Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Journal of Econometrics 134 (2006) 507–551
JOURNAL OF Econometrics
www.elsevier.com/locate/jeconom

Saddlepoint approximations for continuous-time Markov processes

Yacine Aït-Sahalia[a,*], Jialin Yu[b]

[a]Department of Economics, Princeton University and NBER, Princeton, NJ 08544, USA
[b]Department of Finance and Economics, Columbia University, New York, NY 10027, USA

Available online 30 August 2005

Abstract

This paper proposes saddlepoint expansions as a means to generate closed-form approximations to the transition densities and cumulative distribution functions of Markov processes. This method is applicable to a large class of models considered in finance, for which a Laplace or characteristic functions, but not the transition density, can be found in closed form. But even when such a computation is not possible explicitly, we go one step further by showing how useful approximations can be obtained by replacing the Laplace or characteristic functions by an expansion in small time.

JEL classification: C13; C22; C32

Keywords: Transition density; Infinitesimal generator; Characteristic function; Closed-form approximation

1. Introduction

Transition densities play a crucial role in continuous time finance, at both the theoretical model building and econometric inference and testing levels: see, the survey by Aït-Sahalia et al. (2002). Arbitrage considerations in finance make many

*Corresponding author. Tel.: +1 609 258 4015; fax: +1 609 258 0719.
E-mail addresses: yacine@princeton.edu (Y. Aït-Sahalia), jy2167@columbia.edu (J. Yu).

doi:10.1016/j.jeconom.2005.07.004

pricing problems linear; as a result, they depend upon the computation of conditional expectations for which knowledge of the transition function is essential. Inference strategies relying on maximum-likelihood or Bayesian methods can be greatly simplified if the transition function is known. Unfortunately, such an object is rarely known in closed form.

This has given rise to the development of a method designed to approximate it in *closed form*: see, Aït-Sahalia (1999, 2001, 2002) for the original work on diffusions. This method is based on expanding the transition density of the process in orthogonal polynomials around a Gaussian leading term. While writing down a Hermite series can be done for any model, the key idea is to exploit the specificity afforded by the diffusion hypothesis in order to obtain the expressions for the coefficients of the series fully explicitly, as functions of the state vectors at the present and future dates, the time interval that separates them and the parameters of the assumed stochastic differential equation.

Jensen and Poulsen (2002) conducted an extensive comparison of different techniques for approximating transition function and demonstrated that the method is both the most accurate and the fastest to implement for the types of problems one encounters in finance. The method has been extended to time inhomogeneous processes by Egorov et al. (2003) and to jump-diffusions by Schaumburg (2001) and Yu (2003). DiPietro (2001) has extended the methodology to make it applicable in a Bayesian setting. Bakshi and Yu (2005) proposed a refinement to the method.

In this paper, we propose an alternative strategy for constructing closed-form approximations to the transition density of a continuous time Markov process. Instead of expanding the transition function in orthogonal polynomials around a leading term, we rely on the *saddlepoint method*, which originates in the work of Daniels (1954). In addition to the transition function, saddlepoint approximations can be usefully applied to compute tail probabilities: see, Rogers and Zane (1999) for an application to option pricing. Jensen (1995) provides a detailed survey of the saddlepoint method.

To understand our approach, it is useful to contrast it with the Hermite-based method in Aït-Sahalia (2002). Consider a standardized sum of random variables to which the central limit theorem (CLT) apply. Often, one is willing to approximate the actual sample size n by infinity and use the $\mathrm{N}(0,1)$ limiting distribution for the properly standardized transformation of the data. If not, higher-order terms of the limiting distribution (for example, the classical Edgeworth expansion based on Hermite polynomials) can be calculated to improve the small sample performance of the approximation. The basic idea is to create an analogy between this situation and that of approximating the transition density of a diffusion. The sampling interval Δ plays the role of the sample size n in the CLT. If we properly standardize the data, we can find out the limiting distribution of the standardized data as Δ tends to 0 (by analogy with what happens in the CLT when n tends to ∞). Properly standardizing the data in the CLT means summing them and dividing by $n^{1/2}$; here, it involves transforming the original process into another one. In both cases, the appropriate standardization makes $\mathrm{N}(0,1)$ the leading term. This $\mathrm{N}(0,1)$ approximation is then refined by correcting for the fact that Δ is not 0 (just like in practical applications of the CLT n is not infinity), i.e., by computing the higher-order terms. So this expansion can be viewed as analogous to a small sample correction to the CLT. As in the CLT case, it is natural to consider higher-order terms based on Hermite polynomials, which are orthogonal with respect to the leading $N(0, 1)$ term. This is an Edgeworth (or Gram–Charlier, depending upon how the terms are gathered) type of expansion.

By contrast, saddlepoint expansions rely on first tilting the original density—transforming it into another one—and then applying an Edgeworth-like expansion to the tilted density. If the tilted density is chosen wisely, the resulting approximation can be quite accurate in the tails, and applicable fairly generally. In order to be able to calculate a saddlepoint approximation, one needs to be able to calculate the Laplace transform or characteristic function of the process of interest. This requirement is a restriction on the applicability of the method, but as we will see, one that is possible to satisfy in many cases in our context of Markov processes. But even when such a computation is not possible explicitly, we go one step further by showing how useful approximations can be obtained by replacing the characteristic function by an expansion in small time. Expansions in small time, which involve the infinitesimal generator of the Markov process, are a key element shared with the Hermite-based expansions described above.

The paper is organized as follows. We start in Section 2 with some brief preliminaries on Markov processes, focusing on their transition properties, and on characteristic functions. We then present in Section 3 the main results of the saddlepoint theory and then specialize its results to our context. We first construct saddlepoint approximations for diffusion processes in Section 4 and give examples in Section 5. Next, we turn to jump-diffusions and give examples in Section 6. Finally, we present results for more general Lévy processes and give examples in Sections 7. Section 8 concludes. All proofs are in Appendices.

2. Preliminaries

We consider a Markov process, specified in continuous time. The premise of this paper is that the characteristic function of the process is tractable, often known in closed form, while the transition function of the process over discrete time intervals is not. Yet we are interested in the transition function for a variety of purposes, ranging from likelihood inference about the parameters of the model (where the discrete time interval is the data sampling interval) to option pricing (where the discrete time interval is the option's time to expiration).

2.1. The transition density of a Markov process

In what follows, we consider a stochastic process X on $\mathbb{R}^m$, with the standard probability space $(\Omega, \mathscr{F}, \Pr)$ and σ-field of Borelians $\mathscr{E}$ in $\mathbb{R}^m$ such that for each $t \geqslant 0$, $X_t : (\Omega, \mathscr{F}, \Pr) \to (E, \mathscr{E})$ is a measurable function. Consider the family of (time-homogeneous) conditional probability functions $P : (\mathbb{R}_+ \times \mathscr{E} \times \mathbb{R}^m) \to [0, \infty]$, where

$P(\Delta, \cdot|x)$ is a probability measure in $\mathscr{E}$ for each x in $\mathbb{R}^m$ and time interval $\Delta \geqslant 0$ with the limit $P(\Delta, \cdot|x) \to \delta$ as $\Delta \to 0$. A transition function is a family of conditional probability functions that satisfies for each (Δ, τ) in $\mathbb{R}_+^2$ the Chapman–Kolmogorov equation

$$P(\Delta + \tau, B|x) = \int P(\Delta, B|y)P(\tau, \mathrm{d}y|x). \tag{2.1}$$

Let now $\mathscr{F}_t \subset \mathscr{F}$ be an increasing family of σ-algebras and X a stochastic process that is adapted to $\mathscr{F}_t$. X is Markov with transition function P if for each non-negative, Borel measurable $\psi : \mathbb{R}^m \to \mathbb{R}$ and each $(s, t) \in \mathbb{R}_+^2, s < t$

$$\mathrm{E}[\psi(X_t)|\mathscr{F}_s] = \int \psi(y)P(t - s, \mathrm{d}y|X_s). \tag{2.2}$$

Kolmogorov's construction (for example, Revuz and Yor, 1994, Chapter III, Theorem 1.5) allows one to parameterize Markov processes using transition functions. Namely, given a transition function P and a probability measure v on $\mathbb{R}^m$ serving as the initial distribution, there exists a unique probability measure such that the coordinate process X is Markovian with respect to $\sigma(X_u, u \leqslant t)$, has transition function P, and X_0 has v as its distribution.

2.2. Characteristic functions

In what follows, we will assume that for each $(\Delta, x) \in \mathbb{R}_+ \times \mathbb{R}^m$ the probability measure $P(\Delta, \cdot|x)$ admits a probability density $p(\Delta, \cdot|x)$ with respect to Lebesgue measure. The (conditional) Laplace transform of the process X is defined for $u \in \mathbb{R}^m$ as

$$\varphi(\Delta, u|x) = \mathrm{E}[\exp(u^{\mathrm{T}} X_\Delta)|X_0 = x], \tag{2.3}$$

where $^{\mathrm{T}}$ denotes transposition, for all values of (Δ, u, x). We assume that for each (Δ, x), the region of convergence in $\mathbb{R}^m$ of the integral $\varphi(\Delta, u|x)$ (i.e., the set Λ of $u \in \mathbb{R}^m$ such that $\varphi(\Delta, u|x) < \infty$) is a product of m non-vanishing intervals which contain the origin. That is, $\Lambda = \prod_{i=1}^m \Lambda_i$, with $\Lambda_i = (-c_i, d_i)$, $c_i \geqslant 0$, $d_i \geqslant 0$ and $c_i + d_i > 0$. Note that the existence of the density p insures that $\varphi(\Delta, u|x) \to 0$ as $|u| \to \infty$.

The cumulant transform (or cumulant generating function) of X is the function

$$K(\Delta, u|x) = \ln \varphi(\Delta, u|x). \tag{2.4}$$

For given (Δ, x), K is a closed-convex function of u in $\mathbb{R}^m$ and if the variance matrix $\mathrm{Var}(X)$ is positive definite, K is strictly convex on Λ. Note that $K(\Delta, 0|x) = 0$.

Derivatives of all order of φ with respect to u exist and are given by

$$\frac{\partial^{r_1 + \cdots + r_m} \varphi(\Delta, u|x)}{\partial u_1^{r_1} \ldots \partial u_m^{r_m}} = \mathrm{E}[X_{1\Delta}^{r_1} \ldots X_{m\Delta}^{r_m} \exp(u \cdot X_\Delta)|X_0 = x].$$

The characteristic function of X is the function $\varphi(\Delta, \mathbf{i}u|x)$, where $\mathbf{i}^2 = -1$. The cumulants of X are the arrays $\kappa_r \equiv \{\kappa_{r_1,\ldots,r_m} : r_1 + \cdots + r_m = r\}$, where

$$\kappa_{r_1,\ldots,r_m} = \left.\frac{\partial^{r_1+\cdots+r_m} K(\Delta, u|x)}{\partial u_1^{r_1} \ldots \partial u_m^{r_m}}\right|_{u=0}$$

and we leave the dependence on (Δ, x) implicit. In particular,

$$\kappa_2 = \left.\frac{\partial^2 K(\Delta, u|x)}{\partial u \partial u^{\mathrm{T}}}\right|_{u=0}$$

is an $m \times m$ matrix.

We will use a number of facts about characteristic functions which can be applied in our context of Markovian transition functions. In particular, the density function and characteristic functions are linked by the Fourier inversion formula

$$\begin{aligned} p(\Delta, y|x) &= (2\pi)^{-m} \int_{-\infty}^{+\infty} \exp(-\mathbf{i}u \cdot y)\varphi(\Delta, \mathbf{i}u|x)\,\mathrm{d}u \\ &= (2\pi\mathbf{i})^{-m} \int_{\hat{u}-\mathbf{i}\infty}^{\hat{u}+\mathbf{i}\infty} \exp(-u \cdot y)\varphi(\Delta, u|x)\,\mathrm{d}u \end{aligned} \tag{2.5}$$

provided that $\varphi(\Delta, \cdot|x) \in L^1(\mathbb{R}^m)$. The vector $\hat{u} \in \mathbb{R}^m$ in the integration limits is chosen in such a way that $-c_i < \hat{u}_i < d_i$ so that $-c_i < \mathrm{Re}(u_i) < d_i$ on the path of integration.

3. The saddlepoint method

We now provide a brief review of the saddlepoint method, which originated with Daniels (1954), before specializing the results to our context of Markov transition densities. The new aspect introduced by Markov processes is the dynamic dimension, i.e., the dependence of all quantities on Δ and the analysis of their asymptotic behavior in small time, i.e., as Δ gets small.

3.1. Expansion of the cumulant generating function

The key to the saddlepoint method is to choose the path of integration, i.e., $\hat{u}$ in (2.5), well. Consider the following choice: set $\hat{u} = \hat{u}(\Delta, y|x) \in \mathbb{R}^m$ as the solution in u of the equation

$$\frac{\partial K(\Delta, u|x)}{\partial u} = y. \tag{3.1}$$

Such a solution exists and is unique because of the convexity in u of the function K. Often, the solution $\hat{u}$ to Eq. (3.1) must be computed numerically either because the function K is not known explicitly or because Eq. (3.1) is too involved: see, Phillips (1986) for a discussion of this issue in the context of large deviations, Easton and

Ronchetti (1986) in the context of L-statistics and Lieberman (1994) for an approach involving replacing $\hat{u}$ by an approximation.

Despite this difficulty, why does the selection of $\hat{u}$ of Eq. (3.1) make sense? Rewrite Eq. (2.5) in the form

$$p(\Delta, y|x) = (2\pi\mathbf{i})^{-m} \int_{\hat{u}-\mathbf{i}\infty}^{\hat{u}+\mathbf{i}\infty} \exp(K(\Delta, u|x) - u \cdot y)\, \mathrm{d}u \tag{3.2}$$

and Taylor expand the function $u \longmapsto K(\Delta, u|x) - u \cdot y$ around its minimum $\hat{u}$:

$$K(\Delta, u|x) - u \cdot y = K(\Delta, \hat{u}|x) - \hat{u} \cdot y + \frac{1}{2}(u - \hat{u})^{\mathrm{T}} \frac{\partial^2 K(\Delta, \hat{u}|x)}{\partial u \partial u^{\mathrm{T}}}(u - \hat{u}) + \mathrm{O}(\|u - \hat{u}\|^3). \tag{3.3}$$

Incidentally, the name "saddlepoint" comes from the shape of the right-hand side of (3.3) for u in a neighborhood of $\hat{u}$. Consider the univariate case $m = 1$. If $u - \hat{u}$ is the complex number $c = a + \mathbf{i}b$, then the real part of the right-hand side of (3.3) is of the form $(a, b) \longmapsto \alpha + \beta(a^2 - b^2)$, where α and β are real-valued constants. This function has the shape of a saddle.

On the path of integration relevant for (3.2), we have $u = \hat{u} + \mathbf{i}v$ with $v \in \mathbb{R}^m$, hence $u - \hat{u}$ is a purely imaginary complex vector. Thus,

$$K(\Delta, u|x) - u \cdot y = K(\Delta, \hat{u}|x) - \hat{u} \cdot y - \frac{1}{2} v^{\mathrm{T}} \frac{\partial^2 K(\Delta, \hat{u}|x)}{\partial u \partial u^{\mathrm{T}}} v + \mathrm{O}(\|v\|^3).$$

From this it follows that the leading term of an approximation to $p(\Delta, y|x)$ can be taken to be

$$p^{(0)}(\Delta, y|x) = (2\pi)^{-m} \exp(K(\Delta, \hat{u}|x) - \hat{u} \cdot y) \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2} v^{\mathrm{T}} \frac{\partial^2 K(\Delta, \hat{u}|x)}{\partial u \partial u^{\mathrm{T}}} v\right) \mathrm{d}v.$$

Then, we have

$$(2\pi)^{-m} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2} v^{\mathrm{T}} \frac{\partial^2 K(\Delta, \hat{u}|x)}{\partial u \partial u^{\mathrm{T}}} v\right) \mathrm{d}v = (2\pi)^{-m/2} \det\left(\frac{\partial^2 K(\Delta, \hat{u}|x)}{\partial u \partial u^{\mathrm{T}}}\right)^{-1/2}$$

from which it follows that

$$p^{(0)}(\Delta, y|x) = (2\pi)^{-m/2} \det\left(\frac{\partial^2 K(\Delta, \hat{u}|x)}{\partial u \partial u^{\mathrm{T}}}\right)^{-1/2} \exp(K(\Delta, \hat{u}|x) - \hat{u} \cdot y). \tag{3.4}$$

This leading term, while everywhere positive, will not in general integrate to 1. Dividing $p^{(0)}(\Delta, y|x)$ in (3.4) by its integral over y insures the latter.

3.2. Higher-order expansions

Higher-order saddlepoint expansions can be obtained by expanding the function $K(\Delta, u|x) - u \cdot y$ around $\hat{u}$ at a higher order. To avoid complicating matters with the tensor notation for multivariate series (see, McCullagh, 1987), we write this expansion in the univariate case

$$K(\Delta,u|x)-u\cdot y = K(\Delta,\hat{u}|x)-\hat{u}\cdot y-\frac{1}{2}\frac{\partial^2 K(\Delta,\hat{u}|x)}{\partial u^2}v^2 \\ -\frac{1}{6}\frac{\partial^3 K(\Delta,\hat{u}|x)}{\partial u^3}\mathbf{i}v^3+\frac{1}{24}\frac{\partial^4 K(\Delta,\hat{u}|x)}{\partial u^4}v^4+\mathrm{O}(v^5) \tag{3.5}$$

from which it follows that

$$\exp(K(\Delta,u|x)-u\cdot y) = \exp(K(\Delta,\hat{u}|x)-\hat{u}\cdot y)\exp\left(-\frac{1}{2}\frac{\partial^2 K(\Delta,\hat{u}|x)}{\partial u^2}v^2\right) \\ \times\left\{1-\frac{1}{6}\frac{\partial^3 K(\Delta,\hat{u}|x)}{\partial u^3}\mathbf{i}v^3+\frac{1}{24}\frac{\partial^4 K(\Delta,\hat{u}|x)}{\partial u^4}v^4 \right.\\ \left.-\frac{1}{72}\left(\frac{\partial^3 K(\Delta,\hat{u}|x)}{\partial u^3}\right)^2 v^6+\mathrm{O}(v^7)\right\},$$

where the last term comes from the quadratic term in expanding $\mathrm{e}^x=1+x+\frac{1}{2}x^2+\mathrm{O}(x^3)$.

