Ann Inst Stat Math (2009) 61:181–195
DOI 10.1007/s10463-007-0131-7

Notes on estimating inverse-Gaussian and gamma subordinators under high-frequency sampling

Hiroki Masuda

Received: 14 September 2006 / Revised: 9 January 2007 / Published online: 14 April 2007

Abstract We study joint efficient estimation of two parameters dominating either the inverse-Gaussian or gamma subordinator, based on discrete observations sampled at $(t_i^n)_{i=1}^n$ satisfying $h_n := \max_{i \le n}(t_i^n - t_{i-1}^n) \to 0$ as $n \to \infty$. Under the condition that $T_n := t_n^n \to \infty$ as $n \to \infty$ we have two kinds of optimal rates, $\sqrt{n}$ and $\sqrt{T_n}$. Moreover, as in estimation of diffusion coefficient of a Wiener process the $\sqrt{n}$-consistent component of the estimator is effectively workable even when T_n does not tend to infinity. Simulation experiments are given under several h_n's behaviors.

Keywords Efficient estimation · Gamma subordinator · High-frequency sampling · Inverse-Gaussian subordinator · Optimal rate

1 Introduction

A subordinator $Z = (Z_t)_{t \in \mathbb{R}_+}$ is a one-dimensional non-decreasing càdlàg (right continuous and having left hand side limits) process a.s. starting from the origin with independent and stationary increments. For any subordinator without drift, there corresponds a Lévy measure ν satisfying $\int_0^1 |z|\nu(\mathrm{d}z) < \infty$ and supported by $\mathbb{R}_+$ for which

$$\varphi_{Z_t}(u) = \exp\left\{t \int (\mathrm{e}^{\mathrm{i}uz} - 1)\nu(\mathrm{d}z)\right\}, \quad u \in \mathbb{R}, \ t \in \mathbb{R}_+. \tag{1}$$

This is a special case of the so called Lévy–Khintchine formula. Here and in the sequel $u \mapsto \varphi_\xi(u)$ stands for the characteristic function of ξ, a random variable or a

H. Masuda (✉)
Graduate School of Mathematics, Kyushu University, 6-10-1 Hakozaki, Higashi-ku, Fukuoka 812-8581, Japan
e-mail: hiroki@math.kyushu-u.ac.jp

distribution. Given a subordinator Z the law at time 1, say $\mathcal{L}(Z_1)$, is uniquely associated with an infinitely divisible distribution whose support is contained in $\mathbb{R}_+ = [0, \infty)$; see, e.g., Bertoin (1996) for a systematic account of subordinators. A subordinator plays a role as a natural continuous-time analogue of independent and identically distributed (iid) sequence of positive random variables, which appears in, e.g., theory of dams (Moran 1959) and insurance theory (Huzak et al. 2004).

In this note we shall present two case studies of estimating a subordinator based on a kind of high-frequency discrete data. We shall consider two specific subordinators such that $\mathcal{L}(Z_1) = IG(\delta, \gamma)$ and $\Gamma(\delta, \gamma)$ admitting the density (w.r.t. the Lebesgue measure) given by

$$p(x; \delta, \gamma) = \frac{\delta e^{\delta\gamma}}{\sqrt{2\pi}} x^{-3/2} \exp\left\{ -\frac{1}{2}\left(\gamma^2 x + \frac{\delta^2}{x}\right)\right\} \mathbf{1}_{\mathbb{R}_+}(x), \tag{2}$$

$$p(x; \delta, \gamma) = \frac{\gamma^\delta}{\Gamma(\delta)} x^{\delta-1} \exp(-\gamma x) \mathbf{1}_{\mathbb{R}_+}(x), \tag{3}$$

respectively, where δ and γ are positive constants. The Lévy measure ν in the formula (1) of the inverse-Gaussian subordinator (resp. the gamma subordinator) admits a density (w.r.t. the Lebesgue measure) given by

$$g(z; \delta, \gamma) = \frac{\delta}{\sqrt{2\pi}} z^{-3/2} \exp\left(-\frac{\gamma^2 z}{2}\right) \mathbf{1}_{\mathbb{R}_+}(z),$$
$$\left(\text{resp. } g(z; \delta, \gamma) = \frac{\delta}{z} \exp(-\gamma z) \mathbf{1}_{\mathbb{R}_+}(z)\right),$$

hence a.s. has infinitely many jumps over each finite time interval as $\nu(\mathbb{R}) = \infty$. In each case we are interested in estimating $\theta = (\delta, \gamma)$ when Z is *discretely* observed with available data being

$$Z_{t_0^n}, Z_{t_1^n}, \ldots, Z_{t_n^n},$$

where $(t_i^n)_{i=0}^n$ is a nonrandom positive sequence satisfying

$$0 \equiv t_0^n < t_1^n < \cdots < t_n^n =: T_n$$

for each $n \in \mathbb{N}$. *Throughout this note we suppose*

$$\begin{cases} h_n := \max_{1 \le i \le n}(t_i^n - t_{i-1}^n) \to 0, \\ T_n \asymp n h_n, \end{cases} \tag{4}$$

as $n \to \infty$, where $a_n \asymp b_n$ means that there exists a constant $c > 0$ such that $c^{-1} \le a_n/b_n \le c$ for every n large enough. For joint estimation of δ and γ we shall additionally suppose $T_n \to \infty$; then the sampling scheme comes near the ideal but unrealistic *continuous* observation $(Z_t)_{t \le T}$ with $T \to \infty$.

Our main goal is to derive asymptotic behaviors of the corresponding maximum-likelihood estimators (MLE) of $\theta := (\delta, \gamma)$, say $\hat{\theta}_n = (\hat{\delta}_n, \hat{\gamma}_n)$. *In both cases we suppose that the parameter space* $\Theta \subset (0, \infty)^2$ *is a bounded domain whose closure,* *say* Θ^-, *is contained in* $(0,\infty)^2$ *and that there is a true parameter which lies in* Θ. Denote by P_θ^n the image measure of $(Z_{t_i^n})_{i=0}^n$ associated with θ. We shall derive the local asymptotic normality (LAN) as well as the asymptotic normality of the MLEs with rate $\mathrm{diag}(\sqrt{n},\sqrt{T_n})$ when $T_n\to\infty$, both uniform in Θ. Here the asymptotic normality with rate $\mathrm{diag}(\sqrt{n},\sqrt{T_n})$ means that

$$\begin{pmatrix}\sqrt{n}(\hat{\delta}_n-\delta)\\ \sqrt{T_n}(\hat{\gamma}_n-\gamma)\end{pmatrix}\Rightarrow\mathcal{N}_2(0,I(\theta)^{-1}):\tag{5}$$

especially, in both of the inverse-Gaussian and gamma cases, $I(\theta)$ turns out to be diagonal, which implies that *the joint ML estimation of* δ *and* γ *is asymptotically mutually independent under the asymptotics* (4) *with* $T_n\to\infty$; see the expressions (11) and (15) below for specified expressions of $I(\theta)$. This is a similar phenomenon to the well known case where Z is a Wiener process such that $\mathcal{L}(Z_1)=\mathcal{N}_1(\gamma,\delta)$, or, more generally, a diffusion process. Also, δ can be consistently estimated even when (T_n) is bounded in n; in this case γ may be unknown, hence a nuisance parameter.

