Nonparametric Bayesian inference for multidimensional compound Poisson processes

Shota Gugushvili[a,*]**, Frank van der Meulen**[b]**, Peter Spreij**[c]

[a]*Mathematical Institute, Leiden University, P.O. Box 9512,
2300 RA Leiden, The Netherlands*
[b]*Delft Institute of Applied Mathematics, Faculty of Electrical Engineering,
Mathematics and Computer Science, Delft University of Technology, Mekelweg 4,
2628 CD Delft, The Netherlands*
[c]*Korteweg–de Vries Institute for Mathematics, University of Amsterdam,
P.O. Box 94248, 1090 GE Amsterdam, The Netherlands*

shota.gugushvili@math.leidenuniv.nl (S. Gugushvili), f.h.vandermeulen@tudelft.nl (F. van der Meulen),
spreij@uva.nl (P. Spreij)

Abstract   Given a sample from a discretely observed multidimensional compound Poisson process, we study the problem of nonparametric estimation of its jump size density r_0 and intensity λ_0. We take a nonparametric Bayesian approach to the problem and determine posterior contraction rates in this context, which, under some assumptions, we argue to be optimal posterior contraction rates. In particular, our results imply the existence of Bayesian point estimates that converge to the true parameter pair (r_0, λ_0) at these rates. To the best of our knowledge, construction of nonparametric density estimators for inference in the class of discretely observed multidimensional Lévy processes, and the study of their rates of convergence is a new contribution to the literature.

*Corresponding author.

VEX

1 Introduction

Let $N = (N_t)_{t \geq 0}$ be a Poisson process of constant intensity $\lambda > 0$, and let $\{Y_j\}$ be independent and identically distributed (i.i.d.) $\mathbb{R}^d$-valued random vectors defined on the same probability space and having a common distribution function R, which is assumed to be absolutely continuous with respect to the Lebesgue measure with density r. Assume that N and $\{Y_j\}$ are independent and define the $\mathbb{R}^d$-valued process $X = (X_t)_{t \geq 0}$ by

$$X_t = \sum_{j=1}^{N_t} Y_j.$$

The process X is called a compound Poisson process (CPP) and forms a basic stochastic model in a variety of applied fields, such as, for example, risk theory and queueing; see [10, 21].

Suppose that, corresponding to the true parameter pair (λ_0, r_0), a sample X_Δ, $X_{2\Delta}, \ldots, X_{n\Delta}$ from X is available, where the sampling mesh $\Delta > 0$ is assumed to be fixed and thus independent of n. The problem we study in this note is nonparametric estimation of r_0 (and of λ_0). This is referred to as decompounding and is well studied for one-dimensional CPPs; see [2, 3, 6, 9, 24]. Some practical situations in which this problem may arise are listed in [9, p. 3964]. However, the methods used in the above papers do not seem to admit (with the exception of [24]) a generalization to the multidimensional setup. This is also true for papers studying nonparametric inference for more general classes of Lévy processes (of which CPPs form a particular class), such as, for example, [4, 5, 19]. In fact, there is a dearth of publications dealing with nonparametric inference for multidimensional Lévy processes. An exception is [1], where the setup is however specific in that it is geared to inference in Lévy copula models and that, unlike the present work, the high-frequency sampling scheme is assumed ($\Delta = \Delta_n \to 0$ and $n\Delta_n \to \infty$).

In this work, we will establish the posterior contraction rate in a suitable metric around the true parameter pair (λ_0, r_0). This concerns study of asymptotic frequentist properties of Bayesian procedures, which has lately received considerable attention in the literature (see, e.g., [14, 15]), and is useful in that it provides their justification from the frequentist point of view. Our main result says that for a β-Hölder regular density r_0, under some suitable additional assumptions on the model and the prior, the posterior contracts at the rate $n^{-\beta/(2\beta+d)} (\log n)^\ell$, which, perhaps up to a logarithmic factor, is arguably the optimal posterior contraction rate in our problem. Finally, our Bayesian procedure is adaptive: the construction of our prior does not require knowledge of the smoothness level β in order to achieve the posterior contraction rate given above.

The proof of our main theorem employs certain results from [14, 22] but involves a substantial number of technicalities specifically characteristic of decompounding.

We remark that a practical implementation of the Bayesian approach to decompounding lies outside the scope of the present paper. Preliminary investigations and a small scale simulation study we performed show that it is feasible and under certain conditions leads to good results. However, the technical complications one has to deal

with are quite formidable, and therefore the results of our study of implementational aspects of decompounding will be reported elsewhere.

The rest of the paper is organized as follows. In the next section, we introduce some notation and recall a number of notions useful for our purposes. Section 3 contains our main result, Theorem 2, and a brief discussion on it. The proof of Theorem 2 is given in Section 4. Finally, Section 5 contains the proof of the key technical lemma used in our proofs.

2 Preliminaries

Assume without loss of generality that $\Delta = 1$, and let $Z_i = X_i - X_{i-1}, i = 1, \ldots, n$. The $\mathbb{R}^d$-valued random vectors Z_i are i.i.d. copies of a random vector

$$Z = \sum_{j=1}^{T} Y_j,$$

where $\{Y_j\}$ are i.i.d. with distribution function R_0, whereas T, which is independent of $\{Y_j\}$, has the Poisson distribution with parameter λ_0. The problem of decompounding the jump size density r_0 introduced in Section 1 is equivalent to estimation of r_0 from observations $\mathcal{Z}_n = \{Z_1, Z_2, \ldots, Z_n\}$, and we will henceforth concentrate on this alternative formulation. We will use the following notation:

$\mathbb{P}_r$ law of Y_1,

$\mathbb{Q}_{\lambda,r}$ law of Z_1,

$\mathbb{R}_{\lambda,r}$ law of $X = (X_t, \ t \in [0, 1])$.

2.1 Likelihood

We will first specify the dominating measure for $\mathbb{Q}_{\lambda,r}$, which allows us to write down the likelihood in our model. Define the random measure μ by

$$\mu(B) = \big\{ \#t : (t, X_t - X_{t-}) \in B \big\}, \quad B \in \mathcal{B}\big([0, 1]\big) \otimes \mathcal{B}\big(\mathbb{R}^d \setminus \{0\}\big).$$

Under $\mathbb{R}_{\lambda,r}$, the random measure μ is a Poisson point process on $[0, 1] \times (\mathbb{R}^d \setminus \{0\})$ with intensity measure $\Lambda(\mathrm{d}t, \mathrm{d}x) = \lambda \mathrm{d}t r(x) \mathrm{d}x$. Provided that $\lambda, \widetilde{\lambda} > 0$, and $\widetilde{r} > 0$, by formula (46.1) on p. 262 in [23] we have

$$\frac{\mathrm{d}\mathbb{R}_{\lambda,r}}{\mathrm{d}\mathbb{R}_{\widetilde{\lambda},\widetilde{r}}}(X) = \exp\bigg(\int_0^1 \int_{\mathbb{R}^d} \log\bigg(\frac{\lambda r(x)}{\widetilde{\lambda}\widetilde{r}(x)} \bigg) \mu(\mathrm{d}t, \mathrm{d}x) - (\lambda - \widetilde{\lambda}) \bigg). \qquad (1)$$

