Saddlepoint Approximation for the Generalized Inverse Gaussian Lévy Process

Mimi Zhang[1,3], Matthew Revie[2] and John Quigley[2]

[1]School of Computer Science and Statistics, Trinity College Dublin, Ireland
[2]Department of Management Science, University of Strathclyde, UK
[3]I-Form Advanced Manufacturing Research Centre, Science Foundation Ireland

ABSTRACT

The generalized inverse Gaussian (GIG) Lévy process is a limit of compound Poisson processes, including the stationary gamma process and the stationary inverse Gaussian process as special cases. However, fitting the GIG Lévy process to data is computationally intractable due to the fact that the marginal distribution of the GIG Lévy process is not convolution-closed. The current work reveals that the marginal distribution of the GIG Lévy process admits a simple yet extremely accurate saddlepoint approximation. Particularly, we prove that if the order parameter of the GIG distribution is greater than or equal to -1, the marginal distribution can be approximated accurately – no need to normalize the saddlepoint density. Accordingly, maximum likelihood estimation is simple and quick, random number generation from the marginal distribution is straightforward by using Monte Carlo methods, and goodness-of-fit testing is undemanding to perform. Therefore, major numerical impediments to the application of the GIG Lévy process are removed. We demonstrate the accuracy of the saddlepoint approximation via various experimental setups.

KEY WORDS: Metropolis-Hastings algorithm; Modified Bessel functions of the second kind; Parametric bootstrap; Saddlepoint approximation.

1 Introduction

In the family of pure-jump increasing Lévy processes, both the gamma process and the inverse Gaussian process have wide applications. This is mainly because their marginal distributions, namely the gamma and the inverse Gaussian distributions, are convolution-closed and infinitely divisible. Therefore, these two Lévy processes can be easily extended to model non-stationary time-series data (see, e.g., Zhou et al., 2017; Cholette et al., 2019). This work introduces and studies a very general Lévy process, called the generalized inverse Gaussian (GIG) Lévy process, which includes the gamma and inverse Gaussian processes as special cases.

The GIG distribution was proposed by Étienne Halphen in 1941 and popularized by Ole Barndorff-Nielsen in the 1970s. Barndorff-Nielsen et al. (1978) proved that any GIG distribution with a non-positive power parameter is the distribution of the first hitting time to level 0 for a time-homogeneous diffusion process with state space $[0,\infty)$. This fact suggests the potential use of the GIG distribution as a lifetime distribution or the distribution for times between successive events in a renewal process (Embrechts, 1983). Halgreen (1979) further showed that the GIG distribution is self-decomposable. Therefore, all GIG probability density functions are unimodal (Yamazato, 1978). The self-decomposability makes the GIG distribution suitable for option pricing (see, e.g., Carr et al., 2007). The GIG distribution is also a conjugate prior for the normal distribution when serving as the mixing distribution in a normal variance-mean mixture (Barndorff-Nielsen, 1997).

Barndorff-Nielsen and Halgreen (1977) proved that the GIG distribution has infinite divisibility, which implies that we can construct a Lévy process from the GIG distribution, i.e., the GIG Lévy process. Applications of the GIG Lévy process are reported in Alexandrov and Lacis (2000) for cloud/aerosol particle size modelling, Protassov (2004) and Vilca et al. (2014) for constructing mixture distributions of heavy tail and skewness, Luciano and Semeraro (2010) for modelling return processes in finance, Themelis et al. (2016) for time-adaptive group sparse signal estimation, etc. However, unlike the gamma and inverse Gaussian processes, the GIG Lévy process has received very limited attention. This is mainly because the GIG distribution is not convolution-closed. In other words, for a GIG Lévy process $\{X_t;\ t \geq 0\}$, if X_t has a GIG density, then for any $s \neq t$, the random variable X_s does not have a GIG density. Moreover, the density function of X_s even does not have an analytic form. Hence, applying the GIG Lévy process to areas where the gamma and inverse Gaussian processes have been adopted is prohibitively daunting. The current work reveals that this problem can be solved by employing the saddlepoint approximation.

Saddlepoint methods provide approximations to densities and probabilities, which are very ac-

curate in a wide range of settings. In particular, it is often the case that relative errors of these approximations stay bounded in the extreme tails, a desirable property that is not shared by most other types of approximation. Saddlepoint approximations are constructed by performing various operations on the cumulant generating function of a random variable. For the development and discussion of saddlepoint methodology, see Daniels (1954) for details of the density approximation, Lugannani and Rice (1980) and Daniels (1987) for the discussion of a tail area approximation which has a uniform relative error, and Reid (1988) and Goutis and Casella (1999) for a review of saddlepoint techniques.

The main objective of this paper is to show that the marginal distribution of the GIG Lévy process can be well approximated by an analytical function and hence that the GIG Lévy process can be readily applied to model time series data. The remainder of this paper is organized as follows. Section 2 introduces the GIG distribution and the GIG Lévy process. Section 3 gives a detailed explanation of the saddlepoint approximation and its uniqueness. Section 4 reveals that, although the saddlepoint approximation is fairly accurate, it is not exact, even after normalization. The saddlepoint density is then modified to provide an improved approximation. Section 5 addresses the problems of parameter estimation, random number generation and goodness-of-fit test. Section 6 explores the accuracy of the saddlepoint approximation via simulation. Finally, Section 7 concludes with a summary and remarks.

2 GIG Distribution & GIG Lévy Process

The density function of the GIG distribution is given by

$$f(x;\lambda,a,b) = \frac{(\sqrt{a})^{\lambda}}{2(\sqrt{b})^{\lambda}K_{\lambda}(\sqrt{ab})}x^{\lambda-1}\exp\left(-\frac{1}{2}(ax+bx^{-1})\right), \qquad x>0, \tag{1}$$

where $a>0$, $b>0$, and the order parameter $\lambda \in \mathbb{R}$; $K_{\lambda}(\cdot)$ is a normalizing constant (called modified Bessel function of the second kind):

$$K_{\lambda}(v) = \frac{1}{2}\int_0^{\infty} x^{\lambda-1}\exp\left(-\frac{v}{2}(x+x^{-1})\right)dx.$$

$K_{\lambda}(v)$ is an exponentially decaying function of v, diverges for all orders at $v=0$, and has the property that $K_{-\lambda}(v) = K_{\lambda}(v)$. Modified Bessel functions of the second kind of order $\{0, 1, 2, 3, 4, 5\}$ are shown in Figure 1. We let $GIG(\lambda,a,b)$ represent the GIG distribution (1). GIG distributions enjoy several nice probabilistic features. For example, if X follows the GIG distribution

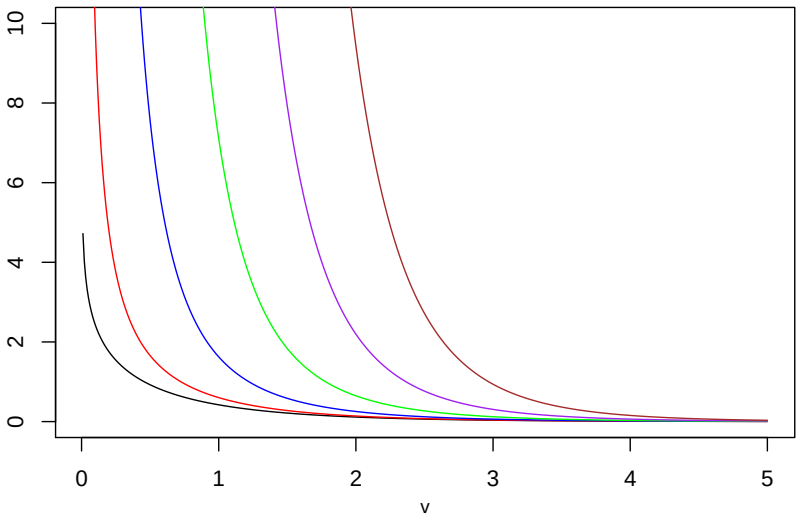

Figure 1: Plot of the modified Bessel functions of the second kind: $K_0(v)$ (black), $K_1(v)$ (red), $K_2(v)$ (blue), $K_3(v)$ (green), $K_4(v)$ (purple), $K_5(v)$ (brown).

$GIG(\lambda,a,b)$, then its reciprocal $1/X$ follows $GIG(-\lambda,b,a)$. The GIG distribution includes as special cases the gamma distribution ($b=0$ and $\lambda>0$), the inverse gamma distribution ($a=0$ and $\lambda<0$), the inverse Gaussian distribution ($\lambda=-0.5$) and the hyperbolic distribution ($\lambda=0$).

Owning to its infinite divisibility, we can construct a Lévy process from the GIG distribution, herein called GIG Lévy process. We say that the process $\{X_t;\ t\geq 0\}$ is a GIG Lévy process, if the law of X_1 is the GIG distribution $GIG(\lambda,a,b)$. A Lévy process can be fully determined by the characteristic function of X_t which is given by the Lévy-Khintchine formula. In the manner of Dufresne et al. (1991), it is easy to prove that

$$\mathrm{E}[e^{iuX_t}]=\exp\left(t\int_0^\infty(e^{iux}-1)\Pi(dx)\right),$$

where $\Pi(\cdot)$ is called the Lévy measure. According to Barndorff-Nielsen and Shephard (2001), the Lévy measure of the GIG Lévy process is absolutely continuous with density

$$\Pi(dx)=\frac{1}{x}\left[b\int_0^\infty \exp(-xz)g_\lambda(2bz)dz+\max\{0,\lambda\}\right]\exp(-\frac{a}{2}x)dx,$$

where $g_\lambda(y)=\left\{\frac{1}{2}\pi^2 y\left[J^2_{|\lambda|}(\sqrt{y})+Y^2_{|\lambda|}(\sqrt{y})\right]\right\}^{-1}$, $y>0$. $J_{|\lambda|}(\cdot)$ is the Bessel function of the first kind, and $Y_{|\lambda|}(\cdot)$ is the Bessel function of the second kind (see Chapter 9 of Abramovitz and Segun, 1970). For the GIG Lévy process, the arrival rate $\int_0^\infty\Pi(dx)$ is infinite (Morales, 2004). Hence, the

GIG Lévy process is a limit of compound Poisson processes, composing of an infinite number of infinitesimal jumps.

The GIG Lévy process includes the stationary gamma process and inverse Gaussian process as special cases. If $\{X_t;\ t \geq 0\}$ is a stationary gamma process (resp., inverse Gaussian process), $\forall$ $t > 0$, X_t always follows the gamma distribution (resp., the inverse Gaussian distribution). However, for the GIG Lévy process, only X_1 follows the GIG distribution. $\forall\, 0 \leq s < t$, if $t - s \neq 1$, then $X_t - X_s$ does not follow the GIG distribution. This is because the GIG distribution is not closed under convolution. In other words, if two random variables Z_1 and Z_2 both follow the GIG distribution $GIG(\lambda, a, b)$, then their sum $Z_1 + Z_2$ does not follow a GIG distribution. This undesirable feature has restrained the application of the GIG Lévy process in areas where the gamma process and inverse Gaussian process have been employed.

To make the GIG Lévy process a practical model, we need to formulate the density function of X_t for any $t > 0$, which can be derived via the Fourier inversion formula. $\forall\, t > 0$, let $f_{X_t}(x;\boldsymbol{\theta})$ denote the density function of X_t, where $\boldsymbol{\theta} = (\lambda, a, b)$. The characteristic function of a GIG random variable X is

$$\mathrm{E}[e^{iuX}] = \left(\frac{a}{a-2iu}\right)^{\frac{\lambda}{2}} \frac{K_\lambda(\sqrt{(a-2iu)b})}{K_\lambda(\sqrt{ab})},$$

where $u \in \mathbb{R}$. Hence, we have

$$f_{X_t}(x;\boldsymbol{\theta}) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \exp(t\varphi(u) - iux)du, \tag{2}$$

where $\varphi(u)$ is the logarithm of $\mathrm{E}[e^{iuX}]$:

$$\varphi(u) = \frac{\lambda}{2}[\log(a) - \log(a-2iu)] + \log(K_\lambda(\sqrt{(a-2iu)b})) - \log(K_\lambda(\sqrt{ab})).$$

Apparently, recovering the density function $f_{X_t}(x;\boldsymbol{\theta})$ from its characteristic function is not possible explicitly. Hence, in the following section we introduce the saddlepoint method for constructing a closed-form approximation to $f_{X_t}(x;\boldsymbol{\theta})$.

