinkage prior for all λ. This example illustrates how shrinkage priors improve Bayes estimators based on non-shrinkage priors.

IV. DISCUSSION AND CONCLUSION

A class of improper shrinkage priors for nonparametric Bayesian inference of the nonhomogeneous Poisson processes with kernel mixture is investigated. The Bayesian predictive densities and the Bayes estimators based on the priors are admissible under the Kullback–Leibler loss. The class of improper priors could be useful as objective priors for nonhomogeneous Poisson models.

The information theoretic properties of the Kullback–Leibler loss play essential roles in our theory. Our results can be easily generalized to various methods for intensity estimation including models based on kernels with parameters.

Although few studies on admissibility of Bayesian prediction and inference based on improper priors for infinite-dimensional models have been carried out, our approach could be useful for various infinite-dimensional problems.

One important direction of future research is the generalization of the present results for nonhomogeneous Poisson process models to those for other stochastic process models such as nonhomogeneous negative binomial process models.

For finite dimensional models, [15], [16] extended the theory for finite dimensional Poisson models to finite dimensional negative binomial models and negative multinomial models. These generalizations require techniques not used in the theory for the Poisson models.

Hazard rate models closely related to nonhomogeneous Poisson models are widely used in applications and kernel mixtures with general random measures have been investigated ([9], [17]). The use of various random measures other than the gamma random measure has been studied. The relation between theories on hazard rate models with general random measures and the present study is also an important topic for future study.

Nonhomogeneous Poisson process models have a variety of applications. For example, they have recently been used in neural information processing ([29], [30]). The development of data analyzing methods for such applications based on the theory presented here is also a future challenge. In applications such as survival analysis, weighted gamma processes are often used as priors. By extending the theorems in [23] to those for infinite-dimensional models, we could generalize our results to improve the limits of weighted gamma process priors.

Furthermore, estimation and prediction with infinite dimensional statistical models based on Brownian motions could be studied in line with the approach adopted in the present paper.

APPENDIX A
PROOFS OF THEOREMS AND LEMMAS

Proof of Lemma 2: Since

$$
\begin{aligned}
f(r, s) &:= \mathrm{E}\big[h(\boldsymbol{z}_r, s)\big] \\
&= \sum_{n=0}^{\infty} \Bigg[\frac{(rw)^n}{n!} \exp(-rw) \\
&\quad \times \int \cdots \int \Big\{\prod_{i=1}^{n} \overline{\lambda}(z_i)\Big\} h_n(z_1, \ldots, z_n, s) \mathrm{d}z_1 \cdots \mathrm{d}z_n\Bigg]
\end{aligned}
$$

and $\frac{\partial}{\partial s}\mathrm{E}[h(\boldsymbol{z}_r, s)] = \mathrm{E}[\frac{\partial}{\partial s}h(\boldsymbol{z}_r, s)]$, we have

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}\tau}\mathrm{E}\big[h(\boldsymbol{z}_\tau, \tau)\big] &= \frac{\mathrm{d}}{\mathrm{d}\tau} f(\tau, \tau) \\
&= \frac{\partial}{\partial r} f(r, \tau)\Big|_{r=\tau} + \frac{\partial}{\partial s} f(\tau, s)\Big|_{s=\tau} \\
&= w \sum_{n=1}^{\infty}\Bigg[\frac{(\tau w)^{n-1}}{(n-1)!}\exp(-\tau w) \\
&\quad \times \int \cdots \int \Big\{\prod_{i=1}^{n}\overline{\lambda}(z_i)\Big\}h_n(z_1, \ldots, z_n, \tau)\mathrm{d}z_1 \cdots \mathrm{d}z_n\Bigg] \\
&\quad - w\sum_{n=0}^{\infty}\Bigg[\frac{(\tau w)^n}{n!}\exp(-\tau w) \\
&\quad \times \int \cdots \int \Big\{\prod_{i=1}^{n}\overline{\lambda}(z_i)\Big\}h_n(z_1, \ldots, z_n, \tau)\mathrm{d}z_1 \cdots \mathrm{d}z_n\Bigg] \\
&\quad + \frac{\partial}{\partial r}\mathrm{E}\big[h(\boldsymbol{z}_\tau, r)\big]\Big|_{r=\tau} \\
&= \int_U \lambda(y)\mathrm{E}_\lambda\big[h(\boldsymbol{z}_\tau + \delta_y, \tau) - h(\boldsymbol{z}_\tau, \tau)\big]\mathrm{d}y
\end{aligned}
$$

$$
+ \mathrm{E}\left[\frac{\partial h}{\partial r}(\boldsymbol{z}_\tau, r)\Big|_{r=\tau}\right]. \qquad \square
$$

