SPRINGER BRIEFS IN QUANTITATIVE FINANCE

Yue Kuen Kwok · Wendong Zheng

Saddlepoint Approximation Methods in Financial Engineering

Springer

SpringerBriefs in Quantitative Finance

More information about this series at http://www.springer.com/series/8784

Yue Kuen Kwok · Wendong Zheng

Saddlepoint Approximation Methods in Financial Engineering

Springer

Yue Kuen Kwok
Department of Mathematics
Hong Kong University of Science and Technology
Hong Kong
China

Wendong Zheng
Department of Mathematics
Hong Kong University of Science and Technology
Hong Kong
China

ISSN 2192-7006 ISSN 2192-7014 (electronic)
SpringerBriefs in Quantitative Finance
ISBN 978-3-319-74100-0 ISBN 978-3-319-74101-7 (eBook)
https://doi.org/10.1007/978-3-319-74101-7

Library of Congress Control Number: 2017964441

Mathematics Subject Classification (2010): 62P05, 62E17, 44A10

Printed on acid-free paper

This Springer imprint is published by the registered company Springer International Publishing AG part of Springer Nature
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

Objectives and Audience

Financial institutions always strive for effective valuation of prices of exotic financial derivatives and risk positions of portfolios of risky instruments. Most problems in pricing financial derivatives and risk measures calculation in credit portfolios involve computation of tail probabilities and tail expectations of the corresponding underlying state variables. It is desirable to derive efficient, reliable and accurate analytic approximation formulas for derivatives pricing and risk measures computation. The saddlepoint approximation approach has been shown to be a versatile tool in statistics that is based on the steepest descent method to derive analytic approximation formulas of the Laplace inversion integrals for density functions, tail probabilities and tail expectation of a sum of independent and identically distributed random variables. There has been a growing literature in recent decades in applying the saddlepoint approximation methods in pricing exotic financial derivatives and calculating risk measures and risk contributions in credit portfolios under various default correlation models.

In this book, we summarize recent research results in applying the saddlepoint approximation methods in pricing exotic derivatives and calculating risk measures of credit portfolios. We start with the presentation of derivation of a variety of saddlepoint approximation formulas under different applications so that researchers new to the topic are shown the fine technicalities in various saddlepoint approximation approaches. The exposition and style are made rigorous by providing formal proofs of most of the results. Also, this book is made to be self-contained by providing the relevant background knowledge of the analytic tools used in the derivation procedures found in Appendices of the chapters. Illustrative numerical examples are included to help readers make assessment on accuracy and effectiveness of the saddlepoint approximation formulas under various applications. In a few cases, we show how some saddlepoint approximation formulas may fail to produce accurate results.

This book will be valuable to researchers in saddlepoint approximation since it offers a single source for most of the saddlepoint approximation results in financial engineering, with different sets of ready-to-use approximation formulas. Many of these results may otherwise only be found in the original research publications. It is also well suited as a textbook on the subject for courses in financial mathematics at universities.

Guide to the Chapters

This book contains five chapters. The foundations are laid in Chap. 1 in terms of generalized Fourier transform (or characteristic function) of the probability density function of a random variable and the inverse Fourier transform, by which the probability density function can be recovered from the characteristic function. The Bromwich integral formulas of the density, tail probability and tail expectation are expressed in terms of the cumulant generating function of the underlying random variable. The steepest descent method is applied to obtain the asymptotic expansion of a Fourier or Laplace integral. As the basic essence of the saddlepoint approximation approach, a Bromwich contour in the complex plane is deformed to pass through a saddlepoint along the path of steepest descent.

In Chap. 2, we present the detailed procedures of deriving various saddlepoint approximation formulas for the density functions, tail probabilities and tail expectations of random variables under both the Gaussian and non-Gaussian base distribution, using the steepest descent method and the technique of exponentially tilted Edgeworth expansion. In this chapter, the saddlepoint approximation formulas are derived based on the assumption that the cumulant generating function of the underlying random variable is known in closed from. Beyond the theoretical derivation and numerical examples, we explore the scenarios under which some saddlepoint approximation formulas may fail. Since an Edgeworth expansion is expanded around the mean, it is envisioned that expansion may fail in the tail of the distribution. For this reason, the exponentially tilted Edgeworth expansions are considered, in which one performs the expansion in the region of interest. This would lead to more accurate saddlepoint approximations for tail probabilities and tail expectations, which are of great use in credit risk portfolio models and option pricing in later chapters.

In Chap. 3, the saddlepoint approximation methods from two research papers are presented for continuous time Markov processes of affine jump diffusion, in which case a closed-form solution for the cumulant generating function is not available. In the first section of the chapter, a Taylor expansion in small time is used to obtain the cumulant generating function and the saddlepoint. An error analysis of the derived approximation is provided. In the second section, the focus is on the affine jump-diffusion model in which the characteristic function of the process has an exponential affine structure. The characteristic function can be solved in terms of a system of Riccati ordinary differential equations, either explicitly or numerically. On this basis, the saddlepoint can be approximated by combining a root-finding algorithm together with the series inversion techniques.

In Chap. 4, we show the decomposition of the price of a European call option as the difference of two tail expectations under the risk neutral measure and share measure. Provided that closed-form formulas of the cumulant generating function for the underlying asset price process under the two measures exist, the celebrated Lugannani–Rice formula can be applied to obtain the saddlepoint approximation for the European call option. This technique is extended to pricing European options under stochastic volatility and interest rate. We also show how to use the saddlepoint approximation method for pricing VIX futures and options under stochastic volatility model with jumps. The highlight of the chapter is the derivation of the saddlepoint approximation formulas for pricing options on discrete realized variance under the Lévy processes and stochastic volatility processes with jumps.

In Chap. 5, the saddlepoint approximation methods are applied to risk measures calculations in credit risk portfolios and pricing of the tranched Collateralized Debt Obligations. Two classes of default correlation models are considered in detail, the CreditRisk^{+} and Gaussian copula models. Risk measures are introduced in the axiomatic setting of coherent measure. Various saddlepoint approximation formulas are derived for calculating the Value-at-Risk and expected shortfall of the losses in credit portfolios, together with the computation of risk contributions of the obligors to these two risk measures. As an alternative to the computationally intensive recursive scheme to generate the loss distribution in the CreditRisk^{+} model, efficient analytical saddlepoint approximations are derived. Illustrative numerical examples are presented to demonstrate numerical performance of different saddlepoint approximation formulas in computing risk measures of credit portfolios under the Vasicek one-factor default model. We also show how the calculations of the fair spread rates of the different tranches in the Collateralized Debt Obligations can be effected by computing the tail expectations of the credit losses using appropriate saddlepoint approximation formulas.

Acknowledgements

This book benefits from various dialogues with colleagues and students in research seminars and classroom discussions. Some materials in the books are research results with former students, and special thanks go to Mr. Yue Wang and Mr. Yuantao Zhang. The book also benefits from the helpful comments from the editors and anonymous referees. In addition, we would like to thank Ms. Angela Yao for her painstaking jobs of careful typing of the manuscript. She has

entertained seemingly endless changes in the editing process. The first author would like to thank his wife, Oi Chun, for her love and care that have always been his source of support in life.

Hong Kong, China
March 2018

Yue Kuen Kwok
Wendong Zheng

Contents

Chapter 1
Cumulant Generating Functions and Steepest Descent Method

The density functions of most stochastic processes that model the stochastic evolution of the underlying state variables in financial engineering models do not admit closed form analytic representation while their generalized Fourier transform and Laplace transform may have closed form formulas. The pricing of derivatives and computation of risk measures in general involve the evaluation of Fourier or Laplace type integrals. First, we discuss the fundamental properties related to the characteristic functions and cumulant generating functions of random variables. We then present the steepest descent method and illustrate how to deform the contour of integration in the complex plane to pass through a saddlepoint. These mathematical techniques are used to derive the asymptotic expansion of complex integrals.

1.1 Characteristic Functions and Cumulant Generating Functions

Recall the definition of the Fourier transform $\widehat{f}(u)$ of a function $f(x)$, $x \in \mathbb{R}$, where

$$\widehat{f}(u) = \int_{-\infty}^{\infty} e^{iux} f(x)\,\mathrm{d}x, \quad u \in \mathbb{R}.$$

It is natural to ask under what conditions that $\widehat{f}(u)$ exists. Suppose f is Riemann integrable on every interval $[a, b]$ and $\int_{-\infty}^{\infty} |f(x)|\,\mathrm{d}x$ converges, then $\widehat{f}(u)$ is well defined for all $u \in \mathbb{R}$. The proof is quite straightforward. Note that $\left|e^{iux} f(x)\right| = |f(x)|$ and since any absolutely convergent (Riemann) infinite integral is convergent, so $\displaystyle\int_{-\infty}^{\infty} e^{iux} f(x)\,\mathrm{d}x$ converges.

Y. K. Kwok and W. Zheng, *Saddlepoint Approximation Methods in Financial Engineering*, SpringerBriefs in Quantitative Finance, https://doi.org/10.1007/978-3-319-74101-7_1

In the context of pricing derivatives, we may be required to find the Fourier transform of the terminal payoff of a call option. It is easily seen that the terminal payoff function may fail to satisfy the integrability condition for the existence of its Fourier transform. It is necessary to extend the Fourier transform variable from the real variable u to the complex variable z. The corresponding extended version is known as the *generalized Fourier transform*.

1.1.1 Generalized Fourier Transform and Characteristic Functions

Suppose that $f(x)$ is *Fourier integrable* in a strip (a, b), where $a, b \in \mathbb{R}$, then we observe

$$\int_{-\infty}^{\infty} e^{-ax}\,|f(x)|\,\mathrm{d}x < \infty \quad \text{and} \quad \int_{-\infty}^{\infty} e^{-bx}\,|f(x)|\,\mathrm{d}x < \infty.$$

The *generalized Fourier transform* of $f(x)$, $x \in \mathbb{R}$, is defined by

$$\hat{f}(z) = \mathscr{F}\{f(x)\}(z) = \int_{-\infty}^{\infty} e^{izx} f(x)\,\mathrm{d}x, \tag{1.1}$$

where $z \in \mathbb{C}$. In financial engineering applications, $f(x)$ may represent an option payoff function. In this case, the generalized Fourier transform $\hat{f}(z)$ exists when $\mathrm{Im}\, z$ is restricted to a horizontal strip, where $a < \mathrm{Im}\, z < b$ in the complex z-plane. Correspondingly, $\hat{f}(z)$ can be inverted by integrating within this horizontal strip along a straight line parallel to the real axis. The inverse generalized Fourier transform of $\hat{f}(z)$, $z \in \mathbb{C}$ with $a < \mathrm{Im}\, z < b$, is given by

$$f(x) = \mathscr{F}^{-1}\{\hat{f}(z)\}(x) = \frac{1}{2\pi}\int_{i\omega-\infty}^{i\omega+\infty} e^{-ixz}\hat{f}(z)\,\mathrm{d}z, \quad a < \omega < b, \tag{1.2}$$

where $x \in \mathbb{R}$.

The *complex characteristic function* of a random variable X is defined to be the generalized Fourier transform of the probability density function of X, where

$$\phi_X(z) = E\left[e^{izX}\right] = \int_{-\infty}^{\infty} e^{izx} p_X(x)\,\mathrm{d}x, \quad a < \mathrm{Im}\, z < b. \tag{1.3}$$

Here, $p_X(x)$ is the density function of X that is Fourier integrable in (a, b). By the Fourier inversion formula (1.2), we can recover the probability density function $p_X(x)$ from the complex characteristic function $\phi_X(z)$ via

$$p_X(x) = \frac{1}{2\pi}\int_{i\omega-\infty}^{i\omega+\infty} e^{-izx}\phi_X(z)\,\mathrm{d}z, \quad a < \omega < b. \tag{1.4}$$

Parseval Identity

Let $f : \mathbb{R} \mapsto \mathbb{R}$ and $g : \mathbb{R} \mapsto \mathbb{R}$ be integrable functions and one of them is bounded. Assume that $\int_{-\infty}^{\infty} \overline{f(x-y)}g(x)\,\mathrm{d}x$ is continuous at $y = 0$. The Parseval identity states that

$$\int_{-\infty}^{\infty} \overline{f(x)}g(x)\,\mathrm{d}x = \frac{1}{2\pi}\int_{i\omega-\infty}^{i\omega+\infty} \overline{\hat{f}(z)}\hat{g}(z)\,\mathrm{d}z, \quad i\omega \in \overline{\mathscr{S}_f} \cap \mathscr{S}_g, \tag{1.5}$$

where $\overline{\mathscr{S}_f}$ and $\mathscr{S}_g$ are the respective horizontal strip of regularity of $\overline{\hat{f}(z)}$ and $\hat{g}(z)$ in the complex plane. Since $f(x)$ is a real-valued function, we have $\overline{f(x)} = f(x)$. Moreover, $\overline{\hat{f}(z)} = \int_{-\infty}^{\infty} e^{-izx} f(x)\,\mathrm{d}x = \hat{f}(-z)$, the Parseval identity (1.5) can be expressed as

$$\int_{-\infty}^{\infty} f(x)g(x)\,\mathrm{d}x = \frac{1}{2\pi}\int_{i\omega-\infty}^{i\omega+\infty} \hat{f}(-z)\hat{g}(z)\,\mathrm{d}z, \quad i\omega \in \mathscr{S}_f \cap \mathscr{S}_g. \tag{1.6}$$

As an application in pricing financial derivatives, by regarding $f(x)$ as the transition density of the underlying price process and $g(x)$ as the terminal payoff function of a derivative, the undiscounted expectation of the terminal payoff of the derivative admits the Fourier representation as shown in (1.6) (Kwok et al. 2012). The existence of the integral in (1.6) requires the intersection of the strips of regularity of the generalized Fourier transform of the terminal payoff function and the complex characteristic function to be nonempty.

1.1.2 Laplace Transform and Cumulant Generating Functions

Let the function $f(x)$, $x \in \mathbb{R}$, be piecewise continuous on every finite interval in $[0, \infty)$ and satisfies $|f(x)| \le Me^{ax}$ for $x \in [0, \infty)$, where $M, a \in \mathbb{R}_+$ are constants independent of x. The *generalized Laplace transform* of $f(x)$ is defined by

$$\tilde{f}(z) = \mathscr{L}\{f(x)\}(z) = \int_0^{\infty} e^{-zx} f(x)\,\mathrm{d}x, \tag{1.7}$$

where $z \in \mathbb{C}$ and $\mathrm{Re}z > a$. The *inverse Laplace transform* of $\tilde{f}(z)$, $z \in \mathbb{C}$, is given by Kwok (2010)

$$f(x) = \mathscr{L}^{-1}\{\tilde{f}(z)\}(x) = \frac{1}{2\pi i}\int_{\omega-i\infty}^{\omega+i\infty} e^{xz}\tilde{f}(z)\,\mathrm{d}z, \tag{1.8}$$

which is a Bromwich integral along a vertical line parallel to the imaginary axis. Here, $\omega > \max\left\{\mathrm{Re} z_s : z_s \text{is a singular point of} \tilde{f}(z)\right\}$ and $x > 0$.

Sometimes, it may be instructive to consider the *bilateral Laplace transform* as defined by

$$\tilde{\tilde{f}}(z) = \mathscr{B}\{f(x)\}(z) = \int_{-\infty}^{\infty} e^{-zx} f(x)\,\mathrm{d}x, \tag{1.9}$$

where the imaginary axis is contained in the vertical strip of regularity of $\tilde{\tilde{f}}(z)$. It can be shown that the inversion formula of the bilateral Laplace transform takes the same form as given by (1.8), with the change of domain of x from $x > 0$ to $x \in \mathbb{R}$.

For any random variable X, its *moment generating function* $M_X(z) = E\left[e^{zX}\right]$, and *cumulant generating function* $\kappa_X(z) = \log M_X(z)$, are closely related to the bilateral Laplace transform of its probability density $p_X(x)$, where

$$M_X(-z) = e^{\kappa_X(-z)} = \int_{-\infty}^{\infty} e^{-zx} p_X(x)\,\mathrm{d}x = \mathscr{B}\{p_X(x)\}(z). \tag{1.10}$$

Suppose M_X exists, then all positive moments of X exist; that is, $E[|X|^n] < \infty$ for any positive integer n. Also, the moments of X can be obtained by

$$E[X^n] = \left.\frac{\mathrm{d}^n}{\mathrm{d}z^n} M_X(z)\right|_{z=0}. \tag{1.11}$$

In a similar manner, the nth order derivative of $\kappa_X(z)$ evaluated at $z = 0$ is called the nth order cumulant of X. The first few orders of cumulant are given by

$$\kappa_1 = \mu,\ \kappa_2 = \mu_2,\ \kappa_3 = \mu_3 \text{ and } \kappa_4 = \mu_4 - 3\mu_2^2, \tag{1.12}$$

where $\mu = E[X]$ and $\mu_n = E[(X - \mu)^n]$, $n = 2, 3, \ldots$.

Conversely, one may obtain the probability density and other related quantities by inverting the cumulant generating function. The relevant formulas are summarized in the following proposition.

Proposition 1.1 *Let X be any random variable with probability density $p_X(x)$ and cumulant generating function $\kappa_X(z)$. Suppose that $\kappa_X(z)$ is analytic in some open vertical strip*[1] $\{z \in \mathbb{C} : \alpha_- < \mathrm{Re} z < \alpha_+\}$*, where $\alpha_- < 0$ and $\alpha_+ > 0$. For any $K \in \mathbb{R}$, we have*

$$p_X(K) = \frac{1}{2\pi i}\int_{\gamma - i\infty}^{\gamma + i\infty} e^{\kappa_X(z) - zK}\,\mathrm{d}z, \qquad \gamma \in (\alpha_-, \alpha_+); \tag{1.13a}$$

$$P[X > K] = \frac{1}{2\pi i}\int_{\gamma - i\infty}^{\gamma + i\infty} \frac{e^{\kappa_X(z) - zK}}{z}\,\mathrm{d}z, \qquad \gamma \in (0, \alpha_+); \tag{1.13b}$$

[1]This vertical strip always contains the imaginary axis since $\kappa_X(0) = 0$.

$$E\left[(X-K)^+\right] = \frac{1}{2\pi i}\int_{\gamma-i\infty}^{\gamma+i\infty} \frac{e^{\kappa_X(z)-zK}}{z^2}\,\mathrm{d}z, \qquad \gamma \in (0,\alpha_+). \tag{1.13c}$$

Proof By observing that $e^{\kappa_X(z)}$ is the moment generating function and swapping the sign of z in $M_X(z)$ (see 1.9 and 1.10), then applying inversion formula (1.8), we obtain (1.13a). Next, by imposing $\mathrm{Re}\, z = \gamma \in (0,\alpha_+)$ so that $\int_K^\infty e^{-zx}\,\mathrm{d}x$ exists, we have

$$\begin{aligned}
P[X>K] &= \int_K^\infty \frac{1}{2\pi i}\int_{\gamma-i\infty}^{\gamma+i\infty} e^{\kappa_X(z)-zx}\,\mathrm{d}z\,\mathrm{d}x \\
&= \frac{1}{2\pi i}\int_{\gamma-i\infty}^{\gamma+i\infty} e^{\kappa_X(z)}\left(\int_K^\infty e^{-zx}\,\mathrm{d}x\right)\mathrm{d}z = \frac{1}{2\pi i}\int_{\gamma-i\infty}^{\gamma+i\infty}\frac{e^{\kappa_X(z)-zK}}{z}.
\end{aligned}$$

In a similar manner, we choose $\gamma \in (0,\alpha_+)$ and consider

$$\begin{aligned}
E\left[(X-K)^+\right] &= \int_K^\infty (x-K)p_X(x)\,\mathrm{d}x \\
&= \int_K^\infty (x-K)\frac{1}{2\pi i}\int_{\gamma-i\infty}^{\gamma+i\infty} e^{\kappa_X(z)-zx}\,\mathrm{d}z\,\mathrm{d}x \\
&= \frac{1}{2\pi i}\int_{\gamma-i\infty}^{\gamma+i\infty} e^{\kappa_X(z)}\left(\int_K^\infty (x-K)e^{-zx}\,\mathrm{d}x\right)\mathrm{d}z \\
&= \frac{1}{2\pi i}\int_{\gamma-i\infty}^{\gamma+i\infty}\frac{e^{\kappa_X(z)-zK}}{z^2}\,\mathrm{d}z.
\end{aligned}$$

□

Suppose we write $F_X(K) = P[X \le K]$ as the distribution function of X, it is seen that

$$\frac{1}{2\pi i}\int_{\gamma-i\infty}^{\gamma+i\infty}\frac{e^{\kappa_X(z)-zK}}{z}\,\mathrm{d}z = \begin{cases} -F_X(K) & \text{if } \gamma \in (\alpha_-,0) \\ 1-F_X(K) & \text{if } \gamma \in (0,\alpha_+) \end{cases}. \tag{1.13d}$$

The conditional expectation in the tail region $E[X\mathbf{1}_{\{X\ge K\}}]$ is related to the expected shortfall in risk management. The Bromwich integral representation of $E[X\mathbf{1}_{\{X\ge K\}}]$ can be shown to be (see Theorem 7.1 in Huang 2009 and 5.21 in Sect. 5.2.1)

$$E[X\mathbf{1}_{\{X\ge K\}}] = \frac{1}{2\pi i}\int_{\gamma-i\infty}^{\gamma+i\infty}\frac{\mathrm{d}\kappa_X(z)}{\mathrm{d}z}\frac{e^{\kappa_X(z)-zK}}{z}\,\mathrm{d}z, \quad \gamma \in (0,\alpha_+). \tag{1.13e}$$

1.2 Steepest Descent Method

In this section, we present the steepest descent method for deriving the asymptotic expansion of a Fourier or Laplace integral. In the procedure, one deforms a contour in the complex plane to pass through a saddlepoint along the path of steepest descent (Nayfeh 1981). The steepest descent method provides the basic tool for deriving various forms of the saddlepoint approximation formulas. First, we recall the Cauchy-Riemann equations in complex function theory and use them to derive two fundamental properties on stationary points and orthogonality of the level curves of the real part and imaginary part of a complex function (Kwok 2010).

Theorem 1.1 *Suppose that* $f(z) = u(x, y) + iv(x, y)$, *where* $z = x + iy$, $u(x, y)$ *and* $v(x, y)$ *are real-valued functions. Inside the domain of analyticity of* $f(z)$, *the Cauchy-Riemann equations hold:*

$$u_x = v_y, \quad u_y = -v_x, \tag{1.14a}$$

The first order derivative of $f(z)$ *is given by*

$$f'(z) = u_x + iv_x = u_x - iu_y = v_y + iv_x. \tag{1.14b}$$

1.2.1 Saddlepoint and Steepest Descent Path

The properties of a saddlepoint and steepest descent path can be deduced from the following corollaries of Theorem 1.1.

Corollary 1.1 *The real part* $u(x, y)$ *does not have a local maximum or minimum inside the domain of analyticity of* $f(z)$. *Any stationary point of* $u(x, y)$ *is a saddlepoint.*

Proof At a stationary point of $u(x, y)$, we have $u_x = u_y = 0$. Consider the determinant of the Hessian matrix, where

$$H = \begin{vmatrix} u_{xx} & u_{xy} \\ u_{yx} & u_{yy} \end{vmatrix} = u_{xx}u_{yy} - (u_{xy})^2 = -(v_{xy})^2 - (u_{xy})^2 < 0.$$

Note that we have used the Cauchy-Riemann equations (1.14a). The negativity of H implies that the stationary point of $u(x, y)$ is a saddlepoint and cannot be a local extremum. $\square$

Corollary 1.2 *The level curves of* $u(x, y)$ *is everywhere orthogonal to the level curves of* $v(x, y)$.

Proof Consider the gradients of u and v, they are orthogonal to the level curves of u and v, respectively. By the Cauchy-Riemann equations (1.14a), we have

$$\nabla u \cdot \nabla v = (u_x, u_y) \cdot (v_x, v_y) = u_x v_x + u_y v_y = 0.$$

This implies orthogonality of ∇u and ∇v, and equivalently, orthogonality of the level curves of u and v. $\square$

Since the level curves of $u = \text{Re } f$ and $v = \text{Im } f$ form families of orthogonal trajectories, the curves of steepest descent (and ascent) of Re f are the level curves of Im f, and vice versa. The saddlepoint z_0 of a complex function $f(z)$ is said to be simple if $f'(z_0) = 0$ but $f''(z_0) \neq 0$. Depending on the path that the saddlepoint is approached, by virtue of the property of a saddlepoint, Re f exhibits maximum value or minimum value at the saddlepoint.

To find the steepest descent path at the simple saddlepoint z_0, where $f''(z_0) \neq 0$, we write $f''(z_0) = |f''(z_0)|e^{i\psi}$ and $z - z_0 = re^{i\theta}$. The quadratic power term in the Taylor expansion of $f(z)$ around $z = z_0$ can be expressed as

$$\frac{f''(z_0)}{2}(z - z_0)^2 = \frac{|f''(z_0)|}{2} r^2 e^{i(2\theta+\psi)}.$$

The real part Re f attains its steepest descent at z_0 if we choose θ such that $e^{i(2\theta+\psi)} = -1$, or along the path where $\theta = -\frac{\psi}{2} \pm \frac{\pi}{2}$.

As an example, consider the complex function $f(z) = i\left(z + \frac{z^3}{3}\right)$, whose saddlepoints are the zeros of $f'(z) = (z^2 + 1)i$; that is, $z = \pm i$. Both saddlepoints are simple since $f''(\pm i) = \mp 2 \neq 0$. We consider the saddlepoint $z_0 = i$ and observe that $f(i) = -\frac{2}{3}$ so that Im $f(z_0) = 0$. Since Im $f(z_0) = 0$ at the simple saddlepoint $z_0 = i$, the steepest descent and ascent paths passing through $z_0 = i$ are the level curves: $v(x, y) = x\left(\frac{x^2}{3} - y^2 + 1\right) = v(0, 1) = 0$. The upper branch of the hyperbola: $\frac{x^2}{3} - y^2 + 1 = 0$ and the vertical line: $x = 0$ both pass through the saddlepoint $z_0 = i$. It is easy to check that the path along the vertical line: $x = 0$ is the steepest descent path while the orthogonal path along the hyperbola: $\frac{x^2}{3} - y^2 + 1 = 0$ is the steepest ascent path (see Fig. 1.1).

As a check, we observe $\psi = \arg f''(i) = 0$ so that the steepest descent path is along $\theta = \pm\frac{\pi}{2}$, or along the line: $x = 0$.

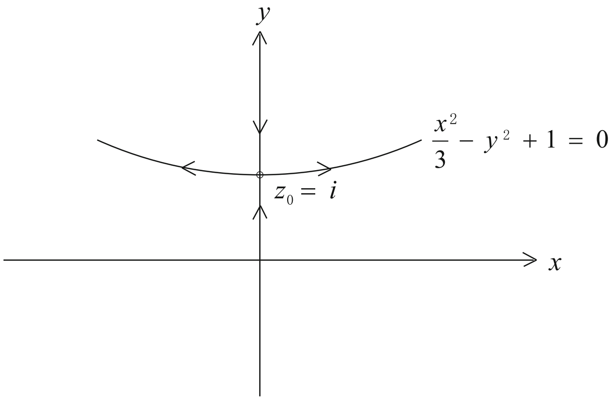


Fig. 1.1 Plot of the steepest descent and ascent paths approaching the simple saddlepoint $z_0 = i$ of $f(z) = i(z + \frac{z^3}{3})$. The arrows indicate the direction in which Re f decreases in value

1.2.2 Asymptotic Expansion of Complex Integrals

We would like to find an asymptotic expansion of a complex integral of the form

$$I(\lambda) = \int_C h(z)e^{\lambda f(z)}\,\mathrm{d}z, \tag{1.15}$$

where C is a contour in the complex plane, $h(z)$ and $f(z)$ are analytic functions in a domain $\mathscr{D}$ containing C. Recall that the zeros of $f'(z)$ are the saddlepoints of $f(z)$. Here, λ is taken to be real positive and large in value. Suppose λ is complex, where $\lambda = |\lambda|e^{i\alpha}$, we can always absorb $e^{i\alpha}$ into $h(z)$. We are interested to find an asymptotic expansion of $I(\lambda)$ as $\lambda \to \infty$ using the steepest descent method.

By virtue of the Cauchy Theorem in complex analysis, we deform the contour C to another contour $\widetilde{C}$ with the same end-points and lying in $\mathscr{D}$ such that $\max_{z\in\widetilde{C}} \text{Re } f(z)$ is attained only at the saddlepoints or at the end-points of $\widetilde{C}$. The asymptotic value of $I(\lambda)$ as $\lambda \to \infty$ is obtained by summing the contributions from the saddlepoints. We expect that the value of $I(\lambda)$ is dominated by the saddlepoint point with the largest value in Re f. Let that be the simple saddlepoint z_0.

For z on $\widetilde{C}$ that is near z_0, we approximate $f(z)$ and $h(z)$ by

$$f(z) \approx f(z_0) + \frac{f''(z_0)}{2}(z - z_0)^2 \text{ and } h(z) \approx h(z_0).$$

We set

$$z - z_0 = re^{i\theta} \text{ and } f''(z_0) = |f''(z_0)|e^{i\psi},$$

and choose θ along the steepest descent path where $e^{i(2\theta+\psi)} = -1$ or $\theta = -\dfrac{\psi}{2} \pm \dfrac{\pi}{2}$ (the sign is chosen based on the orientation of $\widetilde{C}$). We extend the limits of integration with respect to r to be infinity (reduced to a Gaussian integral for ease of evaluation). As $\lambda \to \infty$ and on the steepest descent path, the dominant contribution to the complex integral can be obtained by a local computation in the neighborhood of z_0 since $|e^{\lambda f(z)}|$ is negligible elsewhere, so

$$I(\lambda) \approx h(z_0)e^{\lambda f(z_0)} \int_{-\infty}^{\infty} e^{\frac{\lambda}{2}|f''(z_0)|r^2 e^{i(2\theta+\psi)}} e^{i\theta}\, \mathrm{d}r.$$

By evaluating the Gaussian integral, we obtain the asymptotic expansion

$$I(\lambda) \approx h(z_0)e^{\lambda f(z_0)}e^{i\theta} \int_{-\infty}^{\infty} e^{-\frac{\lambda}{2}|f''(z_0)|r^2}\, \mathrm{d}r = h(z_0)e^{\lambda f(z_0)}e^{i\theta}\sqrt{\frac{2\pi}{\lambda|f''(z_0)|}}. \quad (1.16)$$

The asymptotic formula is based on the contribution to $I(\lambda)$ from the single simple saddlepoint z_0 that has the largest value of $\mathrm{Re}\, f(z)$. The formula involves $e^{i\theta}$, the value of which can be found once the steepest descent path is identified.

Chapter 2
Saddlepoint Approximations to Density Functions, Tail Probabilities and Tail Expectations

In many applications in financial engineering, it is necessary to consider approximating the tail probabilities and tail expectations of random variables whose closed forms may not be tractable. Suppose the first few moments of the underlying distribution of a random variable are known, one may apply the Edgeworth type expansion up to the order of the known moments. However, it is known that the Edgeworth expansion may fail to give an accurate approximation when the tail region of the distribution is considered since the Edgeworth approximation is taken around the mean of the distribution. Intuitively, improved accuracy of an asymptotic approximation can be achieved if one performs approximation around the region of interest in the underlying distribution. This idea forms the basic principle of exponentially tilted Edgeworth expansion in deriving the saddlepoint approximation and explains why the saddlepoint approximation is highly effective and accurate for approximating tail probabilities and tail expectations of random variables, where various applications can be found in option pricing and credit portfolio calculations. Instead of limiting to the normal base approximation, the saddlepoint approximation approach can be extended to an arbitrary base distribution that fits better the underlying distribution.

In this book, we place emphasis on the uses of the saddlepoint methods in financial engineering and report some of the latest developments of the methods beyond the treatise in the two comprehensive saddlepoint approximation texts by Jensen (1995) and Butler (2007). A recent comprehensive review of saddlepoint approximations and their applications in financial risk management can be found in Broda and Paolella (2012). Saddlepoint approximations also find wide applications in maximum likelihood estimation. Readers interested in the applications of saddlepoint applications in statistical inference may read the classical paper by Barndorff-Nielsen and Cox (1979) and the comprehensive review paper by Reid (1988).

In this chapter, we discuss various saddlepoint approximation formulas for the density functions, tail probabilities and tail expectations of continuous random variables under both the Gaussian and non-Gaussian base distribution. The proofs of

Y. K. Kwok and W. Zheng, *Saddlepoint Approximation Methods in Financial Engineering*, SpringerBriefs in Quantitative Finance, https://doi.org/10.1007/978-3-319-74101-7_2

these formulas use the steepest descent method and the technique of exponentially tilted Edgeworth expansion. The saddlepoint approximation for finding the density function of the mean of a large number of independent and identically distributed (i.i.d.) random variables was first derived by Daniels (1954) using the steepest descent method. We consider the relation between the derivation approach using the statistical approach of exponentially titled Edgeworth expansion and the steepest descent method in approximating density functions. We present the detailed review of the latest innovative approaches of obtaining the saddlepoint approximations of the tail probabilities and tail expectations that are beyond the renowned Lugannani-Rice approach (1980). Besides the theoretical derivation of the approximation formulas, we also report the numerical tests performed for assessment of accuracy of these formulas. Some discussion on error analysis of saddlepoint approximations and cases of potential failure of the saddlepoint approximations are included.

2.1 Density Functions

The saddlepoint approximation was first applied in statistics for approximating the density function of a random variable with analytic moment generating function but nontractable distribution function. Daniels (1954) uses the steepest descent method to derive the saddlepoint approximation for the density function of the mean of a large number of i.i.d. random variables. The saddlepoint approximation formula produces a high level of accuracy compared with the Edgeworth expansion, especially in approximating the tail behavior of the density function.

Proposition 2.1 (Daniels 1954) *Let $M(z) = E[e^{zX}]$ and $\kappa(z) = \log M(z)$ be the respective moment generating function and cumulant generating function of a random variable X that are analytic in some open vertical strip $\{z \in \mathbb{C} : \alpha_- < \mathrm{Re}\, z < \alpha_+\}$, where $\alpha_- < 0$ and $\alpha_+ > 0$. Let $X_1, \ldots, X_n$ be independent and identically distributed random variables that share the same moment generating function $M(z)$.*

Let $p_{\bar{X}}(x)$ denote the density function of the sample mean $\bar{X}$, where $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.