3 Saddlepoint Method

3.1 Brief Explanation

For readability, we introduce here the formal calculations to derive the saddlepoint approximation. Suppose X is a continuous random variable with density $f(x)$. Let $\psi(u)$ denote the moment

generating function: $\psi(u) = \int_{-\infty}^{\infty} e^{ux} f(x)dx$. Via the Fourier transform, we have

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-iux}\psi(iu)du = \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp(\log(\psi(iu)) - iux)du.$$

Substituting iu with u and applying the Closed Curve Theorem, we have

$$f(x) = \frac{1}{2\pi i}\int_{-i\infty}^{i\infty} \exp(\log(\psi(u)) - ux)du = \frac{1}{2\pi i}\int_{\tau-i\infty}^{\tau+i\infty} \exp(\log(\psi(u)) - ux)du,$$

where τ is within the interval of convergence for $\psi(u)$ which we assume to contain the origin as an interior point. Define $k(u,x)$ as $k(u,x) = \log(\psi(u)) - ux$. In what follows, we let $k'(u,x)$ and $k''(u,x)$ respectively denote the first and second derivative w.r.t. u. Approximate $k(u,x)$ by its Taylor expansion:

$$k(u,x) \approx k(\hat{u},x) + k''(\hat{u},x)\frac{(u-\hat{u})^2}{2},$$

where $\hat{u}$ satisfies $k'(\hat{u},x) = 0$ and $k''(\hat{u},x) > 0$. Then we have

$$f(x) \approx \frac{1}{2\pi i}\int_{\tau-i\infty}^{\tau+i\infty} \exp(k(\hat{u},x) + k''(\hat{u},x)\frac{(u-\hat{u})^2}{2})du = \frac{1}{2\pi i}\int_{-i\infty}^{i\infty} \exp(k(\hat{u},x) + k''(\hat{u},x)\frac{u^2}{2})du,$$

in which, again availing the Closed Curve Theorem, we have set $\tau = \hat{u}$. Substituting u with iu and employing the idea of Laplace approximation, we have

$$f(x) \approx \exp(k(\hat{u},x))\frac{1}{2\pi}\int_{-\infty}^{+\infty} \exp(-k''(\hat{u},x)\frac{u^2}{2})du = \frac{1}{\sqrt{2\pi k''(\hat{u},x)}}\exp(k(\hat{u},x)).$$

That is,

$$f(x) \approx \frac{1}{\sqrt{2\pi k''(\hat{u},x)}}\exp(\log(\psi(\hat{u})) - \hat{u}x), \tag{3}$$

which is the saddlepoint approximation for $f(x)$. Note that $\log(\psi(u))$ is called the cumulant-generating function of the random variable X.

3.2 Saddlepoint Density for $f_{X_t}(x;\boldsymbol{\theta})$

Let $H_\lambda(u)$ denote the cumulant generating function of X_1 (i.e., $t = 1$): for $u < \frac{a}{2}$,

$$H_\lambda(u) = \log(\mathrm{E}[e^{uX_1}]) = \frac{\lambda}{2}[\log(a) - \log(a-2u)] + \log(K_\lambda(\sqrt{(a-2u)b})) - \log(K_\lambda(\sqrt{ab})). \tag{4}$$

Then the cumulant generating function of X_t is $tH_\lambda(u)$. Following (3), the saddlepoint density approximation to $f_{X_t}(x;\boldsymbol{\theta})$ is:

$$\hat{f}_{X_t}(x;\boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi t H''_\lambda(\hat{u})}}\exp(tH_\lambda(\hat{u}) - \hat{u}x), \tag{5}$$

where $H''_\lambda(u)$ represents the second derivative of $H_\lambda(u)$ w.r.t. u, and $\hat{u}$ is the saddlepoint satisfying $tH'_\lambda(\hat{u}) - x = 0$. Note that $\hat{u} = \hat{u}(x,t)$ is a function of x and t. The first derivative of $H_\lambda(u)$ is

$$H'_\lambda(u) = \frac{\lambda}{a-2u} - \frac{K'_\lambda(\sqrt{(a-2u)b})}{K_\lambda(\sqrt{(a-2u)b})}\sqrt{\frac{b}{a-2u}},$$

where

$$K'_\lambda(v) = -\frac{1}{4}\int_0^\infty (x^\lambda + x^{\lambda-2})\exp\left(-\frac{v}{2}(x+x^{-1})\right)dx = -\frac{1}{2}[K_{\lambda+1}(v) + K_{\lambda-1}(v)]. \tag{6}$$

Hence, $\hat{u}$ is obtained by solving w.r.t. u

$$\frac{\lambda}{a-2u} - \frac{K'_\lambda(\sqrt{(a-2u)b})}{K_\lambda(\sqrt{(a-2u)b})}\sqrt{\frac{b}{a-2u}} = \frac{x}{t}. \tag{7}$$

The second derivative of $H_\lambda(u)$ is

$$\begin{aligned} H''_\lambda(u) =& \frac{2\lambda}{(a-2u)^2} - \frac{K'_\lambda(\sqrt{(a-2u)b})}{K_\lambda(\sqrt{(a-2u)b})}\sqrt{\frac{b}{(a-2u)^3}} \\ &+ \frac{K''_\lambda(\sqrt{(a-2u)b})K_\lambda(\sqrt{(a-2u)b}) - K'_\lambda(\sqrt{(a-2u)b})^2}{K_\lambda(\sqrt{(a-2u)b})^2}\frac{b}{a-2u}, \end{aligned} \tag{8}$$

where

$$K''_\lambda(v) = -\frac{1}{2}[K'_{\lambda+1}(v) + K'_{\lambda-1}(v)] = \frac{1}{4}[K_{\lambda+2}(v) + 2K_\lambda(v) + K_{\lambda-2}(v)].$$

The pseudo code in Algorithm 1 summarizes the steps for evaluating the saddlepoint density function $\hat{f}_{X_t}(x;\boldsymbol{\theta})$ at any point x.

Algorithm 1 Evaluating the saddlepoint density $\hat{f}_{X_t}(x;\boldsymbol{\theta})$.

1: Solve the saddlepoint equation (7), w.r.t. u, to obtain the saddlepoint $\hat{u}$;
2: Replace u in Equation (8) with $\hat{u}$ to calculate $H''_\lambda(\hat{u})$;
3: Replace u in Equation (4) with $\hat{u}$ to calculate $H_\lambda(\hat{u})$;
4: Calculate $\hat{f}_{X_t}(x;\boldsymbol{\theta})$ by evaluating the right hand side of Equation (5).

3.3 Existence and Uniqueness of the Saddlepoint

The feasibility of the saddlepoint approximation depends on the existence and uniqueness of the solution $\hat{u}$ to $tH'_\lambda(u) = x$ and on $\hat{u}$ satisfying $H''_\lambda(\hat{u}) > 0$. In this section we discuss the existence and properties of the real root of the equation

$$H'_\lambda(u) = \xi.$$

Proposition 1. *There is no real root of the equation $H'_\lambda(u) = \xi$ whenever $\xi \leq 0$.*

Proof. Define a function $M(u,\xi)$ as

$$M(u,\xi) = e^{H_\lambda(u)-u\xi} = \int_0^{+\infty} e^{u(x-\xi)} f(x;\lambda,a,b)dx,$$

which exists only for $u < \frac{a}{2}$. Taking partial derivative of $M(u,\xi)$ w.r.t. u, we have that

$$M'(u,\xi) = [H'_\lambda(u) - \xi]e^{H_\lambda(u)-u\xi}, \tag{9}$$

and that

$$M'(u,\xi) = \int_0^{+\infty} (x-\xi)e^{u(x-\xi)} f(x;\lambda,a,b)dx. \tag{10}$$

Here and in Proposition 2, we implicitly utilize the dominated convergence theorem to exchange derivatives and integrals. It is clear that the integrand $e^{u(x-\xi)} f(x;\lambda,a,b)$ and its partial derivative w.r.t. u are integrable functions of x.

From Equation (10) we know that, when $\xi \leq 0$, $M'(u,\xi) > 0$ for any $u < \frac{a}{2}$. Hence, from Equation (9) we know that, when $\xi \leq 0$, $H'_\lambda(u) - \xi > 0$ for any $u < \frac{a}{2}$; that is, $H'_\lambda(u) - \xi = 0$ has no real root when $\xi \leq 0$. □

Proposition 2. *For any $\xi > 0$, if there exists a root of the equation $H'_\lambda(u) = \xi$, then the root is simple and unique and satisfies $H''_\lambda(\hat{u}) > 0$. A necessary and sufficient condition for the equation to have a root for all $\xi > 0$ is that $\lim_{u\to\frac{a}{2}} H'_\lambda(u) = +\infty$.*

Proof. Note that $M'(u,\xi)$ is strictly increasing with u, because

$$M''(u,\xi) = \int_0^{+\infty} (x-\xi)^2 e^{u(x-\xi)} f(x;\lambda,a,b)dx > 0.$$

Then for any root $\hat{u}$ of the equation $H'_\lambda(u) = \xi$, we have $M''(\hat{u},\xi) = H''_\lambda(\hat{u})e^{H_\lambda(\hat{u})-\hat{u}\xi} > 0$, and therefore $H''_\lambda(\hat{u}) > 0$, and $\hat{u}$ is a simple root.

When $\xi > 0$, we rewrite Equation (10) into:

$$M'(u,\xi) = \int_0^{\xi} (x-\xi)e^{u(x-\xi)} f(x;\lambda,a,b)dx + \int_\xi^{+\infty} (x-\xi)e^{u(x-\xi)} f(x;\lambda,a,b)dx.$$

For the first integration, we have

$$\lim_{u\to-\infty} \int_0^{\xi} (x-\xi)e^{u(x-\xi)} f(x;\lambda,a,b)dx = -\infty.$$

For the second integration, let m denote the maximum value of $f(x;\lambda,a,b)$ over the interval $(\xi,+\infty)$. Because $f(x;\lambda,a,b)$ is integrable over the interval $(0,+\infty)$, we must have $\lim_{x\to+\infty} f(x;\lambda,a,b) = 0$, and hence m is finite. Then

$$0 \le \lim_{u\to-\infty} \int_{\xi}^{+\infty} (x-\xi) e^{u(x-\xi)} f(x;\lambda,a,b)dx \le \lim_{u\to-\infty} m \int_{\xi}^{+\infty} (x-\xi) e^{u(x-\xi)} dx = 0.$$

Therefore, $\lim_{u\to-\infty} M'(u,\xi) = -\infty$ for any $\xi > 0$, and we conclude that $M'(u,\xi)$ is strictly increasing from $-\infty$. If $\lim_{u\to\frac{a}{2}} M'(u,\xi) < 0$, then $M'(u,\xi) < 0$ (and hence $H'_\lambda(u) - \xi < 0$) for all $u < \frac{a}{2}$. If $\lim_{u\to\frac{a}{2}} M'(u,\xi) > 0$, then there is only one point $\hat{u}$ at which $M'(\hat{u},\xi) = 0$; that is, if $\lim_{u\to\frac{a}{2}} M'(u,\xi) > 0$, then the equation $H'_\lambda(u) - \xi = 0$ has one and only one root.

Equation (9) indicates that $\lim_{u\to\frac{a}{2}} M'(u,\xi) > 0$ for all $\xi > 0$ if and only if $\lim_{u\to\frac{a}{2}} H'_\lambda(u) = +\infty$. □

Proposition 3. *We have* $\lim_{u\to-\infty} H'_\lambda(u) = 0$ *for any fixed* λ*, and that* $H'_\lambda(u)$ *is a strictly increasing function of* u*.*

Proof. We note that, for large values of v, the asymptotic approximation of $K_\lambda(v)$ is $\sqrt{\frac{\pi}{2v}}\exp(-v)$, and therefore

$$\lim_{v\to\infty} \frac{K_{\lambda+1}(v)}{K_\lambda(v)} = \lim_{v\to\infty} \frac{K_{\lambda-1}(v)}{K_\lambda(v)} = 1.$$

Then it follows from (6) that:

$$\begin{aligned}\lim_{u\to-\infty} H'_\lambda(u) &= \lim_{u\to-\infty} \left\{ \frac{\lambda}{a-2u} + \frac{1}{2} \frac{K_{\lambda+1}(\sqrt{(a-2u)b}) + K_{\lambda-1}(\sqrt{(a-2u)b})}{K_\lambda(\sqrt{(a-2u)b})} \sqrt{\frac{b}{a-2u}} \right\} \\ &= \lim_{u\to-\infty} \frac{\lambda}{a-2u} + \lim_{u\to-\infty} \sqrt{\frac{b}{a-2u}} = 0.\end{aligned}$$

According to Proposition 2, for any $\xi > 0$, if there exists a root of the equation $H'_\lambda(u) = \xi$, then the root is unique and satisfies $H''_\lambda(\hat{u}) > 0$. To prove that $H'_\lambda(u)$ is a strictly increasing function, we only need to prove that $H''_\lambda(u) > 0$ for any $u < \frac{a}{2}$. Let ξ_0 be the point at which $\lim_{u\to\frac{a}{2}} M'(u,\xi_0) = 0$. On one hand, for any $0 < \xi < \xi_0$, we have

$$\lim_{u\to\frac{a}{2}} M'(u,\xi) = \lim_{u\to\frac{a}{2}} \int_0^{+\infty} (x-\xi) e^{u(x-\xi)} f(x;\lambda,a,b)dx > \lim_{u\to\frac{a}{2}} M'(u,\xi_0) = 0,$$

and therefore there is a unique root to the equation $H'_\lambda(u) = \xi$ for any $0 < \xi < \xi_0$. On the other hand, recall that $M'(u,\xi)$ is strictly increasing with u; if $\lim_{u\to\frac{a}{2}} M'(u,\xi_0) = 0$, then $M'(u,\xi_0) < 0$ for all $u < \frac{a}{2}$, and therefore $H'_\lambda(u) - \xi_0 < 0$ for all $u < \frac{a}{2}$; that is, we have $0 < H'_\lambda(u) < \xi_0$ for any

u. Therefore, we can claim that $H'_\lambda(u)$ is a bijection from $(-\infty, \frac{a}{2})$ to $(0, \xi_0)$. Combining with the fact that $H''_\lambda(\hat{u}) > 0$ everywhere, we can conclude that $H'_\lambda(u)$ is a strictly increasing function of u, mapping $(-\infty, \frac{a}{2})$ to $(0, \xi_0)$, where ξ_0 is the point at which $\lim_{u\to\frac{a}{2}} M'(u, \xi_0) = 0$ (or, equivalently, $\xi_0 = \lim_{u\to\frac{a}{2}} H'_\lambda(u)$). □