The leading term, called the saddlepoint approximation, is the term

$$p^{(0)}(\Delta,y|x)=\frac{\exp(K(\Delta,\hat{u}|x)-\hat{u}\cdot y)}{\sqrt{2\pi}(\partial^2 K(\Delta,\hat{u}|x)/\partial u^2)^{1/2}}. \tag{3.6}$$

Next, stopping the expansion (3.5) at order 4 in v, we have that

$$\begin{aligned} &p^{(1)}(\Delta,y|x)\\ &=(2\pi)^{-1}\exp(K(\Delta,\hat{u}|x)-\hat{u}\cdot y)\int_{-\infty}^{+\infty}\exp\left(-\frac{1}{2}\frac{\partial^2 K(\Delta,\hat{u}|x)}{\partial u^2}v^2\right)\\ &\quad\times\left\{1-\frac{1}{6}\frac{\partial^3 K(\Delta,\hat{u}|x)}{\partial u^3}\mathbf{i}v^3+\frac{1}{24}\frac{\partial^4 K(\Delta,\hat{u}|x)}{\partial u^4}v^4-\frac{1}{72}\left(\frac{\partial^3 K(\Delta,\hat{u}|x)}{\partial u^3}\right)^2 v^6\right\}\mathrm{d}v\\ &=\frac{\exp(K(\Delta,\hat{u}|x)-\hat{u}\cdot y)}{2\pi(\partial^2 K(\Delta,\hat{u}|x)/\partial u^2)^{1/2}}\left\{\int_{-\infty}^{+\infty}\exp\left(-\frac{1}{2}w^2\right)\mathrm{d}w\right.\\ &\quad-\frac{1}{6}\left(\frac{\partial^2 K(\Delta,\hat{u}|x)}{\partial u^2}\right)^{-3/2}\frac{\partial^3 K(\Delta,\hat{u}|x)}{\partial u^3}\mathbf{i}\int_{-\infty}^{+\infty}\exp\left(-\frac{1}{2}w^2\right)w^3\,\mathrm{d}w\\ &\quad+\frac{1}{24}\left(\frac{\partial^2 K(\Delta,\hat{u}|x)}{\partial u^2}\right)^{-2}\frac{\partial^4 K(\Delta,\hat{u}|x)}{\partial u^4}\int_{-\infty}^{+\infty}\exp\left(-\frac{1}{2}w^2\right)w^4\,\mathrm{d}w\\ &\quad\left.-\frac{1}{72}\left(\frac{\partial^2 K(\Delta,\hat{u}|x)}{\partial u^2}\right)^{-3}\left(\frac{\partial^3 K(\Delta,\hat{u}|x)}{\partial u^3}\right)^2\int_{-\infty}^{+\infty}\exp\left(-\frac{1}{2}w^2\right)w^6\,\mathrm{d}w\right\} \end{aligned}$$

so that

$$p^{(1)}(\Delta, y|x) = \frac{\exp(K(\Delta, \hat{u}|x) - \hat{u} \cdot y)}{\sqrt{2\pi}(\partial^2 K(\Delta, \hat{u}|x)/\partial u^2)^{1/2}} \left\{ 1 + \frac{1}{8} \left(\frac{\partial^2 K(\Delta, \hat{u}|x)}{\partial u^2} \right)^{-2} \frac{\partial^4 K(\Delta, \hat{u}|x)}{\partial u^4} - \frac{5}{24} \left(\frac{\partial^2 K(\Delta, \hat{u}|x)}{\partial u^2} \right)^{-3} \left(\frac{\partial^3 K(\Delta, \hat{u}|x)}{\partial u^3} \right)^2 \right\}, \tag{3.7}$$

where we have used the change of variable $w = (\partial^2 K(\Delta, \hat{u}|x)/\partial u^2)^{1/2} v$. The result then follows from the facts that

$$\int_{-\infty}^{+\infty} \exp\left(-\frac{1}{2}w^2\right) \mathrm{d}w = \sqrt{2\pi}, \quad \int_{-\infty}^{+\infty} \exp\left(-\frac{1}{2}w^2\right) w^3 \, \mathrm{d}w = 0,$$

$$\int_{-\infty}^{+\infty} \exp\left(-\frac{1}{2}w^2\right) w^4 \, \mathrm{d}w = 3\sqrt{2\pi}, \quad \int_{-\infty}^{+\infty} \exp\left(-\frac{1}{2}w^2\right) w^6 \, \mathrm{d}w = 15\sqrt{2\pi}.$$

In the case of diffusions, for instance, we shall see that both correction terms $(\partial^2 K(\Delta, \hat{u}|x)/\partial u^2)^{-2} \, \partial^4 K(\Delta, \hat{u}|x)/\partial u^4$ and $(\partial^2 K(\Delta, \hat{u}|x)/\partial u^2)^{-3} \, (\partial^3 K(\Delta, \hat{u}|x)/\partial u^3)^2$ are of order $O(\Delta)$ as shown in the proof of Theorem 3. A higher-order expansion, $p^{(2)}(\Delta, y|x)$, can be obtained similarly. To simplify the notation, let $K_{(n)}$ denote $\partial^{(n)} K(\Delta, \hat{u}|x)/\partial u^n$ so that

$$K(\Delta, u|x) - u \cdot y = K(\Delta, \hat{u}|x) - \hat{u} \cdot y - \tfrac{1}{2} K_{(2)} v^2 - \tfrac{1}{6} K_{(3)} \mathbf{i} v^3 + \tfrac{1}{24} K_{(4)} v^4 + \tfrac{1}{120} K_{(5)} \mathbf{i} v^5 - \tfrac{1}{720} K_{(6)} v^6 + O(\|v\|^7).$$

We will see in the proof of Theorem 3 that, for diffusions, we have

$$p^{(2)}(\Delta, y|x) = \frac{\exp(K(\Delta, \hat{u}|x) - \hat{u} \cdot y)}{\sqrt{2\pi} K_{(2)}^{1/2}} \left\{ 1 + \frac{1}{8} \frac{K_{(4)}}{K_{(2)}^2} - \frac{5}{24} \frac{K_{(3)}^2}{K_{(2)}^3} - \frac{1}{48} \frac{K_{(6)}}{K_{(2)}^3} + \frac{35}{384} \frac{K_{(4)}^2}{K_{(2)}^4} + \frac{7}{48} \frac{K_{(3)} K_{(5)}}{K_{(2)}^4} - \frac{35}{64} \frac{K_{(3)}^2 K_{(4)}}{K_{(2)}^5} + \frac{385}{1152} \frac{K_{(3)}^4}{K_{(2)}^6} \right\}. \tag{3.8}$$

Different Markov processes will differ in that their corresponding $K_{(n)}$ can be of different orders in Δ.

3.3. Non-Gaussian leading terms

It is possible to use a local approximation, in the neighborhood of the saddlepoint $\hat{u}$, different from (3.3). Any cumulant generating function K_0 that is analytic at 0 can serve that role. A non-Gaussian choice can be useful, in particular, when focusing on the small time behavior of jump processes (more on that later in Sections 6 and 7).

To understand how this works, it is useful to first derive the classical saddlepoint approximation using a change of variable inside the integral (3.2). Consider for simplicity the univariate case. The idea is to find an approximation for the function $K(\Delta, u|x) - uy$, that is, valid *both* when u is in a neighborhood of $\hat{u}$ and when u is in a neighborhood of 0. The function vanishes at 0, while its first derivative vanishes at $\hat{u}$ (recalling from (3.1) that $y = \partial K(\Delta, \hat{u}|x)/\partial u$). Let us approximate the quadratic behavior of $K(\Delta, u|x) - u \cdot y$ near $\hat{u}$, given by (3.3), by the same quadratic behavior of w near some $\hat{w}$. That is,

$$\{K(\Delta, u|x) - uy\} - \{K(\Delta, \hat{u}|x) - \hat{u}y\} = \tfrac{1}{2}(w - \hat{w})^2$$

for an arbitrary real $\hat{w}$. In particular, suppose that we select $\hat{w}$ to satisfy

$$K(\Delta, \hat{u}|x) - \hat{u}y = -\tfrac{1}{2}\hat{w}^2, \tag{3.9}$$

that is,

$$\hat{w} = \{2(\hat{u}y - K(\Delta, \hat{u}|x))\}^{1/2}\,\mathrm{sgn}(\hat{u}), \tag{3.10}$$

where $\mathrm{sgn}(\hat{u}) = -1, 0$ or $+1$ depending upon whether $\hat{u}$ is negative, zero or positive. Then, we obtain

$$K(\Delta, u|x) - uy = \tfrac{1}{2}w^2 - w\hat{w} \tag{3.11}$$

and the minima on the left- and right-hand side are indeed identical. Eq. (3.11) defines implicitly the change of variable $u(w)$.

Applying this change of variable in the inverse Fourier representation (3.2) for the density yields

$$\begin{aligned} p(\Delta, y|x) &= (2\pi\mathbf{i})^{-1} \int_{\hat{u}-\mathbf{i}\infty}^{\hat{u}+\mathbf{i}\infty} \exp(K(\Delta, u|x) - u \cdot y)\,\mathrm{d}u \\ &= (2\pi\mathbf{i})^{-1} \int_{\widehat{w}-\mathbf{i}\infty}^{\widehat{w}+\mathbf{i}\infty} \exp\left(\frac{1}{2}w^2 - w\hat{w}\right) \frac{\mathrm{d}u(w)}{\mathrm{d}w}\,\mathrm{d}w, \end{aligned}$$

where the term $\mathrm{d}u(w)/\mathrm{d}w$ can now be approximated near $w = \hat{w}$ as $\mathrm{d}u(w)/\mathrm{d}w = \mathrm{d}u(\hat{w})/\mathrm{d}w + \mathrm{O}(w - \hat{w})$. To compute $\mathrm{d}u(\hat{w})/\mathrm{d}w$, we differentiate both sides of (3.11) with respect to w, once then twice:

$$\frac{\partial K(\Delta, u(w)|x)}{\partial u} \frac{\mathrm{d}u(w)}{\mathrm{d}w} - \frac{\mathrm{d}u(w)}{\mathrm{d}w} y = w - \hat{w},$$

$$\frac{\partial^2 K(\Delta, u(w)|x)}{\partial u^2} \left(\frac{\mathrm{d}u(w)}{\mathrm{d}w}\right)^2 + \frac{\partial K(\Delta, u(w)|x)}{\partial u} \frac{\mathrm{d}^2 u(w)}{\mathrm{d}w^2} - \frac{\mathrm{d}^2 u(w)}{\mathrm{d}w^2} y = 1.$$

Evaluating this at $w = \hat{w}$, and recalling that $u(\hat{w}) = \hat{u}$ and $\partial K(\Delta, \hat{u}|x)/\partial u = y$ gives

$$\frac{\mathrm{d}u(\hat{w})}{\mathrm{d}w} = \left(\frac{\partial^2 K(\Delta, \hat{u}|x)}{\partial u^2}\right)^{-1/2}. \tag{3.12}$$

Thus, we obtain as the leading term

$$\begin{aligned} p^{(0)}(\Delta, y|x) &= (2\pi\mathbf{i})^{-1} \int_{\widehat{w}-\mathbf{i}\infty}^{\widehat{w}+\mathbf{i}\infty} \exp\left(\frac{1}{2}w^2 - w\hat{w}\right) \mathrm{d}w \frac{\mathrm{d}u(\hat{w})}{\mathrm{d}w} \\ &= \phi(\hat{w})\left(\frac{\partial^2 K(\Delta, \hat{u}|x)}{\partial u^2}\right)^{-1/2}, \end{aligned} \tag{3.13}$$

where ϕ is the standard normal density function. It is easy to check that (3.13) coincides with (3.6). Indeed, from (3.9)

$$\begin{aligned} \phi(\hat{w})\left(\frac{\partial^2 K(\Delta, \hat{u}|x)}{\partial u^2}\right)^{-1/2} &= (2\pi)^{-1/2}\left(\frac{\partial^2 K(\Delta, \hat{u}|x)}{\partial u^2}\right)^{-1/2} \exp\left(-\frac{1}{2}\hat{w}^2\right) \\ &= (2\pi)^{-1/2}\left(\frac{\partial^2 K(\Delta, \hat{u}|x)}{\partial u \partial u^{\mathrm{T}}}\right)^{-1/2} \exp(K(\Delta, \hat{u}|x) - \hat{u}y). \end{aligned}$$

To allow for a non-Gaussian leading term, we replace the cumulant generating function $w^2/2$ with another choice K_0, which is also analytic at 0. Let F_0 (resp., f_0) be the explicit cumulative distribution function (resp., density function) corresponding to K_0. The idea is to again approximate the quadratic behavior of $u \longmapsto (\Delta, u|x) - uy$ near $\hat{u}$, but this time with a function involving the base K_0. Consider a fixed point y_0 and its associated saddlepoint $\hat{w}$ for K_0, which by definition solves $K_0'(\hat{w}) = y_0$. The function $w \longmapsto K_0(w) - wy_0$ is quadratic near $\hat{w}$ and should therefore approximate the original function $u \longmapsto K(\Delta, u|x) - uy$ near $\hat{u}$:

$$\{K(\Delta, u|x) - uy\} - \{K(\Delta, \hat{u}|x) - \hat{u}y\} = \{K_0(w) - wy_0\} - \{K_0(\hat{w}) - \hat{w}y_0\}. \tag{3.14}$$

Suppose now that, for a given y, we select $y_0 = y$. Then, instead of (3.11), the change of variable from u to w will be given by

$$\{K(\Delta, u|x) - uy\} - \{K(\Delta, \hat{u}|x) - \hat{u}y\} = \{K_0(w) - wy\} - \{K_0(\hat{w}) - \hat{w}y\}. \tag{3.15}$$

The two functions have the same local behavior near $\hat{u}$. As in the Gaussian case, we now have

$$\begin{aligned} p(\Delta, y|x) &= (2\pi\mathbf{i})^{-1} \int_{\hat{u}-\mathbf{i}\infty}^{\hat{u}+\mathbf{i}\infty} \exp(K(\Delta, u|x) - uy)\, \mathrm{d}u \\ &= (2\pi\mathbf{i})^{-1} \exp(\{K(\Delta, \hat{u}|x) - \hat{u}y\} - \{K_0(\hat{w}) - \hat{w}y\}) \\ &\quad \times \int_{\hat{w}-\mathbf{i}\infty}^{\hat{w}+\mathbf{i}\infty} \exp(K_0(w) - wy) \frac{\mathrm{d}u(w)}{\mathrm{d}w}\, \mathrm{d}w \end{aligned}$$

to be approximated by the leading term

$$
\begin{aligned}
p^{(0)}(\Delta, y|x) = (2\pi \mathbf{i})^{-1} \exp(\{K(\Delta, \hat{u}|x) - \hat{u}y\} - \{K_0(\hat{w}) - \hat{w}y\}) \\
\times \int_{\hat{w}-\mathbf{i}\infty}^{\hat{w}+\mathbf{i}\infty} \exp(K_0(w) - wy)\, \mathrm{d}w \frac{\mathrm{d}u(\hat{w})}{\mathrm{d}w} \\
= f_0(y) \exp(\{K(\Delta, \hat{u}|x) - \hat{u}y\} - \{K_0(\hat{w}) - \hat{w}y\}) \\
\times (K_0''(\hat{w}))^{1/2} \left(\frac{\partial^2 K(\Delta, \hat{u}|x)}{\partial u \partial u^{\mathrm{T}}} \right)^{-1/2}
\end{aligned}
\tag{3.16}
$$

since (3.15) now yields

$$
\frac{\mathrm{d}u(\hat{w})}{\mathrm{d}w} = (K_0''(\hat{w}))^{1/2} \left(\frac{\partial^2 K(\Delta, \hat{u}|x)}{\partial u^2} \right)^{-1/2}
$$

instead of (3.12). Of course, these expressions coincide with their Gaussian counterparts if $K_0(w) = w^2/2$ and $f_0 = \phi$.

3.4. Approximating the cumulative transition density function

In some situations in finance, the need arises to approximate not the (conditional) density function of the process, but rather the corresponding cumulative distribution function. The calculation of Value at Risk (VaR) is the first example. In VaR calculations, one needs to evaluate the probability that the value S_T at date T of a given portfolio will be lower than some preassigned cutoff a, i.e., $\Pr(S_T < a)$. The cutoff is often set as a function of the value of the portfolio at the current date t, say a loss not to exceed 20%, i.e., the calculation required is of the form $\Pr(S_T < aS_t)$ with, for example, $a = 0.8$. All these probabilities are conditional on the information set at t, which in this case reduces to S_t, so the calculation in the relative cutoff case involves no additional difficulty.

A second example is represented by the calculation of the expected loss on a portfolio, given that a loss of a certain magnitude occurs. This is sometimes advocated as a better risk measure than the simple VaR. The calculation to be performed in this case is $\mathrm{E}[S_T|S_T < a]$, with the expectation again taken conditionally on S_t.

A combination of both types of calculations arise in a third situation, option pricing. Considering the risk-neutral version of the process, the price at time t of a European put option with strike price X, time to maturity $\Delta = T - t$, riskless rate r, written on an asset with price S, is

$$
\begin{aligned}
P_t &= \exp(-r\Delta)\mathrm{E}[\max(0, X - S_T)] \\
&= \exp(\ln X - r\Delta) \Pr(S_T < X) - \exp(-r\Delta)\mathrm{E}[S_T|S_T < X] \Pr(S_T < X),
\end{aligned}
$$

where the expected value and probability are both conditioned on the current value S_t of the underlying asset.

We can express the option price in terms of cumulative probabilities and the corresponding cumulant generating functions, as in Rogers and Zane (1999). Let

$Y = \ln(S)$ and K denote the conditional cumulant generating function of Y_T, that is, $K(\varDelta, u|x) = \ln(\mathrm{E}[\exp(uY_T)])$. Note that $\mathrm{E}[S_T|S_T < X]\Pr(S_T < X) = \exp(K(1))\widetilde{\Pr}(S_T < X)$, where the probability $\widetilde{\Pr}$ is defined by

$$\widetilde{\mathrm{E}}[\exp(uY_T)] = \mathrm{E}[\exp((u+1)Y_T)]\exp(-K(\varDelta, 1|x)).$$

The cumulant generating function of $\widetilde{\Pr}$ is given by

$$\widetilde{K}(u) = K(u+1) - K(1).$$

Therefore, provided that the function K is known, we can express the option price as

$$P_t = \exp(\ln X - r\varDelta)\Pr(Y_T < \ln X) - \exp(-r\varDelta + K(1))\widetilde{\Pr}(Y_T < \ln X)$$

and all that remains to be done is to approximate the cumulative probabilities $\Pr(Y_T < \ln X)$ and $\widetilde{\Pr}(Y_T < \ln X)$ in order to approximate the option price. In all these situations, we need to be able to estimate the cumulative distribution function of a given Markov process.