On and after the next section, we shall present our asymptotic results in Sect. 2, then some simulation results in Sect. 3, and finally the proofs in Sect. 4. Section 4.1 contains a useful statement, which may apply to much more general situations than ours.

We end this section with some remarks.

Remark 1 If we were able to observe continuous data $(Z_t)_{t\in[0,T]}$, the likelihood theory has been already established: see, e.g., Akritas and Johnson (1981). Denote by P_θ^T the law of a sample path $(Z_t)_{t\in[0,T]}$ on the Skorohod space (i.e., the space of càdlàg processes endowed with the Skorohod topology), and fix any $\theta^i=(\delta^i,\gamma^i)$, $i=1,2$, and $T>0$. Then, in both of the inverse-Gaussian and gamma cases, $P_{\theta^1}^T$ and $P_{\theta^2}^T$ fail to be mutually absolutely continuous as soon as $\delta^1\neq\delta^2$ (see, e.g., Akritas and Johnson 1981, Theorem 4.1), so that we *cannot* consider the likelihood estimation from a continuous record while it makes sense in discrete-observation cases as the likelihood does exist.

Remark 2 Recall that, for general Lévy processes the likelihood function can be written down only up to the Fourier inversion formula. This fact makes it difficult to develop a general feasible procedure for likelihood estimation of a multi-dimensional parameter contained in a Lévy process (not necessarily a subordinator, of course) from high-frequency data. As a matter of fact, specification of the parametric optimal rates in estimating a general Lévy process seems to be an intricate problem. For example, Masuda (2006) previously studied the LAN property for discretely observed non-Gaussian stable Lévy processes, where various optimal rates were found for each component: there the scaling property, which is inherent in the stable case among general Lévy processes, was fully utilized, and it has been shown that the Fisher information matrix is always degenerate as long as joint estimation of scale and index parameters is concerned. Such a phenomenon does not arise in the present context.

Remark 3 If we suppose $t_i^n-t_{i-1}^n\equiv h>0$ instead of (4), then the situation is nothing but the classical iid framework as $\{Z_{ih}-Z_{(i-1)h}\}_{i\leq n}$ forms an iid sequence of random variables. In this case, Woerner (2001) systematically studied the LAN property for much more general Lévy processes when parameter's dimension is one, and Jongbloed and van der Meulen (2006) studied the parametric estimation of subordinators and induced Ornstein-Uhlenbeck processes based on the empirical characteristic function (their estimator is more robust than the MLE, but not efficient). The convergence rate of the MLE is then of course $\sqrt{n}$ for both of δ and γ, and we can readily get the closed form of the corresponding Fisher information matrices (depending on h in this case); more precisely, when $t_i^n - t_{i-1}^n \equiv h > 0$ the forms of the asymptotic Fisher information matrices are different from (11) and (15) in our framework. To be clearer, let us mention the gamma case. Suppose $t_i^n - t_{i-1}^n \equiv h > 0$ so that $T_n = nh \to \infty$. Then, by a routine argument (e.g., van der Vaart 1998) we can see that the MLE fulfils

$$\sqrt{n}\begin{pmatrix} \hat{\delta}_n - \delta \\ \hat{\gamma}_n - \gamma \end{pmatrix} \Rightarrow \mathcal{N}_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} h^2\psi'(\delta h) & -h/\gamma \\ -h/\gamma & h\delta/\gamma^2 \end{pmatrix}^{-1} \right),$$

where $\psi(x) := \partial_x \Gamma(x)/\Gamma(x)$ denotes the digamma function. Actually, the form of the asymptotic variance is different from that of (15), especially, non-diagonal for every $h > 0$. Though we have $\sqrt{n}$-consistency for both of $\hat{\delta}_n$ and $\hat{\gamma}$, we cannot accommodate the bounded-domain asymptotics, i.e., $T_n = O(1)$.

Remark 4 Finally we mention Basawa and Brockwell (1978, 1980), which studied estimation of the gamma subordinator (also, the stable subordinator) for a very different asymptotics from (4). Concerning the gamma subordinator Z associated with the density (3), they considered the following situation. Suppose we observe *all jump sizes* of Z greater than or equal to some $\epsilon > 0$, on an interval $[0, T]$. They considered the MLE $(\tilde{\delta}_{T,\epsilon}, \tilde{\gamma}_{T,\epsilon})$ of (δ, γ) based on the available sample $D(T, \epsilon) := \{(\tau_i, U_i(\epsilon)); i = 1, 2, \ldots, N(T, \epsilon)\}$, where τ_i denotes the ith jump time, $U_i(\epsilon)$ the jump size of ith observation, and $N(T, \epsilon)$ the number of the jumps in $[0, T]$ (all of these three quantities are random). Note that such an asymptotics are possible only when we can observe the whole path $(Z_t)_{t \le T}$; compared with this, our framework may be more realistic. Here are two remarks.