Theorem 1. *If $\lim_{u\to\frac{a}{2}} H'_\lambda(u) = +\infty$, then with ξ increasing from 0 to $+\infty$, the unique root $\hat{u}$ increases monotonically from $-\infty$ to $\frac{a}{2}$. We have*

$$\lim_{u\to\frac{a}{2}} H'_\lambda(u) = \begin{cases} +\infty, & \text{if } \lambda \geq -1; \\ \frac{-b}{4(\lambda+1)}, & \text{if } \lambda < -1. \end{cases}$$

For equation $H'_\lambda(u) = \xi$,

- *when $\lambda \geq -1$, there is a unique simple root for any $\xi > 0$;*
- *when $\lambda < -1$, there is a unique simple root for any $0 < \xi < \frac{-b}{4(\lambda+1)}$, but no root for any $\xi \geq \frac{-b}{4(\lambda+1)}$.*

Proof. When $\lambda > 0$, we have

$$\begin{aligned} \lim_{u\to\frac{a}{2}} H'_\lambda(u) &= \lim_{u\to\frac{a}{2}} \left\{ \frac{\lambda}{a-2u} + \frac{1}{2} \frac{K_{\lambda+1}(\sqrt{(a-2u)b}) + K_{\lambda-1}(\sqrt{(a-2u)b})}{K_\lambda(\sqrt{(a-2u)b})} \sqrt{\frac{b}{a-2u}} \right\} \\ &\geq \lim_{u\to\frac{a}{2}} \frac{\lambda}{a-2u} = +\infty. \end{aligned}$$

When $\lambda = 0$, we have

$$\begin{aligned} \lim_{u\to\frac{a}{2}} H'_\lambda(u) &= \lim_{u\to\frac{a}{2}} \left\{ \frac{1}{2} \frac{K_{\lambda+1}(\sqrt{(a-2u)b}) + K_{\lambda-1}(\sqrt{(a-2u)b})}{K_\lambda(\sqrt{(a-2u)b})} \sqrt{\frac{b}{a-2u}} \right\} \\ &\geq \lim_{u\to\frac{a}{2}} \left\{ \frac{1}{2} \frac{K_{\lambda+1}(\sqrt{(a-2u)b})}{K_\lambda(\sqrt{(a-2u)b})} \sqrt{\frac{b}{a-2u}} \right\} \\ &\geq \lim_{u\to\frac{a}{2}} \frac{1}{2} \sqrt{\frac{b}{a-2u}} = +\infty. \end{aligned}$$

When $\lambda < 0$ and $\lambda \neq -1$, define $v = \sqrt{(a-2u)b}$ and we have

$$H'_\lambda(u) = \frac{b\lambda}{v^2} + \frac{1}{2} \frac{K_{|\lambda+1|}(v) + K_{|\lambda-1|}(v)}{K_{|\lambda|}(v)} \frac{b}{v},$$

where we have utilized the property that $K_{-\lambda}(v) = K_{\lambda}(v)$. For small values of v, the asymptotic approximation of $K_{\lambda}(v)$ is $\frac{1}{2}\Gamma(\lambda)(\frac{1}{2}v)^{-\lambda}$ for $\lambda > 0$, and therefore

$$\lim_{v\to 0} \frac{K_{|\lambda+1|}(v) + K_{|\lambda-1|}(v)}{K_{|\lambda|}(v)} = \lim_{v\to 0} \frac{\Gamma(|\lambda+1|)(\frac{1}{2}v)^{-|\lambda+1|} + \Gamma(|\lambda-1|)(\frac{1}{2}v)^{-|\lambda-1|}}{\Gamma(|\lambda|)(\frac{1}{2}v)^{-|\lambda|}}.$$

Hence, when $-1 < \lambda < 0$, we have

$$\begin{aligned}\lim_{u\to\frac{a}{2}} H'_{\lambda}(u) &= \lim_{v\to 0} \left\{ \frac{b\lambda}{v^2} + \frac{1}{2} \frac{\Gamma(\lambda+1)(\frac{1}{2}v)^{-\lambda-1} + \Gamma(1-\lambda)(\frac{1}{2}v)^{\lambda-1}}{\Gamma(-\lambda)(\frac{1}{2}v)^{\lambda}} \frac{b}{v} \right\} \\ &= \lim_{v\to 0} \frac{\Gamma(\lambda+1)4^{\lambda}}{\Gamma(-\lambda)v^{2\lambda+2}} b = +\infty.\end{aligned}$$

When $\lambda < -1$, we have

$$\begin{aligned}\lim_{u\to\frac{a}{2}} H'_{\lambda}(u) &= \lim_{v\to 0} \left\{ \frac{b\lambda}{v^2} + \frac{1}{2} \frac{\Gamma(-\lambda-1)(\frac{1}{2}v)^{\lambda+1} + \Gamma(1-\lambda)(\frac{1}{2}v)^{\lambda-1}}{\Gamma(-\lambda)(\frac{1}{2}v)^{\lambda}} \frac{b}{v} \right\} \\ &= \frac{b}{4} \frac{\Gamma(-\lambda-1)}{\Gamma(-\lambda)} < +\infty.\end{aligned}$$

When $\lambda = -1$, we have

$$H'_{\lambda=-1}(u) = -\frac{1}{a-2u} + \frac{1}{2} \frac{K_0(\sqrt{(a-2u)b}) + K_2(\sqrt{(a-2u)b})}{K_1(\sqrt{(a-2u)b})} \sqrt{\frac{b}{a-2u}}.$$

For small values of v, the asymptotic approximation of $K_0(v)$ is $-\log(v)$. Therefore, we have

$$\lim_{u\to\frac{a}{2}} H'_{\lambda=-1}(u) = \lim_{v\to 0} \left\{ -\frac{b}{4}\log(v) \right\} = +\infty.$$

To conclude, we have

$$\lim_{u\to\frac{a}{2}} H'_{\lambda}(u) = \begin{cases} +\infty, & \lambda \geq -1; \\ \frac{-b}{4(\lambda+1)}, & \lambda < -1. \end{cases}$$

□

Theorem 1 also explains why the inverse Gaussian distribution, i.e. $\lambda = -0.5$, can be (exactly) approximated by the saddlepoint density (Daniels, 1980).

4 Saddlepoint Density Modification

4.1 Approximation Error

In many applications, the saddlepoint density does not integrate to one, and hence needs to be normalized. We here point out that generally the saddlepoint approximation for $f_{X_t}(x;\boldsymbol{\theta})$ is not exact, even after normalization.

We prove by calculating the ratio $\frac{\hat{f}_{X_t}(x;\boldsymbol{\theta})}{f_{X_t}(x;\boldsymbol{\theta})}$. If the ratio is not 1, then we can conclude that the saddlepoint approximation $\hat{f}_{X_t}(x;\boldsymbol{\theta})$ is not exact. Moreover, if the ratio changes with x, then we can conclude that even the normalized saddlepoint approximation is not exact. Note that, on one hand, only $f_{X_1}(x;\boldsymbol{\theta})$ has an explicit expression. On the other hand, if the normalized $\hat{f}_{X_1}(x;\boldsymbol{\theta})$ is not exact, then for any $t > 0$, the normalized $\hat{f}_{X_t}(x;\boldsymbol{\theta})$ is not exact either. Therefore, we only need to examine the ratio $\frac{\hat{f}_{X_1}(x;\boldsymbol{\theta})}{f_{X_1}(x;\boldsymbol{\theta})}$ for $x > 0$.

In Figure 2 we plot the ratio for different values of (a,b), with λ fixed at value 2. In each row, b

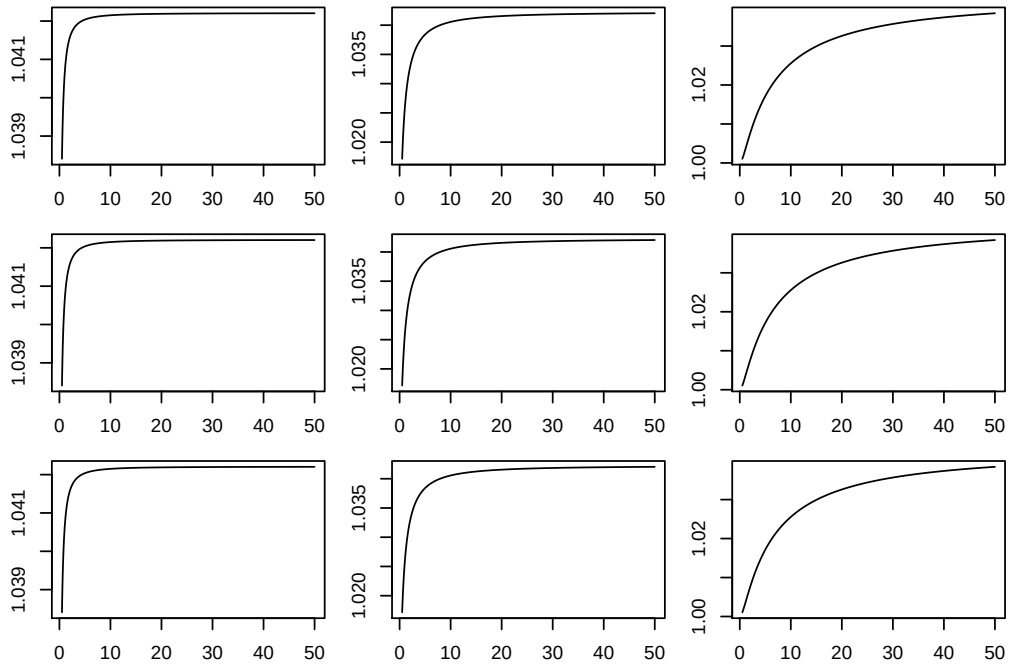

Figure 2: Plot of the ratio $\frac{\hat{f}_{X_1}(x;\boldsymbol{\theta})}{f_{X_1}(x;\boldsymbol{\theta})}$ for different values of a and b, with λ fixed at 2. In each row, b takes a value from $\{0.1, 1, 10\}$; in each column, a takes a value from $\{0.1, 1, 10\}$. The x-axis represents the value of x.

takes a value from $\{0.1, 1, 10\}$, while in each column, a takes a value from $\{0.1, 1, 10\}$. The x-axis

represents the value of x, and the y-axis represents the value of the ratio $\frac{\hat{f}_{X_1}(x;\boldsymbol{\theta})}{f_{X_1}(x;\boldsymbol{\theta})}$. Figure 3 repeats the procedure with λ fixed at value -0.75. Both figures indicate that the saddlepoint approximation

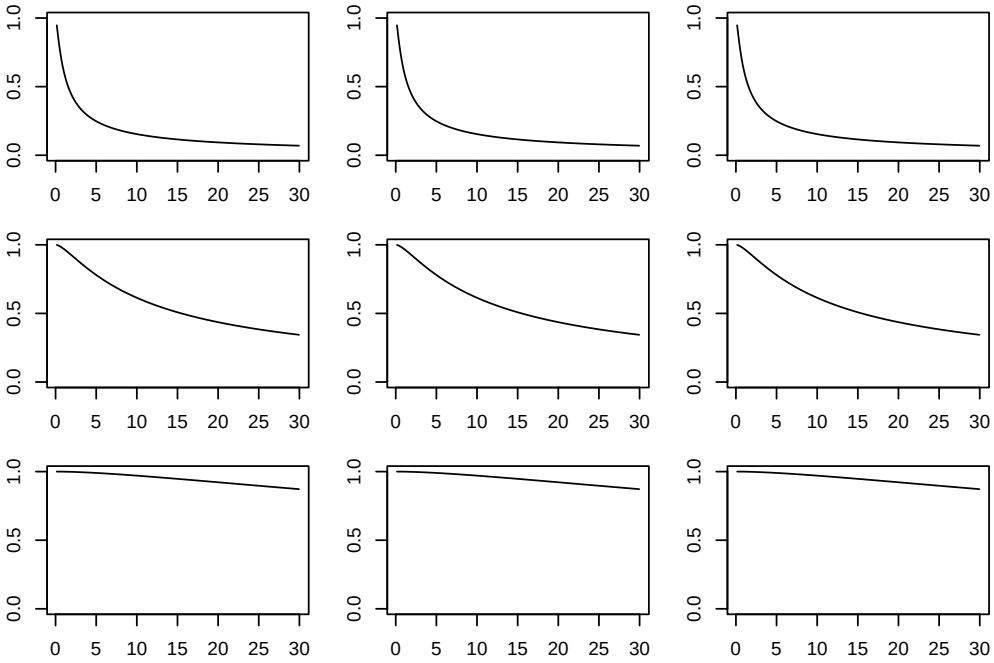

Figure 3: Plot of the ratio $\frac{\hat{f}_{X_1}(x;\boldsymbol{\theta})}{f_{X_1}(x;\boldsymbol{\theta})}$ for different values of a and b, with λ fixed at -0.75.