The density $k_{\lambda,r}$ of $\mathbb{Q}_{\lambda,r}$ with respect to $\mathbb{Q}_{\widetilde{\lambda},\widetilde{r}}$ is then given by the conditional expectation

$$k_{\lambda,r}(x) = \mathbb{E}_{\widetilde{\lambda},\widetilde{r}}\bigg(\frac{\mathrm{d}\mathbb{R}_{\lambda,r}}{\mathrm{d}\mathbb{R}_{\widetilde{\lambda},\widetilde{r}}}(X) \,\Big|\, X_1 = x \bigg), \qquad (2)$$

where the subscript in the conditional expectation operator signifies the fact that it is evaluated under $\mathbb{R}_{\widetilde{\lambda},\widetilde{r}}$; see Theorem 2 on p. 245 in [23] and Corollary 2 on p. 246

there. Hence, the likelihood (in the parameter pair (λ, r)) associated with the sample $\mathcal{Z}_n$ is given by

$$L_n(\lambda, r) = \prod_{i=1}^{n} k_{\lambda, r}(Z_i). \tag{3}$$

2.2 Prior

We will use the product prior $\Pi = \Pi_1 \times \Pi_2$ for (λ_0, r_0). The prior Π_1 for λ_0 will be assumed to be supported on the interval $[\underline{\lambda}, \overline{\lambda}]$ and to possess a density π_1 with respect to the Lebesgue measure.

The prior for r_0 will be specified as a Dirichlet process mixture of normal densities. Namely, introduce a convolution density

$$r_{F, \Sigma}(x) = \int \phi_\Sigma(x - z) F(\mathrm{d}z), \tag{4}$$

where F is a distribution function on $\mathbb{R}^d$, Σ is a $d \times d$ positive definite real matrix, and ϕ_Σ denotes the density of the centered d-dimensional normal distribution with covariance matrix Σ. Let α be a finite measure on $\mathbb{R}^d$, and let $\mathcal{D}_\alpha$ denote the Dirichlet process distribution with base measure α (see [11] or, alternatively, [13] for a modern overview). Recall that if $F \sim \mathcal{D}_\alpha$, then for any Borel-measurable partition $B_1, \ldots, B_k$ of $\mathbb{R}^d$, the distribution of the vector $(F(B_1), \ldots, F(B_k))$ is the k-dimensional Dirichlet distribution with parameters $\alpha(B_1), \ldots, \alpha(B_k)$. The Dirichlet process location mixture of normals prior Π_2 is obtained as the law of the random function $r_{F, \Sigma}$, where $F \sim \mathcal{D}_\alpha$ and $\Sigma \sim G$ for some prior distribution function G on the set of $d \times d$ positive definite matrices. For additional information on Dirichlet process mixtures of normal densities, see, for example, the original papers [12] and [18], or a recent paper [22] and the references therein.

2.3 Posterior

Let $\mathcal{R}$ denote the class of probability densities of the form (4). By Bayes' theorem, the posterior measure of any measurable set $A \subset (0, \infty) \times \mathcal{R}$ is given by

$$\Pi(A|\mathcal{Z}_n) = \frac{\iint_A L_n(\lambda, r) \mathrm{d}\Pi_1(\lambda) \mathrm{d}\Pi_2(r)}{\iint L_n(\lambda, r) \mathrm{d}\Pi_1(\lambda) \mathrm{d}\Pi_2(r)}.$$

The priors Π_1 and Π_2 indirectly induce the prior $\Pi = \Pi_1 \times \Pi_2$ on the collection of densities $k_{\lambda, r}$. We will use the symbol Π to signify both the prior on (λ_0, r_0) and the density k_{λ_0, r_0}. The posterior in the first case will be understood as the posterior for the pair (λ_0, r_0), whereas in the second case as the posterior for the density k_{λ_0, r_0}. Thus, setting $\overline{A} = \{k_{\lambda, r} : (\lambda, r) \in A\}$, we have

$$\Pi(\overline{A}|\mathcal{Z}_n) = \frac{\int_{\overline{A}} L_n(k) \mathrm{d}\Pi(k)}{\int L_n(k) \mathrm{d}\Pi(k)}.$$

In the Bayesian paradigm, the posterior encapsulates all the inferential conclusions for the problem at hand. Once the posterior is available, one can next proceed with computation of other quantities of interest in Bayesian statistics, such as Bayes point estimates or credible sets.

2.4 Distances

The Hellinger distance $h(\mathbb{Q}_0, \mathbb{Q}_1)$ between two probability laws $\mathbb{Q}_0$ and $\mathbb{Q}_1$ on a measurable space $(\Omega, \mathfrak{F})$ is given by

$$h(\mathbb{Q}_0, \mathbb{Q}_1) = \left(\int \left(d\mathbb{Q}_0^{1/2} - d\mathbb{Q}_1^{1/2} \right)^2 \right)^{1/2}.$$

Assuming that $\mathbb{Q}_0 \ll \mathbb{Q}_1$, the Kullback–Leibler divergence $K(\mathbb{Q}_0, \mathbb{Q}_1)$ is

$$K(\mathbb{Q}_0, \mathbb{Q}_1) = \int \log \left(\frac{d\mathbb{Q}_0}{d\mathbb{Q}_1} \right) d\mathbb{Q}_0.$$

We also define the V-discrepancy by

$$V(\mathbb{Q}_0, \mathbb{Q}_1) = \int \log^2 \left(\frac{d\mathbb{Q}_0}{d\mathbb{Q}_1} \right) d\mathbb{Q}_0.$$

In addition, for positive real numbers x and y, we put

$$K(x, y) = x \log \frac{x}{y} - x + y,$$

$$V(x, y) = x \log^2 \frac{x}{y},$$

$$h(x, y) = \left| \sqrt{x} - \sqrt{y} \right|.$$

Using the same symbols K, V, and h is justified as follows. Suppose that Ω is a singleton $\{\omega\}$ and consider the Dirac measures δ_x and δ_y that put masses x and y, respectively, on Ω. Then $K(\delta_x, \delta_y) = K(x, y)$, and similar equalities are valid for the V-discrepancy and the Hellinger distance.

2.5 Class of locally β-Hölder functions

For any $\beta \in \mathbb{R}$, by $\lfloor \beta \rfloor$ we denote the largest integer strictly smaller than β, by $\mathbb{N}$ the set of natural numbers, whereas $\mathbb{N}_0$ stands for the union $\mathbb{N} \cup \{0\}$. For a multiindex $k = (k_1, \dots, k_d) \in \mathbb{N}_0^d$, we set $k_\cdot = \sum_{i=1}^d k_i$. The usual Euclidean norm of a vector $y \in \mathbb{R}^d$ is denoted by $\|y\|$.