- Consider the asymptotics where $T \to \infty$ and then $\epsilon \to 0$. Then Basawa and Brockwell (1978, Eq. 10) states that

$$\sqrt{T}\begin{pmatrix} \tilde{\delta}_{T,\epsilon} - \delta \\ \tilde{\gamma}_{T,\epsilon} - \gamma \end{pmatrix} \Rightarrow \mathcal{N}_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & \gamma^2/\delta \end{pmatrix} \right),$$

  which reveals an important fact, namely, we can estimate δ at a *faster* rate than $\sqrt{T}$. Turning to our asymptotics (4), a faster optimal convergence rate $\sqrt{n}$ of $\hat{\delta}_n$ than $\sqrt{T_n}$ of $\hat{\gamma}_n$ is specified in (5).
- Actually Basawa and Brockwell (1978, 1980) did not make any further inquiry concerning the asymptotics $T \to \infty$ and $\epsilon \to 0$. Instead, they focused on the case where $T > 0$ *is fixed* and $\epsilon \to 0$. In this case we cannot estimate γ consistently (more precisely, the corresponding observed information is stochastically bounded in ϵ); this point is also seen in our setting, namely, we *cannot* estimate γ consistently, as soon as T_n is bounded. Then they proved that $|\log \epsilon|^{1/2}(\hat{\delta}_{T,\epsilon} - \delta) \Rightarrow$

$\mathcal{N}_1(0, \delta T)$ as $\epsilon \to 0$, whereas $\sqrt{n}(\hat{\delta}_n - \delta) \Rightarrow \mathcal{N}_1(0, \delta^2)$ in our framework (see Corollary 2).

Although the asymptotics used in Basawa and Brockwell (1978, 1980) and this note are of different types, as above stated, a similar phenomenon occurs, namely, we can construct a consistent estimator of δ from a high-frequency sample over any fixed $[0, T]$. Thus we may say that sufficiently high-frequency data of Z on any nonempty time interval is enough to estimate δ consistently; we might expect that this is the case for any "scale parameter" of Woerner' definition (cf. Woerner 2001, Sect. 3.2).

2 Statement of results

We use asymptotic symbols for $n \to \infty$ unless otherwise stated. Write $\Delta_i^n t = t_i^n - t_{i-1}^n$, $\Delta_i^n Z = Z_{t_i^n} - Z_{t_{i-1}^n}$, and $\partial_\theta = \partial/\partial\theta$. Note that the sequence $(\Delta_i^n Z)_{i=1}^n$ forms a rowwise independent triangular array fulfilling

$$\mathcal{L}(\Delta_i^n Z) = \mathcal{L}(Z_{\Delta_i^n t}) \tag{6}$$

for each $i \in \{1, 2, \ldots, n\}$. Let $|a| := (\sum_{k,l} a_{kl}^2)^{1/2}$ and $a^\top$ denote the transpose for any matrix $a = [a_{kl}]_{k,l}$. For a $\sigma(Z_{t_i^n} : i \leq n)$-measurable random variables $X_n(\theta)$, $n \in \mathbb{N}$, and a (possibly random) function $\theta \mapsto X(\theta)$, we write:

(i) "$X_n(\theta) \Rightarrow_u X(\theta)$" if
$$|P^{X_n(\theta_n)} f - P^{X(\theta)} f| \to 0$$
for every bounded continuous function f and nonrandom sequence $(\theta_n) \subset \Theta^-$ such that $\theta_n \to \theta$, where P^ξ denotes the law of ξ;

(ii) "$X_n(\theta) \to_u^p X(\theta)$" if for every $\epsilon > 0$ we have
$$P_{\theta_n}^n [|X_n(\theta_n) - X(\theta)| > \epsilon] \to 0$$
for every nonrandom sequence $(\theta_n) \subset \Theta^-$ such that $\theta_n \to \theta$.

By the definitions we see that $X_n(\theta) \to_u^p X(\theta)$ implies $X_n(\theta) \Rightarrow_u X(\theta)$.

Given a log-likelihood function $\theta \mapsto \ell_n(\theta)$ of class $C^2(\Theta)$, we write

$$\mathcal{S}_n(\theta) = \partial_\theta \ell_n(\theta) \quad \text{and} \quad \mathcal{I}_n(\theta) = -\partial_\theta^2 \ell_n(\theta), \tag{7}$$

the score function and the observed information matrix, respectively.

With the above-mentioned notation, we formulate the uniform LAN property in our context as follows. Write

$$A_n = \operatorname{diag}(\sqrt{n}, \sqrt{T_n}), \tag{8}$$

so that $A_n^{-1} \to 0$. We shall say that the experiments (P_θ^n) is "uniformly A_n-LAN with Fisher information $I(\theta)$" if:

[U1] $\ell_n(\theta + A_n^{-1}u_n) - \ell_n(\theta) - u_n^\top A_n^{-1}\mathcal{S}_n(\theta) + \frac{1}{2}u_n^\top A_n^{-1}\mathcal{I}_n(\theta)A_n^{-1}u_n \to_u^p 0$ for any nonrandom bounded sequence $(u_n) \subset \mathbb{R}^2$ such that $u_n \to u$;

[U2] There exists a nonrandom $I(\theta) \in \mathbb{R}^{2\otimes 2}$ positive definite for any $\theta \in \Theta^-$, such that $A_n^{-1}\mathcal{S}_n(\theta) \Rightarrow_u \mathcal{N}_2(0, I(\theta))$;

[U3] $A_n^{-1}\mathcal{I}_n(\theta)A_n^{-1} \to_u^p I(\theta)$, with the same $I(\theta)$ as in [U2].

The forthcoming asymptotic results reveal that the MLE $\hat{\theta}_n$ is asymptotically efficient in both of inverse-Gaussian and gamma cases (in the sense of Hajék and Le Cam; see van der Vaart 1998). If $\hat{\theta}_n \in \Theta^-$ is not well-defined, we may assign any number $\theta \in \Theta^-$ to $\hat{\theta}_n$; asymptotically, this does not matter.

2.1 Inverse-Gaussian case

When $\mathcal{L}(Z_1) = IG(\delta, \gamma)$ whose density is given by (2), we have $\mathcal{L}(Z_t) = IG(\delta t, \gamma)$ for each $t > 0$ since

$$\varphi_{IG(\delta,\gamma)}(u) = \exp\{\delta(\gamma - \sqrt{\gamma^2 - 2iu})\}.$$

On account of (6), the target log-likelihood function of $(Z_{t_i^n})_{i=0}^n$ is given by

$$\ell_n(\theta) = \sum_{i=1}^n \left\{ \log\delta + \delta\gamma\Delta_i^n t - \frac{1}{2}\left(\frac{\delta^2(\Delta_i^n t)^2}{\Delta_i^n Z} + \gamma^2\Delta_i^n Z\right)\right\}. \tag{9}$$

Solving $\partial_\theta \ell_n(\theta) = 0$, we get the explicit MLE:

$$\hat{\delta}_n = \left[\frac{1}{n}\left\{\sum_{i=1}^n \frac{(\Delta_i^n t)^2}{\Delta_i^n Z} - \frac{T_n^2}{Z_{T_n}}\right\}\right]^{-1/2}, \quad \hat{\gamma}_n = \frac{T_n\hat{\delta}_n}{Z_{T_n}}. \tag{10}$$

For the joint estimation we have the following.