$\hat{f}_{X_t}(x;\boldsymbol{\theta})$, even after normalization, is not exact. Moreover, in Figure 2, all the ratios are above 1, while in Figure 3, all the ratios are below 1. All the ratio curves are bounded. Note that the relative difference between $\hat{f}_{X_1}(x;\boldsymbol{\theta})$ and $f_{X_1}(x;\boldsymbol{\theta})$, as measured by the ratio, increases with x; however, with x increasing, the true value $f_{X_1}(x;\boldsymbol{\theta})$ quickly converges to 0. Hence, if we plot the two density functions, they are visually the same (see Section 6). In fact, via exhaustive numerical study, we find that the saddlepoint approximation $\hat{f}_{X_t}(x;\boldsymbol{\theta})$ has the following property:

Proposition 4. *The saddlepoint approximation $\hat{f}_{X_t}(x;\boldsymbol{\theta})$ is exact only when $\lambda = -0.5$. When $\lambda \neq -0.5$, it is not exact even after normalization. When $\lambda > -0.5$, the ratio $\frac{\hat{f}_{X_t}(x;\boldsymbol{\theta})}{f_{X_t}(x;\boldsymbol{\theta})}$ is larger than one for any $x > 0$, and increases with x. When $\lambda < -0.5$, the inverse ratio $\frac{f_{X_t}(x;\boldsymbol{\theta})}{\hat{f}_{X_t}(x;\boldsymbol{\theta})}$ is larger than one for any $x > 0$, and increases with x.*

Remark 1. *We can relate the GIG density (1) to the inverse Gaussian density $f(x;-0.5,a,b)$ by writing*

$$f(x;\lambda,a,b) = \frac{x^{\lambda+0.5}}{E[X^{\lambda+0.5}]} f(x;-0.5,a,b),$$

where

$$f(x;-0.5,a,b)=\sqrt{\frac{b}{2\pi}}\exp(\sqrt{ab})x^{-3/2}\exp\left(-\frac{1}{2}(ax+bx^{-1})\right),$$

and the expectation is taken w.r.t. $f(x;-0.5,a,b)$*. The saddlepoint density can exactly approximate* $f(x;-0.5,a,b)$*, and* $E[X^{\lambda+0.5}]$ *is independent of x. Therefore, the approximation error is introduced by the exponentiation* $x^{\lambda+0.5}$*. When* λ *increases from -1 to infinity, the change of the difference* $\hat{f}_{X_t}(x;\boldsymbol{\theta})-f_{X_t}(x;\boldsymbol{\theta})$ *from negative to positive may be caused by the exponentiation* $x^{\lambda+0.5}$*, which changes from a decreasing function to an increasing function.*

4.2 A Modified Approximation

Proposition 4 indicates that if we want to improve the approximation $\hat{f}_{X_t}(x;\boldsymbol{\theta})$, we have to multiply it by a non-constant factor. Following Section 3.1, let $\psi_0(w)$ denote the moment generating function of an appropriate distribution that admits an analytic density function $f_0(x)$. Define $k_0(w,x)=\log(\psi_0(w))-wx$, and let $\hat{w}$ be the unique root of $k_0'(w,x)=0$. Instead of approximating $k(u,x)$ by a truncated Taylor expansion, Ait-Sahalia and Yu (2006) approximated it by $k(\hat{u},x)+k_0(w,x)-k_0(\hat{w},x)$:

$$k(u,x)-k(\hat{u},x)\approx k_0(w,x)-k_0(\hat{w},x).$$

Both $k(u,x)$ and $k_0(w,x)$ are strictly convex. Hence, $u=\hat{u}$ if and only if $w=\hat{w}$. Now it is clear that the appropriateness of the benchmark density $f_0(x)$ means that $\psi_0(w)$ is defined on a non-trivial interval, and $\hat{w}$ exists whenever $\hat{u}$ exists. Moreover, the two local functions, $k(u,x)-k(\hat{u},x)$ around $\hat{u}$ and $k_0(w,x)-k_0(\hat{w},x)$ around $\hat{w}$, are expected to behave alike.

Now we can treat u as a function of w. By differentiating twice the above equation and setting w to be $\hat{w}$, we have

$$u'(\hat{w})=\sqrt{\frac{k_0''(\hat{w},x)}{k''(\hat{u},x)}}.$$

Then we have

$$\begin{aligned} f(x) &= \frac{1}{2\pi i}\int_{\hat{u}-i\infty}^{\hat{u}+i\infty} \exp(\log(\psi(u)) - ux)du \\ &\approx \exp\left(k(\hat{u},x) - k_0(\hat{w},x)\right)\frac{1}{2\pi i}\int_{\hat{w}-i\infty}^{\hat{w}+i\infty} \exp(\log(\psi_0(w)) - wx)u'(w)dw \\ &\approx u'(\hat{w})\exp\left(k(\hat{u},x) - k_0(\hat{w},x)\right)\frac{1}{2\pi i}\int_{\hat{w}-i\infty}^{\hat{w}+i\infty} \exp(\log(\psi_0(w)) - wx)dw \\ &= \sqrt{\frac{k_0''(\hat{w},x)}{k''(\hat{u},x)}}\exp\left([\log(\psi(\hat{u})) - \hat{u}x] - [\log(\psi_0(\hat{w})) - \hat{w}x]\right) f_0(x), \end{aligned}$$

or, equivalently,

$$f(x) \approx \frac{1}{\sqrt{2\pi k''(\hat{u},x)}}\exp\left(\log(\psi(\hat{u})) - \hat{u}x\right)\frac{f_0(x)}{\frac{1}{\sqrt{2\pi k_0''(\hat{w},x)}}\exp(\log(\psi_0(\hat{w})) - \hat{w}x)}.$$

We notice that $\frac{1}{\sqrt{2\pi k_0''(\hat{w},x)}}\exp(\log(\psi_0(\hat{w})) - \hat{w}x)$ is the saddlepoint approximation for $f_0(x)$.

For the GIG Lévy process, one candidate of $f_0(x)$ for $f_{X_t}(x;\boldsymbol{\theta})$ is given by

$$f_0(x) = \frac{(\sqrt{a})^\lambda}{2(\sqrt{bt^2})^\lambda K_\lambda(\sqrt{abt^2})}x^{\lambda-1}\exp\left(-\frac{1}{2}(ax + bt^2x^{-1})\right),$$

which simply replaces b in Equation (1) with bt^2. Hence, the quantities $\hat{w}$, $\psi_0(\hat{w})$ and $k_0''(\hat{w},x)$ can be readily calculated following Section 3.2. Let $\bar{f}_{X_t}(x;\boldsymbol{\theta})$ denote the modified saddlepoint approximation for $f_{X_t}(x;\boldsymbol{\theta})$. According to Section 3.2, we have

$$\begin{aligned} \bar{f}_{X_t}(x;\boldsymbol{\theta}) =& \sqrt{\frac{k_0''(\hat{w},x)}{tH_\lambda''(\hat{u}(x,t))}}\frac{(\sqrt{a})^\lambda}{2(\sqrt{bt^2})^\lambda K_\lambda(\sqrt{abt^2})}x^{\lambda-1}\exp(-\frac{1}{2}(ax + bt^2x^{-1})) \\ &\times \exp\left([tH_\lambda(\hat{u}(x,t)) - \hat{u}(x,t)x] - [\log(\psi_0(\hat{w})) - \hat{w}x]\right). \end{aligned}$$

When $t = 1$, $f_0(x)$ is identical to $f_{X_1}(x;\boldsymbol{\theta})$, and hence $\bar{f}_{X_1}(x;\boldsymbol{\theta})$ is exact. Note that, when $\lambda = -0.5$, $f_0(x)$ is the marginal density function of the inverse Gaussian process.

We examine the similarity between $f_0(x)$ and $f_{X_t}(x;\boldsymbol{\theta})$, and for ease of exposition we let $t = m$ be an integer. Let i.i.d. random variables $\{Y_1,\ldots,Y_m\}$ follow the GIG distribution (1), and i.i.d. random variables $\{X_1,\ldots,X_m\}$ follow the inverse Gaussian distribution $f(x;-0.5,a,b)$. Then the summand $\sum_{i=1}^m Y_i$ follows the distribution $f_{X_m}(x;\boldsymbol{\theta})$ with the moment generating function given by

$$\mathrm{E}[\exp(u\sum_{i=1}^m Y_i)] = \prod_{i=1}^m \mathrm{E}[\exp(uY_i)] = \prod_{i=1}^m \frac{\mathrm{E}[X_i^{\lambda+0.5}\exp(uX_i)]}{\mathrm{E}[X_i^{\lambda+0.5}]} = \frac{\mathrm{E}[(\prod_{i=1}^m X_i)^{\lambda+0.5}\exp(u\sum_{i=1}^m X_i)]}{\mathrm{E}[(\prod_{i=1}^m X_i)^{\lambda+0.5}]}.$$

The corresponding $f_0(x)$ for $f_{X_m}(x;\boldsymbol{\theta})$ is

$$f_0(x) = \frac{(\sqrt{a})^\lambda}{2(\sqrt{bm^2})^\lambda K_\lambda(\sqrt{abm^2})} x^{\lambda-1} \exp\left(-\frac{1}{2}(ax + bm^2 x^{-1})\right).$$

Let Y be a random variable from $f_0(x)$, and X a random variable from $f_{X_m}(x;-0.5,a,b)$ – a special case of $f_{X_m}(x;\boldsymbol{\theta})$ and also the density function of $\sum_{i=1}^m X_i$. The moment generating function of Y is

$$\mathrm{E}[\exp(wY)] = \frac{\mathrm{E}[X^{\lambda+0.5}\exp(wX)]}{\mathrm{E}[X^{\lambda+0.5}]} = \frac{\mathrm{E}[(\sum_{i=1}^m X_i)^{\lambda+0.5}\exp(w\sum_{i=1}^m X_i)]}{\mathrm{E}[(\sum_{i=1}^m X_i)^{\lambda+0.5}]}.$$

We observe that the moment generating function of $f_{X_m}(x;\boldsymbol{\theta})$ and that of $f_0(x)$ are alike. Hence, it is expected that the modified saddlepoint density $\bar{f}_{X_t}(x;\boldsymbol{\theta})$ will yield a better approximation.

5 Other Issues

5.1 Parameter Estimation

We now fit the GIG Lévy process to a time series dataset and estimate the unknown parameters. Let $\hat{\boldsymbol{\theta}}$ denote the maximum likelihood (ML) estimate of $\boldsymbol{\theta}$, and Θ a restricted parameter set: $\Theta = \{(\lambda,a,b)|\lambda > -1, a > 0, b > 0\}$. Here we presume that the true value of $\boldsymbol{\theta}$ falls in Θ, as otherwise the saddlepoint method will fail. Although Θ is a restricted set, it is still larger than the parameter sets of the gamma process and the inverse Gaussian process. Starting from time 0, suppose we have data $\{x_0,x_1,x_2,\cdots,x_m\}$ collected at time points $0 = t_0 < t_1 < t_2 < \cdots < t_m$. Define $\Delta x_i = x_i - x_{i-1}$ and $\Delta t_i = t_i - t_{i-1}$ for $i = 1,\ldots,m$. Then the likelihood for observing Δx_i shall be $f_{X_{\Delta t_i}}(\Delta x_i;\boldsymbol{\theta})$ which, according to Equation (2), is difficult to evaluate.

We notice from Figure 2 that the approximation error is quite small and uniformly bounded. In Section 6.1, we will corroborate that the saddlepoint density approximates the true density nearly exactly. In fact, due to the round-off error, numerical integration of $\hat{f}_{X_t}(x;\boldsymbol{\theta})$ over the interval $(0,+\infty)$ even gives the value 1. Hence we might perform ML estimation by directly maximizing the log-likelihood $\sum_{i=1}^m \log(\hat{f}_{X_{\Delta t_i}}(\Delta x_i;\boldsymbol{\theta}))$:

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}\in\Theta} \sum_{i=1}^m \left\{ -\frac{1}{2}\log(H''_\lambda(\hat{u}(\Delta x_i,\Delta t_i))) + \Delta t_i H_\lambda(\hat{u}(\Delta x_i,\Delta t_i)) - \hat{u}(\Delta x_i,\Delta t_i)\Delta x_i \right\}.$$

Here, we use the notation $\hat{u}(x,t)$ to highlight that the root is a function of x and t. Maximizing the above log-likelihood function is undemanding: the root $\hat{u}(\Delta x_i,\Delta t_i)$ can be quickly found using, e.g., the bisection method, because $H'_\lambda(u)$ is a strictly increasing function of u.

5.2 Random Number Generation for $f_{X_t}(x;\boldsymbol{\theta})$

We generate data from $\hat{f}_{X_t}(x;\boldsymbol{\theta})$ and treat them as sampled from $f_{X_t}(x;\boldsymbol{\theta})$. Though we can evaluate $\hat{f}_{X_t}(x;\boldsymbol{\theta})$ for any x, the root $\hat{u}(x,t)$ does not have an analytic expression. Hence, there is no method available to directly draw i.i.d. samples from $\hat{f}_{X_t}(x;\boldsymbol{\theta})$. We here propose to adopt the Markov chain Monte Carlo (MCMC) technique to generate a sequence of dependent samples, denoted by $\{\mathscr{X}_i, i = 1,2,\cdots\}$, which is a Markov chain with the equilibrium distribution $\hat{f}_{X_t}(x;\boldsymbol{\theta})$. Then the Nth element (with N being sufficiently large), $\mathscr{X}_N$, can be used as a random sample from $\hat{f}_{X_t}(x;\boldsymbol{\theta})$. Readers are referred to the excellent texts of Chen et al. (2012) and Meyn and Tweedie (2012) and review papers by Tierney (1994) and Andrieu et al. (2003) for more information on MCMC. We herein develop a sampler based on the Metropolis-Hastings (MH) algorithm.