Let $\beta > 0$ and $\tau_0 \geq 0$ be constants, and let $L : \mathbb{R}^d \to \mathbb{R}_+$ be a measurable function. We define the class $\mathcal{C}^{\beta, L, \tau_0}(\mathbb{R}^d)$ of locally β-Hölder regular functions as the set of all functions $r : \mathbb{R}^d \to \mathbb{R}$ such that all mixed partial derivatives $D^k r$ of r up to order $k_\cdot \leq \lfloor \beta \rfloor$ exist and, for every k with $k_\cdot = \lfloor \beta \rfloor$, satisfy

$$\left| (D^k r)(x + y) - (D^k) r(x) \right| \leq L(x) \exp(\tau_0 \|y\|^2) \|y\|^{\beta - \lfloor \beta \rfloor}, \quad x, y \in \mathbb{R}^d.$$

See p. 625 in [22] for this class of functions.

3 Main result

Define the complements of the Hellinger-type neighborhoods of (λ_0, r_0) by

$$A(\varepsilon_n, M) = \left\{ (\lambda, r) : h(\mathbb{Q}_{\lambda_0, r_0}, \mathbb{Q}_{\lambda, r}) > M\varepsilon_n \right\},$$

where $\{\varepsilon_n\}$ is a sequence of positive numbers. We say that ε_n is a posterior contraction rate if there exists a constant $M > 0$ such that

$$\Pi\big(A(\varepsilon_n, M)\big|\mathcal{Z}_n\big) \to 0$$

as $n \to \infty$ in $\mathbb{Q}_{\lambda_0,r_0}^n$-probability.

The ε-covering number of a subset B of a metric space equipped with the metric ρ is the minimum number of ρ-balls of radius ε needed to cover it. Let $\mathcal{Q}$ be a set of CPP laws $\mathbb{Q}_{\lambda,r}$. Furthermore, we set

$$B(\varepsilon, \mathbb{Q}_{\lambda_0,r_0}) = \big\{(\lambda, r) : \mathrm{K}(\mathbb{Q}_{\lambda_0,r_0}, \mathbb{Q}_{\lambda,r}) \leq \varepsilon^2, \mathrm{V}(\mathbb{Q}_{\lambda_0,r_0}, \mathbb{Q}_{\lambda,r}) \leq \varepsilon^2\big\}. \quad (5)$$

We recall the following general result on posterior contraction rates.

Theorem 1 ([14]). *Suppose that for positive sequences $\overline{\varepsilon}_n, \widetilde{\varepsilon}_n \to 0$ such that $n \min(\overline{\varepsilon}_n^2, \widetilde{\varepsilon}_n^2) \to \infty$, constants $c_1, c_2, c_3, c_4 > 0$, and sets $\mathcal{Q}_n \subset \mathcal{Q}$, we have*

$$\log N(\overline{\varepsilon}_n, \mathcal{Q}_n, h) \leq c_1 n \overline{\varepsilon}_n^2, \quad (6)$$

$$\Pi(\mathcal{Q} \setminus \mathcal{Q}_n) \leq c_3 e^{-n\widetilde{\varepsilon}_n^2(c_2+4)}, \quad (7)$$

$$\Pi\big(B(\widetilde{\varepsilon}_n, \mathbb{Q}_{\lambda_0,r_0})\big) \geq c_4 e^{-c_2 n \widetilde{\varepsilon}_n^2}. \quad (8)$$

Then, for $\varepsilon_n = \max(\overline{\varepsilon}_n, \widetilde{\varepsilon}_n)$ and a constant $M > 0$ large enough, we have that

$$\Pi\big(A(\varepsilon_n, M)\big|\mathcal{Z}_n\big) \to 0 \quad (9)$$

as $n \to \infty$ in $\mathbb{Q}_{\lambda_0,r_0}^n$-probability, assuming that the i.i.d. observations $\{Z_j\}$ have been generated according to $\mathbb{Q}_{\lambda_0,r_0}$.

In order to derive the posterior contraction rate in our problem, we impose the following conditions on the true parameter pair (λ_0, r_0).

Assumption 1. Denote by (λ_0, r_0) the true parameter values for the compound Poisson process.

(i) λ_0 is in a compact set $[\underline{\lambda}, \overline{\lambda}] \subset (0, \infty)$;

(ii) The true density r_0 is bounded, belongs to the set $\mathcal{C}^{\beta, L, \tau_0}(\mathbb{R}^d)$, and additionally satisfies, for some $\varepsilon > 0$ and all $k \in \mathbb{N}_0^d$, $k \leq \beta$,

$$\int \left(\frac{L}{r_0}\right)^{(2\beta+\varepsilon)/\beta} r_0 < \infty, \qquad \int \left(\frac{|D^k r_0|}{r_0}\right)^{(2\beta+\varepsilon)/k} r_0 < \infty.$$

Furthermore, we assume that there exist strictly positive constants a, b, c, and τ such that

$$r_0(x) \leq c \exp\big(-b\|x\|^\tau\big), \quad \|x\| > a.$$

The conditions on r_0 come from Theorem 1 in [22] and are quite reasonable. They simplify greatly when r_0 has a compact support.

We also need to make some assumptions on the prior Π defined in Section 2.2.

Assumption 2. The prior $\Pi = \Pi_1 \times \Pi_2$ on (λ_0, r_0) satisfies the following assumptions:

(i) The prior Π_1 on λ has a density π_1 (with respect to the Lebesgue measure) that is supported on the finite interval $[\underline{\lambda}, \overline{\lambda}] \subset (0, \infty)$ and is such that

$$0 < \underline{\pi}_1 \le \pi_1(\lambda) \le \overline{\pi}_1 < \infty, \quad \lambda \in [\underline{\lambda}, \overline{\lambda}], \tag{10}$$

for some constants $\underline{\pi}_1$ and $\overline{\pi}_1$;

(ii) The base measure α of the Dirichlet process prior $\mathcal{D}_\alpha$ is finite and possesses a strictly positive density on $\mathbb{R}^d$ such that for all sufficiently large $x > 0$ and some strictly positive constants a_1, b_1, and C_1,

$$1 - \overline{\alpha}\big([-x, x]^d\big) \le b_1 \exp\big(-C_1 x^{a_1}\big),$$

where $\overline{\alpha}(\cdot) = \alpha(\cdot)/\alpha(\mathbb{R}^d)$;

(iii) There exist strictly positive constants κ, a_2, a_3, a_4, a_5, b_2, b_3, b_4, C_2, C_3 such that for all $x > 0$ large enough,

$$G\big(\Sigma : \text{eig}_d\big(\Sigma^{-1}\big) \ge x\big) \le b_2 \exp\big(-C_2 x^{a_2}\big),$$

for all $x > 0$ small enough,

$$G\big(\Sigma : \text{eig}_1\big(\Sigma^{-1}\big) < x\big) \le b_3 x^{a_3},$$

and for any $0 < s_1 \le \cdots \le s_d$ and $t \in (0, 1)$,

$$G\big(\Sigma : s_j < \text{eig}_j\big(\Sigma^{-1}\big) < s_j(1+t), j = 1, \ldots, d\big) \ge b_4 s_1^{a_4} t^{a_5} \exp\big(-C_3 s_d^{\kappa/2}\big).$$

Here $\text{eig}_j(\Sigma^{-1})$ denotes the jth smallest eigenvalue of the matrix Σ^{-1}.