The second order saddlepoint approximation to $p_{\bar{X}}(x)$ is given by

$$p_{\bar{X}}(x) \approx p_{\bar{X}}^{(2)}(x) = \frac{\sqrt{n}e^{n[\kappa(\hat{z})-\hat{z}x]}}{\sqrt{2\pi\kappa''(\hat{z})}}\left\{1 + \frac{1}{8n}\left[\lambda_4(\hat{z}) - \frac{5}{3}\lambda_3^2(\hat{z})\right]\right\}, \tag{2.1}$$

where $\lambda_3(\hat{z}) = \frac{\kappa'''(\hat{z})}{\kappa''(\hat{z})^{3/2}}$ and $\lambda_4(\hat{z}) = \frac{\kappa''''(\hat{z})}{\kappa''(\hat{z})^2}$; and $\hat{z}$ is the unique simple real root that solves the saddlepoint equation:

$$\kappa'(z) = x. \tag{2.2}$$

Proof The moment generating function (mgf) of the sample mean $\bar{X}$ of n i.i.d. random variables is given by

$$M_{\bar{X}}(z) = E[e^{z\bar{X}}] = E[e^{\frac{z}{n}\sum_{i=1}^{n} X_i}] = \prod_{i=1}^{n} E[e^{\frac{z}{n}X_i}] = \left[M\left(\frac{z}{n}\right)\right]^n = e^{n\kappa\left(\frac{z}{n}\right)}. \tag{2.3a}$$

Accordingly, $\mathrm{var}(\bar{X})$ and the cumulant generating function (cgf) of X are related by

$$\mathrm{var}(\bar{X}) = \left.\frac{\mathrm{d}^2}{\mathrm{d}z^2}\log M_{\bar{X}}(z)\right|_{z=0} = \left.\frac{\mathrm{d}^2}{\mathrm{d}z^2 n}\kappa\left(\frac{z}{n}\right)\right|_{z=0} = \frac{\kappa''(0)}{n}. \tag{2.3b}$$

The density function of $\bar{X}$ is given by the Laplace inversion of $M_{\bar{X}}(z)$ as follows (see 1.13a):

$$\begin{aligned} p_{\bar{X}}(x) &= \frac{1}{2\pi i}\int_{\gamma-i\infty}^{\gamma+i\infty} M_{\bar{X}}(z)e^{-zx}\,\mathrm{d}z \\ &= \frac{1}{2\pi i}\int_{\gamma-i\infty}^{\gamma+i\infty} e^{n\kappa\left(\frac{z}{n}\right)}e^{-zx}\,\mathrm{d}z = \frac{n}{2\pi i}\int_{\frac{\gamma}{n}-i\infty}^{\frac{\gamma}{n}+i\infty} e^{n[\kappa(z)-zx]}\,\mathrm{d}z, \end{aligned}$$

where $\alpha_- < \gamma < \alpha_+$. When n is large, we may apply the steepest descent method to derive an asymptotic expansion of the above Bromwich integral. The procedures include the determination of the saddlepoint $\hat{z}$ that solves the saddlepoint Eq. (2.2) and the deformation of the contour to pass through the saddlepoint $\hat{z}$. Under certain mild conditions on the distribution function of X, Daniels (1954, Theorem 6.2) shows that there exists unique simple real root of the saddlepoint equation that lies within (α_-, α_+). We replace the original contour by the steepest descent path that passes through the real simple saddlepoint $\hat{z}$, which becomes the vertical line along $\hat{z} - i\infty$ to $\hat{z} + i\infty$ parallel to the imaginary axis. According to the principle of the steepest descent method, the principal contribution to the above Bromwich integral derives from the analytic behavior of the integrand in the immediate neighborhood of the saddlepoint $\hat{z}$.

To derive the second order saddlepoint approximation formula of the density function of the sample mean, we consider the Taylor expansion of the exponent of the integrand $\kappa(z) - zx$ about $\hat{z}$. We use the change of variable:

$$w = -i\sqrt{n\kappa''(\hat{z})}(z - \hat{z}).$$

The corresponding Taylor expansion of $\kappa(z) - zx$ at $\hat{z}$ up to the fourth order is given by

$$\kappa(z) - zx = \kappa(\hat{z}) - \hat{z}x + \frac{\kappa''(\hat{z})}{2!}\left[\frac{iw}{\sqrt{n\kappa''(\hat{z})}}\right]^2 + \frac{\kappa'''(\hat{z})}{3!}\left[\frac{iw}{\sqrt{n\kappa''(\hat{z})}}\right]^3 + \frac{\kappa''''(\hat{z})}{4!}\left[\frac{iw}{\sqrt{n\kappa''(\hat{z})}}\right]^4 + \cdots.$$

Substituting into the Bromwich integral of $p_{\bar{X}}(x)$, we obtain

$$p_{\bar{X}}(x) \approx \frac{\sqrt{n}}{2\pi\sqrt{\kappa''(\hat{z})}}\int_{-\infty}^{\infty} e^{n[\kappa(\hat{z})-\hat{z}x]-w^2/2}\left\{1 - \frac{i}{\sqrt{n\kappa''(\hat{z})^3}}\kappa'''(\hat{z})\frac{w^3}{3!} + \frac{1}{n}\left[\frac{\kappa''''(\hat{z})}{\kappa''(\hat{z})^2}\frac{w^4}{4!} - \frac{\kappa'''(\hat{z})^2}{\kappa''(\hat{z})^3}\frac{w^6}{2!3!^2} + \cdots\right]\right\}\mathrm{d}w.$$

In the above expansion, when the exponentiation of the third order derivative term in $\kappa(z) - zx$ is considered, the last term involving w^6 is added in the above integral when we seek for expansion up to $O(\frac{1}{n})$. More specifically, we use the power series: $e^x \approx 1 + x + \frac{x^2}{2}$, where x stands for $\left(\frac{iw}{\sqrt{n\kappa''(\hat{z})}}\right)^3$. By observing the following identities:

$$\int_{-\infty}^{\infty} e^{-\frac{w^2}{2}}dw = \sqrt{2\pi}, \quad \int_{-\infty}^{\infty} w^3 e^{-\frac{w^2}{2}}\,dw = 0,$$
$$\int_{-\infty}^{\infty} w^4 e^{-\frac{w^2}{2}}dw = 3\sqrt{2\pi}, \quad \int_{-\infty}^{\infty} w^6 e^{-\frac{w^2}{2}}dw = 15\sqrt{2\pi},$$

we obtain the second order saddlepoint approximation $p_{\bar{X}}^{(2)}(x)$ to the density function $p_{\bar{X}}(x)$ as follows:

$$\begin{aligned} p_{\bar{X}}(x) &\approx p_{\bar{X}}^{(2)}(x) \\ &= \frac{\sqrt{n}e^{n[\kappa(\hat{z})-\hat{z}x]}}{2\pi\sqrt{\kappa''(\hat{z})}}\left[\sqrt{2\pi} + \frac{3\sqrt{2\pi}}{24n}\frac{\kappa''''(\hat{z})}{\kappa''(\hat{z})^2} - \frac{15\sqrt{2\pi}}{72n}\frac{\kappa'''(\hat{z})^2}{\kappa''(\hat{z})^3}\right] \\ &= \frac{\sqrt{n}e^{n[\kappa(\hat{z})-\hat{z}x]}}{\sqrt{2\pi\kappa''(\hat{z})}}\left\{1 + \frac{1}{8n}\left[\lambda_4(\hat{z}) - \frac{5}{3}\lambda_3^2(\hat{z})\right]\right\}. \end{aligned}$$

□

We may deduce uniqueness of the solution to the saddlepoint Eq. (2.2) by establishing that $\kappa'(z)$ is a monotonically nondecreasing function. This is seen by observing

$$\kappa'(z) = \frac{d}{dz}\log(E[e^{zX}]) = \frac{E[Xe^{zX}]}{E[e^{zX}]},$$
$$\kappa''(z) = \frac{E[X^2e^{zX}]E[e^{zX}] - (E[Xe^{zX}])^2}{(E[e^{zX}])^2} \geq 0.$$

Nonnegativity of $\kappa''(z)$ is obtained by the Cauchy-Schwarz inequality: $E[W^2]$ $E[Y^2] \geq (E[WY])^2$, where $W = X\sqrt{e^{zX}}$ and $Y = \sqrt{e^{zX}}$.

Daniels (1980) shows that the saddlepoint approximation becomes exact for three distributions, namely, the normal, gamma and inverse normal distribution. For example, we consider the gamma distribution with density function

$$p(x) = \frac{1}{\Gamma(\alpha)}x^{\alpha-1}e^{-x}, \quad x > 0,$$

where $\Gamma(\alpha) = (\alpha - 1)!$. Note that $\kappa(z) = -\alpha\log(1-z)$ for the gamma distribution, so the corresponding saddlepoint equation is

$$\kappa'(z) = \frac{\alpha}{1-z} = x.$$

This gives the unique simple saddlepoint: $\hat{z} = 1 - \frac{\alpha}{x}$, $x > 0$. Note that $\kappa''(\hat{z}) = \frac{x^2}{\alpha}$. By taking $n = 1$ and using the first order term in (2.1), the first order saddlepoint approximation of the gamma distribution is given by

$$p_{\text{app}}(x) = \frac{1}{\sqrt{2\pi x^2/\alpha}}\exp(-\alpha\log(1-\hat{z}) - \hat{z}x) = \frac{x^{\alpha-1}e^{-x}}{\sqrt{2\pi}\alpha^{\alpha-\frac{1}{2}}e^{-\alpha}}, \ x > 0.$$

Note that $p_{\text{app}}(x)$ and $p(x)$ differ by a constant multiple. If we normalize the approximating density function $p_{\text{app}}(x)$ to have an integrated sum of one, then the saddlepoint approximation becomes exact. One may note that the approximation of $\Gamma(\alpha)$ by $\sqrt{2\pi}\alpha^{\alpha-\frac{1}{2}}e^{-\alpha}$ is indeed the well known Stirling's approximation for $\Gamma(\alpha)$. Additional numerical examples of illustrating accuracy of the saddlepoint approximation formula for density functions of the student's t distribution and noncentral chi-squared distribution can be found in Goutis and Casella (1999).

Higher order saddlepoint approximation formulas for the densities can be found in a similar manner by performing higher order Taylor expansion of $\kappa(z) - zx$. The next order saddlepoint approximation formula for the density function is known to be (García-Céspedes 2014)

$$p_{\bar{X}}(x) = \frac{\sqrt{n}e^{n[\kappa(\hat{z})-\hat{z}x]}}{\sqrt{2\pi\kappa''(\hat{z})}}\left\{1 + \frac{1}{8n}\left[\lambda_4(\hat{z}) - \frac{5}{3}\lambda_3^2(\hat{z})\right]\right.$$
$$\left. + \frac{5}{48n^2}\left[-\frac{\lambda_6(\hat{z})}{5} + \frac{7\lambda_4^2(\hat{z})}{8} - \frac{21\lambda_3^2(\hat{z})\lambda_4(\hat{z})}{4}\right]\right\} + O\left(\frac{1}{n^3}\right),$$

where $\lambda_6(\hat{z}) = \dfrac{\kappa^{(6)}(\hat{z})}{\kappa''(\hat{z})^3}$. Here, $\kappa^{(6)}(z)$ denotes the sixth order derivative of $\kappa(z)$. As a remark, though the theoretical derivation of the saddlepoint approximation formula (2.1) is based on the asymptotic limit as $n \to \infty$, numerical tests show that the formula provides surprisingly good approximation even for $n = 1$.

The mathematical formalism for an asymptotic series requires the error in the truncated series to be asymptotically equal to the first omitted term in the asymptotic series. For a given value of x, there are certain number of terms used in the asymptotic expansion that improve accuracy. Beyond which, point accuracy may deteriorate since an asymptotic series may not be a convergent series.

2.1.1 Exponentially Tilted Edgeworth Expansion

The saddlepoint approximation method can be interpreted in statistical sense as the Edgeworth expansion of the exponentially tilted distribution of the underlying random variable. We illustrate this connection via an alternative proof of Proposition 2.1 using the exponentially tilted Edgeworth expansion (Barndorff-Nielsen and Cox 1979). A brief review of the exponential tilting technique and Edgeworth expansion of the distribution of a random variable is presented in the Appendix.

Recall that $\bar{X}$ is the mean of n i.i.d. random variables with density function $p_{\bar{X}}(x)$. All these i.i.d. random variables share the same cgf $\kappa(z)$. Consider the exponential tilting θ-family derived from $p_{\bar{X}}(x)$ as defined by

$$p_{\bar{X}}(x;\theta) = \exp(n[\theta x - \kappa(\theta)])p_{\bar{X}}(x). \tag{2.4}$$

Recall that the mean and variance of $\bar{X}$ under the P_θ measure are given by $E_\theta[\bar{X}] = \kappa'(\theta)$ and $\mathrm{var}_\theta(\bar{X}) = \kappa''(\theta)/n$ (see 2.3b and Appendix). By fitting an Edgeworth expansion to $p_{\bar{X}}(x;\theta)$ (see Appendix), we have

$$\begin{aligned} p_{\bar{X}}(x;\theta) = \frac{\sqrt{n}}{\sqrt{\kappa''(\theta)}}\,\phi(y)\bigg[1 + \frac{\lambda_3(\theta)H_3(y)}{6\sqrt{n}} + \frac{\lambda_4(\theta)H_4(y)}{24n} \\ + \frac{\lambda_3^2(\theta)H_6(y)}{72n}\bigg] + O\left(\frac{1}{n}\right), \end{aligned} \tag{2.5}$$

where

$$y = \frac{x - E_\theta[\bar{X}]}{\sqrt{\mathrm{var}_\theta(\bar{X})}} = \sqrt{n}\,\frac{x - \kappa'(\theta)}{\sqrt{\kappa''(\theta)}}.$$

Here, $\phi(z)$ is the standard normal density function and the Hermite polynomials are

$$H_3(y) = y^3 - 3y,\ H_4(y) = y^4 - 6y^2 + 3,\ H_6(y) = y^6 - 15y^4 + 45y^2 - 15. \tag{2.6}$$

In general, the normal approximation to the distribution of a random variable is accurate near the mean of the random variable, but accuracy deteriorates in the tail of the distribution. Therefore, we are motivated to consider the Edgeworth expansion of the exponentially titled distribution by choosing θ such that x is close to the mean of the titled distribution. Suppose we choose $\theta = \hat{\theta}$ such that $\hat{\theta}$ solves the saddlepoint equation: $\kappa'(\hat{\theta}) = x$, thus giving $y = 0$. With this choice of $\hat{\theta}$, we are essentially considering the Edgeworth expansion of $\bar{X}$ around $E_\theta[\bar{X}]$. By observing $H_3(0) = 0$, $H_4(0) = 3$, $H_6(0) = -15$, $\phi(0) = \dfrac{1}{\sqrt{2\pi}}$, and substituting $\hat{\theta}$ by $\hat{z}$ as the saddlepoint, together with the rearrangement of the terms, we obtain

$$\begin{aligned} p_{\bar{X}}(x) &= e^{n[\kappa(\hat{z})-\hat{z}x]} p_{\bar{X}}(x;\hat{\theta}) \\ &= \frac{\sqrt{n}e^{n[\kappa(\hat{z})-\hat{z}x]}}{\sqrt{2\pi\kappa''(\hat{z})}} \left\{ 1 + \frac{1}{8n}\left[\lambda_4(\hat{z}) - \frac{5}{3}\lambda_3^2(\hat{z})\right] + O(n^{-2}) \right\}, \end{aligned} \tag{2.7}$$

which agrees with (2.1) up to $O(n^{-2})$.

2.1.2 *Extension to the Non-Gaussian Base*

With the choice of the quadratic function in approximating the analytic behavior of the exponent term: $\kappa(z) - zx$, and the same choice of the normal function in the Edgeworth approximation of the exponentially tilted distribution in the derivation of the saddlepoint formula (2.1), we essentially use the Gaussian distribution as the base distribution in the saddlepoint approximation procedure. Therefore, the saddlepoint approximation formula (2.1) works well for distributions whose higher moments are less significant. However, one would expect accuracy of the saddlepoint approximation may deteriorate when one deals with the density function of a distribution with large skewness and kurtosis, like jump processes (Aït-Sahalia and Yu 2006).

We would like to extend the saddlepoint approximation with a general base distribution. To motivate the derivation procedure, we consider the reformulation of the proof of formula (2.1). For ease of presentation, we take $n = 1$. Consider the Laplace inversion formulation of the density function $p(x)$ in terms of its cgf $\kappa(z)$:

$$p(x) = \frac{1}{2\pi i}\int_{\gamma-i\infty}^{\gamma+i\infty} e^{\kappa(z)-zx}\,\mathrm{d}z, \tag{2.8}$$

where $\gamma \in (\alpha_-, \alpha_+)$ (see 1.13a). Recall that the cgf of a Gaussian distribution is a quadratic function in its argument. Consistent with the notion of the Gaussian base approximation, we set

$$[\kappa(z) - zx] - [\kappa(\hat{z}) - \hat{z}x] = \frac{(w-\hat{w})^2}{2}, \tag{2.9a}$$

where $\hat{z}$ is the saddlepoint and $\hat{w}$ is set to be $\hat{z}x - \kappa(\hat{z}) = \frac{\hat{w}^2}{2}$. This may be visualized as a transformation of variable from z to w, where

$$\kappa(z) - zx = \frac{w^2}{2} - w\hat{w}. \tag{2.9b}$$

Note that the two functions match their values at $w = 0$ with $z = 0$ and $w = \hat{w}$ with $z = \hat{z}$. By forcing $\hat{w}$ and $\hat{z}$ to have the same sign, we obtain

$$\hat{w} = \operatorname{sgn}(\hat{z})\sqrt{2[\hat{z}x - \kappa(\hat{z})]}. \tag{2.9c}$$

We choose the contour in the Bromwich integral in (2.8) to be the vertical line through $\hat{z}$. Since w equals $\hat{w}$ when z equals $\hat{z}$, by applying the change of variable defined in (2.9b), the density function can be expressed as

$$p(x) = \frac{1}{2\pi i}\int_{\hat{w}-i\infty}^{\hat{w}+i\infty} e^{\frac{w^2}{2}-w\hat{w}}\frac{\mathrm{d}z}{\mathrm{d}w}\,\mathrm{d}w. \tag{2.10}$$

Applying the same principle used in the steepest descent method, we argue that the main contribution of $\frac{\mathrm{d}z}{\mathrm{d}w}$ to the above complex integral arises from the region near the transformed saddlepoint $\hat{w}$. One may adopt the approximation (though apparently crude):

$$\frac{\mathrm{d}z}{\mathrm{d}w} \approx \left.\frac{\mathrm{d}z}{\mathrm{d}w}\right|_{w=\hat{w}}. \tag{2.11a}$$

To evaluate $\left.\frac{\mathrm{d}z}{\mathrm{d}w}\right|_{w=\hat{w}}$, we differentiate (2.9b) with respect to w on both sides to obtain

$$\begin{aligned}
&\kappa'(z)\frac{\mathrm{d}z}{\mathrm{d}w} - \frac{\mathrm{d}z}{\mathrm{d}w}x = w - \hat{w}\\
&\kappa''(z)\left(\frac{\mathrm{d}z}{\mathrm{d}w}\right)^2 + \kappa'(z)\frac{\mathrm{d}^2z}{\mathrm{d}w^2} - \frac{\mathrm{d}^2z}{\mathrm{d}w^2}x = 1.
\end{aligned}$$

Note that at $w = \hat{w}$, we have $z = \hat{z}$ and $\kappa'(\hat{z}) = x$, so we obtain

$$\left.\frac{\mathrm{d}z}{\mathrm{d}w}\right|_{w=\hat{w}} = \frac{1}{\sqrt{\kappa''(\hat{z})}}. \tag{2.11b}$$

We now substitute the approximation in (2.11a) and the result in (2.11b) into the complex integral (2.10). We let $w = \hat{w} + iv$ so that

$$\frac{w^2}{2} - w\hat{w} = \frac{(\hat{w}+iv)^2}{2} - (\hat{w}+iv)\hat{w} = -\frac{v^2}{2} - \frac{\hat{w}^2}{2}.$$

The density function is approximated by

$$
\begin{aligned}
p(x) &\approx \frac{1}{2\pi i}\int_{\hat{w}-i\infty}^{\hat{w}+i\infty} e^{\frac{w^2}{2}-w\hat{w}}\frac{1}{\sqrt{\kappa''(\hat{z})}}\,\mathrm{d}w \\
&= \frac{1}{\sqrt{2\pi}}e^{-\frac{\hat{w}^2}{2}}\frac{1}{\sqrt{\kappa''(\hat{z})}}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{v^2}{2}}\,\mathrm{d}v \\
&= \frac{\phi(\hat{w})}{\sqrt{\kappa''(\hat{z})}} = \frac{e^{\kappa(\hat{z})-\hat{z}x}}{\sqrt{2\pi\kappa''(\hat{z})}},
\end{aligned}
\tag{2.12}
$$

where $\hat{z}$ solves the saddlepoint Eq. (2.2) and $\hat{w}$ is given by (2.9c). It is seen that the above approximation agrees with (2.1) up to the first order approximation with $n = 1$ (see Proposition 2.1).

The extension of the saddlepoint approximation of the density function to the non-Gaussian base follows in a similar manner. For an arbitrary base distribution with cumulant generating function $\kappa_0(w)$, Aït-Sahalia and Yu (2006) use the function $\kappa_0(w) - wx$ in w to approximate $\kappa(z) - zx$ by matching their values at $w = \hat{w}$ with $z = \hat{z}$, where $\hat{w}$ is the saddlepoint for κ_0 that solves $\kappa_0'(w) = x$. In a similar manner, the corresponding change of variable formula from z to w is given by

$$
\{\kappa(z) - zx\} - \{\kappa(\hat{z}) - \hat{z}x\} = \{\kappa_0(w) - wx\} - \{\kappa_0(\hat{w}) - \hat{w}x\}. \tag{2.13a}
$$

Similar to the above Gaussian base case, by performing twice differentiation with respect to w of (2.13a) and setting $z = \hat{z} \Leftrightarrow w = \hat{w}$, we obtain

$$
\left.\frac{\mathrm{d}z}{\mathrm{d}w}\right|_{w=\hat{w}} = \sqrt{\frac{\kappa_0''(\hat{w})}{\kappa''(\hat{z})}}. \tag{2.13b}
$$

By following the same approximation procedure as in (2.11a) and using the result in (2.13b), the approximation of the density function under a non-Gaussian base distribution κ_0 is given by

$$
\begin{aligned}
p(x) &= e^{[\kappa(\hat{z})-\hat{z}x]-[\kappa_0(\hat{w})-\hat{w}x]}\frac{1}{2\pi i}\int_{\hat{w}-i\infty}^{\hat{w}+i\infty} e^{\kappa_0(w)-wx}\frac{\mathrm{d}z}{\mathrm{d}w}\,\mathrm{d}w \\
&\approx e^{[\kappa(\hat{z})-\hat{z}x]-[\kappa_0(\hat{w})-\hat{w}x]}\left(\left.\frac{\mathrm{d}z}{\mathrm{d}w}\right|_{w=\hat{w}}\right)\frac{1}{2\pi i}\int_{\hat{w}-i\infty}^{\hat{w}+i\infty} e^{\kappa_0(w)-wx}\,\mathrm{d}w \\
&= e^{[\kappa(\hat{z})-\hat{z}x]-[\kappa_0(\hat{w})-\hat{w}x]}\sqrt{\frac{\kappa_0''(\hat{w})}{\kappa''(\hat{z})}}f_0(x),
\end{aligned}
\tag{2.14}
$$

where $\hat{w}$ is the saddlepoint for κ_0 that solves $\kappa_0'(w) = x$, and $f_0(x)$ is the density function of the base distribution κ_0. To check whether the Gaussian base result for approximating the density function given in (2.12) is recovered when we choose

$\kappa_0(w) = \dfrac{w^2}{2}$ and $f_0(x) = \phi(x)$, we observe that the saddlepoint equation for $\hat{w}$ gives $\hat{w} = x$. Furthermore, we observe $\kappa_0''(\hat{w}) = 1$ and

$$e^{-[\kappa_0(\hat{w})-\hat{w}x]} f_0(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(\hat{w}-x)^2}{2}} = \frac{1}{\sqrt{2\pi}},$$

so the same result in (2.12) is obtained.

2.2 Tail Probabilities

Theoretically, once the approximation to density function has been obtained, the approximations to tail probability and tail expectation can be derived by integration. For the saddlepoint approximation formula for the density function, though its value is everywhere positive, the approximating density function would not in general integrate to one. Therefore, one needs to perform normalization by dividing the saddlepoint approximation by its integration sum. As shown in the numerical examples reported in Goutis and Casella (1999), the normalization procedure improves numerical accuracy in the saddlepoint approximation. This procedure may lead to quite complicated saddlepoint approximation formulas.

In their pioneering work, Lugannani and Rice (1980) propose an effective saddlepoint approximation formula for computing the tail probability in the form of $P[X > x]$. Their approximation formula works surprisingly well over a wide range of x, while the earlier versions perform poorly for x near $E[\bar{X}]$ and in the tails (Strawderman 2000). Analogous to the saddlepoint approximation to density function, the Lugannani-Rice formula for tail probabilities remains accurate for $n = 1$ (Daniels 1987, Aït-Sahalia and Yu 2006). The Lugannani-Rice saddlepoint approximation for the tail probability formula for $n = 1$ is presented in the following proposition.

Proposition 2.2 (Lugannani and Rice 1980) *Let $\kappa(z)$ denote the cumulant generating function of X that satisfies the condition in Proposition 2.1. The saddlepoint approximation formula for $P[X > x]$ is given by*

$$P(X > x) \approx \begin{cases} 1 - \Phi(\hat{w}) + \phi(\hat{w}) \left[\dfrac{1}{\hat{z}\kappa''(\hat{z})^{1/2}} - \dfrac{1}{\hat{w}} \right], & x \neq E[X] = \kappa'(0) \\ \dfrac{1}{2} - \dfrac{1}{6\sqrt{2\pi}} \dfrac{\kappa'''(0)}{\kappa''(0)^{3/2}}, & x = E[X] = \kappa'(0) \end{cases} \tag{2.15}$$

where $\hat{z}$ solves the saddlepoint Eq. (2.2), and $\hat{w}$ is given by (2.9c). Note that when $x = E[X] = \kappa'(0)$, we have $\hat{z} = \hat{w} = 0$. The alternative formula under this scenario is derived by taking the asymptotic limits: $\hat{z} \to 0$ and $\hat{w} \to 0$.

Proof The proof starts from the inverse Laplace transform (1.13b). Provided that the saddlepoint $\hat{z}$ exists in $(0, \alpha_+)$, we observe

$$P[X > x] = \frac{1}{2\pi i} \int_{\hat{z}-i\infty}^{\hat{z}+i\infty} \frac{e^{\kappa(z)-zx}}{z} \, \mathrm{d}z.$$

Under the choice of the Gaussian base distribution, we adopt the change of variable from z to w as defined in (2.9b). This gives

$$P[X > x] = \frac{1}{2\pi i} \int_{\hat{w}-i\infty}^{\hat{w}+i\infty} e^{\frac{w^2}{2}-w\hat{w}} \frac{1}{z(w)} \frac{\mathrm{d}z(w)}{\mathrm{d}w} \, \mathrm{d}w.$$

Since $z = 0$ when $w = 0$, the term $\frac{1}{z(w)} \frac{\mathrm{d}z(w)}{\mathrm{d}w}$ has a singularity at $w = 0$. Near $z = \hat{z}$ and $w = \hat{w}$, it can be shown that z and w have linear relationship (see Aït-Sahalia and Yu 2006, p. 518). We perform the following decomposition to isolate the singularity at $w = 0$, where

$$P[X > x] = \frac{1}{2\pi i} \int_{\hat{w}-i\infty}^{\hat{w}+i\infty} \frac{e^{\frac{w^2}{2}-w\hat{w}}}{w} \, \mathrm{d}w + \frac{1}{2\pi i} \int_{\hat{w}-i\infty}^{\hat{w}+i\infty} e^{\frac{w^2}{2}-w\hat{w}} \left[\frac{1}{z(w)} \frac{\mathrm{d}z(w)}{\mathrm{d}w} - \frac{1}{w} \right] \mathrm{d}w. \quad (2.16)$$

By virtue of (1.13b), the value of the first integral is known to be $1 - \Phi(\hat{w})$, where $\Phi(\cdot)$ is the standard normal distribution function. For the second integral, the integrand is analytic in the vertical strip in the complex plane. One can apply the steepest descent method where the principal contribution of the integral comes from the behavior of the integrand near the immediate neighborhood of the saddlepoint (see Sect. 1.2). We freeze the value of the terms inside the bracket in the integrand in the second integral in (2.16) at the saddlepoint $z = \hat{z}$ (equivalently, $w = \hat{w}$) and apply the result in (2.11b) to obtain:

$$\frac{1}{z(w)} \frac{\mathrm{d}z(w)}{\mathrm{d}w} - \frac{1}{w} \approx \frac{1}{\hat{z}} \frac{\mathrm{d}z(\hat{w})}{\mathrm{d}w} - \frac{1}{\hat{w}} = \frac{1}{\hat{z}} \frac{1}{\kappa''(\hat{z})^{1/2}} - \frac{1}{\hat{w}},$$

when $\hat{z} \neq 0$ [which is equivalent to $\hat{w} \neq 0 \Leftrightarrow x \neq E[X] = \kappa'(0)$]. Using a similar derivation procedure shown in (2.12), we obtain the first result in (2.15).

When $\hat{z}$ is equal to zero (so does $\hat{w}$), we have $x = E[X] = \kappa'(0)$. It is necessary to consider the asymptotic limit of the saddlepoint approximation as $\hat{w} \to 0$ and $\hat{z} \to 0$ simultaneously. We start with the following Taylor expansion of $\kappa(z)$ at $\hat{z}$ in powers of $\hat{z}$ and evaluated at $z = \hat{z}$:

$$\kappa(0) = \kappa(\hat{z}) - \kappa'(\hat{z})\hat{z} + \frac{\kappa''(\hat{z})}{2}\hat{z}^2 - \frac{\kappa'''(\hat{z})}{6}\hat{z}^3 + \cdots$$

By observing that $\kappa'(\hat{z}) = x$ and $\kappa(0) = 0$, we find the leading power expansion in $\hat{z}$ of $\hat{w}$ as follows:

$$\hat{w}^2 = 2[\hat{z}x - \kappa(\hat{z})] = \kappa''(\hat{z})\hat{z}^2 - \frac{\kappa'''(\hat{z})}{3}\hat{z}^3 + \cdots$$

and

$$\hat{w} = \kappa''(\hat{z})^{1/2}\hat{z}\left[1 - \frac{\kappa'''(\hat{z})}{3\kappa''(\hat{z})}\hat{z} + \cdots\right]^{1/2} = \kappa''(\hat{z})^{1/2}\hat{z}\left[1 - \frac{\kappa'''(\hat{z})}{6\kappa''(\hat{z})}\hat{z} + \cdots\right].$$

Next, we consider

$$\frac{1}{\kappa''(\hat{z})^{1/2}\hat{z}} - \frac{1}{\hat{w}} = \frac{1}{\kappa''(\hat{z})^{1/2}\hat{z}}\left[1 - \frac{1}{1 - \frac{\kappa'''(\hat{z})}{6\kappa''(\hat{z})}\hat{z}} + \cdots\right] = -\frac{\kappa'''(\hat{z})}{6\kappa''(\hat{z})^{3/2}} + O(\hat{z}).$$

In the limit $\hat{w} \to 0$ and $\hat{z} \to 0$, we obtain

$$\lim_{\substack{\hat{w}\to 0\\ \hat{z}\to 0}} 1 - \Phi(\hat{w}) + \phi(\hat{w})\left[\frac{1}{\hat{z}\kappa''(\hat{z})^{1/2}} - \frac{1}{\hat{w}}\right] = \frac{1}{2} - \frac{1}{6\sqrt{2\pi}}\frac{\kappa'''(0)}{\kappa''(0)^{3/2}}.$$

□

Extension to n i.i.d. random variables

The above derivation is based on $n = 1$ and only the leading order approximation of the tail probability is obtained. For the general case of finding the tail probability of the mean $\bar{X}$ of n i.i.d. random variables and higher order approximation, one can perform similar calculations to arrive at the following Lugannani-Rice higher order approximation formulas under two separate cases.

(i) For $\hat{z}$ far away from zero, we have

$$\begin{aligned} P[\bar{X} > \bar{x}] = 1 - \Phi(\hat{w}) + \frac{\phi(\hat{w})}{\sqrt{n}}\Bigg\{&\frac{1}{\hat{z}\kappa''(\hat{z})^{1/2}} - \frac{1}{\hat{w}} \\ &+ \frac{1}{n}\left[\frac{1}{\hat{w}^3} - \frac{1}{\hat{z}^3\kappa''(\hat{z})^{3/2}} - \frac{\lambda_3}{2\hat{z}^2\kappa''(\hat{z})}\right. \\ &\left.+ \frac{1}{\hat{z}\kappa''(\hat{z})^{1/2}}\left(\frac{\lambda_4(\hat{z})}{8} - \frac{5\lambda_3^2(\hat{z})}{24}\right)\right] + O\left(\frac{1}{n^2}\right)\Bigg\}, \end{aligned} \tag{2.17a}$$

where $\hat{w} = \mathrm{sgn}(\hat{z})\sqrt{2n[\hat{z}\bar{x} - \kappa(\hat{z})]}$.

(ii) When $\hat{z}$ is close to zero, so does $\hat{w}$, an alternative higher order saddlepoint formula has to be used. By following a similar Taylor expansion of $\kappa(z) - zx$ (as in the proof of Proposition 2.1) and substituting into the Laplace inversion integral of $P[\bar{X} > \bar{x}]$, we obtain (see formula 2.2 in Lieberman 1994 and detailed derivation in García-Céspedes 2014)

$$P[\bar{X} > \bar{x}] = e^{n[\kappa(\hat{z})-\hat{z}x]+\frac{\hat{z}^2}{2}} \left\{ (1-\Phi(\hat{z})) \left[1 - \frac{\lambda_3(\hat{z})\hat{z}^3}{6\sqrt{n}} + \frac{1}{n} \left(\frac{\lambda_4(\hat{z})\hat{z}^4}{24} + \frac{\lambda_3^2 \hat{z}^6}{72} \right) \right] \right.$$
$$\left. + \phi(\hat{z}) \left[\frac{\lambda_3(\hat{z})}{6\sqrt{n}}(\hat{z}^2-1) - \frac{\lambda_4(\hat{z})}{24n}(\hat{z}^3-\hat{z}) - \frac{\lambda_3^2 \hat{z}}{72n}(\hat{z}^5-\hat{z}^3+3\hat{z}) \right] \right\}$$
$$+ O\left(\frac{1}{n^{3/2}}\right). \tag{2.17b}$$

As a final remark, Barndorff-Nielsen (1991) proposes the following alternative saddlepoint approximation formula for $P[\bar{X} > \bar{x}]$:

$$P[\bar{X} > \bar{x}] = 1 - \Phi\left[\hat{w} - \frac{1}{\hat{w}} \log \frac{1}{\hat{z}\sqrt{n\kappa''(\hat{z})}} \right],$$

where $\hat{w} = \text{sgn}(\hat{z})\sqrt{2n[\hat{z}x - \kappa(\hat{z})]}$. This formula has the nice property that its value is guaranteed to lie between 0 and 1, consistent with value of a probability.

2.2.1 Extension to Non-Gaussian Base

Similar to the saddlepoint approximation of density functions with a non-Gaussian base distribution, the Lugannani-Rice saddlepoint approximation formula for tail probabilities can be extended to the non-Gaussian base distribution, the details of which are stated in Proposition 2.3.

Proposition 2.3 (Wood et al. 1993) *Let $\kappa(z)$ denote the cumulant generating function of X that satisfies the technical conditions in Proposition 2.1. Let $\kappa_0(z)$, $f_0(x)$ and $F_0(x)$ be the cumulant generating function, density function and distribution function of the new base distribution, respectively. The saddlepoint approximation formula for $P[X > x]$ under the general base distribution is given by*

$$P[X > x] \approx \begin{cases} 1 - F_0\big(\kappa_0'(\hat{w})\big) + f_0\big(\kappa_0'(\hat{w})\big) \left\{ \frac{1}{\hat{z}} \left[\frac{\kappa_0''(\hat{w})}{\kappa''(\hat{z})} \right]^{1/2} - \frac{1}{\hat{w}} \right\}, & \\ \qquad\qquad \textit{when } x \neq E[X] = \kappa'(0) & \\ 1 - F_0\big(\kappa_0'(0)\big) + \frac{\kappa_0''(0)^{1/2} f_0\big(\kappa_0'(0)\big)}{6} \left[\frac{\kappa_0'''(0)}{\kappa_0''(0)^{3/2}} - \frac{\kappa'''(0)}{\kappa''(0)^{3/2}} \right], & \\ \qquad\qquad \textit{when } x = E[X] = \kappa'(0) & \end{cases} \tag{2.18}$$

where $\hat{z}$ solves the saddlepoint Eq. (2.2), and $\hat{w}$ is the solution to

$$\kappa(\hat{z}) - \hat{z}\kappa'(\hat{z}) = \kappa_0(w) - w\kappa_0'(w)$$

that is chosen to have the same sign as $\hat{z}$.