Let $\mathscr{X}_i = x$ denote the current state of the Markov chain. The MH sampler is composed of three steps: (1) Generate a proposal sample $\ddot{x}$ from a proposal distribution $g(\ddot{x}|x)$. (2) Compute the acceptance probability α: $\alpha = \min\{1, \frac{\hat{f}_{X_t}(\ddot{x};\boldsymbol{\theta})}{\hat{f}_{X_t}(x;\boldsymbol{\theta})} \frac{g(x|\ddot{x})}{g(\ddot{x}|x)}\}$. (3) Accept the candidate sample with probability α. If $\ddot{x}$ is accepted, set $\mathscr{X}_{i+1} = \ddot{x}$; otherwise, set $\mathscr{X}_{i+1} = x$. The proposal distribution $g(\ddot{x}|x)$ is a conditional distribution that represents the probability of moving from x to $\ddot{x}$. If the proposal distribution satisfies the regularity conditions: irreducibility and aperiodicity, then the generated Markov chain converges to the target distribution, i.e., $\hat{f}_{X_t}(x;\boldsymbol{\theta})$ (Tierney, 1994; Mengersen and Tweedie, 1996).

To fully develop an MH sampler, we need to specify the proposal distribution $g(\ddot{x}|x)$. Here we work with the Gaussian distribution centered at the current value x with standard deviation $\sigma(>0)$. The value of σ should be subjectively determined to maintain the acceptance rate of proposals in a reasonable range. Note that the target distribution, $\hat{f}_{X_t}(x;\boldsymbol{\theta})$, does not have full support, while the Gaussian proposal distribution does. Hence, we need to work with a slightly different proposal distribution – the truncated Gaussian distribution: $g(\ddot{x}|x) = \frac{\phi(\frac{\ddot{x}-x}{\sigma})}{\Phi(\frac{x}{\sigma})}$, $\ddot{x} > 0$, where $\phi(\cdot)$ and $\Phi(\cdot)$ are respectively the density function and cumulative distribution function of the standard normal distribution. Then the acceptance probability is simply $\alpha = \min\{1, \frac{\hat{f}_{X_t}(\ddot{x};\boldsymbol{\theta})}{\hat{f}_{X_t}(x;\boldsymbol{\theta})} \frac{\Phi(\frac{x}{\sigma})}{\Phi(\frac{\ddot{x}}{\sigma})}\}$.

Remark 2. *If the value of λ in $\hat{f}_{X_t}(x;\boldsymbol{\theta})$ is not very large, we can apply the importance sampling with the proposal distribution being the inverse Gaussian distribution ($\lambda = -0.5$) or the hyperbolic distribution ($\lambda = 0$). Importance sampling is faster than the MH sampler, as the samples are independent.*

Godsill and Kndap (2021) recently developed a data-generation technique for the GIG Lévy process. They first constructed a bivariate point process having the GIG Lévy process as its

marginal, and then developed an acceptance-rejection sampling method for the bivariate point process. By contrast, our simulation method is much simpler.

5.3 Goodness-of-Fit Test

To test the goodness of fit, we propose to employ empirical-distribution-function test statistics, e.g., the Kolmogorov-Smirnov test. The idea is to invoke the probability integral transformation and calculate $y_i = F_{X_{\Delta t_i}}(\Delta x_i; \hat{\boldsymbol{\theta}})$ for $i = 1, \ldots, m$. To calculate $\{y_1, \cdots, y_m\}$, we need to be able to approximate $F_{X_t}(x; \boldsymbol{\theta})$ for any $t > 0$. Again, this can be accomplished by employing the saddlepoint method (Lugannani and Rice, 1980; Daniels, 1987):

$$\hat{F}_{X_t}(x; \boldsymbol{\theta}) = \begin{cases} \Phi(z) + \phi(z)\left(\frac{1}{z} - \frac{1}{\hat{u}(x,t)\sqrt{tH''_\lambda(\hat{u}(x,t))}}\right), & \text{for } x \neq \mathrm{E}[X_t], \\ \frac{1}{2} + \frac{tH'''_\lambda(0)}{6\sqrt{2\pi}[tH''_\lambda(0)]^{3/2}}, & \text{for } x = \mathrm{E}[X_t], \end{cases}$$

where $z = \mathrm{sgn}(\hat{u}(x,t))\sqrt{2[\hat{u}(x,t)x - tH_\lambda(\hat{u}(x,t))]}$.

We might assume $\{y_1, \cdots, y_m\}$ are in ascending order. Denote by $\hat{F}_m(y)$ the empirical distribution function of the data $\{y_1, \cdots, y_m\}$. The Kolmogorov-Smirnov statistic is defined by

$$\hat{K}_m^2 = \sqrt{m} \sup_{0<y<1} |\hat{F}_m(y) - y| = \sqrt{m}\ \max\left(\max_{1\leq i\leq m}(\frac{i}{m} - y_i),\ \max_{1\leq i\leq m}(y_i - \frac{i-1}{m})\right);$$

the Cramér-von Mises statistic is defined by

$$\hat{W}_m^2 = m\int_0^1 [\hat{F}_m(y) - y]^2 dy = \sum_{i=1}^m \left(y_i - \frac{i-0.5}{m}\right)^2 + \frac{1}{12m};$$

and the Anderson-Darling (AD) statistic is defined by

$$\hat{A}_m^2 = m\int_0^1 \frac{[\hat{F}_m(y) - y]^2}{y(1-y)} dy = -m - \frac{1}{m}\sum_{i=1}^m (2i-1)\left[\log(y_i) + \log(1 - y_{m+1-i})\right].$$

We then employ the parametric bootstrap technique (Stute et al., 1993) to calculate p-values:

1. For $i = 1, \ldots, m$, draw an observation x_i^* from $\hat{f}_{X_{\Delta t_i}}(x; \hat{\boldsymbol{\theta}})$ via the MH sampler.
2. Compute $\hat{\boldsymbol{\theta}}^* = \arg\max_{\boldsymbol{\theta}\in\Theta} \sum_{i=1}^m \log(\hat{f}_{X_{\Delta t_i}}(x_i^*; \boldsymbol{\theta}))$.
3. Compute $y_i^* = \hat{F}_{X_{\Delta t_i}}(x_i^*; \hat{\boldsymbol{\theta}}^*)$ for $i = 1, \cdots, m$, and then compute the values of the test statistics.
4. Repeat steps 1 to 3 for a large number of times to obtain the corresponding p-values.

6 Numerical Study

6.1 Performance of the Saddlepoint Approximation

We examine the performance of the saddlepoint approximation by approximating $f_{X_2}(x;\boldsymbol{\theta})$, i.e., $t = 2$. To simulate a random value from $f_{X_2}(x;\boldsymbol{\theta})$, we simulate two random values from $f_{X_1}(x;\boldsymbol{\theta})$ and then add them together to obtain a realization of X_2. Repeat in this manner to simulate a dataset with size 100,000 from $f_{X_2}(x;\boldsymbol{\theta})$. Then the kernel density plot of the simulated data provides an accurate graphical representation of $f_{X_2}(x;\boldsymbol{\theta})$. To examine the accuracy of $\hat{f}_{X_2}(x;\boldsymbol{\theta})$, we just need to plot $\hat{f}_{X_2}(x;\boldsymbol{\theta})$ within the kernel density plot, which is illustrated in Figures 4, 5 and 6. In Figures 4-7 and 12, the black curve represents the kernel density estimate, the red curve

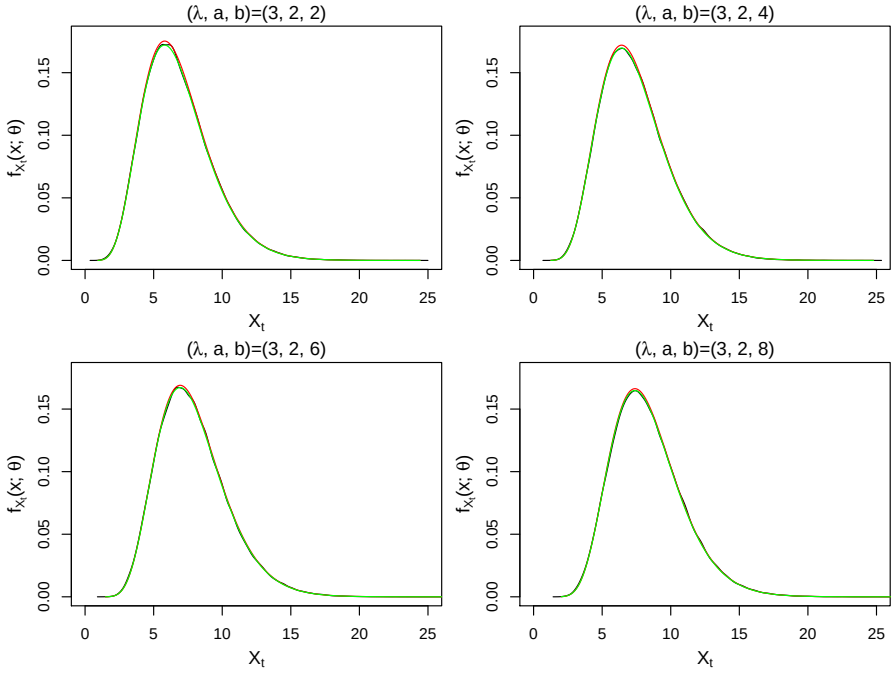

Figure 4: Density plots of $f_{X_2}(x;\boldsymbol{\theta})$ (black), $\hat{f}_{X_2}(x;\boldsymbol{\theta})$ (red) and $\bar{f}_{X_2}(x;\boldsymbol{\theta})$ (green), when b increases.

represents the saddlepoint density $\hat{f}_{X_2}(x;\boldsymbol{\theta})$, and the green curve represents the modified saddlepoint density $\bar{f}_{X_2}(x;\boldsymbol{\theta})$. In Figure 4, we gradually increase b, with the other two parameters being fixed. Similarly, in Figures 5 and 6, we respectively gradually increase a and λ. In every panel, the three density plots are virtually indistinguishable, confirming that the saddlepoint density $\hat{f}_{X_2}(x;\boldsymbol{\theta})$ can accurately approximate the true density $f_{X_2}(x;\boldsymbol{\theta})$. The modified saddlepoint density $\bar{f}_{X_2}(x;\boldsymbol{\theta})$ locates between $\hat{f}_{X_2}(x;\boldsymbol{\theta})$ and $f_{X_2}(x;\boldsymbol{\theta})$, yielding a slightly better approximation than $\hat{f}_{X_2}(x;\boldsymbol{\theta})$.

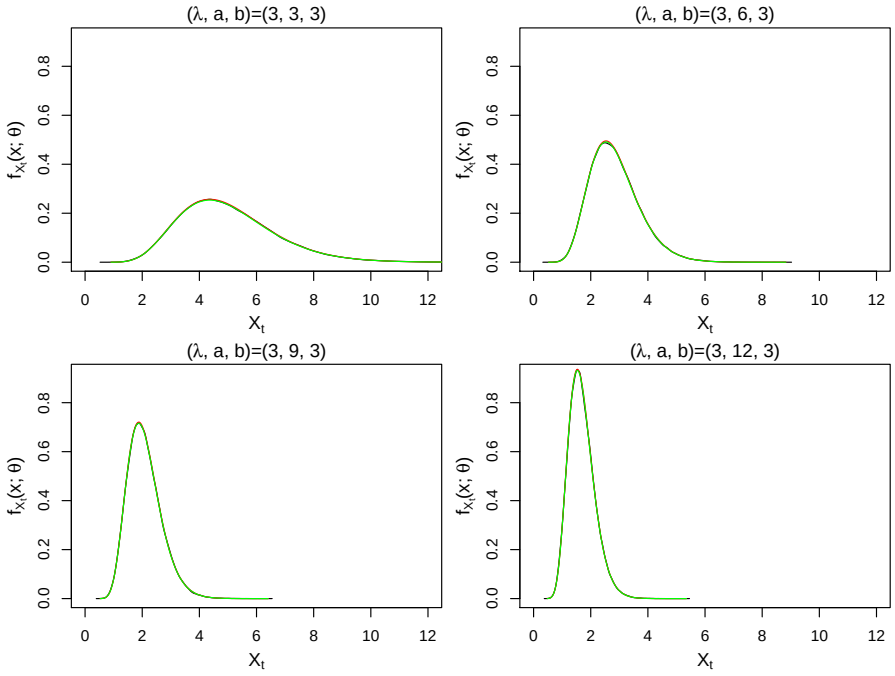

Figure 5: Density plots of $f_{X_2}(x;\boldsymbol{\theta})$ (black), $\hat{f}_{X_2}(x;\boldsymbol{\theta})$ (red) and $\bar{f}_{X_2}(x;\boldsymbol{\theta})$ (green), when a increases.