Theorem 1 (Unbounded-domain asymptotics) *Let Z be a subordinator such that $\mathcal{L}(Z_1) = IG(\delta, \gamma)$ with $(\delta, \gamma) \in \Theta$, let $\ell_n(\theta)$ and $\hat{\theta}_n = (\hat{\delta}_n, \hat{\gamma}_n)$ be as in* (9) *and* (10)*, respectively, and suppose* (4) *and $T_n \to \infty$. Then (P_θ^n) is uniformly A_n-LAN with Fisher information*

$$I_{IG}(\theta) = \begin{pmatrix} 2/\delta^2 & 0 \\ 0 & \delta/\gamma \end{pmatrix}, \quad \theta \in \Theta, \tag{11}$$

and we have $A_n(\hat{\theta}_n - \theta) \Rightarrow_u \mathcal{N}_2(0, I_{IG}(\theta)^{-1})$.

If T_n does not tends to infinity, then the observed information associated with γ is stochastically bounded in n, and this is the case also for the gamma Lévy process; see (22) and (23). Nevertheless, this is not the case for estimating δ, and actually we may use the same estimate as in (10).

Corollary 1 (Bounded-domain asymptotics) *Let Z be a subordinator such that $\mathcal{L}(Z_1) = IG(\delta, \gamma)$ with $(\delta, \gamma) \in \Theta$, where $\gamma > 0$ is fixed while it may be unknown, let $\hat{\delta}_n$ be given by* (10)*, and suppose* (4) *and $T_n = O(1)$. Moreover suppose* *that $\delta \in (a, b)$ for some $0 < a < b < \infty$. Then (P^n_δ) is uniformly $\sqrt{n}$-LAN with Fisher information $2/\delta^2$, and $\hat{\delta}_n$ fulfils $\sqrt{n}(\hat{\delta}_n - \delta) \Rightarrow_u \mathcal{N}_1(0, \delta^2/2)$.*

2.2 Gamma case

When $\mathcal{L}(Z_1) = \Gamma(\delta, \gamma)$ whose density is given by (3), we get $\mathcal{L}(Z_t) = \Gamma(\delta t, \gamma)$ for each $t > 0$ since

$$\varphi_{\Gamma(\delta,\gamma)}(u) = (1 - iu/\gamma)^{-\delta}.$$

Thus the log-likelihood function of $(Z_{t_i^n})_{i=0}^n$ is given by

$$\ell_n(\theta) = \sum_{i=1}^n \left\{ \delta \Delta_i^n t \log \gamma - \log \Gamma(\delta \Delta_i^n t) + \delta \Delta_i^n t \log(\Delta_i^n Z) - \gamma \Delta_i^n Z \right\}. \tag{12}$$

The corresponding MLE solves

$$\sum_{i=1}^n (\Delta_i^n t)\{\log(\delta \Delta_i^n t) - \psi(\delta \Delta_i^n t)\} = T_n \log\left(\frac{Z_{T_n}}{T_n}\right) - \sum_{i=1}^n (\Delta_i^n t) \log\left(\frac{\Delta_i^n Z}{\Delta_i^n t}\right), \tag{13}$$

$$\gamma = \delta \frac{T_n}{Z_{T_n}}, \tag{14}$$

where $\psi(x) := \partial_x \Gamma(x)/\Gamma(x)$ denotes the digamma function.

For each $n \in \mathbb{N}$ the left-hand side of (13), say $f_n(\delta)$, is a smooth, positive, and strictly decreasing function of $\delta \in (0, \infty)$: $f_n(\delta) \to 0$ (resp. $\to \infty$) as $\delta \to \infty$ (resp. $\to 0$). So (13) admits a unique root $\hat{\delta}_n$ a.s. on the event where the right-hand side of (13) is positive, and we can simply apply, e.g., the bisection search in order to find the root of (13) readily.

Theorem 2 (Unbounded-domain asymptotics) *Let Z be a subordinator such that $\mathcal{L}(Z_1) = \Gamma(\delta, \gamma)$ with $(\delta, \gamma) \in \Theta$, and let $\ell_n(\theta)$ and $\hat{\theta}_n = (\hat{\delta}_n, \hat{\gamma}_n)$ be as in (12) and the solution to (13) and (14), respectively, and suppose (4) and $T_n \to \infty$. Then (P^n_θ) is uniformly A_n-LAN with Fisher information*

$$I_\Gamma(\theta) = \begin{pmatrix} 1/\delta^2 & 0 \\ 0 & \delta/\gamma^2 \end{pmatrix}, \quad \theta \in \Theta, \tag{15}$$

and we have $A_n(\hat{\theta}_n - \theta) \Rightarrow_u \mathcal{N}_2(0, I_\Gamma(\theta)^{-1})$.

We also have an analogue to Corollary 1.

Corollary 2 (Bounded-domain asymptotics) *Let Z be a subordinator such that $\mathcal{L}(Z_1) = \Gamma(\delta, \gamma)$ with $(\delta, \gamma) \in \Theta$, where $\gamma > 0$ is fixed while it may be unknown, let $\hat{\delta}_n$ be a solution of (13), and suppose (4) and $T_n = O(1)$. Moreover, suppose that $\delta \in (a, b)$ for some $0 < a < b < \infty$. Then (P^n_δ) is uniformly $\sqrt{n}$-LAN with Fisher information $1/\delta^2$, and $\hat{\delta}_n$ fulfils $\sqrt{n}(\hat{\delta}_n - \delta) \Rightarrow_u \mathcal{N}_1(0, \delta^2)$.*

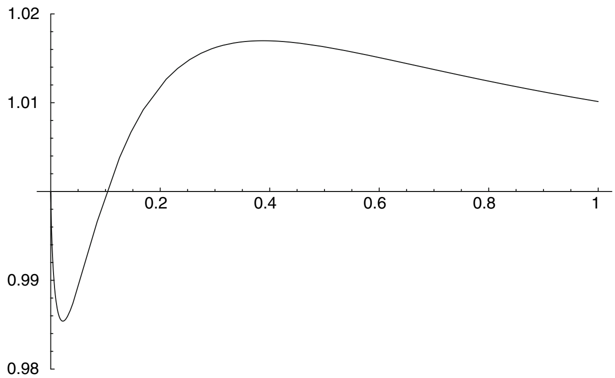

Fig. 1 The plot of the bounded function $(0, \infty) \ni \alpha \mapsto \{\log \alpha - \psi(\alpha)\}/\{(3\alpha + 1)/(6\alpha^2 + \alpha)\}$, which is increasing to 1 as $\alpha \to 0$ near the origin