The proof of Proposition 2.3 involves the combined sets of techniques used in deriving the earlier saddlepoint formulas (2.14) (density function under the general base distribution) and (2.15) (tail probability under the Gaussian base distribution). For the details of the proof, interested readers may refer to Wood et al. (1993), Sect. 2 and Appendix B. Booth and Wood (1995) report an example in which the Lugannani-Rice saddlepoint formula fails when the Gaussian base is used while an excellent approximation is obtained when the Gaussian base is replaced by the inverse Gaussian base.

One can check that the Lugannani-Rice formula (2.15) is a special case of the above formula (2.18) when $\kappa_0(w) = \frac{w^2}{2}$ and $f_0(x) = \phi(x)$. Wood et al. (1993) also establish the following invariance property of the saddlepoint approximation formula (2.18), the details of which are stated in Proposition 2.4.

Proposition 2.4 (Wood et al. 1993) *Let $F_{a,b}(x)$ be the shifted and scaled distribution function of the original base distribution $F_0(x)$: $F_{a,b}(x) = F_0\big((x-a)/b\big)$. Suppose $F_{a,b}$ is used as the base distribution in* (2.18)*, then the resulting saddlepoint approximation formula does not depend on a or b, provided that $b > 0$.*

The proof can be found in Appendix A of Wood et al. (1993). Obviously, the saddlepoint approximation becomes exact when κ_0 in (2.18) is chosen to be the cgf of X. Together with Proposition 2.4, we obtain the following corollary.

Corollary 2.1 *If the base distribution used in Proposition 2.3 is a shifted and scaled transformation of the target distribution of X, then the saddlepoint approximation formula* (2.18) *is exact.*

Various numerical examples on comparison of accuracy between the saddlepoint approximations using the Gaussian and non-Gaussian bases are available in Wood et al. (1993). As a rule of thumb, the approximation would be more accurate if we use a base distribution that resembles the distribution of the target random variable X. This resemblance can be attained by matching the low order derivatives of the cumulant generating function of the base distribution evaluated at some chosen value with those corresponding to the target random variable X. When the chosen point at which the derivatives are evaluated is taken to be the zero value, it is equivalent to the simple moment matching method. In their applications in option pricing, Carr and Madan (2009) take the chosen evaluation point to be the saddlepoint of the target distribution.

2.2.2 Lattice Variables

A lattice variable assumes discrete set of values defined by $\alpha + nh$, where $h > 0$, α is a real number and n is an integer. We consider the extension of the Lugganani-Rice formula to lattice variables. For simplicity of our discussion, we assume that X only takes integer values k with probability p_k, where $P[X = k] = p_k$. The mgf and cgf of X are given by

$$M(z) = \exp(\kappa(z)) = \sum_k p_k e^{kz}.$$

Let S be the sum of n i.i.d. lattice variables $X_j, j = 1, 2, \ldots, n$ sharing the same cgf $\kappa(z)$, where $S = \sum_{j=1}^{n} X_j$. Let $p_n(s) = P[S = s]$, the probability mass function $p_n(s)$ is given by

$$p_n(s) = \frac{1}{2\pi i} \int_{-i\pi}^{i\pi} e^{n\kappa(z)-zs}\, \mathrm{d}z.$$

The integration interval is $[-i\pi, i\pi]$ due to the use of discrete Fourier transform for a lattice random variable. The tail probability of S is given by

$$\begin{aligned} P[S > s] = \sum_{m=s}^{\infty} p_n(m) &= \frac{1}{2\pi i} \int_{c-i\pi}^{c+i\pi} e^{n\kappa(z)-zs} \sum_{\ell=0}^{\infty} e^{-z\ell}\, \mathrm{d}z \\ &= \frac{1}{2\pi i} \int_{c-i\pi}^{c+i\pi} e^{n\kappa(z)-zs} \frac{1}{1-e^{-z}}\, \mathrm{d}z, \end{aligned} \tag{2.19}$$

where $c > 0$. To produce the saddlepoint approximation formula similar to that of a continuous random variable, we change the integration limit from $c \pm i\pi$ to $c \pm i\infty$ as an approximation since this only leads to a negligible error that is exponentially small in n. Next, we define

$$\overline{X} = S/n \text{ and } \overline{x} = s/n.$$

We also write

$$\frac{1}{1-e^{-z}} = \frac{z}{1-e^{-z}} \frac{1}{z}$$

and expand $\frac{z}{1-e^{-z}}$ about $z = \hat{z}$, where $\hat{z}$ is the saddlepoint obtained by solving $\kappa'(z) = \bar{x}$. The derivation of the corresponding Luggannani-Rice formula would be similar to that of the continuous case, except that a multiplier $\frac{\hat{z}}{1-e^{-\hat{z}}}$ appears in the term $\frac{1}{\hat{z}\kappa''(\hat{z})^{1/2}}$. As deduced from (2.15) and (2.17a), we obtain (Daniels 1987)

$$P[\bar{X} > \bar{x}] = 1 - \Phi(\hat{w}) + \frac{\phi(\hat{w})}{\sqrt{n}} \left[\frac{\hat{z}}{1-e^{-\hat{z}}} \frac{1}{\hat{z}\kappa''(\hat{z})^{1/2}} - \frac{1}{\hat{w}} \right], \tag{2.20}$$

where $\hat{w} = \mathrm{sgn}(\hat{z})\sqrt{2n[\hat{z}\bar{x} - \kappa(\hat{z})]}$.

2.3 Tail Expectations

Many pricing problems in financial engineering, like European option pricing and expected shortfall calculations in credit portfolios, are related to evaluation of tail expectations. In option pricing and insurance, we visualize $E[(X - K)^+]$ as the payoff of a call option and stop-loss premium, respectively. In this section, we use three different approaches to derive the saddlepoint approximations for tail expectations. The first type of saddlepoint approximation methods involve the conversion of tail expectations into tail probabilities by adopting the change of measure. The second type of methods make use of the Esscher exponential tilting and Edgeworth expansion techniques. The last type of methods start with the Laplace inversion representation and consider various methods of approximating the complex integral arising from the Laplace inversion integral.

2.3.1 Change of Measure Approach

This approach requires boundedness of the random variable. Suppose X is bounded below by $-L$ for some $L > 0$. Using the change of measure technique (Studer 2001), we show how to express the tail expectation $E[(X - K)^+]$ in terms of a tail probability. To achieve this, we define $Y = X + L$ and let $\mu_X = E[X]$, such that Y is nonnegative and $\mu_Y = \mu_X + L$. We then have

$$E[(X - K)^+] = E[(Y - L - K)\mathbf{1}_{\{Y > L+K\}}] = E[Y\mathbf{1}_{\{Y > L+K\}}] - (L + K)P[Y > L + K].$$

To express the first term as a tail probability, we define a new measure Q by

$$\frac{\mathrm{d}Q}{\mathrm{d}P} = \frac{Y}{\mu_Y}.$$

It is easy to verify that the Radon-Nikodym derivative defined above is nonnegative and integrates into one. As a result, we obtain

$$E[Y\mathbf{1}_{\{Y > L+K\}}] = \mu_Y E_Q[\mathbf{1}_{\{Y > L+K\}}] = \mu_Y Q[Y > L + K].$$

Combining the above results, we have

$$E[(X - K)^+] = \mu_Y Q[X > K] - (L + K)P[X > K]. \tag{2.21}$$

The cgf of X under the original measure P is given by $\kappa(z) = \log(E[e^{zX}])$ and recall $\kappa'(z) = \dfrac{E[Xe^{zX}]}{E[e^{zX}]}$. Under the measure Q, the cgf of X is found to be

$$\begin{aligned}
\kappa_Q(z) &= \log\left[\frac{E[(X+L)e^{zX}]}{\mu_Y}\right] \\
&= \log\left(\frac{E[Xe^{zX}] + LE[e^{zX}]}{E[e^{zX}]}\right) + \log E[e^{zX}] - \ln\mu_Y \\
&= \log(\kappa'(z) + L) + \kappa(z) - \ln\mu_Y.
\end{aligned}$$

Given the cgf's under both measures, we can obtain the saddlepoint approximation of $E[(X-K)^+]$ by simply applying the earlier approximation formulas discussed in the previous section for tail probabilities. In this approach, there are two saddlepoint equations to be solved and the technical condition of boundedness of the underlying random variable is required.

2.3.2 Esscher Exponential Tilting and Edgeworth Expansion

As an improvement to the change of measure approach by Studer (2001), Zheng and Kwok (2014) and Huang and Oosterlee (2011) provide two other saddlepoint approximation methods for calculating tail expectations using the Esscher exponential tilting technique. Zheng and Kwok (2014) derive the saddlepoint approximation of the tail expectation from that of the tail probability. Their formula is deduced from the renowned Lugannani-Rice approximation formula by differentiating with respect to the tilting parameter. On the other hand, Huang and Oosterlee (2011) apply the traditional Edgeworth expansion method to derive their saddlepoint approximation formulas for tail expectations.

Differentiation of the Lugannani-Rice formula

The general form $E[(X-K)^+]$ can be taken as $E[\tilde{X}^+]$, where $\tilde{X} = X - K$. If we let $F(x)$ and $\kappa(z)$ be the distribution function and cgf of X, respectively, then the distribution function and cgf of $\tilde{X}$ are given by

$$\tilde{F}(x) = F(x+K), \quad \tilde{\kappa}(z) = \kappa(z) - Kz.$$

Without loss of generality, we consider the saddlepoint approximation of the tail expectation $E[X^+]$.

To obtain the saddlepoint approximation of $E[X^+]$, we use $F(x;\theta)$ to denote the distribution function of the exponentially θ-tilted distribution of X such that

$$\mathrm{d}F(x;\theta) = e^{\theta x - \kappa(\theta)}\mathrm{d}F(x). \tag{2.22}$$

Note that $F(x; 0) = F(x)$. The cgf of the θ-tilted distribution is related to the original one by (see Appendix)

$$\kappa_\theta(z) = \kappa(z + \theta) - \kappa(\theta). \tag{2.23}$$

The tail expectation can be calculated as follows:

$$\begin{aligned} E[X^+] &= E\left[\frac{\partial}{\partial\theta} e^{\theta X} \mathbf{1}_{\{X>0\}}\Big|_{\theta=0}\right] \\ &= \frac{\partial}{\partial\theta}\left[e^{\kappa(\theta)} \int_0^\infty e^{\theta x - \kappa(\theta)} \mathrm{d}F(x)\right]\Big|_{\theta=0} \\ &= \frac{\partial}{\partial\theta}\left\{e^{\kappa(\theta)}[F(\infty; \theta) - F(0; \theta)]\right\}\Big|_{\theta=0} \\ &= \left\{\kappa'(\theta) e^{\kappa(\theta)}[1 - F(0; \theta)]\right\}\Big|_{\theta=0} - \left[e^{\kappa(\theta)} \frac{\partial F(0; \theta)}{\partial\theta}\right]\Bigg|_{\theta=0} \\ &= \kappa'(0)[1 - F(0)] - \frac{\partial F(0; \theta)}{\partial\theta}\Big|_{\theta=0}, \end{aligned} \tag{2.24}$$

where $F(\infty; \theta) = 1$ for any value of θ and $\kappa(0) = 0$. The tail expectation involves the derivative of the θ-tilted distribution function $F(x; \theta)$ with respect to the parameter θ. We then approximate $F(x; \theta)$ by the following Lugannani-Rice approximation $\hat{F}(x; \theta)$:

$$\hat{F}(x; \theta) \approx \begin{cases} \Phi(\hat{w}_\theta) + \phi(\hat{w}_\theta)\left(\dfrac{1}{\hat{w}_\theta} - \dfrac{1}{\hat{u}_\theta}\right), & x \neq \kappa'(\theta) \\ \dfrac{1}{2} + \dfrac{1}{6\sqrt{2\pi}} \dfrac{\kappa'''(\theta)}{\kappa''(\theta)^{3/2}}, & x = \kappa'(\theta), \end{cases} \tag{2.25}$$

where $\Phi(\cdot)$ and $\phi(\cdot)$ denote the standard normal distribution function and density function, respectively. The transformed saddlepoint $\hat{w}$ (see 2.9c) and $\hat{z}\kappa''(\hat{z})$ under the exponentially tilted distribution become

$$\begin{aligned} \hat{w}_\theta &= \text{sgn}(\hat{z}_x - \theta)\{2[(\hat{z}_x - \theta)x - \kappa(\hat{z}_x) + \kappa(\theta)]\}^{1/2}, \\ \hat{u}_\theta &= (\hat{z}_x - \theta)\sqrt{\kappa''(\hat{z}_x)}. \end{aligned}$$

Here, $\hat{z}_x$ denotes the unique solution (with dependence on x) to the usual saddlepoint equation:

$$\kappa'(z) = x.$$

The saddlepoint approximation to $\frac{\partial F(x;\theta)}{\partial\theta}$ is obtained by differentiating (2.25) with respect to θ, where

$$\frac{\partial \hat{F}(x;\theta)}{\partial\theta} \approx \begin{cases} \phi(\hat{w}_\theta)\left\{[x-\kappa'(\theta)]\left[\frac{1}{\hat{w}_\theta^3}-\frac{1}{\hat{u}_\theta}\right]-\frac{1}{(\hat{z}_x-\theta)\hat{u}_\theta}\right\}, & x\neq\kappa'(\theta) \\ \frac{1}{\sqrt{2\pi}}\left\{\frac{1}{24}\left[\frac{\kappa''''(\theta)}{\kappa''(\theta)^{\frac{3}{2}}}-\frac{\kappa'''(\theta)^2}{\kappa''(\theta)^{\frac{5}{2}}}\right]-\sqrt{\kappa''(\theta)}\right\}, & x=\kappa'(\theta). \end{cases} \tag{2.26}$$

The formula for the degenerate case $x=\kappa'(\theta)$ is obtained by taking the limit $\hat{z}_x\to\theta$, the details of which are provided in the Appendix. Finally, we obtain the following approximation to (2.24) as follows:

$$E[X^+]\approx\kappa'(0)[1-\hat{F}(0)]-\frac{\partial\hat{F}(0;\theta)}{\partial\theta}\Big|_{\theta=0}. \tag{2.27}$$

Lastly, by replacing X with $\tilde{X}=X-K$, we obtain the following saddlepoint approximation formula for the usual tail expectation:

$$\begin{aligned} &E[(X-K)^+] \\ &\approx[\kappa'(0)-K][1-\hat{F}(K)]-\frac{\partial\hat{F}(K;\theta)}{\partial\theta}\Big|_{\theta=0} \\ &=\begin{cases} [\kappa'(0)-K][1-\hat{F}(K)]+\phi(\hat{w})\left\{[K-\kappa'(0)]\left[\frac{1}{\hat{u}}-\frac{1}{\hat{w}^3}\right]+\frac{1}{\hat{z}_K\hat{u}}\right\}, & K\neq\kappa'(0) \\ \frac{1}{\sqrt{2\pi}}\left\{\frac{1}{24}\left[\frac{\kappa'''(0)^2}{\kappa''(0)^{5/2}}-\frac{\kappa''''(0)}{\kappa''(0)^{3/2}}\right]+\sqrt{\kappa''(0)}\right\}, & K=\kappa'(0), \end{cases} \end{aligned} \tag{2.28}$$

where $\hat{z}_K$ denotes the unique solution to $\kappa'(z)=K$, $\hat{w}=\text{sgn}(\hat{z}_K)\{2[\hat{z}_Kx-\kappa(\hat{z}_K)]\}^{1/2}$ and $\hat{u}=\hat{z}_K\sqrt{\kappa''(\hat{z}_K)}$.

The above saddlepoint approximation method has the advantage that it only requires computation of one saddlepoint and there is no boundedness constraint on X. Also, it is a compact saddlepoint approximation formula that only involves the first two cumulants for the general case $K\neq\kappa'(0)$. Higher order saddlepoint approximations can be obtained if one uses a higher order version of the Lugannani-Rice approximation. However, further differentiation with respect to θ to obtain the corresponding approximation could be notoriously tedious. As a final remark, an alternative derivation of approximation (2.27) can also be derived if one uses the saddlepoint approximation results for the moments of the truncated random variable, $X|X>0$ (Butler and Wood 2004).

By following a similar derivation approach using the differentiation of the Lugannani-Rice formula (see 2.18 and 2.28), one can obtain the following saddlepoint approximation formula for the tail expectation under non-Gaussian base distribution with density function f_0 (Zhang and Kwok 2018):

$$
E[(X-K)^+]
$$

$$
= \begin{cases} [\kappa'(0)-K][1-\hat{F}(K)] \\ +f_0(\kappa_0'(\hat{w}))\left\{[K-\kappa'(0)]\left[\frac{1}{\hat{w}}-\frac{1}{\hat{w}^3\kappa_0''(\hat{w})}-\frac{\kappa_0'''(\hat{w})}{2\hat{w}\kappa_0''(\hat{w})^{3/2}\hat{u}}\right]+\frac{\sqrt{\kappa_0''(\hat{w})}}{\hat{z}_K\hat{u}}\right\} \\ +f_0'(\kappa_0'(\hat{w}))[K-\kappa'(0)]\left[\frac{1}{\hat{w}^2}-\frac{\sqrt{\kappa_0''(\hat{w})}}{\hat{w}\hat{u}}\right], \qquad K\neq\kappa'(0); \\ f_0(\kappa_0'(0))\left\{\frac{\sqrt{\kappa_0''(0)}}{24\sqrt{\kappa''(0)}}\left[\frac{\kappa'''(0)^2}{\kappa''(0)^2}-\frac{\kappa''''(0)}{\kappa''(0)}\right]\right. \\ \qquad +\frac{\sqrt{\kappa''(0)}}{8\sqrt{\kappa_0''(0)}}\left[\frac{\kappa_0''''(0)^2}{\kappa_0''(0)^2}-\frac{\kappa_0''''(0)}{\kappa_0''(0)}\right] \\ \qquad \left.+\frac{1}{12}\frac{\kappa_0'''(0)}{\kappa_0''(0)}\frac{\kappa'''(0)}{\kappa''(0)}+\sqrt{\kappa''(0)\kappa_0''(0)}\right\} \\ +\frac{f_0'(\kappa_0'(0))\sqrt{\kappa''(0)}\kappa_0''(0)}{6}\left[\frac{\kappa'''(0)}{\kappa''(0)^{3/2}}-\frac{\kappa_0'''(0)}{\kappa_0''(0)^{3/2}}\right], \quad K=\kappa'(0). \end{cases} \tag{2.29}
$$

Here, κ_0 is the cgf of the non-Gaussian base distribution.

Edgeworth expansion

Huang and Oosterlee (2011) consider the approximation of $E[(X-K)^+]$ for the sum $X=\sum_{i=1}^n X_i$, where $X_i,\ i=1,2,\ldots,n$, are independent and identically distributed random variables that share $M_1(z)$ and $\kappa_1(z)$ as the common mgf and cgf. Obviously, the mgf and cgf of X are given by $M(z)=M_1(z)^n$ and $\kappa(z)=n\kappa_1(z)$, respectively.

Let $f(x)$ be the density function of X and define $\hat{z}$ to be the unique solution to the saddlepoint equation $\kappa'(z)=K$. We first consider the case where $\kappa'(0)\le K$ so that the saddlepoint solution satisfies $\hat{z}\ge 0$. The tail expectation is reformulated under an exponentially tilted probability measure as follows:

$$
E[(X-K)^+]=\int_K^\infty (x-K)f(x)\,\mathrm{d}x = e^{\kappa(\hat{z})-K\hat{z}}\int_K^\infty (x-K)e^{-\hat{z}(x-K)}\tilde{f}(x)\,\mathrm{d}x,
$$

where $\tilde{f}(x)=e^{\hat{z}x-\kappa(\hat{z})}f(x)$. Under this $\hat{z}$-tilted measure, the mgf of X is given by $\tilde{M}(z)=M(\hat{z}+z)/M(\hat{z})$. Recall that $\tilde{\kappa}(z)=\kappa(z+\hat{z})-\kappa(z)$, so $\tilde{\kappa}'(0)=\kappa'(\hat{z})=K$, and $\tilde{\kappa}''(0)=\kappa''(\hat{z})=n\kappa_1''(\hat{z})$. Therefore, the mean and variance of X under the $\hat{z}$-tilted measure are K and $n\kappa_1''(\hat{z})$, respectively. Let

$$
\Sigma=\sqrt{\kappa''(\hat{z})}=\sqrt{n\kappa_1''(\hat{z})}
$$

and define

$$
Y=(X-K)/\Sigma
$$

so that Y has zero mean and unit variance. Under the $\hat{z}$-tilted measure, the density function of Y can be expressed as

$$
g(y)=\Sigma\tilde{f}(\Sigma y+K).
$$

As a result, we have

$$E[(X-K)^+] = e^{\kappa(\hat{z})-K\hat{z}}\Sigma \int_0^\infty y e^{-\hat{z}\Sigma y} g(y)\, \mathrm{d}y. \tag{2.30a}$$

Suppose we use the lowest order Edgeworth expansion under the Gaussian base distribution to approximate $g(y)$ by

$$g(y) = \phi(y)[1 + O(n^{-1/2})], \tag{2.30b}$$

where $\phi(\cdot)$ is the normal density function. Substituting (2.30b) into (2.30a) and evaluating the Gaussian integral yields

$$E[(X-K)^+] = e^{\kappa(\hat{z})-K\hat{z}}\left[\frac{\Sigma}{\sqrt{2\pi}} - \hat{z}\Sigma^2 e^{\frac{(\hat{z}\Sigma)^2}{2}}\Phi\left(-\hat{z}\Sigma\right)\right][1 + O(n^{-1/2})],$$

where $\Phi(\cdot)$ is the normal distribution function. By deleting the error term, we obtain the following first order saddlepoint approximation to the tail expectation of order $O(n^{-1/2})$:

$$C_1 = e^{\kappa(\hat{z})-K\hat{z}}\left[\frac{\Sigma}{\sqrt{2\pi}} - \hat{z}\Sigma^2 e^{\frac{(\hat{z}\Sigma)^2}{2}}\Phi\left(-\hat{z}\Sigma\right)\right]. \tag{2.31}$$

Higher order saddlepoint approximations

Higher order saddlepoint approximations can be obtained analogously if one uses a higher order version of the Edgeworth expansion of $g(y)$. Instead of using the lowest order Edgeworth expansion as defined in (2.30b), we use the following higher order Edgeworth expansion:

$$g(y) = \phi(y)\left[1 + \frac{\kappa'''(\hat{z})}{\kappa''(\hat{z})^{3/2}}\frac{y^3-3y}{6} + O(n^{-1})\right].$$

After some tedious calculations, we obtain

$$\begin{aligned} E[(X-K)^+] = C_1[1 + O(n^{-1})] + e^{(\hat{z}\Sigma)^2/2+\kappa(\hat{z})-K\hat{z}}\frac{\kappa'''(\hat{z})}{6\kappa''(\hat{z})} \\ \left[\Phi(-\hat{z}\Sigma)(\hat{z}^2\Sigma^2+3)\hat{z}^2\Sigma^2 - \phi(\hat{z}\Sigma)(\hat{z}^2\Sigma^2+2)\hat{z}\Sigma\right]. \end{aligned}$$

Dropping the error term, we obtain the following second order saddlepoint approximation to the tail expectation:

$$C_2 = C_1 + e^{(\hat{z}\Sigma)^2/2+\kappa(\hat{z})-K\hat{z}} \frac{\kappa'''(\hat{z})}{6\kappa''(\hat{z})} \left[\Phi(-\hat{z}\Sigma)(\hat{z}^2\Sigma^2+3)\hat{z}^2\Sigma^2 - \phi(\hat{z}\Sigma)(\hat{z}^2\Sigma^2+2)\hat{z}\Sigma\right], \tag{2.32}$$

where C_1 is given in (2.31).

Negative saddlepoint

The derivation of C_1 and C_2 relies on the assumption that $\kappa'(0) \leq K$, or equivalently, $\hat{z} \geq 0$. The violation of this assumed condition may lead to erroneous approximation since accuracy of (2.30a) with $g(y)$ being replaced by (2.30b) is strongly affected by the exponential factor $e^{-\hat{z}\Sigma y}$. When $\hat{z} < 0$, the exponential factor no longer serves as a decay control for large n, but instead amplifies the residual error entailed by higher order terms in (2.30b). As a result, the Edgeworth expansion fails.

To modify the saddlepoint approximation formulas for the negative saddlepoint case, we consider $W = -X$ and utilize the put-call parity stated as

$$E[(X-K)^+] = \kappa'(0) - K + E[(K-X)^+],$$

where

$$E[(K-X)^+] = E[(-W-(-K))^+].$$

Since $\kappa_W'(z) = -\kappa'(z)$, the solution to the saddlepoint equation $\kappa_W'(z) = -K$ would be $-\hat{z}$, where $\hat{z}$ satisfies $\kappa'(\hat{z}) = K$. Therefore, the saddlepoint approximation of $E[(-W-(-K))^+]$ is readily known from C_1 and C_2 with (X, K) being replaced by $(-W, -K)$.

The respective first order and second order modified saddlepoint approximations to $E[(X-K)^+]$ for $K < \kappa'(0)$ are given by

$$\tilde{C}_1 = \kappa'(0) - K + e^{\kappa(\hat{z})-K\hat{z}} \left[\frac{\Sigma}{\sqrt{2\pi}} + \hat{z}\Sigma^2 e^{\frac{(\hat{z}\Sigma)^2}{2}} \Phi\left(\hat{z}\Sigma\right)\right], \tag{2.33}$$

$$\tilde{C}_2 = \tilde{C}_1 - e^{(\hat{z}\Sigma)^2/2+\kappa(\hat{z})-K\hat{z}} \frac{\kappa'''(\hat{z})}{6\kappa''(\hat{z})} \left[\Phi(\hat{z}\Sigma)(\hat{z}^2\Sigma^2+3)\hat{z}^2\Sigma^2 + \phi(\hat{z}\Sigma)(\hat{z}^2\Sigma^2+2)\hat{z}\Sigma\right]. \tag{2.34}$$

The above derivation of the saddlepoint approximations C_1 and C_2 ($\tilde{C}_1$ and $\tilde{C}_2$) (Huang and Oosterlee 2011) inherits the statistical approach for deriving the saddlepoint approximation of density function. There is only one saddlepoint equation to be solved. Moreover, the rate of convergence of the approximation can be determined and it is relatively straightforward to derive higher order saddlepoint approximations.

Table 2.1 Summary of the characteristics of the saddlepoint approximation formulas. The numbers with brackets in the column labeled "Highest order cumulant" represent the order of the highest order cumulant involved in the formulas under the degenerate case where $K = \mu$

Formula	Number of saddlepoints	Highest order cumulant	Removable singularity at μ
(2.21)	2	2 (3)	Yes
(2.28)	1	2 (4)	Yes
(2.31)/(2.33)	1	2 (2)	No
(2.32)/(2.34)	1	3 (2)	No

Numerical Examples

We would like to present studies on the numerical performance of the above saddlepoint approximation formulas. The important features of various methods are summarized in Table 2.1.

In terms of computational complexity, the first order approximation formulas (2.31)/(2.33) is apparently more desirable since it involves only the calculation of one saddlepoint and the first two cumulants. Also, no special provision is required to take special care of the degenerate case when $K = \mu$. Under usual scenarios, formula (2.28) is seen to be as efficient as (2.31)/(2.33). To have a close scrutiny of accuracy of these approximation formulas, we show numerical tests in which X is taken to be a gamma distribution with shape parameter α and scale parameter β. For any $x > 0$, the probability density of X is known to be

$$f(x; \alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta},$$

where $\Gamma(\alpha) = (\alpha - 1)! = \int_0^\infty e^{-x} x^{\alpha-1} \, \mathrm{d}x$ is the gamma function. The cumulative distribution function of X is found to be

$$F(x; \alpha, \beta) = \frac{\gamma(x/\beta; \alpha)}{\Gamma(\alpha)},$$

where $\gamma(x; \alpha) = \int_0^x e^{-t} t^{\alpha-1} \, \mathrm{d}t$ is the lower incomplete gamma function. Consequently, the tail expectation admits the following closed form expression

$$\begin{aligned} E[(X-K)^+] &= \int_K^\infty (x-K) f(x; \alpha, \beta) \, \mathrm{d}x \\ &= \alpha\beta[1 - F(K; \alpha+1, \beta)] - K[1 - F(K; \alpha, \beta)]. \end{aligned} \tag{2.35}$$

Meanwhile, the mgf of X is also explicitly known to be

Table 2.2 Calculations of tail expectations based on various saddlepoint approximation formulas. Moneyness is defined to be the ratio of K to the mean μ. The means for Panel A and Panel B are $\mu_A = 2$ and $\mu_B = 5$, respectively

Moneyness	Panel A: $\alpha = 1, \beta = 2$			Panel B: $\alpha = 5, \beta = 1$		
	0.2	1	1.8	0.2	1	1.8
(2.21)	1.638211	0.736611	0.329941	4.000690	0.877202	0.083668
(2.28)	1.633749	0.731394	0.328540	4.000682	0.877194	0.083758
(2.31)/(2.33)	1.629473	0.797885	0.481997	4.000568	0.892062	0.101290
(2.32)/(2.34)	1.639141	0.797885	0.323893	4.000676	0.892062	0.082014
Exact	1.637462	0.735759	0.330598	4.000689	0.877337	0.083780

$$M(z; \alpha, \beta) = \frac{1}{(1 - \beta z)^{\alpha}}, \quad z < 1/\beta.$$

Our numerical tests compare the numerical results produced by the saddlepoint approximation formulas listed in Table 2.1 with the exact result calculated using (2.35). In Table 2.2, we show the numerical tests under two scenarios: gamma distribution in Panel A that exhibits relatively large skewness and kurtosis, and gamma distribution in Panel B that resembles closer to the normal distribution. Overall speaking, we observe that the approximations in Panel B are slightly better than Panel A, suggesting that these Gaussian-based saddlepoint approximations perform better for distributions that are close to the Gaussian distribution. Accuracy depends sensibly on the relative magnitude of K to the mean of the random variable μ. We define moneyness to be the ratio of K to μ. In terms of accuracy, we observe that (2.21) dominates the other approximation formulas, and it is closely followed by (2.28), for all chosen values of K. Formulas (2.31)/(2.33) in general have worse performance in accuracy and fail to provide sufficient accuracy for the at-the-money (moneyness equals 1) and out-of-the-money (moneyness equals 1.8) cases. Higher order approximation given by C_2 ($\tilde{C}_2$) does improve accuracy significantly for the out-of-the-money case, yet it remains to be unsatisfactory for the at-the-money case. We expect to observe poor accuracy for the at-the-money case, since C_1 and C_2 reduce to the same formula when $K = \mu$. The above observations are also illustrated by the various plots of the percentage errors of the approximation results shown in Fig. 2.1.

2.3.3 Laplace Inversion Representation

The third approach to derive saddlepoint approximations of tail expectation is based on the Laplace inversion integral representation of the tail expectation. The approximation techniques include the Taylor expansion method, frozen integrand method, local quadratic approximation method, and a variant of the steepest descent method.

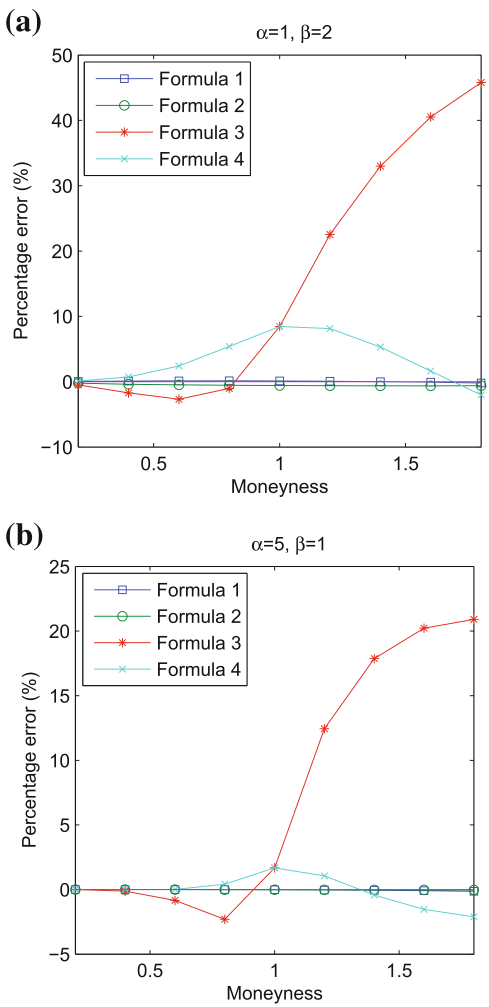


Fig. 2.1 Plots of the percentage errors of various saddlepoint approximations with varying values of K/μ (moneyness). "Formula 1" stands for (2.21), "Formula 2" stands for (2.28), "Formula 3" stands for (2.31)/(2.33), and "Formula 4" stands for (2.32)/(2.34). Formula 3 fails to provide satisfactory approximations at high values of moneyness

Like the Lugannani-Rice formula, the resulting approximation formulas are surprisingly accurate, even for $n = 1$.

Recall the following Laplace inversion formula for the tail expectation, $E[(X - K)^+]$ (see 1.13c):

$$E[(X - K)^+] = \frac{1}{2\pi i} \int_{\tau - i\infty}^{\tau + i\infty} \frac{e^{\kappa(z) - Kz}}{z^2} \, \mathrm{d}z, \quad \tau > 0, \tag{2.36}$$

where the convergence strip of $\kappa(z)$ contains at least some small neighborhood of the origin and τ lies within the convergence strip of $\kappa(z)$. By Cauchy's Theorem, we are allowed to shift the contour of integration to the left of the origin provided that the complex integral is properly adjusted by the residue at $z = 0$, whose value is $\kappa'(0) - K$. An alternative Laplace inversion formula for the tail expectation is given by

$$E[(X-K)^+] = \kappa'(0) - K + \frac{1}{2\pi i}\int_{\tilde{\tau}-i\infty}^{\tilde{\tau}+i\infty} \frac{e^{\kappa(z)-Kz}}{z^2}\,\mathrm{d}z, \quad \tilde{\tau} < 0, \tag{2.37a}$$

where the vertical line through $\tilde{\tau}$ represents the new contour within the convergence strip of $\kappa(z)$. Since $\kappa'(0) = E[X]$ and recall the following Laplace inversion formula for $E[(K-X)^+]$:

$$E[(K-X)^+] = \frac{1}{2\pi i}\int_{\hat{\tau}-i\infty}^{\hat{\tau}+i\infty} \frac{e^{\kappa(z)-Kz}}{z^2}\,\mathrm{d}z, \quad \hat{\tau} < 0, \tag{2.37b}$$

the combination of (2.37a) and (2.37b) resemble the put-call parity in option pricing. The saddlepoint approximation methods discussed below represent the various types of approximation of the complex integral in (2.36) or (2.37a).