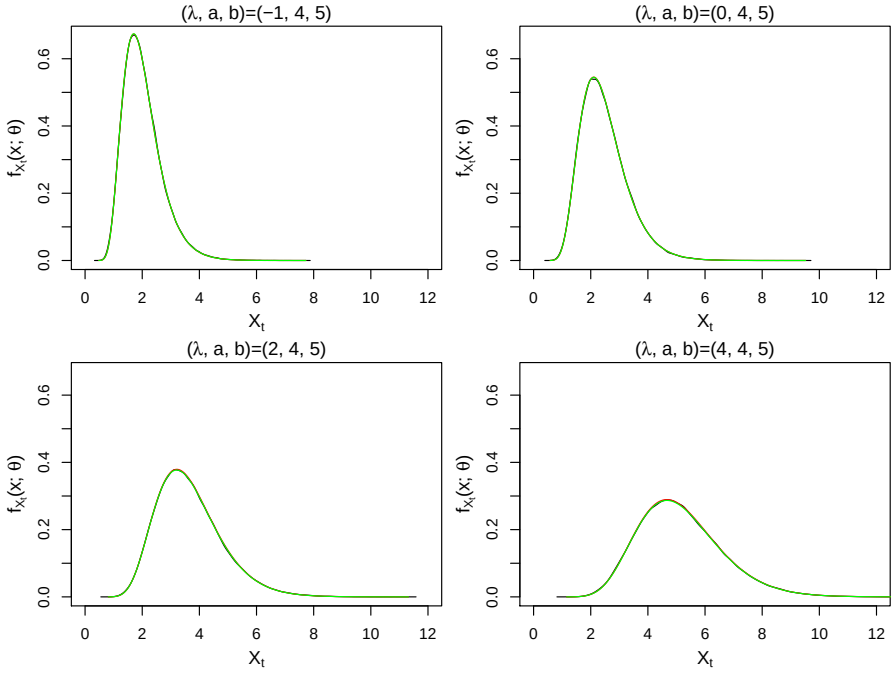

Figure 6: Density plots of $f_{X_2}(x;\boldsymbol{\theta})$ (black), $\hat{f}_{X_2}(x;\boldsymbol{\theta})$ (red) and $\bar{f}_{X_2}(x;\boldsymbol{\theta})$ (green), when λ increases.

We now investigate the impact of t on the performance of the saddlepoint approximation by fixing (λ, a, b) at (2, 2, 6). Likewise, to simulate an observation of X_t, we simulate t observations of X_1 and then add them together. To plot the kernel density estimate, we simulate 100,000 data points for each value of t. Figure 7 shows the results. Again, it is observed that the saddlepoint density

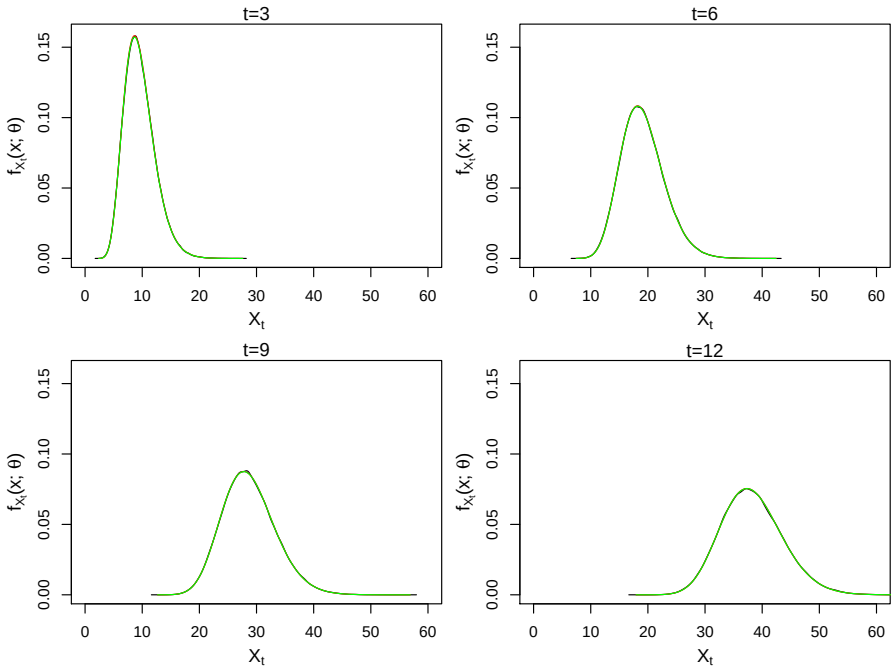

Figure 7: Density plots of $f_{X_t}(x;\boldsymbol{\theta})$ (black), $\hat{f}_{X_t}(x;\boldsymbol{\theta})$ (red) and $\bar{f}_{X_t}(x;\boldsymbol{\theta})$ (green), when t increases.

$\hat{f}_{X_t}(x;\boldsymbol{\theta})$ is fairly accurate for each value of t. Moreover, numerical integration of $\hat{f}_{X_t}(x;\boldsymbol{\theta})$ over the interval $(0,+\infty)$ gives the value 1 (due to round-off error). Figure 12 covers more exhaustive parameter settings. Figures 4-7 and 12 verify the competence of the saddlepoint approximation, which shall greatly simplify the inference procedure of the GIG Lévy process.

6.2 Parameter Estimation

In this section, we examine the feasibility of directly maximizing $\sum_{i=1}^{m}\log(\hat{f}_{X_{\Delta t_i}}(\Delta x_i;\boldsymbol{\theta}))$ for ML estimation. The time series data with size $m = 100$ are simulated as follows. The time increments $\{\Delta t_1, \Delta t_2, \cdots, \Delta t_m\}$ are randomly sampled with replacement from the set $\{1, 2, ..., 9, 10\}$. Then randomly generate Δt_i $(i = 1, \ldots, m)$ samples from the GIG distribution $GIG(\lambda, a, b)$ and set their sum as the realization of the increment Δx_i. A variety of parameter settings are examined: $a \in \{0.5, 1, 3, 5, 7, 9\}$, $b \in \{0.5, 1, 3, 5, 7, 9\}$ and $\lambda \in \{-0.75, -0.25, 0.5, 1, 3, 5, 7, 9\}$. Hence, there are

in total 288 different parameter settings. For each parameter setting, we repeat for 5000 times and hence obtain 5000 ML estimates of the parameter vector (a, b, λ). The relative error of every ML estimate is calculated, which is the difference (between the estimate and the true value) divided by the true value.

We remark that, with any one of the three parameters $\{a, b, \lambda\}$ being known, the other two parameters can be accurately estimated by directly maximizing $\sum_{i=1}^{m} \log(\hat{f}_{X_{\Delta t_i}}(\Delta x_i; \boldsymbol{\theta}))$. We corroborate this statement via the relative-error box plots in Figures 8, 9 and 10.

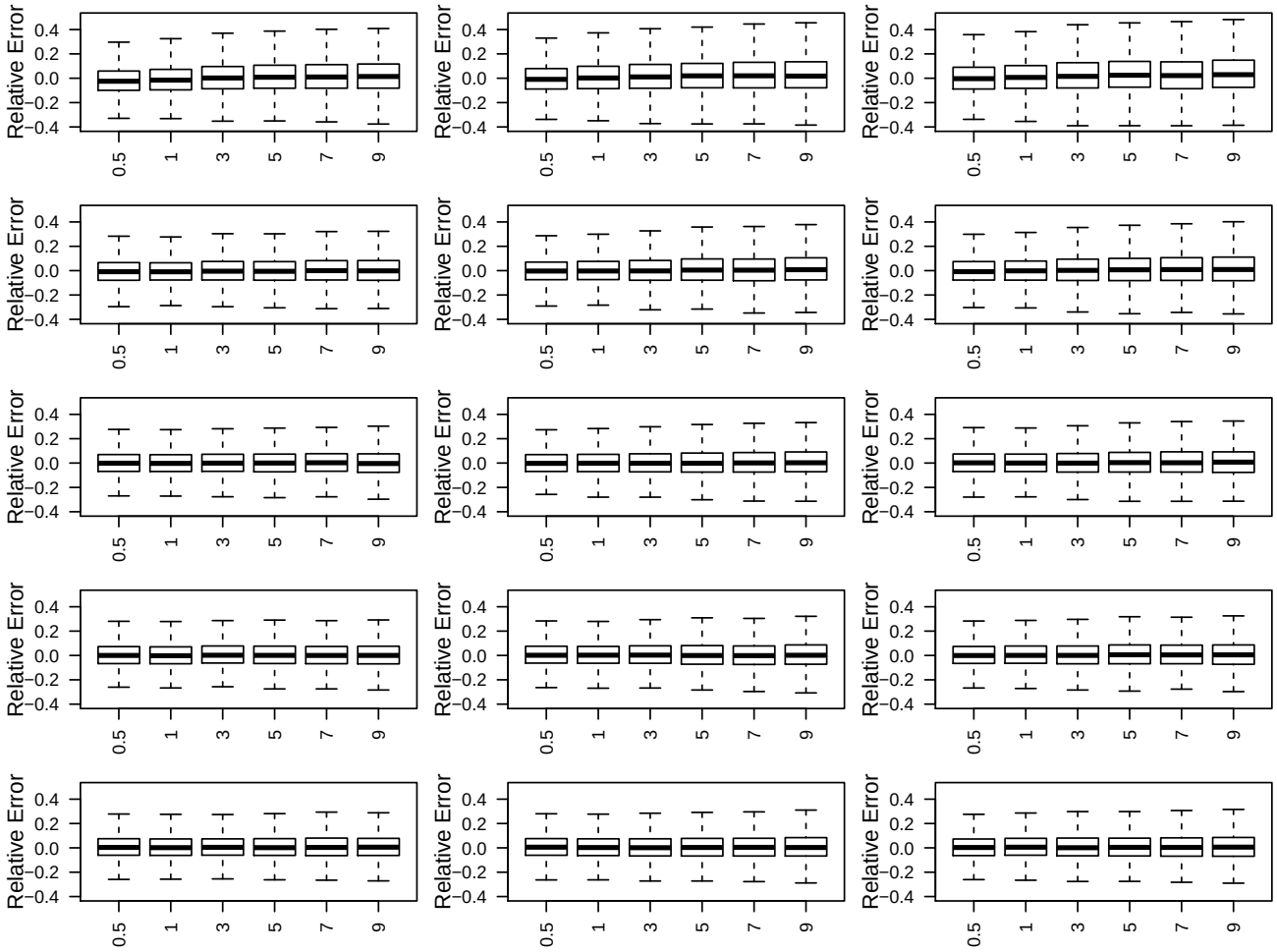

Figure 8: Box plots of the relative errors of the ML estimates $\hat{a}$, with the value of b being known. In each column, λ takes in turn a value from $\{1, 3, 5, 7, 9\}$. In each row, b takes in turn a value from $\{1, 3, 5\}$. In each panel, a takes in turn a value from $\{0.5, 1, 3, 5, 7, 9\}$.

- The box plots in Figure 8 characterize the variation of the relative error $\frac{\hat{a}-a}{a}$, assuming that

