LOCAL ASYMPTOTIC NORMALITY FOR NORMAL INVERSE GAUSSIAN LÉVY PROCESSES WITH HIGH-FREQUENCY SAMPLING *

Reiichiro Kawai[1] and Hiroki Masuda[2]

Abstract. We prove the local asymptotic normality for the full parameters of the normal inverse Gaussian Lévy process X, when we observe high-frequency data $X_{\Delta_n}, X_{2\Delta_n}, \ldots, X_{n\Delta_n}$ with sampling mesh $\Delta_n \to 0$ and the terminal sampling time $n\Delta_n \to \infty$. The rate of convergence turns out to be $(\sqrt{n\Delta_n}, \sqrt{n\Delta_n}, \sqrt{n}, \sqrt{n})$ for the dominating parameter $(\alpha, \beta, \delta, \mu)$, where α stands for the heaviness of the tails, β the degree of skewness, δ the scale, and μ the location. The essential feature in our study is that the suitably normalized increments of X in small time is approximately Cauchy-distributed, which specifically comes out in the form of the asymptotic Fisher information matrix.

1. Introduction

Lévy processes have been recognized as building blocks for analyzing realistic data structure, which most often loses touch with the conventional Gaussianity especially when dealing with high-frequency data, such as intraday stock returns. For a stochastic-process model based on high-frequency data, one of the most fundamental, yet in no way obvious, issues is estimation of the dominating parameters involved, where we observe discrete-time sample $X_{\Delta_n}, X_{2\Delta_n}, \ldots, X_{n\Delta_n}$ stemming from a stochastic process $X = (X_t)_{t \in \mathbb{R}_+}$, where $\Delta_n \to 0$ denotes a diminishing sampling mesh. This high-frequency data framework often leads to a better understanding of fine data structure and estimation performance than in case of targeting the classical independent and identically distributed data with $\Delta_n \equiv \Delta > 0$, a fixed constant. Moreover, due to a wide variety of the class of Lévy processes, it is a rather difficult matter to formulate a unified and versatile parametric estimation procedure for the whole class of Lévy processes discretely observed at high frequency. In general, the marginal distribution $\mathcal{L}(X_t)$ for $t \neq 1$ do not belong to the same distribution family as $\mathcal{L}(X_1)$. Even if the existence of the Lebesgue density of $\mathcal{L}(X_{\Delta_n})$ is guaranteed, the corresponding likelihood function does not have a closed form, making the likelihood analysis intractable and difficult.

Among others, the normal inverse Gaussian (NIG) Lévy process exhibits attractive features: the tractability and the availability of a simple simulation method at arbitrary sampling frequencies (*i.e.*, for any $\Delta_n > 0$). The NIG distribution is a four-parameter family, derived as a special case of the highly flexible five-parameter

[1] School of Mathematics and Statistics, University of Sydney NSW 2006, Australia. `reiichiro.kawai@maths.usyd.edu.au`

[2] Institute of Mathematics for Industry, Kyushu University, 744 Motooka, Nishi-ku, Fukuoka 819-0395, Japan. `hiroki@imi.kyushu-u.ac.jp`

generalized hyperbolic (GH) distribution introduced by Barndorff–Nielsen [5] for investigating a distribution of size of wind-blown particles of sand. The GH distribution is known to be infinitely divisible (more strongly, selfdecomposable), hence, we can associate with it a Lévy process such that $\mathcal{L}(X_1)$ is the GH distribution. However, the GH Lévy processes has a drawback that $\mathcal{L}(X_t)$ no longer belongs to the GH family for $t \neq 1$, and moreover, the characteristic function of the GH distribution is rather complicated, so that the Fourier inversion formula is also intractable (see, *e.g.*, Prause [19], Lem. 1.18). One may consult Prause [19] in Section 1, for estimation of the GH distributions (hence contains the NIG case) in the classical iid setting, together with a series of empirical analysis concerning financial asset data. Although he studied explicit derivation of the score function and several numerical recipes to obtain maximum likelihood estimates, no theoretical background, such as its asymptotic distribution, was given. Prause [19] in Section 1.5, reported that the GH model is also appropriate for financial intraday (referred to as high-frequency) data, nevertheless, the rigorous analysis of the estimation based on the high-frequency setting $\Delta_n \to 0$ we have adopted in this paper is extremely difficult, due to the lack of the reproducing property.

Within the GH family, NIG and normal gamma (NG) distributions are known to have the reproducing property, which entails that, if the distribution at unit time for a Lévy process is NIG or NG, then its marginal distribution at any time belongs to the same distribution family. Under discrete sampling, the reproducing property combined with the Markov property helps to simplify the expression of the likelihood function, and its further asymptotic analysis as well. Besides, toward optimal inference and testing hypothesis concerning the dominating parameters of the NIG Lévy process, a fundamental step is to investigate asymptotic behavior of the likelihood-ratio random fields based on an available record $X_{\Delta_n}, X_{2\Delta_n}, \ldots, X_{n\Delta_n}$. In this paper, we investigate Local Asymptotic Normality (LAN) for NIG Lévy process observed at discrete time points under large-term and high-frequency sampling design, where $\Delta_n \to 0$ and $n\Delta_n \to \infty$. The concept of LAN was introduced by Le Cam (1924–2000) in [15] in order to study approximations (simplifications) of statistical tests for large sample, and nowadays has become a vital concept to establish asymptotic optimalities of estimation and test in large-sample framework. For a systematic account concerning the LAN theory, we refer to, among others, Le Cam and Yang [16], Strasser [23] and van der Vaart [24]. Also, Jacod [9] presents a nice concise review in this direction, with a particular focus on the case of diffusion processes. An earlier attempt at systematic study of the LAN for discretely observed Lévy processes was made by Woerner [25], where various LAN results were "individually" provided for each specific parameter, such as drift, diffusion, scale, and skeweness. (See Sect. 3 for more details.)

The objective of this paper is to derive the "full-parameter" LAN for NIG Lévy processes discretely observed at high frequency. Previously, on the one hand, in the high-frequency data framework, Masuda [18] (see also Aït-Sahalia and Jacod [2]) derived the full-parameter LAN for the non-Gaussian stable Lévy process with drift and symmetric Lévy density, and Kawai and Masuda [14] derived the full-parameter LAN for the Meixner Lévy process. In both cases, the Fisher information matrices are constantly degenerate, rendering that there is no guarantee for usual good asymptotic behavior of the corresponding maximum likelihood estimators. On the other hand, we have clarified that the Fisher information matrix in the NIG case is always non-degenerate. To the best of our knowledge, Theorem 3.1 below is the first fruitful "full-parameter" LAN for a high frequently observed infinite activity pure jump Lévy process which is not subordinator.

The rest of this paper is organized as follows. Section 2 is devoted to a brief review of basic facts on the normal inverse Gaussian Levy process and the LAN under high-frequency sampling. Section 3 states our main result, which provides the rate of convergence and the Fisher information matrix in closed form concerning the LAN for NIG Lévy processes discretely observed at high frequency. Also, we partly compare our result with the case of continuous observation, and clarify big differences between them. To maintain the flow of the paper, we collect proofs in Section 4. Our result requires rather lengthy proofs of somewhat routine nature. To avoid overloading the paper, we omit nonessential details in some instances.

2. Preliminaries

2.1. **Basic notation**

Throughout this paper, the following basic notation is used:

- $I(A)$ denotes the indicator function of any event A;
- $\mathcal{L}(X)$ denotes the distribution of a random element X;
- φ_a denotes the characteristic function of a, which indicates a distribution or a random variable;
- $\partial_x := (\frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_k})^\top$ and $\partial_x^2 := \partial_x \partial_x^\top$ for a vector $x = (x_j)_{j \leq k}$ with $\top$ denoting transpose, and also we sometimes use the notation f' for the derivative of a function f, when no confusion may occur for the differentiating variable;
- $M^{\otimes 2} := MM^\top$ for any matrix M;
- C denotes a generic positive constant which may vary at each appearance;
- $a_n \lesssim b_n$ and $a_n \sim b_n$ indicate that $a_n \leq C b_n$ for every n large enough and that $a_n/b_n \to 1$ as $n \to \infty$, respectively.

2.2. **Normal inverse Gaussian Lévy process**

A univariate Lévy process $X = (X_t)_{t \in \mathbb{R}_+}$ with finite mean has the Lévy-Khintchine representation

$$\varphi_{X_t}(u) = \exp\left\{ t\left(i\mu_0 u - \frac{1}{2}\sigma^2 u^2 + \int (\mathrm{e}^{iuz} - 1 - iuz)\nu(\mathrm{d}z) \right) \right\}, \tag{2.1}$$

where $\mu_0 \in \mathbb{R}$, $\sigma^2 \geq 0$, and $\nu(\mathrm{d}z)$ is a Lévy measure, *i.e.*, a σ-finite measure on $\mathbb{R}$ such that $\nu(\{0\}) = 0$ and $\int (1 \wedge |x|^2)\nu(\mathrm{d}z) < \infty$. When the generating triplet $(\mu_0, \sigma^2, \nu(\mathrm{d}z))$ depends on a finite-dimensional parameter $\theta \in \Theta \subset \mathbb{R}^p$, we denote by P_θ the distribution of X on the Skorohod space. We refer the reader to Sato [21] for a detailed account of Lévy processes.

The univariate normal inverse Gaussian (NIG) distribution, denoted by $NIG(\alpha, \beta, \delta, \mu)$, is the selfdecomposable (hence infinitely divisible) distribution admitting a density

$$p(y; \alpha, \beta, \sigma, \mu) = \frac{\alpha\delta}{\pi} \exp\left\{ \delta\sqrt{\alpha^2 - \beta^2} + \beta(y - \mu) \right\} \frac{K_1(\alpha\sqrt{\delta^2 + (y-\mu)^2})}{\sqrt{\delta^2 + (y-\mu)^2}}, \quad y \in \mathbb{R}, \tag{2.2}$$

where $K_w(y)$, $w \in \mathbb{R}$, $y > 0$, denotes the modified Bessel function of the third kind[1] with index w:

$$K_w(y) = \frac{1}{2} \int_0^\infty x^{w-1} \exp\left\{ -\frac{y}{2}\left(x + \frac{1}{x} \right) \right\}\mathrm{d}x. \tag{2.3}$$

We write

$$\theta = (\alpha, \beta, \delta, \mu) \in \Theta \subset \mathbb{R}^4,$$

where the parameter space Θ is a bounded convex domain such that

$$\Theta^- \subset \{(\alpha, \beta, \delta, \mu) | \alpha > 0, \alpha > |\beta| \geq 0, \delta > 0, \mu \in \mathbb{R}\}, \tag{2.4}$$

where Θ^- denotes the closure of Θ. (Throughout, we rule out the case where $\alpha = |\beta| \geq 0$.) The distribution $NIG(\alpha, \beta, \delta, \mu)$ exhibits semi-heavy tails in the sense that the density behaves as a constant multiple of

---

[1]Here, we follow the terminology used in [7]. Note that some authors call this function "modified Bessel function of the second kind" or "modified Hankel function".