This assumption comes from [22, p. 626], to which we refer for an additional discussion. In particular, it is shown there that an inverse Wishart distribution (a popular prior distribution for covariance matrices) satisfies the assumptions on G with $\kappa = 2$. As far as α is concerned, we can take it such that its rescaled version $\overline{\alpha}$ is a nondegenerate Gaussian distribution on $\mathbb{R}^d$.

Remark 1. Assumption (10) requiring that the prior density π_1 is bounded away from zero on the interval $[\underline{\lambda}, \overline{\lambda}]$ can be relaxed to allowing it to take the zero value at the end points of this interval, provided that λ_0 is an interior point of $[\underline{\lambda}, \overline{\lambda}]$.

We now state our main result.

Theorem 2. *Let Assumptions 1 and 2 hold. Then there exists a constant $M > 0$ such that, as $n \to \infty$,*

$$\Pi\big(A\big((\log n)^\ell n^{-\gamma}, M\big)\big|\mathcal{Z}_n\big) \to 0$$

in $\mathbb{Q}_{\lambda_0, r_0}^n$-probability. Here

$$\gamma = \frac{\beta}{2\beta + d^*}, \qquad \ell > \ell_0 = \frac{d^*(1 + 1/\tau + 1/\beta) + 1}{2 + d^*/\beta}, \qquad d^* = \max(d, \kappa).$$

We conclude this section with a brief discussion on the obtained result: the logarithmic factor $(\log n)^\ell$ is negligible for practical purposes. If $\kappa = 1$, then the posterior

contraction rate obtained in Theorem 2 is essentially $n^{-2\beta/(2\beta+d)}$, which is the minimax estimation rate in a number of nonparametric settings. This is arguably also the minimax estimation rate in our problem as well (cf. Theorem 2.1 in [16] for a related result in the one-dimensional setting), although here we do not give a formal argument. Equally important is the fact that our result is adaptive: the posterior contraction rate in Theorem 2 is attained without the knowledge of the smoothness level β being incorporated in the construction of our prior Π. Finally, Theorem 2, in combination with Theorem 2.5 and the arguments on pp. 506–507 in [15], implies the existence of Bayesian point estimates achieving (in the frequentist sense) this convergence rate.

Remark 2. After completion of this work, we learned about the paper [8] that deals with nonparametric Bayesian estimation of intensity functions for Aalen counting processes. Although CPPs are in some sense similar to the latter class of processes, they are not counting processes. An essential difference between our work and [8] lies in the fact that, unlike [8], ours deals with discretely observed multidimensional processes. Also [8] uses the log-spline prior, or the Dirichlet mixture of uniform densities, and not the Dirichlet mixture of normal densities as the prior.

4 Proof of Theorem 2

The proof of Theorem 2 consists in verification of the conditions in Theorem 1. The following lemma plays the key role.

Lemma 1. *The following estimates are valid:*

$$K(\mathbb{Q}_{\lambda_0,r_0}, \mathbb{Q}_{\lambda,r}) \leq \lambda_0 K(\mathbb{P}_{r_0}, \mathbb{P}_r) + K(\lambda_0, \lambda), \tag{11}$$

$$V(\mathbb{Q}_{\lambda_0,r_0}, \mathbb{Q}_{\lambda,r}) \leq 2\lambda_0(1+\lambda_0)V(\mathbb{P}_{r_0}, \mathbb{P}_r) + 4\lambda_0 K(\mathbb{P}_{r_0}, \mathbb{P}_r)$$
$$+ 2V(\lambda_0, \lambda) + 4K(\lambda_0, \lambda) + 2K(\lambda_0, \lambda)^2, \tag{12}$$

$$h(\mathbb{Q}_{\lambda_0,r_0}, \mathbb{Q}_{\lambda,r}) \leq \sqrt{\lambda_0}\, h(\mathbb{P}_{r_0}, \mathbb{P}_r) + h(\lambda_0, \lambda). \tag{13}$$

Moreover, there exists a constant $\overline{C} \in (0, \infty)$, depending on $\underline{\lambda}$ and $\overline{\lambda}$ only, such that for all $\lambda_0, \lambda \in [\underline{\lambda}, \overline{\lambda}]$,

$$K(\mathbb{Q}_{\lambda_0,r_0}, \mathbb{Q}_{\lambda,r}) \leq \overline{C}\big(K(\mathbb{P}_{r_0}, \mathbb{P}_r) + |\lambda_0 - \lambda|^2\big), \tag{14}$$

$$V(\mathbb{Q}_{\lambda_0,r_0}, \mathbb{Q}_{\lambda,r}) \leq \overline{C}\big(V(\mathbb{P}_{r_0}, \mathbb{P}_r) + K(\mathbb{P}_{r_0}, \mathbb{P}_r) + |\lambda_0 - \lambda|^2\big), \tag{15}$$

$$h(\mathbb{Q}_{\lambda_0,r_0}, \mathbb{Q}_{\lambda,r}) \leq \overline{C}\big(|\lambda_0 - \lambda| + h(\mathbb{P}_{r_0}, \mathbb{P}_r)\big). \tag{16}$$

The proof of the lemma is given in Section 5. We proceed with the proof of Theorem 2.

Let $\varepsilon_n = n^{-\gamma}(\log n)^{\ell}$ for γ and $\ell > \ell_0$ as in the statement of Theorem 2. Set $\overline{\varepsilon}_n = 2\overline{C}\varepsilon_n$, where $\overline{C}$ is the constant from Lemma 1. We define the sieves of densities $\mathcal{F}_n$ as in Theorem 5 in [22]:

$$\mathcal{F}_n = \left\{ r_{F,\Sigma} \text{ with } F = \sum_{i=1}^{\infty} \pi_i \delta_{z_i} : z_i \in [-\alpha_n, \alpha_n]^d, \forall i \leq I_n; \sum_{i > I_n} \pi_i < \varepsilon_n; \right.$$

$$\left. \sigma_{0,n}^2 \leq \text{eig}_j(\Sigma) < \sigma_{0,n}^2\big(1 + \varepsilon_n^2/d\big)^{J_n} \right\},$$

where
$$I_n = \lfloor n\varepsilon_n^2 / \log n \rfloor, \qquad J_n = \alpha_n^{a_1} = \sigma_{0,n}^{-2a_2} = n,$$

and a_1 and a_2 are as in Assumption 2. We also put

$$\mathcal{Q}_n = \{\mathbb{Q}_{\lambda,r} : r \in \mathcal{F}_n, \lambda \in [\underline{\lambda}, \overline{\lambda}]\}. \tag{17}$$

In [22], sieves of the type $\mathcal{F}_n$ are used to verify conditions of Theorem 1 and to determine posterior contraction rates in the standard density estimation context. We further will show that these sieves also work in the case of decompounding by verifying the conditions of Theorem 1 for the sieves $\mathcal{Q}_n$ defined in (17).