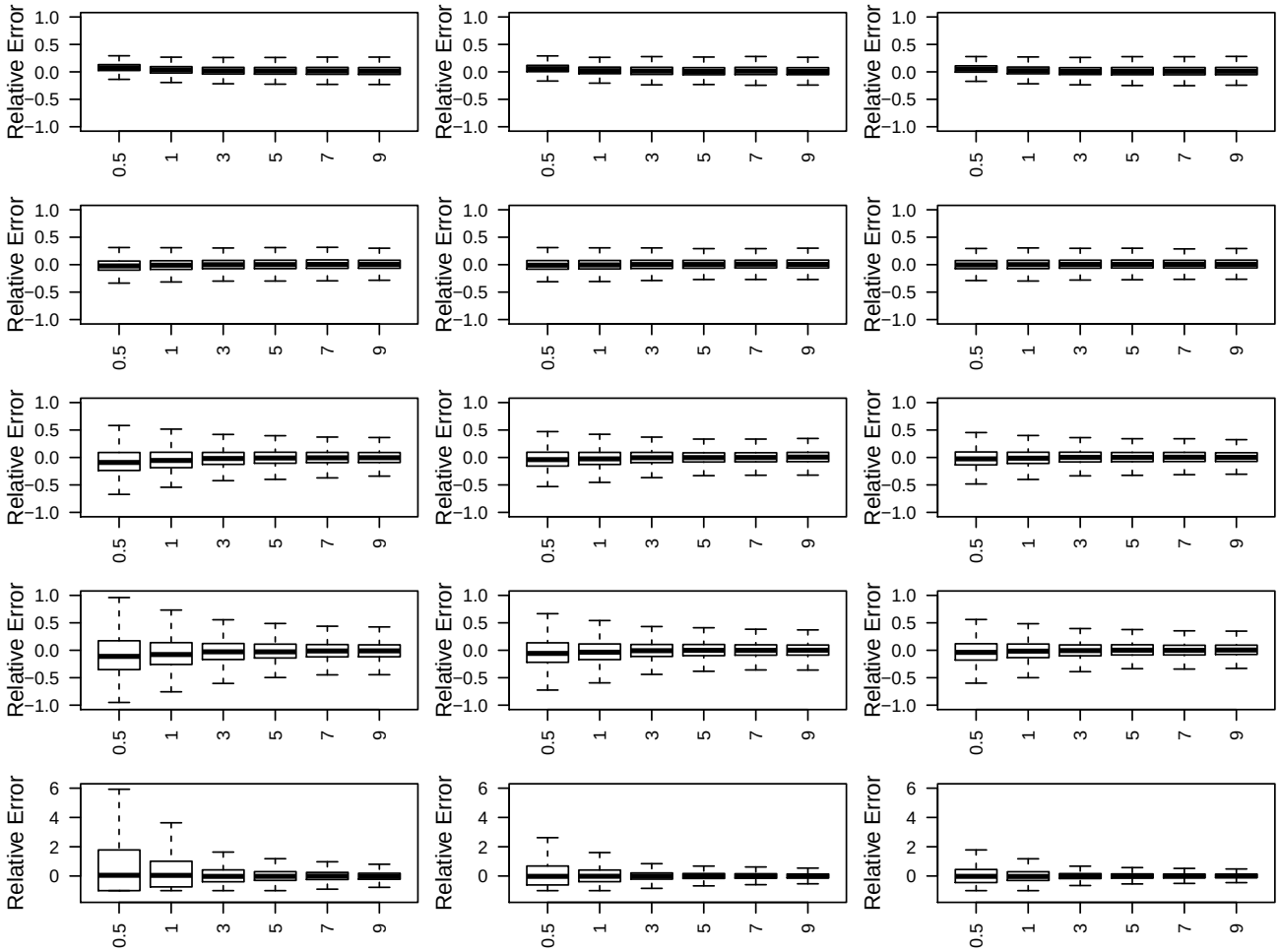

Figure 9: Box plots of the relative errors of the ML estimates $\hat{b}$, with the value of λ being known. In each column, λ takes in turn a value from {-0.75, -0.25, 0.5, 1, 3}. In each row, a takes in turn a value from {1, 3, 5}. In each panel, b takes in turn a value from {0.5, 1, 3, 5, 7, 9}.

the true value of b is known. For the five panels in each column, λ takes in turn a value from {1, 3, 5, 7, 9}. For the three panels in each row, b takes in turn a value from {1, 3, 5}. For the six box plots in each panel, a takes in turn a value from {0.5, 1, 3, 5, 7, 9}. It is observed that, for every box plot in Figure 8, the distance between the 1st quantile and 3rd quantile is quite small. For each combination of the values of the three parameters, with 5000 repetitions, the 5000 ML estimates scatter symmetrically at the two sides of the true parameter value, and the median of the 5000 relative errors is virtually zero.

- The box plots in Figure 9 describe the variation of the relative error $\frac{\hat{b}-b}{b}$, assuming that the

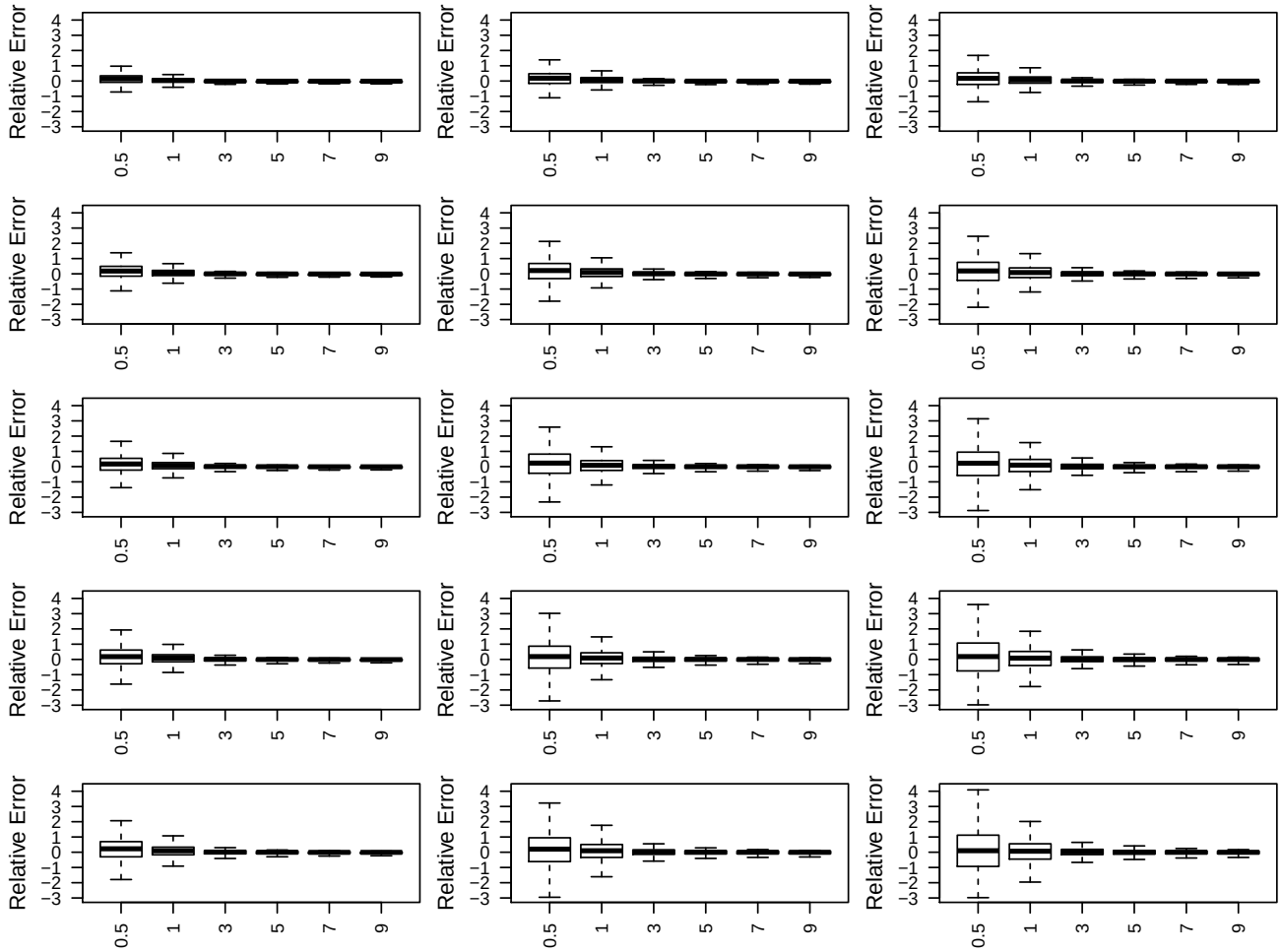

Figure 10: Box plots of the relative errors of the ML estimates $\hat{\lambda}$, with the value of a being known. In each column, a takes in turn a value from $\{1, 3, 5, 7, 9\}$. In each row, b takes in turn a value from $\{1, 3, 5\}$. In each panel, λ takes in turn a value from $\{0.5, 1, 3, 5, 7, 9\}$.

true value of λ is known. For the five panels in each column, λ takes in turn a value from {-0.75, -0.25, 0.5, 1, 3}. For the three panels in each row, a takes in turn a value from {1, 3, 5}. For the six box plots in each panel, b takes in turn a value from $\{0.5, 1, 3, 5, 7, 9\}$. We notice that, with a decreasing and λ increasing, the variance of the relative error increases. (This is because, as verified in Figure 11, when a is small and λ is large, the value of b has little impact on the GIG density function). However, the median of the 5000 relative errors is still close to zero, implying that the ML estimate $\hat{b}$ is unbiased.

- The box plots in Figure 10 describe the variation of the relative error $\frac{\hat{\lambda}-\lambda}{\lambda}$, assuming that the

true value of a is known. For the five panels in each column, a takes in turn a value from $\{1, 3, 5, 7, 9\}$. For the three panels in each row, b takes in turn a value from $\{1, 3, 5\}$. For the six box plots in each panel, λ takes in turn a value from $\{0.5, 1, 3, 5, 7, 9\}$. Again, the 5000 ML estimates $\hat{\lambda}$ scatter symmetrically at the two sides of the true parameter value, the distance between the 1st quantile and 3rd quantile is quite small, and the median of the 5000 relative errors is virtually zero.

Figures 8-10 reveal that, with any one of the three parameters $\{a, b, \lambda\}$ being known, the ML estimates of the other two parameters obtained by directly maximizing $\sum_{i=1}^{m} \log(\hat{f}_{X_{\Delta t_i}}(\Delta x_i; \boldsymbol{\theta}))$ are unbiased and efficient. Therefore, by fixing any of the parameters $\{a, b, \lambda\}$ at an arbitrary value within its domain, we can obtain a new two-parameter Lévy process which, via the saddlepoint technique, can be readily applied to practical problems. In other words, the set of applicable pure-jump increasing Lévy processes has been significantly enriched.

In the general case, when all the three parameters $\{a, b, \lambda\}$ are unknown, the ML estimate $\hat{\boldsymbol{\theta}}$ undoubtedly will have a larger variance. Through comprehensive numerical study, we found that, when $\lambda > 1$, the ML estimate $\hat{\boldsymbol{\theta}}$ obtained by directly maximizing $\sum_{i=1}^{m} \log(\hat{f}_{X_{\Delta t_i}}(\Delta x_i; \boldsymbol{\theta}))$ is unbiased. However, when $\lambda \leq 1$, the ML estimate $\hat{\lambda}$ tends to be larger than the true value, and accordingly the ML estimate $\hat{b}$ tends to be smaller than the true value. We will tackle this problem through felicitous modifications of the saddlepoint approximation, which is left for future work.

7 Conclusions

We uncovered the simplicity of the GIG Lévy process by proving that, when $\lambda \geq -1$, the marginal distribution of the GIG Lévy process admits an explicit form, which is a highly accurate approximation. The availability of the analytic and accurate approximation greatly simplifies the problems of parameter estimation, goodness-of-fit testing, random number generation, etc. Particularly, if any one of the three parameters is known, or if $\lambda > 1$, the unknown parameters can be accurately and efficiently estimated by directly maximizing the saddlepoint-approximation log-likelihood function. Due to the generality of the GIG Lévy process, the set of practicable pure-jump increasing Lévy processes has been significantly enriched. Our continued work on this process will propose a well-grounded modified saddlepoint approximation.

Acknowledgements

This publication has emanated from research supported by a research grant from Science Foundation Ireland (SFI) under grant number 16/RC/3872 and is co-funded under the European Regional Development Fund.

Appendix A Sensitivity Analysis

We here examine the sensitivity of the GIG density function w.r.t. the parameter b. Density plots are given in Figure 11: for the eight panels in each row, λ takes in turn a value from {-0.75, -0.25, 0.5, 1, 3, 5, 7, 9}; for the six panels in each column, a takes in turn a value from {0.5, 1, 3, 5, 7, 9}; in each panel, b takes in turn a value from {0.5, 1, 3, 5, 7, 9}. It is clear from Figure 11 that, when a is small and λ is large, the value of b has little impact on the GIG density function. For example, when $\lambda \geq 5$ and $a \leq 1$, all the density curves for b in {0.5, 1, 3, 5, 7, 9} are virtually the same. Consequently, the estimate of b, produced by any parameter estimation method, will have a large variance. Examination of the sensitivity of the GIG density function w.r.t. the parameter λ reveals that (not shown here) when both a and b are large, a small value of λ (e.g., $\lambda \leq 1$) has little impact on the GIG density curve. To put it briefly, if the condition number of the log-likelihood function $\sum_{i=1}^{m} \log(f_{X_{\Delta_i}}(\Delta x_i; \boldsymbol{\theta}))$ w.r.t. $\boldsymbol{\theta}$ is large, the ML estimate $\hat{\boldsymbol{\theta}}$ will behave erratically.

Corresponding to Figure 11, Figure 12 includes saddlepoint approximations $\hat{f}_{X_1}(x; \boldsymbol{\theta})$ (red) and $\bar{f}_{X_1}(x; \boldsymbol{\theta})$ (green) to $GIG(\lambda, a, b)$, i.e., $f_{X_1}(x; \boldsymbol{\theta})$. In Figure 12, for better visualization, b only takes two values: 0.5 and 9. Figure 12 further verifies the exceptional performance of the saddlepoint approximation, with the modified saddlepoint density $\bar{f}_{X_1}(x; \boldsymbol{\theta})$ yielding a slightly better approximation than $\hat{f}_{X_1}(x; \boldsymbol{\theta})$. According to Figure 12, we can claim that the Kullback-Leibler divergence from $\hat{f}_{X_t}(x; \hat{\boldsymbol{\theta}})$ to $f_{X_t}(x; \boldsymbol{\theta})$ will be trivial, even if the condition number of the log-likelihood function $\sum_{i=1}^{m} \log(f_{X_{\Delta_i}}(\Delta x_i; \boldsymbol{\theta}))$ w.r.t. $\boldsymbol{\theta}$ is large.

References

Abramovitz, M. and Segun, I. (1970). *Handbook of Mathematical Functions*. Dover Publications Inc.