Here is a remark for finding the root of (13) in the equidistant-sampling case, $h_n = t_i^n - t_{i-1}^n$ for each n: in this case (13) can be rewritten as

$$\log(\delta h_n) - \psi(\delta h_n) = \log\left(\frac{1}{n}\sum_{i=1}^{n} \Delta_i^n Z\right) - \frac{1}{n}\sum_{i=1}^{n} \log(\Delta_i^n Z).$$

Write this right-hand side as Y_n. By a standard argument it is not difficult to show that $h_n Y_n \to_u^p \delta^{-1} > 0$, hence Y_n becomes positive with probability tending to 1. Now, using the approximation (see Fig. 1)

$$\log(\alpha) - \psi(\alpha) \sim \frac{3\alpha + 1}{6\alpha^2 + \alpha}, \quad \alpha \searrow 0,$$

and taking the positivity of δ into account, we get the approximate MLE $\tilde{\delta}_n$ of δ given by

$$\tilde{\delta}_n = \frac{3 - Y_n}{12 h_n Y_n} + \left\{\left(\frac{3 - Y_n}{12 h_n Y_n}\right)^2 + \frac{1}{6 h_n Y_n}\right\}^{1/2},$$

which, together with $\tilde{\gamma}_n := \tilde{\delta}_n T_n / Z_{T_n}$ in case of $T_n \to \infty$, enables us to bypass the numerical optimization procedure.

Remark 5 As a naive estimator, we may consider moment estimator based on the first and second sample moments, utilizing the convergences

$$\frac{1}{T_n}\sum_{i=1}^{n} \Delta_i^n Z \quad \to_u^p \frac{\delta}{\gamma},$$
$$\frac{1}{T_n}\sum_{i=1}^{n} (\Delta_i^n Z)^2 \to_u^p \frac{\delta}{\gamma^2}.$$

Table 1 *Simulation* 1. Means and standard deviations (s.d.) of the estimate based on 1,000 independent trajectories. Here $h_n = n^{-0.7}$, i.e. $T_n = n^{0.3} \to \infty$, and $(\delta, \gamma) = (3, 2)$

n	T_n	$\hat{\delta}_n^{IG}$-mean (s.d.)	$\hat{\gamma}_n^{IG}$-mean (s.d.)	$\tilde{\delta}_n^{\Gamma}$-mean (s.d.)	$\tilde{\gamma}_n^{\Gamma}$-mean (s.d.)
50	3.23	3.0910 (0.3195)	2.1514 (0.5217)	3.0674 (0.4832)	2.2386 (0.8505)
100	3.98	3.0276 (0.2243)	2.0865 (0.4657)	3.0604 (0.3249)	2.2331 (0.7651)
300	5.54	3.0142 (0.1233)	2.0560 (0.3547)	3.0499 (0.1798)	2.1650 (0.6089)
500	6.45	3.0021 (0.0929)	2.0657 (0.3410)	3.0426 (0.1321)	2.1241 (0.5064)

However, the asymptotic behavior of the moment estimator, say $\hat{\theta}_{M,n}$, obtained from the above convergences is far from being optimal: using the delta method we can obtain

$$\sqrt{T_n}(\hat{\theta}_{M,n} - \theta) \Rightarrow \mathcal{N}_2\left(0, \begin{pmatrix} 2\delta & 2\gamma \\ 2\gamma & 3\gamma^2/\delta \end{pmatrix}\right).$$

This reveals that we cannot use $\hat{\theta}_{M,n}$ even as an initial estimate in applying Fisher's scoring or one-step improvement because of the slower convergence rate of $\hat{\delta}_{M,n}$.

3 Simulation experiments

We here report some numerical results. For simplicity we carried out equidistant high-frequency sampling cases, $\Delta_i^n t = h_n$ for each $i \leq n$. For gamma cases, we adopted the approximate MLE $\tilde{\theta}_n := (\tilde{\delta}_n, \tilde{\gamma}_n)$; the results will show that even $\tilde{\theta}_n$ performs well. In each simulation we simulated 1,000 independent discrete sample paths of Z, and then computed mean and standard deviation (s.d.) of the obtained 1,000 estimates. Throughout the true value is $(\delta, \gamma) = (3, 2)$. For generating pseudorandom-$\Gamma(p, q)$ numbers with $p \in (0, 1)$, we used the algorithm developed in Michael et al. (1976).

In Tables 1 and 2, we distinguish inverse-Gaussian and gamma cases by the superscripts "IG" and "Γ".

Simulation 1. We set $h_n = n^{-0.7}$, so that $T_n = n^{0.3} \to \infty$ and the jointly consistent estimation of δ and γ can be done. The results are given in Table 1.

Simulation 2. We set $h_n = n^{-0.3}$, so that $T_n = n^{0.7} \to \infty$; the total observation-time domain diverges faster than the previous case. It is observed that:

- accuracy of estimating δ is slightly worse than Simulation 1, which implies that δ can be estimated more accurately with more high-frequency data; and
- performance of estimating γ is much better than Simulation 1, because of the faster increase of T_n.

The results are given in Table 2.

Finally, let us look at a case of $T_n = O(1)$ in the inverse-Gaussian case.

Simulation 3. Set $h_n = 1/n$, so that $T_n \equiv 1$. In this case only δ can be consistently estimated. The results are given in Table 3: just for reference we also give estimates $\hat{\gamma}_n$, which badly behaved and have severe inevitable bias, as was expected.