$|y|^{-3/2} \exp(-\alpha|y| + \beta y)$ for $|y| \to \infty$, so that moments of any order are finite. The mean and variance of $NIG(\alpha, \beta, \delta, \mu)$ are respectively given by

$$\mu + \frac{\beta\delta}{\sqrt{\alpha^2 - \beta^2}} \quad \text{and} \quad \frac{\alpha^2\delta}{(\alpha^2 - \beta^2)^{3/2}},$$

and the characteristic function by

$$\varphi_{NIG(\alpha,\beta,\delta,\mu)}(u) = \exp\left\{iu\mu + \delta(\sqrt{\alpha^2 - \beta^2} - \sqrt{\alpha^2 - (iu + \beta)^2})\right\}. \tag{2.5}$$

Now, the univariate NIG Lévy process is defined to be a Lévy process X starting from the origin such that $\mathcal{L}(X_1) = NIG(\alpha, \beta, \delta, \mu)$. It is clear from (2.5) that for any $\Delta_n > 0$ and $a \neq 0$

$$\mathcal{L}(a(X_{\Delta_n} - \mu\Delta_n)) = NIG\left(\frac{\alpha}{|a|}, \frac{\beta}{a}, \delta|a|\Delta_n, 0\right). \tag{2.6}$$

The generating triplet of X is given by $\mu_0 = \mu + \beta\delta/\sqrt{\alpha^2 - \beta^2}$, $\sigma^2 = 0$, and $\nu(\mathrm{d}z; \theta) = g(z; \alpha, \beta, \delta)\mathrm{d}z$ with

$$g(z; \alpha, \beta, \delta) = \frac{\alpha\delta}{\pi|z|}\mathrm{e}^{\beta z}K_1(\alpha|z|), \quad z \neq 0. \tag{2.7}$$

One can consult Barndorff–Nielsen [6,7] for more analytical facts concerning the NIG distribution and the NIG Lévy process.

2.3. LAN under high-frequency sampling

In this subsection, Θ denotes a convex domain in $\mathbb{R}^p$, and X a Lévy process (not necessarily the NIG one). For later use and convenience, let us introduce a general framework for LAN under high-frequency sampling.

Fix a $\theta \in \Theta$, and let X be observed at $t_j = t_j^n$, $j \leq n$, with $t_0^n < t_1^n < \cdots < t_n^n$ for each n. We denote by $x_j = x_{nj}$ the successive increments:

$$x_j = x_{nj} := X_{t_j} - X_{t_{j-1}}. \tag{2.8}$$

Because of the independent-increments property of X, the sequence $(x_j)_{j \leq n}$ for each n forms an independent array. For simplicity, we set $t_j = j\Delta_n$ for some $\Delta_n > 0$, so that $\mathcal{L}(x_j) = \mathcal{L}(X_{\Delta_n})$ under P_θ for every $j \leq n$. Then, we denote by P_θ^n the distribution of $(X_{t_j})_{j \leq n}$ under P_θ.

Suppose that $\mathcal{L}(X_{\Delta_n})$ under P_θ admits an everywhere positive Lebesgue density, say $p_{\Delta_n}(y; \theta)$, which is of the class $\mathcal{C}^2(\Theta)$ for each $y \in \mathbb{R}$ as a function of $\theta \in \Theta$. According to the stationarity and independence of increments of X, the log-likelihood function takes the form

$$\ell_n(\theta) = \sum_{j=1}^n \log p_{\Delta_n}(x_j; \theta),$$

where $p_{\Delta_n}(x; \theta)$ denotes the density of X_{Δ_n} under P_θ.

Let (r_n) be a nonrandom positive definite diagonal matrices tending to 0 in norm, and $\mathcal{I}(\theta)$ a nonnegative definite symmetric $\mathbb{R}^p \otimes \mathbb{R}^p$ matrix. Pick any $h \in \mathbb{R}^p$. We may suppose that $\theta_n := \theta + r_n h \in \Theta$. We say that *LAN holds true at θ with rate r_n and Fisher information matrix $\mathcal{I}(\theta)$*, if the stochastic expansion

$$\log \frac{\mathrm{d}P_{\theta_n}^n}{\mathrm{d}P_\theta^n} = \ell_n(\theta_n) - \ell_n(\theta) = h^\top \mathcal{S}_n(\theta) - \frac{1}{2}h^\top \mathcal{I}(\theta)h + o_{P_\theta}(1) \tag{2.9}$$

holds true, where $\mathcal{S}_n(\theta) := r_n \partial_\theta \ell_n(\theta) \to \mathcal{N}_p(0, \mathcal{I}(\theta))$ weakly under P_θ, where $\mathcal{N}_p(0, \mathcal{I}(\theta))$ stands for the p-variate normal distribution with mean 0 and covariance $\mathcal{I}(\theta)$. Let us note that, in order to apply the general asymptotic

optimality theory based on the LAN, the matrix $\mathcal{I}(\theta)$ has to be positive definite; if not, the LAN is not of much help to clarify asymptotic optimality criteria.

If we have the LAN, then it is known that general criteria for asymptotic optimalities of estimation and testing hypotheses follows from the LAN. Here, let us briefly mention the following (see Le Cam and Yang [16], Strasser [23] and van der Vaart [24] for more details): if one has asymptotically normally distributed estimator $\hat{\theta}_n$ of θ, say $c_n^{-1}(\hat{\theta}_n - \theta) \to \mathcal{N}_p(0, \mathcal{V}(\theta))$ weakly under P_θ where $c_n^{-1} \to \infty$ and $\mathcal{V}(\theta) \in \mathbb{R}^p \otimes \mathbb{R}^p$ is positive definite, then the maximal rate of convergence and the minimal asymptotic covariance matrix are given by r_n^{-1} and $\mathcal{I}(\theta)^{-1}$, respectively. Namely, the optimal quantities are explicitly provided by the form of the LAN obtained. In our main result (Thm. 3.1 below), the rate and the Fisher information matrix are specified by (3.3) and (3.2), respectively, where the latter turns out to be positive definite for each $\theta \in \Theta$.

3. MAIN RESULT

Let X be a Lévy process such that $\mathcal{L}(X_1) = NIG(\alpha, \beta, \delta, \mu)$ (recall (2.6) and suppose that available data is $(X_{j\Delta_n})_{j \leq n}$ with

$$\Delta_n \to 0 \quad \text{and} \quad n\Delta_n \to \infty. \tag{3.1}$$

Define the matrix $\mathcal{I}(\theta) = [\mathcal{I}_{kl}(\theta)]_{k,l=1}^4$ for $\theta = (\alpha, \beta, \delta, \mu) \in \Theta$ as follows:

$$\mathcal{I}(\theta) = \begin{pmatrix} \mathcal{I}_{11}(\theta) & \mathcal{I}_{12}(\theta) & 0 & 0 \\ & \mathcal{I}_{22}(\theta) & 0 & 0 \\ & & \mathcal{I}_{33}(\theta) & 0 \\ \text{sym.} & & & \mathcal{I}_{44}(\theta) \end{pmatrix}, \tag{3.2}$$

where

$$\mathcal{I}_{11}(\theta) := \frac{\delta}{\alpha\pi} \int_0^\infty \left(e^{(\beta/\alpha)y} + e^{-(\beta/\alpha)y} \right) y \frac{\{K_0(y)\}^2}{K_1(y)} dy,$$

$$\mathcal{I}_{12}(\theta) := -\frac{2\alpha\delta}{\pi(\alpha^2 - \beta^2)} \left\{ 1 + \frac{\beta}{\sqrt{\alpha^2 - \beta^2}} \arctan\left(\frac{\beta}{\sqrt{\alpha^2 - \beta^2}} \right) \right\},$$

$$\mathcal{I}_{22}(\theta) := \frac{\alpha^2\delta}{(\alpha^2 - \beta^2)^{3/2}},$$

$$\mathcal{I}_{33}(\theta) := \frac{1}{2\delta^2},$$

$$\mathcal{I}_{44}(\theta) := \frac{1}{2\delta^2}.$$

Here, the integral in $\mathcal{I}_{11}(\theta)$ is indeed finite; see Lemma 4.8[2].

Let

$$r_n = \text{diag}(r_{1n}, r_{2n}, r_{3n}, r_{4n}) := \text{diag}\left(\frac{1}{\sqrt{n\Delta_n}}, \frac{1}{\sqrt{n\Delta_n}}, \frac{1}{\sqrt{n}}, \frac{1}{\sqrt{n}} \right). \tag{3.3}$$

Our main result is the following, which clarifies a crucial contrast between *discrete* and *continuous* observations (see Cor. 3.3 for the latter case).

Theorem 3.1. Let X be as above and suppose (2.4) and (3.1). Then LAN holds true at each $\theta \in \Theta$ with rate r_n and the Fisher information matrix $\mathcal{I}(\theta)$. In particular, $\mathcal{I}(\theta)$ is positive definite for each $\theta \in \Theta$.

---

[2]We note that Woerner ([26], Thm. 2 and Ex. 1) contains a slight typo concerning the expression of $\mathcal{I}_{22}(\theta)$: the exponent in the denominator is not 5/2 but 3/2.

Thus we have seen that the rate $\sqrt{n}$ for (δ, μ) is faster than $\sqrt{n\Delta_n}$ for (α, β). Such a phenomenon is known to arise in some specific cases of Lévy processes under high-frequency sampling. A prime example is the scaled Wiener process with drift, say $X_t = \mu t + \sigma w_t$, where w denotes the standard Wiener process. In this case, we have the LAN for each (μ, σ) at rate $(\sqrt{n\Delta_n}, \sqrt{n})$, which can be easily seen by, *e.g.*, applying Proposition 4.3 in Section 4.1 to the corresponding (fully Gaussian) likelihood function. The rate $\sqrt{n\Delta_n}$ is the discrete-sampling analogue to $\sqrt{T}$ in the case of continuous observation $(X_t)_{t \leq T}$ as $T \to \infty$; see, *e.g.*, Akritas and Johnson [3] for details. See also Masuda [17] for the cases of the gamma and the inverse Gaussian subordinators.

For non-Gaussian stable Lévy processes with drift and symmetric Lévy density, it turns out that the Fisher information matrix is singular at "every" θ whenever both the stability index and scale parameters are included in θ (see Aït-Sahalia and Jacod [2] and Masuda [18] for details). In the present NIG case, normalized small time increment $(X_{\Delta_n} - \mu\Delta_n)/(\delta\Delta_n)$ is approximately Cauchy distributed (see Lem. 4.6 below). If X is the Cauchy Lévy process such that $\mathcal{L}(X_1)$ admits the Lebesgue density $x \mapsto (\delta/\pi)\{\delta^2 + (x - \mu)^2\}^{-1}$, then, by a direct application of Masuda ([18], Thm. 2.1), we see that the LAN holds true at each (δ, μ) with rate $\sqrt{n}$ and Fisher information matrix $\mathrm{diag}\{1/(2\delta^2), 1/(2\delta^2)\}$; we here do not suffer from the singularity of Fisher information, since the stability index is fixed at 1 and is not the parameter to be estimated. Returning to the present NIG case, we note that the last expression is exactly the same as the lower right 2×2 submatrix of $\mathcal{I}(\theta)$ in Theorem 3.1, that is, $\mathrm{diag}\{\mathcal{I}_{33}(\theta), \mathcal{I}_{44}(\theta)\} = \mathrm{diag}\{1/(2\delta^2), 1/(2\delta^2)\}$: informally speaking, this implies that we can extract information of the parameter (δ, μ) of the NIG Lévy process even in small time (in view of asymptotic Fisher information), and the information amount can be the same as if we observe the Cauchy Lévy process. Although we have additionally α and β for the NIG case, Theorem 3.1 implies that we can derive the LAN jointly for the full parameter θ as soon as $n\Delta_n \to \infty$. Moreover, in view of the block diagonal form of $\mathcal{I}(\theta)$, we may expect various possibilities of approximate conditional inference, simplified estimation procedure, and so on (see, *e.g.*, Cox and Reid [8] and Jørgensen and Knudsen [10]).