Proof of Theorem 3: Let

$$
\begin{aligned}
h(\boldsymbol{z}_r, s) &:= \log \frac{p_{s,\lambda}(\boldsymbol{z}_r)}{p_{s,\pi}(\boldsymbol{z}_r)} \\
&= \log \frac{\Big\{\prod_{i=1}^{N_r} s\lambda(z_{r,i})\Big\}\frac{1}{N_r!}\exp(-sw)}{\int\Big\{\prod_{i=1}^{N_r} s\lambda'(z_{r,i})\Big\}\frac{1}{N_r!}\exp(-sw')\pi(\mathrm{d}\lambda')} \\
&= -\log \int\Big\{\prod_{i=1}^{N_r}\frac{\lambda'(z_{r,i})}{\lambda(z_{r,i})}\Big\}\exp(-sw')\pi(\mathrm{d}\lambda') - sw.
\end{aligned}
$$

From Lemma 2, the integrand of (24) is represented by

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}\tau}\mathrm{E}_\lambda\left[\log \frac{p_{\tau,\lambda}(\boldsymbol{z}_\tau)}{p_{\tau,\pi}(\boldsymbol{z}_\tau)}\right] &= \frac{\mathrm{d}}{\mathrm{d}\tau}\mathrm{E}_\lambda\big[h(\boldsymbol{z}_\tau, \tau)\big] \\
&= -\frac{\mathrm{d}}{\mathrm{d}\tau}\mathrm{E}_\lambda\left[\log\int\Big\{\prod_{i=1}^{N_\tau}\frac{\lambda'(z_{\tau,i})}{\lambda(z_{\tau,i})}\Big\}\exp(-\tau w')\pi(\mathrm{d}\lambda')\right] - w \\
&= -\int_U \lambda(y)\,\mathrm{E}_\lambda\Bigg[\log\int\Big\{\prod_{i=1}^{N_\tau}\frac{\lambda'(z_{\tau,i})}{\lambda(z_{\tau,i})}\Big\}\frac{\lambda'(y)}{\lambda(y)}\exp(-\tau w')\pi(\mathrm{d}\lambda') \\
&\quad - \log\int\Big\{\prod_{i=1}^{N_\tau}\frac{\lambda'(z_{\tau,i})}{\lambda(z_{\tau,i})}\Big\}\exp(-\tau w')\pi(\mathrm{d}\lambda')\Bigg]\mathrm{d}y \\
&\quad - \mathrm{E}_\lambda\left[\frac{-\int w'\Big\{\prod_{i=1}^{N_\tau}\lambda'(z_{\tau,i})\Big\}\exp(-\tau w')\pi(\mathrm{d}\lambda')}{\int\Big\{\prod_{i=1}^{N_\tau}\lambda'(z_{\tau,i})\Big\}\exp(-\tau w')\pi(\mathrm{d}\lambda')}\right] - w \\
&= -\int_U \lambda(y)\,\log\frac{\lambda_{\pi,\boldsymbol{z}_\tau,\tau}(y)}{\lambda(y)}\mathrm{d}y + w_{\pi,\boldsymbol{z}_\tau,\tau} - w \\
&= w_{\pi,\boldsymbol{z}_\tau,\tau} - w - w\log\frac{w_{\pi,\boldsymbol{z}_\tau,\tau}}{w} \\
&\quad + w\int_U \overline{\lambda}(y)\,\log\frac{\overline{\lambda}(y)}{\overline{\lambda}_{\pi,\boldsymbol{z}_\tau,\tau}(y)}\mathrm{d}y. \qquad (34)
\end{aligned}
$$

From (24) and (34), we have the desired result. $\qquad \square$

Proof of Theorem 4: We prove the first part (1) for Bayesian predictive densities. The second part (2) for Bayes estimators can be proved in a similar manner.