Taylor expansion method

Antonov et al. (2007) consider the power series expansion of the exponential part of the integrand so that the complex integral in (2.36) is decomposed into analytically tractable Gaussian integrals. Let $\hat{z}$ be the unique solution to $\kappa'(z) = K$. We consider the following power series expansion:

$$\begin{aligned} e^{\kappa(z)-Kz} &= e^{\kappa(\hat{z})-K\hat{z}+\frac{\kappa''(\hat{z})}{2}(z-\hat{z})^2} e^{\frac{\kappa'''(\hat{z})}{6}(z-\hat{z})^3+\cdots} \\ &\approx e^{\kappa(\hat{z})-K\hat{z}+\frac{\kappa''(\hat{z})}{2}(z-\hat{z})^2}\left[1 + \frac{\kappa'''(\hat{z})}{6}(z-\hat{z})^3\right], \end{aligned} \tag{2.38}$$

where we have made use of the Taylor series of $\kappa(z) - Kz$ and e^z, respectively. Substituting (2.38) into (2.36), we obtain the approximation to the tail expectation as follows:

$$E[(X-K)^+] \approx \frac{e^{\kappa(\hat{z})-K\hat{z}}}{2\pi i}\int_{\tau-i\infty}^{\tau+i\infty} \frac{e^{\frac{\kappa''(\hat{z})}{2}(z-\hat{z})^2}}{z^2}\left[1 + \frac{\kappa'''(\hat{z})}{6}(z-\hat{z})^3\right]\mathrm{d}z. \tag{2.39}$$

To obtain the saddlepoint approximation formula, thanks to the following lemma, the above approximating integral can be evaluated explicitly.

Lemma 2.1 *For any $\lambda > 0$ and $\tau \neq 0$, define*

$$J_k(\lambda, z_0) = \frac{1}{2\pi i}\int_{\tau-i\infty}^{\tau+i\infty} \frac{e^{\frac{\lambda}{2}(z-z_0)^2}}{z^k}\,\mathrm{d}z, \quad k = -1, 0, 1, 2, \ldots.$$

The first few members admit the following closed form formulas:

$$
\begin{aligned}
J_0(\lambda, z_0) &= \frac{1}{\sqrt{2\pi\lambda}}, \quad J_{-1}(\lambda, z_0) = \frac{z_0}{\sqrt{2\pi\lambda}}, \\
J_1(\lambda, z_0) &= \operatorname{sgn}(\tau) e^{\frac{\lambda}{2} z_0^2} \Phi(-\operatorname{sgn}(\tau)\sqrt{\lambda} z_0), \\
J_2(\lambda, z_0) &= \sqrt{\frac{\lambda}{2\pi}} - \operatorname{sgn}(\tau)\lambda z_0 e^{\frac{\lambda}{2} z_0^2} \Phi(-\operatorname{sgn}(\tau)\sqrt{\lambda} z_0).
\end{aligned} \tag{2.40}
$$

The proof of Lemma 2.1 is presented in the Appendix. We now express the approximating complex integral in (2.39) in terms of $J_k(\kappa''(\hat{z}), \hat{z})$, $k = -1, 0, 1, 2$. First, by noting

$$
1 + \frac{\kappa'''(\hat{z})}{6}(z - \hat{z})^3 = 1 + \frac{\kappa'''(\hat{z})}{6}(z^3 - 3\hat{z}z^2 + 3\hat{z}^2 z - \hat{z}^3),
$$

and using Lemma 2.1, we obtain

$$
\begin{aligned}
&\frac{1}{2\pi i} \int_{\tau - i\infty}^{\tau + i\infty} \frac{e^{\frac{\kappa''(\hat{z})}{2}(z-\hat{z})^2}}{z^2} \left[1 + \frac{\kappa'''(\hat{z})}{6}(z - \hat{z})^3\right] \mathrm{d}z \\
&= \frac{\kappa'''(\hat{z})}{6} J_{-1}(\kappa''(\hat{z}), \hat{z}) - \frac{\hat{z}\kappa'''(\hat{z})}{2} J_0(\kappa''(\hat{z}), \hat{z}) \\
&\quad + \frac{\hat{z}^2\kappa'''(\hat{z})}{2} J_1(\kappa''(\hat{z}), \hat{z}) + \left[1 - \frac{\hat{z}^3\kappa'''(\hat{z})}{6}\right] J_2(\kappa''(\hat{z}), \hat{z}).
\end{aligned}
$$

Substituting the above result into (2.39), we obtain the following saddlepoint approximation formula for the tail expectation:

$$
\begin{aligned}
&E[(X - K)^+] \\
&\approx e^{\kappa(\hat{z}) - K\hat{z}} \left\{ \frac{\kappa'''(\hat{z})}{6} J_{-1}(\kappa''(\hat{z}), \hat{z}) - \frac{\hat{z}\kappa'''(\hat{z})}{2} J_0(\kappa''(\hat{z}), \hat{z}) \right. \\
&\quad \left. + \frac{\hat{z}^2\kappa'''(\hat{z})}{2} J_1(\kappa''(\hat{z}), \hat{z}) + \left[1 - \frac{\hat{z}^3\kappa'''(\hat{z})}{6}\right] J_2(\kappa''(\hat{z}), \hat{z}) \right\}.
\end{aligned} \tag{2.41}
$$

Antonov et al.'s approach shares a common drawback with other saddlepoint approximation methods that one cannot directly estimate the residual error of the approximation formula. The derivation of this approach does not treat negative saddlepoint separately. As indicated by numerical tests, formula (2.41) does not work well for large K.

Frozen integrand method

Martin (2006) converts the complex integral in the Laplace inversion formula with a double pole at $z = 0$ into one with a simple pole via integration by parts. It then follows by freezing the integrand of the new complex integral at the saddlepoint. He then expresses the tail expectation in terms of the corresponding tail probability and density, for which saddlepoint approximation formulas are available from the previous sections.

Using integration by parts, (2.36) can be rewritten as

$$E[(X-K)^+] = -\frac{e^{\kappa(z)-Kz}}{z}\bigg|_{\tau-i\infty}^{\tau+i\infty} + \frac{1}{2\pi i}\int_{\tau-i\infty}^{\tau+i\infty} \frac{e^{\kappa(z)-Kz}}{z}[\kappa'(z)-K]\,\mathrm{d}z.$$

Here, τ is chosen to be within the convergence strip of $\kappa(z)$, so that $e^{\kappa(z)-Kz}$ is bounded. Hence, $\frac{e^{\kappa(z)-Kz}}{z}$ vanishes at $|z| \to \infty$. We then have

$$E[(X-K)^+] = \frac{1}{2\pi i}\int_{\tau-i\infty}^{\tau+i\infty} \frac{e^{\kappa(z)-Kz}}{z}[\kappa'(z)-K]\,\mathrm{d}z.$$

Martin (2006) proceeds to decompose the above complex integral into the following two parts:

$$E[(X-K)^+] = \frac{1}{2\pi i}\int_{\tau-i\infty}^{\tau+i\infty} e^{\kappa(z)-zK}\frac{\mu-K}{z}\,\mathrm{d}z + \frac{1}{2\pi i}\int_{\tau-i\infty}^{\tau+i\infty} e^{\kappa(z)-zK}\frac{\kappa'(z)-\mu}{z}\,\mathrm{d}z, \tag{2.42}$$

where $\mu = E[X] = \kappa'(0)$. The first integral equals $(\mu - K)P[X > K]$. The second integral is the regular part, which is approximated using the steepest descent method. We deform the contour of integration in the second integral to pass through the saddlepoint $z = \hat{z}$ that solves the saddlepoint equation

$$\kappa'(z) = K.$$

Next, we freeze $\frac{\kappa'(z)-\mu}{z}$ in the integrand at the above saddlepoint by setting

$$\frac{\kappa'(z)-\mu}{z} \approx \frac{\kappa'(\hat{z})-\mu}{\hat{z}}, \tag{2.43}$$

provided that $\hat{z} \neq 0 \Leftrightarrow K \neq \mu$. Otherwise, when $\hat{z}$ converges to 0^+, we apply asymptotic analysis to obtain

$$\frac{\kappa'(z)-\mu}{z} \approx \kappa''(0). \tag{2.44}$$

Finally, we substitute (2.43) and (2.44) into the second integral in (2.42). Direct integration of the complex integral gives

$$E[(X-K)^+] \approx \begin{cases} (\mu-K)P[X > K] + \dfrac{K-\mu}{\hat{z}} f_X(K), & \text{if } K \neq \mu \\ (\mu-K)P[X > K] + \kappa''(0) f_X(K), & \text{if } K = \mu, \end{cases} \tag{2.45}$$

where f_X denotes the density function of X.

The saddlepoint approximation formulas for density $f_X(K)$ and tail probability $P[X > K]$ developed in the preceding subsections, either with the Gaussian base or non-Gaussian base, can be substituted into (2.45) to produce the corresponding saddlepoint approximations for tail expectations. Martin's approach is an indirect saddlepoint approximation method that converts the tail expectation into a combination of tail probability and density function, for which further saddlepoint approximations are required. There is only one saddlepoint equation to be solved in the numerical implementation procedure.

Local quadratic approximation method

Huang and Oosterlee (2011) apply the local quadratic approximation to the exponent, $\kappa(z) - Kz$, to derive the saddlepoint approximation formulas for the tail expectation $E[(X-K)^+]$, where $X = \sum_{i=1}^{n} X_i$. Here, X_i, $i = 1, 2, \ldots, n$, are independent and identically distributed random variables that share $\kappa_1(z)$ as the common cgf. We let $k = K/n$ and recall the saddlepoint equation: $\kappa(\hat{z}) = K \Leftrightarrow \kappa_1(\hat{z}) = k$. The key idea of the local quadratic approximation approach is to approximate $\kappa_1(z) - kz$ in (2.36) over an interval with 0 and $\hat{z}$ as the endpoints by a quadratic function, where

$$\frac{(w-\hat{w})^2}{2} = \kappa_1(z) - \kappa_1(\hat{z}) - k(z-\hat{z}), \tag{2.46}$$

such that when z goes from 0 to $\hat{z}$, the new variable w goes from 0 to $\hat{w}$. By setting $w = 0$ (hence, $z = 0$), we determine $\hat{w}$ via the relation:

$$\frac{\hat{w}^2}{2} = k\hat{z} - \kappa_1(\hat{z}), \tag{2.47a}$$

so that

$$e^{\kappa(z)-Kz} = e^{n[\kappa_1(z)-kz]} = e^{n\left(\frac{w^2}{2}-\hat{w}w\right)}.$$

Among the two choices of value with opposite signs, we choose the one with the same sign of $\hat{z}$, so

$$\hat{w} = \operatorname{sgn}(\hat{z})\sqrt{2[k\hat{z} - \kappa_1(\hat{z})]}. \tag{2.47b}$$

Differentiating (2.46) with respect to w once and twice, respectively, we obtain

$$w - \hat{w} = [\kappa_1'(z) - k]\frac{\mathrm{d}z}{\mathrm{d}w}, \tag{2.48a}$$

$$1 = [\kappa_1'(z) - k]\frac{\mathrm{d}^2z}{\mathrm{d}w^2} + \kappa_1''(z)\left(\frac{\mathrm{d}z}{\mathrm{d}w}\right)^2. \tag{2.48b}$$

At $w = \hat{w}$ (or equivalently, $z = \hat{z}$), we observe $\kappa_1(\hat{z}) = k$ in (2.48b), so we have

$$\left.\frac{\mathrm{d}z}{\mathrm{d}w}\right|_{w=\hat{w}} = \frac{1}{\sqrt{\kappa_1''(\hat{z})}}. \tag{2.49}$$

On the other hand, at $w = 0$ (or equivalently, $z = 0$), we deduce from (2.49) and (2.48a), respectively, to obtain

$$\left.\frac{\mathrm{d}z}{\mathrm{d}w}\right|_{w=0} = \begin{cases} \dfrac{1}{\sqrt{\kappa_1''(0)}}, & \text{if } \hat{z} = 0, \\ \dfrac{\hat{w}}{\kappa_1'(\hat{z}) - \kappa_1'(0)}, & \text{if } \hat{z} \neq 0. \end{cases} \tag{2.50}$$

It can be easily seen that the case $\hat{z} = 0$ is just the limiting result as $\hat{z} \to 0$. To proceed, we first rewrite (2.36) with a change of variable:

$$E[(X - K)^+] = \frac{1}{2\pi i} \int_{\zeta - i\infty}^{\zeta + i\infty} e^{n(\frac{w^2}{2} - \hat{w}w)} \frac{1}{z^2} \frac{\mathrm{d}z}{\mathrm{d}w} \mathrm{d}w, \quad \zeta > 0. \tag{2.51}$$

Taking the first three terms of the Laurent expansion of $\frac{1}{z^2} \frac{\mathrm{d}z}{\mathrm{d}w}$ at $w = 0$, we consider the approximation

$$\frac{1}{z^2} \frac{\mathrm{d}z}{\mathrm{d}w} \approx A_0 + A_1 w^{-1} + A_2 w^{-2}, \tag{2.52}$$

where

$$\begin{aligned} A_1 &= \frac{1}{2\pi i} \oint_C \frac{1}{z^2} \frac{\mathrm{d}z}{\mathrm{d}w} \, \mathrm{d}w = \frac{1}{2\pi i} \oint_C \frac{1}{z^2} \, \mathrm{d}z, \\ A_2 &= \frac{1}{2\pi i} \oint_C \frac{1}{z^2} \frac{\mathrm{d}z}{\mathrm{d}w} w \, \mathrm{d}w = \frac{1}{2\pi i} \oint_C \frac{w}{z^2} \, \mathrm{d}z. \end{aligned}$$

Here, the contour of integration C traces out a circle around 0 in the counterclockwise sense. Note that in any small neighborhood of $z = 0$, w is almost linear in z, $\frac{w}{z^2}$ and $\frac{1}{z^2}$ have poles of order 1 and 2 at $z = 0$, respectively. By virtue of (2.50), we obtain

$$A_1 = 0, \quad A_2 = \lim_{z \to 0} \frac{w}{z} = \left.\frac{\mathrm{d}w}{\mathrm{d}z}\right|_{z=0} = \frac{\kappa_1'(\hat{z}) - \kappa_1'(0)}{\hat{w}}.$$

Finally, the constant term A_0 is chosen in such a way that the approximation given by (2.52) becomes exact at $z = \hat{z}$. Using (2.49) and (2.52), we obtain

$$A_0 = \frac{1}{\hat{z}^2 \sqrt{\kappa_1''(\hat{z})}} + \frac{\kappa_1'(0) - \kappa_1'(\hat{z})}{\hat{w}^3}.$$

Substituting (2.52) into (2.51) and changing the integration variable to $y = \sqrt{n}w$ yields

$$E[(X-K)^+] \approx \frac{A_0}{2\pi i\sqrt{n}} \int_{\xi-i\infty}^{\xi+i\infty} e^{\frac{y^2}{2}-\sqrt{n}\,\hat{w}y}\,\mathrm{d}y + \frac{\sqrt{n}A_2}{2\pi i} \int_{\xi-i\infty}^{\xi+i\infty} \frac{e^{\frac{y^2}{2}-\sqrt{n}\,\hat{w}y}}{y^2}\,\mathrm{d}y, \tag{2.53}$$

where $\xi > 0$. The first integral in (2.53) can be seen as the Laplace inversion formula for the standard normal density function:

$$\frac{A_0}{2\pi i\sqrt{n}} \int_{\xi-i\infty}^{\xi+i\infty} e^{\frac{y^2}{2}-\sqrt{n}\,\hat{w}y}\,\mathrm{d}y = \frac{A_0}{\sqrt{n}}\phi(\sqrt{n}\hat{w}) = \frac{A_0}{\sqrt{2\pi n}}\, e^{-\frac{n}{2}\hat{w}^2}.$$

For the second integral in (2.53), we recognize that

$$\frac{1}{2\pi i} \int_{\xi-i\infty}^{\xi+i\infty} \frac{e^{\frac{y^2}{2}-\sqrt{n}\,\hat{w}y}}{y^2}\,\mathrm{d}y = E[(Z-\sqrt{n}\,\hat{w})^+],$$

where Z is a standard normal random variable. It is easily seen that

$$E[(Z-\sqrt{n}\,\hat{w})^+] = \phi(\sqrt{n}\,\hat{w}) - \sqrt{n}\,\hat{w}\Phi(-\sqrt{n}\,\hat{w}).$$

Putting all the results together, we have

$$E[(X-K)^+] \approx C_4 = \begin{cases} n[\kappa_1'(0)-k]\left[\Phi(-\sqrt{n}\,\hat{w}) - \frac{\phi(\sqrt{n}\,\hat{w})}{\sqrt{n}\,\hat{w}}\right] & \\ + \frac{\phi(\sqrt{n}\,\hat{w})}{\sqrt{n}}\left[\frac{1}{\hat{z}^2\sqrt{\kappa_1''(\hat{z})}} + \frac{\kappa_1'(0)-\kappa_1'(\hat{z})}{\hat{w}^3}\right], & \kappa_1'(0) \neq k, \\ \sqrt{\frac{n\kappa_1''(0)}{2\pi}} + \frac{1}{24\sqrt{2\pi n}}\left[\frac{\kappa_1'''(0)^2}{\kappa_1''(0)^{5/2}} - \frac{\kappa_1''''(0)}{\kappa_1''(0)^{3/2}}\right], & \kappa_1'(0) = k. \end{cases} \tag{2.54}$$

When we set $n = 1$, formula (2.54) becomes identical to formula (2.28) if we take

$$1-\hat{F}(k) = \Phi(-\hat{w}) - \frac{\phi(\hat{w})}{\hat{w}}.$$

Though the two derivation procedures of the two saddlepoint approximation formulas are quite different, it is quite remarkable that the two final analytic forms can be the same.

The above procedure does not offer an estimation of the order of error in the approximation formula (2.54). In order to perform an order analysis of the error, an alternative procedure (Huang and Oosterlee 2011) is presented below.

Error analysis and order of accuracy

Using the transformed variable w defined in (2.46), (2.41) can be reformulated as

$$
\begin{aligned}
E[(X-K)^+] &= \frac{1}{2\pi i}\int_{\zeta-i\infty}^{\zeta+i\infty} n[\kappa_1'(z)-k]e^{n(\frac{w^2}{2}-\hat{w}w)}\frac{1}{z}\frac{\mathrm{d}z}{\mathrm{d}w}\,\mathrm{d}w \\
&= \frac{n[\kappa_1'(0)-k]}{2\pi i}\int_{\zeta-i\infty}^{\zeta+i\infty}\frac{e^{n(\frac{w^2}{2}-\hat{w}w)}}{w}\,\mathrm{d}w \\
&\quad + \frac{n}{2\pi i}\int_{\zeta-i\infty}^{\zeta+i\infty} e^{n(\frac{w^2}{2}-\hat{w}w)}\left[\frac{\kappa_1'(z)-k}{z}\frac{\mathrm{d}z}{\mathrm{d}w}-\frac{\kappa_1'(0)-k}{w}\right]\mathrm{d}w.
\end{aligned}
\tag{2.55}
$$

By substituting $y=\sqrt{n}\,w$, the first integral is recognized as

$$
\frac{1}{2\pi i}\int_{\zeta-i\infty}^{\zeta+i\infty}\frac{e^{\frac{y^2}{2}-\sqrt{n}\,\hat{w}y}}{y}\,\mathrm{d}y = P[Z\geq\sqrt{n}\,\hat{w}] = \Phi(-\sqrt{n}\,\hat{w}), \tag{2.56}
$$

where Z is a standard normal random variable. For the second integral in (2.55), the major contribution to the complex integral comes from the saddlepoint $\hat{z}$. We define

$$
f(w)=\frac{\kappa_1'(z)-k}{z}\frac{\mathrm{d}z}{\mathrm{d}w}-\frac{\kappa_1'(0)-k}{w}
$$

and expand $f(w)$ around $\hat{w}$ (or equivalently, around $\hat{z}$) in power series expansion in $w-\hat{w}$ to obtain

$$
f(w)=f(\hat{w})+f'(\hat{w})(w-\hat{w})+f''(\hat{w})\frac{(w-\hat{w})^2}{2}+\ldots.
$$

By substituting $y=\sqrt{n}w$, we have

$$
f(y/\sqrt{n})=f(\hat{w})+f'(\hat{w})\frac{y-\sqrt{n}\,\hat{w}}{\sqrt{n}}+f''(\hat{w})\frac{(y-\sqrt{n}\,\hat{w})^2}{2n}+\ldots. \tag{2.57}
$$

Note that $f(\hat{w})=-\dfrac{\kappa_1'(0)-k}{\hat{w}}$. To derive the lower order saddlepoint approximation formula, we take the first two terms in $f(w)$ and substitute into the second integral in (2.55) so that

$$
\frac{n}{2\pi i}\int_{\xi-i\infty}^{\xi+i\infty} e^{\frac{y^2}{2}-\sqrt{n}\,\hat{w}y}\frac{1}{\sqrt{n}}\left[-\frac{\kappa_1'(0)-k}{\hat{w}}+f'(\hat{w})\frac{y-\sqrt{n}\,\hat{w}}{\sqrt{n}}+O(n^{-1})\right]\mathrm{d}y.
$$

Since integrating the Gaussian integral with the term that is linear in $y-\sqrt{n}\,\hat{w}$ gives the value zero, the above integral can be simplified as

$$
n\left[-\frac{\kappa_1'(0)-k}{\sqrt{n}\,\hat{w}}\phi(\sqrt{n}\,\hat{w})+O(n^{-3/2})\right].
$$

Putting all these results together yields

$$E[(X-K)^+] \approx n[\kappa_1'(0)-k]\left[\Phi(-\sqrt{n}\,\hat{w}) - \frac{\phi(\sqrt{n}\,\hat{w})}{\sqrt{n}\,\hat{w}}\right] + O(n^{-1/2}). \tag{2.58}$$

By dropping the error term, we obtain the following approximation formula

$$E[(X-K)^+] \approx C_3 = n[\kappa_1'(0)-k]\left[\Phi(-\sqrt{n}\,\hat{w}) - \frac{\phi(\sqrt{n}\,\hat{w})}{\sqrt{n}\,\hat{w}}\right]. \tag{2.59}$$

In our notation, we let C_3 be the first term of C_4 in (2.54). When $k = \kappa_1'(0)$, we have $\hat{z} = 0$. Using (2.50), we deduce that

$$\lim_{\hat{z}\to 0} \frac{\hat{w}}{\kappa_1'(\hat{z}) - \kappa_1'(0)} = \frac{1}{\sqrt{\kappa_1''(0)}}$$

so that (2.59) degenerates into

$$C_3 = \sqrt{\frac{n\kappa_1''(0)}{2\pi}}.$$

Indeed, C_3 as the leading term of C_4 in (2.54) has order of accuracy $O(n^{-1/2})$. To obtain C_4 completely, one needs to include more terms in (2.57) to approximate $f(y/\sqrt{n})$. It is necessary to consider the expansion of $f(y/\sqrt{n})$ up to the third order term so that

$$\begin{aligned} f\left(\frac{y}{\sqrt{n}}\right) = & f(\hat{w}) + f'(\hat{w})\frac{y-\sqrt{n}\,\hat{w}}{\sqrt{n}} + \frac{f''(\hat{w})}{2}\frac{(y-\sqrt{n}\,\hat{w})^2}{n} \\ & + \frac{f'''(\hat{w})}{6}\frac{(y-\sqrt{n}\,\hat{w})^3}{n^{3/2}} + O(n^{-2}). \end{aligned} \tag{2.60}$$

To compute $f''(\hat{w})$, we first differentiate (2.48b) with respect to w once again and evaluate the equation at $w = \hat{w}$, yielding

$$[\hat{z}\kappa_1''(\hat{z})^{1/2} + \kappa_1''(\hat{z})]\frac{\mathrm{d}^2 z}{\mathrm{d}w^2}\bigg|_{w=\hat{w}} + \frac{\kappa_1'''(\hat{z})}{\kappa_1''(\hat{z})^{3/2}} = 0.$$

After tedious calculations and applying the above equation and (2.48b), we obtain

$$f''(\hat{w}) = -\frac{2}{\hat{z}^2\sqrt{\kappa_1''(\hat{z})}} - \frac{2[\kappa_1'(0)-k]}{\hat{w}^3}.$$

If we substitute (2.60) into the second integral in (2.55), then the Gaussian integrals with the terms that involve odd powers of $y - \sqrt{n}\,\hat{w}$ are integrated to zero. To be more specific, by shifting the contour so that $\zeta = \sqrt{n}\hat{w}$, the transformed integral becomes a Gaussian integral of odd powers, which is integrated to give zero value. The complex integral has the following remaining terms:

$$n\left[-\frac{\kappa_1'(0)-k}{\sqrt{n}\,\hat{w}}\phi(\sqrt{n}\,\hat{w})\right] + \frac{nf''(\hat{w})}{4\pi i}\int_{\zeta-i\infty}^{\zeta+i\infty} e^{\frac{y^2}{2}-\sqrt{n}\,\hat{w}y}\frac{(y-\sqrt{n}\,\hat{w})^2}{n\sqrt{n}}\,\mathrm{d}y + O(n^{-3/2}).$$

By observing

$$\frac{1}{2\pi i}\int_{\zeta-i\infty}^{\zeta+i\infty} e^{\frac{1}{2}(y-\sqrt{n}\,\hat{w})^2}(y-\sqrt{n}\,\hat{w})^2\,\mathrm{d}y = \phi''(0) = -\frac{1}{\sqrt{2\pi}},$$

so that the last integral in the above expression becomes

$$-\frac{f''(\hat{w})e^{-\frac{n}{2}\hat{w}^2}}{2\sqrt{2\pi n}} = \frac{\phi(\sqrt{n}\,\hat{w})}{\sqrt{n}}\left[\frac{1}{\hat{z}^2\sqrt{\kappa_1''(\hat{z})}} + \frac{\kappa_1'(0)-k}{\hat{w}^3}\right].$$

Putting all these results together, we obtain

$$\begin{aligned} E[(X-K)^+] \approx\; & n[\kappa_1'(0)-k]\left[\Phi(-\sqrt{n}\,\hat{w}) - \frac{\phi(\sqrt{n}\,\hat{w})}{\sqrt{n}\,\hat{w}}\right] \\ & + \frac{\phi(\sqrt{n}\,\hat{w})}{\sqrt{n}}\left[\frac{1}{\hat{z}^2\sqrt{\kappa_1''(\hat{z})}} + \frac{\kappa_1'(0)-\kappa_1'(\hat{z})}{\hat{w}^3}\right] + O(n^{-3/2}). \end{aligned}$$

By dropping the error term $O(n^{-3/2})$, we retrieve the saddlepoint approximation formula C_4 given by (2.54).

Alternative saddlepoint approximations

The saddlepoint approximations developed earlier rely on the existence of unique real root of the saddlepoint equation: $\kappa'(z) = K$. One may be concerned with the potential failure of the methods due to the nonexistence of a real root to the saddlepoint equation within the domain of definition of the cgf. The occurrences of these scenarios are not uncommon in option pricing problems where the mgf is defined on the negative axis only.

In the sequel, the saddlepoint approximation methods associated with the solution of the saddlepoint equation: $\kappa'(z) = K$ are referred as the classical saddlepoint approximation methods. By observing the simple relation that $E[(X-K)\mathbf{1}_{\{X>K\}}] + E[(X-K)\mathbf{1}_{\{X\le K\}}] = \mu - K$ (analogous to the put-call parity in option pricing), we summarize the performance of the classical saddlepoint approximation methods in Table 2.3.

The potential failure of the classical saddlepoint approximation methods induces one to think beyond the original framework as characterized by the saddlepoint equation: $\kappa'(z) = K$. The saddlepoint approximation methods discussed earlier can

Table 2.3 Summary of the performance of the classical saddlepoint approximation methods for tail expectations. The domain of definition of the cgf is (α_-, α_+), where $\alpha_- < 0$ and $\alpha_+ > 0$

Saddlepoint $\hat{z}$	$E[(X-K)^+]$	$E[(K-X)^+]$
$\hat{z} \le \alpha_-$	Failure of the method	
$\alpha_- < \hat{z} \le 0$	By put-call parity	Direct approximation
$0 \le \hat{z} < \alpha_+$	Direct approximation	By put-call parity
$\hat{z} \ge \alpha_+$	Failure of the method	

be regarded as different variants of the steepest descent method being applied to the exponent $\kappa(z) - zK$. As an alternative approach for deriving the saddlepoint approximation formulas, Yang et al. (2006) and Zheng and Kwok (2014) apply the steepest descent method to different forms of the complex integrand. These give rise to various forms of the modified saddlepoint equation. The modified saddlepoint equation is more robust in the sense that it always admits a root even if the mgf is only defined on the negative axis.

Without loss of generality, we consider the saddlepoint approximation of tail expectation $E[X^+]$. This is a special case of (2.36), where $K = 0$, so

$$\Xi_1 = E[X^+] = E[X\mathbf{1}_{\{X>0\}}] = \frac{1}{2\pi i}\int_{\gamma-i\infty}^{\gamma+i\infty} \frac{e^{\kappa(z)}}{z^2}\,\mathrm{d}z, \quad \gamma \in (0, \alpha_+). \tag{2.61}$$

While the classical saddlepoint approximation methods consider the exponent in the numerator only, the alternative approach expresses the whole integrand in an exponential form. We write

$$\Xi_1 = \frac{1}{2\pi i}\int_{\gamma-i\infty}^{\gamma+i\infty} e^{\kappa(z)-2\log z}\,\mathrm{d}z, \quad \gamma \in (0, \alpha_+), \tag{2.62}$$

where the new exponent to be approximated is now $\kappa(z) - 2\log z$. This representation resembles closer to the standard formulation of the steepest descent method. Following the derivation in Sect. 1.2, we deform the contour of integration of integral (2.62) to become the steepest descent path from $\hat{z} - i\infty$ to $\hat{z} + i\infty$ that passes through the saddlepoint $\hat{z}$, which is obtained by solving the *modified* saddlepoint equation

$$\kappa'(z) - \frac{2}{z} = 0. \tag{2.63}$$

Recall that the new saddlepoint equation is derived by taking the first order derivative of the exponent of the integrand of (2.62) and setting it to be 0. It can be shown that the modified saddlepoint Eq. (2.63) always admits two real roots, one positive $\hat{z}_1$ and one negative $\hat{z}_2$ (Zheng and Kwok 2014). The modified saddlepoint approximation method fails only when both $\hat{z}_1 > \alpha_+$ and $\hat{z}_2 < \alpha_+$, so the chance of failure is less compared to that of the classical saddlepoint approximation method (see Table 2.3).

To proceed with the modified saddlepoint approximation for Ξ_1, we choose the positive saddlepoint $\hat{z}_1$ when $\hat{z}_1$ exists in $(0, \alpha_+)$, where

$$\Xi_1 = \frac{1}{2\pi i} \int_{\hat{z}_1 - i\infty}^{\hat{z}_1 + i\infty} e^{\kappa(z) - 2\log z}\, \mathrm{d}z. \tag{2.64}$$

By virtue of the steepest descent method, we perform the Taylor expansion of the exponent in the integrand around $\hat{z}_1$:

$$\kappa(z) - 2\log z \approx \kappa(\hat{z}_1) - 2\ln \hat{z}_1 + \left[\kappa''(\hat{z}_1) + \frac{2}{\hat{z}_1^2}\right] \frac{(z - \hat{z}_1)^2}{2}. \tag{2.65}$$

Substituting (2.65) into (2.64) and evaluating the resulting Gaussian integral, we obtain the first order saddlepoint approximation formula:

$$\begin{aligned}
\Xi_1 \approx \Xi_1^{(1)} &= \frac{1}{2\pi i} \int_{\hat{z}_1 - i\infty}^{\hat{z}_1 + i\infty} e^{\kappa(\hat{z}_1) - 2\ln \hat{z}_1 + \left[\kappa''(\hat{z}_1) + \frac{2}{\hat{z}_1^2}\right] \frac{(z - \hat{z}_1)^2}{2}}\, \mathrm{d}z \\
&= e^{\kappa(\hat{z}_1) - 2\ln \hat{z}_1} \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-\left[\kappa''(\hat{z}_1) + \frac{2}{\hat{z}_1^2}\right] \frac{y^2}{2}}\, \mathrm{d}y \\
&= \frac{e^{\kappa(\hat{z}_1)} / \hat{z}_1^2}{\sqrt{2\pi \left[\kappa''(\hat{z}_1) + \frac{2}{\hat{z}_1^2}\right]}}.
\end{aligned} \tag{2.66}$$

The same result can be obtained by applying the steepest descent formula (1.16) to the integral (2.62) directly.

In order to obtain the second order saddlepoint approximation formula, Zheng and Kwok (2014) perform the Taylor expansion of the exponent $\kappa(z) - 2\log z$ of the integrand in (2.64) around $\hat{z}_1$ up to the fourth order and perform further Taylor approximation of the resulting exponential function. The second order saddlepoint approximation for Ξ_1 can be obtained as follows:

$$\Xi_1 \approx \Xi_1^{(2)} = \Xi_1^{(1)}(1 + R_1), \tag{2.67}$$

where

$$R_1 = \frac{1}{8} \frac{\kappa''''(\hat{z}_1) + \frac{12}{\hat{z}_1^4}}{\left[\kappa''(\hat{z}_1) + \frac{2}{\hat{z}_1^2}\right]^2} - \frac{5}{24} \frac{\left[\kappa'''(\hat{z}_1) - \frac{4}{\hat{z}_1^3}\right]^2}{\left[\kappa''(\hat{z}_1) + \frac{2}{\hat{z}_1^2}\right]^3}.$$

The details of the derivation can be found in Zheng (2012), Appendix B.1. A similar derivation procedure can be found in the proof of (4.18) (see Appendix in Chap. 4).

For the left-tail expectation, one can start with a similar Laplace inversion integral representation as in (2.61):

$$\Xi_2 = E[(-X)^+] = -E[X\mathbf{1}_{\{X<0\}}] = \frac{1}{2\pi i}\int_{\gamma-i\infty}^{\gamma+i\infty}\frac{e^{\kappa(z)}}{z^2}\,\mathrm{d}z, \quad \gamma \in (\alpha_-, 0). \quad (2.68)$$

By choosing the negative saddlepoint $\hat{z}_2$, the first order and second order saddlepoint approximation formulas can be derived in a similar manner. The respective saddlepoint approximation formulas are given by

$$\Xi_2 \approx \Xi_2^{(1)} = \frac{e^{\kappa(\hat{z}_2)}/\hat{z}_2^2}{\sqrt{2\pi\left[\kappa''(\hat{z}_2)+\frac{2}{\hat{z}_2^2}\right]}}; \quad (2.69)$$

and

$$\Xi_2 \approx \Xi_2^{(2)} = \Xi_2^{(1)}(1+R_2), \quad (2.70)$$

where

$$R_2 = \frac{1}{8}\frac{\kappa''''(\hat{z}_2)+\frac{12}{\hat{z}_2^4}}{\left[\kappa''(\hat{z}_2)+\frac{2}{\hat{z}_2^2}\right]^2} - \frac{5}{24}\frac{\left[\kappa'''(\hat{z}_2)-\frac{4}{\hat{z}_2^3}\right]^2}{\left[\kappa''(\hat{z}_2)+\frac{2}{\hat{z}_2^2}\right]^3}.$$

The new approach provides an easier derivation procedure to extend the saddlepoint approximation formulas to higher orders. Since the derivation involves the Gaussian integral, so this new approach implicitly adopts the Gaussian base distribution. An application of formulas (2.69) and (2.70) in pricing options on discrete realized variance is discussed in Sect. 4.3.2. Lastly, we provide a summary of the performance of the modified saddlepoint approximations in Table 2.4. Note that the relation $\Xi_1 - \Xi_2 = \mu$ (put-call parity) is used.

Numerical examples

Using the same gamma distribution in (2.35) as the test case, we present numerical studies on the assessment of accuracy of the various saddlepoint approximation formulas for tail expectations derived by using the inverse Laplace transform representation. A comparison of the formulas is summarized in Table 2.5.