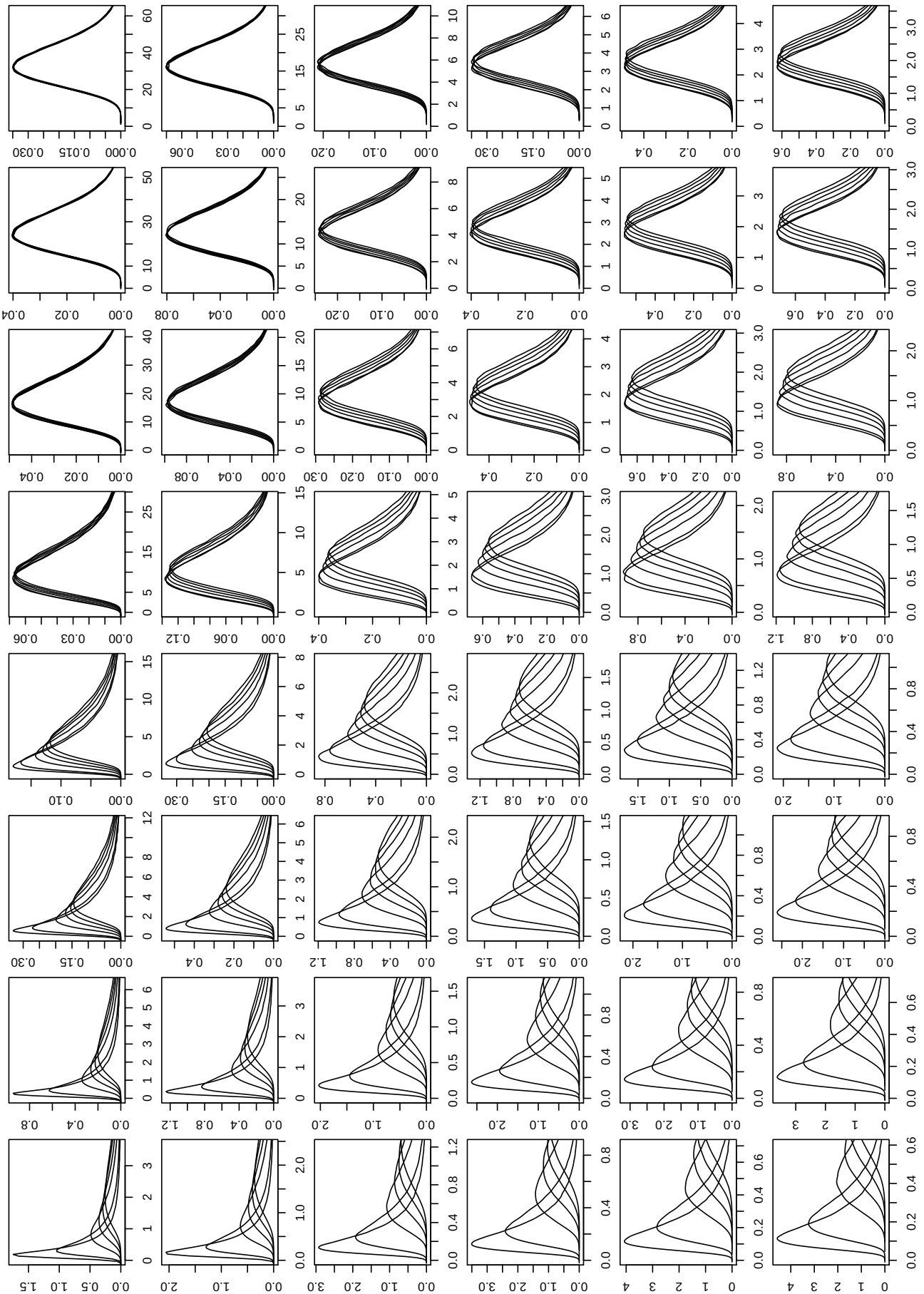

Figure 11: Density plots of $GIG(\lambda, a, b)$. In each row, λ takes in turn a value from $\{-0.75, -0.25, 0.5, 1, 3, 5, 7, 9\}$. In each column, a takes in turn a value from $\{0.5, 1, 3, 5, 7, 9\}$. In each panel, b takes in turn a value from $\{0.5, 1, 3, 5, 7, 9\}$.

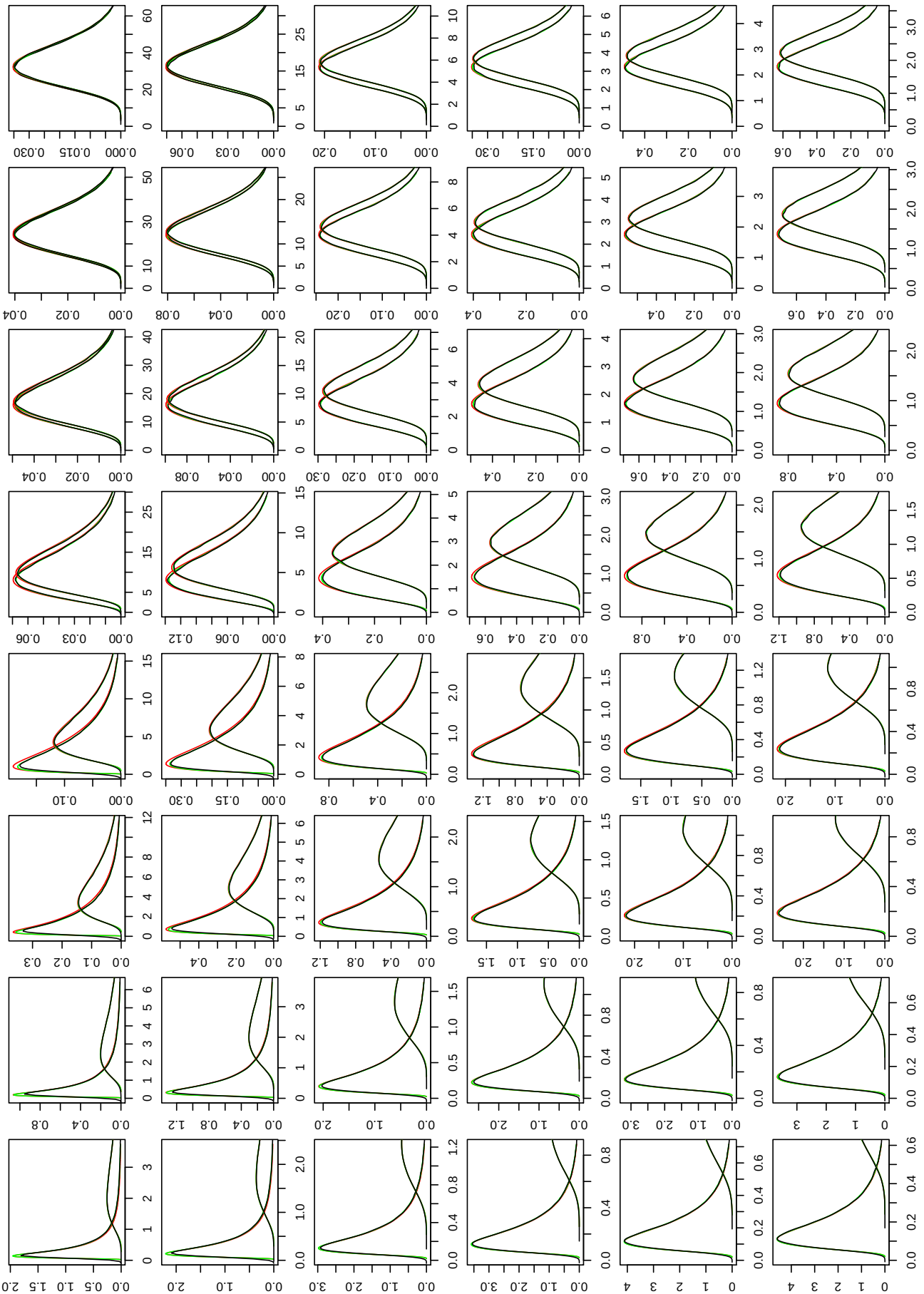

Figure 12: Density plots of $f_{X_1}(x;\boldsymbol{\theta})$ (black), $\hat{f}_{X_1}(x;\boldsymbol{\theta})$ (red) and $\bar{f}_{X_1}(x;\boldsymbol{\theta})$ (green). In each row, λ takes in turn a value from {-0.75, -0.25, 0.5, 1, 3, 5, 7, 9}. In each column, a takes in turn a value from {0.5, 1, 3, 5, 7, 9}. In each panel, b takes values 0.5 and 9.

Ait-Sahalia, Y. and Yu, J. (2006). Saddlepoint approximations for continuous-time markov processes. *Journal of Econometrics*, 134(2):507 – 551.

Alexandrov, M. D. and Lacis, A. A. (2000). A new three-parameter cloud/aerosol particle size distribution based on the generalized inverse gaussian density function. *Applied Mathematics and Computation*, 116(12):153 – 165.

Andrieu, C., de Freitas, N., Doucet, A., and Jordan, M. I. (2003). An introduction to mcmc for machine learning. *Machine Learning*, 50(1):5–43.

Barndorff-Nielsen, O., Blsild, P., and Halgreen, C. (1978). First hitting time models for the generalized inverse gaussian distribution. *Stochastic Processes and their Applications*, 7(1):49 – 54.

Barndorff-Nielsen, O. and Halgreen, C. (1977). Infinite divisibility of the hyperbolic and generalized inverse gaussian distributions. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 38(4):309–311.

Barndorff-Nielsen, O. E. (1997). Normal inverse gaussian distributions and stochastic volatility modelling. *Scandinavian Journal of Statistics*, 24(1):1–13.

Barndorff-Nielsen, O. E. and Shephard, N. (2001). Non-gaussian ornstein-uhlenbeck-based models and some of their uses in financial economics. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(2):167–241.

Carr, P., Geman, H., Madan, D. B., and Yor, M. (2007). Self decomposability and option pricing. *Mathematical Finance*, 17(1):31–57.

Chen, M., Shao, Q., and Ibrahim, J. G. (2012). *Monte Carlo methods in Bayesian computation*. Springer Science & Business Media.

Cholette, M. E., Yu, H., Borghesani, P., Ma, L., and Kent, G. (2019). Degradation modeling and condition-based maintenance of boiler heat exchangers using gamma processes. *Reliability Engineering & System Safety*, 183:184 – 196.

Daniels, H. E. (1954). Saddlepoint approximations in statistics. *The Annals of Mathematical Statistics*, 25(4):631–650.

Daniels, H. E. (1980). Exact saddlepoint approximations. *Biometrika*, 67(1):59–63.

Daniels, H. E. (1987). Tail probability approximations. *International Statistical Review / Revue Internationale de Statistique*, 55(1):37–48.

Dufresne, F., Gerber, H. U., and Shiu, E. S. W. (1991). Risk theory and the gamma process. *ASTIN Bulletin*, 22:177–192.

Embrechts, P. (1983). A property of the generalized inverse gaussian distribution with some applications. *Journal of Applied Probability*, 20(3):537–544.

Godsill, S. and Kndap, Y. (2021). Point process simulation of generalised inverse gaussian processes and estimation of the jaeger integral.

Goutis, C. and Casella, G. (1999). Explaining the saddlepoint approximation. *The American Statistician*, 53(3):216–224.

Halgreen, C. (1979). Self-decomposability of the generalized inverse gaussian and hyperbolic distributions. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 47(1):13–17.

Luciano, E. and Semeraro, P. (2010). A generalized normal mean-variance mixture for return processes in finance. *International Journal of Theoretical and Applied Finance*, 13(3):415–440.

Lugannani, R. and Rice, S. (1980). Saddle point approximation for the distribution of the sum of independent random variables. *Advances in Applied Probability*, 12(2):475–490.

Mengersen, K. L. and Tweedie, R. L. (1996). Rates of convergence of the hastings and metropolis algorithms. *The Annals of Statistics*, 24(1):101–121.

Meyn, S. P. and Tweedie, R. L. (2012). *Markov chains and stochastic stability*. Springer Science & Business Media.

Morales, M. (2004). Risk theory with the generalized inverse gaussian lévy process. *ASTIN Bulletin*, 34:361–377.

Protassov, R. S. (2004). Em-based maximum likelihood parameter estimation for multivariate generalized hyperbolic distributions with fixed λ. *Statistics and Computing*, 14(1):67–77.

Reid, N. (1988). Saddlepoint methods and statistical inference. *Statistical Science*, 3(2):213–227.

Stute, W., Manteiga, W., and Quindimil, M. (1993). Bootstrap based goodness-of-fit-tests. *Metrika*, 40(1):243–256.

Themelis, K. E., Rontogiannis, A. A., and Koutroumbas, K. D. (2016). Variational bayes group sparse time-adaptive parameter estimation with either known or unknown sparsity pattern. *IEEE Transactions on Signal Processing*, 64(12):3194–3206.

Tierney, L. (1994). Markov chains for exploring posterior distributions. *The Annals of Statistics*, 22(4):1701–1728.

Vilca, F., Balakrishnan, N., and Zeller, C. B. (2014). Multivariate skew-normal generalized hyperbolic distribution and its properties. *Journal of Multivariate Analysis*, 128:73 – 85.

Yamazato, M. (1978). Unimodality of infinitely divisible distribution functions of class l. *Annals of Probability*, 6(4):523–531.

Zhou, W., Xiang, W., and Hong, H. (2017). Sensitivity of system reliability of corroding pipelines to modeling of stochastic growth of corrosion defects. *Reliability Engineering & System Safety*, 167:428 – 438.