Since (18), the posterior mean of $\overline{\lambda}$ based on $\pi_{\alpha,\gamma}$ coincides with that based on $\pi_{\alpha,\widetilde{\gamma}}$. Therefore, from Theorem 3, we have

$$
\begin{aligned}
\mathrm{E}_\lambda\Big[D(p(\boldsymbol{y}\mid\lambda), p_{\alpha,\gamma}(\boldsymbol{y}\mid\boldsymbol{x}))\Big] &- \mathrm{E}_\lambda\Big[D(p(\boldsymbol{y}\mid\lambda), p_{\alpha,\widetilde{\gamma}}(\boldsymbol{y}\mid\boldsymbol{x}))\Big] \\
&= \int_s^t w\,\mathrm{E}_\lambda\left[\frac{w_{\alpha,\gamma,\tau}}{w} - 1 - \log\frac{w_{\alpha,\gamma,\tau}}{w}\right]\mathrm{d}\tau \\
&\quad - \int_s^t w\,\mathrm{E}_\lambda\left[\frac{w_{\alpha,\widetilde{\gamma},\tau}}{w} - 1 - \log\frac{w_{\alpha,\widetilde{\gamma},\tau}}{w}\right]\mathrm{d}\tau, \qquad (35)
\end{aligned}
$$

where $w_{\alpha,\gamma,\tau} := w_{\alpha,\gamma,\boldsymbol{z}(\tau),\tau}$. From (19), the posterior means of w with respect to priors $\pi_\alpha = \pi_{\alpha,\gamma=0}$ and $\pi_{\alpha,\widetilde{\gamma}=|\alpha|-1}$ are $w_{\alpha,\tau} = (N_\tau + |\alpha|)/\tau$ and $w_{\alpha,\widetilde{\gamma}=|\alpha|-1,\tau} = (N_\tau + 1)/\tau$, respectively, when we observe N_τ.

Thus, the integrand of (35) with $\gamma = 0$ and $\tilde{\gamma} = |\alpha| - 1$ is

$$
w\mathrm{E}_\lambda\left[\frac{w_{\alpha,\tau}}{w} - 1 - \log\frac{w_{\alpha,\tau}}{w}\right]
$$

$$
-w\mathrm{E}_\lambda\left[\frac{w_{\alpha,\tilde{\gamma}=|\alpha|-1,\tau}}{w} - 1 - \log\frac{w_{\alpha,\tilde{\gamma}=|\alpha|-1,\tau}}{w}\right]
$$

$$
= \frac{|\alpha|}{\tau} - w\,\mathrm{E}_\lambda\left[\log\frac{N_\tau + |\alpha|}{\tau}\right] - \frac{1}{\tau} + w\,\mathrm{E}_\lambda\left[\log\frac{N_\tau + 1}{\tau}\right]
$$

$$
= \frac{1}{\tau}\Big\{|\alpha| - 1 - \tau w\,\mathrm{E}_\lambda\Big[\log(N_\tau + |\alpha|)\Big]
$$

$$
+ \tau w\,\mathrm{E}_\lambda\Big[\log(N_\tau + 1)\Big]\Big\}. \tag{36}
$$

From Lemma 3 below, (36) is positive. $\square$

Proof of Lemma 3:

$$
c - \sum_{n=0}^\infty \exp(-\theta)\frac{\theta^{n+1}}{n!}\Big\{\log(n+1+c) - \log(n+1)\Big\}
$$

$$
\geq c - \sum_{n=0}^\infty \exp(-\theta)\frac{\theta^{n+1}}{n!}\frac{c}{n+1}
$$

$$
= c\exp(-\theta)\Big\{\exp\theta - \sum_{n=0}^\infty \frac{\theta^{n+1}}{(n+1)!}\Big\}
$$

$$
= c\exp(-\theta) > 0.
$$

$\square$

Proof of Theorem 5: We prove the first part (1) for Bayesian predictive densities using Blyth's method with a convex loss. Blyth's method is widely used to prove admissibility usually for problems with a finite-dimensional parameter space. The method with a convex loss also works for our infinite-dimensional problem. The second part (2) for Bayes estimators can be proved in a similar manner.

By the assumption, the distributions $p_\lambda(\boldsymbol{x})$ ($\lambda \in \Lambda$) are absolutely continuous with respect to each other. The action space $\mathscr{P}$ is convex because $aq(\boldsymbol{y}) + (1-a)q'(\boldsymbol{y}) \in \mathscr{P}$ if both $q(\boldsymbol{y})$ and $q'(\boldsymbol{y})$ belong to $\mathscr{P}$ and $0 \leq a \leq 1$. The Kullback–Leibler loss

$$
L(\lambda, q) := \mathrm{E}_\lambda\left[\log\frac{p_\lambda(M, y_1, \ldots, y_M)}{q(M, y_1, \ldots, y_M)}\right]
$$

is a strictly convex function from $q \in \mathscr{P}$ to $[0, \infty]$ for every fixed λ.