Table 2.4 Summary of the performance of the modified saddlepoint approximation methods for tail expectations

Saddlepoints $\hat{z}_1 > 0$ and $\hat{z}_2 < 0$	$\Xi_1 = \mathrm{E}[\mathrm{X}^+] = \mathrm{E}[\mathrm{X}1_{\{\mathrm{X}>0\}}]$	$\Xi_2 = \mathrm{E}[(-\mathrm{X})^+] = -\mathrm{E}[\mathrm{X}1_{\{\mathrm{X}<0\}}]$
$\hat{z}_1 \geq \alpha_+$ and $\hat{z}_2 \leq \alpha_-$	Failure of the method	
$\hat{z}_1 \geq \alpha_+$ and $\alpha_- < \hat{z}_2 < 0$	By put-call parity	By (2.69)/(2.70) directly
$0 < \hat{z}_1 < \alpha_+$ and $\hat{z}_2 \leq \alpha_-$	By (2.66)/(2.67) directly	By put-call parity
$0 < \hat{z}_1 < \alpha_+$ and $\alpha_- < \hat{z}_2 < 0$	By (2.66)/(2.67) directly or by put-call parity	By (2.69)/(2.70) directly or by put-call parity

Table 2.5 Comparison of the saddlepoint approximation formulas for tail expectation. The numbers inside the brackets in the column labeled "Highest order cumulant" represent the order of the highest order cumulant involved in the formulas under the degenerate case where $K = \mu$

Formula	Saddlepoint equation	Highest order cumulant	Removable singularity at μ
(2.41)	Classic	3	No
(2.45)	Classic	4 (4)	Yes
(2.54)	Classic	2 (4)	Yes
(2.59)	Classic	1 (2)	Yes
(2.66)	Modified	2	No
(2.67)	Modified	4	No

Note In order to achieve the same order of accuracy as the Lugannani-Rice formula, the density function in (2.45) has to be approximated by (2.1) with the inclusion of the fourth order cumulant

Table 2.6 Numerical approximation of tail expectations computed by various saddlepoint approximation formulas. Moneyness is defined to be the ratio of K to the mean μ. The means for Panel A and Panel B are $\mu_A = 2$ and $\mu_B = 5$, respectively. The sign (+) [(-)] indicates that the positive (negative) saddlepoint of (2.66)/(2.67) is used in the calculations

Moneyness	Panel A: $\alpha = 1, \beta = 2$			Panel B: $\alpha = 5, \beta = 1$		
	0.2	1	1.8	0.2	1	1.8
(2.41)	1.639141	0.797885	0.323893	4.000676	0.892062	0.082014
(2.45)	1.631106	0.731394	0.326873	4.000659	0.877194	0.083599
(2.54)	1.633749	0.731394	0.328540	4.000682	0.877194	0.083758
(2.59)	1.660701	0.797885	0.380649	4.000919	0.892062	0.088744
(2.66) (+)	1.251346	0.635889	0.307590	3.513619	0.796929	0.081220
(2.66) (-)	1.638508	0.755009	0.377806	4.000697	0.879373	0.089861
(2.67) (+)	1.527176	0.731721	0.334941	3.908128	0.878763	0.084414
(2.67) (-)	1.637444	0.735601	0.329297	4.000689	0.877677	0.067152
Exact	1.637462	0.735759	0.330598	4.000689	0.877337	0.083780

In Table 2.6, we list the numerical approximation results for the three different values of K under two different sets of distribution parameters of the gamma distribution. Since the modified saddlepoint Eq. (2.63) can admit two roots, which leads to two choices of the approximation formula of the tail expectation, we would like to discuss the general rule of choosing either one saddlepoint via some numerical tests. From Table 2.6, Antonov et al.'s approximation formula (2.41) is seen to be fairly good for the in-the-money (moneyness equals 0.2) and out-of-the-money (moneyness equals 1.8) cases, but it is unable to provide satisfactory approximation for the at-the-money (moneyness equals 1) case. On the other hand, Martin's approximation formula (2.45) seems to perform well for varying levels of moneyness. While the first order approximation C_3 given by (2.59) is not quite accurate in approximating tail expectations, the second order approximation C_4 given by (2.54) does show good improvement on accuracy. Indeed, C_4 has the best performance among the classical saddlepoint approximation methods.

For the alternative saddlepoint approximation formulas, we observe that the choice of the saddlepoint does have an impact on accuracy of the approximation. From the numerical results in Table 2.6, it seems that the negative saddlepoint is a better choice for small K, whereas the positive saddlepoint works better for large K. This may be related to the magnitude of the saddlepoint. Recall that the first order approximation exhibits the asymptotic order:

$$\frac{e^{\kappa(\hat{z}_1)}/\hat{z}_1^2}{\sqrt{2\pi\left[\kappa''(\hat{z}_1)+\frac{2}{\hat{z}_1^2}\right]}} \approx O(|\hat{z}_1|^{-1}),$$

when $\hat{z}_1$ is close to zero. It is expected that both (2.66) and (2.67) are bound to exhibit numerical instabilities when the saddlepoint has a small magnitude. The next challenge is how to switch optimally from the choice of one saddlepoint to the other in order to achieve better numerical performance. Recall that when there are multiple saddlepoints, the useful rule of thumb in the deepest descent method is to use the saddlepoint with the largest modulus. In our numerical tests, we adopt the ad hoc switching rule of choosing the saddlepoint with a larger magnitude, which seems to work well in most cases. Figure 2.2 demonstrates how this ad hoc rule helps improve the overall numerical performance of (2.66) and (2.67).

Finally, we perform a comparison of percentage errors of all the saddlepoint approximation formulas presented in Table 2.6, where the ad hoc rule for optimal switching is adopted for (2.66) and (2.67). The numerical tests indicate that Martin's formula (2.45) and Huang and Oosterlees' second order approximation formula (2.54) are favorable choices as they exhibit high degree of reliable performance. Zheng and Kwok's second order approximation formula is also quite accurate, provided that an appropriate saddlepoint is used in the approximation.

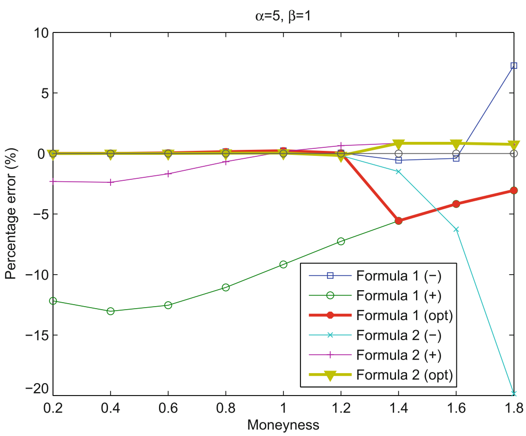
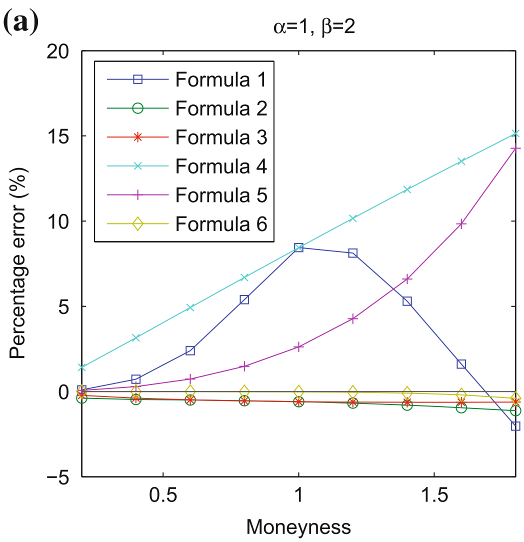


Fig. 2.2 Plots of the percentage errors of Zheng and Kwok's saddlepoint approximation formulas. "Formula 1" stands for (2.66) and "Formula 2" stands for (2.67). The symbols (+), (-) are interpreted in the same manner as in Table 2.6. The notation (opt) stands for the ad hoc switching rule of choosing the saddlepoint according to the magnitude of the saddlepoint



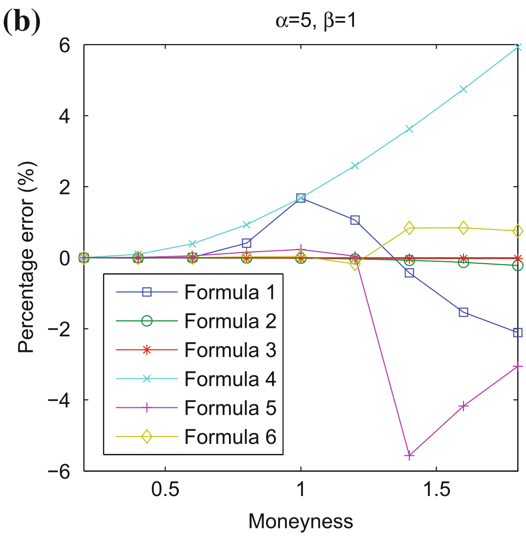


Fig. 2.3 Plots of the percentage errors of various saddlepoint approximation formulas. “Formula 1” stands for (2.41), “Formula 2” stands for (2.45), “Formula 3” stands for (2.54), and “Formula 4” stands for (2.59), “Formula 5” stands for (2.66), and “Formula 6” stands for (2.67). Saddlepoint approximation formulas (2.45) and (2.54), both considered as variants of the Lugganni-Rice formula, give the best numerical performance

As a final conclusion, formula (2.21) remains to be the best performer, followed by (2.26) (or equivalently 2.54) and Martin’s formula (2.45). All these saddlepoint approximation formulas for tail expectations are seen as variants of the Lugannani-Rice formula (Fig. 2.3).

Appendix

Exponential Tilting Technique

Given a random variable X with density function $p_X(x)$ and moment generating function (mgf) $M_X(z) = E[e^{zX}]$ under the probability measure P, the exponentially tilted measure P_θ is defined by

$$P_\theta[X \in \mathrm{d}x] = \frac{E[e^{\theta X}\mathbf{1}_{\{X \in \mathrm{d}x\}}]}{M_X(\theta)} = e^{\theta x - \kappa_X(\theta)} P[X \in \mathrm{d}x],$$

where the cumulant generating function (cgf) is $\kappa_X(\theta) = \log E[e^{\theta X}] = \log M_X(\theta)$. Consequently, the corresponding density function $p_X(x;\theta)$ under P_θ is given by

$$p_X(x;\theta) = e^{\theta x - \kappa_X(\theta)} p_X(x).$$

For example, suppose X is the normal distribution $N(\mu, \sigma^2)$ with mean μ and variance σ^2 under P, then X under P_θ after exponential tilting becomes $N(\mu + \theta\sigma^2, \sigma^2)$. Note that the mgf and cgf of the exponentially tilted X are given by $\dfrac{M_X(x+\theta)}{M_X(\theta)}$ and $\kappa_X(x+\theta) - \kappa_X(\theta)$, respectively. The mean and variance of the exponentially tilted X are given by the first order and second order derivative of the corresponding cgf evaluated at $x = 0$, and they are found to be $\kappa_X'(\theta)$ and $\kappa_X''(\theta)$, respectively.

Edgeworth Expansion

Let M_X and M_A be the respective mgf of the random variable X whose density function is to be approximated by the distribution of another random variable A. We consider the respective power series expansion of the cgf as follows:

$$\log M_X(t) = \sum_{j=1}^{n} \kappa_j(X) \frac{(it)^j}{j!} + o(t^n)$$

$$\log M_A(t) = \sum_{j=1}^{n} \kappa_j(A) \frac{(it)^j}{j!} + o(t^n),$$

where $\kappa_j(X)$ and $\kappa_j(A)$ are the respective jth order cumulant of X and A. Formally, we obtain

$$M_X(t) = \exp\left(\sum_{j=1}^{n} \frac{\kappa_j(X) - \kappa_j(A)}{j!}(it)^j\right) M_A(t) + o(t^n).$$

Recall from the properties of Fourier transform that $(it)^j M_A(t)$ is the Fourier transform of $(-1)^j \dfrac{\mathrm{d}^j}{\mathrm{d}t^j} p_A(t)$, where $p_A(t)$ is the density function of the random variable A. Taking the inversion of Fourier transform, the density function of X is approximated by

$$p_X(t) = \exp\left(\sum_{j=1}^{n} \frac{\kappa_j(X) - \kappa_j(A)}{j!}(-1)^j \frac{\mathrm{d}^j}{\mathrm{d}t^j} p_A(t)\right) + o(t^n).$$

Suppose the approximating density function $p_A(t)$ is chosen to be the normal density with mean and variance set equal to those of X. Recall that the mean and variance of X are $\mu = \kappa_1(X)$ and $\sigma^2 = \kappa_2(X)$, respectively. By setting $\kappa_1(X) = \kappa_1(A)$ and $\kappa_2(X) = \kappa_2(A)$, and performing the straightforward differentiation of the normal density function to higher orders, the above approximation formula reduces to the so-called Edgeworth expansion as follows:

$$\begin{aligned} p_X(t) \approx \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(t-\mu)^2}{2\sigma^2}\right) \Bigg\{ 1 &+ \frac{\kappa_3(X)}{3!\sigma^3} H_3\left(\frac{t-\mu}{\sigma}\right) + \frac{\kappa_4(X)}{4!\sigma^4} H_4\left(\frac{t-\mu}{\sigma}\right) \\ &+ \frac{1}{2}\left[\frac{\kappa_3(X)}{3!\sigma^3}\right]^2 H_6\left(\frac{t-\mu}{6}\right)\Bigg\} + o(t^6), \end{aligned}$$

where the Hermite polynomials are defined by

$$H_j(t) = (-1)^j \frac{\frac{\mathrm{d}^j}{\mathrm{d}t^j}\phi(t)}{\phi(t)}, \quad j = 3, 4, \ldots.$$

Here, $\phi(t)$ is the standard normal density function. The first few members of the Hermite polynomials are found to be

$$H_3(t) = t^3 - 3t,\ H_4(t) = t^4 - 6t^2 + 3 \text{ and } H_6(t) = t^6 - 15t^4 + 45t^2 - 15.$$

As a remark, the term with the quadratic power of $\dfrac{\kappa_3(X)}{3!\sigma^3}$ in the Edgeworth expansion arises from the use of the power series expansion: $e^x \approx 1 + x + \dfrac{x^2}{2}$, where x stands for $\dfrac{\kappa_3(X)}{3!\sigma^3}$.

Suppose $\{X_i\}$ is a sequence of independent and identically distributed random variables with common μ and σ^2, and let S_n be their standardized sum

$$S_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{X_i - \mu}{\sigma}.$$

Let F_n denote the cumulative distribution function of S_n. The Edgeworth expansion of F_n can be deduced to be

$$F_n(x) = \Phi(x) - \frac{\lambda_3}{6\sqrt{n}}\Phi^{(3)}(x) + \frac{1}{n}\left[\frac{\lambda_4}{24}\Phi^{(4)}(x) + \frac{\lambda_3^2}{72}\Phi^{(6)}(x)\right] + O\left(\frac{1}{n^{3/2}}\right),$$

where $\Phi(x)$ is the cumulative function of the standard normal variable defined by

$$\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-t^2/2}\,\mathrm{d}t,$$

$\Phi^{(n)}(x)$ is the nth order derivative of $\Phi(x)$ and $\lambda_j = \kappa_j/\sigma^j, j = 3, 4$. The above Edgeworth expansion formula can be considered as a refinement to the result from the central limit theorem stated as

$$\lim_{n\to\infty} F_n(x) = \Phi(x).$$

Proof of the Degenerate Case in (2.26)

A convenient approach of deriving the approximation formula for the degenerate case is via the power expansion of $\frac{1}{\hat{w}_\theta}$ and $x - \kappa'(\theta)$ and other composite functions as a function of z around θ up to an appropriate order. For notational convenience, we replace $\hat{z}_x$ by z in the sequel and assume $z \to \theta^+$, so $\operatorname{sgn}(z - \theta^+) = 1$. First, we consider the power series expansion of $\hat{w}_\theta$ in powers of $z - \theta$ as follows:

$$\hat{w}_\theta = \sqrt{\kappa''(\theta)}(z-\theta)\left[1 + \frac{2}{3}\frac{\kappa'''(\theta)}{\kappa''(\theta)}(z-\theta) + \frac{1}{4}\frac{\kappa''''(\theta)}{\kappa''(\theta)}(z-\theta)^2 + O((z-\theta)^3)\right]^{\frac{1}{2}}.$$

The power series expansion of $1/w_\theta^3$ can be expressed as

$$\frac{1}{\hat{w}_\theta^3} = \frac{1}{\kappa''(\theta)^{\frac{3}{2}}(z-\theta)^3}\left[1 + \frac{2}{3}\frac{\kappa'''(\theta)}{\kappa''(\theta)}(z-\theta) + \frac{1}{4}\frac{\kappa''''(\theta)}{\kappa''(\theta)}(z-\theta)^2 + O((z-\theta)^3)\right]^{-\frac{3}{2}}.$$

To improve analytic tractability, we further expand the terms inside the square root bracket into a power series of the form:

$$1 + b(z-\theta) + c(z-\theta)^2 + O((z-\theta)^3).$$

Using the binomial expansion and after some tedious calculations, we obtain

$$b = -\frac{\kappa'''(\theta)}{\kappa''(\theta)}, \quad c = \frac{5}{6}\frac{\kappa'''(\theta)^2}{\kappa''(\theta)^2} - \frac{3}{8}\frac{\kappa''''(\theta)}{\kappa''(\theta)}.$$

Recall $x = \kappa'(z)$ and we expand $\kappa'(z)$ at $z = \theta$ in power series expansion of $z - \theta$ as follows:

$$x - \kappa'(\theta) = \kappa''(\theta)(z-\theta) + \frac{\kappa'''(\theta)}{2}(z-\theta)^2 + \frac{\kappa''''(\theta)}{6}(z-\theta)^3 + O((z-\theta)^4).$$

Putting these power series expansions together, we obtain

$$\begin{aligned} &\frac{x - \kappa'(\theta)}{\hat{w}_\theta^3} \\ &= \frac{1 + b(z-\theta) + c(z-\theta)^2 + O((z-\theta)^3)}{\kappa''(\theta)^{\frac{3}{2}}(z-\theta)^2} \\ &\quad \left[\kappa''(\theta) + \frac{\kappa'''(\theta)}{2}(z-\theta) + \frac{\kappa''''(\theta)}{6}(z-\theta)^2 + O((z-\theta)^3)\right] \\ &= \frac{1}{\kappa''(\theta)^{1/2}(z-\theta)^2} - \frac{\kappa'''(\theta)}{2\kappa''(\theta)^{\frac{3}{2}}(z-\theta)} + \frac{1}{3}\frac{\kappa'''(\theta)^2}{\kappa''(\theta)^{\frac{5}{2}}} - \frac{5}{24}\frac{\kappa''''(\theta)}{\kappa''(\theta)^{\frac{3}{2}}} + O(z-\theta). \end{aligned}$$

Apparently, there are two singular terms in negative powers of $z - \theta$ involved in the above expansion. Fortunately, these two negative power terms are seen to be canceled by the other singular terms in the expansion of $\frac{1}{(z-\theta)\hat{u}_\theta}$, which admits the following expansion:

$$\begin{aligned} &\frac{1}{(z-\theta)\hat{u}_\theta} \\ &= \frac{1}{\kappa''(\theta)^{1/2}(z-\theta)^2} - \frac{\kappa'''(\theta)}{2\kappa''(\theta)^{\frac{3}{2}}(z-\theta)} + \frac{3}{8}\frac{\kappa'''(\theta)^2}{\kappa''(\theta)^{\frac{5}{2}}} - \frac{\kappa''''(\theta)}{4\kappa''(\theta)^{\frac{3}{2}}} + O(z-\theta). \end{aligned}$$

By putting all the results together and noting that

$$\frac{x - \kappa'(\theta)}{\hat{u}_\theta} \to \sqrt{\kappa''(\theta)}, \quad \phi(w_\theta) \to \frac{1}{\sqrt{2\pi}},$$

we obtain the formula for the degenerate case in (2.26).

Proof of Lemma 2.1

Observing that $\frac{\lambda}{2}(z-z_0)^2 = \frac{1}{2}(\sqrt{\lambda}z)^2 - \lambda z_0 z + \frac{\lambda}{2}z_0^2$, so

$$\begin{aligned} J_0(\lambda, z_0) &= e^{\frac{\lambda}{2}z_0^2}\frac{1}{2\pi i}\int_{\tau-i\infty}^{\tau+i\infty} e^{\frac{1}{2}(\sqrt{\lambda}z)^2-\lambda z_0 z}\,\mathrm{d}z \\ &= \frac{e^{\frac{\lambda}{2}z_0^2}}{\sqrt{\lambda}}\frac{1}{2\pi i}\int_{\sqrt{\lambda}\tau-i\infty}^{\sqrt{\lambda}\tau+i\infty} e^{\frac{y^2}{2}-\sqrt{\lambda}z_0 y}\,\mathrm{d}y \\ &= \frac{e^{\frac{\lambda}{2}z_0^2}}{\sqrt{\lambda}}\phi(\sqrt{\lambda}z_0) = \frac{1}{\sqrt{2\pi\lambda}}, \end{aligned}$$

where we have used the Laplace inversion formula for the standard normal density function. Similarly, we have

$$J_{-1}(\lambda, z_0) = \frac{e^{\frac{\lambda}{2}z_0^2}}{\lambda}\frac{1}{2\pi i}\int_{\sqrt{\lambda}\tau-i\infty}^{\sqrt{\lambda}\tau+i\infty} e^{\frac{y^2}{2}-\sqrt{\lambda}z_0 y}y\,\mathrm{d}y = -\frac{e^{\frac{\lambda}{2}z_0^2}}{\lambda}\phi'(\sqrt{\lambda}z_0).$$

As a remark, the results for J_0 and J_{-1} do not actually rely on the condition $\tau \neq 0$, as $z = 0$ is not a pole for both integrands. In an analogous manner, we obtain

$$J_1(\lambda, z_0) = e^{\frac{\lambda}{2}z_0^2}\frac{1}{2\pi i}\int_{\sqrt{\lambda}\tau-i\infty}^{\sqrt{\lambda}\tau+i\infty} \frac{e^{\frac{y^2}{2}-\sqrt{\lambda}z_0 y}}{y}\,\mathrm{d}y.$$

Let Z be a standard normal random variable. By using the Laplace inversion formula for the tail probability of Z, we find that

$$\frac{1}{2\pi i}\int_{\sqrt{\lambda}\tau-i\infty}^{\sqrt{\lambda}\tau+i\infty} \frac{e^{\frac{y^2}{2}-\sqrt{\lambda}z_0 y}}{y}\,\mathrm{d}y = \begin{cases} P[Z > \sqrt{\lambda}z_0], & \text{if } \tau > 0 \\ -P[Z \le \sqrt{\lambda}z_0], & \text{if } \tau < 0. \end{cases}$$

We obtain

$$\begin{aligned} J_1(\lambda, z_0) &= \mathrm{sgn}(\tau)e^{\frac{\lambda}{2}z_0^2}\Phi(-\mathrm{sgn}(\tau)\sqrt{\lambda}z_0) \\ J_2(\lambda, z_0) &= \sqrt{\lambda}e^{\frac{\lambda}{2}z_0^2}\frac{1}{2\pi i}\int_{\sqrt{\lambda}\tau-i\infty}^{\sqrt{\lambda}\tau+i\infty} \frac{e^{\frac{y^2}{2}-\sqrt{\lambda}z_0 y}}{y^2}\mathrm{d}y \\ &= \begin{cases} \sqrt{\lambda}e^{\frac{\lambda}{2}z_0^2}E[(Z-\sqrt{\lambda}z_0)^+], & \text{if } \tau > 0 \\ \sqrt{\lambda}e^{\frac{\lambda}{2}z_0^2}E[(\sqrt{\lambda}z_0 - Z)^+], & \text{if } \tau < 0. \end{cases} \end{aligned}$$

Lastly, evaluation of the expectations yields the formula for $J_2(\lambda, z_0)$.

Chapter 3
Extended Saddlepoint Approximation Methods

In the previous chapter, we assume that the cumulant generating function (cgf) of the underlying random variable X is known in closed form. The saddlepoint equation involves the first order derivative of the cgf and one can solve for the saddlepoint by a root finding algorithm. However, availability of analytic closed form of the cgf is limited to a small class of random processes. The effective implementation of the saddlepoint approximation becomes challenging when the cgf of the underlying process does not admit closed form formula. One has to resort to computing numerical approximation of the cgf as part of the procedure for various saddlepoint approximation methods.

In this chapter, we consider how to obtain the saddlepoint approximation formulas for density functions of continuous time pure diffusion and jump-diffusion processes under small time expansion. We also show how to determine the saddlepoint effectively under the affine jump-diffusion process without direct solution of the saddlepoint equation. We focus on the two research papers that discuss saddlepoint approximation methods under nonavailability of closed form formula of the cgf. For general continuous time Markov processes, including the pure diffusion and jump-diffusion processes, Aït-Sahalia and Yu (2006) use the Taylor expansion in small time (time interval of the transition density of time-homogeneous process) to obtain approximations of the cgf and saddlepoints. They also perform a detailed error analysis of their approximations with respect to the small time expansion. Glasserman and Kim (2009) consider the affine jump-diffusion models under which the cgf of the state variable admits an exponential affine form and the parameter functions in the exponent are completely determined by solving a Riccati system of ordinary differential equations (ODEs). The Riccati system is explicitly solvable for only a small subclass of the affine family, such as the Heston stochastic volatility models. To implement the saddlepoint approximation under the general affine models, Glasserman and Kim (2009) develop a procedure of finding saddlepoint approximations by

Y. K. Kwok and W. Zheng, *Saddlepoint Approximation Methods in Financial Engineering*, SpringerBriefs in Quantitative Finance, https://doi.org/10.1007/978-3-319-74101-7_3

numerically solving the Riccati system of ODEs and approximating the saddlepoints using the series inversion technique (Lieberman 1994).

3.1 Small Time Expansion

Aït-Sahalia and Yu (2006) consider the saddlepoint approximation formulas for density functions of continuous time Markov processes under pure diffusion and jump-diffusion by performing an expansion of the moment generating function (mgf) in small time.

3.1.1 Pure Diffusion Processes

We consider the pure diffusion process governed by the following stochastic differential equation

$$\mathrm{d}X_t = \mu(X_t)\mathrm{d}t + \sigma(X_t)\mathrm{d}W_t. \tag{3.1}$$

Define the infinitesimal generator $\mathscr{A}$ of the diffusion process as follows:

$$\mathscr{A} f = \frac{\partial f}{\partial \delta} + \mu(y)\frac{\partial f}{\partial y} + \frac{\sigma^2(y)}{2}\frac{\partial^2 f}{\partial y^2}, \tag{3.2}$$

where the function $f(\delta, y, x)$ is assumed to be continuously differentiable once in δ and twice in y. Under some technical conditions [see Sect. 4 of Aït-Sahalia and Yu (2006)], the Taylor expansion with respect to the small time increment Δ is given by

$$E[f(\Delta, Y, X)|X = x] = \sum_{k=0}^{N} \frac{\Delta^k}{k!}\mathscr{A}^k f(0, x, x) + O(\Delta^{N+1}). \tag{3.3}$$

This is the key result in the approximation of the mgf of X, so does its transition density function and other related quantities.

In the sequel, we let $f_{[n]}$ be the small time approximation at order n in Δ for any function f that admits an expansion in the form of (3.3). Suppose we take $f(\delta, y, x) = e^{uy}$, then the mgf of X_Δ is given by

$$M(\Delta, u|x) = E[e^{uX_\Delta}|X_0 = x] = E[f(\Delta, X_\Delta, X_0)|X_0 = x].$$

Applying (3.3) with $N = 1$ to $M(\Delta, u|x)$ yields the first order approximation

$$M_{[1]}(\Delta, u|x) = e^{ux}\left\{1 + [\mu(x)u + \frac{\sigma^2(x)}{2}u^2]\Delta\right\},$$

with an error of order $O(\Delta^2)$. The corresponding first order approximation of the cgf is found to be

$$\kappa_{[1]}(\Delta, u|x) = \log M_{[1]}(\Delta, u|x) = ux + \left[\mu(x)u + \frac{\sigma^2(x)}{2}u^2\right]\Delta. \tag{3.4}$$

Using the above approximation of the cgf, it is straightforward to obtain the saddlepoint $\hat{u}_{[1]}$ to the first order accuracy by solving

$$\frac{\partial \kappa_{[1]}(\Delta, u|x)}{\partial u} = y.$$

This gives

$$\hat{u}_{[1]}(\Delta, y|x) = \frac{y - [x + \mu(x)\Delta]}{\sigma^2(x)\Delta}.$$

We consider the increment in units of $\Delta^{1/2}$ since the differential of a Brownian motion that drives X is of order $\Delta^{1/2}$. When evaluated at $y = x + z\Delta^{1/2}$, the first order approximation of the saddlepoint is given by

$$\hat{u}_{[1]}(\Delta, x + z\Delta^{1/2}|x) = \frac{z}{\sigma^2(x)\Delta^{1/2}} + O(1). \tag{3.5}$$

A second order expansion of the cgf can be obtained by expanding the square term in (3.3). It can be shown that

$$\begin{aligned}\kappa_{[2]}(\Delta, u|x) = ux &+ \left[\mu(x)u + \frac{\sigma^2(x)}{2}u^2\right]\Delta + \frac{u}{8}\Big\{4\mu(x)\frac{\mathrm{d}\mu(x)}{\mathrm{d}x} + 2\sigma^2(x)\frac{\mathrm{d}^2\mu(x)}{\mathrm{d}x^2} \\ &+ u[4\sigma^2(x)\frac{\mathrm{d}\mu(x)}{\mathrm{d}x} + 2\mu(x)\frac{\mathrm{d}\sigma^2(x)}{\mathrm{d}x} + \sigma^2(x)\frac{\mathrm{d}^2\sigma^2(x)}{\mathrm{d}x^2}] \\ &+ 2u^2\sigma^2(x)\frac{\mathrm{d}\sigma^2(x)}{\mathrm{d}x}\Big\}\Delta^2,\end{aligned} \tag{3.6}$$

with error of $O(\Delta^3)$. In this case, the saddlepoint equation

$$\frac{\partial \kappa_{[2]}(\Delta, u|x)}{\partial u} = y$$

is a quadratic equation. When we take $y = x + z\Delta^{1/2}$ and substitute a solution of the form

$$\hat{u} = \frac{u_{-1/2}}{\Delta^{1/2}} + u_0 + O(\Delta^{1/2})$$

with undetermined parameters $u_{-1/2}$ and u_0 into the above saddlepoint equation, we obtain

$$\hat{u}_{[2]}(\Delta, y + z\Delta^{1/2}|x) = \frac{z}{\sigma^2(x)\Delta^{1/2}} - \left[\frac{\mu(x)}{\sigma^2(x)} + \frac{3\frac{\mathrm{d}\sigma^2(x)}{\mathrm{d}x}}{4\sigma^4(x)}z^2\right] + O(\Delta^{1/2}). \quad (3.7)$$

Consistency of the expansion is verified since the first term of $\hat{u}_{[2]}$ is the leading term of $\hat{u}_{[1]}$.

With availability of the small time expansion of the cgf and the saddlepoint, the saddlepoint approximation of the transition density function can be obtained at relative ease. For convenience, we use the notation $p_{[n_1,n_2]}$ to indicate the saddlepoint approximation of the transition density function of order n_1 using a Taylor expansion in Δ of $M(\Delta, u|x)$, with order of accuracy n_2 in Δ. Under certain regularity conditions on the diffusion coefficients $\mu(x)$ and $\sigma^2(x)$, Aït-Sahalia and Yu (2006) obtain the following theorem on the saddlepoint approximation of the transition density function.

Theorem 3.1 (Aït-Sahalia and Yu 2006) *The leading term of the saddlepoint approximation of the transition density function at the first order in Δ and with a Gaussian base coincides with the classical Euler approximation of the transition density, that is,*

$$p_{[0,1]}(\Delta, u|x) = \frac{1}{\sqrt{2\pi\Delta\sigma^2(x)}} \exp\left(-\frac{[y - x - \mu(x)\Delta]^2}{\sigma^2(x)\Delta}\right). \quad (3.8)$$

The first order saddlepoint approximation at the first order in Δ with a Gaussian base is given by

$$\begin{aligned} &p_{[1,1]}(\Delta, x + z\Delta^{1/2}|x) \\ &= \frac{\exp(-\frac{z^2}{2\sigma^2(x)} + e_{1/2}(z|x)\Delta^{1/2} + e_1(z|x)\Delta)}{\sqrt{2\pi}\sigma(x)\Delta^{1/2}[1 + d_{1/2}(z|x)\Delta^{1/2} + d_1(z|x)\Delta]}[1 + c_1(z|x)\Delta], \end{aligned} \quad (3.9)$$

where

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

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

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

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

$$c_1(z|x) = \frac{\frac{d^2\sigma^2(x)}{dx^2}}{4} - \frac{3}{32\sigma^2(x)}\left[\frac{d\sigma^2(x)}{dx}\right]^2.$$

For $n_1 \geq 0$ and $n_2 \geq 1$, we have

$$p(\Delta, x + z\Delta^{1/2}|x) = p_{[n_1,n_2]}(\Delta, x + z\Delta^{1/2})[1 + O(\Delta^{\min(n_1+1,n_2/2)})].$$

Specifically, if $\kappa(\Delta, u|X_0)$ is analytic at $\Delta = 0$, then for $n_1 \geq 0$, we have

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

The proof of the above results can be found in Aït-Sahalia and Yu (2006). No numerical solution of the saddlepoint equation for the saddlepoint is required in the above formula since it is "ready-to-use" approximation that depends only on the coefficient functions of the underlying diffusion process.

Extensive numerical examples are given in Aït-Sahalia and Yu (2006) that test accuracy of the proposed saddlepoint approximation formulas for the transition density functions. In particular, they tested the Heston stochastic volatility model for which Δ is taken to be 0.1. Very small log-density error is observed even under the Euler approximation and first order approximation. The error almost becomes invisible for the second order approximation [see Fig. 4 in Aït-Sahalia and Yu (2006)]. Note that the mgf of X_t under the Heston stochastic volatility model is available in closed form, so it is not necessary to introduce any small time expansion approximation in this case.

As a remark, the above saddlepoint approximation formulas with a Gaussian base perform well for pure diffusion processes. However, it becomes a challenging task when jump processes are included. In fact, the choice of an appropriate non-Gaussian base distribution becomes essential when we consider saddlepoint approximation of the transition density of a jump-diffusion process. The details of which are discussed in the following subsection.

3.1.2 Jump-Diffusion Processes

The small time expansion technique discussed above can be extended to derive the saddlepoint approximation of the transition density of a jump-diffusion process. Suppose the random process X_t is governed by the stochastic differential equation

$$\mathrm{d}X_t = \mu(X_t)\mathrm{d}t + \sigma(X_t)\mathrm{d}W_t + J_t\,\mathrm{d}N_t, \tag{3.10}$$

where N_t represents an independent Poisson process with constant arrival rate λ and J_t is the random jump size with probability density $\nu(\cdot)$. As commonly adopted in the financial engineering literature, J_t is assumed to be independent of W_t and N_t. We let $M_J(u) = E[e^{uJ_t}]$.

Analogous to Theorem 3.1, we can apply the small time expansion technique to obtain the saddlepoint approximation formulas. Note that the infinitesimal generator

$\mathscr{A}$ of the jump-diffusion process is given by

$$\mathscr{A} f = \frac{\partial f}{\partial \delta} + \mu(y)\frac{\partial f}{\partial y} + \frac{\sigma^2(y)}{2}\frac{\partial^2 f}{\partial y^2} + \lambda E[f(\delta, y + J, x) - f(\delta, y, x)]. \quad (3.11)$$

Saddlepoint approximation with a Gaussian base

Following a similar approach as in Theorem 3.1, we obtain the following theorem on the saddlepoint approximation of the transition density function for the jump-diffusion process.