Table 2 *Simulation 2*. Means and standard deviations (s.d.) of the estimate based on 1, 000 independent trajectories. Here $h_n = n^{-0.3}$, i.e. $T_n = n^{0.7} \to \infty$, $(\delta, \gamma) = (3, 2)$

n	T_n	$\hat{\delta}_n^{IG}$-mean (s.d)	$\hat{\gamma}_n^{IG}$-mean (s.d.)	$\tilde{\delta}_n^{\Gamma}$-mean (s.d)	$\tilde{\gamma}_n^{\Gamma}$-mean (s.d.)
50	15.46	3.0812 (0.3190)	2.0630 (0.2941)	3.1331 (0.5908)	2.1335 (0.5235)
100	25.12	3.0384 (0.2152)	2.0487 (0.2248)	3.0319 (0.3988)	2.0470 (0.3557)
300	54.20	3.0058 (0.1224)	2.0095 (0.1392)	2.9866 (0.2059)	2.0030 (0.2130)
500	77.50	3.0115 (0.0933)	2.0087 (0.1138)	2.9713 (0.1534)	1.9911 (0.1741)

Table 3 *Simulation 3*. Means and standard deviations (s.d.) of the estimate based on 1, 000 independent trajectories. Here $h_n = 1/n$, i.e. $T_n \equiv 1$, $(\delta, \gamma) = (3, 2)$

n	$\hat{\delta}_n^{IG}$-mean (s.d)	$\hat{\gamma}_n^{IG}$-mean (s.d.)
50	3.0760 (0.3220)	2.3424 (0.9575)
100	3.0431 (0.2200)	2.4193 (0.9441)
300	3.0211 (0.1230)	2.3300 (0.9481)
500	3.0114 (0.0935)	2.3123 (0.9044)
1000	3.0011 (0.0674)	2.3603 (0.9735)

4 Proofs

4.1 Preliminary

We shall utilize the results of Sweeting (1980). For convenience, in Theorem 3 below we shall rephrase them in a more convenient form, which can apply to much more general situations than our present setting. *In this subsection we shall forget our main context, and target at a log-likelihood function $\ell_n(\theta)$ associated with any (possibly dependent) observation stemming from a sequence of dominated experiments (P_θ^n).*

Let $\Theta \subset \mathbb{R}^p$ be a convex domain with compact closure Θ^-, and suppose $\theta \mapsto \ell_n(\theta)$ is of class $C^2(\Theta)$ for each $n \in \mathbb{N}$. Let $(A_n(\theta)) \subset \mathbb{R}^{p\otimes p}$ be a sequence of nonrandom matrices, which are continuous and positive definite in Θ^-, such that $|A_n^{-1}(\theta)| \to_u 0$, where $\to_u$ means the usual uniform convergence in Θ^-. Define $\mathcal{S}_n(\theta)$ and $\mathcal{I}_n(\theta)$ along with (7), let

$$H_n(\theta) = [H_n^{kl}(\theta)]_{k,l=1}^n := A_n(\theta)^{-1}\mathcal{I}_n(\theta)A_n(\theta)^{-1\top},$$

the normalized observed information matrix, and put $\mathcal{I}_n(\theta) = [\mathcal{I}_n^{kl}(\theta)]_{k,l=1}^p$. For $\Gamma := (\rho^k)_{k=1}^p \subset \Theta$, we introduce the notation

$$\mathcal{I}_n(\Gamma) := [\mathcal{I}_n^{kl}(\rho^k)]_{k,l=1}^p,$$

and, for each constant $a > 0$,

$$F_n^a(\theta) := \sup_{\rho^k \in \Theta^- : |\rho^k - \theta)| \le a|A_n(\theta)^{-1}|, k \le p} |A_n(\theta)^{-1}\{\mathcal{I}_n(\Gamma) - \mathcal{I}_n(\theta)\}A_n(\theta)^{-1\top}|.$$

Let I_p denote the p-dimensional identity matrix. As well as the uniform $A_n(\theta)$-LAN property introduced in Sect. 2, we say the experiments (P_θ^n) is "uniformly $A_n(\theta)$-LAMN (locally asymptotically mixed normal) with random Fisher information $I(\theta)$" if the following $[U1']$ and $[U2']$ are fulfilled:

$[U1']$ $\ell_n(\theta + A_n(\theta)^{-1\top}u_n) - \ell_n(\theta) - u_n^\top A_n(\theta)^{-1}\mathcal{S}_n(\theta) + \frac{1}{2}u_n^\top A_n(\theta)^{-1}\mathcal{I}_n(\theta) A_n(\theta)^{-1\top}u_n \to_u^p 0$ for any nonrandom bounded sequence $(u_n) \subset \mathbb{R}^p$ such that $u_n \to u$;

$[U2']$ There exist random vector $\Delta(\theta) \in \mathbb{R}^p$ and random matrix $H(\theta) \in \mathbb{R}^{p\otimes p}$ positive definite for any $\theta \in \Theta^-$, such that $(A_n(\theta)^{-1}\mathcal{S}_n(\theta), H_n(\theta)) \Rightarrow_u (\Delta(\theta), H(\theta))$, where $\mathcal{L}\{\Delta(\theta)|H(\theta) = H\} = \mathcal{N}_p(0, H)$.

See, e.g., van der Vaart (1998, Sect. 9) for a brief account of the general LAMN property.

By means of Sweeting (1980, Theorems 1 and 2) we then obtain

Theorem 3 *Suppose there exists a random matrix $H(\theta)$ defined on some probability space $(\Omega, \mathcal{A}, P)$, such that $H(\theta)$ is P-a.s. positive definite in $\theta \in \Theta^-$ and that*

$$H_n(\theta) \Rightarrow_u H(\theta). \tag{16}$$

Moreover, suppose that for every $a > 0$ we have

$$\sup_{\rho\in\Theta^-:|\rho-\theta|\le a|A_n(\theta)^{-1}|} |A_n(\theta)^{-1}A_n(\rho) - I_p| \to_u 0 \tag{17}$$

and

$$F_n^a(\theta) \to_u^p 0. \tag{18}$$

Then, we have

$$(A_n(\theta)^{-1}\mathcal{S}_n(\theta), H_n(\theta)) \Rightarrow_u (H(\theta)^{1/2}Z, H(\theta)), \tag{19}$$

where $\mathcal{L}(Z) = \mathcal{N}_p(0, I_p)$ and Z is independent of $H(\theta)$. Also, there exists a local maximizer $\hat{\theta}_n$ of $\ell_n(\theta)$ with probability tending to one, for which

$$A_n(\theta)^{-1}\mathcal{S}_n(\theta) - H_n(\theta)A_n(\theta)^\top(\hat{\theta}_n - \theta) \to_u^p 0. \tag{20}$$

In particular, (P_θ^n) is uniformly $A_n(\theta)$-LAMN with Fisher information $H(\theta)$ and

$$\left(H_n(\theta)^{1/2}A_n(\theta)^\top(\hat{\theta}_n - \theta), H_n(\theta)\right) \Rightarrow_u (Z, H(\theta)). \tag{21}$$

Remark 6 We have put the conditions (17) and (18) with seemingly stronger forms than C2 of Sweeting (1980). They may be equivalent if, e.g., $\sup_{n\in\mathbb{N}}\sup_{\theta\in\Theta^-}|A_n(\theta)||A_n(\theta)^{-1}| < \infty$, which is not the case when components of $A_n(\theta)$ increase with different rates as in our case (8).