We use a monotonically increasing sequence of proper priors defined by $\pi_{\alpha,\gamma}^{[l]}(\mathrm{d}\mu) = \pi_{\alpha,\gamma}(\mathrm{d}\mu)\frac{1}{2}h_l^2(w)$ ($l = 1, 2, \ldots$), where $\pi_{\alpha,\gamma}(\mathrm{d}\mu) = p_\alpha^{\mathscr{D}}(\mathrm{d}\overline{\mu})\,p_{|\alpha|-\gamma,\infty}^{\mathrm{Ga}}(w)\mathrm{d}w = p_\alpha^{\mathscr{D}}(\mathrm{d}\overline{\mu})\,w^{|\alpha|-\gamma-1}\mathrm{d}w$ and

$$
h_l(w) = \begin{cases} 1 & \text{if } 0 \leq w \leq 1 \\ 1 - \dfrac{\log w}{\log l} & \text{if } 1 < w \leq l \\ 0 & \text{if } l < w. \end{cases}
$$

Function sequences of this kind were introduced by [6] and are used to prove admissibility of linear estimators ([14]) and admissibility of predictive densities ([21]) for finite-dimensional independent Poisson models. Then, $\pi_{\alpha,\gamma}^{[l]} \ll \pi_{\alpha,\gamma}$ for all $l = 1, 2, \ldots$. Let $C := \{(w, \overline{\lambda}) \mid 0 < w < 1, \overline{\lambda} \in \overline{\Lambda}\}$. Then, $\pi(C) > 0$ and $\mathrm{d}\pi^{[l]}/\mathrm{d}\pi = 1/2$ if $\mu \in C$.

In order to prove admissibility of $p_{\alpha,\gamma}(\boldsymbol{y} \mid \boldsymbol{x})$ based on $\pi_{\alpha,\gamma}$ with $0 < |\alpha| - \gamma \leq 1$, it suffices to show

$$
\lim_{l\to\infty}\int \pi_{\alpha,\gamma}^{[l]}(\mathrm{d}\lambda)\Big\{\mathrm{E}_\lambda\big[D(p_\lambda(\boldsymbol{y}), p_{\alpha,\gamma}(\boldsymbol{y} \mid \boldsymbol{x}))\big]
$$

$$
- \mathrm{E}_\lambda\big[D(p_\lambda(\boldsymbol{y}), p_{\alpha,\gamma}^{[l]}(\boldsymbol{y} \mid \boldsymbol{x}))\big]\Big\} = 0, \tag{37}
$$

where $p^{[l]}$ is the Bayesian predictive density based on $\pi_{\alpha,\gamma}^{[l]}$. The proof of the sufficiency of (37) for admissibility is given in Appendix B.

We set $c = \gamma - |\alpha| + 1$ ($0 \leq c < 1$), and $g_l(w) = (1/2)h_l^2(w)$. From Theorem 3 and (35), we have the expression

$$
\int \pi_{\alpha,\beta}^{[l]}(\mathrm{d}\lambda)\Big\{\mathrm{E}_\lambda\big[D(p(\boldsymbol{y} \mid \lambda), p_{\alpha,\beta}(\boldsymbol{y} \mid \boldsymbol{x}))\big]
$$

$$
- \mathrm{E}_\lambda\big[D(p(\boldsymbol{y} \mid \lambda), p_{\alpha,\beta}^{[l]}(\boldsymbol{y} \mid \boldsymbol{x}))\big]\Big\}
$$

$$
= \int_s^{s+t}\left[\int_0^\infty g_l(w)w^{-c}\Big\{\sum_{n=0}^\infty \exp(-\tau w)\frac{(\tau w)^n}{n!}\right.
$$

$$
\left.\times\left(\frac{n+1-c}{\tau} - \widehat{w}_{l,\tau}(n) - w\log\frac{n+1-c}{\tau\widehat{w}_{l,\tau}(n)}\right)\Big\}\mathrm{d}w\right]\mathrm{d}\tau, \tag{38}
$$

where

$$
\widehat{w}_{l,\tau}(n) = \frac{\int_0^\infty w\exp(-\tau w)\frac{(\tau w)^n}{n!}w^{-c}g_l(w)\mathrm{d}w}{\int_0^\infty \exp(-\tau w)\frac{(\tau w)^n}{n!}w^{-c}g_l(w)\mathrm{d}w}
$$

$$
= \frac{\int_0^\infty \exp(-\tau w)(\tau w)^{n+1-c}g_l(w)\mathrm{d}w}{\tau\int_0^\infty \exp(-\tau w)(\tau w)^{n-c}g_l(w)\mathrm{d}w}
$$

$$
= \frac{n+1-c}{\tau} + \frac{\int_0^\infty \exp(-\tau w)(\tau w)^{n+1-c}g_l'(w)\mathrm{d}w}{\tau^2\int_0^\infty \exp(-\tau w)(\tau w)^{n-c}g_l(w)\mathrm{d}w}, \tag{39}
$$

because $\overline{\lambda}_{\alpha,\gamma} = \overline{\lambda}_{\alpha,\gamma}^{[l]}$ and it does not depend on γ.