An obvious approximation of the cumulative distribution function can be obtained by integrating the density approximation. Even if the cumulant generating function is not explicit, this is less tedious than it first seems because the monotonicity of $\partial K/\partial u$ means that the saddlepoint need not be recomputed at each point. In any event, this can be bypassed altogether. An explicit saddlepoint approximation for the cumulative distribution function can be constructed using a method proposed by Lugannani and Rice (1980). Start from the Fourier inversion formula, applied to the cumulative distribution function

$$\Pr(Y_\varDelta > y|Y_0 = x) \equiv P(\varDelta, y|x) = (2\pi\mathbf{i})^{-1}\int_{\hat{u}-\mathbf{i}\infty}^{\hat{u}+\mathbf{i}\infty}\exp(K(\varDelta, u|x) - u\cdot y)\frac{\mathrm{d}u}{u}. \tag{3.17}$$

Changing the variable u to w, as defined in (3.11), yields

$$P(\varDelta, y|x) = (2\pi\mathbf{i})^{-1}\int_{\hat{w}-\mathbf{i}\infty}^{\hat{w}+\mathbf{i}\infty}\exp\left(\frac{1}{2}w^2 - w\hat{w}\right)\left(\frac{1}{u(w)}\frac{\mathrm{d}u(w)}{\mathrm{d}w}\right)\mathrm{d}w. \tag{3.18}$$

Near $u = w = 0$, $K(\varDelta, u|x) \approx u\partial K(\varDelta, 0|x)/\partial u$, hence

$$w \approx \left\{\frac{y - \partial K(\varDelta, 0|x)/\partial u}{\hat{w}}\right\}u$$

provided that $y \neq \partial K(\varDelta, 0|x)/\partial u = \mathrm{E}[X_\varDelta|X_0 = x]$, and

$$w \approx \{\partial^2 K(\varDelta, 0|x)/\partial u^2\}^{-1/2}u$$

otherwise (since $\hat{u} = \hat{w} = 0$ when $y = \partial K(\varDelta, 0|x)/\partial u = \mathrm{E}[X_\varDelta|X_0 = x]$). Thus, near 0,

$$\frac{1}{u}\frac{\mathrm{d}u}{\mathrm{d}w} \approx \frac{1}{w}.$$

To isolate the singularity in $u^{-1}\,\mathrm{d}u/\mathrm{d}w$ near $w=0$, it is therefore necessary to decompose the integral (3.18) into two pieces

$$P(\Delta, y|x) = (2\pi\mathbf{i})^{-1}\int_{\hat{w}-\mathbf{i}\infty}^{\hat{w}+\mathbf{i}\infty} \exp\left(\frac{1}{2}w^2 - w\hat{w}\right)\frac{\mathrm{d}w}{w} + (2\pi\mathbf{i})^{-1}\int_{\hat{w}-\mathbf{i}\infty}^{\hat{w}+\mathbf{i}\infty} \exp\left(\frac{1}{2}w^2 - w\hat{w}\right)\left(\frac{1}{u}\frac{\mathrm{d}u}{\mathrm{d}w} - \frac{1}{w}\right)\mathrm{d}w. \quad (3.19)$$

The first integral in (3.19) above is simply $1-\Phi(\hat{w})$, where Φ is the standard normal cumulative distribution function. In the second integral, the function $u^{-1}\,\mathrm{d}u/\mathrm{d}w - w^{-1}$ is now analytic in a neighborhood of 0 (since $\mathrm{d}u/\mathrm{d}w$ is) and there is no singularity. It can be Taylor expanded around $\hat{w}$.

The expansion can be calculated to an arbitrary order using the definition of the function $u(w)$ implicit in (3.11); the leading term is

$$\begin{aligned}\frac{1}{u(w)}\frac{\mathrm{d}u(w)}{\mathrm{d}w} - \frac{1}{w} &= \frac{1}{\hat{u}}\frac{\mathrm{d}u(\hat{w})}{\mathrm{d}w} - \frac{1}{\hat{w}} + \mathrm{O}((w-\hat{w}))\\ &= \frac{1}{\hat{u}}\left(\frac{\partial^2 K(\Delta,\hat{u}|x)}{\partial u^2}\right)^{-1/2} - \frac{1}{\hat{w}} + \mathrm{O}((w-\hat{w}))\end{aligned}$$

when $y \neq \mathrm{E}[X_\Delta|X_0 = x]$. This yields the following leading term for the cumulative distribution function approximation:

$$\begin{aligned}P^{(0)}(\Delta, y|x) &= 1-\Phi(\hat{w}) + (2\pi\mathbf{i})^{-1}\int_{\hat{w}-\mathbf{i}\infty}^{\hat{w}+\mathbf{i}\infty}\exp\left(\frac{1}{2}w^2 - w\hat{w}\right)\mathrm{d}w\left\{\frac{1}{\hat{u}}\left(\frac{\partial^2 K(\Delta,\hat{u}|x)}{\partial u^2}\right)^{-1/2} - \frac{1}{\hat{w}}\right\}\\ &= 1-\Phi(\hat{w}) + \phi(\hat{w})\left\{\frac{1}{\hat{u}}\left(\frac{\partial^2 K(\Delta,\hat{u}|x)}{\partial u^2}\right)^{-1/2} - \frac{1}{\hat{w}}\right\}. \quad (3.20)\end{aligned}$$

When $y = \mathrm{E}[X_\Delta|X_0 = x]$, we have $\hat{u} = \hat{w} = 0$ and

$$\frac{1}{u(w)}\frac{\mathrm{d}u(w)}{\mathrm{d}w} - \frac{1}{w} = -\frac{\partial^3 K(\Delta,0|x)/\partial u^3}{6(\partial^2 K(\Delta,0|x)/\partial u^2)^{3/2}} + \mathrm{O}(w)$$

from which it follows as in (3.20) that the leading term of the approximation is

$$\begin{aligned}P^{(0)}(\Delta, \mathrm{E}[X_\Delta|X_0 = x]|x) &= 1-\Phi(0) + \phi(0)\left\{-\frac{\partial^3 K(\Delta,0|x)/\partial u^3}{6(\partial^2 K(\Delta,0|x)/\partial u^2)^{3/2}}\right\}\\ &= \frac{1}{2} - \frac{1}{6\sqrt{2\pi}}\frac{\partial^3 K(\Delta,0|x)/\partial u^3}{(\partial^2 K(\Delta,0|x)/\partial u^2)^{3/2}}. \quad (3.21)\end{aligned}$$

As shown by Wood et al. (1993), the natural generalizations of formulae (3.20)–(3.21) to non-Gaussian leading terms are

$$P^{(0)}(\Delta, y|x) = 1 - F_0(y_0) + f_0(y_0)\left\{\frac{1}{\hat{u}}(K_0''(\hat{w}))^{1/2}\left(\frac{\partial^2 K(\Delta, \hat{u}|x)}{\partial u^2}\right)^{-1/2} - \frac{1}{\hat{w}}\right\} \tag{3.22}$$

and

$$P^{(0)}(\Delta, \mathrm{E}[X_\Delta|X_0 = x]|x) = 1 - F_0(K_0'(0)) + f_0(K_0'(0))\frac{K_0''(0)}{6} \times\left\{\frac{K_0'''(0)}{K_0''(0)^{3/2}} - \frac{\partial^3 K(\Delta, 0|x)/\partial u^3}{(\partial^2 K(\Delta, 0|x)/\partial u^2)^{3/2}}\right\} \tag{3.23}$$

when $y \neq \mathrm{E}[X_\Delta|X_0 = x]$ and $y = \mathrm{E}[X_\Delta|X_0 = x]$, respectively.

4. The limiting behavior of the saddlepoint approximation for diffusions

We now turn to studying the behavior of the saddlepoint approximation when applied to diffusions. Consider the stochastic differential equation

$$\mathrm{d}X_t = \mu(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}W_t \tag{4.1}$$

and for simplicity of notation consider the scalar case. We let $\mathcal{S} = (\underline{x}, \bar{x})$ denote the domain of the diffusion X_t. Under regularity assumptions on the drift and diffusion function of the process (covering smoothness of the coefficients and boundary behavior), the stochastic differential equation (4.1) admits a weak solution which is unique in probability law. We will make use of the scale density of the process, defined as

$$s(x) \equiv \exp\left\{-2\int^x \frac{\mu(y)}{\sigma^2(y)}\,\mathrm{d}y\right\} \tag{4.2}$$

for a given lower bound of integration whose choice is irrelevant in what follows. We will denote by L^2 the Hilbert space of measurable real-valued functions f on $\mathcal{S}$ such that $\|f\|^2 \equiv \mathrm{E}[f(X_0)^2] < \infty$. When f is a function of other variables, in addition to the state variable y, we say that $f \in L^2$ if it satisfies the integrability condition for every given value of the other variables.

Expansions in Δ are obtained by iterations of the infinitesimal generator A of the process, that is, the operator which returns

$$A \cdot f = \frac{\partial f}{\partial \delta} + \mu(y)\frac{\partial f}{\partial y} + \frac{1}{2}\sigma^2(y)\frac{\partial^2 f}{\partial y^2} \tag{4.3}$$

when applied to functions $f(\delta, y, x)$ that are continuously differentiable once in δ, twice in y and such that $\partial f/\partial y$ and $A \cdot f$ are both in L^2 and satisfy

$$\lim_{y\to\underline{x}} \frac{\partial f/\partial y}{s(y)} = \lim_{y\to\bar{x}} \frac{\partial f/\partial y}{s(y)} = 0. \tag{4.4}$$

We define $\mathscr{D}$ to be the set of functions f which have these properties. For instance, functions f that are polynomial in y near the boundaries of $\mathscr{S}$, and their iterates by repeated application of the generator, retain their polynomial growth characteristic near the boundaries provided that (μ, σ^2) and their successive derivatives with respect to y have polynomial growth there; so they are all in L^2. They then satisfy (4.4) if $s(y)$ diverges exponentially near both boundaries of $\mathscr{S}$, whereas polynomials and their iterates diverge at most polynomially (multiplying and adding functions with polynomial growth yields a function still with polynomial growth). If f exhibits exponential divergence (for example, $f^{\sim}\exp(y)$), then s must also diverge exponentially and faster (for example, $f^{\sim}\exp(y^2)$). Finally, let us define $\mathscr{D}^J$ as the set of functions f which with $J+2$ continuous derivatives in δ, $2(J+2)$ in y, such that f and its first J iterates by repeated applications of A all remain in $\mathscr{D}$. For such an f, we have

$$\mathrm{E}[f(\varDelta, Y, X)|X = x] = \sum_{j=0}^{J} \frac{\varDelta^j}{j!} A^j \cdot f(0, x, x) + \mathrm{O}(\varDelta^{J+1}) \tag{4.5}$$

which can be viewed as an expansion in small time.

4.1. Expansions for the saddlepoint

The key to our approach is to approximate the Laplace transform of the process, and the resulting saddlepoint, as a Taylor series in $\varDelta$ around their continuous-time limit. This will result in an approximation (in $\varDelta$) to the saddlepoint (which itself is an approximation to the true but unknown transition density of the process). By applying (4.5) to the function $f(\delta, y, x) = \exp(uy)$, u treated as a fixed parameter, we can compute the expansion of the Laplace transform $\varphi(\varDelta, u|x)$ in $\varDelta$. At order $n_2 = 1$, the result is

$$\varphi^{(1)}(\varDelta, u|x) = \mathrm{e}^{ux}(1 + (\mu(x)u + \tfrac{1}{2}\sigma^2(x)u^2)\varDelta).$$

Then, by taking its log, we see that the expansion at order $\varDelta$ of the cumulant transform K is simply

$$\begin{aligned} K^{(1)}(\varDelta, u|x) &= ux + \frac{\partial}{\partial \varDelta} K(\varDelta, u|x)\bigg|_{\varDelta=0} \varDelta \\ &= ux + \left(\mu(x)u + \frac{1}{2}\sigma^2(x)u^2\right)\varDelta. \end{aligned}$$

The first-order saddlepoint $\hat{u}^{(1)}$ solves $\partial K^{(1)}(\varDelta, u|x)/\partial u = y$, that is,

$$\hat{u}^{(1)}(\varDelta, y|x) = \frac{y - (x + \mu(x)\varDelta)}{\sigma^2(x)\varDelta},$$

and, when evaluated at $y = x + z\varDelta^{1/2}$, we have

$$\hat{u}^{(1)}(\varDelta, x + z\varDelta^{1/2}|x) = \frac{z}{\sigma^2(x)\varDelta^{1/2}} + \mathrm{O}(1) \tag{4.6}$$

and

$$K^{(1)}(\Delta, \hat{u}^{(1)}(\Delta, x + z\Delta^{1/2}|x)|x) - \hat{u}^{(1)}(\Delta, x + z\Delta^{1/2}|x)(x + z\Delta^{1/2})$$
$$= -\frac{z^2}{2\sigma^2(x)} + \mathrm{O}(\Delta^{1/2}).$$

Similarly, a second-order expansion in Δ of φ is

$$\varphi^{(2)}(\Delta, u|x) = \exp\left[ux + (\mu(x)u + \frac{1}{2}\sigma^2(x)u^2)\Delta\right] \times \left\{1 + \frac{1}{2}\frac{\partial^2}{\partial\Delta^2}K(\Delta, u|x)\bigg|_{\Delta=0}\Delta^2\right\}, \tag{4.7}$$

$$K^{(2)}(\Delta, u|x) = ux + \Delta\left(\mu(x)u + \frac{1}{2}\sigma^2(x)u^2\right) + \frac{\Delta^2 u}{8}\{4\mu(x)\mu'(x) + 2\sigma^2(x)\mu''(x) + u(4\sigma^2(x)\mu'(x) + 2\mu(x)(\sigma^2)'(x) + \sigma^2(x)(\sigma^2)''(x)) + 2u^2\sigma^2(x)(\sigma^2)'(x)\} + \mathrm{O}(\Delta^3).$$

The second-order saddlepoint $\hat{u}^{(2)}$ solves $\partial K^{(2)}(\Delta, u|x)/\partial u = y$, which is a quadratic equation explicitly solvable in u, and we see after some calculations that

$$\hat{u}^{(2)}(\Delta, x + z\Delta^{1/2}|x) = \frac{z}{\sigma^2(x)\Delta^{1/2}} - \left\{\frac{\mu(x)}{\sigma^2(x)} + \frac{3(\sigma^2)'(x)}{4\sigma^4(x)}z^2\right\} + \mathrm{O}(\Delta^{1/2}) \tag{4.8}$$

and

$$K^{(2)}(\Delta, \hat{u}^{(2)}(\Delta, x + z\Delta^{1/2}|x)|x) - \hat{u}^{(2)}(\Delta, x + z\Delta^{1/2}|x)(x + z\Delta^{1/2})$$
$$= -\frac{z^2}{2\sigma^2(x)} + \left\{\frac{\mu(x)}{\sigma^2(x)}z + \frac{(\sigma^2)'(x)}{4\sigma^4(x)}z^3\right\}\Delta^{1/2} + \mathrm{O}(\Delta).$$

The way the correction terms in $\varphi^{(2)}(\Delta, u|x)$ are grouped is similar to that of an Edgeworth expansion. The proof of Theorem 3 discusses how higher-order approximate Laplace transforms are constructed.

4.2. Explicit expressions for the saddlepoint approximation

In what follows, we use the notation $p^{(n_1,n_2)}$ to indicate a saddlepoint approximation of order n_1 (in the sense of Section 3.2) using a Taylor expansion in Δ of the Laplace transform φ, that is, correct at order n_2 in Δ. When the expansions in Δ are analytic at zero (see, Aït-Sahalia, 2002), then $p^{(n_1,\infty)} = p^{(n_1)}$.

Theorem 1. *Assume that* (μ, σ^2) *are such that the function* $f(y) \equiv \exp(uy)$ *is in* $\mathscr{D}^2$. *Then, the leading term of the saddlepoint approximation at the first order in* Δ *and with a Gaussian base coincides with the classical Euler approximation of the transition* *density*

$$p^{(0,1)}(\Delta, y|x) = (2\pi\Delta\sigma^2(x))^{-1/2} \exp\left(-\frac{(y - x - \mu(x)\Delta)^2}{\sigma^2(x)\Delta}\right). \tag{4.9}$$

The next order saddlepoint approximation, which will obviously be more useful in empirical applications than the term $p^{(0,1)}$, is given by the following:

Theorem 2. *Under the same assumption as Theorem* 1, *the first-order saddlepoint approximation at the first order in* Δ *and with a Gaussian base is given by* (3.7)

$$p^{(1,1)}(\Delta, x + z\Delta^{1/2}|x) = \frac{\exp(-z^2/2\sigma^2(x) + e_{1/2}(z|x)\Delta^{1/2} + e_1(z|x)\Delta)}{\sqrt{2\pi}\sigma(x)\Delta^{1/2}\{1 + d_{1/2}(z|x)\Delta^{1/2} + d_1(z|x)\Delta\}}\{1 + c_1(z|x)\Delta\}, \tag{4.10}$$

where

$$e_{1/2}(z|x) = \frac{z\mu(x)}{\sigma^2(x)} + \frac{z^3(\sigma^2)'(x)}{4\sigma^4(x)},$$

$$\begin{aligned} e_1(z|x) = {} & -\frac{\mu(x)^2}{2\sigma^2(x)} + \frac{z^2(12\sigma^2(x)(4\mu'(x) + (\sigma^2)''(x)) - 48\mu(x)(\sigma^2)'(x))}{96\sigma^4(x)} \\ & + \frac{z^4(8\sigma^2(x)(\sigma^2)''(x) - 15(\sigma^2)'(x)^2)}{96\sigma^6(x)}, \end{aligned} \tag{4.11}$$

$$d_{1/2}(z|x) = \frac{3z\sigma'(x)}{2\sigma(x)},$$

$$d_1(z|x) = \frac{\mu'(x)}{2} - \frac{\mu(x)\sigma'(x)}{\sigma(x)} + \frac{\sigma'(x)^2}{4} + \frac{\sigma(x)\sigma''(x)}{4} + z^2\left(\frac{5\sigma'(x)^2}{8\sigma(x)^2} + \frac{\sigma''(x)}{\sigma(x)}\right),$$

$$c_1(z|x) = \frac{1}{4}(\sigma^2)''(x) - \frac{3}{32}\frac{(\sigma^2)'(x)^2}{\sigma^2(x)}.$$

The expression (4.10) *provides an alternative gathering of the correction terms beyond the leading term that is equivalent at order* Δ *to the irreducible expansion of the transition density in* Aït-Sahalia (2001).

Note also that there is no need to compute the saddlepoint in the two theorems above, as the results already incorporate this calculation and represent the final "ready-to-use" saddlepoint approximation as a function of the specification of the diffusion model, $(\mu(\cdot), \sigma^2(\cdot))$ and their derivatives.