Proof of Theorem 3 Equations (19) and (20) follow from Theorems 1 and 2 of Sweeting (1980), respectively, and (21) is then a direct consequence of the continuous mapping theorem. As for the uniform $A_n(\theta)$-LAMN property it suffices to verify $[U1']$ as we already have (19) and (20), but from Taylor's formula and (18) we see that

$$\left|\ell_n(\theta + A_n(\theta)^{-1}u_n) - \ell_n(\theta) - u_n^\top \mathcal{S}_n(\theta) + \frac{1}{2}u_n^\top H_n(\theta)u_n\right| \lesssim F_n^{\|u\|_\infty}(\theta) \to_u^p 0$$

for any bounded sequence $u_n \to u$ in $\mathbb{R}^p$. □

The following simple corollary of Theorem 3 is enough to prove our main results.

Corollary 3 *Suppose that $\theta \mapsto \ell_n(\theta)$ is three times differentiable in Θ, that $A_n(\theta)$ is independent of θ and diagonal, say*

$$A_n(\theta) = A_n = \mathrm{diag}(A_{1n}, A_{2n}, \dots, A_{pn}),$$

and that the following statements hold true for any constants $a > 0$ and $\epsilon > 0$ and any nonrandom sequence $(\theta_n) \subset \Theta^-$ such that $\theta_n \to \theta$:

(a) $E_{\theta_n}^n[H_n(\theta_n)] \to H(\theta)$, *where $H(\theta)$ is positive definite for each $\theta \in \Theta^-$;*
(b) $\mathrm{Var}_{\theta_n}^n[H_n^{kl}(\theta_n)] \to 0$ *for each $k, l \in \{1, \dots, p\}$;*
(c) $A_{kn}^{-1}A_{ln}^{-1}E_{\theta_n}^n\left[\sup_{\theta\in\Theta^-}|\partial_\theta \mathcal{I}_n^{kl}(\theta)|\right] = O(1)$ *for each $k, l \in \{1, \dots, p\}$.*

Then (19), (20) and (21) hold true. In particular, the experiments $\{P_\theta^n : n \in \mathbb{N}\}$ is uniformly A_n-LAN with the Fisher information matrix $H(\theta)$, and $A_n(\hat{\theta}_n - \theta) \Rightarrow_u \mathcal{N}_p(0, H(\theta)^{-1})$.

Proof of Corollary 3 Denote $a_n \lesssim b_n$ if there exists a constant $c > 0$ for which $a_n \le cb_n$ for every n large enough. First we show (16) and (18) under (a) to (c); clearly (17) is automatic. From Markov's inequality we have for each $\epsilon > 0$

$$P_{\theta_n}^n[|H_n(\theta_n) - H(\theta)| > \epsilon] \lesssim \sum_{k,l}\left\{\mathrm{Var}_{\theta_n}^n[H_n^{kl}(\theta_n)] + \left(E_{\theta_n}^n[H_n^{kl}(\theta_n)] - H^{kl}(\theta)\right)^2\right\},$$

the right-hand side tending to zero by means of (a) and (b). This shows that $H_n(\theta) \to_u^p H(\theta)$, hence (16). Next, let $G_n^a := P_{\theta_n}^n[F_n^a(\theta_n) > \epsilon]$ for each $a > 0$, where $\epsilon > 0$. As for (18), it suffices to observe that

$$\begin{aligned} G_n^a &\lesssim \sum_{k,l}\left\{\left(\sup_{\rho\in\Theta^-:|\rho-\theta_n|\le a|A_n^{-1}|}|\rho - \theta_n|\right)\cdot A_{kn}^{-1}A_{ln}^{-1}E_{\theta_n}^n\left[\sup_{\theta\in\Theta^-}|\partial_\theta \mathcal{I}_n^{kl}(\theta)|\right]\right\} \\ &\lesssim O(|A_n^{-1}|) = o(1) \end{aligned}$$

on account of Markov's inequality, Taylor's formula, and (c). □

Now let us return to our main framework, where the norming matrix is given by (8) and the maps $\theta \mapsto \ell_n(\theta)$ are smooth in both of the inverse-Gaussian and gamma cases. It suffices to prove the conditions (a) to (c) of Corollary 3. In the sequel, we shall consistently use the notation introduced above.

4.2 Inverse-Gaussian case

Proof of Theorem 1. Direct computations yield

$$H_n(\theta) = \begin{pmatrix} \dfrac{1}{\delta^2} + \dfrac{1}{n}\sum_{i=1}^{n} \dfrac{(\Delta_i^n t)^2}{\Delta_i^n Z} & -\sqrt{\dfrac{T_n}{n}} \\ \text{sym.} & \dfrac{Z_{T_n}}{T_n} \end{pmatrix}. \tag{22}$$

Fix any $(\theta_n) \subset \Theta^-$ such that $\theta_n = (\delta_n, \gamma_n) \to (\delta, \gamma)$, and observe that

$$\begin{aligned} E_{\theta_n}^n[\Delta_i^n Z] &= \frac{\delta_n \Delta_i^n t}{\gamma_n}, \\ E_{\theta_n}^n[(\Delta_i^n Z)^2] &= \frac{\delta_n \Delta_i^n t}{\gamma_n^3} + \left(\frac{\delta_n \Delta_i^n t}{\gamma_n}\right)^2, \\ E_{\theta_n}^n[(\Delta_i^n Z)^{-1}] &= \frac{1}{(\delta_n \Delta_i^n t)^2} + \frac{\gamma_n}{\delta_n \Delta_i^n t}, \\ E_{\theta_n}^n[(\Delta_i^n Z)^{-2}] &= \frac{1}{(\delta_n \Delta_i^n t)^4} + \frac{3\gamma_n}{(\delta_n \Delta_i^n t)^3} + \frac{2+\gamma_n^2}{(\delta_n \Delta_i^n t)^2}. \end{aligned}$$