The integral (38) coincides with (22) in [21] and converges to 0 as l goes to infinity as shown in the following.

We show that (38) converges to 0 in the same manner as the evaluation of (22) in [21]. We include the proof here for self-containedness.

We have

$$
\int_0^\infty g_l(w)w^{-c}\Big\{\sum_{z=0}^\infty \exp(-\tau w)\frac{(\tau w)^z}{z!}
$$

$$
\times\left(\frac{z+1-c}{\tau} - \widehat{w}_{l,\tau} - w\log\frac{z+1-c}{\tau\widehat{w}_{l,\tau}}\right)\Big\}\mathrm{d}w
$$

$$
\leq \int_0^\infty g_l(w)w^{-c}\left[\sum_{z=0}^\infty \exp(-\tau w)\frac{(\tau w)^z}{z!}\right.
$$

$$
\left.\times\Big\{\frac{z+1-c}{\tau} - \widehat{w}_{l,\tau} + w\frac{\tau\widehat{w}_{l,\tau} - (z+1-c)}{z+1-c}\Big\}\right]\mathrm{d}w
$$

$$
= \sum_{z=0}^\infty\left[\frac{\tau^{c-1}}{z!}\Big\{\int_0^\infty \exp(-\tau w)g_l(w)\frac{(\tau w)^{z+1-c}}{z+1-c}\mathrm{d}w\right.
$$

$$
\left.- \int_0^\infty \exp(-\tau w)g_l(w)(\tau w)^{z-c}\mathrm{d}w\Big\}\right.
$$

$$\times \left\{ \tau \widehat{w}_{l,\tau} - (z+1-c) \right\} \Bigg]$$

$$= \sum_{z=0}^{\infty} \Bigg[\frac{\tau^{c-2}}{z!} \int_0^{\infty} \exp(-\tau w) g_l'(w) \frac{(\tau w)^{z+1-c}}{z+1-c} dw$$

$$\times \left\{ \tau \widehat{w}_{l,\tau} - (z+1-c) \right\} \Bigg]. \tag{40}$$

Using (39) and the inequality $(z+1)/(z+1-c) \le 1/(1-c)$, where $0 \le c < 1$, we have

$$\int_0^{\infty} g_l(w) w^{-c} \Bigg\{ \sum_{z=0}^{\infty} \exp(-\tau w) \frac{(\tau w)^z}{z!}$$

$$\times \left(\frac{z+1-c}{\tau} - \widehat{w}_{l,\tau} - w \log \frac{z+1-c}{\tau \widehat{w}_{l,\tau}} \right) \Bigg\} dw$$

$$\le \frac{\tau^{c-3}}{1-c} \sum_{z=0}^{\infty} \frac{1}{(z+1)!} \frac{\left\{ \int_0^{\infty} \exp(-\tau w)(\tau w)^{z+1-c} g_l'(w) dw \right\}^2}{\int \exp(-\tau w)(\tau w)^{z-c} g_l(w) dw}$$

$$= \frac{\tau^{c-3}}{1-c} \Bigg[\sum_{z=0}^{\infty} \frac{2}{(z+1)!}$$

$$\times \frac{\left\{ \int_0^{\infty} \exp(-\tau w)(\tau w)^{z-c} (\tau w h_l'(w)) h_l(w) dw \right\}^2}{\int \exp(-\tau \overline{\mu})(\tau w)^{z-c} h_l^2(w) dw} \Bigg]$$

$$\le \frac{\tau^{c-3}}{1-c} \Bigg[\sum_{z=0}^{\infty} \frac{2}{(z+1)!}$$

$$\times \left\{ \int_0^{\infty} \exp(-\tau w)(\tau w)^{z-c} (\tau w h_l'(w))^2 dw \right\}$$

$$\times \left\{ \int_0^{\infty} \exp(-\tau w)(\tau w)^{z-c} h_l^2(w) dw \right\}$$