4.1 Verification of (6)

Introduce the notation

$$\overline{h}_1(\lambda_1, \lambda_2) = \overline{C}|\lambda_1 - \lambda_2|, \qquad \overline{h}_2(r_1, r_2) = \overline{C}h(\mathbb{P}_{r_1}, \mathbb{P}_{r_2}).$$

Let $\{\lambda_i\}$ be the centers of the balls from a minimal covering of $[\underline{\lambda}, \overline{\lambda}]$ with $\overline{h}_1$-intervals of size $\overline{C}\varepsilon_n$. Let $\{r_j\}$ be centers of the balls from a minimal covering of $\mathcal{F}_n$ with $\overline{h}_2$-balls of size $\overline{C}\varepsilon_n$. By Lemma 1, for any $\mathbb{Q}_{\lambda,r} \in \mathcal{Q}_n$,

$$h(\mathbb{Q}_{\lambda,r}, \mathbb{Q}_{\lambda_i,r_j}) \le \overline{h}_1(\lambda, \lambda_i) + \overline{h}_2(r, r_j) \le \overline{\varepsilon}_n$$

by appropriate choices of i and j. Hence,

$$N(\overline{\varepsilon}_n, \mathcal{Q}_n, h) \le N(\overline{C}\varepsilon_n, [\underline{\lambda}, \overline{\lambda}], \overline{h}_1) \times N(\overline{C}\varepsilon_n, \mathcal{F}_n, \overline{h}_2),$$

and so

$$\log N(\overline{\varepsilon}_n, \mathcal{Q}_n, h) \le \log N(\overline{C}\varepsilon_n, [\underline{\lambda}, \overline{\lambda}], \overline{h}_1) + \log N(\overline{C}\varepsilon_n, \mathcal{F}_n, \overline{h}_2).$$

By Proposition 2 and Theorem 5 in [22], there exists a constant $c_1 > 0$ such that for all n large enough,

$$\log N(\overline{C}\varepsilon_n, \mathcal{F}_n, \overline{h}_2) = \log N(\varepsilon_n, \mathcal{F}_n, h) \le c_1 n\varepsilon_n^2 = \frac{c_1}{4\overline{C}^2} n\overline{\varepsilon}_n^2.$$

On the other hand,

$$\log N(\overline{C}\varepsilon_n, [\underline{\lambda}, \overline{\lambda}], \overline{h}_1) = \log N(\varepsilon_n, [\underline{\lambda}, \overline{\lambda}], |\cdot|),$$
$$\lesssim \log\left(\frac{1}{\varepsilon_n}\right)$$
$$\lesssim \log\left(\frac{1}{\overline{\varepsilon}_n}\right).$$

With our choice of $\overline{\varepsilon}_n$, for all n large enough, we have

$$\frac{c_1}{4\overline{C}^2} n\overline{\varepsilon}_n^2 \ge \log\left(\frac{1}{\overline{\varepsilon}_n}\right),$$

so that for all n large enough,

$$\log N(\overline{\varepsilon}_n, \mathcal{Q}_n, h) \leq \frac{c_1}{2\overline{C}^2} n\overline{\varepsilon}_n^2.$$

We can simply rename the constant $c_1/(2\overline{C}^2)$ in this formula into c_1, and thus (6) is satisfied with that constant.

4.2 Verification of (7) and (8)

We first focus on (8). Introduce

$$\widetilde{B}(\varepsilon, \mathbb{Q}_{\lambda_0, r_0}) = \left\{(\lambda, r) : K(\mathbb{P}_{r_0}, \mathbb{P}_r) \leq \varepsilon^2, V(\mathbb{P}_{r_0}, \mathbb{P}_r) \leq \varepsilon^2, |\lambda_0 - \lambda| \leq \varepsilon \right\}.$$

Suppose that $(\lambda, r) \in \tilde{B}(\varepsilon, \mathbb{Q}_{\lambda_0, r_0})$. From (14) we obtain

$$K(\mathbb{Q}_{\lambda_0, r_0}, \mathbb{Q}_{\lambda, r}) \leq \overline{C}K(\mathbb{P}_{r_0}, \mathbb{P}_r) + \overline{C}|\lambda - \lambda_0|^2 \leq 2\overline{C}\varepsilon^2.$$

Furthermore, using (15), we have

$$V(\mathbb{Q}_{\lambda_0, r_0}, \mathbb{Q}_{\lambda, r}) \leq \overline{C}V(\mathbb{P}_{r_0}, \mathbb{P}_r) + \overline{C}K(\mathbb{P}_{r_0}, \mathbb{P}_r) + \overline{C}|\lambda - \lambda_0|^2 \leq 3\overline{C}\varepsilon^2.$$

Combination of these inequalities with the definition of the set $B(\varepsilon, \mathbb{Q}_{\lambda_0, r_0})$ in (5) yields

$$\widetilde{B}(\varepsilon, \mathbb{Q}_{\lambda_0, r_0}) \subset B(\sqrt{3\overline{C}}\varepsilon, \mathbb{Q}_{\lambda_0, r_0}).$$

Consequently,

$$\begin{aligned}
\Pi\big(B(\sqrt{3\overline{C}}\varepsilon, \mathbb{Q}_{\lambda_0, r_0})\big) &\geq \Pi\big(\tilde{B}(\varepsilon, \mathbb{Q}_{\lambda_0, r_0})\big) \\
&= \Pi_1(|\lambda_0 - \lambda| \leq \varepsilon) \\
&\quad \times \Pi_2\big(r_{f, \Sigma} : K(\mathbb{P}_{r_0}, \mathbb{P}_{r_{F, \Sigma}}) \leq \varepsilon^2, \; V(\mathbb{P}_{r_0}, \mathbb{P}_{r_{F, \Sigma}}) \leq \varepsilon^2\big).
\end{aligned}$$
$$(18)$$

By Assumption 2(i),

$$\Pi_1(|\lambda_0 - \lambda| \leq \varepsilon) \geq \underline{\pi}_1 \varepsilon.$$