With these quantities we see that

$$E_{\theta_n}^n[H_n(\theta_n)] = \begin{pmatrix} 2/\delta_n^2 + T_n\gamma_n/(n\delta_n) & -\sqrt{T_n/n} \\ \text{sym.} & \delta_n/\gamma_n \end{pmatrix} \to I_{IG}(\theta),$$

hence (a); recall (11). Also, observe that

$$\begin{aligned} \mathrm{Var}_{\theta_n}^n[H_n^{11}(\theta_n)] &\lesssim \frac{1}{n}\sum_{i=1}^{n} E_{\theta_n}^n\left[\left\{\frac{(\Delta_i^n t)^2}{\Delta_i^n Z} - \frac{\gamma_n \Delta_i^n t}{\delta_n} - \frac{1}{\delta_n^2}\right\}^2\right] \lesssim O(h_n) = o(1), \\ \mathrm{Var}_{\theta_n}^n[H_n^{22}(\theta_n)] &= \frac{\delta_n}{T_n\gamma_n^3} = o(1), \end{aligned}$$

where we applied Hölder's inequality to the former, from which (b) follows. (c) is clear in view of (22). □

Proof of Corollary 1. Note that in the proof of Theorem 1 the condition $T_n \to \infty$ was not used for $H_n^{11}(\theta)$, so that the claim follows on applying the continuous mapping theorem. □

4.3 Gamma case

Proof of Theorem 2. In this case the observed information is nonrandom and

$$H_n(\theta) = \begin{pmatrix} \frac{1}{\delta^2 n}\sum_{i=1}^{n}(\delta\Delta_i^n t)^2\psi'(\delta\Delta_i^n t) & -\sqrt{\frac{T_n}{\gamma n}} \\ \text{sym.} & \frac{\delta}{\gamma^2} \end{pmatrix}, \tag{23}$$

ψ' denoting the derivative of ψ, hence (b) is automatic. It suffices to only consider $H_n^{11}(\theta)$. Again fix any $(\theta_n) \subset \Theta^-$ such that $\theta_n = (\delta_n, \gamma_n) \to (\delta, \gamma)$. On account of (4) we can find $n_1 \in \mathbb{N}$ for which

$$\sup_{n\geq n_1}\sup_{i\leq n}(\delta_n\Delta_i^n t) < 1.$$

It follows from Abramowitz and Stegun (1992, Formula 6.4.10) that

$$|z^{m+1}\psi^{(m)}(z) - (-1)^{m+1}m!| = \left|(-1)^{m+1}m!\sum_{k=1}^{\infty}\left(\frac{z}{z+k}\right)^{m+1}\right| \lesssim |z|^{m+1} \tag{24}$$

for each $m \in \mathbb{N}$ and $z \in (0,1)$. Now fix any $\epsilon > 0$. By (24) with $m = 1$ we can find an integer $n_2 \geq n_1$ for which

$$\sup_{n\geq n_2}\sup_{i\leq n}\left\{\delta_n^{-2}|(\delta_n\Delta_i^n t)^2\psi'(\delta_n\Delta_i^n t) - 1|\right\} < \epsilon/2$$

and $|\delta^{-2} - \delta_n^{-2}| < \epsilon/2$. Then, for every $n \geq n_2$ we see

$$\left|H_n^{11}(\theta_n) - \frac{1}{\delta^2}\right| \leq \frac{1}{n}\sum_{i=1}^{n}\left\{\delta_n^{-2}\left|(\delta_n\Delta_i^n t)^2\psi'(\delta_n\Delta_i^n t) - 1\right|\right\} + |\delta^{-2} - \delta_n^{-2}| < \epsilon,$$

yielding (a). Turning to (c), it follows from (24) with $m = 2$ that

$$|\partial_\delta H_n^{11}(\theta)| \lesssim \frac{1}{n}\sum_{i=1}^{n}\left\{|(\delta\Delta_i^n t)^3\psi''(\delta\Delta_i^n t) + 2| + 2\right\} \lesssim 1$$

uniformly in Θ^- for every n large enough, hence we are done. □

Proof of Corollary 2. This follows every bit as Corollary 1. □

Acknowledgments The author is grateful to the two anonymous referees for their valuable comments, which improved the presentation of the first version. This work was partly supported by Grant-in-Aid for Scientific Research from the Ministry of Education, Japan, and by the twenty-first century COE Program "Development of Dynamic Mathematics with High Functionality" at Kyushu University.

References

Abramowitz, M., Stegun, I. A. (eds.) (1992). *Handbook of mathematical functions with formulas, graphs, and mathematical tables.* Reprint of the 1972 edition. New York: Dover. Publications, Inc.

Akritas, M. G., Johnson, R. A. (1981). Asymptotic inference in Lévy processes of the discontinuous type. *Annals of Statistics*, *9*, 604–614.

Basawa, I. V., Brockwell, P. J. (1978). Inference for gamma and stable processes. *Biometrika*, *65*, 129–133.

Basawa, I. V., Brockwell, P. J. (1980). A note on estimation for gamma and stable processes. *Biometrika*, *67*, 234–236.

Bertoin, J. (1996). *Lévy processes.* Cambridge: Cambridge University Press.

Jongbloed, G., van der Meulen, F. H. (2006). Parametric estimation for subordinators and induced OU processes. *Scandinavian Journal of Statistics*, *33*, 825–847.

Masuda, H. (2006). Likelihood estimation of stable Lévy processes from discrete data. MHF preprint series 2006-18, Kyushu Univeresity.

Michael, J. R., Schucany, W. R., Haas, R.W. (1976). Generating random variates using transformations with multiple roots. *The American Statistician*, *30*, 88–90.

Moran, P. A. P. (1959). *The theory of storage.* New York: Wiley.

Huzak, M., Perman, M., Šikić, H., Vondraček, Z. (2004). Ruin probabilities and decompositions for general perturbed risk processes. *Annals of Applied Probability*, *14*, 1378–1397.

Sweeting, T. J. (1980). Uniform asymptotic normality of the maximum likelihood estimator. *Annals of Statistics*, *8*, 1375–1381. [Corrections: (1982) *Annals of Statistics*, *10*, 320.]

van der Vaart, A. W. (1998). *Asymptotic statistics.* Cambridge: Cambridge University Press.

Woerner, J. H. C. (2001). *Statistical analysis for discretely observed Lévy processes.* PhD thesis, University of Freiburg.