$$\Bigg/ \left\{ \int \exp(-\tau w)(\tau w)^{z-c} h_l^2(w) dw \right\} \Bigg]$$

$$= \frac{\tau^{c-3}}{1-c} \sum_{z=0}^{\infty} \frac{2}{(z+1)!} \int_0^{\infty} \exp(-\tau w)(\tau w)^{z+2-c} (h_l'(w))^2 dw$$

$$= \frac{2\tau^{c-3}}{1-c} \int_0^{\infty} \left\{ 1 - \exp(-\tau w) \right\} (\tau w)^{1-c} (h_l'(w))^2 dw$$

$$\le \frac{2}{(1-c)\tau^2} \int_0^{\infty} w^{1-c} (h_l'(w))^2 dw \tag{41}$$

The derivative of $h_l(w)$ is

$$h_l'(w) = \begin{cases} 0 & \text{if } 0 < w < 1 \\ -\dfrac{1}{w \log l} & \text{if } 1 < w < l \\ 0 & \text{if } l < w. \end{cases} \tag{42}$$

From (38), (41) and (42), we have

$$\int \pi_{\alpha,\beta}^{[l]}(d\lambda) E_\lambda \Big[D(p_\lambda(\boldsymbol{y}), p_{\alpha,\beta}(\boldsymbol{y} \mid \boldsymbol{x}))$$

$$- D(p_\lambda(y), p_{\alpha,\beta}^{[l]}(\boldsymbol{y} \mid \boldsymbol{x})) \Big]$$

$$\le \int_s^{s+t} \left\{ \frac{2}{(1-c)\tau^2} \int_0^{\infty} w^{1-c} (h_l'(w))^2 dw \right\} d\tau$$

$$= \int_s^{s+t} \left\{ \frac{2}{(1-c)\tau^2} \int_1^{l} \frac{1}{w^{1+c}(\log l)^2} dw \right\} d\tau$$

$$\le \int_s^{s+t} \frac{2}{(1-c)\tau^2} \frac{1}{\log l} d\tau = \frac{2}{(1-c)\log l} \left(\frac{1}{s} - \frac{1}{s+t} \right)$$

$$\to 0 \quad \text{as} \quad l \to \infty. \qquad \square$$

APPENDIX B
SUFFICIENCY OF (37) FOR ADMISSIBILITY

We show that (37) is sufficient for admissibility. The proof for our infinite-dimensional problem parallels the proof for finite-dimensional problems with a convex loss (see [28] Chapter 3).

We denote a map from $\boldsymbol{x}$ to a probability distribution of $\boldsymbol{y}$ by q, and denote the probability density that is the image of $\boldsymbol{x}$ under q by $q_{\boldsymbol{x}}$. The map q corresponds to a predictive density.

Let $L(\lambda, q_{\boldsymbol{x}})$ be a loss function, which is strictly convex with respect to $q_{\boldsymbol{x}}$, and let $R(\lambda, q) := E_\lambda[L(\lambda, q_{\boldsymbol{x}})]$, where $\boldsymbol{x}$ is distributed according to a probability density p_λ. Assume that a predictive density q satisfying

$$\lim_{l \to \infty} \left\{ R(\pi^{[l]}, q) - R(\pi^{[l]}, q^{[l]}) \right\} = 0, \tag{43}$$

where $q^{[l]}$ is the Bayesian predictive density based on $\pi^{[l]}$ and $R(\pi^{[l]}, q^{[l]}) := \int R(\lambda, q^{[l]}) \pi^{[l]}(d\lambda)$ is the Bayes risk of $q^{[l]}$ with respect to the prior $\pi^{[l]}$, is inadmissible. Assume that $d\pi^{[l]}/d\pi \ge a > 0$ if $\mu \in C$ for a subset C of the parameter space. We assume that all probability measures in $\{p_\lambda : \lambda \in \Lambda\}$ are mutually absolutely continuous with respect to each other.

In our problem, (43) corresponds to (37), L is the Kullback–Leibler loss, $\pi^{[l]} = \pi_{\alpha,\gamma}^{[l]}$, and $d\pi^{[l]}/d\pi = 1/2$ if $\mu \in C := \{ (w, \overline{\lambda}) \mid 0 < w < 1, \overline{\lambda} \in \overline{\Lambda} \}$.