Furthermore, Theorem 4 in [22] yields that for some $A, C > 0$ and all sufficiently large n,

$$\begin{aligned}
\Pi_2\big(r_{F, \Sigma} : K(\mathbb{P}_{r_0}, \mathbb{P}_{r_{F, \Sigma}}) &\leq An^{-2\gamma}(\log n)^{2\ell_0}, V(\mathbb{P}_{r_0}, \mathbb{P}_{r_{F, \Sigma}}) \leq An^{-2\gamma}(\log n)^{2\ell_0}\big) \\
&\geq \exp\big(-Cn\{n^{-\gamma}(\log n)^{\ell_0}\}^2\big).
\end{aligned}$$

We substitute ε with $\sqrt{A}n^{-\gamma}(\log n)^{\ell_0}$ and write $\widetilde{\varepsilon}_n = \sqrt{3A\overline{C}}n^{-\gamma}(\log n)^{\ell_0}$ to arrive at

$$\Pi\big(B(\widetilde{\varepsilon}_n, \mathbb{Q}_{\lambda_0, r_0})\big) \geq \underline{\pi}_1\sqrt{A}n^{-\gamma}(\log n)^{\ell_0} \times \exp\left(-\frac{C}{3A\overline{C}}n\widetilde{\varepsilon}_n^2\right).$$

Now, since $\gamma < \frac{1}{2}$, for all n large enough, we have

$$\underline{\pi}_1\sqrt{A}n^{-\gamma}(\log n)^{\ell_0} \geq \exp\big(-n^{1-2\gamma}(\log n)^{2\ell_0}\big).$$

Consequently, for all n large enough,

$$\Pi(B(\widetilde{\varepsilon}_n, \mathbb{Q}_{\lambda_0, f_0}) \geq \exp\left(-\left(\frac{C+1}{3A\overline{C}}\right)n\widetilde{\varepsilon}_n^2\right). \tag{19}$$

Choosing $c_2 = \frac{C+1}{3A\overline{C}}$, we have verified (8) (with $c_4 = 1$).

For the verification of (7), we use the constants c_2 and $\widetilde{\varepsilon}_n$ as above. Note first that

$$\Pi(\mathcal{Q} \setminus \mathcal{Q}_n) = \Pi_2(\mathcal{F}_n^c).$$

By Theorem 5 in [22] (see also p. 627 there), for some $c_3 > 0$ and any constant $c > 0$, we have

$$\Pi_2(\mathcal{F}_n^c) \leq c_3 \exp\left(-(c+4)n\{n^{-\gamma}(\log n)^{\ell_0}\}^2\right),$$

provided that n is large enough. Thus,

$$\Pi(\mathcal{Q} \setminus \mathcal{Q}_n) \leq c_3 \exp\left(-\frac{c+4}{3A\overline{C}} n\widetilde{\varepsilon}_n^2\right).$$

Without loss of generality, we can take the positive constant c greater than $3A\overline{C}(c_2 + 4) - 4$. This gives

$$\Pi(\mathcal{Q} \setminus \mathcal{Q}_n) \leq c_3 \exp\left(-(c_2+4)n\widetilde{\varepsilon}_n^2\right),$$

which is indeed (7).

We have thus verified conditions (6)–(8), and the statement of Theorem 2 follows by Theorem 1 since $\bar{\varepsilon}_n \geq \widetilde{\varepsilon}_n$ (eventually).

5 Proof of Lemma 1

We start with a lemma from [7], which will be used three times in the proof of Lemma 1. Consider a probability space $(\Omega, \mathfrak{F}, \mathbb{P})$. Let $\mathbb{P}_0$ be a probability measure on $(\Omega, \mathfrak{F})$ and assume that $\mathbb{P}_0 \ll \mathbb{P}$ with Radon–Nikodym derivative $\zeta = \frac{d\mathbb{P}_0}{d\mathbb{P}}$. Furthermore, let $\mathfrak{G}$ be a sub-σ-algebra of $\mathfrak{F}$. The restrictions of $\mathbb{P}$ and $\mathbb{P}_0$ to $\mathfrak{G}$ are denoted $\mathbb{P}'$ and $\mathbb{P}'_0$, respectively. Then $\mathbb{P}'_0 \ll \mathbb{P}'$ and $\frac{d\mathbb{P}'_0}{d\mathbb{P}'} = \mathbb{E}_{\mathbb{P}}[\zeta | \mathfrak{G}] =: \zeta'$.

Lemma 2. *Let $g : [0, \infty) \to \mathbb{R}$ be a convex function. Then*

$$\mathbb{E}_{\mathbb{P}'} g(\zeta') \leq \mathbb{E}_{\mathbb{P}} g(\zeta).$$

The proof of the lemma consists in an application of Jensen's inequality for conditional expectations. This lemma is typically used as follows. The measures $\mathbb{P}$ and $\mathbb{P}_0$ are possible distributions of some random element X. If $X' = T(X)$ is some measurable transformation of X, then we consider $\mathbb{P}'$ and $\mathbb{P}'_0$ as the corresponding distributions of X'. Here T may be a projection. In the present context, we take $X = (X_t, t \in [0, 1])$ and $X' = X_1$, and so $\mathbb{P}$ in the lemma should be taken as $\mathbb{R} = \mathbb{R}_{\lambda, r}$ and $\mathbb{P}'$ as $\mathbb{Q} = \mathbb{Q}_{\lambda, r}$.

In the proof of Lemma 1, for economy of notation, a constant $c(\underline{\lambda}, \overline{\lambda})$ depending on $\underline{\lambda}$ and $\overline{\lambda}$ may differ from line to line. We also abbreviate $\mathbb{Q}_{\lambda_0, r_0}$ and $\mathbb{Q}_{\lambda, r}$ to $\mathbb{Q}_0$ and $\mathbb{Q}$, respectively. The same convention will be used for $\mathbb{R}_{\lambda_0, r_0}$, $\mathbb{R}_{\lambda, r}$, $\mathbb{P}_{r_0}$, and $\mathbb{P}_r$.

Proof of inequalities (11) and (14). Application of Lemma 2 with $g(x) = (x \log x)1_{\{x \geq 0\}}$ gives $K(\mathbb{Q}_0, \mathbb{Q}) \leq K(\mathbb{R}_0, \mathbb{R})$. Using (1) and the expression for the mean of a stochastic integral with respect to a Poisson point process (see, e.g., property 6 on p. 68 in [23]), we obtain that

$$
\begin{aligned}
K(\mathbb{R}_0, \mathbb{R}) &= \int \log\left(\frac{d\mathbb{R}_0}{d\mathbb{R}}\right) d\mathbb{R}_0 \\
&= \lambda_0 \int \log\left(\frac{\lambda_0 r_0}{\lambda r}\right) r_0 - (\lambda_0 - \lambda) \\
&= \lambda_0 K(\mathbb{P}_0, \mathbb{P}) + \left(\lambda_0 \log\left(\frac{\lambda_0}{\lambda}\right) - [\lambda_0 - \lambda]\right) \\
&= \lambda_0 K(\mathbb{P}_0, \mathbb{P}) + K(\lambda_0, \lambda).
\end{aligned}
$$