Theorem 3.2 (Aït-Sahalia and Yu 2006) *The leading term of the saddlepoint approximation of the transition density function at the first order in Δ with a Gaussian base is given by*

$$p_{[0,1]}(\Delta, y|x) = \frac{\exp(-\frac{\sigma^2(x)}{2}\hat{u}_{[1]}^2\Delta)\exp(\lambda[M_J(\hat{u}_{[1]}) - \hat{u}_{[1]}M_J'(\hat{u}_{[1]}) - 1]\Delta)}{\{2\pi[\sigma^2(x)\Delta + \Delta\lambda M_J''(\hat{u}_{[1]})]\}^{1/2}}, \quad (3.12)$$

where $\hat{u}_{[1]}$ is the saddlepoint that satisfies the following equation:

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

To address the potential failure of the saddlepoint approximation with a Gaussian base in the presence of jumps, Aït-Sahalia and Yu (2006) provide a simple example in which they test the saddlepoint approximation of the transition density of a jump-diffusion process defined by

$$\mathrm{d}X_t = \sigma\, \mathrm{d}W_t + J_t\, \mathrm{d}N_t, \quad (3.13)$$

with $X_0 = 0$ and where the jump J has exponential distribution $f_J(y) = \gamma e^{-\gamma y}$. The Gaussian saddlepoint approximation has the leading term

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

whose corresponding cgf is

$$\kappa_0(u) = ux + \frac{\sigma^2 u^2}{2}\Delta.$$

The saddlepoint approximation of the transition density admits [see also (3.15)]

$$p_{[0]}(\Delta, y|x) = f_0(y)\exp(\kappa(\Delta, \hat{u}|x) - \hat{u}y - \kappa_0(\hat{w}) + \hat{w}y) \\ \kappa_0''(\hat{w})^{1/2}\left[\frac{\partial^2\kappa(\Delta, \hat{u}|x)}{\partial u^2}\right]^{-1/2}. \quad (3.14)$$

To evaluate $p_{[0]}$, one can use the following expansion for $\hat{u}$ and $\hat{w}$:

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

from which one can further deduce that

$$\kappa(\Delta, \hat{u}|x) - \hat{u}y - \kappa_0(\hat{w}) + \hat{w}y = \frac{y^2}{2\sigma^2\Delta} - y\gamma + O(\sqrt{\Delta}),$$
$$\kappa_0''(\hat{w})^{1/2}\left[\frac{\partial^2\kappa(\Delta, \hat{u}|x)}{\partial u^2}\right]^{-1/2} = \frac{1}{\sqrt{2}}\left(\frac{\gamma\lambda}{y^3}\right)^{1/4}\sigma\Delta^{3/4} + o(\Delta^{3/4}).$$

Using these expressions, we are able to obtain an expansion of $p_{[0]}$ at order $O(\Delta^{3/4})$.

It is obvious that using $f_0(y)$ alone leads to a very poor approximation since $f_0(y)$ decays exponentially in $1/\Delta$ at the tails, whereas the jump density is of order $O(\Delta)$. Also, $p_{[0]}$ involves a correction term which can be shown to cancel the exponential factor of $f_0(y)$. Given a reasonable choice of the model parameters and $\Delta = 1/52$, numerical tests show that $p_{[0]}$ performs significantly better than $f_0(y)$. However, at the tails it remains at order $O(\Delta^{1/4})$, which is still inadequate to reflect the true behavior of the density function. It indicates that the Gaussian base may not be the appropriate choice. We need to consider a non-Gaussian base distribution that captures the jump feature better.

Saddlepoint approximation with a non-Gaussian base distribution

The potential failure of the saddlepoint approximation with a Gaussian base leads one to explore a non-Gaussian base distribution that would better reflect the jump component. We refer to Proposition 2.3 for a review of the saddlepoint approximation with a non-Gaussian base distribution. To maintain analytic tractability, we assume that the base distribution has a closed form mgf:

$$\exp(\kappa_0(\Delta, u|x)) = \exp(ux)\left[(1-\lambda\Delta)\exp\left(\mu(x)u\Delta + \frac{\sigma^2(x)u^2}{2}\Delta\right) + \lambda\Delta M_J(u)\right].$$

The above equation approximates the true mgf of the underlying process up to the first order in Δ. This is easily seen by noting that

$$\begin{aligned}
&\exp(\kappa_0(\Delta, u|x)) \\
&= \exp(ux)\left\{(1-\lambda\Delta)\left[1 + \mu(x)u\Delta + \frac{\sigma^2(x)u^2}{2}\Delta + O(\Delta^2)\right] + \lambda\Delta M_J(u)\right\}, \\
&= \exp(ux)\left\{1 + \left[\mu(x)u + \frac{\sigma^2(x)u^2}{2} - \lambda\right]\Delta + \lambda\Delta M_J(u)\right\} + O(\Delta^2).
\end{aligned}$$

The leading term is recognized as the first order approximation of $M(\Delta, u|x)$ according to the small time expansion in (3.3). Statistically, this choice of mgf corresponds to the base distribution of a process that can have at most one jump in a time interval of length Δ and the jump probability is $\lambda\Delta$. The density function $f_0(y)$ corresponding to $\kappa_0(\Delta, u|x)$ is

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

where $\phi(y; \mu, \sigma^2)$ is the normal density function with mean μ and variance σ^2. In the sequel, we let $p_{[n]}$ be the saddlepoint approximation with the non-Gaussian base κ_0 using a small time expansion of the mgf at order n in Δ. The expression of $p_{[n]}$ is given by the following theorem.

Theorem 3.3 (Aït-Sahalia and Yu 2006) *The saddlepoint approximation of the transition density function with the non-Gaussian base κ_0 and with the true cgf replaced by its nth order aproximation $\kappa_{[n]}$ is given by*

$$\begin{aligned} p_{[n]}(\Delta, y|x) = f_0(y) \exp(\kappa_{[n]}(\Delta, \hat{u}|x) - \hat{u}y) \\ - [\kappa_0(\hat{w}) + \hat{w}y]\kappa_0''(\hat{w})^{1/2}\left[\frac{\partial^2\kappa_{[n]}(\Delta, \hat{u}|x)}{\partial u^2}\right]^{-1/2}, \end{aligned} \tag{3.15}$$

where the saddlepoints $\hat{u}$ and $\hat{w}$ can be obtained by solving

$$\frac{\partial\kappa_{[n]}(\Delta, u|x)}{\partial u} = y, \quad \kappa_0'(w) = y,$$

respectively.

Unlike the previous saddlepoint approximations, the above formula (3.15) requires an extra effort in computing the saddlepoints $\hat{u}$ and $\hat{w}$ by Newton's iterative root finding method. One may also use the series inversion method (Lieberman, 1994) to find an analytic approximation solution without resort to numerical root finding procedure.

Aït-Sahalia and Yu (2006) tested the saddlepoint approximation with a non-Gaussian base on the same process given by (3.13). By examining a plot over $[-5, 5]$ which covers a 10 standard deviation region, they find that the saddlepoint approximation with a non-Gaussian base significantly improves accuracy by reducing the log approximation error to almost one tenth of the Gaussian approximation. More importantly, the tail behavior of the true density function has been largely captured by the approximation (with percentage error bounded by 5%).

Lévy processes

Lévy processes are natural generalization of the Brownian motion with Poisson jumps, which have stationary and independent increments. A Lévy process X_t is fully characterized by the triplet (μ, σ^2, ν), where μ is the drift coefficient, σ^2 is the diffusion coefficient, and $\nu(\cdot)$ is the Lévy measure that drives the jump distribution of the process. Unlike the Brownian motion and Poisson process, the transition density of a Lévy process is analytically intractable. On the other hand, the characteristic function of X_t is usually available and explicitly given by the Lévy-Khintchine formula (Bertoin 1998):

$$E[e^{iuX_\Delta}|X_0 = 0] = e^{\Delta\Psi(u)}, \tag{3.16}$$

where

$$\Psi(u) = i\mu u - \frac{\sigma^2}{2}u^2 + \int_{\mathbb{R}\setminus\{0\}} [e^{iuw} - 1 - iuw\mathbf{1}_{|w|<1}]\nu(\mathrm{d}w).$$

Here, the Lévy measure satisfies $\int_{\mathbb{R}}(1 \wedge |w|^2)\nu(\mathrm{d}w) < \infty$. In practice, it is often assumed that the characteristic function is analytically extensible to a complex region that contains the origin as an interior point. Accordingly, the mgf of X_t is well defined over a neighborhood of the origin. It is related to the characteristic function as follows:

$$M(\Delta, u|X_0 = 0) = E[e^{uX_\Delta}|X_0 = x] = e^{\Delta\Psi(-iu)}.$$

It is well known that the density function can be obtained by the Fourier inversion method once the characteristic function is available. Yet, the numerical Fourier inversion procedure tends to lose information about the potential infinitely many small jumps of the Lévy process, due to the truncation made to facilitate the numerical integration. Fortunately, there is no such truncation needed in saddlepoint approximation. This may make the identification of extremely small jumps possible in the density function.

Using (3.16), the cgf of X_t is given by

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

By choosing $\varepsilon \in (0, 1)$, we obtain

$$\begin{aligned}\int_{-\varepsilon}^{\varepsilon} (e^{ux} - 1 - ux\mathbf{1}_{\{|x|<1\}})\nu(\mathrm{d}x) &= \int_{-\varepsilon}^{\varepsilon} (e^{ux} - 1 - ux)\nu(\mathrm{d}x)\\ &= \int_{-\varepsilon}^{\varepsilon}\left[\frac{u^2x^2}{2} + o(x^2)\right]\nu(\mathrm{d}x).\end{aligned}$$

Since $\int_{\mathbb{R}}(1 \wedge |x|^2)\nu(\mathrm{d}x) < \infty$, we have the following approximation:

$$\int_{-\varepsilon}^{\varepsilon} (e^{ux} - 1 - ux\mathbf{1}_{\{|x|<1\}})\nu(\mathrm{d}x) \approx \frac{u^2}{2}\int_{-\varepsilon}^{\varepsilon} x^2\nu(\mathrm{d}x) = \frac{u^2}{2}\sigma_J^2,$$

where σ_J^2 can be regarded as an adjustment to the diffusion coefficient due to small jumps. As a result, the cgf can be approximated as follows:

$$\kappa(\Delta, u|X_0 = 0) \approx \Delta\left[\mu u + \frac{u^2}{2}(\sigma^2 + \sigma_J^2) + \int_{|x|\geq\varepsilon} (e^{ux} - 1)\nu(\mathrm{d}x)\right]. \tag{3.17}$$

In a similar spirit of choosing the non-Gaussian base for jump-diffusion processes, we consider the base distribution whose cgf is given by

$$\kappa_0(\Delta, u|X_0 = 0) = \ln\left((1-\lambda\Delta)\exp\left(\mu u\Delta + \frac{u^2}{2}(\sigma^2+\sigma_J^2)\Delta\right) + \Delta\int_{|x|\geq\varepsilon} e^{ux}\nu(\mathrm{d}x)\right),$$

where $\lambda = \int_{|x|\geq\varepsilon}\nu(\mathrm{d}x)$. It is observed that the difference between κ_0 and the approximate κ is of order $O(\Delta^2)$. Moreover, the density function f_0 that corresponds to κ_0 is found to be

$$f_0(x) = (1-\lambda\Delta)\phi(x; \mu\Delta, (\sigma^2+\sigma_J^2)\Delta) + \Delta\nu(x)\mathbf{1}_{\{|x|\geq\varepsilon\}}. \tag{3.18}$$

Theorem 3.4 (Aït-Sahalia and Yu 2006) *With the approximate cgf in (3.17), the saddlepoint approximation of the transition density of the Lévy process X_t with a non-Gaussian base is given by*

$$p(\Delta, x|0) = f_0(x)\exp(\kappa(\Delta, \hat{u}|0) - \hat{u}x - \kappa_0(\Delta, \hat{w}|0) + \hat{w}x)$$
$$\left[\frac{\partial^2\kappa_0(\Delta, \hat{w}|0)}{\partial w^2}\Big/\frac{\partial^2\kappa(\Delta, \hat{u}|0)}{\partial u^2}\right]^{1/2}, \tag{3.19}$$

where the saddlepoints $\hat{u}$ and $\hat{w}$ are determined, respectively, by solving the following saddlepoint equations:

$$\frac{\partial\kappa(\Delta, u|0)}{\partial u} = x, \quad \frac{\partial\kappa_0(\Delta, w|0)}{\partial w} = x.$$

The saddlepoint approximation formula (3.19) was tested on a variance gamma process in Aït-Sahalia and Yu (2006). Over a large range of values of x, (3.19) is seen to provide very satisfactory approximations.

3.2 Affine Jump-Diffusion Processes

The affine jump-diffusion (ADJ) processes are renowned for nice tractability in the sense that the characteristic function of the underlying process exhibits an exponential affine structure and can be obtained by either solving a Riccati system of ordinary differential equations (ODEs) (Duffie et al. 2000) explicitly or numerically. In practice, it is desirable for the ADJ model to admit a closed form characteristic function so as to facilitate calculations such as inverse Fourier transform. This requirement on analytic tractability limits the choice of ADJ models to just a small class of processes, like the Heston stochastic volatility model.

The saddlepoint approximation method introduced by Glasserman and Kim (2009) represents one of ingenious means to overcome the difficulty of nonavailability of an explicit solution imposed by the general ODE system. They consider the n-dimensional ADJ process $\mathbf{X}_t \in \mathbb{R}^n$ that is defined by the following stochastic differential equation:

$$\mathrm{d}\mathbf{X}_t = \boldsymbol{\mu}(\mathbf{X}_t)\mathrm{d}t + \sigma(\mathbf{X}_t)\mathrm{d}\mathbf{W}_t + \mathrm{d}\mathbf{J}_t,$$

where $\mathbf{W}_t$ is an $(\mathscr{F}_t)$-adapted Brownian motion in $\mathbb{R}^n$, $\mathscr{F}_t$ stands for the σ-field of information sets available up to time t, and $\mathbf{J}_t$ is a pure jump process whose jumps have a fixed probability distribution $\nu(\cdot)$ in $\mathbb{R}^n$ and arrives with intensity $\lambda(\mathbf{X}_t)$. The affine feature is reflected by the following specification of the functional forms of the coefficients that are linear in $\mathbf{x}$:

$$\begin{aligned}
&\boldsymbol{\mu}(\mathbf{x}) = \mathbf{K}_0 + K_1\mathbf{x}, \quad K = (\mathbf{K}_0, K_1) \in \mathbb{R}^n \times \mathbb{R}^{n\times n},\ \mathbf{x} \in \mathbb{R}^n \\
&(\sigma(\mathbf{x})\sigma(\mathbf{x})^T)_{ij} = (H_0)_{ij} + (H_1)_{ij} \cdot \mathbf{x}, \quad H = (H_0, H_1) \in \mathbb{R}^{n\times n} \times \mathbb{R}^{n\times n\times n} \\
&\lambda(\mathbf{x}) = l_0 + \mathbf{l}_1 \cdot \mathbf{x}, \quad l = (l_0, \mathbf{l}_1) \in \mathbb{R} \times \mathbb{R}^n.
\end{aligned}$$

Also, let $\theta(\mathbf{c}) = \int_{\mathbb{R}^n} \exp(\mathbf{c}\cdot\mathbf{z})\,\mathrm{d}\nu(\mathbf{z})$ for $\mathbf{c} \in \mathbb{C}^n$. For generality, the interest rate process is assumed to be stochastic and affine in $\mathbf{X}_t$, where

$$r(\mathbf{x}) = \rho_0 + \boldsymbol{\rho}_1 \cdot \mathbf{x}, \quad \rho = (\rho_0, \boldsymbol{\rho}_1) \in \mathbb{R} \times \mathbb{R}^n.$$

The process $\mathbf{X}_t$ is said to have the characteristic (K, H, l, θ, ρ).

System of ordinary differential equations for extended transforms

We define the kth extended transform to be

$$M_k(\mathbf{v}, \mathbf{u}, \mathbf{X}_t, t, T) = E\left[\exp\left(-\int_t^T r(\mathbf{X}_s)\,\mathrm{d}s\right)(\mathbf{v}\cdot\mathbf{X}_T)^k e^{\mathbf{u}\cdot\mathbf{X}_T}\middle|\mathscr{F}_t\right], \quad k = 0, 1, 2, \cdots,$$

for $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$. It has been shown in Duffie et al. (2000) that M_0 and M_1 admit the following affine structures:

$$M_0(\mathbf{u}, \mathbf{X}_t, t, T) = e^{\alpha(T-t,u)+\boldsymbol{\beta}(T-t,u)\cdot\mathbf{X}_t}, \tag{3.20a}$$
$$M_1(\mathbf{v}, \mathbf{u}, \mathbf{X}_t, t, T) = M_0(\mathbf{u}, X_t, t, T)[A(T-t, \mathbf{u}, \mathbf{v}) + \mathbf{B}(T-t, \mathbf{u}, \mathbf{v})\cdot\mathbf{X}_t], \tag{3.20b}$$

where the parameter functions α, $\boldsymbol{\beta}$, A and $\mathbf{B}$ satisfy the following Ricatti system of ordinary differential equations (ODEs):

$$\frac{\partial\boldsymbol{\beta}}{\partial\tau} = -\boldsymbol{\rho}_1 + K_1^T\boldsymbol{\beta}(\tau) + \frac{1}{2}\boldsymbol{\beta}(\tau)^T H_1\boldsymbol{\beta}(\tau) + \mathbf{l}_1[\theta(\boldsymbol{\beta}(\tau)) - 1], \tag{3.21a}$$
$$\frac{\partial\alpha}{\partial\tau} = -\rho_0 + \mathbf{K}_0\cdot\boldsymbol{\beta}(\tau) + \frac{1}{2}\boldsymbol{\beta}(\tau)^T H_0\boldsymbol{\beta}(\tau) + l_0[\theta(\boldsymbol{\beta}(\tau)) - 1], \tag{3.21b}$$
$$\frac{\partial\mathbf{B}}{\partial\tau} = K_1^T\mathbf{B}(\tau) + \boldsymbol{\beta}(\tau)^T H_1\mathbf{B}(\tau) + \mathbf{l}_1\nabla\theta(\boldsymbol{\beta}(\tau))\mathbf{B}(\tau), \tag{3.21c}$$

$$\frac{\partial A}{\partial \tau} = \mathbf{K}_0 \cdot \mathbf{B}(\tau) + \boldsymbol{\beta}(\tau)^T H_0 \mathbf{B}(\tau) + l_0 \nabla\theta(\boldsymbol{\beta}(\tau))\mathbf{B}(\tau), \tag{3.21d}$$

with $\boldsymbol{\beta}(0) = \mathbf{u}$, $\alpha(0) = 0$, $\mathbf{B}(0) = \mathbf{v}$, $A(0) = 0$, and $\nabla\theta(\mathbf{c})$ denotes the gradient of $\theta(\mathbf{c})$. Here, we have suppressed the dependency on $\mathbf{u}$ and $\mathbf{v}$ for brevity.

Higher order cumulants normally appear in the saddlepoint approximation formulas. Glasserman and Kim (2009) manage to obtain the following formulas for higher order extended transforms:

$$\begin{aligned} M_2(\mathbf{v}, \mathbf{u}, \mathbf{X}_t, t, T) = M_0(\mathbf{u}, X_t, t, T)\{&[A(T-t) + \mathbf{B}(T-t)\cdot\mathbf{X}_t]^2 \\ &+ C(T-t) + \mathbf{D}(T-t)\cdot\mathbf{X}_t\} \end{aligned} \tag{3.22a}$$

$$\begin{aligned} M_3(\mathbf{v}, \mathbf{u}, X_t, t, T) = M_0(u, X_t, t, T)\{&[A(T-t) + \mathbf{B}(T-t)\cdot\mathbf{X}_t]^3 \\ &+ 3[A(T-t) + \mathbf{B}(T-t)\cdot\mathbf{X}_t][C(T-t) + \mathbf{D}(T-t)\cdot\mathbf{X}_t] \\ &+ E(T-t) + \mathbf{F}(T-t)\cdot\mathbf{X}_t\}, \end{aligned} \tag{3.22b}$$

where C, $\mathbf{D}$, E and $\mathbf{F}$ satisfy the following ODE system:

$$\begin{aligned}
\frac{\partial \mathbf{D}}{\partial \tau} &= K_1^T \mathbf{D}(\tau) + \boldsymbol{\beta}(\tau)^T H_1 \mathbf{D}(\tau) + \mathbf{l}_1 \nabla\theta(\boldsymbol{\beta}(\tau))\mathbf{D}(\tau) + \mathbf{B}(\tau)^T H_1 \mathbf{B}(\tau) \\
&\quad + \mathbf{l}_1 \mathbf{B}(\tau)^T \nabla^2\theta(\boldsymbol{\beta}(\tau))\mathbf{B}(\tau), \\
\frac{\partial C}{\partial \tau} &= \mathbf{K}_0 \cdot \mathbf{D}(\tau) + \boldsymbol{\beta}(\tau)^T H_0 \mathbf{D}(\tau) + l_0 \nabla\theta(\boldsymbol{\beta}(\tau))\mathbf{D}(\tau) + \mathbf{B}(\tau)^T H_0 \mathbf{B}(\tau) \\
&\quad + l_0 \mathbf{B}(\tau)^T \nabla^2\theta(\boldsymbol{\beta}(\tau))\mathbf{B}(\tau), \\
\frac{\partial \mathbf{F}}{\partial \tau} &= K_1^T \mathbf{F}(\tau) + \boldsymbol{\beta}(\tau)^T H_1 \mathbf{F}(\tau) + \mathbf{l}_1 \nabla\theta(\boldsymbol{\beta}(\tau))\mathbf{F}(\tau) + 3\mathbf{B}(\tau)^T H_1 \mathbf{D}(\tau) \\
&\quad + 3\mathbf{l}_1 \mathbf{B}(\tau)^T \nabla^2\theta(\boldsymbol{\beta}(\tau))\mathbf{D}(\tau) + \mathbf{l}_1 \int_{\mathbb{R}^n} e^{\mathbf{z}\cdot\boldsymbol{\beta}(\tau)} [\mathbf{z}\cdot\mathbf{B}(\tau)]^3 \, d\nu(\mathbf{z}), \\
\frac{\partial E}{\partial \tau} &= \mathbf{K}_0 \cdot \mathbf{F}(\tau) + \boldsymbol{\beta}(\tau)^T H_0 F(\tau) + l_0 \nabla\theta(\boldsymbol{\beta}(\tau)) F(\tau) + 3\mathbf{B}(\tau)^T H_0 \mathbf{D}(\tau) \\
&\quad + 3 l_0 \mathbf{B}(\tau)^T \nabla^2\theta(\boldsymbol{\beta}(\tau))\mathbf{D}(\tau) + l_0 \int_{\mathbb{R}^n} e^{\mathbf{z}\cdot\boldsymbol{\beta}(\tau)} [\mathbf{z}\cdot\mathbf{B}(\tau)]^3 \, d\nu(\mathbf{z}),
\end{aligned}$$

with $C(0) = E(0) = 0$, $\mathbf{D}(0) = \mathbf{F}(0) = \mathbf{0}$, and $\nabla^2\theta(\mathbf{c})$ is the Hessian of $\theta(\mathbf{c})$.

In practice, we are more interested in one-dimensional underlying processes constructed as $x_t = \mathbf{b}\cdot\mathbf{X}_t$ for some scalar vector $\mathbf{b} \in \mathbb{R}^n$. For example, x_t can be used to model the log asset price process. As a result, we restrict our discussion to the special version of M_k that is parameterized by the parameter vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$ and scalar z:

$$\bar{M}_k(z, \mathbf{X}_t, t, T) = M_k(\mathbf{b}, \mathbf{a} + z\mathbf{b}, \mathbf{X}_t, t, T).$$

Analogously, we introduce the following notations for convenience:

$$\bar{\alpha}(\tau, z) = \alpha(\tau, \mathbf{a} + z\mathbf{b}), \qquad \bar{\boldsymbol{\beta}}(\tau, z) = \boldsymbol{\beta}(\tau, \mathbf{a} + z\mathbf{b})$$

$$\bar{A}(\tau, z) = A(\tau, \mathbf{a} + z\mathbf{b}, \mathbf{b}), \quad \bar{\mathbf{B}}(\tau, z) = \mathbf{B}(\tau, \mathbf{a} + z\mathbf{b}, \mathbf{b})$$
$$\bar{C}(\tau, z) = C(\tau, \mathbf{a} + z\mathbf{b}, \mathbf{b}), \quad \bar{\mathbf{D}}(\tau, z) = \mathbf{D}(\tau, \mathbf{a} + z\mathbf{b}, \mathbf{b})$$
$$\bar{E}(\tau, z) = E(\tau, \mathbf{a} + z\mathbf{b}, \mathbf{b}), \quad \bar{\mathbf{F}}(\tau, z) = \mathbf{F}(\tau, \mathbf{a} + z\mathbf{b}, \mathbf{b}).$$

It can be readily checked that

$$\frac{\partial \bar{\alpha}(\tau, z)}{\partial z} = \bar{A}(\tau, z), \quad \frac{\partial^2 \bar{\alpha}(\tau, z)}{\partial z^2} = \bar{C}(\tau, z), \quad \frac{\partial^3 \bar{\alpha}(\tau, z)}{\partial z^3} = \bar{E}(\tau, z),$$
$$\frac{\partial \bar{\boldsymbol{\beta}}(\tau, z)}{\partial z} = \bar{\mathbf{B}}(\tau, z), \quad \frac{\partial^2 \bar{\boldsymbol{\beta}}(\tau, z)}{\partial z^2} = \bar{\mathbf{D}}(\tau, z), \quad \frac{\partial^3 \bar{\boldsymbol{\beta}}(\tau, z)}{\partial z^3} = \bar{\mathbf{F}}(\tau, z).$$

In the context of option pricing under stochastic interest rate, we employ the change of measure to change the expectation under the risk neutral measure Q of a tail probability under an equivalent measure $\tilde{Q}$ so that the saddlepoint approximation formulas for tail probability can be applied (see Sect. 4.1.1). Glasserman and Kim (2009) define the equivalent measure $\tilde{Q}$ by

$$\left.\frac{d\tilde{Q}}{dQ}\right|_{\mathscr{F}_t} = \frac{\xi_t}{\xi_0}, \quad \xi_t = \exp\left(-\int_0^t r(\mathbf{X}_s)\, ds\right) e^{\alpha(T-t) + \boldsymbol{\beta}(T-t)\cdot \mathbf{X}_t},$$

where $\alpha(\tau)$ and $\boldsymbol{\beta}(\tau)$ satisfy (3.21b) and (3.21a), respectively, with boundary conditions given by $\alpha(0) = 0$ and $\boldsymbol{\beta}(0) = \mathbf{a}$. Under the measure $\tilde{Q}$, the mgf of x_T is closely related to $\bar{M}_0(z, \mathbf{X}_t, t, T)$ as follows:

$$E_{\tilde{Q}}[e^{zx_T}] = \frac{1}{\xi_0} E_Q\left[\exp\left(-\int_0^T r(\mathbf{X}_s)\, ds\right) e^{(\mathbf{a}+z\mathbf{b})\cdot \mathbf{X}_T}\right] = \frac{\bar{M}_0(z, \mathbf{X}_0, 0, T)}{\bar{M}_0(0, \mathbf{X}_0, 0, T)}$$
$$= \exp(\bar{\alpha}(T, z) - \bar{\alpha}(T, 0) + [\bar{\boldsymbol{\beta}}(T, z) - \bar{\boldsymbol{\beta}}(T, 0)] \cdot \mathbf{X}_0). \tag{3.23}$$

For other applications, such as risk management, one may be interested in finding the mgf of x_T under the risk neutral measure Q. This can be easily derived by setting $\boldsymbol{\rho}_1 = \mathbf{0}$ and $\mathbf{a} = \mathbf{0}$ in (3.23), so that $\tilde{Q}$ is the same as Q.

In summary, the cgf of x_t and its first three derivatives can be deduced from (3.23), which are found to be

$$\kappa_{\tilde{Q}}(z) = \bar{\alpha}(T, z) - \bar{\alpha}(T, 0) + [\bar{\boldsymbol{\beta}}(T, z) - \bar{\boldsymbol{\beta}}(T, 0)] \cdot \mathbf{X}_0$$
$$\kappa'_{\tilde{Q}}(z) = \bar{A}(T, z) + \bar{\mathbf{B}}(T, z) \cdot \mathbf{X}_0$$
$$\kappa''_{\tilde{Q}}(z) = \bar{C}(T, z) + \bar{\mathbf{D}}(T, z) \cdot \mathbf{X}_0$$
$$\kappa'''_{\tilde{Q}}(z) = \bar{E}(T, z) + \bar{\mathbf{F}}(T, z) \cdot \mathbf{X}_0.$$

All these functions can be evaluated by solving a set of ODEs numerically when explicit solutions to the Ricatti systems do not exist.

Approximation of the saddlepoint

Although the cgf and its first three derivatives can be numerically evaluated by solving the ODE systems (3.21a) to (3.21d), solving the saddlepoint equation $\kappa'(z) = y$ by the root finding iteration procedure could be quite time consuming. Therefore, it is crucial to find an accurate proxy for the true root $\hat{z}$ of $\kappa'(z) = y$ to circumvent the computationally intensive root finding iteration.

Lieberman (1994) proposes the series inversion technique to approximate the saddlepoint. Starting with the initial guess $z_0 = 0$, Lieberman (1994) takes the next Newton-Raphson iteration to be

$$z_1 = \frac{y - \kappa'(0)}{\kappa''(0)} = \frac{y - \mu}{\sigma^2}.$$

One more iteration yields

$$z_2 = z_1 - \frac{\kappa'(z_1) - y}{\kappa''(z_1)}.$$

Since the derivatives of the cgf are not known in an explicit form, Lieberman (1994) further adopts the Taylor expansion around $z = 0$ for both $\kappa'(z)$ and $\kappa''(z)$. As a result, we obtain

$$z_2 = z_1 + \frac{y - \kappa'(z_1)}{\kappa''(z_1)} = z_1 - \frac{\frac{\kappa'''(0)}{2} z_1^2 + \frac{\kappa''''(0)}{6} z_1^3 + \cdots}{\sigma^2 + \kappa'''(0) z_1 + \frac{\kappa''''(0)}{2} z_1^2 + \cdots}.$$

By using the series inversion technique, the above expression can be rewritten as

$$\begin{aligned} z_2 &= z_1 - \frac{\kappa'''(0)}{2\sigma^2} z_1^2 + \left[\frac{\kappa'''(0)^2}{2\sigma^4} - \frac{\kappa''''(0)}{6\sigma^2}\right] z_1^3 + \cdots \\ &= \frac{y-\mu}{\sigma^2} - \frac{\kappa'''(0)}{2\sigma^2}\left(\frac{y-\mu}{\sigma^2}\right)^2 + \left[\frac{\kappa'''(0)^2}{2\sigma^4} - \frac{\kappa''''(0)}{6\sigma^2}\right]\left(\frac{y-\mu}{\sigma^2}\right)^3 + \cdots \end{aligned}$$

Lieberman (1994) then takes

$$\hat{z}_3 = \frac{y-\mu}{\sigma^2} - \frac{\kappa'''(0)}{2\sigma^2}\left(\frac{y-\mu}{\sigma^2}\right)^2 + \left[\frac{\kappa'''(0)^2}{2\sigma^4} - \frac{\kappa''''(0)}{6\sigma^2}\right]\left(\frac{y-\mu}{\sigma^2}\right)^3, \quad (3.24)$$

to be an approximation of the true saddlepoint.

As a remark, Lieberman (1994) also proposes a different saddlepoint approximation formula that is supposed to work better than others, including the Lugannani-Rice formula, with the approximate saddlepoint. We let

$$\hat{v} = \hat{z}_3\sqrt{n\kappa''(\hat{z}_3)},\ \hat{\lambda}_3 = \kappa'''(\hat{z}_3)/\kappa''(\hat{z}_3)^{3/2} \text{ and } \hat{\lambda}_4 = \kappa''''(\hat{z}_3)/\kappa''(\hat{z}_3)^2.$$

Let the generalized Heavside function be defined by

$$H(x) = \mathbf{1}_{\{x>0\}} + \frac{1}{2}\mathbf{1}_{\{x=0\}},$$

Liberman's saddlepoint approximation formula for tail probability is given by

$$\begin{aligned} P[\bar{X} > y] =& H(-\hat{v}) + \exp\left(n[\kappa(\hat{z}_3) - y\hat{z}_3] + \hat{v}^2/2\right) \\ & \Bigg\{ [H(\hat{v}) - \Phi(\hat{v})] \left[1 - \frac{\hat{\lambda}_3 \hat{v}^3}{6\sqrt{n}} + \frac{1}{n}\left(\frac{\hat{\lambda}_4 \hat{v}^4}{24} + \frac{\hat{\lambda}_3^2 \hat{v}^6}{72} \right) \right] \\ & + \phi(\hat{v}) \left[\frac{\hat{\lambda}_3(\hat{v}^2 - 1)}{6\sqrt{n}} - \frac{1}{n}\left(\hat{\lambda}_4 \frac{\hat{v}^3 - \hat{v}}{24} + \hat{\lambda}_3^2 \frac{\hat{v}^5 - \hat{v}^3 + 3\hat{v}}{72} \right) \right] \Bigg\} \\ & [1 + O(n^{-3/2})]. \end{aligned}$$

Lieberman's approximation is not guaranteed to be accurate when y deviates from μ. To improve accuracy of Lieberman's approximation, Glasserman and Kim (2009) propose to start the iteration with guess $z = \hat{z}_3$ (instead of 0) and expand $\kappa'(z)$ and $\kappa''(z)$ around $z = \hat{z}_3$ to the third order to obtain an improved iterate:

$$\begin{aligned} \hat{z}_4 = \hat{z}_3 + \frac{y - \kappa'(\hat{z}_3)}{\kappa''(\hat{z}_3)} - \frac{\kappa'''(\hat{z}_3)}{2\kappa''(\hat{z}_3)} \left[\frac{y - \kappa'(\hat{z}_3)}{\kappa''(\hat{z}_3)} \right]^2 \\ + \left[\frac{\kappa'''(\hat{z}_3)}{2\kappa''(\hat{z}_3)^2} - \frac{\kappa''''(\hat{z}_3)}{6\kappa''(\hat{z}_3)} \right] \left[\frac{y - \kappa'(\hat{z}_3)}{\kappa''(\hat{z}_3)} \right]^3 . \end{aligned} \tag{3.25}$$

Glasserman and Kim (2009) tested the four saddlepoint approximation formulas against the exact analytic value of the option price under a variety of affine stochastic volatility models. These four formulas are the Lugannani-Rice formula with numerically solved saddlepoint (LR), the Lugannani-Rice formula with approximate saddlepoint $\hat{z}_3$ (LLR), Lieberman's formula, and the Lugannani-Rice formula with approximate saddlepoint $\hat{z}_4$ (GKLR). The LR method and GKLR method are shown to outperform the other two over the whole range of maturity and moneyness. The Lieberman formula works best for at-the-money options while the LLR method produces the smallest errors for deep in-the-money option prices. The LR method yields uniformly accurate results for all the models considered therein, whereas the GKLR method occasionally breaks down at the short-maturity and deep-in-the-money region, due to inaccurate approximation to the true saddlepoint by $\hat{z}_4$ under the models with jumps. In terms of accuracy and stability, the LR method remains to be the best performers, followed by the GKLR method. Yet, as we have mentioned previously, the LR method involves a time consuming root finding procedure. Fortunately, Glasserman and Kim (2009) show that by using $\hat{z}_3$ as the initial guess, the number of iterations can be reduced by 66–84% on average. Nevertheless, when efficiency is the top priority and extreme cases are not the main concern, the GKLR method is quite a good alternative.