Woerner [25,26] previously studied LAN for discretely observed Lévy processes. Especially for the NIG Lévy process, she essentially provided LAN results "individual" for: β when $\Delta_n \equiv \Delta > 0$ and $\Delta_n \to 0$ ([25, Ex. 1] and ([26], Ex. 1); δ when $\Delta_n \equiv \Delta > 0$ ([25], Thm. 3.3); and finally, μ when $\Delta_n \equiv \Delta > 0$ ([25], Thm. 3.5), while no LAN for α was given. In contrast, our Theorem 3.1 provides the information of asymptotically optimal "full-parameter" estimation.

It is interesting to compare Theorem 3.1 with the case of *continuous* observation. In order to state the continuous-observation LAN result for the NIG Lévy processes, let us first recall a general characterization of the absolute continuity. Let X be a Lévy process admitting the Lévy-Khintchine representation (2.1) with $(\mu_0, \sigma^2, \nu(\mathrm{d}z)) = (\mu_0(\theta), \sigma^2(\theta), \nu(\mathrm{d}z; \theta))$ for $\theta \in \Theta \subset \mathbb{R}^p$, and suppose that we observe $(X_t)_{t \in [0,T]}$. Let $P_\theta^{(T)}$ denote the restriction of P_θ to $\mathcal{F}_T$, the natural filtration generated by the continuous-time record $(X_t)_{t \leq T}$. The equivalence of $P_\theta^{(T)}$ and $P_{\theta'}^{(T)}$ for $\theta \neq \theta'$ is characterized by the following proposition borrowed from Raible ([20], Prop. 2.19). See also Sato ([21], Thm. 33.1.)

Proposition 3.2. Fix any $T > 0$ and $\theta, \theta' \in \Theta$. Then $P_\theta^{(T)}$ and $P_{\theta'}^{(T)}$ are equivalent iff the following conditions are fulfilled.

(a) $\nu(\mathrm{d}z; \theta') = \gamma(z; \theta, \theta')\nu(\mathrm{d}z, \theta)$ for some Borel function $\gamma(\cdot; \theta, \theta') : \mathbb{R} \to (0, \infty)$.
(b) $\mu_0(\theta') = \mu_0(\theta) + \int_\mathbb{R} z(\gamma(z; \theta, \theta') - 1)\nu(\mathrm{d}z; \theta) + \sqrt{\sigma^2(\theta)}b$ for some $b \in \mathbb{R}$.
(c) $\int_\mathbb{R}(1 - \sqrt{\gamma(z; \theta, \theta')})^2\nu(\mathrm{d}z; \theta) < \infty$.
(d) $\sigma^2(\theta') = \sigma^2(\theta)$.

As a corollary to Raible ([20], Prop. 2.20) based on Proposition 3.2, we have

Corollary 3.3. Let P_{θ_k}, $k = 1, 2$, denote the distribution of the NIG Lévy process with parameters $\theta_k = (\alpha_k, \beta_k, \delta_k, \mu_k) \in \Theta$, and fix any $T > 0$. Then $P_{\theta_1}^{(T)}$ and $P_{\theta_2}^{(T)}$ are equivalent iff $\delta_1 = \delta_2$ and $\mu_1 = \mu_2$.

Corollary 3.3 clears up an essential difference between the cases of continuous and high-frequency sampling for the NIG Lévy processes. Indeed, Corollary 3.3 enables us to study the LAN for (α, β) in the continuous-observation case, where the asymptotics are taken as $T \to \infty$. We do not touch the details in order not to

digress from the main topic, but only refer to Akritas and Johnson [3] for possible LAN for (α, β) at rate $\sqrt{T}$; specifically, they derived sufficient conditions under which a parametric family of continuously observed pure jump Lévy processes fulfills a LAN property, and also, they provided the general form of the Fisher information matrix for continuously observed pure pump Lévy processes, the matrix being represented as some integral with respect to the Lévy measure (see Eq. (2.4) in Akritas and Johnson [3]). On the contrary, as in the present NIG case, the full-parameter likelihood function may exist when we deal with the high-frequency (discrete-time) sample, so that the maximum-likelihood estimation of (δ, μ) becomes meaningful. Finally, let us mention that, as the rate of convergence of (δ, μ) is $\sqrt{n}$ free of Δ_n, it may not be necessary to impose that $n\Delta_n \to \infty$ for estimation of (δ, μ). In this case, we regard (α, β) as a nuisance parameter as in the case of estimation of σ in the aforementioned Wiener case.

4. Proof

We proceed as follows. First, in Section 4.1 we provide a useful general tool (Prop. 4.3) for proving Theorem 3.1. Next, we prepare some preliminary lemmas in Section 4.2 for investigating the likelihood function in question, whose expression together with its derivatives up to the second order are specified in Section 4.3. Finally, Sections 4.4 to 4.7 are devoted to verifications of the conditions of Proposition 4.3.

4.1. A tool for proving LAN under high-frequency sampling

In this section, as a continuation of Section 2.3, we prepare a useful tool for proving our main result. Our setup here covers general Lévy processes discretely observed at high frequency.

Write $g_{nj}(\theta) = \partial_\theta \log p_{\Delta_n}(x_j; \theta)$. The random fields $\log(\mathrm{d}P_{\theta_n}^n / \mathrm{d}P_\theta^n)$ on Θ admits the asymptotically quadratic structure

$$\log \frac{\mathrm{d}P_{\theta_n}^n}{\mathrm{d}P_\theta^n} = h^\top \sum_{j=1}^n r_n \{g_{nj}(\theta) - E_\theta[g_{nj}(\theta)]\} - \frac{1}{2}h^\top \sum_{j=1}^n E_\theta[\{r_n g_{nj}(\theta)\}^{\otimes 2}]h + o_{P_\theta}(1) \tag{4.1}$$

if it holds that

$$\limsup_{n \to \infty} \sum_{j=1}^n E_\theta[|r_n g_{nj}(\theta)|^2] < \infty, \tag{4.2}$$

$$\sum_{j=1}^n E_\theta\left[|r_n g_{nj}(\theta)|^2 I(|r_n g_{nj}(\theta)| \geq \epsilon)\right] \to 0 \quad \text{for every } \epsilon > 0, \tag{4.3}$$

$$\sum_{j=1}^n E_\theta\left[\left\{\sqrt{\frac{p_{\Delta_n}(x_j; \theta_n)}{p_{\Delta_n}(x_j; \theta)}} - 1 - \frac{1}{2}h^\top r_n g_{nj}(\theta)\right\}^2\right] \to 0. \tag{4.4}$$

If (4.2) and (4.3) hold true with $g_{nj}(\theta)$ replaced by $g_{nj}^*(\theta)$, where $g_{nj}^*(\theta) := 2\{\sqrt{p_{\Delta_n}(x_j; \theta_n)/p_{\Delta_n}(x_j; \theta)} - 1\}$, then we have the asymptotically quadratic structure (4.1) with $g_{nj}(\theta)$ replaced by $g_{nj}^*(\theta)$ in the right-hand side (Strasser [23], Thm. 74.2). Moreover, Strasser ([23], Cor. 74.4) in turn ensures that (4.1) itself holds true under (4.4). Thus, (4.1) indeed follows on verifying the conditions (4.2) to (4.4).

Now we assume that

$$\sum_{j=1}^n E_\theta[\{r_n g_{nj}(\theta)\}^{\otimes 2}] \to \mathcal{I}(\theta), \tag{4.5}$$

$$\sum_{j=1}^n |r_n E_\theta[g_{nj}(\theta)]|^2 \to 0. \tag{4.6}$$

Note that (4.5) implies (4.2); in Section 4.5, we in fact will check that (17) holds in our framework.

To treat the first term in the right-hand side of (4.1), we prepare the following.

Lemma 4.1. Suppose the conditions (4.3), (4.5), and (4.6) hold. Then the first term in the right-hand side of (4.1) weakly under P_θ tends to $\mathcal{N}_p(0, h^\top \mathcal{I}(\theta)h)$.

Proof. Introduce the centered variables $\chi_{nj} = \chi_{nj}(\theta) := r_n\{g_{nj}(\theta) - E_\theta[g_{nj}(\theta)]\}$. By means of the central limit theorem for rowwise independent triangular arrays (*e.g.* Kallenberg [11], Thm. 5.12 combined with the Cramér–Wald device), the claim follows from the convergence of the cumulative variance and the Lindeberg condition, that is, $\sum_{j=1}^n E_\theta[\chi_{nj}^{\otimes 2}] \to \mathcal{I}(\theta)$ and $\sum_{j=1}^n E_\theta[|\chi_{nj}|^2 I(|\chi_{nj}| \geq \epsilon')] \to 0$ for every $\epsilon' > 0$, respectively. The former is obtained by noting that

$$\sum_{j=1}^n E_\theta[\chi_{nj}^{\otimes 2}] = \sum_{j=1}^n E_\theta[\{r_n g_{nj}(\theta)\}^{\otimes 2}] - \sum_{j=1}^n \{r_n E_\theta[g_{nj}(\theta)]\}^{\otimes 2}$$

and then applying (4.5) and (4.6). Now fix any $\epsilon' > 0$. Since (4.6) entails that $|\sqrt{n} r_n E_\theta[g_{n1}(\theta)]| \to 0$, we can find $\epsilon > 0$ such that $\epsilon' - |r_n E_\theta[g_{n1}(\theta)]| \geq \epsilon$ for every n large enough. Accordingly,

$$\sum_{j=1}^n E_\theta[|\chi_{nj}|^2 I(|\chi_{nj}| \geq \epsilon')]$$

$$\lesssim \sum_{j=1}^n E_\theta[|r_n g_{nj}(\theta)|^2 I(|r_n g_{nj}(\theta)| + |r_n E_\theta[g_{nj}(\theta)]| \geq \epsilon')] + \sum_{j=1}^n |r_n E_\theta[g_{nj}(\theta)]|^2$$

$$\lesssim \sum_{j=1}^n E_\theta[|r n g_{nj}(\theta)|^2 I(|r_n g_{nj}(\theta)| \geq \epsilon)] + |\sqrt{n} r_n E_\theta[g_{n1}(\theta)]|^2 \to 0$$

by virtue of (4.3). The proof is complete. $\qquad\square$

Thus we have seen that the desired property (2.9) can be derived under (4.3), (4.4), (4.5), and (4.6). Nevertheless, it is convenient to replace (4.4) by an alternative, which is easier to verify. We prepare the following lemma.