Now

$$
\begin{aligned}
\lambda_0 \log\left(\frac{\lambda_0}{\lambda}\right) - (\lambda_0 - \lambda) &= \lambda_0 \left| \log\left(\frac{\lambda}{\lambda_0}\right) - \left(\frac{\lambda}{\lambda_0} - 1\right) \right| \\
&\leq c(\underline{\lambda}, \overline{\lambda}) |\lambda_0 - \lambda|^2,
\end{aligned}
$$

where $c(\underline{\lambda}, \overline{\lambda})$ is some constant depending on $\underline{\lambda}$ and $\overline{\lambda}$. The result follows. $\qquad \square$

Proof of inequalities (12) and (15). We have

$$
\begin{aligned}
V(\mathbb{Q}_0, \mathbb{Q}) &= \mathbb{E}_{\mathbb{Q}_0}\left[\log^2\left(\frac{d\mathbb{Q}_0}{d\mathbb{Q}}\right)1_{\{\frac{d\mathbb{Q}_0}{d\mathbb{Q}} \geq 1\}}\right] + \mathbb{E}_{\mathbb{Q}_0}\left[\log^2\left(\frac{d\mathbb{Q}_0}{d\mathbb{Q}}\right)1_{\{\frac{d\mathbb{Q}_0}{d\mathbb{Q}} < 1\}}\right] \\
&= I + II.
\end{aligned}
$$

Application of Lemma 2 with $g(x) = (x \log^2(x))1_{\{x \geq 1\}}$ (which is a convex function) gives

$$
I \leq \mathbb{E}_{\mathbb{R}_0}\left[\log^2\left(\frac{d\mathbb{R}_0}{d\mathbb{R}}\right)1_{[\frac{d\mathbb{R}_0}{d\mathbb{R}} \geq 1]}\right] \leq V(\mathbb{R}_0, \mathbb{R}). \tag{20}
$$

As far as II is concerned, for $x \geq 0$, we have the inequalities

$$
\frac{x^2}{2} \leq e^x - 1 - x \leq 2(e^{x/2} - 1)^2.
$$

The first inequality is trivial, and the second is a particular case of inequality (8.5) in [15] and is equally elementary. The two inequalities together yield

$$
e^{-x} x^2 \leq 4(e^{-x/2} - 1)^2.
$$

Applying this inequality with $x = -\log \frac{d\mathbb{Q}_0}{d\mathbb{Q}}$ (which is positive on the event $\{\frac{d\mathbb{Q}_0}{d\mathbb{Q}} < 1\}$) and taking the expectation with respect to $\mathbb{Q}$ give

$$
II = \mathbb{E}_{\mathbb{Q}}\left[\frac{d\mathbb{Q}_0}{d\mathbb{Q}} \log^2 \frac{d\mathbb{Q}_0}{d\mathbb{Q}} 1_{\{\frac{d\mathbb{Q}_0}{d\mathbb{Q}} < 1\}}\right]
$$

$$\leq 4 \int \left(\sqrt{\frac{d\mathbb{Q}_0}{d\mathbb{Q}}} - 1 \right)^2 d\mathbb{Q}$$

$$= 4h^2(\mathbb{Q}_0, \mathbb{Q}) \leq 4K(\mathbb{Q}_0, \mathbb{Q}).$$

For the final inequality, see [20], p. 62, formula (12).

Combining the estimates on I and II, we obtain that

$$V(\mathbb{Q}_0, \mathbb{Q}) \leq V(\mathbb{R}_0, \mathbb{R}) + 4K(\mathbb{Q}_0, \mathbb{Q}). \tag{21}$$

After some long and tedious calculations employing (1) and the expressions for the mean and variance of a stochastic integral with respect to a Poisson point process (see, e.g., property 6 on p. 68 in [23] and Lemma 1.1 in [17]), we get that

$$V(\mathbb{R}_0, \mathbb{R}) = \lambda_0 \int \left\{ \log\left(\frac{\lambda_0}{\lambda}\right) + \log\left(\frac{r_0}{r}\right) \right\}^2 f_0$$

$$+ \lambda_0^2 \left\{ \int \log\left(\frac{r_0}{r}\right) r_0 + \log\left(\frac{\lambda_0}{\lambda}\right) - \left(1 - \frac{\lambda}{\lambda_0}\right) \right\}^2$$

$$= \text{III} + \text{IV}.$$

By the c_2-inequality $(a + b)^2 \leq 2a^2 + 2b^2$ we have

$$\text{III} \leq 2\lambda_0 \log^2\left(\frac{\lambda_0}{\lambda}\right) + 2\lambda_0 \int \log^2\left(\frac{r_0}{r}\right) r_0$$

$$= 2V(\lambda_0, \lambda) + 2\lambda_0 V(\mathbb{P}_0, \mathbb{P}), \tag{22}$$

from which we deduce

$$\text{III} \leq c(\underline{\lambda}, \overline{\lambda})|\lambda_0 - \lambda|^2 + 2\overline{\lambda}V(\mathbb{P}_0, \mathbb{P}) \tag{23}$$

for some constant $c(\underline{\lambda}, \overline{\lambda})$ depending on $\underline{\lambda}$ and $\overline{\lambda}$ only. As far as IV is concerned, the c_2-inequality and the Cauchy–Schwarz inequality give that

$$\text{IV} \leq 2\lambda_0^2 \left(\int \log\left(\frac{r_0}{r}\right) r_0 \right)^2 + 2\lambda_0^2 \left(\log\left(\frac{\lambda_0}{\lambda}\right) - \left[1 - \frac{\lambda}{\lambda_0}\right] \right)^2$$

$$\leq 2\lambda_0^2 V(\mathbb{P}_0, \mathbb{P}) + 2K(\lambda_0, \lambda)^2, \tag{24}$$

from which we find the upper bound

$$\text{IV} \leq 2\overline{\lambda}^2 V(\mathbb{P}_0, \mathbb{P}) + c(\underline{\lambda}, \overline{\lambda})|\lambda_0 - \lambda|^2 \tag{25}$$

for some constant $c(\underline{\lambda}, \overline{\lambda})$ depending on $\underline{\lambda}$ and $\overline{\lambda}$. Combining estimates (22) and (24) on III and IV with inequalities (21) and (11) yields (12). Similarly, the upper bounds (23) and (25), combined with (21) and (11), yield (15). $\square$

Proof of inequalities (13) and (16). First, note that for $g(x) = (\sqrt{x} - 1)^2 1_{[x \geq 0]}$,

$$h^2(\mathbb{Q}_0, \mathbb{Q}) = \mathbb{E}_{\mathbb{Q}}\left[\left(\sqrt{\frac{d\mathbb{Q}_0}{d\mathbb{Q}}} - 1 \right)^2 \right] = \mathbb{E}_{\mathbb{Q}}\left[g\left(\frac{d\mathbb{Q}_0}{d\mathbb{Q}}\right) \right].$$