Chapter 4
Saddlepoint Approximation Formulas for Pricing Options

Though the saddlepoint approximations have been commonly used in computing the tail probability of a random variable whose cumulant generating function (cgf) is known, the use of saddlepoint approximations for option pricing is relatively thin in the literature. Rogers and Zane (1999) first introduce the saddlepoint approximation method to price options under the exponential Lévy models. Nice analytic tractability is exhibited since the cgf of log stock price under a Lévy model is known in closed form. By expressing the option price as the difference of two tail expectations under the risk neutral measure and share measure, they apply the Lugannani-Rice formula (2.15) to obtain the corresponding saddlepoint approximation of the European option price. Xiong et al. (2005) extend the saddlepoint approximation methods to option pricing under stochastic interest rates and volatility. Carr and Madan (2009) adopt the Gaussian minus exponential distribution as the base distribution in the saddlepoint approximation procedure and show that this gives the exact call option price formula under the Black-Scholes model. Zhang and Chan (2016) consider the saddlepoint approximation for pricing European options in a regime-switching model under a Markov-modulated geometric Brownian motion. In the case of two-state Markov chains, they manage to derive an explicit analytic formula of the cgf of the underlying process. When the regime-switching model has more than two states, an approximate formula of the cgf is found using a splitting method. They then use the saddlepoint approximation formula for tail probability based on the approximate saddlepoint (Lieberman 1994) to obtain approximate option prices. As for potential extensions, they claim that their approach can be used to price barrier options and lookback options under a multi-state regime-switching model. For more exotic financial derivatives, like VIX products and variance derivatives, we discuss how the saddlepoint approximations can be applied to pricing VIX futures and options, and options on discrete realized variance. For VIX products, the price formulas involve tail expectation of the square root of the random terminal value of the underlying price process under the stochastic volatility model. For options on discrete realized

Y. K. Kwok and W. Zheng, *Saddlepoint Approximation Methods in Financial Engineering*, SpringerBriefs in Quantitative Finance, https://doi.org/10.1007/978-3-319-74101-7_4

variance, an innovative analytic technique using small time asymptotic approximation is used to find the approximate cgf of the underlying Lévy process and stochastic volatility process.

4.1 Option Prices as Complementary Probabilities

We consider a European call option with maturity T, strike $K = e^k$ and underlying asset price process $S_t = e^{X_t}$. Let Q denote a risk neutral pricing measure under which the discounted price process is a Q-martingale. We assume that the cgf of X_T defined as $\kappa(z) = \log E_t^Q\left[e^{zX_T}\right]$ is available in closed form and analytic in some open vertical strip $\{z \in \mathbb{C} : \alpha_- < \mathrm{Re}\, z < \alpha_+\}$, where $\alpha_- < 0$ and $0 < \alpha_+$. Since the discounted value of a financial derivative is also a martingale under Q, the time-t call price is given by

$$\begin{aligned} C_t &= e^{-r(T-t)} E_t^Q\left[(e^{X_T} - e^k)^+\right] \\ &= e^{-r(T-t)} E_t^Q\left[(e^{X_T} - e^k)\mathbf{1}_{X_T > k}\right] \\ &= e^{-r(T-t)} E_t^Q\left[e^{X_T}\mathbf{1}_{X_T > k}\right] - e^{k-r(T-t)} Q[X_T > k]. \end{aligned} \tag{4.1}$$

In order to express the first expectation term in (4.1) as tail probability, Rogers and Zane (1999) define a probability measure Q_θ via the Radon-Nikodym derivative:

$$\frac{\mathrm{d}Q_\theta}{\mathrm{d}Q} = \frac{e^{\theta X_T}}{E_t^Q[e^{\theta X_T}]} = e^{\theta X_T - \kappa(\theta)}, \tag{4.2}$$

for any $\theta \in (\alpha_-, \alpha_+)$. Recall that such a change of measure is equivalent to exponential θ-tilting discussed in Sect. 2.3.2. The cgf of X_T under Q_θ is related to that under Q via the following relations:

$$E_t^{Q_\theta}\left[e^{zX_T}\right] = E_t^Q\left[e^{zX_T}\frac{\mathrm{d}Q_\theta}{\mathrm{d}Q}\right] = E_t^Q\left[e^{zX_T}e^{\theta X_T - \kappa(\theta)}\right] = E_t^Q\left[e^{(z+\theta)X_T}\right]e^{-\kappa(\theta)},$$

so that $\kappa^{Q_\theta}(z) = \kappa(z+\theta) - \kappa(\theta)$. Provided that $1 < \alpha_+$, by taking $\theta = 1$ in (4.2), we obtain

$$\frac{\mathrm{d}Q_1}{\mathrm{d}Q} = \frac{S_T}{E_t^Q[S_T]}.$$

Indeed, Q_1 is the well known share measure. The first expectation term (without the discount term) in (4.1) can be rewritten as

$$E_t^Q\left[e^{X_T}\mathbf{1}_{X_T>k}\right] = E_t^{Q_1}\left[e^{X_T}\mathbf{1}_{X_T>k}\frac{\mathrm{d}Q}{\mathrm{d}Q_1}\right] = E_t^{Q_1}\left[e^{X_T}\mathbf{1}_{X_T>k}\frac{E_t^Q[e^{X_T}]}{e^{X_T}}\right]$$
$$= E_t^Q[e^{X_T}]E_t^{Q_1}\left[\mathbf{1}_{X_T>k}\right] = e^{\kappa(1)}Q_1[X_T > k].$$

As a result, the time-t call price can be expressed as

$$C_t = e^{\kappa(1)-r(T-t)}Q_1[X_T > k] - e^{k-r(T-t)}Q[X_T > k], \tag{4.3}$$

where $\kappa(z)$ is the cgf of X_T under Q.

Provided that the closed form representation of the cgf of X_T under Q and Q_1 are available, one can apply the Lugannani-Rice formula (2.15) to evaluate the approximate values of the tail probabilities $Q_1[X_T > k]$ and $Q[X_T > k]$, and obtain the saddlepoint approximation for the European option price.

As demonstrated by the numerical tests for pricing European call options under several exponential-Lévy models such as the Merton jump-diffusion model, gamma model and hyperbolic model, Rogers and Zane (1999) conclude that the saddlepoint approximation methods can produce rather accurate and fast approximation results. By virtue of the central limit theorem, we observe that numerical accuracy improves as option's expiry increases due to the asymptotic Gaussian behavior of the asset return distribution.

Next, we discuss how the saddlepoint approximation can be extended to option pricing under stochastic volatility and interest rate (Xiong et al. 2005). Also, we show how numerical accuracy of the saddlepoint approximation for option pricing can be enhanced by a judicious choice of the base distribution (Carr and Madan 2009).

4.1.1 Extension to Stochastic Volatility and Interest Rate

Xiong et al. (2005) extend the approach of Rogers and Zane (1999) to price European options under stochastic volatility and interest rate, allowing discontinuous jumps in the underlying asset price as well.

We define $R_{t,T} = \int_t^T r_u \, \mathrm{d}u$ and write $B(t, T) = E_t^Q[e^{-R_{t,T}}]$ as the time-t price of T-maturity zero-coupon bond. Under the zero dividend yield assumption, the martingale property of discounted asset price process gives

$$S_t = E_t^Q\left[e^{-R_{r,T}+X_T}\right],$$

where $S_T = e^{X_T}$. Under stochastic interest rate, the share measure Q_S is defined by

$$\frac{\mathrm{d}Q_S}{\mathrm{d}Q} = \frac{S_T/e^{R_{t,T}}}{S_t} = \frac{e^{-R_{t,T}+X_T}}{E_t^Q[e^{-R_{t,T}+X_T}]}. \tag{4.4a}$$

This share measure reduces to the form of Q_1 defined earlier under constant interest rate. Also, the T-forward measure Q_T is defined by

$$\frac{\mathrm{d}Q_T}{\mathrm{d}Q} = \frac{e^{-R_{t,T}}}{B(t,T)} = \frac{e^{-R_{t,T}}}{E_t^Q[e^{-R_{t,T}}]}. \tag{4.4b}$$

Similarly, the European option price can also be expressed as the difference of two tail expectations under the share measure Q_S and T-forward measure Q_T. The time-t call option price is given by

$$\begin{aligned}
C_t &= E_t^Q\left[e^{-R_{t,T}}(e^{X_T} - e^k)^+\right] \\
&= E_t^Q\left[e^{-R_{t,T}+X_T}\mathbf{1}_{X_T>k}\right] - e^k E_t^Q\left[e^{-R_{t,T}}\mathbf{1}_{X_T>k}\right] \\
&= E_t^{Q_S}\left[e^{-R_{t,T}+X_T}\mathbf{1}_{X_T>k}\frac{\mathrm{d}Q}{\mathrm{d}Q_S}\right] - e^k E_t^{Q_T}\left[e^{-R_{t,T}}\mathbf{1}_{X_T>k}\frac{\mathrm{d}Q}{\mathrm{d}Q_T}\right] \\
&= E_t^{Q_S}\left[e^{-R_{t,T}+X_T}\mathbf{1}_{X_T>k}\frac{E_t^Q[e^{-R_{t,T}+X_T}]}{e^{-R_{t,T}+X_T}}\right] - e^k E_t^{Q_T}\left[e^{-R_{t,T}}\mathbf{1}_{X_T>k}\frac{E_t^Q[e^{-R_{t,T}}]}{e^{-R_{t,T}}}\right] \\
&= E_t^Q[e^{-R_{t,T}+X_T}]Q_S[X_T > k] - e^k E_t^Q[e^{-R_{t,T}}]Q_T[X_T > k] \\
&= S_t Q_S[X_T > k] - e^k B(t,T) Q_T[X_T > k].
\end{aligned} \tag{4.5}$$

The remaining task is to compute the corresponding cgf under both measures Q_S and Q_T. Let $\psi(z)$ be the mgf of the discounted log stock price process X_T under the risk neutral measure Q, where

$$\psi(z) = E_t^Q\left[e^{-R_{t,T}+zX_T}\right].$$

Note that $S_t = E_t^Q[e^{-R_{t,T}+X_T}] = \psi(1)$ and $B(t,T) = E_t^Q[e^{-R_{t,T}}] = \psi(0)$. The mgf of X_T under Q_S is given by

$$\begin{aligned}
E_t^{Q_S}\left[e^{zX_T}\right] &= E_t^Q\left[e^{zX_T}\frac{\mathrm{d}Q_S}{\mathrm{d}Q}\right] = E_t^Q\left[e^{zX_T}\frac{e^{-R_{t,T}+X_T}}{E_t^Q[e^{-R_{t,T}+X_T}]}\right] \\
&= \frac{E_t^Q[e^{-R_{t,T}+(z+1)X_T}]}{E_t^Q[e^{-R_{t,T}+X_T}]} = \frac{\psi(z+1)}{\psi(1)}
\end{aligned}$$

so that the corresponding cgf under Q_S is

$$\kappa^{Q_S}(z) = \log\frac{\psi(z+1)}{\psi(1)} = \kappa(z+1) - \ln S_t. \tag{4.6a}$$

In a similar manner, we consider

$$E_t^{Q_T}\left[e^{zX_T}\right] = E_t^Q\left[e^{zX_T}\frac{\mathrm{d}Q_T}{\mathrm{d}Q}\right] = E_t^Q\left[e^{zX_T}\frac{e^{-R_{t,T}}}{E_t^Q[e^{-R_{t,T}}]}\right]$$
$$= \frac{E_t^Q[e^{-R_{t,T}+zX_T}]}{E_t^Q[e^{-R_{t,T}}]} = \frac{\psi(z)}{\psi(0)}$$

so that the cgf of X_T under Q_T is

$$\kappa^{Q_T}(z) = \log\frac{\psi(z)}{\psi(0)} = \kappa(z) - \ln B(t,T). \tag{4.6b}$$

Suppose $\kappa(z)$ is available in closed form, we may apply the Lugannani-Rice formula to obtain the saddlepoint approximations for the tail probabilities $Q_S[X_T > k]$ and $Q_T[X_T > k]$ in (4.5).

Based on their numerical tests, Xiong et al. (2005) conclude that the saddlepoint approximation method for pricing European options under stochastic volatility and interest rate is generally accurate, and computationally efficient when compared with numerical Fourier inversion method and Monte Carlo simulation.

4.1.2 Non-Gaussian Base

Carr and Madan (2009) manage to find a judicious non-Gaussian base distribution that produces an exact result for pricing a call option under the Black-Scholes model. While most option price formulas involve two tail expectation calculations, they reduce option valuation to single expectation. They argue that the call price can be written as a complementary probability under the share measure Q_S such that the log stock price $X_T = \ln S_T$ exceeds the log strike $k = \ln K$ by an independent exponential random variable Y with parameter value $\lambda = 1$. This is a distribution free result. To show their claim, we consider the call price as a function of the strike K using the share measure Q_S where the stock price is used as the numeraire. This gives

$$\frac{C_t(K)}{S_t} = E_t^{Q_S}\left[\left(1 - \frac{K}{S_T}\right)^+\right].$$

Define U to be $\ln\frac{S_T}{K}$ and write its cumulative distribution function as $F(u)$ and density function as $f(u)$ under the share measure Q_S. We obtain

$$\frac{C_t(K)}{S_t} = E_t^{Q_S}\left[(1-e^{-U})^+\right] = \int_0^\infty [1-F(u)]e^{-u}\,\mathrm{d}u = \int_0^\infty Q_S[U>u]e^{-u}\,\mathrm{d}u$$
$$= \int_0^\infty Q_S[\log S_T - k > u]e^{-u}\,\mathrm{d}u = Q_S[\log S_T - k > Y], \tag{4.7}$$

with $Y \sim \text{Exp}(1)$ that is independent of S_T. The last step is obtained by noting that e^{-u} is the density of an exponential random variable with $\lambda = 1$ and then applying the convolution formula.

When applied to the classical Black-Scholes model, assuming zero dividend, the log stock price under the share measure Q_S is given by

$$X_T = \ln S_T = \ln S_t + \big(r - \frac{\sigma^2}{2}\big)(T-t) + \sigma\sqrt{T-t}Z \quad \text{with } Z \sim N(0,1).$$

Allowing shifting and scaling in the distribution function, (4.7) implies that the underlying distribution of the Black-Scholes call price is a Gaussian variate $N\big(\frac{\sigma^2}{2}(T-t), \sigma^2(T-t)\big)$ minus an independent exponential variate Exp (1). In order to obtain an exact saddlepoint formula (2.18) for the Black-Scholes case, Carr and Madan (2009) choose the base distribution to be a family $\{a + bB\}$ of zero-mean Gaussian minus exponential (GME) random variables up to shift a and scaling b. We have

$$B = Z + \frac{1}{\lambda} - Y, \tag{4.8}$$

where $Z \sim N(0,1) \perp Y \sim \text{Exp}(\lambda)$, and the parameters a, b and λ are free to be chosen. Recall that the parameter λ of an exponential random variable Y observes $E[Y] = 1/\lambda$ and $\text{var}(Y) = 1/\lambda^2$.

It can be shown that by choosing $\lambda = b = \sigma\sqrt{T-t}$ and $a = \frac{\sigma^2(T-t)}{2} - 1$, the cgf of the base $a + bB$ is equal to that of the underlying distribution of the Black-Scholes case under the share measure Q_S. The corresponding cgf under Q_S is found to be

$$\kappa^{Q_S}(z) = \frac{\sigma^2}{2}(T-t)z + \frac{\sigma^2}{2}(T-t)z^2 - \log(1+z).$$

When applied to general models where the log stock price X_T admits closed form representation of cgf, the parameters a, b and λ are determined by matching the cgf and its derivatives of the base distribution $a + bB$ with those of the underlying target distribution $X_T - Y$ under the share measure Q_S. These cumulants are evaluated at the respective saddlepoints. Finally, after a, b and λ have been determined, Wood et al. (1993)'s generalized Lugannani-Rice formula (2.18) with the choice of the GME base can be used to approximate the tail probability $Q_S[X_T - Y > k]$ in (4.7), yielding the corresponding saddlepoint approximation for the European option price.

In their numerical tests, Carr and Madan (2009) verify that the GME-base saddlepoint approximation method is exact for the classical Black-Scholes model. For a variety of other asset price models, this method is seen to produce accurate and reliable approximations for both near-the-money and deep out-of-the-money European options. This circumvents the common difficulties in pricing out-of-the-money options using the fast Fourier transform method (Carr and Madan 1999), where negative numerical option prices may result.

4.2 VIX Derivatives

In this section, we apply the saddlepoint approximation methods to price VIX futures and option under the stochastic volatility model with simultaneous jumps (SVJJ). Thanks to the affine structure of the SVJJ model, the mgf of the instantaneous variance V_T is available in closed form, while the transition density does not admit closed form. Provided that the cgf of X is available in closed form, the saddlepoint approximation methods are versatile enough to encompass approximations to tail expectations that admit the forms that involve $\sqrt{X}$, like $E[\sqrt{X}]$ or $E[(\sqrt{X}-K)^+]$, K being a constant. For these forms of expectations, it is necessary to employ the alternative saddlepoint approximation procedure where the power function in z in the integrand has to be absorbed into the exponent in the exponential term. We show how to price VIX futures and options using these saddlepoint approximation formulas.

4.2.1 Pricing VIX Futures

Under the SVJJ model, the underlying asset price process and its stochastic volatility process are governed by

$$\begin{aligned} \frac{\mathrm{d}S_t}{S_t} &= (r - \bar{\mu}\lambda)\mathrm{d}t + \sqrt{V_t}\,\mathrm{d}W_t^S + (e^{J^S} - 1)\mathrm{d}N_t \\ \mathrm{d}V_t &= \kappa(\theta - V_t)\mathrm{d}t + \sigma_V\sqrt{V_t}\,\mathrm{d}W_t^V + J^V\mathrm{d}N_t \end{aligned}, \tag{4.9}$$

where W_t^S and W_t^V are a pair of correlated standard Brownian motions with $\mathrm{d}W_t^S\mathrm{d}W_t^V = \rho\,\mathrm{d}t$, and N_t is a Poisson process with constant intensity that is uncorrelated to the two Brownian motions. We assume simultaneous jumps in S_t and V_t so that the dynamics equations of S_t and V_t share the common N_t. We let J^S and J^V denote the random jump size of the log asset price and variance, respectively. These random jump sizes are assumed to be independent of W_t^S, W_t^V and N_t. Also, we let r denote the constant riskless interest rate. To achieve nice analytical tractability, we assume J^V to be exponentially distributed while $J^S|J^V$ is normally distributed. To be specific, we assume $J^V \sim \mathrm{Exp}(1/\eta)$, which is an exponential distribution with parameter rate $1/\eta$; and $J^S|J^V$ follows:

$$J^S|J^V = N(\mu_S + \rho_J J^V, \sigma_S^2),$$

that is, the normal distribution with mean $\mu_S + \rho_J J^V$ and variance σ_S^2. One can obtain

$$\bar{\mu} = E_t^Q[e^{J^S} - 1] = \frac{e^{\mu_S} + \sigma_S^2/2}{1 - \eta\rho_J} - 1,$$

provided that $\eta\rho_J < 1$ (Zhu and Lian 2012).

As specified by the Chicago Board Options Exchange (CBOE 2003), the formal definition of $\mathrm{VIX}(t, t+\hat{\tau})$ is given by

$$\mathrm{VIX}^2(t, t+\hat{\tau}) = \frac{2}{\hat{\tau}} \sum_i \frac{\Delta X_i}{X_i^2} e^{r\hat{\tau}} Q_i(X_i) - \frac{1}{\hat{\tau}} \left[\frac{F_t(t+\hat{\tau})}{X_0} - 1 \right]^2,$$

where $\hat{\tau} = 30/365$, $Q_i(X_i)$ is the price of the out-of-the-money SPX option with strike X_i, and X_0 is the highest strike below the index forward price $F_t(t+\hat{\tau})$. Here, all index options and forward are maturing at $t+\hat{\tau}$. By taking the continuous limit of the above discretized sum of the out-of-the-money SPX options, we obtain

$$\mathrm{VIX}_t^2 = -\frac{2}{\hat{\tau}} E_t^Q \left[\ln \frac{S_{t+\hat{\tau}}}{S_t e^{r\hat{\tau}}} \right]. \tag{4.10}$$

The proof of formula (4.10) is presented in Appendix. The above conditional expectation can be shown to be (Zhu and Lian 2012)

$$\mathrm{VIX}_t^2 = a V_t + b, \tag{4.11}$$

where the $\hat{\tau}$-dependent parameter functions a and b are found to be

$$a = \frac{1 - e^{-\kappa\hat{\tau}}}{\kappa\hat{\tau}}, \; b = 2\lambda\big[\bar{\mu} - (\mu_S + \rho_J \eta)\big] + \left(\theta + \frac{\eta\lambda}{\kappa}\right)(1-a).$$

The time-t price of the T-maturity VIX futures is given by

$$F(V_t, t) = E_t^Q[\mathrm{VIX}_T] = E_t^Q[\sqrt{aV_T + b}].$$

In order to derive the saddlepoint approximation formulas for the VIX futures price $F(V_t, t)$, the first step is to derive the corresponding approximation of the expectation $E_t^Q[\sqrt{X}]$ in (4.11), where $X = aV_T + b$.

It is desirable to have analytic closed forms of the mgf and cgf of X. Let $f(z; \tau, V_t)$ denote the mgf of V_T, where $\tau = T - t$ is the time to maturity. Thanks to the affine structure of the governing dynamics equation of V_t, we are able to obtain $f(z; \tau, V_t)$ in analytic form:

$$f(z; \tau, V_t) = E_t^Q\left[e^{zV_T}\right] = e^{B(z;\tau)V_t + \Gamma(z;\tau) + \Lambda(z;\tau)}, \quad \mathrm{Re}\, z < \alpha_+, \tag{4.12}$$

where

$$B(z; \tau) = \frac{2\kappa z}{\sigma_V^2(1 - e^{\kappa\tau})z + 2\kappa e^{\kappa\tau}}$$

$$\Gamma(z; \tau) = -\frac{2\kappa\theta}{\sigma_V^2} \log\left(1 + \frac{\sigma_V^2 z}{2\kappa}\left(e^{-\kappa\tau} - 1\right)\right)$$

$$\Lambda(z;\tau) = \frac{2\lambda\mu_V}{2\kappa\mu_V - \sigma_V^2}\log\left(1 + \frac{z(\sigma_V^2 - 2\kappa\mu_V)}{2\kappa(1-\mu_V z)}(e^{-\kappa\tau}-1)\right),$$

and α_+ is determined by requiring the arguments of the above logarithm terms to be greater than zero. In the sequel, we drop the superscript Q and subscript t in E_t^Q for notational simplicity. The corresponding mgf of X is seen to be

$$E\left[e^{zX}\right] = e^{bz}E\left[e^{azV_T}\right] = e^{bz}f(az;\tau,V_t), \quad \text{Re}\, z < \alpha_+; \tag{4.13}$$

and the cgf of X is given by

$$\begin{aligned}\kappa(z) &= \log E\left[e^{zX}\right] = bz + \log f(az;\tau,V_t)\\ &= bz + B(az;\tau)V_t + \Gamma(az;\tau) + \Lambda(az;\tau), \quad \text{Re}\, z < \alpha_+.\end{aligned} \tag{4.14}$$

To find the saddlepoint approximation of $E[\sqrt{X}]$, we start with the Bromwich integral representation of $E[\sqrt{X}]$. It can be shown that (see Appendix)

$$\begin{aligned}E[\sqrt{X}] &= \frac{1}{4\sqrt{\pi}i}\int_{\gamma-i\infty}^{\gamma+i\infty}\frac{e^{\kappa(z)}}{z^{3/2}}\,\mathrm{d}z\\ &= \frac{1}{4\sqrt{\pi}i}\int_{\gamma-i\infty}^{\gamma+i\infty}e^{\kappa(z)-\frac{3}{2}\log z}\,\mathrm{d}z, \quad \gamma\in(0,\alpha_+).\end{aligned} \tag{4.15}$$

Next, we find the positive saddlepoint $\hat{z}$ within the domain $(0,\alpha_+)$ that solves the modified saddlepoint equation:

$$\kappa'(z) - \frac{3}{2z} = 0, \tag{4.16a}$$

where

$$\kappa'(z) = b + a\frac{\mathrm{d}B}{\mathrm{d}z}(az;\tau)V_t + a\frac{\mathrm{d}\Gamma}{\mathrm{d}z}(az;\tau) + a\frac{\mathrm{d}\Lambda}{\mathrm{d}z}(az;\tau). \tag{4.16b}$$

The corresponding first order derivatives of B, Γ and Λ are found to be

$$\begin{aligned}\frac{dB}{dz}(z;\tau) &= \frac{4\kappa^2 e^{\kappa\tau}}{\left[\sigma_V^2(1-e^{\kappa\tau})z + 2\kappa e^{\kappa\tau}\right]^2}\\ \frac{d\Gamma}{dz}(z;\tau) &= \frac{2\kappa\theta(e^{\kappa\tau}-1)}{\sigma_V^2(1-e^{\kappa\tau})z + 2\kappa e^{\kappa\tau}}\\ \frac{d\Lambda}{dz}(z;\tau) &= \frac{2\lambda\mu_V(e^{\kappa\tau}-1)}{\left\{\left[\sigma_V^2(1-e^{\kappa\tau}) - 2\kappa\mu_V\right]z + 2\kappa e^{\kappa\tau}\right\}\left(1-\mu_V z\right)}.\end{aligned}$$

By following similar derivation steps used in deriving the saddlepoint approximation formulas (2.64)–(2.70), the first order saddlepoint approximation formula for computing the VIX futures price $F(V_t, t)$ is given by

$$E[\sqrt{X}] \approx \frac{\sqrt{2}}{4} \frac{e^{\kappa(\hat{z})}/\hat{z}^{3/2}}{\sqrt{\kappa''(\hat{z}) + \frac{3}{2\hat{z}^2}}}. \tag{4.17}$$

Also, the corresponding second order saddlepoint approximation formula is given by

$$E[\sqrt{X}] \approx \frac{\sqrt{2}}{4} \frac{e^{\kappa(\hat{z})}/\hat{z}^{3/2}}{\sqrt{\kappa''(\hat{z}) + \frac{3}{2\hat{z}^2}}}\big(1 + R\big), \tag{4.18}$$

where

$$R = \frac{1}{8} \frac{\kappa''''(\hat{z}) + \frac{9}{\hat{z}^4}}{\left[\kappa''(\hat{z}) + \frac{3}{2\hat{z}^2}\right]^2} - \frac{5}{24} \frac{\left[\kappa'''(\hat{z}) - \frac{3}{\hat{z}^3}\right]^2}{\left[\kappa''(\hat{z}) + \frac{3}{2\hat{z}^2}\right]^3}.$$

The proof of the above saddlepoint approximation formulas for computing $F(V_t, t)$ is presented in the Appendix.

Numerical Tests on Accuracy

We performed numerical tests to examine accuracy of the saddlepoint approximation formulas (4.17) and (4.18) for pricing VIX futures with varying maturities. The parameter values of the SVJJ model used in the numerical tests are listed in Table 4.1.

We take $\sqrt{V_0} = 8.7\%$. The "exact" values (served as benchmark for comparison of accuracy) are obtained using numerical integration of the Fourier integral (4.15) with very fine integration steps. Table 4.2 lists the VIX futures values obtained using the first order formula (4.17) and second order formula (4.18), the "exact" values and percentage errors under varying maturities.

It is observed that the percentage errors using the second order approximation formula (4.18) are within 0.32%, which are considered to be fairly accurate. The first order approximation formula (4.17) gives approximation values with percentage errors ranging from 2.39 to 4.74%.

Table 4.1 Parameter values of the SVJJ model

κ	θ	ρ	σ_V	λ	η	σ_S	μ_S	ρ_J	r
0.008	1.541	−0.577	0.045	0.0007	0.374	2.305	−0.736	0.422	0.0319

Table 4.2 Comparison of numerical approximation values of VIX futures with varying maturities under the SVJJ model using the saddlepoint approximation formulas (4.17) and (4.18)

Maturity (year)	0.2	0.4	0.6	0.8	1
First order formula (4.17)	15.2739	16.0059	16.7024	17.3687	18.0087
Second order formula (4.18)	14.5473	15.3521	16.1141	16.8390	17.5316
"Exact" value	14.5823	15.3906	16.1574	16.8883	17.5879
Percentage error of first order approximation	4.74	4.00	3.37	2.84	2.39
Percentage error of second order approximation	−0.24	−0.25	−0.26	−0.29	−0.32

4.2.2 Pricing VIX Options

Analytic pricing models of VIX options under stochastic volatility with jumps of the stock index have been well explored in the literature (Goard and Mazur 2013; Lian and Zhu 2013; Lin and Chang 2009; Sepp 2008a). The VIX option prices are expressed in terms of Fourier integrals and numerical Fourier inversion algorithms are employed to perform numerical valuation of the option prices. Here, we would like to show how to derive the saddlepoint approximation pricing formulas for the VIX options under the SVJJ model. From the linear relationship between VIX_T^2 and V_T, the VIX call option price is

$$\begin{aligned} C(V_t, t) &= e^{-r(T-t)} E_t^Q\big[(VIX_T - K)^+\big] \\ &= e^{-r(T-t)} E_t^Q\big[(\sqrt{aV_T + b} - K)^+\big] \\ &= e^{-r(T-t)} E_t^Q\big[(\sqrt{X} - K)^+\big], \end{aligned} \tag{4.19}$$

where $X = aV_T + b$, a and b are given by (4.11). In order to derive the saddlepoint approximation formulas for the VIX call price $C(V_t, t)$, it is necessary to derive the corresponding approximation to $E^Q\big[(\sqrt{X} - K)^+\big]$ in (4.19), using $\kappa'(z)$ of X given in (4.16b). We first consider $E\big[(\sqrt{X} - K)^+\big]$ for any nonnegative random variable X with known cgf, then apply the result to pricing the VIX call option.

Lemma 4.1 *Let X be a nonnegative random variable, whose cgf as denoted by $\kappa(z)$ is analytic in some open vertical strip $\{z \in \mathbb{C} : \alpha_- < \text{Re } z < \alpha_+\}$, where $\alpha_- < 0$ and $\alpha_+ > 0$. The Bromwich integral representation of $E[(\sqrt{X} - K)^+]$ is given by*

$$E\big[(\sqrt{X} - K)^+\big] = \frac{1}{4\sqrt{\pi} i} \int_{\gamma - i\infty}^{\gamma + i\infty} \frac{e^{\kappa(z)}\big[1 - \text{erf}(\sqrt{z} K)\big]}{z^{3/2}}\, \mathrm{d}z, \quad \gamma \in (0, \alpha_+), \tag{4.20}$$

where $\text{erf}(x) = \dfrac{2}{\sqrt{\pi}} \displaystyle\int_0^x e^{-t^2}\, dt$.

The proof of Lemma 4.1 is shown in the Appendix. To derive the saddlepoint approximation formula, we start by expressing the integrand of the Bromwich integral (4.20) as an exponential function as follows:

$$E\big[(\sqrt{X}-K)^+\big]=\frac{1}{4\sqrt{\pi}i}\int_{\gamma-i\infty}^{\gamma+i\infty}e^{\kappa(z)+g(z)-\frac{3}{2}\log z}\,\mathrm{d}z,$$

where $g(z)=\log\big(1-\mathrm{erf}(\sqrt{z}K)\big)$. Accordingly, the modified saddlepoint equation is

$$\kappa'(z)+g'(z)-\frac{3}{2z}=0, \tag{4.21}$$

where

$$g'(z)=-\frac{Ke^{-zK^2}}{\sqrt{\pi}\sqrt{z}\big[1-\mathrm{erf}(\sqrt{z}K)\big]}.$$

Suppose that there exists a saddlepoint $\hat{z}$ that lies within the domain $(0,\alpha_+)$, by deforming the contour to the vertical line that passes through $\hat{z}$, we obtain an alternative integral representation as follows:

$$E\big[(\sqrt{X}-K)^+\big]=\frac{1}{4\sqrt{\pi}i}\int_{\hat{z}-i\infty}^{\hat{z}+i\infty}e^{\kappa(z)+g(z)-\frac{3}{2}\log z}\,\mathrm{d}z,\quad \hat{z}\in(0,\alpha_+). \tag{4.22}$$

Proposition 4.1 *Suppose that the asset price process S_t follows the stochastic volatility model with simultaneous jumps as specified by* (4.9)*. Suppose that there exists a positive saddlepoint $\hat{z}$ within the domain $(0,\alpha_+)$ that solves the modified saddlepoint Eq.* (4.21)*, where α_+ is determined by requiring the argument of the logarithm functions in* (4.12) *to be greater than zero. The first order saddlepoint approximation pricing formula for the time-t price of the VIX call option with maturity T and strike K is given by*

$$C(V_t,t)\approx e^{-r(T-t)}\frac{\sqrt{2}}{4}\frac{e^{\kappa(\hat{z})+g(\hat{z})}/\hat{z}^{3/2}}{\sqrt{\kappa''(\hat{z})+g''(\hat{z})+\frac{3}{2\hat{z}^2}}}; \tag{4.23}$$

and the second order approxzimation is given by

$$C(V_t,t)\approx e^{-r(T-t)}\frac{\sqrt{2}}{4}\frac{e^{\kappa(\hat{z})+g(\hat{z})}/\hat{z}^{3/2}}{\sqrt{\kappa''(\hat{z})+g''(\hat{z})+\frac{3}{2\hat{z}^2}}}\big(1+R\big) \tag{4.24}$$

where

$$R=\frac{1}{8}\frac{\kappa''''(\hat{z})+g''''(\hat{z})+\frac{9}{\hat{z}^4}}{\big[\kappa''(\hat{z})+g''(\hat{z})+\frac{3}{2\hat{z}^2}\big]^2}-\frac{5}{24}\frac{\big[\kappa'''(\hat{z})+g'''(\hat{z})-\frac{3}{\hat{z}^3}\big]^2}{\big[\kappa''(\hat{z})+g''(\hat{z})+\frac{3}{2\hat{z}^2}\big]^3}.$$

The proof of Proposition 4.1 is similar to that of formulas (4.17) and (4.18) except that $\kappa(z)$ in (4.15) is replaced by $\kappa(z) + g(z)$ in (4.22). The saddlepoint approximation pricing formulas for the VIX put option can be obtained by invoking the put-call parity relation, where

$$P(V_t, t) = C(V_t, t) - e^{-r(T-t)} F(V_t, t) + K e^{-r(T-t)}.$$

4.3 Options on Discrete Realized Variance

Given N monitoring dates: $0 = t_0 < t_1 < \cdots < t_N = T$, the discrete realized variance $I(0, T; N)$ of the underlying asset price process S_t over the time period $[0, T]$ is defined by

$$I(0, T; N) = \frac{A}{N} \sum_{k=1}^{N} \left(\ln \frac{S_{t_k}}{S_{t_{k-1}}} \right)^2 = \frac{A}{N} \sum_{k=1}^{N} \left(X_{t_k} - X_{t_{k-1}} \right)^2, \tag{4.25}$$

where $X_t = \ln S_t$ is the log asset price process and A is the annualized factor. It is common to take $A = 252$ for daily monitoring; and there holds $A/N = 1/T$. There are various types of derivatives on discrete realized variance that have been structured, like the variance swaps, volatility swaps, and options on discrete realized variance or volatility.