Lemma 4.2. The condition (4.4) holds true if

$$\sum_{j=1}^n \sup_{\{\rho \in \Theta : |r_n^{-1}(\rho - \theta)| \leq a\}} E_\rho\left[|r_n \partial_\theta[g_{nj}(\rho)^\top] r_n|^2 + |r_n g_{nj}(\rho)|^4\right] \to 0. \tag{4.7}$$

for any $a > 0$.

Proof. Write $\sum_{j=1}^n e_{nj}(\theta)$ for the left-hand side of (4.4), and let

$$H_n(x; \theta) := \{p_{\Delta_n}(x; \theta_n)^{1/2} - p_{\Delta_n}(x; \theta)^{1/2} - (\theta_n - \theta)^\top \partial_\theta[p_{\Delta_n}(x; \theta)^{1/2}]\}^2.$$

Then

$$e_{nj}(\theta) = E_\theta\left[p_{\Delta_n}(x_j; \theta)^{-1} H_n(x_j; \theta)\right] = \int_\mathbb{R} H_n(x; \theta)\mathrm{d}x. \tag{4.8}$$

We note that

$$\left|r \partial_\theta^2[f(\theta)^{1/2}] r\right|^2 \lesssim f(\theta)\{|r[\partial_\theta^2 \log f(\theta)] r|^2 + |r \partial_\theta \log f(\theta)|^4\} \tag{4.9}$$

for any nonnegative $\mathcal{C}^2(\Theta)$ function f and diagonal $p \times p$ matrix r: indeed, using the identities $\partial_{\theta_k} f(\theta) = f(\theta)\partial_{\theta_k} \log f(\theta)$ and $\partial_{\theta_k}\partial_{\theta_l} f(\theta) = f(\theta)\{\partial_{\theta_k}\partial_{\theta_l} \log f(\theta) + (\partial_{\theta_k} \log f(\theta))(\partial_{\theta_l} \log f(\theta))\}$, which are valid for any

$k, l \in \{1, \ldots, p\}$, we get

$$\partial_{\theta_k} \partial_{\theta_l} [f(\theta)^{1/2}] = 2^{-1} f(\theta)^{-1} \{ f(\theta)^{1/2} \partial_{\theta_k} \partial_{\theta_l} f(\theta) - 2^{-1} f(\theta)^{-1/2} (\partial_{\theta_k} f(\theta))(\partial_{\theta_l} f(\theta)) \}$$
$$= f(\theta)^{1/2} \{ 2^{-1} \partial_{\theta_k} \partial_{\theta_l} \log f(\theta) + 4^{-1} (\partial_{\theta_k} \log f(\theta))(\partial_{\theta_l} \log f(\theta)) \},$$

from which the inequality (4.9) readily follows. Using the last inequality, we get for each x

$$H_n(x; \theta) \lesssim \int_0^1 \left| r_n \partial_\theta^2 [p_{\Delta_n}(x; \theta + s r_n h)^{1/2}] r_n \right|^2 \mathrm{d}s$$
$$\lesssim \int_0^1 \left\{ \left| r_n \partial_\theta^2 \log p_{\Delta_n}(x; \rho_n'(s)) r_n \right|^2 + \left| r_n \partial_\theta \log p_{\Delta_n}(x; \rho_n'(s)) \right|^4 \right\} p_{\Delta_n}(x; \rho_n'(s)) \mathrm{d}s, \qquad (4.10)$$

where we wrote $\rho_n'(s) = \theta + s r_n h$, which belongs to Θ for every n large enough. Now, by substituting (4.10) in (4.8) and then applying Fubini's theorem for interchanging the $\mathrm{d}s$ and $\mathrm{d}x$ integrals, we have

$$\sum_{j=1}^n e_{nj}(\theta) \lesssim \sum_{j=1}^n \int_0^1 \int_{\mathbb{R}} \left\{ \left| r_n \partial_\theta^2 \log p_{\Delta_n}(x; \rho_n'(s)) r_n \right|^2 \right.$$
$$\left. + \left| r_n \partial_\theta \log p_{\Delta_n}(x; \rho_n'(s)) \right|^4 \right\} p_{\Delta_n}(x; \rho_n'(s)) \mathrm{d}x \mathrm{d}s$$
$$\leq \sum_{j=1}^n \sup_{\{\rho \in \Theta : |r_n^{-1}(\rho - \theta)| \leq |h|\}} \int_{\mathbb{R}} \left\{ \left| r_n \partial_\theta^2 \log p_{\Delta_n}(x; \rho) r_n \right|^2 \right.$$
$$\left. + \left| r_n \partial_\theta \log p_{\Delta_n}(x; \rho) \right|^4 \right\} p_{\Delta_n}(x; \rho) \mathrm{d}x \to 0$$

by means of (4.7); recall that $h \in \mathbb{R}^p$ here is fixed arbitrarily. This completes the proof. $\qquad \square$

To sum up we have derived the following proposition, which serves as our basic tool for proving LAN.

Proposition 4.3. Suppose that (4.3), (4.5), (4.6), and (4.7) hold true. Then we have (2.9), that is, LAN holds true at each θ with rate r_n and the Fisher information matrix $\mathcal{I}(\theta)$.

Remark 4.4. Of course, the concept of LAN is defined for much more general statistical experiments than ours, such as a discrete-time sample from an ergodic process. Also in such cases we can provide a set of conditions analogous to the ones in Proposition 4.3, as a useful tool for proving the LAN.

4.2. Preliminary lemmas

We are here going to check that, in our framework, the conditions (4.3), (4.5), (4.6), and (4.7) hold and that $\mathcal{I}(\theta)$ is positively definite for each $\theta \in \Theta$.

For later use, we prepare some lemmas. We consistently use the notation (2.8). For $j \leq n$, we introduce

$$\epsilon_{nj} = \epsilon_{nj}(\delta, \mu) := \frac{x_j - \mu \Delta_n}{\delta \Delta_n}.$$

Clearly we have $\mathcal{L}(\epsilon_{nj}) = \mathcal{L}(\epsilon_{n1})$ for each $n \in \mathbb{N}$ and $j \leq n$.

Lemma 4.5. It holds that for each $n \in \mathbb{N}$, $\mathcal{L}(\epsilon_{n1}) = NIG(\alpha \delta \Delta_n, \beta \delta \Delta_n, 1, 0)$.

Proof. Obvious from (2.6) with taking $a = (\delta \Delta_n)^{-1}$. $\qquad \square$

An important point in our study is that the normalized increments of X in small time can be approximated by the Cauchy distribution having Blumenthal–Getoor index 1. In what follows, let

$$\phi_1(y) := \frac{1}{\pi(1 + y^2)},$$

the standard symmetric Cauchy density corresponding to the characteristic function $u \mapsto \exp(-|u|)$.

Lemma 4.6. Denote by $f_{\Delta_n} : \mathbb{R} \to (0, \infty)$ the smooth density of $\mathcal{L}(\epsilon_{n1})$. For any nonnegative integer k, we have

$$\lim_{\Delta_n \to 0} \sup_{y \in \mathbb{R}} \left| \partial_y^k f_{\Delta_n}(y) - \partial_y^k \phi_1(y) \right| = 0.$$

Proof. In view of (2.5), we have $\varphi_{\epsilon_{n1}}(u) = \exp\{\delta\Delta_n\sqrt{m} - \sqrt{(\alpha\delta\Delta_n)^2 - (iu + \beta\delta\Delta_n)^2}\}$. Here and in what follows, we write

$$m = \alpha^2 - \beta^2 > 0.$$

Simple manipulation of complex numbers gives

$$\varphi_{\epsilon_{n1}}(u) = e^{\delta\Delta_n\sqrt{m}} \exp\left\{-\sqrt{\frac{1}{2}(A + \sqrt{A^2 + B^2})} - \frac{iB}{\sqrt{2(A + \sqrt{A^2 + B^2})}}\right\}, \qquad (4.11)$$

where we wrote $A := (\delta\Delta_n)^2 m + u^2$ ($A > 0$ always) and $B := -2\beta\delta\Delta_n u$ for brevity. Hence, we clearly have $\varphi_{\epsilon_{n1}}(u) \to \exp(-|u|)$ for each $u \in \mathbb{R}$. The expression (4.11) also leads to the estimate

$$\left|\varphi_{\epsilon_{n1}}(u)\right| \lesssim \exp\left\{-\sqrt{\frac{1}{2}(A + \sqrt{A^2 + B^2})}\right\} \le e^{-\sqrt{A}} \le e^{-|u|}. \qquad (4.12)$$

By means of the Fourier inversion formula we have

$$\sup_{y \in \mathbb{R}} \left|\partial_y^k f_{\Delta_n}(y) - \partial_y^k \phi_1(y)\right| \lesssim \int |u|^k \left|\varphi_{\epsilon_{n1}}(u) - e^{-|u|}\right| du. \qquad (4.13)$$

Under (4.12), we can apply the dominated convergence theorem to the upper bound of (4.13). This completes the proof. $\qquad\square$

In particular, note that the limit of $\mathcal{L}(\epsilon_{n1})$ in total variation is symmetric even if $\beta \neq 0$. As a matter of fact, since the Lévy density g of $NIG(\alpha, \beta, \delta, \mu)$ admits the expansion $z^2 g(z) = (1/\pi) + (\delta\beta/\pi)|z| + o(|z|)$ as $|z| \to 0$ (see (2.7) together with (4.15) below, or, more generally, Raible [20], Prop. 2.18), Lemma 4.6 can also be deduced from the behavior of the Lévy density of $\mathcal{L}(\epsilon_{n1})$ around the origin; note that the standard Cauchy Lévy density equals $z \mapsto (1/\pi)|z|^{-2}$.

We introduce the following functions:

$$\begin{aligned}
\eta(y) &:= \phi_1'(y)/\phi_1(y), \quad y \in \mathbb{R}, \\
\zeta(y) &:= K_1'(y)/K_1(y), \quad y \in [0, \infty), \\
H(y) &:= y^{-1}\{1 + y\zeta(y)\} = -K_0(y)/K_1(y), \quad y \in [0, \infty), \qquad (4.14)
\end{aligned}$$

where we used the identity $K_w'(y) = -K_{w-1}(y) - (w/y)K_w(y)$ for (4.14). The functions ζ, H, and their derivatives are to be defined at $y = 0$ as limits from the right.

Lemma 4.7. (a) The functions $y \mapsto \eta(y)$, $y\eta(y)$, and $y^2\eta'(y)$ are bounded in $\mathbb{R}$.

(b) $y \mapsto H(y)$ is bounded and continuous in $[0, \infty)$. Moreover, $H(y) \sim -y\log(1/y)$ as $y \to 0$ and $H(y) = -1 + 1/(2y) + 3/(8y^2) + O(y^{-3})$ as $y \to \infty$.