Since g is convex, an application of Lemma 2 yields $h(\mathbb{Q}_0, \mathbb{Q}) \leq h(\mathbb{R}_0, \mathbb{R})$. Using (1) and invoking Lemma 1.5 in [17], in particular, using formula (1.30) in its statement, we get that

$$\begin{aligned} h(\mathbb{R}_0, \mathbb{R}) &\leq \|\sqrt{\lambda_0 r_0} - \sqrt{\lambda r}\| \\ &\leq \|\sqrt{\lambda_0 r_0} - \sqrt{\lambda_0 r}\| + \|\sqrt{\lambda_0 r} - \sqrt{\lambda r}\| \\ &\leq \sqrt{\lambda_0}\|\sqrt{r_0} - \sqrt{r}\| + |\sqrt{\lambda_0} - \sqrt{\lambda}| \\ &= \sqrt{\lambda_0} h(\mathbb{P}_0, \mathbb{P}) + h(\lambda_0, \lambda), \end{aligned}$$

where $\|\cdot\|$ denotes the L^2-norm. This proves (13). Furthermore, from this we obtain the obvious upper bound

$$h(\mathbb{R}_0, \mathbb{R}) \leq \sqrt{\overline{\lambda}}\, h(\mathbb{P}_0, \mathbb{P}) + \frac{1}{2\sqrt{\underline{\lambda}}}|\lambda_0 - \lambda|,$$

which yields (16). □

Acknowledgments

References

[1] Bücher, A., Vetter, M.: Nonparametric inference on Lévy measures and copulas. Ann. Stat. **41**(3), 1485–1515 (2013). MR3113819. doi:10.1214/13-AOS1116

[2] Buchmann, B., Grübel, R.: Decompounding: An estimation problem for Poisson random sums. Ann. Stat. **31**(4), 1054–1074 (2003). MR2001642. doi:10.1214/aos/1059655905

[3] Buchmann, B., Grübel, R.: Decompounding Poisson random sums: recursively truncated estimates in the discrete case. Ann. Inst. Stat. Math. **56**(4), 743–756 (2004). MR2126809. doi:10.1007/BF02506487

[4] Comte, F., Genon-Catalot, V.: Non-parametric estimation for pure jump irregularly sampled or noisy Lévy processes. Stat. Neerl. **64**(3), 290–313 (2010). MR2683462. doi:10.1111/j.1467-9574.2010.00462.x

[5] Comte, F., Genon-Catalot, V.: Estimation for Lévy processes from high frequency data within a long time interval. Ann. Stat. **39**(2), 803–837 (2011). MR2816339. doi:10.1214/10-AOS856

[6] Comte, F., Duval, C., Genon-Catalot, V.: Nonparametric density estimation in compound Poisson processes using convolution power estimators. Metrika **77**(1), 163–183 (2014). MR3152023. doi:10.1007/s00184-013-0475-3

[7] Csiszár, I.: Eine informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität von Markoffschen Ketten. Magy. Tud. Akad. Mat. Kut. Intéz. Közl. **8**, 85–108 (1963). MR0164374

[8] Donnet, S., Rivoirard, V., Rousseau, J., Scricciolo, C.: Posterior concentration rates for counting processes with Aalen multiplicative intensities (2014). arXiv:1407.6033 [stat.ME]

[9] Duval, C.: Density estimation for compound Poisson processes from discrete data. Stoch. Process. Appl. **123**(11), 3963–3986 (2013). MR3091096. doi:10.1016/j.spa.2013.06.006

[10] Embrechts, P., Klüppelberg, C., Mikosch, T.: Modelling Extremal Events: For Insurance and Finance. Appl. Math., vol. 33, p. 645. Springer, New York (1997). MR1458613. doi:10.1007/978-3-642-33483-2

[11] Ferguson, T.S.: A Bayesian analysis of some nonparametric problems. Ann. Stat. **1**, 209–230 (1973). MR0350949

[12] Ferguson, T.S.: Bayesian density estimation by mixtures of normal distributions. In: Recent Advances in Statistics, pp. 287–302. Academic Press, New York (1983). MR0736538

[13] Ghosal, S.: The Dirichlet process, related priors and posterior asymptotics. In: Bayesian Nonparametrics. Camb. Ser. Stat. Probab. Math., pp. 35–79. Cambridge Univ. Press, Cambridge (2010). MR2730660

[14] Ghosal, S., van der Vaart, A.W.: Entropies and rates of convergence for maximum likelihood and Bayes estimation for mixtures of normal densities. Ann. Stat. **29**(5), 1233–1263 (2001). MR1873329. doi:10.1214/aos/1013203453

[15] Ghosal, S., Ghosh, J.K., van der Vaart, A.W.: Convergence rates of posterior distributions. Ann. Stat. **28**(2), 500–531 (2000). MR1790007. doi:10.1214/aos/1016218228

[16] Gugushvili, S.: Nonparametric inference for partially observed Lévy processes. PhD thesis, University of Amsterdam (2008)

[17] Kutoyants, Y.A.: Statistical Inference for Spatial Poisson Processes. Lect. Notes Stat., vol. 134, p. 276. Springer (1998). MR1644620. doi:10.1007/978-1-4612-1706-0

[18] Lo, A.Y.: On a class of Bayesian nonparametric estimates. I. Density estimates. Ann. Stat. **12**(1), 351–357 (1984). MR0733519. doi:10.1214/aos/1176346412

[19] Neumann, M.H., Reiß, M.: Nonparametric estimation for Lévy processes from low-frequency observations. Bernoulli **15**(1), 223–248 (2009). MR2546805. doi:10.3150/08-BEJ148

[20] Pollard, D.: A User's Guide to Measure Theoretic Probability. Camb. Ser. Stat. Probab. Math., vol. 8, p. 351. Cambridge University Press, Cambridge (2002). MR1873379

[21] Prabhu, N.U.: Stochastic Storage Processes: Queues, Insurance Risk, Dams, and Data Communication, 2nd edn. Appl. Math., vol. 15, p. 206. Springer, New York (1998). MR1492990. doi:10.1007/978-1-4612-1742-8

[22] Shen, W., Tokdar, S.T., Ghosal, S.: Adaptive Bayesian multivariate density estimation with Dirichlet mixtures. Biometrika **100**(3), 623–640 (2013). MR3094441. doi:10.1093/biomet/ast015

[23] Skorohod, A.V.: Random Processes with Independent Increments, Nauka, Moscow (1964) (in Russian); English translation: Kluwer (1991). MR0182056

[24] Van Es, B., Gugushvili, S., Spreij, P.: A kernel type nonparametric density estimator for decompounding. Bernoulli **13**(3), 672–694 (2007). MR2348746. doi:10.3150/07-BEJ6091