As for the next order, two, the expression can be computed similarly: the saddlepoint expansion is obtained from the expansion of the moment generating function

$$\begin{aligned}
&\exp(K(\Delta,u|x)-u\cdot y)\\
&\quad\approx \exp(K(\Delta,\hat{u}|x)-\hat{u}\cdot y)\exp(-\tfrac{1}{2}K_{(2)}v^2)\\
&\qquad\times\{1-\tfrac{1}{6}K_{(3)}\mathbf{i}v^3+\tfrac{1}{24}K_{(4)}v^4+\tfrac{1}{120}K_{(5)}\mathbf{i}v^5-\tfrac{1}{720}K_{(6)}v^6\\
&\qquad+\tfrac{1}{2}[-\tfrac{1}{36}K_{(3)}^2v^6-\tfrac{1}{72}K_{(3)}K_{(4)}\mathbf{i}v^7+\tfrac{1}{576}K_{(4)}^2v^8+\tfrac{1}{360}K_{(3)}K_{(5)}v^8]\\
&\qquad+\tfrac{1}{6}[\tfrac{1}{216}K_{(3)}^3\mathbf{i}v^9-\tfrac{1}{288}K_{(3)}^2K_{(4)}v^{10}]+\tfrac{1}{24}\tfrac{1}{1296}K_{(3)}^4v^{12}\},
\end{aligned}$$

where, to simplify notation, $K_{(n)}$ denotes $\partial^{(n)}K(\Delta,\hat{u}|x)/\partial u^n$ and only terms of the relevant order in Δ are kept. The specific terms below are retained in the expansion so that the saddlepoint approximation will have the desired order in Δ (rather than the irrelevant v):

$$\begin{aligned}
p^{(2,2)}(\Delta,x+z\Delta^{1/2}|x)=&\frac{\exp(K^{(4)}(\Delta,\hat{u}^{(2)}|x)-\hat{u}^{(2)}(x+z\Delta^{1/2}))}{\sqrt{2\pi}K_{(2)}^{1/2}}\\
&\times\Bigg\{1+\frac{1}{8}\underbrace{\frac{K_{(4)}}{K_{(2)}^2}}_{O(\Delta)}-\frac{5}{24}\underbrace{\frac{K_{(3)}^2}{K_{(2)}^3}}_{O(\Delta)}-\frac{1}{48}\underbrace{\frac{K_{(6)}}{K_{(2)}^3}}_{O(\Delta^2)}+\frac{35}{384}\underbrace{\frac{K_{(4)}^2}{K_{(2)}^4}}_{O(\Delta^2)}\\
&+\frac{7}{48}\underbrace{\frac{K_{(3)}K_{(5)}}{K_{(2)}^4}}_{O(\Delta^2)}-\frac{35}{64}\underbrace{\frac{K_{(3)}^2K_{(4)}}{K_{(2)}^5}}_{O(\Delta^2)}+\frac{385}{1152}\underbrace{\frac{K_{(3)}^4}{K_{(2)}^6}}_{O(\Delta^2)}\Bigg\}
\end{aligned}$$

by proceeding as in the proof of Theorem 2.

We now turn to expansions of higher order in Δ, and have the following:

Theorem 3. *For* $n_1\geqslant 0$ *and* $n_2\geqslant 1$, $p(\Delta,x+z\Delta^{1/2}|x)=p^{(n_1,n_2)}(\Delta,x+z\Delta^{1/2}|x)(1+O(\Delta^{\min(n_1+1,n_2/2)}))$.

Corollary 1. *If* (μ,σ^2) *are such that* $K(\Delta,u|X_0)$ *is analytic at* $\Delta=0$, *then*

$$p(\Delta,x+z\Delta^{1/2}|x)=p^{(n_1)}(\Delta,x+z\Delta^{1/2}|x)(1+O(\Delta^{n_1+1}))$$

for $n_1\geqslant 0$.

5. Some diffusion examples

5.1. The Ornstein– Uhlenbeck process

Consider an Ornstein–Uhlenbeck process

$$\mathrm{d}X_t=[A(t)X_t+a(t)]\mathrm{d}t+\sigma(t)\,\mathrm{d}W_t. \tag{5.1}$$

For this process, $X_\Delta - X_0 \sim \mathrm{N}(m(\Delta), V(\Delta))$, where

$$m(t) = \Gamma(t)\int_0^t \Gamma^{-1}(s)a(s)\,\mathrm{d}s,$$
$$V(t) = \Gamma(t)\left\{\int_0^t \Gamma^{-1}(s)\sigma(s)[\Gamma^{-1}(s)\sigma(s)]^{\mathrm{T}}\,\mathrm{d}s\right\}\Gamma^{\mathrm{T}}(t)$$

and $\Gamma(t)$ solves $\Gamma'(t) = A(t)\Gamma(t)$ subject to $\Gamma(0) = I$.

The saddlepoint is

$$\hat{u}(\Delta, X_\Delta|X_0) = [V(\Delta)]^{-1}[X_\Delta - X_0 - m(\Delta)]. \tag{5.2}$$

5.2. *The log-normal model*

In the Black–Scholes geometric Brownian motion specification

$$\mathrm{d}X_t = \mu X_t\,\mathrm{d}t + \sigma X_t\,\mathrm{d}W_t. \tag{5.3}$$

Replacing X by $Y = \ln(X)$ and applying Itô's lemma, we have

$$\mathrm{d}Y_t = (\mu - \sigma^2/2)\,\mathrm{d}t + \sigma\,\mathrm{d}W_t$$

which takes us back to the Gaussian case. Since $Y_\Delta - Y_0 \sim \mathrm{N}(\Delta(\mu - \sigma^2/2), \sigma^2\Delta)$, the saddlepoint for this model is given by

$$\hat{u}(\Delta, Y_\Delta|Y_0) = \frac{Y_\Delta - Y_0 - \Delta(\mu - \sigma^2/2)}{\Delta\sigma^2}. \tag{5.4}$$

5.3. *The square-root process*

Consider Feller's square-root process (used in the Cox et al. (1985) interest rate model):

$$\mathrm{d}X_t = \kappa(\theta - X_t)\,\mathrm{d}t + \sigma\sqrt{X_t}\,\mathrm{d}W_t. \tag{5.5}$$

Let

$$c = \frac{2\kappa}{(1 - \mathrm{e}^{-\Delta\kappa})\sigma^2}, \quad b = \frac{2\kappa X_0}{\mathrm{e}^{\Delta\kappa}(1 - \mathrm{e}^{-\Delta\kappa})\sigma^2}, \quad q = \frac{2\theta\kappa}{\sigma^2} - 1.$$

Then $2c(X_\Delta - X_0)$ is distributed with a non-central χ^2 with $2q+2$ degrees of freedom and non-centrality parameter $2b$. The Laplace transform for X_Δ given $X_0 = x$ is

$$\varphi(\Delta, u|x) = \left(1 - \frac{u}{c}\right)^{-q-1}\exp\left(ux + \frac{bu}{c-u}\right).$$

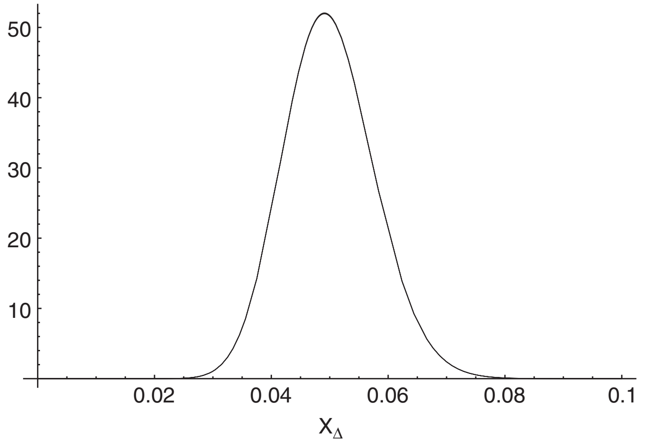

Fig. 1. Transition density $p(\Delta, X_\Delta|X_0)$ for Feller's square-root model.

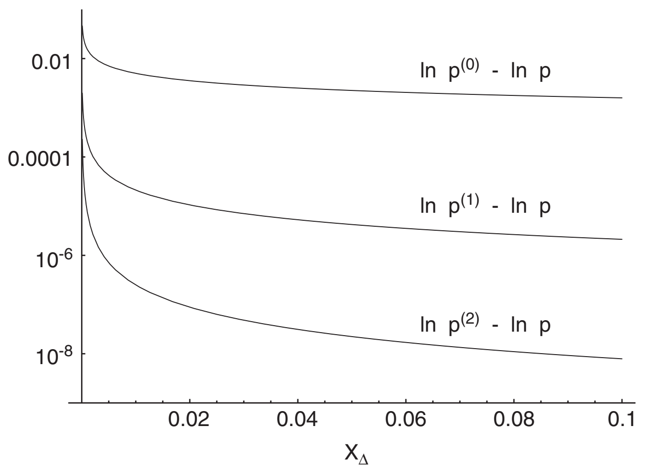

Fig. 2. Log-approximation error of the saddlepoint expansions of order 0, 1 and 2 for Feller's square-root model.

The saddlepoint is then

$$\hat{u}(\Delta, X_\Delta|X_0) = \frac{-\left(1 + q - 2c(X_\Delta - X_0) + \sqrt{1 + 2q + q^2 + 4cb(X_\Delta - X_0)}\right)}{2(X_\Delta - X_0)}. \tag{5.6}$$

We now illustrate the accuracy of saddlepoint approximation for the square-root process. The parameters used are $\kappa = 0.5$, $\theta = 0.05$, $\sigma = 0.25$, $x = 0.05$, $\Delta = \frac{1}{52}$ (weekly frequency). The left boundary 0 can be reached with this set of parameters. The graph in Fig. 1 plots the true transition density $p(\Delta, y|x)$. Fig. 2 shows that the approximation error is small. Further, the approximation error decreases rapidly when higher-order expansions $p^{(1)}(\Delta, y|x)$ and $p^{(2)}(\Delta, y|x)$ are used.

5.4. A stochastic volatility model

The following model was proposed by Heston (1993)

$$\mathrm{d}X_t = (r + kv_t)\,\mathrm{d}t + \sqrt{v_t}\,\mathrm{d}W_{1t}, \tag{5.7}$$

$$\mathrm{d}v_t = (a - bv_t)\,\mathrm{d}t + \sigma\sqrt{v_t}\,\mathrm{d}W_{2t}, \tag{5.8}$$

where W_1 and W_2 are two Brownian motions with correlation coefficient ρ.

The Laplace transform of X_Δ given $X_0 = x$ and v_0 is given by

$$\varphi(\Delta, u|x, v_0) = \exp(C(u, \Delta) + D(u, \Delta)v_0 + uX_0), \tag{5.9}$$

where

$$\begin{aligned}
C(u,\tau) &= ru\tau + \frac{a}{\sigma^2}\left[(b - \rho\sigma u + d)\tau - 2\ln\left(\frac{1 - g\mathrm{e}^{d\tau}}{1 - g}\right)\right],\\
D(u,\tau) &= \frac{b + d - \rho\sigma u}{\sigma^2}\left(\frac{1 - \mathrm{e}^{d\tau}}{1 - g\mathrm{e}^{d\tau}}\right),\\
g &= \frac{b - \rho\sigma u + d}{b - \rho\sigma u - d},\\
d &= \sqrt{(\rho\sigma u - b)^2 - \sigma^2(2ku + u^2)}.
\end{aligned} \tag{5.10}$$

We now illustrate the accuracy of the saddlepoint approximation for this stochastic volatility process. The parameters used are $r = 0.05$, $k = -0.01$, $a = 0.3$, $b = 0.15$, $\sigma = 1$, $\rho = 0.1$, $x = 1$, $v_0 = 0.25$, $\Delta = 0.1$. Fig. 3 plots the transition density from numerical Fourier inversion which is treated as the true transition

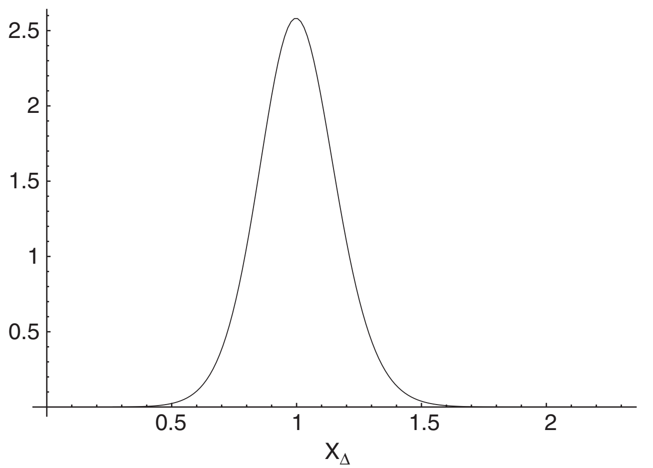

Fig. 3. Transition density $p(\Delta, X_\Delta|X_0)$ for Heston's stochastic volatility model.

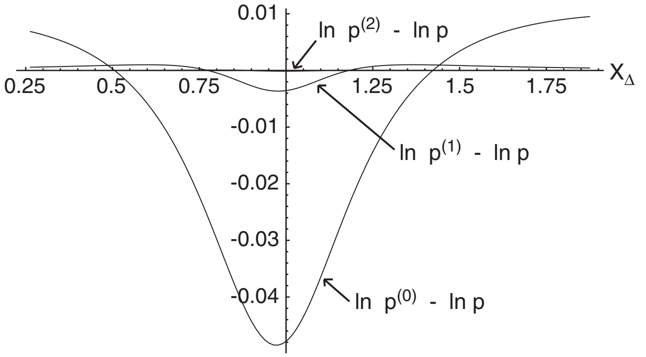

Fig. 4. Log-approximation error of the saddlepoint expansions of order 0, 1 and 2 for Heston's stochastic volatility model.

density. Fig. 4 shows that the log-approximation error is very small. By the time a second-order expansion $p^{(2)}(\Delta, y|x)$ is used, the log-approximation error virtually disappears. Of course, the transition density that we are approximating here is $p(X_\Delta, v_\Delta|X_0, v_0)$: this object cannot be directly used for likelihood inference on the stochastic volatility model because the volatility state variable is unobservable. Inference would have to be based on the integrated out density $p(X_\Delta|X_0)$.

6. The limiting behavior of the saddlepoint approximation for jump-diffusions

Consider now the stochastic differential equation

$$\mathrm{d}X_t = \mu(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}W_t + J_t\,\mathrm{d}N_t$$

and, for simplicity of notation, consider the scalar case. N is a Poisson process with arrival rate λ and the probability density $v(\cdot)$ of the jump size J_t has a moment-generating function denoted as $\theta(u) \equiv \mathrm{E}[\exp(uJ_t)]$. A typical example would be the model of Merton (1976), in which X_t denotes the log-return derived from an asset,

$$\mathrm{d}X_t = \mu\,\mathrm{d}t + \sigma\,\mathrm{d}W_t + J_t\,\mathrm{d}N_t, \tag{6.1}$$

and the log-jump size J_t is a Gaussian random variable with mean β and variance η, so that

$$\theta(u) = \exp(\tfrac{1}{2}\eta^2u^2 + \beta u). \tag{6.2}$$

Other typical examples would be one where J_t has an exponential distribution with mean 0 and variance $2/\kappa^2$ (i.e., density $\kappa\exp(-\kappa|J|)/2$), in which case, for $|u| < \kappa$ we have

$$\theta(u) = \frac{\kappa^2}{\kappa^2 - u^2}, \tag{6.3}$$

or an exponential distribution with all jumps positive of mean $1/\kappa$ and variance $1/\kappa^2$ (i.e., density $\kappa\exp(-\kappa J)$), in which case, again for $|u|<\kappa$, we have

$$\theta(u) = \frac{\kappa}{\kappa - u}. \tag{6.4}$$

In the presence of jumps, we will see first that selecting a Gaussian base for the saddlepoint approximation leads to poor performance in the tails. Accurate approximation in the tails requires the use of a better leading term, one that reflects the presence of the jumps.

6.1. Saddlepoint approximation with a Gaussian base

If we attempt to use the same approach as in the purely diffusive case, that is, use a Gaussian base for the saddlepoint approximation, we obtain at the leading order:

Lemma 1. *The saddlepoint approximation at the first order with a Gaussian base is*

$$p^{(0,1)}(\varDelta, y|x) = \frac{\exp(-\frac{1}{2}\sigma^2(x)\hat{u}^{(1)2}\varDelta)}{(2\pi(\sigma^2(x)\varDelta + \Delta\lambda\theta''(\hat{u}^{(1)})))^{1/2}}\exp(\Delta\lambda(\theta(\hat{u}^{(1)}) - \hat{u}^{(1)}\theta'(\hat{u}^{(1)}) - 1)), \tag{6.5}$$

where the saddlepoint $\hat{u}^{(1)}$ is the solution of the equation

$$\hat{u}^{(1)} = \frac{y - x - \mu(x)\varDelta}{\sigma^2(x)\varDelta} - \frac{\lambda\theta'(\hat{u}^{(1)})}{\sigma^2(x)}.$$

When $\lambda = 0$, there is no jump, and the saddlepoint density (6.5) coincides with the first-order saddlepoint density $p^{(0,1)}(\varDelta, y|x)$ from the diffusion case, given in (4.9).

As an example, consider the jump-diffusion process

$$\mathrm{d}X_t = \sigma\,\mathrm{d}W_t + J_t\,\mathrm{d}N_t$$

with $X_0 = 0$ and where the jump J_t has exponential distribution with moment generating function (6.4). The Gaussian saddlepoint approximation has the leading term

$$f_0(y) = \frac{1}{\sqrt{2\pi\sigma^2\varDelta}}\exp\left(-\frac{y^2}{2\sigma^2\varDelta}\right)$$

whose corresponding cumulant generating function is

$$K_0(u) = ux + \tfrac{1}{2}\sigma^2\Delta u^2$$

and the saddlepoint approximation to the transition density $X_\varDelta|X_0 = 0$ is then

$$p^{(0)}(\varDelta, y|x=0) = f_0(y)\exp(\{K(\varDelta,\hat{u}|x) - \hat{u}y\} - \{K_0(\hat{w}) - \hat{w}y\})(K_0''(\hat{w}))^{1/2}\left(\frac{\partial^2 K(\varDelta,\hat{u}|x)}{\partial u\partial u^{\mathrm{T}}}\right)^{-1/2}.$$

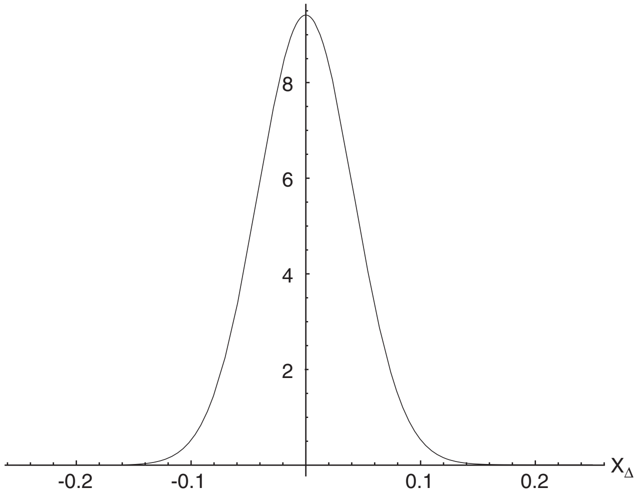

Fig. 5. Transition density $p(\Delta, X_\Delta|X_0)$ for a jump-diffusion model with exponentially distributed jumps.