(c) $H'(y) \sim -\log(1/y)$ as $y \to 0$ and $y^2 H'(y) = -1/2 + O(y^{-1})$ as $y \to \infty$. In particular, $y \mapsto yH'(y)$ is bounded and continuous in $[0, \infty)$.

Proof. The claim (a) readily follows from the well known fact

$$\sup_{y \in \mathbb{R}} \frac{|y|^k |\partial_y^k \phi_1(y)|}{\phi_1(y)} < \infty$$

for each $k \in \mathbb{Z}_+$; this is valid too for ϕ_1 replaced by the a general symmetric non-Gaussian β-stable density.

As for (b), the continuity of H is clear. We first note the asymptotic behaviors:

$$K_w(y) \sim \begin{cases} \log(1/y) + \log 2 - \mathfrak{C} & \text{if } w = 0, \\ \Gamma(|w|) 2^{|w|-1} y^{-|w|} & \text{if } w \neq 0, \end{cases} \quad \text{as } y \to 0, \tag{4.15}$$

$$K_w(y) = \sqrt{\frac{\pi}{2y}} e^{-y} \left\{ 1 + \frac{\kappa - 1}{8y} + \frac{(\kappa-1)(\kappa-9)}{(8y)^2 2!} + O(y^{-3}) \right\} \quad \text{as } y \to \infty, \tag{4.16}$$

where $\mathfrak{C} \approx 0.5772$ is the Euler–Mascheroni constant and $\kappa := 4w^2$. The desired behavior of $H(y)$ as $y \to 0$ is trivial from (4.15). Next, by applying (4.16) for $w = 0, 1$ and then expanding the fraction $-K_0(y)/K_1(y)$ as a power series of y^{-1}, straightforward computations lead to the desired behavior of $H(y)$ as $y \to \infty$. Now the boundedness of H is trivial.

We turn to (c). Using the known identity $K_w(y) = K_{-w}(y)$ valid for each $w, y > 0$ and also the identity $K_w'(y) = -K_{w-1}(y) - (w/y)K_w(y)$ once again, we get

$$H'(y) = 1 + H(y)/y - \{H(y)\}^2,$$

and so $y^2 H'(y) = y^2 + yH(y) - y^2\{H(y)\}^2$. These expressions combined with (b) lead to the claims. $\qquad\square$

Now we define

$$q_{nj} = q_{nj}(\alpha, \delta, \mu) := \alpha\delta\Delta_n(1 + \epsilon_{nj}^2)^{1/2}$$

and

$$A_k(\theta) := (-1)^k \frac{\alpha\delta}{\pi} \int_0^\infty \left(e^{(\beta/\alpha)y} + e^{-(\beta/\alpha)y} \right) y^{k-1} K_1(y) \left\{ \frac{K_0(y)}{K_1(y)} \right\}^k dy,$$

$$A'(\theta) := -\frac{1}{\pi} \int_0^\infty \left(e^{(\beta/\alpha)y} + e^{-(\beta/\alpha)y} \right) y K_0(y) dy.$$

We need the following lemmas to specify the Fisher information matrix $\mathcal{I}(\theta)$, and to estimate the remainder term in the stochastic expansion of the likelihood ratio random fields (see (4.21) below).

Lemma 4.8. For any $k \in \mathbb{N}$ we have

$$\lim_{n \to \infty} \frac{1}{\Delta_n} E_\theta \left[\{q_{n1} H(q_{n1})\}^k \right] = A_k(\theta),$$

with $A_k(\theta)$ being finite. In particular, $\limsup_{n \to \infty} \Delta_n^{-1} E_\theta[|q_{n1} H(q_{n1})|] < \infty$.

Proof. Reminding Lemma 4.5 and (4.14), we have

$$
\frac{1}{\Delta_n} E_\theta \left[\{ q_{n1} H(q_{n1}) \}^k \right]
$$

$$
= \frac{1}{\Delta_n} \int_{\mathbb{R}} \left\{ \alpha \delta \Delta_n \sqrt{1+x^2} H\left(\alpha \delta \Delta_n \sqrt{1+x^2} \right) \right\}^k \frac{\alpha \delta \Delta_n}{\pi} e^{\delta \Delta_n \sqrt{m} + \beta \delta \Delta_n x} \frac{K_1\left(\alpha \delta \Delta_n \sqrt{1+x^2} \right)}{\sqrt{1+x^2}} \mathrm{d}x
$$

$$
= (-1)^k \frac{\alpha \delta}{\pi} e^{\delta \Delta_n \sqrt{m}} \left[\alpha \delta \Delta_n \int_{\mathbb{R}} e^{\beta \delta \Delta_n x} \left(\alpha \delta \Delta_n \sqrt{1+x^2} \right)^{k-1} \left\{ \frac{K_0(\alpha \delta \Delta_n \sqrt{1+x^2})}{K_1(\alpha \delta \Delta_n \sqrt{1+x^2})} \right\}^k K_1(\alpha \delta \Delta_n \sqrt{1+x^2}) \mathrm{d}x \right]
$$

$$
=: (-1)^k \frac{\alpha \delta}{\pi} e^{\delta \Delta_n \sqrt{m}} B_{\Delta_n}^{(k)} \sim (-1)^k \frac{\alpha \delta}{\pi} B_{\Delta_n}^{(k)}. \tag{4.17}
$$

Write $B_{\Delta_n}^{(k)} = \int_0^\infty + \int_{-\infty}^0 =: B_{\Delta_n}^{(k)+} + B_{\Delta_n}^{(k)-}$.

First let us look at $B_{\Delta_n}^{(k)+}$. The change of variables $y = \alpha \delta \Delta_n (\sqrt{1+x^2} - 1)$ leads to $B_{\Delta_n}^{(k)+} = \int_0^\infty b_{\Delta_n}^{(k)+}(y) \mathrm{d}y$, where

$$
b_{\Delta_n}^{(k)+}(y) = e^{(\beta/\alpha)\sqrt{y}\sqrt{y+2\alpha\delta\Delta_n}} \frac{(y + \alpha\delta\Delta_n)^k}{\sqrt{y}\sqrt{y+2\alpha\delta\Delta_n}} \left\{ \frac{K_0(y + \alpha\delta\Delta_n)}{K_1(y + \alpha\delta\Delta_n)} \right\}^k K_1(y + \alpha\delta\Delta_n).
$$

Obviously, for each $y \in (0, \infty)$

$$
b_{\Delta_n}^{(k)+}(y) \to e^{(\beta/\alpha)y} y^{k-1} K_1(y) \left\{ \frac{K_0(y)}{K_1(y)} \right\}^k =: b_0^{(k)+}(y). \tag{4.18}
$$

In order to apply the dominated convergence theorem, we have to look at the behaviors of $b_{\Delta_n}^{(k)+}(y)$ as $y \to 0$ and $y \to \infty$ uniformly in small Δ_n, say $\Delta_n \in (0, 1]$. First, by means of Lemma 4.7(b) and (4.16), we can derive as $y \to \infty$

$$
\sup_{\Delta_n \leq 1} |b_{\Delta_n}^{(k)+}(y)| \lesssim e^{(\beta/\alpha)y} y^{-1/2} (y + \alpha\delta\Delta_n)^{k-1/2} K_1(y + \alpha\delta\Delta_n)
$$

$$
\lesssim e^{-(1-\beta/\alpha)y} y^{k-3/2}, \tag{4.19}
$$

the upper bound being Lebesgue integrable at infinity; here the assumption $|\beta| < \alpha$ comes into effect. On account of (4.15) and Lemma 4.7(b), it holds that $y^{k-1/2}\{K_0(y)/K_1(y)\}^k K_1(y) \sim Cy^{2k-3/2}\{\log(1/y)\}^k \to 0$ as $y \to 0$. This leads to $\sup_{y \in (0,1]} y^{k-1/2}\{K_0(y)/K_1(y)\}^k K_1(y) < \infty$, so that, as $y \to 0$

$$
\sup_{\Delta_n \leq 1} |b_{\Delta_n}^{(k)+}(y)| \lesssim y^{-1/2} \sup_{\Delta_n \leq 1} \left[(y + \alpha\delta\Delta_n)^{k-1/2} \left\{ \frac{K_0(y + \alpha\delta\Delta_n)}{K_1(y + \alpha\delta\Delta_n)} \right\}^k K_1(y + \alpha\delta\Delta_n) \right]
$$

$$
\lesssim y^{-1/2}, \tag{4.20}
$$

the upper bound being Lebesgue integrable near the origin. Having (4.18), (4.19) and (4.20) in hand, the dominated convergence theorem yields that $B_{\Delta_n}^{(k)+} \to \int_0^\infty b_0^{(k)+}(y) \mathrm{d}y < \infty$.

Let $b_0^{(k)-}(y) := e^{-(\beta/\alpha)y} y^{k-1} K_1(y) \{K_0(y)/K_1(y)\}^k$. In the same manner as before, we can deduce that $B_{\Delta_n}^{(k)-} \to \int_0^\infty b_0^{(k)-}(y) \mathrm{d}y$. Thus $B_{\Delta_n}^{(k)} \to \int_0^\infty \{b_0^{(k)+}(y) + b_0^{(k)-}(y)\} \mathrm{d}y$, completing the proof of the first half of the claims. The last half is obvious from (4.17) and what we have seen above. The proof is complete. $\square$

Lemma 4.9. It holds that

$$\lim_{n\to\infty} E_\theta[\epsilon_{n1} q_{n1} H(q_{n1})] = A'(\theta),$$

with $A'(\theta)$ being finite.

Proof. The lemma can be deduced in an analogous way to the proof of Lemma 4.8, so we omit the details. □

The Lemma 4.10 below provides the fully closed form of $A'(\theta)$, which directly leads to the closed form of $\mathcal{I}_{12}(\theta)$, as we will see in Section 4.5. We set aside the integral form of $A'(\theta)$ for later convenience in the proof of positive definiteness of $\mathcal{I}(\theta)$.