We prove admissibility of the predictive density q by contradiction. Since q is inadmissible, there exists another predictive density q' such that, for all $\lambda \in \Lambda$, $R(\lambda, q') \le R(\lambda, q)$, and for at least one parameter value $\lambda_0 \in \Lambda$, $R(\lambda_0, q') < R(\lambda_0, q)$. Since $L(\lambda, q_{\boldsymbol{x}})$ is strictly convex with respect to $q_{\boldsymbol{x}}$ for every λ,

$$L(\lambda, q_{\boldsymbol{x}}'') \le \frac{1}{2} \left\{ L(\lambda, q_{\boldsymbol{x}}') + L(\lambda, q_{\boldsymbol{x}}) \right\},$$

where $q'' := (q + q')/2$, for every λ and $\boldsymbol{x}$. For every $\boldsymbol{x}$ such that $q_{\boldsymbol{x}} \ne q_{\boldsymbol{x}}'$, we have

$$L(\lambda, q_{\boldsymbol{x}}'') < \frac{1}{2} \left\{ L(\lambda, q_{\boldsymbol{x}}') + L(\lambda, q_{\boldsymbol{x}}) \right\}.$$

Here, $P_{\lambda_0}(q_{\boldsymbol{x}}' \ne q_{\boldsymbol{x}}) > 0$ because otherwise $R(\lambda_0, q') < R(\lambda_0, q)$ does not hold. Thus, $P_\lambda(q_{\boldsymbol{x}}' \ne q_{\boldsymbol{x}}) > 0$ for every λ because all probability measures p_λ ($\lambda \in \Lambda$) are mutually absolutely continuous with respect to each other. Thus, for every λ, $R(\lambda, q'') < R(\lambda, q)$. Therefore, for all $l = 1, 2, 3, \dots$,

$$R(\pi^{[l]}, q) - R(\pi^{[l]}, q^{[l]}) \ge R(\pi^{[l]}, q) - R(\pi^{[l]}, q'')$$

$$\ge \int_C \left\{ R(\lambda, q) - R(\lambda, q'') \right\} \pi^{[l]}(d\lambda)$$

$$\ge a \int_C \left\{ R(\lambda, q) - R(\lambda, q'') \right\} \pi(d\lambda) > 0,$$

where the second inequality is because C is a subset of the parameter space Λ. This contradicts (43).

References

[1] J. Aitchison and I. R. Dunsmore, *Statistical Prediction Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1975.

[2] A. R. Barron and T. M. Cover, "Minimum complexity density estimation," *IEEE Trans. Inf. Theory*, vol. 37, no. 4, pp. 1034–1054, Jul. 1991.

[3] A. R. Barron, "Information-theoretic characterization of Bayes performance and the choice of priors in parametric and nonparametric problems (with discussion)," in *Bayesian Statistics*, J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, Eds. London, U.K.: Oxford Univ. Press, 1999, pp. 27–52.

[4] R. Bartoszynski, B. W. Brown, C. M. McBride, and J. R. Thompson, "Some nonparametric techniques for estimating the intensity function of a cancer related nonstationary Poisson process," *Ann. Statist.*, vol. 9, no. 5, pp. 1050–1060, Sep. 1981.

[5] J. O. Berger, *Statistical Decision Theory and Bayesian Analysis*. New York, NY, USA: Springer, 1985.

[6] L. D. Brown and J. T. Hwang, "A unified admissibility proof," in *Statistical Decision Theory and Related Topics III*, S. S. Gupta and J. O. Berger, Eds. New York, NY, USA: Academic, 1982, pp. 205–230.

[7] L. D. Brown, E. I. George, and X. Xu, "Admissible predictive density estimation," *Ann. Statist.*, vol. 36, no. 3, pp. 1156–1170, Jun. 2008.

[8] D. J. Daley and D. Vere-Jones, *An Introduction to the Theory of Point Processes*, vol. 1, 2nd ed. New York, NY, USA: Springer, 2003.

[9] P. De Blasi, G. Peccati, and I. Prünster, "Asymptotics for posterior hazards," *Ann. Statist.*, vol. 37, no. 4, pp. 1906–1945, Aug. 2009.

[10] D. Fourdrinier, W. E. Strawderman, and M. T. Wells, *Shrinkage Estimation*. Cham, Switzerland: Springer, 2018.

[11] S. Geisser, *Predictive Inference: An Introduction*. New York, NY, USA: Chapman & Hall, 1993.

[12] E. I. George, F. Liang, and X. Xu, "Improved minimax predictive densities under Kullback–Leibler loss," *Ann. Statist.*, vol. 34, no. 1, pp. 78–91, Feb. 2006.

[13] J. K. Ghosh and R. V. Ramamoorthi, *Bayesian Nonparametrics*. New York, NY, USA: Springer, 2003.

[14] M. Ghosh and M.-C. Yang, "Simultaneous estimation of Poisson means under entropy loss," *Ann. Statist.*, vol. 16, no. 1, pp. 278–291, Mar. 1988.