The simple form of $\theta(u)$ enables us to compute that

$$\hat{u} = \kappa - \sqrt{\frac{\lambda\kappa}{y}}\sqrt{\Delta} - \frac{1}{2}\sqrt{\frac{\kappa^3\lambda}{y^3}}\sigma^2\Delta^{3/2} + \mathrm{O}(\Delta^2),$$
$$\hat{w} = \frac{y}{\sigma^2\Delta}$$

which imply

$$\{K(\Delta,\hat{u}|x) - \hat{u}y\} - \{K_0(\hat{w}) - \hat{w}y\} = \frac{y^2}{2\sigma^2\Delta} - y\kappa + \mathrm{O}\left(\sqrt{\Delta}\right),$$
$$(K_0''(\hat{w}))^{1/2}\left(\frac{\partial^2 K(\Delta,\hat{u}|x)}{\partial u\partial u^{\mathrm{T}}}\right)^{-1/2} = \frac{1}{\sqrt{2}}\left(\frac{\kappa\lambda}{y^3}\right)^{1/4}\sigma\Delta^{3/4} + \mathrm{o}(\Delta^{3/4}).$$

It can be seen that using the leading term $f_0(y)$ alone will lead to a very poor approximation since f_0 is exponentially small in Δ for y in the tails yet a jump distribution is of order Δ at tails. Fig. 5 plots the true transition density for parameter values $\kappa = 2$, $\Delta = \frac{1}{52}$, $\sigma = 0.3$, $\lambda = 1$, $X_0 = 0$. Fig. 6 plots the log difference between the saddlepoint expansion of order 0 and the true transition density. Exponential distribution allows positive jump size only. To the left of the origin, the saddlepoint approximation is very good as in the diffusion case. To the right of the origin, the saddlepoint approximation with Gaussian base cannot capture the true transition density. It can be seen that Gaussian base cannot generate satisfactory saddlepoint approximations for jump diffusions.

It is worth observing that the saddlepoint correction term in this case is

$$\exp(\{K(\Delta,\hat{u}|x) - \hat{u}y\} - \{K_0(\hat{w}) - \hat{w}y\}) = \exp\left(\frac{y^2}{2\sigma^2\Delta} - y\kappa + \mathrm{O}(\sqrt{\Delta})\right)$$

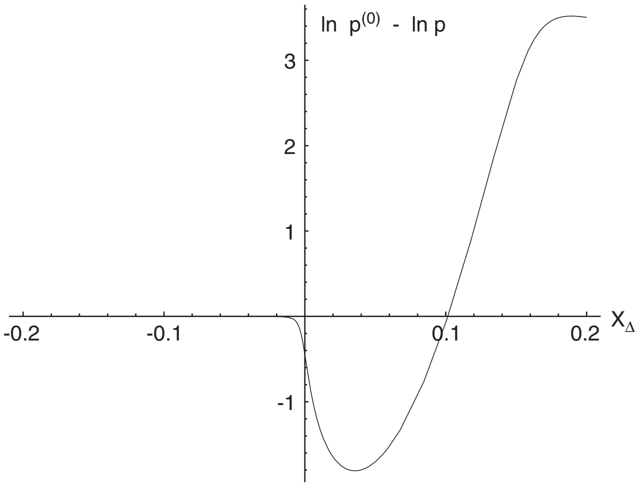

Fig. 6. Log-approximation error of the saddlepoint expansions with a Gaussian base of order 0 for a jump-diffusion model with exponentially distributed jumps. This example illustrates the need for non-Gaussian-based saddlepoint expansions in the case of processes with jumps.

and therefore cancels the exponentially small Gaussian f_0 in order to attempt some correction for the tail behavior that is badly misspecified by the Gaussian f_0. This yields a much better approximation in the tails than using f_0 alone. However, the saddlepoint density remains at order $O(\varDelta^{1/4})$ in the tails which still is inadequate to capture the true transition density. This leads us to consider a non-Gaussian leading term for the saddlepoint approximation when the underlying process can jump.

6.2. Saddlepoint approximation with a non-Gaussian base

So, instead of a Gaussian base, we will now approximate $p(\varDelta, y|x)$ by a saddlepoint expansion with a non-Gaussian leading term and rely on the general formula (3.16). To maintain explicitness, the non-Gaussian leading term we propose to use corresponds to the following Laplace transform:

$$\exp(K_0(\varDelta, u|x)) = \exp(ux)[(1 - \lambda\varDelta)\exp(\mu(x)\Delta u + \tfrac{1}{2}\sigma^2(x)\Delta u^2) + \lambda\Delta\theta(u)]$$

which is an order 1 in $\varDelta$ approximation to the true $\exp(K(\varDelta, u|x))$. This can also be viewed as the Laplace transform of a process that can have at most one jump in an interval of length $\varDelta$ and the jump probability is $1 - \lambda\varDelta$; that is the density function $f_0(y)$ corresponding to $K_0(\varDelta, u|x)$ is

$$f_0(y) = (1 - \lambda\varDelta)\phi(y; x + \mu(x)\varDelta, \sigma^2(x)\varDelta) + \lambda\Delta\nu(y - x).$$

In what follows, we use the notation $p^{(0,n)}$ to indicate a saddlepoint approximation with the non-Gaussian base K_0 (in the sense of (3.16)) using a Taylor expansion in $\varDelta$ of the cumulant generating function K, that is, accurate at order n in $\varDelta$. When the expansions in $\varDelta$ are analytic at zero, then $p^{(0,\infty)} = p^{(0)}$.

The first order in Δ approximation of the cumulant generating function is

$$K^{(1)}(\Delta, u|x) = ux + \Delta[\mu(x)u + \tfrac{1}{2}\sigma(x)^2u^2 + \lambda(\theta(u) - 1)] \tag{6.6}$$

and the second-order approximation of K is

$$\begin{aligned} K^{(2)}(\Delta, u|x) = {} & ux + \Delta\left[\mu(x)u + \frac{1}{2}\sigma(x)^2u^2 + \lambda(\theta(u) - 1)\right] \\ & + \log\left[1 + \frac{1}{2}\Delta^2\frac{\mathrm{d}^2}{\mathrm{d}\Delta^2}K(\Delta, u|x)\bigg|_{\Delta=0}\right], \end{aligned} \tag{6.7}$$

where

$$\begin{aligned} \frac{\mathrm{d}^2}{\mathrm{d}\Delta^2}K(\Delta, u|x)\bigg|_{\Delta=0} = {} & [\mu(x)\mu'(x) + \mu''(x) - \mu(x)]u \\ & + \left[\frac{1}{2}\mu(x)(\sigma^2)'(x) + \sigma^2(x)\mu'(x) + \frac{1}{2}(\sigma^2)''(x) - \frac{1}{2}\sigma^2(x)\right]u^2 \\ & + \frac{1}{2}\sigma^2(x)(\sigma^2)'(x)u^3 \\ & - \lambda(\theta(u) - 1)\left(\mu(x)u + \frac{1}{2}\sigma^2(x)u^2\right) \\ & + \lambda\int \mathrm{e}^{uj}\left[\mu(x+j)u + \frac{1}{2}\sigma^2(x+j)u^2\right]v(\mathrm{d}j). \end{aligned}$$

Higher-order approximations to K can be calculated by iteratively applying the infinitesimal generator as described in (4.5) for the diffusive case. Integrability of the approximate Laplace transform is guaranteed because of the quadratic term $\frac{1}{2}\sigma(x)^2u^2\Delta$. We then obtain that:

Theorem 4. *The saddlepoint approximations with the non-Gaussian base K_0 and with the true cumulant generating function K replaced by its n-th order approximation $K^{(n)}$ are given by*

$$\begin{aligned} p^{(0,n)}(\Delta, y|x) = {} & f_0(y)\exp(\{K^{(n)}(\Delta, \hat{u}|x) - \hat{u}y\} - \{K_0(\hat{w}) - \hat{w}y\}) \\ & \times(K_0''(\hat{w}))^{1/2}\left(\frac{\partial^2 K^{(n)}(\Delta, \hat{u}|x)}{\partial u\partial u^{\mathrm{T}}}\right)^{-1/2}, \end{aligned} \tag{6.8}$$

where the saddlepoint $\widehat{u}$ and $\widehat{w}$ can be solved from

$$y = \frac{\partial}{\partial u}K^{(n)}(\Delta, \hat{u}|x)$$

and from

$$y = \frac{\partial}{\partial u}K_0(\Delta, \hat{w}|x).$$

In this case, the saddlepoint will often need to be computed numerically despite the availability of an explicit Taylor series expansion for the function K and its derivatives (obtained from (6.6) or (6.7)), because Eq. (3.1) will be too involved, in general, to provide a closed-form expression for $\hat{u}$. One could then consider employing the approximate method proposed in Lieberman (1994), which eliminates the needs for a numerical solution.

As an example of the use of a non-Gaussian base for a jump diffusion model, consider the following stochastic process:

$$\mathrm{d}X_t = \sigma\,\mathrm{d}W_t + J_t\,\mathrm{d}N_t,$$

where $X_0 = 0$ and $v(\cdot)$ is $\mathrm{N}(\beta, \eta^2)$, so that $\theta(u)$ is given by (6.2).

Consider a non-Gaussian leading term saddlepoint approximation with the probability density function of the leading term being

$$f_0(y) = (1 - \lambda\Delta)\phi(y; 0, \sigma^2\Delta) + \lambda\Delta\phi(y; \beta, \eta^2),$$

where $\phi(\cdot; m, v)$ is the probability density function of $\mathrm{N}(m, v)$. $f_0(y)$ corresponds to the case where no jump occurs with probability $1 - \lambda\Delta$ and with probability $\lambda\Delta$ the distribution is the same as the jump distribution.

We set the parameter values at $\sigma = 0.3$, $\lambda = 1$, $\beta = 0$, $\eta = 0.5$, $\Delta = \frac{1}{52}$. Fig. 7 shows the log difference between the saddlepoint density and the true density for X_Δ between -0.25 and 0.25, covering a five standard deviation region when there is no jump. Compare this with Fig. 8 which shows the log difference between the leading term f_0 and the true density. Both the saddlepoint density and the leading term f_0 itself provide adequate approximations to the true transition density. In the middle part of the distribution, the saddlepoint correction does not improve the approximation much.

However, things change if we look at the region $[-5, 5]$ which covers a 10 standard deviation region (when a jump occurs). Fig. 9 plots the log difference between the saddlepoint and the true density. Even for observations that are 10 standard deviation away when jumps occur, the saddlepoint approximation provides a very good approximation. Note that at such extreme tails, the probability density is

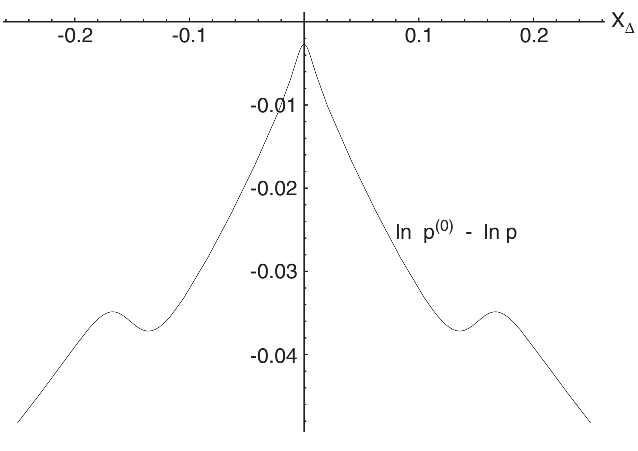

Fig. 7. Log-approximation error of the saddlepoint expansion of order 0 for a jump-diffusion model with normally distributed jumps.

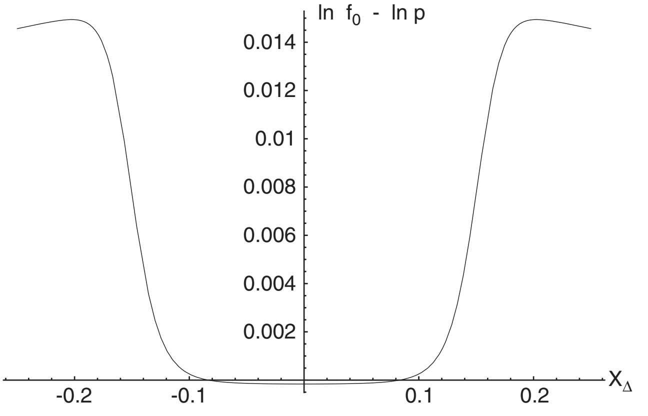

Fig. 8. Log-approximation error of the saddlepoint leading term f_0 for jump-diffusion model with normally distributed jumps, using a non-Gaussian base.

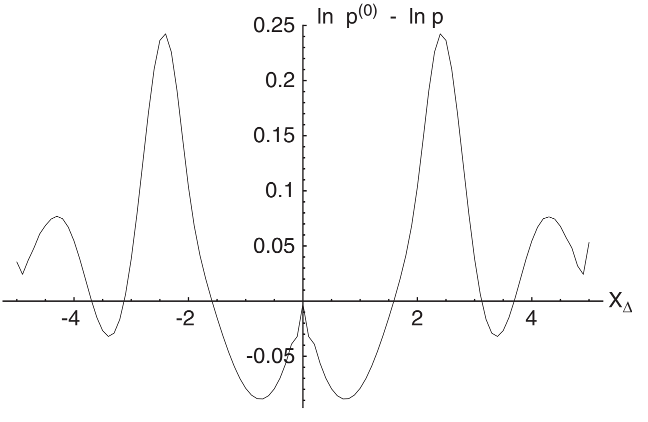

Fig. 9. Tail log-approximation error of the saddlepoint expansion of order 0 for a jump-diffusion model with normally distributed jumps, using a non-Gaussian base.

extremely small and a proportional error of 5% implies an extremely small density approximation error. This is quite different from Fig. 10 which shows the log difference between the leading term f_0 and the true density in the same region of $[-5, 5]$. The approximation error of the leading term f_0 increases dramatically as we move further into the tails.

7. The limiting behavior of the saddlepoint approximation for Lévy processes

We now turn to Lévy processes. Lévy processes are natural generalizations of Brownian motion: like Brownian motion, they have stationary and independent

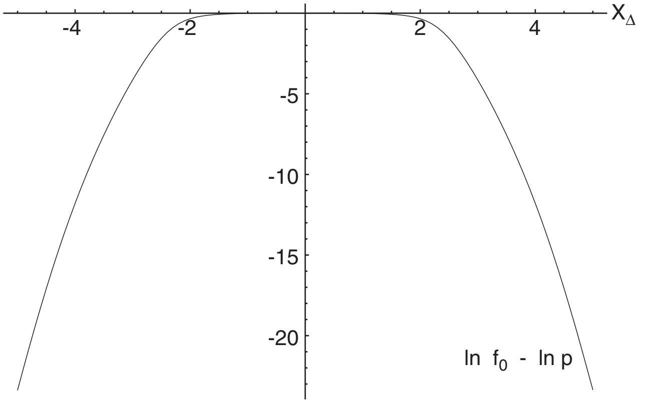

Fig. 10. Tail log-approximation error of the non-Gaussian leading term f_0 (without a further saddlepoint correction) for a jump-diffusion model with normally distributed jumps. This example illustrates the need, when jumps are present, for a saddlepoint correction to the leading term, even when a non-Gaussian base is used.

increments. However, their increments need not be Gaussian (equivalently, their sample paths need not be continuous). By the Lévy–Khintchine formula, a Lévy process X_t has the characteristic function

$$\mathrm{E}[\mathrm{e}^{\mathbf{i}uX_\Delta}|X_0 = 0] = \mathrm{e}^{\Delta\Psi(u)}, \tag{7.1}$$

where $\Psi(u) = \mathbf{i}\mu u - \frac{1}{2}\sigma^2 u^2 + \int_{\mathbb{R}\setminus\{0\}} \mathrm{e}^{\mathbf{i}uw} - 1 - \mathbf{i}uw\mathbf{1}_{\{|w|<1\}}v(\mathrm{d}w)$ and the Lévy measure $v(\cdot)$ satisfies $\int \min\{1, |w|^2\}v(\mathrm{d}w) < \infty$. It is known that a Lévy process can be decomposed into the sum of a Brownian motion with drift, a compound Poisson process with jump size at least 1 and a pure jump process with jump size smaller than 1 (see, e.g., Bertoin (1998) for an introduction to Lévy processes).

For an m-dimensional Lévy process X_t with $X_0 = 0$, the Laplace transform of X_Δ is given by the Lévy–Khintchine formula

$$\varphi(\Delta, u|X_0 = 0) = \mathrm{e}^{-\Delta\Psi(u)},$$

where

$$\begin{aligned}\Psi(u) = & -\frac{1}{2}u^{\mathrm{T}}Qu - d\cdot u + \int_{R^d}(1 - \mathrm{e}^{u\cdot x} + u\cdot h(x))v(\mathrm{d}x)\\ = & -\frac{1}{2}u^{\mathrm{T}}Qu - d\cdot u + \int_{R^d\setminus D}(1 - \mathrm{e}^{u\cdot x})v(\mathrm{d}x) + \int_D (1 - \mathrm{e}^{u\cdot x} + u\cdot h(x))v(\mathrm{d}x)\end{aligned}$$

for some matrix $Q \in \mathbb{R}^{m\times m}$, vector $d \in \mathbb{R}^m$, and truncation function $h : \mathbb{R}^m \to \mathbb{R}^m$ which is bounded, with compact support D, and satisfies $h(x) = x$ in a neighborhood of 0. The Lévy measure v on $\mathbb{R}^m\setminus\{0\}$ satisfies $\int_{\mathbb{R}^m}(1 \wedge |x|^2)v(\mathrm{d}x) < \infty$ and $v(\mathrm{d}x)$ gives the expected number of jumps which fall in dx per unit of time. Each Lévy process is characterized by its (Q, d, v). Its characteristic function is known, and therefore the saddlepoint can be computed.

Remark 1. Numerical Fourier inversion is the numerical integration of $(1/2\pi)\int_{\underline{T}}^{\overline{T}} \mathrm{e}^{-t\cdot x}\varphi(t)\,\mathrm{d}t$ for some truncation points $\underline{T}$ and $\overline{T}$ (see, e.g., Carr and Madan, 1998). For Lévy processes, such truncation does not exploit information about the highest frequency part of the pure-jump component (very small jumps) since

$$\lim_{J\downarrow 0}\int_{|x|\leqslant J}(1-\mathrm{e}^{u\cdot x}+u\cdot h(x))v(\mathrm{d}x)=0$$

for any $u\in[\underline{T},\overline{T}]$. Therefore, over the bounded interval $[\underline{T},\overline{T}]$, the effect of jumps smaller than J on $(1/2\pi)\int_{\underline{T}}^{\overline{T}} \mathrm{e}^{-t\cdot x}\varphi(t)\,\mathrm{d}t$ can be arbitrarily small by making J smaller which implies identification of extremely small jumps is hardly possible after numerical Fourier inversion. In contrast, there is no such truncation in saddlepoint approximation which makes the identification possible.