Lemma 4.10. It holds that

$$A'(\theta) = -\frac{2\alpha^2}{\pi m}\left\{1 + \frac{\beta}{\sqrt{m}}\arctan\left(\frac{\beta}{\sqrt{m}}\right)\right\}.$$

Proof. We note that $(x+1/x)/2 \geq 1$ for any $x \geq 0$, and that, for any $|b| < 1$,

$$\frac{1}{2}\int_0^\infty \frac{1}{x}\left\{\frac{1}{2}\left(x+\frac{1}{x}\right)+b\right\}^{-2}\mathrm{d}x = \frac{1}{1-b^2}\left[\frac{-(bx+1)}{x^2+2bx+1} - \frac{b}{\sqrt{1-b^2}}\arctan\left(\frac{x+b}{\sqrt{1-b^2}}\right)\right]_{x=0}^{x=\infty}$$

$$= \frac{1}{1-b^2}\left[1 - \frac{b}{\sqrt{1-b^2}}\left\{\frac{\pi}{2} - \arctan\left(\frac{b}{\sqrt{1-b^2}}\right)\right\}\right].$$

In view of the definition (2.3) and Fubini's theorem, we see that the leftmost side of the above display equals $\int_0^\infty \mathrm{e}^{-by} y K_0(y)\mathrm{d}y$. Using this fact with $b = \beta/\alpha$ and $-\beta/\alpha$, it is straightforward to deduce the claim. □

4.3. Likelihood, score, and observed information in question

Now, let us look at the log likelihood function $\ell_n(\theta)$ associated with a sample $(X_{t_j})_{j\leq n}$. By (2.2), the density of $\mathcal{L}(x_j)$ under P_θ exists. We use the notation consistent with Sections 2.3 and 4.1:

$$p_{\Delta_n}(x_j;\theta) = \frac{\alpha\delta\Delta_n}{\pi}\exp\{\delta\Delta_n\sqrt{m} + \beta(x_j - \mu\Delta_n)\}\frac{K_1\left(\alpha\sqrt{(\delta\Delta_n)^2 + (x_j - \mu\Delta_n)^2}\right)}{\sqrt{(\delta\Delta_n)^2 + (x_j - \mu\Delta_n)^2}}.$$

Then, we can write down $\ell_n(\theta)$ in terms of $(\epsilon_{nj})_{j\leq n}$ as

$$\ell_n(\theta) = \sum_{j=1}^n \left\{\log\alpha + \delta\Delta_n(\sqrt{m} + \beta\epsilon_{nj}) + \log\phi_1(\epsilon_{nj}) + \frac{1}{2}\log(1+\epsilon_{nj}^2) + \log K_1\left(\alpha\delta\Delta_n\sqrt{1+\epsilon_{nj}^2}\right)\right\}. \quad (4.21)$$

The expression (4.21) may look unnecessarily lengthy, as the term $\log\phi_1(\epsilon_{nj}) + 2^{-1}\log(1+\epsilon_{nj}^2)$ can be obviously simplified. However, we have meaningly transformed it as just described. In fact, the introduction of the standard Cauchy density ϕ_1 above turns out to be convenient in the process of deriving various limiting values as well as deducing estimates of stochastically small terms, to which Lemma 4.7 to 4.9 can effectively be applied.

For studying LAN, we need to look at the score $\theta \mapsto \partial_\theta \ell_n(\theta)$ and the observed information $\theta \mapsto -\partial_\theta^2 \ell_n(\theta)$. Note that $\partial_\mu \epsilon_{nj} = -\delta^{-1}$, $\partial_\mu^2 \epsilon_{nj} = 0$, $\partial_\delta \epsilon_{nj} = -\delta^{-1}\epsilon_{nj}$, and $\partial_\delta^2 \epsilon_{nj} = 2\delta^{-2}\epsilon_{nj}$, and $\partial_\delta\partial_\mu\epsilon_{nj} = \delta^{-2}$. In terms

of (4.21), the first-order partial derivative of $\theta \mapsto \ell_n(\theta)$ are explicitly given as follows:

$$\partial_\alpha \ell_n(\theta) = \sum_{j=1}^n \left\{ \frac{\alpha \delta \Delta_n}{\sqrt{m}} + \frac{1}{\alpha} q_{nj} H(q_{nj}) \right\}, \tag{4.22}$$

$$\partial_\beta \ell_n(\theta) = \sum_{j=1}^n \left\{ \delta \Delta_n \left(\epsilon_{nj} - \frac{\beta}{\sqrt{m}} \right) \right\}, \tag{4.23}$$

$$\partial_\delta \ell_n(\theta) = \sum_{j=1}^n \left\{ -\frac{1}{\delta}(\epsilon_{nj}\eta(\epsilon_{nj}) + 1) + \Delta_n \left(\sqrt{m} + \frac{\alpha}{\sqrt{1 + \epsilon_{nj}^2}} H(q_{nj}) \right) \right\}, \tag{4.24}$$

$$\partial_\mu \ell_n(\theta) = \sum_{j=1}^n \left\{ -\frac{1}{\delta}\eta(\epsilon_{nj}) - \Delta_n \left(\beta + \frac{\alpha \epsilon_{nj}}{\sqrt{1 + \epsilon_{nj}^2}} H(q_{nj}) \right) \right\}. \tag{4.25}$$

We also need to look at the Hessian matrix $\partial_\theta^2 \ell_n(\theta)$: the diagonal elements are

$$\partial_\alpha^2 \ell_n(\theta) = \sum_{j=1}^n \left\{ -\frac{\beta^2 \delta \Delta_n}{m^{3/2}} + \frac{q_{nj}^2}{\alpha^2} H'(q_{nj}) \right\}, \tag{4.26}$$

$$\partial_\beta^2 \ell_n(\theta) = \sum_{j=1}^n \left\{ -\frac{\alpha^2 \delta \Delta_n}{m^{3/2}} \right\} = -\frac{\alpha^2 \delta n \Delta_n}{m^{3/2}}, \tag{4.27}$$

$$\partial_\delta^2 \ell_n(\theta) = \sum_{j=1}^n \left\{ \frac{1}{\delta^2}(1 + 2\epsilon_{nj}\eta(\epsilon_{nj}) + \epsilon_{nj}^2 \eta'(\epsilon_{nj})) + \frac{\alpha \Delta_n}{\delta} \left(\frac{q_{nj}H'(q_{nj})}{(1 + \epsilon_{nj}^2)^{3/2}} + \frac{\epsilon_{nj}^2 H(q_{nj})}{(1 + \epsilon_{nj}^2)^{3/2}} \right) \right\}, \tag{4.28}$$

$$\partial_\mu^2 \ell_n(\theta) = \sum_{j=1}^n \left\{ \frac{1}{\delta^2}\eta'(\epsilon_{nj}) + \frac{\alpha \Delta_n}{\delta} \left(\frac{\epsilon_{nj}^2 q_{nj} H'(q_{nj})}{(1 + \epsilon_{nj}^2)^{3/2}} + \frac{H(q_{nj})}{(1 + \epsilon_{nj}^2)^{3/2}} \right) \right\}, \tag{4.29}$$

and the off-diagonal ones are

$$\partial_\alpha \partial_\beta \ell_n(\theta) = \sum_{j=1}^n \frac{\alpha \beta \delta \Delta_n}{m^{3/2}} = \frac{\alpha \beta \delta n \Delta_n}{m^{3/2}}, \tag{4.30}$$

$$\partial_\alpha \partial_\delta \ell_n(\theta) = \sum_{j=1}^n \left\{ \frac{\alpha \Delta_n}{\sqrt{m}} + \frac{\Delta_n}{\sqrt{1 + \epsilon_{nj}^2}}(H(q_{nj}) + q_{nj}H'(q_{nj})) \right\}, \tag{4.31}$$

$$\partial_\alpha \partial_\mu \ell_n(\theta) = \sum_{j=1}^n \left\{ -\frac{\Delta_n \epsilon_{nj}}{\sqrt{1 + \epsilon_{nj}^2}}(H(q_{nj}) + q_{nj}H'(q_{nj})) \right\}, \tag{4.32}$$

$$\partial_\beta \partial_\delta \ell_n(\theta) = \sum_{j=1}^n \left\{ -\frac{\beta \Delta_n}{\sqrt{m}} \right\} = -\frac{\beta n \Delta_n}{\sqrt{m}}, \tag{4.33}$$

$$\partial_\beta \partial_\mu \ell_n(\theta) = \sum_{j=1}^n \{ -\Delta_n \} = -n\Delta_n, \tag{4.34}$$

$$\partial_\delta \partial_\mu \ell_n(\theta) = \sum_{j=1}^n \left\{ \frac{1}{\delta^2}(\eta(\epsilon_{nj}) + \epsilon_{nj}\eta'(\epsilon_{nj})) - \frac{\alpha \Delta_n}{\delta} \frac{\epsilon_{nj}}{(1 + \epsilon_{nj}^2)^{3/2}}(q_{nj}H'(q_{nj}) - H(q_{nj})) \right\}. \tag{4.35}$$

In what follows, we complete the proof of Theorem 3.1 by verifying the conditions (4.3), (4.5), (4.6), and (4.7) given in Section 2.3, with taking $r_n = (r_{kn})_{k=1}^4$ as in (3.3).

4.4. Lindeberg condition

First we look at (4.3). As is well known, (4.3) is implied by the Lyapunov condition: there exists a constant $\epsilon' > 0$ such that $\sum_{j=1}^n E_\theta[|r_n g_{nj}(\theta)|^{2+\epsilon'}] \to 0$ (e.g., Shiryaev [22]). Here, we set $\epsilon' = 2$ and prove

$$\sum_{k=1}^4 \sum_{j=1}^n E_\theta[|r_{kn} g_{k,nj}(\theta)|^4] \to 0, \tag{4.36}$$

where $g_{k,nj}(\theta)$ denotes the kth component of $g_{nj}(\theta)$. Using Lemma 4.8, we get

$$\sum_{j=1}^n E_\theta[|r_{1n} g_{1,nj}(\theta)|^4] \lesssim \frac{1}{n\Delta_n} \cdot \frac{1}{n} \sum_{j=1}^n \frac{1}{\Delta_n} \{\Delta_n^4 + E_\theta[\{q_{nj} H(q_{nj})\}^4]\} \lesssim \frac{1}{n\Delta_n} \to 0.$$

Next, we note that $E_\theta[|x_j|^q] \lesssim \Delta_n$ for every $q \geq 2$: indeed, since $E_\theta[|X_{\Delta_n} - E_\theta[X_{\Delta_n}]|^q] \lesssim \Delta_n$ (Asmussen and Rosiński [4], Lem. 3.1) and $E_\theta[X_{\Delta_n}] = \Delta_n E_\theta[X_1]$, it follows that $E_\theta[|x_j|^q] \lesssim E_\theta[|X_{\Delta_n} - E_\theta[X_{\Delta_n}]|^q] + |E_\theta[X_{\Delta_n}]|^q \lesssim \Delta_n + \Delta_n^q \lesssim \Delta_n$. Accordingly, noting that $r_{2n} g_{2,nj}(\theta) = (n\Delta_n)^{-1/2}(x_j - \mu\Delta_n - \beta\delta\Delta_n/\sqrt{m})$, we get

$$\sum_{j=1}^n E_\theta[|r_{2n} g_{2,nj}(\theta)|^4] \lesssim \frac{1}{n\Delta_n} \cdot \frac{1}{n} \sum_{j=1}^n \{\Delta_n^{-1} E_\theta[|x_j|^4] + \Delta_n^3\} \lesssim \frac{1}{n\Delta_n} \to 0.$$

In view of the fact $\sup_{y\in\mathbb{R}} |\eta(y)| \vee |y\eta(y)| < \infty$ and $\sup_{y\geq 0} |H(y)| < \infty$ (see Lems. 4.7(a) and 4.7(b)) and the expressions (4.24) and (4.25), we see that $|r_{3n} g_{3,nj}(\theta)|^4 + |r_{4n} g_{4,nj}(\theta)|^4 \lesssim (1 + \Delta_n)/n^2 \lesssim 1/n^2$ (P_θ-a.s.). Therefore, $\sum_{j=1}^n E_\theta[|r_{3n} g_{3,nj}(\theta)|^4] + \sum_{j=1}^n E_\theta[|r_{4n} g_{4,nj}(\theta)|^4] \lesssim 1/n \to 0$. Thus (4.36), hence (4.3), has been obtained.