[15] Y. Hamura and T. Kubokawa, "Bayesian predictive distribution for a negative binomial model," *Math. Methods Statist.*, vol. 28, no. 1, pp. 1–17, Jan. 2019.

[16] Y. Hamura and T. Kubokawa, "Bayesian shrinkage estimation of negative multinomial parameter vectors," *J. Multivariate Anal.*, vol. 179, Sep. 2020, Art. no. 104653.

[17] L. F. James, "Bayesian Poisson process partition calculus with an application to Bayesian Lévy moving averages," *Ann. Statist.*, vol. 33, no. 4, pp. 1771–1799, Aug. 2005.

[18] E. D. Kolaczyk, "Wavelet shrinkage estimation of certain Poisson intensity signals using corrected thresholds," *Statistica Sinica*, vol. 9, pp. 119–135, Jan. 1999.

[19] F. Komaki, "On asymptotic properties of predictive distributions," *Biometrika*, vol. 83, no. 2, pp. 299–313, Jun. 1996.

[20] F. Komaki, "A shrinkage predictive distribution for multivariate Normal observables," *Biometrika*, vol. 88, no. 3, pp. 859–864, Oct. 2001.

[21] F. Komaki, "Simultaneous prediction of independent Poisson observables," *Ann. Statist.*, vol. 32, no. 4, pp. 1744–1769, Aug. 2004.

[22] F. Komaki, "A class of proper priors for Bayesian simultaneous prediction of independent Poisson observables," *J. Multivariate Anal.*, vol. 97, no. 8, pp. 1815–1828, Sep. 2006.

[23] F. Komaki, "Simultaneous prediction for independent Poisson processes with different durations," *J. Multivariate Anal.*, vol. 141, pp. 35–48, Oct. 2015.

[24] A. Y. Lo, "Bayesian nonparametric statistical inference for Poisson point processes," *Zeitschrift Für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, vol. 59, no. 1, pp. 55–66, 1982.

[25] A. Y. Lo, "On a class of Bayesian nonparametric estimates: I. Density estimates," *Ann. Statist.*, vol. 12, no. 1, pp. 351–357, Mar. 1984.

[26] A. Y. Lo and C.-S. Weng, "On a class of Bayesian nonparametric estimates: II. Hazard rate estimates," *Ann. Inst. Stat. Math.*, vol. 41, no. 2, pp. 227–245, Jun. 1989.

[27] E. Phadia, *Prior Processes and Their Applications: Nonparametric Bayesian Estimation*, 2nd ed. Berlin, Germany: Springer, 2016.

[28] M. J. Schervish, *Theory of Statistics*. New York, NY, USA: Springer, 1995.

[29] R. Shibue and F. Komaki, "Firing rate estimation using infinite mixture models and its application to neural decoding," *J. Neurophysiol.*, vol. 118, no. 5, pp. 2902–2913, Nov. 2017.

[30] R. Shibue and F. Komaki, "Deconvolution of calcium imaging data using marked point processes," *PLOS Comput. Biol.*, vol. 16, no. 3, Mar. 2020, Art. no. e1007650.

[31] D. L. Snyder and M. I. Miller, *Random Point Processes in Time and Space*. New York, NY, USA: Springer, 1991.

[32] C. Stein, "Estimation of the mean of a multivariate normal distribution," in *Proc. Prague Symp. Asymptotic Statist.*, J. Hájek, Ed. Prague, Czechia: Universita Karlova, 1974, pp. 345–381.

[33] R. L. Streit, *Poisson Point Processes: Imaging, Tracking, and Sensing*. New York, NY, USA: Springer, 2010.

[34] X. Xu and F. Liang, "Asymptotic minimax risk of predictive density estimation for non-parametric regression," *Bernoulli*, vol. 16, no. 2, pp. 543–560, May 2010.

Fumiyasu Komaki (Member, IEEE) received the B.Eng. and M.Eng. degrees in mathematical engineering from The University of Tokyo, Japan, in 1987 and 1989, respectively, and the Ph.D. degree in statistical science from the Graduate University for Advanced Studies, Japan, in 1992. He is currently a Professor with the Department of Mathematical Informatics, The University of Tokyo, the Director of the Mathematics and Informatics Center, The University of Tokyo, a Unit Leader with the RIKEN Center for Brain Science, and an Affiliated Faculty Member with the International Research Center for Neurointelligence (IRCN), The University of Tokyo. His research interests include prediction theory, Bayesian theory, information geometry, and statistical modeling.