The cumulant generating function of a Lévy process is

$$K(\varDelta,u|X_0=0)=\varDelta\left[\mu u+\frac{1}{2}\sigma^2u^2+\int \mathrm{e}^{ux}-1-ux\mathbf{1}_{\{|x|<1\}}v(\mathrm{d}x)\right].$$

Choose $c\in(0,1)$ in

$$\begin{aligned}\int_{-c}^{c}\mathrm{e}^{ux}-1-ux\mathbf{1}_{\{|x|<1\}}v(\mathrm{d}x)&=\int_{-c}^{c}\mathrm{e}^{ux}-1-uxv(\mathrm{d}x)\\&=\int_{-c}^{c}\frac{1}{2}u^2x^2+\mathrm{o}(x^2)v(\mathrm{d}x).\end{aligned}$$

Since $\int_{\mathbb{R}^m}(1\wedge|x|^2)v(\mathrm{d}x)<\infty$, we approximate $\int_{-c}^{c}\mathrm{e}^{ux}-1-ux\mathbf{1}_{\{|x|<1\}}v(\mathrm{d}x)$ with

$$\frac{1}{2}u^2\int_{-c}^{c}x^2v(\mathrm{d}x)\equiv\frac{1}{2}u^2\sigma_J^2$$

which captures the behavior of the infinitesimal jumps. Let λ denote

$$\lambda=\int_{|x|\geqslant c}v(\mathrm{d}x).$$

We approximate the transition density using the saddlepoint method with the following non-Gaussian leading term

$$f_0(x)=(1-\lambda\varDelta)\phi(x;\mu\varDelta,(\sigma^2+\sigma_J^2)\varDelta)+\Delta v(x)\mathbf{1}_{\{|x|\geqslant c\}}. \tag{7.2}$$

There is a discontinuity of f_0 at $x=c$. In practice, one can pick c small so that all observations in absolute values are larger than c to avoid this discontinuity. The cumulant generating function corresponding to f_0 is

$$K_0(\varDelta,u|X_0=0)=\ln\left((1-\lambda\varDelta)\exp\left(\mu\Delta u+\frac{1}{2}(\sigma^2+\sigma_J^2)\Delta u^2\right)+\varDelta\int_{|x|\geqslant c}\mathrm{e}^{ux}v(\mathrm{d}x)\right).$$

Theorem 5. *The saddlepoint approximation* $p^{(0)}(\Delta, x|0)$ *is given by*

$$p^{(0)}(\Delta, x|0) = f_0(x)\exp(\{K(\Delta, \hat{u}|0) - \hat{u}x\} - \{K_0(\Delta, \hat{w}|0) - \hat{w}x\}) \\ \times \left(\frac{\partial^2}{\partial w \partial w^{\mathrm{T}}} K_0(\Delta, \hat{w}|0)\right)^{1/2} \left(\frac{\partial^2 K(\Delta, \hat{u}|0)}{\partial u \partial u^{\mathrm{T}}}\right)^{-1/2}, \tag{7.3}$$

where the saddlepoint $\widehat{u}$ *and* $\widehat{w}$ *can be solved from*

$$x = \frac{\partial}{\partial u} K(\Delta, \hat{u}|0)$$

and from

$$x = \frac{\partial}{\partial u} K_0(\Delta, \hat{w}|0).$$

As a first example, consider the variance gamma process, a time-changed Brownian motion:

$$X_t = B(\gamma(t, m, \upsilon), \theta, \sigma),$$

where $B(t, \theta, \sigma) = \theta t + \sigma W_t$ and $\gamma(t+h, m, \upsilon) - \gamma(t, m, \upsilon)$ have Gamma distribution with mean mh and variance υh (see, Madan et al., 1998).

When $m = 1$, the Laplace transform for X_Δ given $X_0 = 0$ is

$$\varphi(\Delta, u|0) = (1 - u\upsilon\theta - \sigma^2 u^2 \upsilon/2)^{-\Delta/\upsilon}$$

and the three characteristics (Q, d, v) of the variance gamma Lévy process are $Q = 0$, $d = 0$ and

$$v(\mathrm{d}x) = \frac{\exp(\theta x/\sigma^2)}{\upsilon|x|} \exp\left(-\frac{\sqrt{2/\upsilon + \theta^2/\sigma^2}}{\sigma}|x|\right)\mathrm{d}x. \tag{7.4}$$

Note that the fact that (7.4) has a singularity of order one at zero implies that Y_t is an infinite activity jump process.

For this model, we can obtain the saddlepoint as

$$\hat{u}(\Delta, X_\Delta | X_0 = 0) = \frac{-X_\Delta\theta\upsilon - \Delta\sigma^2 + \sqrt{X_\Delta^2\theta^2\upsilon^2 + 2X_\Delta^2\upsilon\sigma^2 + \Delta^2\sigma^4}}{X_\Delta\upsilon\sigma^2}.$$

For the variance gamma model, $X_t = \theta\gamma(t, 1, v) + \sigma W_t$, where $\gamma(t+\Delta, 1, v) - \gamma(t, 1, v)$ has Gamma distribution with mean Δ and variance $v\Delta$. The transition density for X_Δ

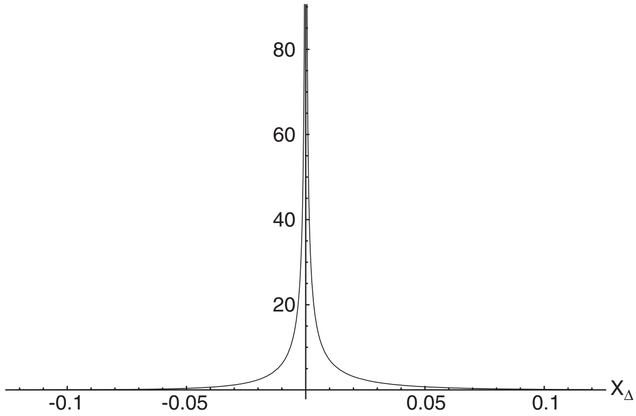

Fig. 11. Transition density $p(\Delta, X_\Delta | X_0)$ for the variance gamma model.

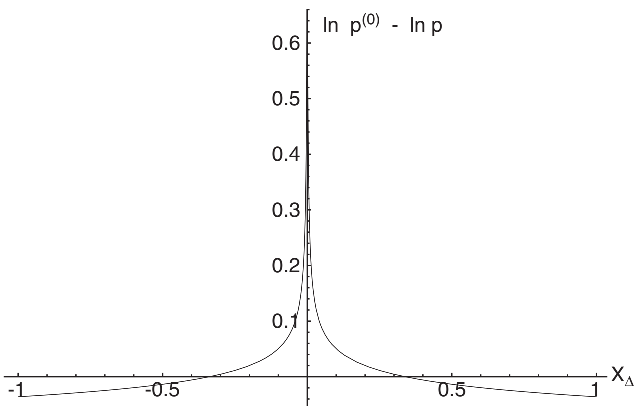

Fig. 12. Log-approximation error of the saddlepoint expansion of order 0 for the variance gamma model, using a non-Gaussian base.

given $X_0 = 0$ is then

$$
\begin{aligned}
& f_{X_\Delta}(x|X_0 = 0) \\
& \quad = \int_0^\infty \frac{1}{\sigma\sqrt{2\pi g}} \exp\left(-\frac{(x-\theta g)^2}{2\sigma^2 g}\right) \frac{g^{\Delta/v-1}\exp(-g/v)}{v^{\Delta/v}\Gamma(\Delta/v)}\, \mathrm{d}g \\
& \quad = \frac{\mathrm{e}^{x\theta/\sigma^2}\sqrt{\frac{2}{\pi}}\left(\frac{\theta^2 v + 2\sigma^2}{x^2 v}\right)^{(-2\Delta+v)/4v} \mathrm{BesselK}_{1/2-\Delta/v}\left(\sqrt{\frac{x^2(\theta^2 v + 2\sigma^2)}{v\sigma^4}}\right)}{v^{\Delta/v}\sigma\Gamma(\frac{\Delta}{v})} \quad \text{when } x \neq 0,
\end{aligned}
$$

where BesselK is the modified Bessel function of the second kind.

As mentioned, the density of the marginals of a Lévy process $Y_t(\vartheta)$ is rarely known in closed form. Instead, it can be obtained by Fourier inversion of the characteristic function whenever the function $\mathrm{e}^{t\psi(u,\vartheta)}$ is integrable, in which case we

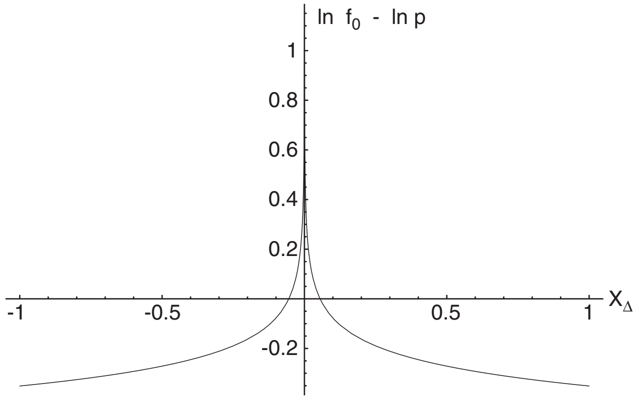

Fig. 13. Log-approximation error of the saddlepoint leading term f_0 for the variance gamma model.

define

$$h(x;\vartheta,\delta) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{e}^{-\mathrm{i}xu}\mathrm{e}^{\delta\psi(u,\vartheta)}\,\mathrm{d}u \tag{7.5}$$

to be the density of the Lévy increment $Y_{t+\delta} - Y_t$.

Setting $\theta = 0.1$, $\sigma = 0.12$, $v = 0.17$, $\Delta = \frac{1}{52}$ and $x = 0$, the true transition density between -0.1 and 0.1 is plotted in Fig. 11. Fig.12 shows the log difference between the saddlepoint density $p^{(0)}(\Delta, y|x)$ and the true transition density. The constant c in (7.2) is set to 0.001. As shown in the plot, the accuracy of saddlepoint approximation is again very good. Fig. 13 plots the log difference between the leading term f_0 and the true transition density. The saddlepoint density $p^{(0)}(\Delta, y|x)$ again improves the tail approximation error relative to the leading term f_0.

8. Conclusions

We have presented a construction of saddlepoint approximations for Markov processes. If the characteristic function is not known explicitly, our method can still be applied by appealing to an expansion in small time of the characteristic function. We have included a number of examples drawn from the recent financial econometrics literature. The examples show that the method can be practically applied to generate transition densities and produce useful results.

Acknowledgements

We are grateful to the Editor and two anonymous referees for their comments. This research was funded in part by the NSF under Grant SES-0111140.

Appendix A. Proof of Theorem 1

Fix u. With $f(y) \equiv \exp(uy)$, we have

$$\varphi(\varDelta, u|x) = \mathrm{E}[\exp(uX_\varDelta)|X_0 = x].$$

At order 1 in $\varDelta$, we obtain the first-order term in the expansion of φ

$$\begin{aligned}\varphi^{(1)}(\varDelta, u|x) &= f(x) + \Delta A \cdot f(x)\\ &= \mathrm{e}^{ux}\{1 + \mu(x)u\varDelta + \tfrac{1}{2}\sigma^2(x)u^2\varDelta\}\end{aligned}$$

and by taking the log at order 1 in $\varDelta$

$$\begin{aligned}K^{(1)}(\varDelta, u|x) &= ux + \mu(x)u\varDelta + \tfrac{1}{2}\sigma^2(x)u^2\varDelta\\ &= (x + \mu(x)\varDelta)u + \tfrac{1}{2}\sigma^2(x)\Delta u^2.\end{aligned}$$

The first-order saddlepoint $\hat{u}^{(1)}$ solves $\partial K^{(1)}(\varDelta, u|x)/\partial u = y$, that is,

$$\hat{u}^{(1)}(\varDelta, y|x) = \frac{y - (x + \mu(x)\varDelta)}{\sigma^2(x)\varDelta}. \tag{A.1}$$

From (3.6), we now have the saddlepoint approximation

$$\begin{aligned}p^{(0,1)}(\varDelta, y|x) &= (2\pi)^{-1/2}\left(\frac{\partial^2 K^{(1)}(\varDelta, \hat{u}^{(1)}|x)}{\partial u^2}\right)^{-1/2} \exp(K^{(1)}(\varDelta, \hat{u}^{(1)}|x) - \hat{u}^{(1)}y)\\ &= (2\pi\Delta\sigma^2(x))^{-1/2}\exp\left(-\frac{(y - x - \mu(x)\varDelta)^2}{\sigma^2(x)\varDelta}\right).\end{aligned}$$

In other words, $p^{(0,1)}$ is the Euler approximation to the transition density of the process.

Appendix B. Proof of Theorem 2

From Theorem 1, we recall that (A.1) and thus, at $y = x + z\varDelta^{1/2}$, we have $u = \widehat{u}^{(1)} + \mathrm{O}(1)$ with

$$\hat{u}^{(1)} = \frac{z}{\sigma^2(x)\varDelta^{1/2}}. \tag{B.1}$$

Now, consider the case $n_2 = 2$. Let $K^{(2)}$ be the cumulant transform at order 2 in $\varDelta$:

$$\begin{aligned}K^{(2)}(\varDelta, u|x) &= \log\varphi^{(2)}(\varDelta, u|x)\\ &= ux + (\mu(x)u + \tfrac{1}{2}\sigma^2(x)u^2)\varDelta + \ln[1 + \tfrac{1}{2}(f_{2,3}u^3 + f_{2,2}u^2 + f_{2,1}u)\varDelta^2]\end{aligned}$$

with $f_{j,k}$ denoting the coefficient of $(\mathbf{i}u)^k$ in $(\partial^j/\partial\varDelta^j)K(\varDelta, \mathbf{i}u|x)|_{\varDelta=0}$; $f_{j,k}$ depends only on x, through $\mu(x)$, $\sigma^2(x)$ and their derivatives. The terms $f_{j,k}$ all have closed-form expressions. Specifically, we have the Taylor expansion in $\varDelta$ of order 2, where we gather the terms according to their dependence in u:

$$K^{(2)}(\Delta, u|x) = u\left(x + \Delta\mu(x) + \frac{\Delta^2\mu(x)\mu'(x)}{2} + \frac{\Delta^2\sigma^2(x)\mu''(x)}{4}\right) + u^2\left(\frac{\Delta\sigma^2(x)}{2} + \frac{\Delta^2\sigma^2(x)\mu'(x)}{2} + \frac{\Delta^2\mu(x)(\sigma^2)'(x)}{4} + \frac{\Delta^2\sigma^2(x)(\sigma^2)''(x)}{8}\right) + \frac{u^3\Delta^2\sigma^2(x)(\sigma^2)'(x)}{4}. \tag{B.2}$$

The saddlepoint $\widehat{u}$ at order 2 solves

$$x + z\Delta^{1/2} = \frac{\partial}{\partial u}K^{(2)}(\Delta, \widehat{u}|x)$$

which in light of (B.2) is a quadratic equation in u. Selecting of the two roots of that quadratic equation the one whose leading term in Δ coincides with (B.1), we obtain that $u = \widehat{u}^{(2)} + \mathrm{O}(\Delta^{1/2})$, where

$$\widehat{u}^{(2)} = \frac{z}{\sigma^2(x)\Delta^{1/2}} - \frac{2\mu(x)\sigma(x) + 3\sigma'(x)z^2}{2\sigma^3(x)}. \tag{B.3}$$

More generally, the form of $K(\Delta, u|x)$ and its Taylor expansion in Δ imply that the partial derivatives of K with respect to u satisfy

$$K_{(n)} \equiv \frac{\partial^k}{\partial u^k}K(\Delta, u|x) = \mathrm{O}(\Delta^{k-1}). \tag{B.4}$$

This follows from

$$\frac{\partial K}{\partial u} = \frac{1}{\varphi}\frac{\partial \varphi}{\partial u}, \quad \frac{\partial^2 K}{\partial u^2} = \frac{1}{\varphi}\frac{\partial^2 \varphi}{\partial u^2} - \left(\frac{1}{\varphi}\frac{\partial \varphi}{\partial u}\right)^2,$$

etc., and therefore by applying (4.5) we have the specific expressions

$$\begin{aligned}
K_{(1)} &= \frac{\partial K(\Delta, u|x)}{\partial u} = x + \mathrm{o}(1),\\
K_{(2)} &= \frac{\partial^2 K(\Delta, u|x)}{\partial u^2} = \Delta\sigma^2(x) + \mathrm{o}(\Delta),\\
K_{(3)} &= \frac{\partial^3 K(\Delta, u|x)}{\partial u^3} = \frac{3}{2}\Delta^2\sigma^2(x)(\sigma^2)'(x) + \mathrm{o}(\Delta^2),\\
K_{(4)} &= \frac{\partial^4 K(\Delta, u|x)}{\partial u^4} = \Delta^3(3\sigma^2(x)(\sigma^2)'(x)^2 + 2\sigma^2(x)^2(\sigma^2)''(x)) + \mathrm{o}(\Delta^3)
\end{aligned} \tag{B.5}$$

so that

$$\begin{aligned}
&\left(\frac{\partial^2 K(\Delta, \hat{u}|x)}{\partial u^2}\right)^{-2}\frac{\partial^4 K(\Delta, \hat{u}|x)}{\partial u^4} = \mathrm{O}(\Delta),\\
&\left(\frac{\partial^2 K(\Delta, \hat{u}|x)}{\partial u^2}\right)^{-3}\left(\frac{\partial^3 K(\Delta, \hat{u}|x)}{\partial u^3}\right)^2 = \mathrm{O}(\Delta).
\end{aligned}$$

Recalling now the general formula (3.7) for the saddlepoint expansion at order $n_1 = 1$, we have

$$p^{(1)}(\Delta, x + z\Delta^{1/2}|x) = \frac{\exp(K(\Delta,\hat{u}|x) - \hat{u}(x + z\Delta^{1/2}))}{\sqrt{2\pi}\left(\frac{\partial^2 K(\Delta,\hat{u}|x)}{\partial u^2}\right)^{1/2}} \left\{1 + \frac{1}{8}\left(\frac{\partial^2 K(\Delta,\hat{u}|x)}{\partial u^2}\right)^{-2} \right.$$
$$\left. \times \frac{\partial^4 K(\Delta,\hat{u}|x)}{\partial u^4} - \frac{5}{24}\left(\frac{\partial^2 K(\Delta,\hat{u}|x)}{\partial u^2}\right)^{-3}\left(\frac{\partial^3 K(\Delta,\hat{u}|x)}{\partial u^3}\right)^2\right\}, \tag{B.6}$$

where, from (B.5), we have

$$1 + \frac{1}{8}\left(\frac{\partial^2 K(\Delta,\hat{u}|x)}{\partial u^2}\right)^{-2}\frac{\partial^4 K(\Delta,\hat{u}|x)}{\partial u^4} - \frac{5}{24}\left(\frac{\partial^2 K(\Delta,\hat{u}|x)}{\partial u^2}\right)^{-3}\left(\frac{\partial^3 K(\Delta,\hat{u}|x)}{\partial u^3}\right)^2$$
$$= 1 + \Delta\left(\frac{1}{4}(\sigma^2)''(x) - \frac{3}{32}\frac{(\sigma^2)'(x)^2}{\sigma^2(x)}\right) + O(\Delta^2) \tag{B.7}$$

and, for the term inside the exponent,

$$K(\Delta,\hat{u}|x) - \hat{u}(x + z\Delta^{1/2})$$
$$= -\frac{z^2}{2\sigma^2(x)} + \left(\frac{z\mu(x)}{\sigma^2(x)} + \frac{z^3(\sigma^2)'(x)}{4\sigma^4(x)}\right)\Delta^{1/2}$$
$$+ \left(-\frac{\mu(x)^2}{2\sigma^2(x)} + \frac{z^2(12\sigma^2(x)(4\mu'(x) + (\sigma^2)''(x)) - 48\mu(x)(\sigma^2)'(x))}{96\sigma^4(x)}\right.$$
$$\left. + \frac{z^4(8\sigma^2(x)(\sigma^2)''(x) - 15(\sigma^2)'(x)^2)}{96\sigma^6(x)}\right)\Delta + O(\Delta^{3/2}). \tag{B.8}$$

Note that this term is obtained as $K(\Delta,\hat{u}|x) - \hat{u}(x + z\Delta^{1/2}) = K^{(3)}(\Delta, \hat{u}^{(2)}|x) - \hat{u}^{(2)}(x + z\Delta^{1/2}) + o(\Delta)$.