For later use, we note the stronger convergence

$$\sum_{j=1}^n \sup_{\theta\in\Theta} E_\theta[|r_n g_{nj}(\theta)|^4] \to 0, \tag{4.37}$$

which directly follows from (4.36) and the boundedness of Θ.

4.5. Fisher information matrix

Next we look at (4.5) and the positive definiteness of the Fisher information matrix $\mathcal{I}(\theta)$.

First we prove (4.5), which amounts to proving that

$$\mathcal{I}_{kl}(\theta) = \lim_{n\to\infty} \sum_{j=1}^n E_\theta[r_{kn} g_{k,nj}(\theta) r_{ln} g_{l,nj}(\theta)], \quad 1 \leq k \leq l \leq 4.$$

Prior to computing the limits, let us recall the expressions (4.22) to (4.25), and the notation $A_k(\theta)$ in Lemma 4.8.

We begin with the diagonal elements. First, we observe that

$$\sum_{j=1}^{n} E_\theta[\{r_{1n}g_{1,nj}(\theta)\}^2] = O(\Delta_n) + \frac{2\delta}{\sqrt{m}}\frac{1}{n}\sum_{j=1}^{n} E_\theta[q_{nj}H(q_{nj})] + \frac{1}{\alpha^2}\frac{1}{n}\sum_{j=1}^{n}\frac{1}{\Delta_n}E_\theta[\{q_{nj}H(q_{nj})\}^2]$$

$$= O(\Delta_n) + \frac{2\delta\Delta_n}{\sqrt{m}}\frac{1}{\Delta_n}E_\theta[q_{n1}H(q_{n1})] + \frac{1}{\alpha^2}\frac{1}{\Delta_n}E_\theta[\{q_{n1}H(q_{n1})\}^2]$$

$$\rightarrow \frac{1}{\alpha^2}A_2(\theta) = \mathcal{I}_{11}(\theta).$$

It follows from Lemma 4.5 that $E_\theta[\epsilon_{nj}] = \beta/\sqrt{m}$ and $E_\theta[(\epsilon_{nj} - \beta/\sqrt{m})^2] = (\delta\Delta_n)^{-1}\alpha^2/m^{3/2}$, so that

$$\sum_{j=1}^{n} E_\theta[\{r_{2n}g_{2,nj}(\theta)\}^2] = \delta\alpha^2/m^{3/2} = \mathcal{I}_{22}(\theta). \tag{4.38}$$

According to the boundedness of H (Lem. 4.7(b)),

$$\sum_{j=1}^{n} E_\theta[\{r_{3n}g_{3,nj}(\theta)\}^2] = \frac{1}{\delta^2 n}\sum_{j=1}^{n} E_\theta[\{\epsilon_{nj}\eta(\epsilon_{nj}) + 1\}^2] + O(\Delta_n) = \frac{1}{\delta^2}E_\theta[\{\epsilon_{n1}\eta(\epsilon_{n1}) + 1\}^2] + O(\Delta_n).$$

Building on Lemmas 4.6 and 4.7(a), we can apply the bounded convergence theorem to the last expectation, so that

$$\sum_{j=1}^{n} E_\theta[\{r_{3n}g_{3,nj}(\theta)\}^2] \rightarrow \frac{1}{\delta^2}\int_\mathbb{R}\left(\frac{y\phi_1'(y)}{\phi_1(y)} + 1\right)^2\phi_1(y)\mathrm{d}y$$

$$= \frac{1}{2\pi\delta^2}\left[\frac{y - y^3}{(1 + y^2)^2} + \arctan y\right]_{y=-\infty}^{y=\infty} = \frac{1}{2\delta^2} = \mathcal{I}_{33}(\theta).$$

In a similar manner, based on the expression (4.25) we can deduce

$$\sum_{j=1}^{n} E_\theta[\{r_{4n}g_{4,nj}(\theta)\}^2] \rightarrow \frac{1}{\delta^2}\int_\mathbb{R}\left(\frac{\phi_1'(y)}{\phi_1(y)}\right)^2\phi_1(y)\mathrm{d}y$$

$$= \frac{1}{2\pi\delta^2}\left[\frac{y^3 - y}{(1 + y^2)^2} + \arctan y\right]_{y=-\infty}^{y=\infty} = \frac{1}{2\delta^2} = \mathcal{I}_{44}(\theta).$$

Now we turn to the off-diagonal elements. First, by means of Lemmas 4.8, 4.9, 4.10, we get

$$\sum_{j=1}^{n} E_\theta[r_{1n}g_{1,nj}(\theta)r_{2n}g_{2,nj}(\theta)] = \delta\frac{1}{n}\sum_{j=1}^{n} E_\theta\left[\left(\epsilon_{nj} - \frac{\beta}{\sqrt{m}}\right)\left(\frac{\alpha\delta\Delta_n}{\sqrt{m}} + \frac{1}{\alpha}q_{nj}H(q_{nj})\right)\right]$$

$$= \frac{\delta}{\alpha}\frac{1}{n}\sum_{j=1}^{n} E_\theta\left[\left(\epsilon_{nj} - \frac{\beta}{\sqrt{m}}\right)q_{nj}H(q_{nj})\right]$$

$$= \frac{\delta}{\alpha}\frac{1}{n}\sum_{j=1}^{n} E_\theta\left[\epsilon_{nj}q_{nj}H(q_{nj})\right] + O(\Delta_n) \rightarrow \frac{\delta}{\alpha}A'(\theta) = \mathcal{I}_{12}(\theta).$$

Next, it follows from Lemma 4.7 that

$$
\begin{aligned}
E_\theta[|r_{2n}g_{2,nj}(\theta)r_{3n}g_{3,nj}(\theta)|] &\leq \frac{1}{n}\left\{\left|E_\theta\left[\sqrt{\Delta_n}\left(\epsilon_{n1}-\frac{\beta}{\sqrt{m}}\right)(1+\epsilon_{n1}\eta(\epsilon_{n1}))\right]\right|\right. \\
&\quad \left.+\delta\Delta_n\left|E_\theta\left[\sqrt{\Delta_n}\left(\epsilon_{n1}-\frac{\beta}{\sqrt{m}}\right)\left(\sqrt{m}+\frac{\alpha}{\sqrt{1+\epsilon_{n1}^2}}H(q_{n1})\right)\right]\right|\right\} \\
&\lesssim \frac{1}{n}E_\theta\left[\left|\sqrt{\Delta_n}\left(\epsilon_{n1}-\frac{\beta}{\sqrt{m}}\right)\right|\right]=:\frac{1}{n}E_\theta[\xi_n].
\end{aligned}
$$

Since $\xi_n = O_{P_\theta}(\sqrt{\Delta_n}) = o_{P_\theta}(1)$ and $\sup_n E_\theta[|\xi_n|^2] = \alpha^2/(\delta m^{3/2}) < \infty$ (recall (4.38)), we deduce that $E_\theta[\xi_n] \to 0$. Thus

$$
\left|\sum_{j=1}^n E_\theta[r_{2n}g_{2,nj}(\theta)r_{3n}g_{3,nj}(\theta)]\right| \lesssim E_\theta[\xi_n] \to 0 = \mathcal{I}_{23}(\theta).
$$

Now let us note that $\int_\mathbb{R} \phi_1'(y)\mathrm{d}y = 0$, and that $\int_\mathbb{R} y\{\phi_1'(y)/\phi_1(y)\}^2\phi_1(y)\mathrm{d}y = 0$ since the integrands are odd and behaves like y^{-3} up to multiplicative constant at infinity. Hence

$$
\sum_{j=1}^n E_\theta[r_{3n}g_{3,nj}(\theta)r_{4n}g_{4,nj}(\theta)] \to \frac{1}{\delta^2}\int_\mathbb{R} \frac{\phi_1'(y)}{\phi_1(y)}\left\{1+y\frac{\phi_1'(y)}{\phi_1(y)}\right\}\phi_1(y)\mathrm{d}y = 0 = \mathcal{I}_{34}(\theta).
$$

The proofs for $\mathcal{I}_{kl}(\theta) = 0$ for the remaining (k,l)s are easier, and we omit them.

Summarizing the above now yields (4.5).

It remains to prove the positive definiteness of $\mathcal{I}(\theta)$ for each $\theta \in \Theta$. In view of the form (3.2), $\mathcal{I}(\theta)$ is positive definite as soon as so is the principal submatrix of second order, say $\mathcal{I}^{\alpha,\beta}(\theta)$. Obviously, $\det[\mathcal{I}^{\alpha,\beta}(\theta)]$ is symmetric as a function of β. Hence, it suffices to prove that, given any $\alpha > 0$, the function $\beta \mapsto \det[\mathcal{I}^{\alpha,\beta}(\theta)] = \mathcal{I}_{11}(\theta)\mathcal{I}_{22}(\theta) - \{\mathcal{I}_{12}(\theta)\}^2$ is positive for $\beta \in [0,\alpha)$. Fix any $\alpha, \delta > 0$ in the sequel. It is convenient to introduce the notation:

$$
C(\beta) = \int_0^\infty (\mathrm{e}^{(\beta/\alpha)y} + \mathrm{e}^{-(\beta/\alpha)y})yK_1(y)\mathrm{d}y,
$$

$$
\Xi(y;\beta) = C(\beta)^{-1}(\mathrm{e}^{(\beta/\alpha)y} + \mathrm{e}^{-(\beta/\alpha)y})yK_1(y), \quad y > 0.
$$

Then $y \mapsto \Xi(y;\beta)$ for each $\beta \in [0,\alpha)$ acts as a probability density function on $(0,\infty)$.

As in the proof of Lemma 4.10, we can derive

$$
\begin{aligned}
C(\beta) &= \frac{1}{2}\int_0^\infty \int_0^\infty y\left(\mathrm{e}^{-\{(x+1/x)/2-\beta/\alpha\}y} + \mathrm{e}^{-\{(x+1/x)/2+\beta/\alpha\}y}\right)\mathrm{d}y\mathrm{d}x \\
&= \frac{1}{2}\left[\int_0^\infty \left\{\frac{1}{2}\left(x+\frac{1}{x}\right)-\frac{\beta}{\alpha}\right\}^{-2}\mathrm{d}x + \int_0^\infty \left\{\frac{1}{2}\left(x+\frac{1}{x}\right)+\frac{\beta}{\alpha}\right\}^{-2}\mathrm{d}x\right] \\
&= \frac{\alpha^3\pi}{m^{3/2}},
\end{aligned}
$$

where we used the identity valid for any $|b| < 1$:

$$\int_0^\infty \left\{ \frac{1}{2}\left(x + \frac{1}{x}\right) + b \right\}^{-2} \mathrm{d}x = \frac{2}{1-b^2}\left[\frac{1}{\sqrt{1-b^2}}\arctan\left(\frac{x+b}{\sqrt{1-b^2}}\right) - \frac{x - 2b^2x - b}{x^2 + 2bx + 1}\right]_{x=0}^{x=\infty}$$
$$= \frac{2}{1-b^2}\left[\frac{1}{\sqrt{1-b^2}}\left\{\frac{\pi}{2} - \arctan\left(\frac{b}{\sqrt{1-b^2}}\right)\right\} - b\right].$$

In particular, we have $C(0) = \pi$. Then, some elementary manipulations and Cauchy–Schwarz's inequality lead to

$$\det[\mathcal{I}^{\alpha,\beta}(\theta)] = \alpha^{-2}A_2(\theta)\cdot\alpha^2\delta m^{-3/2} - \alpha^{-2}\delta^2\{A'(\theta)\}^2$$
$$= \frac{\alpha^4\delta^2}{m^3}\left\{\int_0^\infty \Xi(y;\beta)\left(\frac{K_0(y)}{K_1(y)}\right)^2\mathrm{d}y \cdot \int_0^\infty \Xi(y;\beta)\mathrm{d}y - \left(\int_0^\infty \Xi(y;\beta)\frac{K_0(y)}{K_1(y)}\mathrm{d}y\right)^2\right\}$$
$$> 0,$$

where the last strict inequality does hold true since $y \mapsto K_0(y)/K_1(y)$ is not a constant on $(0,\infty)$. This completes the proof of the positive definiteness of $\mathcal{I}(\theta)$ for each $\theta \in \Theta$.