As for the determinant, we need to expand at order $\Delta^{3/2}$, to obtain an expression that is correct at order Δ after its leading term

$$p^{(0,0)}(\Delta, x + z\Delta^{1/2}|x) = \frac{1}{\sqrt{2\pi}\Delta^{1/2}}\exp\left(-\frac{z^2}{2\sigma^2(x)}\right) \tag{B.9}$$

(note the presence of the term $\Delta^{1/2}$ in the determinant, this is due to the fact that we are evaluating the density of the original diffusion at $y = x + z\Delta^{1/2}$, this is not the density of the transformed variable z, for which the term $\Delta^{1/2}$ goes away as part of the Jacobian formula). Again from (4.5), we have

$$\begin{aligned}
&\sqrt{2\pi}\left(\frac{\partial^2 K(\Delta,\hat{u}|x)}{\partial u^2}\right)^{1/2} \\
&\quad = \sqrt{2\pi}\sigma(x)\Delta^{1/2}\left\{1+\frac{3z\sigma'(x)}{2\sigma(x)}\Delta^{1/2}+\left(\frac{\mu'(x)}{2}-\frac{\mu(x)\sigma'(x)}{\sigma(x)}\right.\right. \\
&\qquad \left.\left.+\frac{\sigma'(x)^2}{4}+\frac{\sigma(x)\sigma''(x)}{4}+z^2\left(\frac{5\sigma'(x)^2}{8\sigma(x)^2}+\frac{\sigma''(x)}{\sigma(x)}\right)\right)\Delta+\mathrm{O}(\Delta^{3/2})\right\}. \qquad \text{(B.10)}
\end{aligned}$$

Putting together (B.7), (B.8) and (B.10) into (B.6), leads to the result (4.10) in the form

$$p^{(1,1)}(\Delta,x+z\Delta^{1/2}|x)=\frac{\exp(-z^2/(2\sigma^2(x))+e_{1/2}(z|x)\Delta^{1/2}+e_1(z|x)\Delta)}{\sqrt{2\pi}\sigma(x)\Delta^{1/2}\{1+d_{1/2}(z|x)\Delta^{1/2}+d_1(z|x)\Delta\}}\{1+c_1(z|x)\Delta\}. \qquad \text{(B.11)}$$

To compare this expression to the one produced by the irreducible expansion of Aït-Sahalia (2001), we Taylor-expand the form (B.11) in Δ, around the leading term (B.9):

$$\begin{aligned}
p^{(1,1)}(\Delta,x+z\Delta^{1/2}|x) = {}& \frac{1}{\sqrt{2\pi}\Delta^{1/2}}\exp\left(-\frac{z^2}{2\sigma^2(x)}\right)\left\{1+(e_{1/2}(z|x)-d_{1/2}(z|x))\Delta^{1/2}\right. \\
&+\left(c_1(z|x)+d_{1/2}(z|x)^2-d_1(z|x)-d_{1/2}(z|x)e_{1/2}(z|x)\right. \\
&\left.\left.+\frac{1}{2}e_{1/2}(z|x)^2+e_1(z|x)\right)\Delta+\mathrm{O}(\Delta^{3/2})\right\}. \qquad \text{(B.12)}
\end{aligned}$$

On the other hand, the irreducible expansion of Aït-Sahalia (2001) at order 1 in Δ is of the form

$$\begin{aligned}
l^{(1)}(x+z\Delta^{1/2}|x,\Delta) = {}& -\frac{1}{2}\ln(2\pi\Delta)-\ln\sigma(x+z\Delta^{1/2})+\frac{C^{(-1)}(x+z\Delta^{1/2}|x)}{\Delta} \\
&+C^{(0)}(x+z\Delta^{1/2}|x_1)+C^{(1)}(x+z\Delta^{1/2}|x)\Delta, \qquad \text{(B.13)}
\end{aligned}$$

where $l=\ln p$, and

$$\begin{aligned}
C^{(-1)}(y|x) = {}& -\frac{(y-x)^2}{2\sigma(x)^2}+\frac{\sigma'(x)}{2\sigma(x)^3}(y-x)^3+\frac{(4\sigma(x)\sigma''(x)-11\sigma'(x)^2)}{24\sigma(x)^4}(y-x)^4 \\
&+\mathrm{O}((y-x)^5),
\end{aligned}$$

$$\begin{aligned}
C^{(0)}(y|x) = {}& \frac{(2\mu(x)-\sigma(x)\sigma'(x))}{2\sigma(x)^2}(y-x) \\
&+\frac{(\sigma(x)(2\mu'(x)+\sigma'(x)^2)-\sigma(x)^2\sigma''(x)-4\mu(x)\sigma'(x))}{4\sigma(x)^3}(y-x)^2 \\
&+\mathrm{O}((y-x)^3),
\end{aligned}$$

$$C^{(1)}(y|x) = \frac{1}{8}\left(2\sigma(x)\sigma''(x) - \frac{4\mu(x)^2}{\sigma(x)^2} - 4\mu'(x) + \frac{8\mu(x)\sigma'(x)}{\sigma(x)} - \sigma'(x)^2\right) + \mathrm{O}((y-x)).$$

Therefore, we have

$$\ln\sigma(x+z\Delta^{1/2}) = \ln\sigma(x) + \frac{z\sigma'(x)}{\sigma(x)}\Delta^{1/2} + \frac{z^2(\sigma(x)\sigma''(x) - \sigma'(x)^2)}{2\sigma(x)^2}\Delta + \mathrm{O}(\Delta^{3/2}),$$

$$\frac{C^{(-1)}(x+z\Delta^{1/2}|x)}{\Delta} = -\frac{z^2}{2\sigma(x)^2} + \frac{z^3\sigma'(x)}{2\sigma(x)^3}\Delta^{1/2} + \frac{z^4(4\sigma(x)\sigma''(x) - 11\sigma'(x)^2)}{24\sigma(x)^4}\Delta + \mathrm{O}(\Delta^{3/2}),$$

$$C^{(0)}(x+z\Delta^{1/2}|x) = \frac{z(2\mu(x) - \sigma(x)\sigma'(x))}{2\sigma(x)^2}\Delta^{1/2} + \frac{z^2(\sigma(x)(2\mu'(x) + \sigma'(x)^2) - \sigma(x)^2\sigma''(x) - 4\mu(x)\sigma'(x))}{4\sigma(x)^3}\Delta + \mathrm{O}(\Delta^{3/2}),$$

$$C^{(1)}(x+z\Delta^{1/2}|x)\Delta = \frac{1}{8}\left(2\sigma(x)\sigma''(x) - \frac{4\mu(x)^2}{\sigma(x)^2} - 4\mu'(x) + \frac{8\mu(x)\sigma'(x)}{\sigma(x)} - \sigma'(x)^2\right)\Delta + \mathrm{O}(\Delta^{3/2}).$$

Gathering the terms above according to their order in powers of Δ, we see that the irreducible expansion in log-form comparable to (B.12) at order Δ is of the form

$$l^{(1)}(x+z\Delta^{1/2}|x,\Delta) = -\frac{1}{2}\ln(2\pi\Delta) - \ln\sigma(x) - \frac{z^2}{2\sigma(x)^2} + l_{1/2}(z|x)\Delta^{1/2} + l_1(z|x)\Delta + \mathrm{O}(\Delta^{3/2}). \tag{B.14}$$

Taking the exponential, the expansion for the transition density itself (instead of its log) is

$$\begin{aligned}\exp l^{(1)}(x+z\Delta^{1/2}|x,\Delta) &= \frac{1}{\sqrt{2\pi}\Delta^{1/2}}\exp\left(-\frac{z^2}{2\sigma^2(x)} + l_{1/2}(z|x)\Delta^{1/2} + l_1(z|x)\Delta\right)\\ &= \frac{1}{\sqrt{2\pi}\Delta^{1/2}}\exp\left(-\frac{z^2}{2\sigma^2(x)}\right)\left\{1 + l_{1/2}(z|x)\Delta^{1/2}\right.\\ &\quad\left. + \left(\frac{1}{2}l_{1/2}(z|x)^2 + l_1(z|x)\right)\Delta + \mathrm{O}(\Delta^{3/2})\right\}.\end{aligned} \tag{B.15}$$

Replacing the coefficients $l_{1/2}$ and l_1 by their expressions, we see that, for the coefficient of order $\Delta^{1/2}$

$$l_{1/2}(z|x) = e_{1/2}(z|x) - d_{1/2}(z|x) = \frac{z(2\mu(x)\sigma(x) + (z^2 - 3\sigma(x)^2)\sigma'(x))}{2\sigma(x)^3} \tag{B.16}$$

and, for the coefficient of order Δ,

$$\begin{aligned}
&\frac{1}{2} l_{1/2}(z|x)^2 + l_1(z|x) \\
&\quad = c_1(z|x) + d_{1/2}(z|x)^2 - d_1(z|x) - d_{1/2}(z|x)e_{1/2}(z|x) + \frac{1}{2}e_{1/2}(z|x)^2 + e_1(z|x) \\
&\quad = \frac{1}{24\sigma(x)^6}\{12\mu(x)\sigma(x)\sigma'(x)(z^4 - 5z^2\sigma(x)^2 + 2\sigma(x)^4) \\
&\qquad + 3z^6\sigma'(x)^2 - 29z^4\sigma(x)^2\sigma'(x)^2 - 3\sigma(x)^6(4\mu'(x) + \sigma'(x)^2) \\
&\qquad + 3z^2\sigma(x)^4(4\mu'(x) + 15\sigma'(x)^2) - 18z^2\sigma(x)^5\sigma''(x) \\
&\qquad + 12\mu(x)^2\sigma(x)^2(z^2 - \sigma(x)^2) + 4z^4\sigma(x)^3\sigma''(x) + 6\sigma(x)^7\sigma''(x)\}
\end{aligned} \tag{B.17}$$

so that the two expansions (B.12) and (B.15) agree up to the relevant design order Δ.

Appendix C. Proof of Theorem 3

Let

$$q^{(n_2)}(\Delta, y|x) = \mathrm{Re}\left[\frac{1}{2\pi}\int_u \mathrm{e}^{-\mathbf{i}uy}\varphi^{(n_2)}(\Delta, \mathbf{i}u|x)\,\mathrm{d}u\right] \tag{C.1}$$

denote the "density" corresponding to the approximate Laplace transform $\varphi^{(n_2)}(\Delta, u|X_0)$ which is itself a small time expansion at order n_2 in Δ of φ. Integrability is assured because the correction term of $\varphi^{(n_2)}$ is of polynomial order in u. Because the Fourier inverse of an approximate Laplace transform need not be real, Re[·] is used to guarantee that $q^{(n_2)}$ is real.

The proof proceeds in two steps: (i) first, we deal with the approximation introduced by the Taylor expansion in Δ of the Laplace transform, and show that

$$q^{(n_2)}(\Delta, x + z\Delta^{1/2}|x) = p(\Delta, x + z\Delta^{1/2}|x)(1 + \mathrm{O}(\Delta^{n_2/2})). \tag{C.2}$$

(ii) Second, we deal with the approximation introduced by the fact that the Fourier inversion is replaced by its saddlepoint approximation, and show that

$$q^{(n_2)}(\Delta, x + z\Delta^{1/2}|x) = p^{(n_1,n_2)}(\Delta, x + z\Delta^{1/2}|x)(1 + \mathrm{O}(\Delta^{n_1+1})). \tag{C.3}$$

For the sake of concreteness, we show below the computations for $n_2 = 1, 2$ and $n_1 = 0, 1, 2$, and outline briefly how the proof in the general case proceeds similarly.

For step (i), $n_2 = 1$ reduces to the case treated in Theorem 1. When $n_2 = 2$, we have

$$\varphi^{(2)}(\Delta, \mathbf{i}u|x) = \exp\left(\mathbf{i}ux + \left(\mathbf{i}\mu(x)u - \frac{1}{2}\sigma^2(x)u^2\right)\Delta\right) \times \left[1 + \frac{1}{2}\frac{\partial^2}{\partial \Delta^2}K(\Delta, \mathbf{i}u|x)\bigg|_{\Delta=0}\Delta^2\right],$$

where

$$\frac{\partial^2}{\partial \Delta^2}K(\Delta, \mathbf{i}u|x)\bigg|_{\Delta=0} = -\mathbf{i}f_{2,3}u^3 - f_{2,2}u^2 + \mathbf{i}f_{2,1}u$$

with $f_{j,k}$ denoting the coefficient of $(\mathbf{i}u)^k$ in $(\partial^j/\partial\Delta^j)K(\Delta, \mathbf{i}u|x)|_{\Delta=0}$; $f_{j,k}$ depends only on x, through $\mu(x)$, $\sigma(x)$ and their derivatives. As an example, $f_{2,3}(x) = \sigma(x)^3\sigma'(x)$.

Evaluating at $y = x + z\Delta^{1/2}$, we have

$$\begin{aligned}
&\frac{1}{2\pi}\mathrm{Re}\left[\int_u \mathrm{e}^{-\mathbf{i}uy}\mathrm{e}^{\mathbf{i}ux+(\mathbf{i}\mu(x)u-(1/2)\sigma^2(x)u^2)\Delta}\frac{\partial^2}{\partial\Delta^2}K(\Delta, \mathbf{i}u|x)\bigg|_{\Delta=0}\Delta^2\,\mathrm{d}u\right]\\
&\quad= \frac{1}{2\pi}\Delta^2\mathrm{Re}\left[\int_u \mathrm{e}^{-\mathbf{i}uz\Delta^{1/2}+(\mathbf{i}\mu(x)u-(1/2)\sigma^2(x)u^2)\Delta}(-\mathbf{i}f_{2,3}u^3 - f_{2,2}u^2 + \mathbf{i}f_{2,1}u)\,\mathrm{d}u\right]\\
&\quad= \frac{1}{2\pi}\Delta^2\mathrm{Re}\left[\int_u (\cos(u(z\Delta^{1/2} - \mu(x)\Delta)) - \mathbf{i}\sin(u(z\Delta^{1/2} - \mu(x)\Delta)))\right.\\
&\qquad\left.\times\mathrm{e}^{-(1/2)\sigma^2(x)u^2\Delta}(-\mathbf{i}f_{2,3}u^3 - f_{2,2}u^2 + \mathbf{i}f_{2,1}u)\,\mathrm{d}u\right]\\
&\quad= \frac{1}{2\pi}\Delta^2\left\{-f_{2,3}\int_u \sin(u(z\Delta^{1/2} - \mu(x)\Delta))\mathrm{e}^{-(1/2)\sigma^2(x)u^2\Delta}u^3\,\mathrm{d}u\right.\\
&\qquad - f_{2,2}\int_u \cos(u(z\Delta^{1/2} - \mu(x)\Delta))\mathrm{e}^{-(1/2)\sigma^2(x)u^2\Delta}u^2\,\mathrm{d}u\\
&\qquad\left. + f_{2,1}\int_u \sin(u(z\Delta^{1/2} - \mu(x)\Delta))\mathrm{e}^{-(1/2)\sigma^2(x)u^2\Delta}u\,\mathrm{d}u\right\}.
\end{aligned}$$

The integrals above have closed-form expressions. Specifically,

$$\begin{aligned}
&\int_{-\infty}^{+\infty}\sin(u(z\Delta^{1/2} - \mu(x)\Delta))\mathrm{e}^{-(1/2)\sigma^2(x)u^2\Delta}u^k\,\mathrm{d}u = \begin{cases}0 & \text{if } k \text{ even},\\ \mathrm{O}(\Delta^{-(k+1)/2}) & \text{if } k \text{ odd},\end{cases}\\
&\int_{-\infty}^{+\infty}\cos(u(z\Delta^{1/2} - \mu(x)\Delta))\mathrm{e}^{-(1/2)\sigma^2(x)u^2\Delta}u^k\,\mathrm{d}u = \begin{cases}\mathrm{O}(\Delta^{-(k+1)/2}) & \text{if } k \text{ even},\\ 0 & \text{if } k \text{ odd}.\end{cases}
\end{aligned} \tag{C.4}$$

For instance,

$$\int_{-\infty}^{+\infty} \sin(u(z\Delta^{1/2} - \mu(x)\Delta))\mathrm{e}^{-(1/2)\sigma^2(x)u^2\Delta} u \,\mathrm{d}u = \mathrm{e}^{-z^2/(2\sigma(x)^2)} \frac{\sqrt{2\pi}z}{\sigma(x)^3} \Delta^{-1} + \mathrm{O}(\Delta^{-1/2}),$$

$$\int_{-\infty}^{+\infty} \cos(u(z\Delta^{1/2} - \mu(x)\Delta))\mathrm{e}^{-(1/2)\sigma^2(x)u^2\Delta} \,\mathrm{d}u = \mathrm{e}^{-z^2/(2\sigma(x)^2)} \frac{\sqrt{2\pi}}{\sigma(x)} \Delta^{-1/2} + \mathrm{O}(1).$$

Using these closed-form expressions, it can be calculated that

$$q^{(1)}(\Delta, x + z\Delta^{1/2}|x) = \frac{\mathrm{e}^{-z^2/(2\sigma(x)^2)}}{\sqrt{2\pi\sigma^2(x)\Delta}} \left(1 + g_1^{(1)}\sqrt{\Delta} + \mathrm{O}(\Delta)\right),$$

$$q^{(2)}(\Delta, x + z\Delta^{1/2}|x) = \frac{\mathrm{e}^{-z^2/(2\sigma(x)^2)}}{\sqrt{2\pi\sigma^2(x)\Delta}} \left(1 + g_1^{(2)}\sqrt{\Delta} + \mathrm{O}(\Delta)\right),$$

where

$$g_1^{(2)} = \frac{2\mu(x)\sigma(x) + (z^2 - 3\sigma^2(x))\sigma'(x)}{2\sigma^3(x)} z \quad \text{and} \quad g_1^{(1)} \neq g_1^{(2)}.$$

On the other hand, we know from Aït-Sahalia (2002) that

$$p(\Delta, x + z\Delta^{1/2}|x) = \frac{\mathrm{e}^{-z^2/(2\sigma(x)^2)}}{\sqrt{2\pi\sigma^2(x)\Delta}} \left(1 + g_1^{(2)}\sqrt{\Delta} + \mathrm{O}(\Delta)\right).$$

Therefore,

$$q^{(1)}(\Delta, x + z\Delta^{1/2}|x) = p(\Delta, x + z\Delta^{1/2}|x)\left(1 + \mathrm{O}\left(\sqrt{\Delta}\right)\right),$$

$$q^{(2)}(\Delta, x + z\Delta^{1/2}|x) = p(\Delta, x + z\Delta^{1/2}|x)(1 + \mathrm{O}(\Delta)).$$

A by-product of the calculation above is to show us how we need to approximate $\varphi^{(n_2)}$ for $n_2 > 2$:

$$K(\Delta, u|x) = ux + \frac{\partial}{\partial\Delta} K(\Delta, u|x)\bigg|_{\Delta=0} \Delta + \frac{1}{2}\frac{\partial^2}{\partial\Delta^2} K(\Delta, u|x)\bigg|_{\Delta=0} \Delta^2 + \cdots + \mathrm{o}(\Delta^{n_2}), \tag{C.5}$$

$$\varphi(\Delta, u|x) = \exp\left(ux + \frac{\partial}{\partial\Delta} K(\Delta, u|x)\bigg|_{\Delta=0} \Delta\right) \times \left[1 + \frac{1}{2}\frac{\partial^2}{\partial\Delta^2} K(\Delta, u|x)\bigg|_{\Delta=0} \Delta^2 + \cdots + \mathrm{o}(\Delta^{n_2})\right]. \tag{C.6}$$

The derivative $(\partial^k/\partial\Delta^k)K(\Delta, u|x)|_{\Delta=0}$ is an order $k+1$ polynomial in u. This, together with (C.4), gives the order information for grouping correction terms. Note, in particular, that in (C.6) the grouping of the first two terms is inside the exponential.