4.6. Negligibility of the centering

Turning to verification of (4.6), it suffices to see that $r_{kn}E_\theta[g_{k,n1}(\theta)] = o(1/\sqrt{n})$ for each k.

Thanks to Lemma 4.8, we have $r_{1n}E_\theta[g_{1,n1}(\theta)] = n^{-1/2}\{O(\sqrt{\Delta_n}) + E_\theta[q_{n1}H(q_{n1})]\} = O(\sqrt{\Delta_n/n}) = o(1/\sqrt{n})$, and it is obvious that $r_{2n}E_\theta[g_{2,nj}(\theta)] = 0$. It follows from Lemmas 4.6 and 4.7(a) that $E_\theta[\epsilon_{n1}\eta(\epsilon_{n1})] \to \int_\mathbb{R} y\phi_1'(y)\mathrm{d}y = -1$ and $E_\theta[\eta(\epsilon_{n1})] \to \int_\mathbb{R} \phi_1'(y)\mathrm{d}y = 0$. Therefore $r_{3n}E_\theta[g_{3,n1}(\theta)] = n^{-1/2}\{o(1) + O(\Delta_n)\} = o(1/\sqrt{n})$ and similarly, $r_{4n}E_\theta[g_{4,n1}(\theta)] = o(1/\sqrt{n})$. Thus we get (4.6).

4.7. Mean-square differentiability

Finally we verify (4.7). For this it only remains to show that $\sum_{j=1}^n \sup_\theta E_\theta[|r_n\partial_\theta^\top g_{nj}(\theta)r_n|^2] \to 0$ since we already have (4.37). To do this, we recall (3.3), and also the summands of the expressions (4.26) to (4.35) for $\partial_\theta^\top g_{nj}(\theta) = \partial_\theta^2 \log p_{\Delta_n}(x_j;\theta)$. It suffices to estimate

$$B_{kl,n}(\theta) := E_\theta\left[|r_{kn}r_{ln}\partial_{\theta_k}\partial_{\theta_l}\log p_{\Delta_n}(x_j;\theta)|^2\right]$$

for $k,l \in \{1,2,3,4\}$ individually, where we wrote $\theta = (\alpha,\beta,\delta,\mu) =: (\theta_1,\theta_2,\theta_3,\theta_4)$ for convenience.

Invoking the boundedness of $y \mapsto yH'(y)$ (*cf.* Lem. 4.7(c)), we get

$$\sup_{\theta\in\Theta} B_{11,nj}(\theta) \lesssim \frac{1}{n^2}\left\{1 + E_\theta[|q_{nj}/\Delta_n|^2]\right\} \lesssim \frac{1}{n^2}\{1 + E_\theta[\epsilon_{nj}^2]\} \lesssim \frac{1}{n^2\Delta_n} = o\left(\frac{1}{n}\right)$$

as soon as $n\Delta_n \to \infty$, so that $\sum_{j=1}^n \sup_{\theta\in\Theta} B_{11,nj}(\theta) \to 0$ according to the boundedness of Θ. For the others, reminding Lemma 4.7 it is not difficult to deduce that

$$\sup_{\theta\in\Theta} B_{kl,nj}(\theta) \lesssim \left\{\begin{array}{ll} O(1/n^2) & \text{for } k = l \neq 1 \text{ and for } \{k,l\} = \{1,2\} \text{ or } \{3,4\}, \\ O(\Delta_n/n) & \text{for all the rest.} \end{array}\right\} = o\left(\frac{1}{n}\right).$$

Therefore $\sum_{j=1}^n \sup_{\theta\in\Theta} B_{kl,nj}(\theta) \to 0$ for each (k,l), completing the proof of (4.7).

5. Concluding remarks

In this paper, we obtained the LAN for the NIG Lévy process discretely observed at high frequency. The rate in the LAN are of two kind: $\sqrt{n}$ for (δ, μ), while $\sqrt{n\Delta_n}$ for (α, β). Furthermore, the Fisher information matrix $\mathcal{I}(\theta)$ turned out to be always positive-definite. Only the element $\mathcal{I}_{11}(\theta)$ involves an integral, however, given any admissible parameter values, we can evaluate it numerically in a small amount of time.

One of the important future tasks is construction of an estimator $\hat{\theta}_n$ of θ, which is asymptotically optimal in the sense that, in view of Theorem 3.1, the normalized estimator $r_n^{-1}(\hat{\theta}_n - \theta)$ is asymptotically distributed as $\mathcal{N}_4(0, \mathcal{I}(\theta)^{-1})$ under the true measure. The maximum likelihood estimator is the first candidate. Nevertheless, direct simultaneous optimization for the four parameters might entail numerical difficulties, so that some numerical procedures must be of practical help. Among others, for possible practical devices for computing estimates, we refer to EM type algorithm (Karlis [12]) and MCMC based Bayesian estimation (Karlis and Lillestöl [13]). Hence, for example, it would then be more convenient to provide a rate-optimal estimator of θ (initial estimator) at first, and then execute the likelihood based one-step improvement in order to attain the minimal asymptotic variance $\mathcal{I}(\theta)^{-1}$ specified in Theorem 3.1.

It would also be interesting to investigate some related classes of stochastic processes. In particular, the class of Lévy-driven Ornstein-Uhlenbeck processes, of increasing importance in various fields of application, is a challenging object, as in general, it does not admit a transition probability density in closed form. Those issues will be addressed in subsequent papers.

References

[1] M. Abramowitz and I.A. Stegun Eds., *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables.* Reprint of the 1972 edition, Dover Publications, Inc., New York (1992)

[2] Y. Aït-Sahalia and J. Jacod, Fisher's information for discretely sampled Lévy processes. *Econometrica* **76** (2008) 727–761.

[3] M.G. Akritas and R.A. Johnson, Asymptotic inference in Lévy processes of the discontinuous type. *Ann. Stat.* **9** (1981) 604–614.

[4] S. Asmussen and J. Rosiński, Approximations of small jumps of Levy processes with a view towards simulation. *J. Appl. Probab.* **38** (2001) 482–493.

[5] O.E. Barndorff-Nielsen, Exponentially decreasing distributions for the logarithm of particle size. *Proc. R. Soc. Lond.* A **353** (1977) 401–419.

[6] O.E. Barndorff-Nielsen, *Normal inverse Gaussian processes and the modelling of stock returns.* Research report 300, Department of Theoretical Statistics, Institute of Mathematics, University of Aarhus (1995)

[7] O.E. Barndorff-Nielsen, Processes of normal inverse Gaussian type. *Finance Stoch.* **2** (1998) 41–68.

[8] D.R. Cox and N. Reid, Parameter orthogonality and approximate conditional inference. With a discussion. *J. R. Stat. Soc., Ser. B* **49** (1987) 1–39.

[9] J. Jacod, Inference for stochastic processes, in *Handbook of Financial Econometrics*, edited by Y. Aït-Sahalia and L.P. Hansen, Amsterdam, North-Holland (2010)

[10] B. Jørgensen and S.J. Knudsen, Parameter orthogonality and bias adjustment for estimating functions. *Scand. J. Statist.* **31** (2004) 93–114.

[11] O. Kallenberg, *Foundations of Modern Probability.* 2nd edition, Springer-Verlag, New York (2002)

[12] D. Karlis, An EM type algorithm for maximum likelihood estimation of the normal-inverse Gaussian distribution. *Stat. Probab. Lett.* **57** (2002) 43–52.

[13] D. Karlis and J. Lillestöl, Bayesian estimation of NIG models via Markov chain Monte Carlo methods. *Appl. Stoch. Models Bus. Ind.* **20** (2004) 323–338.

[14] R. Kawai and H. Masuda, On the local asymptotic behavior of the likelihood function for Meixner Lévy processes under high-frequency sampling. *Stat. Probab. Lett.* **81** (2011) 460–469.

[15] L. Le Cam, Locally asymptotically normal families of distributions. Certain approximations to families of distributions and their use in the theory of estimation and testing hypotheses. *Univ. California Publ. Stat.* **3** (1960) 37–98.

[16] L. Le Cam and G.L. Yang, *Asymptotics in Statistics. Some Basic Concepts.* 2nd edition, Springer-Verlag, New York (2000)

[17] H. Masuda, Notes on estimating inverse-Gaussian and gamma subordinators under high-frequency sampling. *Ann. Inst. Stat. Math.* **61** (2009) 181–195.

[18] H. Masuda, Joint estimation of discretely observed stable Lévy processes with symmetric Lévy density. *J. Japan Stat. Soc.* **39** (2009) 49–75.

[19] K. Prause, *The Generalized Hyperbolic Model: Estimation, Financial Derivatives, and Risk Measures.* Ph.D. thesis, University of Freiburg (1999). Available at http://www.freidok.uni-freiburg.de/volltexte/15/

[20] S. Raible, *Lévy Processes in Finance: Theory, Numerics, and Empirical Facts.* Ph.D. thesis, University of Freiburg (2000). Available at http://www.freidok.uni-freiburg.de/volltexte/51/

[21] K. Sato, *Lévy Processes and Infinitely Divisible Distributions.* Cambridge University Press, Cambridge (1999)

[22] A.N. Shiryaev, *Probability.* 2nd edition, Springer-Verlag, New York (1996)

[23] H. Strasser, *Mathematical Theory of Statistics. Statistical Experiments and Asymptotic Decision Theory.* Walter de Gruyter & Co., Berlin (1985)

[24] A.W. van der Vaart, *Asymptotic Statistics.* Cambridge University Press, Cambridge (1998)

[25] J.H.C. Woerner, *Statistical Analysis for Discretely Observed Lévy Processes.* Ph.D. thesis, University of Freiburg (2001). Available at http://www.freidok.uni-freiburg.de/volltexte/295/

[26] J.H.C. Woerner, Estimating the skewness in discretely observed Lévy processes. *Econ. Theory* **20** (2004) 927–942.