In step (ii), we now show (C.3). When $n_2 = 1$, $\varphi^{(1)}$ and hence $q^{(1)}$ correspond to a normal distribution with mean $x + \mu(x)\Delta$ and variance $\sigma(x)\Delta$. From Theorem 1, we recall that (A.1) and thus, at $y = x + z\Delta^{1/2}$, we have $u = \widehat{u}^{(1)} + \mathrm{O}(1)$ with

$\hat{u}^{(1)} = z/(\sigma^2(x)\Delta^{1/2})$. Furthermore, $q^{(1)}$ coincides with the density from the saddlepoint approximation, and the claim is true.

Now, consider the case $n_2 = 2$. We have

$$\begin{aligned} q^{(2)}(\Delta, x + z\Delta^{1/2}|X_0 = x) &= \mathrm{Re}\left[\frac{1}{2\pi\mathbf{i}}\int_{\widehat{u}-\mathbf{i}\infty}^{\widehat{u}+\mathbf{i}\infty} \mathrm{e}^{K^{(n_2)}(u)-u(x+z\Delta^{1/2})}\,\mathrm{d}u\right] \\ &= \mathrm{Re}\left[\frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{e}^{K^{(n_2)}(\widehat{u}+\mathbf{i}w)-(\widehat{u}+\mathbf{i}w)(x+z\Delta^{1/2})}\,\mathrm{d}w\right]. \end{aligned}$$

It can then be calculated that

$$\begin{aligned} &[K^{(n_2)}(\widehat{u}+\mathbf{i}w) - (\widehat{u}+\mathbf{i}w)(x+z\Delta^{1/2})] - [K^{(n_2)}(\widehat{u}) - \widehat{u}(x+z\Delta^{1/2})] \\ &\quad = -\mathbf{i}w(z\Delta^{1/2} - \mu(x)\Delta - \sigma^2(x)\widehat{u}\Delta) - \frac{1}{2}\sigma^2(x)\Delta w^2 \\ &\qquad + \ln\left(\frac{1+1/2(f_{2,3}(\widehat{u}+\mathbf{i}w)^3 + f_{2,2}(\widehat{u}+\mathbf{i}w)^2 + f_{2,1}(\widehat{u}+\mathbf{i}w))\Delta^2}{1+1/2(f_{2,3}\widehat{u}^3 + f_{2,2}\widehat{u}^2 + f_{2,1}\widehat{u})\Delta^2}\right). \end{aligned}$$

Therefore,

$$\begin{aligned} &q^{(2)}(\Delta, x + z\Delta^{1/2}|X_0 = x) \\ &\quad = \mathrm{e}^{K^{(n_2)}(\widehat{u})-\widehat{u}(x+z\Delta^{1/2})}\mathrm{Re}\left[\frac{1}{2\pi}\int_{-\infty}^{\infty}\exp\left[-\mathbf{i}w(z\Delta^{1/2} - \mu(x)\Delta - \sigma^2(x)\widehat{u}\Delta) - \frac{1}{2}\sigma^2(x)\Delta w^2\right]\right. \\ &\qquad \left.\times\frac{1+1/2(f_{2,3}(\widehat{u}+\mathbf{i}w)^3 + f_{2,2}(\widehat{u}+\mathbf{i}w)^2 + f_{2,1}(\widehat{u}+\mathbf{i}w))\Delta^2}{1+1/2(f_{2,3}\widehat{u}^3 + f_{2,2}\widehat{u}^2 + f_{2,1}\widehat{u})\Delta^2}\,\mathrm{d}w\right]. \end{aligned}$$

For k even,

$$\begin{aligned} &\mathrm{Re}\left[\frac{1}{2\pi}\int_{-\infty}^{\infty}\exp\left[-\mathbf{i}w(z\Delta^{1/2} - \mu(x)\Delta - \sigma^2(x)\widehat{u}\Delta) - \frac{1}{2}\sigma^2(x)\Delta w^2\right](\mathbf{i}w)^k\,\mathrm{d}w\right] \\ &\quad = \frac{(-\mathbf{1})^{k/2}}{2\pi}\int_{-\infty}^{\infty}\cos[w(z\Delta^{1/2} - \mu(x)\Delta - \sigma^2(x)\widehat{u}\Delta)] \\ &\qquad \times\exp\left[-\frac{1}{2}\sigma^2(x)\Delta w^2\right]w^k\,\mathrm{d}w \end{aligned} \tag{C.7}$$

while for k odd,

$$\begin{aligned} &\mathrm{Re}\left[\frac{1}{2\pi}\int_{-\infty}^{\infty}\exp\left[-\mathbf{i}w(z\Delta^{1/2} - \mu(x)\Delta - \sigma^2(x)\widehat{u}\Delta) - \frac{1}{2}\sigma^2(x)\Delta w^2\right](\mathbf{i}w)^k\,\mathrm{d}w\right] \\ &\quad = \frac{(-\mathbf{1})^{(k-1)/2}}{2\pi}\int_{-\infty}^{\infty}\sin[w(z\Delta^{1/2} - \mu(x)\Delta - \sigma^2(x)\widehat{u}\Delta)] \\ &\qquad \times\exp\left[-\frac{1}{2}\sigma^2(x)\Delta w^2\right]w^k\,\mathrm{d}w. \end{aligned} \tag{C.8}$$

Now (B.3) implies

$$z\Delta^{1/2} - \mu(x)\Delta - \sigma^2(x)\widehat{u}\Delta = \frac{3\sigma'(x)z^2\Delta}{2\sigma(x)}(1 + \mathrm{O}(\Delta^{1/2})).$$

From the proof of step (i), recall that (C.7) and (C.8) have closed-form expressions. In particular, it follows that

$$\begin{aligned}
&\int_{-\infty}^{\infty} \sin[w(z\Delta^{1/2} - \mu(x)\Delta - \sigma^2(x)\widehat{u}\Delta)] \exp\left[-\frac{1}{2}\sigma^2(x)\Delta w^2\right] w^k \,\mathrm{d}w \\
&\quad = \begin{cases} 0 & \text{if } k \text{ even}, \\ \mathrm{O}(\Delta^{-k/2}) & \text{if } k \text{ odd}, \end{cases} \\
&\int_{-\infty}^{\infty} \cos[w(z\Delta^{1/2} - \mu(x)\Delta - \sigma^2(x)\widehat{u}\Delta)] \exp\left[-\frac{1}{2}\sigma^2(x)\Delta w^2\right] w^k \,\mathrm{d}w \\
&\quad = \begin{cases} \mathrm{O}(\Delta^{-(k+1)/2}) & \text{if } k \text{ even}, \\ 0 & \text{if } k \text{ odd}. \end{cases}
\end{aligned}$$

For example,

$$\begin{aligned}
&\int_{-\infty}^{\infty} \sin[w(z\Delta^{1/2} - \mu(x)\Delta - \sigma^2(x)\widehat{u}\Delta)] \exp\left[-\frac{1}{2}\sigma^2(x)\Delta w^2\right] w \,\mathrm{d}w = \frac{3\sqrt{2\pi}z^2\sigma'(x)}{2\sqrt{\Delta}\sigma^4(x)}, \\
&\int_{-\infty}^{\infty} \cos[w(z\Delta^{1/2} - \mu(x)\Delta - \sigma^2(x)\widehat{u}\Delta)] \exp\left[-\frac{1}{2}\sigma^2(x)\Delta w^2\right] \mathrm{d}w = \frac{\sqrt{2\pi}}{\sqrt{\Delta}\sigma(x)}.
\end{aligned}$$

It can then be calculated that

$$q^{(2)}(\Delta, x + z\Delta^{1/2}|x) = \frac{1}{\sqrt{2\pi\sigma^2(x)\Delta}} \mathrm{e}^{K^{(2)}(\widehat{u}) - \widehat{u}(x + z\Delta^{1/2})} \left[1 - \frac{3\sigma'(x)z}{2\sigma(x)}\sqrt{\Delta} + \mathrm{O}(\Delta)\right].$$

From the general saddlepoint formula (3.6), it follows that

$$p^{(0,2)}(\Delta, x + z\Delta^{1/2}|x) = \frac{1}{\sqrt{2\pi\sigma^2(x)\Delta}} \mathrm{e}^{K^{(2)}(\widehat{u}) - \widehat{u}(x + z\Delta^{1/2})} \left(1 - \frac{3\sigma'(x)z}{2\sigma(x)}\sqrt{\Delta} + \mathrm{O}(\Delta)\right).$$

Therefore,

$$\begin{aligned}
p^{(0,2)}(\Delta, x + z\Delta^{1/2}|x) &= q^{(2)}(\Delta, x + z\Delta^{1/2}|x)(1 + \mathrm{O}(\Delta)) \\
&= p(\Delta, x + z\Delta^{1/2}|x)(1 + \mathrm{O}(\Delta)).
\end{aligned}$$

Higher-order saddlepoint approximations ($n_1 > 0$) can be constructed, using again (B.4).

Appendix D. Proof of Lemma 1

Fix u. With $f(y) \equiv \exp(uy)$, we have

$$\varphi(\Delta, u|x) = \mathrm{E}[\exp(uX_\Delta)|X_0 = x].$$

At order 1 in Δ, we obtain the first-order term in the expansion of φ:

$$\begin{aligned}\varphi^{(1)}(\Delta, u|x) &= f(x) + \Delta A \cdot f(x) \\ &= \mathrm{e}^{ux}\{1 + \Delta[\mu(x)u + \tfrac{1}{2}\sigma^2(x)u^2 + \lambda(\theta(u) - 1)]\}\end{aligned}$$

and by taking the log at order 1 in Δ:

$$K^{(1)}(\Delta, u|x) = ux + \mu(x)u\Delta + \tfrac{1}{2}\sigma^2(x)u^2\Delta + \Delta\lambda(\theta(u) - 1),$$

$$\frac{\partial K^{(1)}(\Delta, u|x)}{\partial u} = x + \mu(x)\Delta + \sigma^2(x)u\Delta + \Delta\lambda\theta'(u),$$

$$\frac{\partial^2 K^{(1)}(\Delta, u|x)}{\partial u^2} = \sigma^2(x)\Delta + \Delta\lambda\theta''(u).$$

The first-order saddlepoint $\hat{u}^{(1)}$ solves $\partial K^{(1)}(\Delta, u|x)/\partial u = y$, that is,

$$\sigma^2(x)\hat{u}^{(1)}\Delta + \Delta\lambda\theta'(\hat{u}^{(1)}) = y - x - \mu(x)\Delta,$$

$$\hat{u}^{(1)} = \frac{y - x - \mu(x)\Delta}{\sigma^2(x)\Delta} - \frac{\lambda\theta'(\hat{u}^{(1)})}{\sigma^2(x)}.$$

Then, evaluating at $\hat{u}^{(1)}$, we have

$$\begin{aligned}&K^{(1)}(\Delta, \hat{u}^{(1)}|x) - \hat{u}^{(1)}y \\ &\quad= \hat{u}^{(1)}(x - y) + \mu(x)\hat{u}^{(1)}\Delta + \tfrac{1}{2}\sigma^2(x)\hat{u}^{(1)2}\Delta + \Delta\lambda(\theta(\hat{u}^{(1)}) - 1) \\ &\quad= \hat{u}^{(1)}(-\sigma^2(x)u\Delta - \Delta\lambda\theta'(\hat{u}^{(1)})) + \tfrac{1}{2}\sigma^2(x)\hat{u}^{(1)2}\Delta + \Delta\lambda(\theta(\hat{u}^{(1)}) - 1) \\ &\quad= -\tfrac{1}{2}\sigma^2(x)\hat{u}^{(1)2}\Delta + \Delta\lambda(\theta(\hat{u}^{(1)}) - \hat{u}^{(1)}\theta'(\hat{u}^{(1)}) - 1).\end{aligned}$$

We now have the saddlepoint approximation

$$\begin{aligned}p^{(0,1)}(\Delta, y|x) &= (2\pi)^{-1/2}\left(\frac{\partial^2 K^{(1)}(\Delta, \hat{u}^{(1)}|x)}{\partial u^2}\right)^{-1/2} \exp(K^{(1)}(\Delta, \hat{u}^{(1)}|x) - \hat{u}^{(1)}y) \\ &= \frac{\exp(-(1/2)\sigma^2(x)\hat{u}^{(1)2}\Delta)}{(2\pi(\sigma^2(x)\Delta + \Delta\lambda\theta''(\hat{u}^{(1)})))^{1/2}}\exp(\Delta\lambda(\theta(\hat{u}^{(1)}) - \hat{u}^{(1)}\theta'(\hat{u}^{(1)}) - 1)).\end{aligned}$$

References

Aït-Sahalia, Y., 1999. Transition densities for interest rate and other nonlinear diffusions. Journal of Finance 54, 1361–1395.

Aït-Sahalia, Y., 2001. Closed-form likelihood expansions for multivariate diffusions. Technical Report, Princeton University.

Aït-Sahalia, Y., 2002. Maximum-likelihood estimation of discretely-sampled diffusions: a closed-form approximation approach. Econometrica 70, 223–262.

Aït-Sahalia, Y., Hansen, L.P., Scheinkman, J.A., 2002. Operator methods for continuous-time Markov processes. In: Aït-Sahalia, Y., Hansen, L.P. (Eds.), Handbook of Financial Econometrics. North Holland, Amsterdam, The Netherlands, forthcoming.

Bakshi, G.S., Yu, N., 2005. A refinement to Aït-Sahalia's (2002) "Maximum likelihood estimation of discretely sampled diffusions: a closed-form approximation approach". Journal of Business 78, forthcoming.

Bertoin, J., 1998. Lévy Processes. Cambridge University Press, Cambridge, UK.

Carr, P., Madan, D.B., 1998. Option valuation using the fast Fourier transform. Journal of Computational Finance 2, 61–73.

Cox, J.C., Ingersoll, J.E., Ross, S.A., 1985. A theory of the term structure of interest rates. Econometrica 53, 385–408.

Daniels, H., 1954. Saddlepoint approximations in statistics. Annals of Mathematical Statistics 25, 631–650.

DiPietro, M., 2001. Bayesian inference for discretely sampled diffusion processes with financial applications. Ph.D. Thesis, Department of Statistics, Carnegie-Mellon University.

Easton, G.S., Ronchetti, E., 1986. General saddlepoint approximations with applications to L-statistics. Journal of the American Statistical Association 81, 420–430.

Egorov, A.V., Li, H., Xu, Y., 2003. Maximum likelihood estimation of time inhomogeneous diffusions. Journal of Econometrics 114, 107–139.

Heston, S., 1993. A closed-form solution for options with stochastic volatility with applications to bonds and currency options. Review of Financial Studies 6, 327–343.

Jensen, B., Poulsen, R., 2002. Transition densities of diffusion processes: numerical comparison of approximation techniques. Journal of Derivatives 9, 1–15.

Jensen, J., 1995. Saddlepoint Approximations. Oxford University Press, Oxford, UK.

Lieberman, O., 1994. On the approximation of saddlepoint expansions in statistics. Econometric Theory 10, 900–916.

Lugannani, R., Rice, S., 1980. Saddlepoint approximation for the distribution of the sum of independent random variables. Advances in Applied Probability 12, 475–490.

Madan, D.B., Carr, P.P., Chang, E.E., 1998. The variance gamma process and option pricing. European Finance Review 2, 79–105.

McCullagh, P., 1987. Tensor Methods in Statistics. Chapman & Hall, London, UK.

Merton, R.C., 1976. Option pricing when underlying stock returns are discontinuous. Journal of Financial Economics 3, 125–144.

Phillips, P.C., 1986. Large deviation expansions in econometrics. In: Slottje, D.J. (Ed.), Advances in Econometrics, vol. 5. JAI Press, Greenwich, CT, pp. 199–226.

Revuz, D., Yor, M., 1994. Continuous Martingales and Brownian Motion, second ed. Springer, Berlin, Germany.

Rogers, L., Zane, O., 1999. Saddlepoint approximations to option prices. Annals of Applied Probability 9, 493–503.

Schaumburg, E., 2001. Maximum likelihood estimation of jump processes with applications to finance. Ph.D. Thesis, Princeton University.

Wood, A.T.A., Booth, J.G., Butler, R.W., 1993. Saddlepoint approximations to the CDF of some statistics with nonnormal limit distributions. Journal of the American Statistical Association 88, 680–686.

Yu, J., 2003. Closed-form likelihood estimation of jump-diffusions with an application to the realignment risk premium of the Chinese yuan. Ph.D. Thesis, Princeton University.