In most of the earlier pricing models of variance products and volatility derivatives, the discrete realized variance defined in (4.25) is often approximated by the quadratic variation of the log asset price process X_t over $[0, T]$. The quadratic variation can be considered as the asymptotic limit of the discrete realized variance in probability as $N \to \infty$. Suppose one fixes the monitoring frequency of the discrete realized variance to be daily, which means that A is fixed to be 252. The quadratic variation approximation has been widely adopted in pricing variance products and volatility derivatives due to its nice analytic tractability (Carr et al. 2005). For vanilla variance swaps, the quadratic variation approximation is known to work well even for short-maturity swap contracts (Sepp 2008b). However, accuracy of the quadratic variation approximation deteriorates for short-maturity derivatives with non-linear payoffs, like options on discrete realized variance.

4.3.1 Small Time Approximation

The effective implementation of the saddlepoint approximation methods relies crucially on availability of the analytic form of the mgf of the underlying process. For most option pricing models, it is not plausible to derive closed form expression for the mgf of discrete realized variance. Zheng and Kwok (2014) show how the small time asymptotic approximation can be applied to derive accurate approximation of the mgfs (defined in the left half of the complex plane) of discrete realized variance under the exponential Lévy models and stochastic volatility models with jumps.

Lévy Models

Since the Lévy process is known to have independent and stationary increments, so the log asset return $X_{t_k} - X_{t_{k-1}}$, $k = 1, 2, \dots, N$, are independent. In addition, they become identically distributed when the time step Δ is taken to be uniform. Given the independent increment property of X_t, the calculation of the mgf of $I(0, T; N)$ amounts to the calculation of the mgf of the squared process X_t^2 over the time period Δ. We observe that

$$E[e^{-uI}] = (E[e^{-u\frac{A}{N}X_\Delta^2}])^N,$$

where X_Δ^2 denotes the increment of X_t^2 over Δ. For brevity, we write $Y_t = X_t^2$ and $Y = X_\Delta^2$. As a remark, there exists the following integral representation of the mgf of Y_t for all $u \in \mathbb{R}_+$:

$$M_{Y_t}(-u) = E[e^{-uX_t^2}] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{t\psi_X(ix\sqrt{2u}) - x^2/2} dx, \tag{4.26}$$

where $\psi_X(u) = \log E[e^{uX_1}]$ is the Lévy exponent. Though the saddlepoint approximation method only requires knowledge of the mgf on the real axis, the above integral representation is not quite useful since differentiation with respect to u of the cgf [defined by $\log M_{Y_t}(u)$] is quite cumbersome. It would be almost infeasible to derive any tractable analytical expression for the higher order derivatives of the cgf.

To obtain an approximation that is more analytically tractable than the above integral representation, we make use of the small time asymptotic approximation. We consider approximating the Laplace transform of Y by that of the quadratic variation of X_t over the time interval Δ, denoted by $[X, X]_\Delta$. Apparently, Y can be regarded as the realized variance over $[0, \Delta]$. For each $u \in \mathbb{C}_-$, we obtain the asymptotic limit of the Laplace transform of $[X, X]_\Delta$ and Y as $e^{-u\Delta\sigma^2}$ and $\frac{1}{(1-2u\Delta\sigma^2)^{1/2}}$, respectively. We assume that this ratio is preserved for different values of T so that it can be used as an adjustment to deduce the approximate mgf of Y as follows:

$$\hat{M}_Y(u) = M_{[X,X]_\Delta}(u) \frac{e^{-u\Delta\sigma^2}}{(1 - 2u\Delta\sigma^2)^{1/2}}. \tag{4.27}$$

In particular, when X_t is a pure jump process with $\sigma = 0$, (4.27) reduces to

$$\hat{M}_Y(u) = M_{[X,X]_\Delta}(u).$$

In other words, we can take the mgf of the quadratic variation $[X, X]_\Delta$ directly as an approximation without any adjustment under the pure jump model. Moreover, the small time asymptotic approximation becomes exact if X_t is a Gaussian process.

It is well known that the quadratic variation process is still a Lévy process whose characteristic triplet can be derived from that of the original process. Indeed, we have

$$M_{[X,X]_\Delta}(u) = E[e^{u[X,X]_\Delta}] = \exp\left(\Delta\left[\sigma^2 u + \int (e^{ux^2}-1)F(\mathrm{d}x)\right]\right), \tag{4.28}$$

where Re $u \le 0$ (Kallsen and Muhle-Karbe 2011; Carr et al. 2005). Apparently, the mgf of the quadratic variation is still not in a closed form representation when the jump integral cannot be evaluated explicitly. In practice, we find that it is more preferable to choose this form since differentiation with respect to u is rather straightforward when it is performed under this representation.

The combination of (4.27) and (4.28) naturally leads to the following approximation of the cgf of Y:

$$\hat{\kappa}(u) = \Delta g(u) - \frac{1}{2}\ln(1-2\Delta\sigma^2 u). \tag{4.29a}$$

where

$$g(u) = \int (e^{ux^2}-1)F(dx).$$

The corresponding nth order derivatives of $\hat{\kappa}(u)$ are given by

$$\hat{\kappa}^{(n)}(u) = \Delta g^{(n)}(u) + \frac{(n-1)!}{2}\left(\frac{2\Delta\sigma^2}{1-2\Delta\sigma^2 u}\right)^n,\ n=1,2,\ldots, \tag{4.29b}$$

where

$$g^{(n)}(u) = \int e^{ux^2}x^{2n}F(dx),\ n=1,2,\ldots.$$

Stochastic Volatility Models with Jumps

We use stochastic variance as one of the risk factors in the underlying asset price process when we consider pricing of variance products and volatility derivatives. Unfortunately, it is almost intractable to derive an analytic form of the mgf of the discrete realized variance $I(0,T;N)$. The primitive approach uses the quadratic variation as a proxy of $I(0,T;N)$ since the mgf of the quadratic variation, denoted by $I(0,T;\infty)$, under an affine stochastic volatility model can be derived analytically by solving a Ricatti system of ordinary differential equations. In what follows, we derive the small time asymptotic approximation of the mgf of $I(0,T;N)$ under the general framework of stochastic volatility models with jumps.

Recall that the asymptotic limit of $I(0,T;N)$ as $T \to 0^+$ is a gamma distribution with shape parameter $N/2$ and scale parameter $2V_0/N$. The corresponding mgf is given by $\left(1-\frac{2V_0 u}{N}\right)^{-N/2}$. Therefore, for any $u \le 0$, we obtain

$$\lim_{T\to 0^+} M_{I(0,T;\infty)}(u) = e^{uV_0}, \tag{4.30a}$$

$$\lim_{T\to 0^+} M_{I(0,T;N)}(u) = \left(1 - \frac{2V_0 u}{N}\right)^{-N/2}. \tag{4.30b}$$

Assuming that the difference $M_{I(0,T;N)}(u) - M_{I(0,T;\infty)}(u)$ is invariant with respect to T, we use the above difference as a control and propose the following approximate mgf of $I(0,T;N)$:

$$\hat{M}_{I(0,T;N)}(u) = M_{I(0,T;\infty)}(u) + \left(1 - \frac{2V_0 u}{N}\right)^{-N/2} - e^{uV_0}, \quad u \in \mathbb{C}_-. \tag{4.31}$$

Note that the above approximation formula holds under the general stochastic volatility framework. In this case, we do not use the same ratio adjustment as we did in (4.27). The new approximaiton seems less cumbersome with regard to the computation of the cgf and its derivatives. Why do we choose the above approximation of the mgf instead of the one in (4.27)? From the perspective of computational stability, when u takes a very negative value, $e^{-u\Delta\sigma^2}$ grows exponentially and this leads to erosion of the approximation in (4.27). Under the Lévy model, there is a canceling factor in $M_{[X,X]_\Delta}(u)$ and the approximation remains to be stable. Unfortunately, we do not have such a property under the stochastic volatility models. As a result, approximation formulas like (4.27) may likely blow up for very negative values of u.

The approximate cgf and its higher order derivatives of $I(0,T;N)$ are given by

$$\begin{aligned}
\hat{\kappa}_{I(0,T;N)}(u) &= \ln \hat{M}_{I(0,T;N)}(u), \\
\hat{\kappa}'_{I(0,T;N)}(u) &= \frac{M'_{I(0,T;\infty)}(u) + f_1(u)}{\hat{M}_{I(0,T;N)}(u)}, \\
\hat{\kappa}^{(2)}_{I(0,T;N)}(u) &= \frac{M^{(2)}_{I(0,T;\infty)}(u) + f_2(u)}{\hat{M}_{I(0,T;N)}(u)} - \frac{[M'_{I(0,T;\infty)}(u) + f_1(u)]^2}{[\hat{M}_{I(0,T;N)}(u)]^2},
\end{aligned} \tag{4.32}$$

where

$$f_n(u) = V_0^k \frac{\frac{N}{2}\left(\frac{N}{2}+1\right)\cdots\left(\frac{N}{2}+n\right)}{\left(\frac{N}{2}\right)^n} \left(1 - \frac{2V_0 u}{N}\right)^{-N/2-n}, \; n = 1, 2, \ldots .$$

The higher order derivatives can be computed, except that they involve more cumbersome expressions.

Given the analytic expression of the mgf of $I(0,T;N)$, one faces the question of choosing which saddlepoint approximation method to be used. As mentioned before, since the cgf of $I(0,T;N)$ is only defined on the left half plane, the classical saddlepoint approximation methods have no saddlepoint on $(-\infty, 0)$ when $K > \kappa'(0)$, and hence this approach may not work for all strikes. On the other hand, the alternative saddlepoint approximation approach [see formulas (2.69) and (2.70)] can work effectively even under the scenario of $K > \kappa'(0)$ since a negative saddlepoint is guaranteed.

4.3.2 Sample Calculations

In this subsection, we present various numerical tests that were performed for the assessment of accuracy of the various saddlepoint approximation formulas. We consider pricing put options on discrete realized variance under Kou's double exponential model (Kou 2002) and the SVJJ model [see (4.9)]. First, we show the performance of the saddlepoint approximation formulas under Kou's model with different contractual specifications on the sampling frequency and strike rate. Moreover, we also present the results of the small time asymptotic approximation (STAA) by Keller-Ressel and Muhle-Karbe (2013) for comparison of accuracy. Second, we consider pricing of put options on discrete realized variance under the SVJJ model.

Kou's Double Exponential Model

The risk neutral dynamics of S_t under Kou's double exponential model (Kou 2002) is specified as

$$\frac{\mathrm{d}S_t}{S_t} = (r - m\lambda)\mathrm{d}t + \sigma\,\mathrm{d}W_t + (e^Y - 1)\,\mathrm{d}N_t, \tag{4.33}$$

where N_t is a Poisson process with intensity λ that is independent of the Brownian motion W_t, and Y denotes the independent random jump size that has an asymmetric double exponential distribution as specified by

$$Y = \begin{cases} \xi_+ & \text{with probability } p \\ -\xi_- & \text{with probability } 1-p \end{cases},$$

where $\xi_\pm$ are exponential random variables with means $1/\eta_\pm$, respectively. By the martingale property of the underlying asset price process, one can easily infer that

$$m = E[e^Y - 1] = \frac{p}{\eta_+ - 1} - \frac{1-p}{\eta_- + 1}.$$

The mgf of the log return $\ln \frac{S_{t+\Delta}}{S_t}$ is known to be

$$M_\Delta(u) = \exp\left(\Delta\left[(r - m\lambda - \sigma^2/2)u + \frac{\sigma^2 u^2}{2} + \lambda u\left(\frac{p}{\eta_+ - u} - \frac{1-p}{\eta_- + u}\right)\right]\right),$$

for $-\eta_- < u < \eta_+$. The Lévy measure is given by $\lambda f_Y(x)\,\mathrm{d}x$, where

$$f_Y(x) = p\eta_+ e^{-\eta_+ x}\mathbf{1}_{\{x\geq 0\}} + (1-p)\eta_- e^{\eta_- x}\mathbf{1}_{\{x<0\}}.$$

We consider a wide range of strikes of the put options on daily and weekly sampled realized variance. In our sample calculations, we take the riskfree interest rate to be $r = 3\%$ and the initial asset price to be $S_0 = 1$. We use the Monte Carlo simulation results as the benchmark for comparison of accuracy. The number of simulation paths was taken to be 10^6 (Table 4.3).

Numerical results

In Table 4.4, we present the numerical values of the prices of one-year deep out-of-the-money, at-the-money and deep in-the-money put options on discrete realized variance under daily and weekly sampling. The rows labelled "SPA1" and "SPA2" list the option prices calculated with the use of the first order alternative saddle-point approximation (ASPA) formula (2.69) and the second order alternative saddle-point approximation (ASPA) formula (2.70), respectively. The row labelled "STAA" presents the numerical results obtained from the small time asymptotic approximation. Both the STAA and Monte Carlo simulation results are used for benchmark comparison of accuracy.

It is obvious that both the first order and second order ASPA methods perform well under Kou's double exponential model. We do observe a deterioration of accuracy when the sampling frequency becomes lower and the option becomes more out-of-the-money for the approximation methods that have been tested. Surprisingly, the first order ASPA performs better than the second order ASPA in some cases. The

Table 4.3 Model parameter values of Kou's double exponential model

σ	λ	η_+	η_-	p
0.3	3.97	16.67	10	0.15

Table 4.4 Comparison of numerical results obtained using the first order and second order alternative saddlepoint approximation formulas for the prices of one-year put options on discrete realized variance with various strikes and sampling frequencies under Kou's double exponential model. The strikes are chosen to be 0.8 μ, μ, 1.2 μ, where μ is the at-the-money strike. All put option prices are multiplied by a notional value of 100. The last two rows labelled "MC" and "SE" show the Monte Carlo simulation results and the corresponding standard errors, respectively

Frequency (N)	Weekly (52)			Daily (252)		
Strike	0.1295	0.1618	0.1942	0.1294	0.1618	0.1941
SPA1	1.2373	3.0351	5.3350	1.1637	3.0299	5.3649
SPA2	1.2461	3.0255	5.2777	1.1546	2.9740	5.2461
STAA	1.1496	2.9636	5.2441	1.1308	2.9611	5.2409
MC	1.3332	3.1425	5.4032	1.1729	2.9998	5.2748
SE	0.0015	0.0022	0.0027	0.0013	0.0020	0.0025

second order ASPA method consistently outperforms the STAA method across all sampling frequencies and strikes.

Stochastic Volatility Model with Simultaneous Jumps

We use the same set of model parameters in Table 4.3, except that we take $\rho_J = 0$, $r = 3.19\%$ and $S_0 = 1$. As shown in Sect. 4.3.1, the approximate cgf of $I(0, T; N)$ and its higher order derivatives can be expressed in analytic forms if the mgf of the quadratic variation process $I(0, T; \infty)$ is known in closed form. This is possible under the SVJJ model when we impose $\rho_J = 0$, which means the jump size of the asset return J^S is assumed to follow an independent normal distribution with mean μ_S and variance σ_S^2. This lack of dependency between the jump size distributions in general has only minor effect on the SVJJ model in capturing the real asset price dynamics. The derivation of the analytic formula of the mgf of $I(0, T; \infty)$ can be found in Appendix B of Zheng and Kwok (2014).

Numerical results

In Table 4.5, we present the numerical values of the prices of put options on daily sampled realized variance with varying maturities and strike prices. The numerical results indicate that the saddlepoint approximation methods can produce fairly accurate results for the given range of strikes and maturities. Specifically, the approximation results for the short-maturity (5 days) and out-of-the-money (OTM) put options are fairly good. It is interesting that the second order saddlepoint approximation does not necessarily outperform the first order saddlepoint approximation. The numerical results in Table 4.5 show that "SPA2" would in general outperform "SPA1" for short-maturity or out-of-the-money put options. When maturity is lengthened and moneyness becomes more in-the-money, "SPA1" performs better than "SPA2".

Appendix

Proof of VIX_t^2 Formula (4.10)

Recall that $F_t(t + \hat{\tau})$ is the $\hat{\tau}$-period forward price of the underlying SPX at time t. Under the continuous limit of infinite number of strikes, $F_t(t + \hat{\tau})$ and X_0 coincide, so the continuous limit of $\mathrm{VIX}^2(t, t + \tau)$ is seen to be

$$\mathrm{VIX}_t^2 = \frac{2}{\hat{\tau}} \left[\int_0^{F_t} \frac{1}{K^2} e^{r\hat{\tau}} p_t(K, t + \hat{\tau}) \, \mathrm{d}K + \int_{F_t}^{\infty} \frac{1}{K^2} e^{r\hat{\tau}} c_t(K, t + \hat{\tau}) \, \mathrm{d}K \right],$$

where $F_t = S_t e^{r\hat{\tau}}$, $p_t(K, t + \hat{\tau})$ and $c_t(K, t + \hat{\tau})$ denote the time-t price of $(t + \hat{\tau})$-maturity and K-strike out-of-the-money put and call options, respectively. By considering the Taylor series expansion of $\ln \dfrac{S_{t+\hat{\tau}}}{F_t}$ and the integral representation of the remainder term, we have

Table 4.5 Comparison of numerical results obtained using the first order and second order alternative saddlepoint approximation formulas for the prices of put options on daily sampled realized variance with varying strikes and maturities under the SVJJ model with $\rho_J = 0$. The strikes are chosen to be 0.8 μ, μ, 1.2 μ, where μ is the at-the-money strike. All strike prices and option prices are multiplied by a notional value of 100. The last two rows labelled "MC" and "SE" show the Monte Carlo simulation results and the corresponding standard errors, respectively

Maturity (days)	5	10	15	20	40	60
Strike (OTM)	0.9037	0.9222	0.9399	0.9568	1.0174	1.0683
SPA1	0.2885	0.2556	0.2530	0.2577	0.2840	0.3071
SPA2	0.2851	0.2545	0.2500	0.2500	0.2741	0.2986
MC	0.2794	0.2463	0.2404	0.2441	0.2732	0.2992
SE	0.0008	0.0007	0.0006	0.0006	0.0006	0.0007
Strike (ATM)	1.1296	1.1527	1.1748	1.1960	1.2717	1.3354
SPA1	0.4579	0.4334	0.4367	0.4459	0.4865	0.5188
SPA2	0.4500	0.4255	0.4262	0.4336	0.4729	0.5079
MC	0.4490	0.4286	0.4309	0.4406	0.4828	0.5154
SE	0.0010	0.0009	0.0008	0.0008	0.0008	0.0008
Strike (ITM)	1.3555	1.3833	1.4098	1.4352	1.5261	1.6024
SPA1	0.6483	0.6352	0.6455	0.6597	0.7129	0.7517
SPA2	0.6367	0.6240	0.6322	0.6450	0.6978	0.7385
MC	0.6402	0.6330	0.6429	0.6574	0.7094	0.7465
SE	0.0012	0.0010	0.0009	0.0008	0.0008	0.0009

$$\ln \frac{S_{t+\hat{\tau}}}{F_t} = \frac{S_{t+\hat{\tau}} - F_t}{F_t} - \int_0^{F_t} \frac{1}{K^2}(K - S_{t+\hat{\tau}})^+ \mathrm{d}K - \int_{F_t}^{\infty} \frac{1}{K^2}(S_{t+\hat{\tau}} - K)^+ \mathrm{d}K.$$

Next, by taking the expectation under the risk neutral measure Q and observing $E_t^Q[S_{t+\hat{\tau}}] = F_t$, we obtain

$$E_t^Q[\ln \frac{S_{t+\hat{\tau}}}{F_t}] = -\int_0^{F_t} \frac{1}{K^2}\, e^{r\hat{\tau}} p_t(K, t+\hat{\tau})\, \mathrm{d}K - \int_{F_t}^{\infty} \frac{1}{K^2}\, e^{r\hat{\tau}} c_t(K, t+\hat{\tau})\, \mathrm{d}K,$$

hence the result in formula (4.10).

Proof of the Bromwich Representation in (4.15)

Let $p(x)$ be the density function of X. For $\gamma \in (\alpha_-, \alpha_+)$, we have

$$
\begin{aligned}
E[\sqrt{X}] &= \int_0^\infty \sqrt{x}\,p(x)\,\mathrm{d}x = \int_0^\infty \sqrt{x}\,\frac{1}{2\pi i}\int_{\gamma-i\infty}^{\gamma+i\infty} e^{\kappa(z)-zx}\,\mathrm{d}z\mathrm{d}x\\
&= \frac{1}{2\pi i}\int_{\gamma-i\infty}^{\gamma+i\infty} e^{\kappa(z)}\int_0^\infty \sqrt{x}e^{-zx}\,\mathrm{d}z\mathrm{d}x.
\end{aligned}
$$

The inner integral is known to be

$$
\int_0^\infty \sqrt{x}e^{-zx}\,\mathrm{d}x = \frac{\sqrt{\pi}}{2}\frac{1}{z^{3/2}},
$$

where the stronger condition: $\gamma = \mathrm{Re}\ z > 0$ has to be imposed. Combining these results together, we obtain (4.15).

Proof of the Saddlepoint Approximation Formulas (4.17) and (4.18)

We start with the Bromwich integral representation formula (4.15) and the corresponding modified saddlepoint Eq. (4.16a). Suppose that the positive saddlepoint $\hat{z}$ lies within the domain $(0, \alpha_+)$, we perform the deformation of the Bromwich contour through $\hat{z}$ to give

$$
E[\sqrt{X}] = \frac{1}{4\sqrt{\pi}i}\int_{\hat{z}-i\infty}^{\hat{z}+i\infty} e^{\kappa(z)-\frac{3}{2}\log z}\,\mathrm{d}z.
$$

The Taylor expansion of the exponent in the above integrand at $z = \hat{z}$ is seen to be

$$
\kappa(z) - \frac{3}{2}\log z = \kappa(\hat{z}) - \frac{3}{2}\log\hat{z} + \left[\kappa''(\hat{z}) + \frac{3}{2\hat{z}^2}\right]\frac{(z-\hat{z})^2}{2} + \cdots.
$$

Substituting the above Taylor expansion into the Bromwich integral and performing evaluation of the resulting Gaussian integral yields the first order saddlepoint approximation formula:

$$
\begin{aligned}
E[\sqrt{X}] &\approx \frac{1}{4\sqrt{\pi}i}\int_{\hat{z}-i\infty}^{\hat{z}+i\infty} e^{\kappa(\hat{z})-\frac{3}{2}\log\hat{z}+\left[\kappa''(\hat{z})+\frac{3}{2\hat{z}^2}\right]\frac{(z-\hat{z})^2}{2}}\,\mathrm{d}z\\
&= \frac{1}{4\sqrt{\pi}} e^{\kappa(\hat{z})-\frac{3}{2}\log\hat{z}}\int_{-\infty}^{\infty} e^{-\left[\kappa''(\hat{z})+\frac{3}{2\hat{z}^2}\right]\frac{y^2}{2}}\,\mathrm{d}y\\
&= \frac{\sqrt{2}}{4}\frac{e^{\kappa(\hat{z})}/\hat{z}^{3/2}}{\sqrt{\kappa''(\hat{z})+\frac{3}{2\hat{z}^2}}}.
\end{aligned}
$$

To obtain the second order saddlepoint approximation formula, we perform the Taylor expansion of the exponent $\kappa(z) - \frac{3}{2}\log z$ around the saddlepoint $\hat{z}$ up to the fourth order. This gives

$$
\begin{aligned}
e^{\kappa(z)-\frac{3}{2}\log z} &\approx e^{\kappa(\hat{z})-\frac{3}{2}\log\hat{z}+\left[\kappa''(\hat{z})+\frac{3}{2\hat{z}^2}\right]\frac{(z-\hat{z})^2}{2}+\left[\kappa'''(\hat{z})-\frac{3}{\hat{z}^3}\right]\frac{(z-\hat{z})^3}{6}+\left[\kappa''''(\hat{z})+\frac{9}{\hat{z}^4}\right]\frac{(z-\hat{z})^4}{24}} \\
&= e^{\kappa(\hat{z})-\frac{3}{2}\log\hat{z}}e^{\left[\kappa''(\hat{z})+\frac{3}{2\hat{z}^2}\right]\frac{(z-\hat{z})^2}{2}}\left\{e^{\left[\kappa'''(\hat{z})-\frac{3}{\hat{z}^3}\right]\frac{(z-\hat{z})^3}{6}+\left[\kappa''''(\hat{z})+\frac{9}{\hat{z}^4}\right]\frac{(z-\hat{z})^4}{24}}\right\} \\
&\approx e^{\kappa(\hat{z})-\frac{3}{2}\log\hat{z}}e^{\left[\kappa''(\hat{z})+\frac{3}{2\hat{z}^2}\right]\frac{(z-\hat{z})^2}{2}}\left\{1+\left[\kappa'''(\hat{z})-\frac{3}{\hat{z}^3}\right]\frac{(z-\hat{z})^3}{6}\right. \\
&\quad \left.+\left[\kappa''''(\hat{z})+\frac{9}{\hat{z}^4}\right]\frac{(z-\hat{z})^4}{24}+\left[\kappa'''(\hat{z})-\frac{3}{\hat{z}^3}\right]^2\frac{(z-\hat{z})^6}{72}\right\}.
\end{aligned}
$$

By letting $z = \hat{z} + iy$ and substituting the above Taylor expansion into the Bromwich integral, we obtain

$$
\begin{aligned}
&E[\sqrt{X}] \\
&\approx \frac{1}{4\sqrt{\pi}i}\int_{\hat{z}-i\infty}^{\hat{z}+i\infty} e^{\kappa(\hat{z})-\frac{3}{2}\log\hat{z}}e^{\left[\kappa''(\hat{z})+\frac{3}{2\hat{z}^2}\right]\frac{(z-\hat{z})^2}{2}}\left\{1+\left[\kappa'''(\hat{z})-\frac{3}{\hat{z}^3}\right]\frac{(z-\hat{z})^3}{6}\right. \\
&\qquad \left.+\left[\kappa''''(\hat{z})+\frac{9}{\hat{z}^4}\right]\frac{(z-\hat{z})^4}{24}+\left[\kappa'''(\hat{z})-\frac{3}{\hat{z}^3}\right]^2\frac{(z-\hat{z})^6}{72}\right\}\mathrm{d}z \\
&= \frac{1}{4\sqrt{\pi}}e^{\kappa(\hat{z})-\frac{3}{2}\log\hat{z}}\int_{-\infty}^{\infty} e^{-\left[\kappa''(\hat{z})+\frac{3}{2\hat{z}^2}\right]\frac{y^2}{2}}\left\{1-\left[\kappa'''(\hat{z})-\frac{3}{\hat{z}^3}\right]\frac{iy^3}{6}+\left[\kappa''''(\hat{z})+\frac{9}{\hat{z}^4}\right]\frac{y^4}{24}\right. \\
&\qquad \left.-\left[\kappa'''(\hat{z})-\frac{3}{\hat{z}^3}\right]^2\frac{y^6}{72}\right\}\mathrm{d}y, \\
&= \frac{1}{4\sqrt{\pi}}\frac{e^{\kappa(\hat{z})-\frac{3}{2}\log\hat{z}}}{\sqrt{\kappa''(\hat{z})+\frac{3}{2\hat{z}^2}}}\int_{-\infty}^{\infty} e^{-\frac{w^2}{2}}\left\{1-\frac{\kappa'''(\hat{z})-\frac{3}{\hat{z}^3}}{\left[\kappa''(\hat{z})+\frac{3}{2\hat{z}^2}\right]^{3/2}}\frac{iw^3}{6}+\frac{\kappa''''(\hat{z})+\frac{9}{\hat{z}^4}}{\left[\kappa''(\hat{z})+\frac{3}{2\hat{z}^2}\right]^2}\frac{w^4}{24}\right. \\
&\qquad \left.-\frac{\left[\kappa'''(\hat{z})-\frac{3}{\hat{z}^3}\right]^2}{\left[\kappa''(\hat{z})+\frac{3}{2\hat{z}^2}\right]^3}\frac{w^6}{72}\right\}\mathrm{d}w,
\end{aligned}
$$

where $w = \sqrt{\kappa''(\hat{z})+\frac{3}{2\hat{z}^2}}\,y$. By observing the following identities:

$$
\begin{aligned}
&\int_{-\infty}^{\infty} e^{-w^2/2}\,\mathrm{d}w = \sqrt{2\pi}, \quad \int_{-\infty}^{\infty} w^3 e^{-w^2/2}\,\mathrm{d}w = 0, \\
&\int_{-\infty}^{\infty} w^4 e^{-w^2/2}\,\mathrm{d}w = 3\sqrt{2\pi}, \quad \int_{-\infty}^{\infty} w^6 e^{-w^2/2}\,\mathrm{d}w = 15\sqrt{2\pi},
\end{aligned}
$$

we obtain

$$E[\sqrt{X}] \approx \frac{\sqrt{2}}{4} \frac{e^{\kappa(\hat{z})}/\hat{z}^{3/2}}{\sqrt{\kappa''(\hat{z}) + \frac{3}{2\hat{z}^2}}} \left\{ 1 + \frac{1}{8} \frac{\kappa''''(\hat{z}) + \frac{9}{\hat{z}^4}}{\left[\kappa''(\hat{z}) + \frac{3}{2\hat{z}^2}\right]^2} - \frac{5}{24} \frac{\left[\kappa'''(\hat{z}) - \frac{3}{\hat{z}^3}\right]^2}{\left[\kappa''(\hat{z}) + \frac{3}{2\hat{z}^2}\right]^3} \right\}.$$

Proof of Lemma 4.1

We consider

$$\begin{aligned} E\left[(\sqrt{X} - K)^+\right] &= \int_{K^2}^{\infty} (\sqrt{x} - K) p(x)\, \mathrm{d}x \\ &= \int_{K^2}^{\infty} (\sqrt{x} - K) \frac{1}{2\pi i} \int_{\gamma - i\infty}^{\gamma + i\infty} e^{\kappa(z) - zx}\, \mathrm{d}z\, \mathrm{d}x, \\ &= \frac{1}{2\pi i} \int_{\gamma - i\infty}^{\gamma + i\infty} e^{\kappa(z)} \left(\int_{K^2}^{\infty} \sqrt{x} e^{-zx}\, \mathrm{d}x - \int_{K^2}^{\infty} K e^{-zx}\, \mathrm{d}x \right) \mathrm{d}z. \end{aligned}$$

By imposing $\gamma = \mathrm{Re}\, z > 0$, the two inner integrals can be found to be

$$\begin{aligned} I_1 &= \int_K^{\infty} y e^{-zy^2} 2y\, \mathrm{d}y = \frac{K}{z} e^{-zK^2} + \frac{1}{z} \frac{1}{\sqrt{z}} \frac{\sqrt{\pi}}{2} \left[1 - \mathrm{erf}(\sqrt{z} K)\right], \\ I_2 &= -\frac{K}{z} e^{-zx} \Big|_{x=K^2}^{\infty} = \frac{K}{z} e^{-zK^2}. \end{aligned}$$

Putting these results together, we obtain (4.20) in Lemma 4.1.

Chapter 5
Saddlepoint Approximation for Credit Portfolios

We consider a portfolio of loans or bonds, where the loan borrowers or bond issuers may fail to meet the promised cashflows as stated in the loan contracts or bond indentures. These payment defaults lead to credit losses to the holder of the portfolio of these credit instruments or names (loans or bonds). In this chapter, we focus on the credit risk arising from the possibility of default of credit instruments in a credit portfolio. Credit risk is different from other types of risks, like the market risk arising from interest rate fluctuation that affects bond prices in the portfolio. The credit risk models for analyzing the default behavior of a credit obligor are mostly characterized by the occurrence of default through a Bernuolli random variable or the time of default through a positive random variable. For portfolio risk of a basket of credit instruments, we also explore the nature of dependencies of defaults. Correlated defaults occur since corporates operate under common economic environment, or in the same industry or geographical region. Naturally, these corporates are subject to common global and local economic and financial shocks.

We consider the total random loss L from defaults of n credit obligors in a credit portfolio over a time horizon T. Let D_i denote the default indicator for the ith obligor, $i = 1, 2, \ldots, n$, where D_i is a Bernuolli random variable defined by

$$D_i = \begin{cases} 1 \text{ if the } i\text{th obligor defaults within the time horizon } T \\ 0 \text{ if otherwise} \end{cases}$$

The default of the ith obligor is also characterized by the random time of default τ_i, $i = 1, 2, \ldots, n$. For a fixed time horizon T, it is seen that $D_i = \mathbf{1}_{\{\tau_i \leq T\}}$. Let p_i be the marginal probability of default of the ith obligor, then $p_i = P[D_i = 1]$. Let w_i be effective exposure of the defaultable ith obligor. Assuming independence between random losses upon default and default events, the random default loss of the portfolio with n credit obligors within the time horizon T is given by

Y. K. Kwok and W. Zheng, *Saddlepoint Approximation Methods in Financial Engineering*, SpringerBriefs in Quantitative Finance, https://doi.org/10.1007/978-3-319-74101-7_5

$$L = \sum_{i=1}^{n} w_i D_i. \tag{5.1}$$

Since the saddlepoint approximation approach provides accurate estimates to tail probabilities and tall expectations, so it is a convenient tool used in computing risk measures of credit portfolio losses. Some of the earlier works on the use of saddlepoint approximation methods in credit portfolio theory can be found in Martin et al. (2001); Martin (2006, 2013); Giese (2006); Huang et al. (2007); Owen et al. (2009). The more recent comprehensive studies and review of the methodologies are provided by Broda and Paolella (2012); Martin (2013); García-Céspedes (2014); Kim and Kim (2017). In this chapter, we first review the two commonly used default correlation models, CreditRisk$^+$ and the Gaussian copula model, for modeling correlated risk exposures of portfolios of risky obligors. We then discuss how to apply the saddlepoint approximation methods for finding the risk contributions to the two most common risk measures: Value-at-Risk and expected shorfall. We also consider pricing of the tranches of Collateralized Debts Obligations using the saddlepoint approximation approach.

5.1 Default Correlation Models

In risk management of a credit portfolio, the first challenge is the determination of the distribution of the portfolio loss L under correlated defaults of the obligors in the credit portfolio. In simple sense, default correlation means how the default of one obligor may affect the likelihood of defaults of other obligors in the credit portfolio. In the literature, there have been numerous default correlation models that are derived based on various mechanisms that drive the default correlation. In the next two subsections, we discuss the two common default correlation models: CreditRisk$^+$ and Gaussian copula models.

5.1.1 CreditRisk$^+$

CreditRisk$^+$ is an industrial code developed by Credit Suisse Financial Products (Wilde 1997) that applies actuarial mathematics to calculate the loss distribution of a credit portfolio. A collection of research articles on the methodologies and applications of CreditRisk$^+$ can be found in the text edited by Gundlach and Gundlach and Lehrbass (2004). CreditRisk$^+$ requires a limited number of input data and assumptions, and no profits or losses from rating migrations are considered. It provides an analytic based portfolio approach without the use of simulation, so unambiguity of the loss distribution is ensured. Indeed, the resulting portfolio loss distribution can be expressed as a sum of independent negative binomial random variables. Also, efficient evaluation procedures, like the analytic saddlepoint approximation method, are available for fast calculations of the loss distribution.

Summary of assumptions

1. CreditRisk$^+$ uses the Poisson approximation to the Bernuolli event of default. This implicitly allows multiple defaults of a single credit obligor. Given that the expected default probabilities are small, the chance of multiple defaults is extremely low, so the Poisson approximation is well acceptable.
2. CreditRisk$^+$ does not provide any financial modeling for defaults. Instead, default of an obligor is visualized as a purely random event and characterized by the probability of default. Recovery rates are absorbed into the exposure upon default and they are assumed to be independent of default events. The exposures in the credit portfolio are grouped into various bands. Each exposure amount is rounded to the nearest integer multiple of a basic unit of exposure.
3. To generate default correlation among the obligors, CreditRisk$^+$ assumes the probability of default of an obligor to be random and driven by one or more systematic risk factors. We assume conditional independent defaults, where the defaults of obligors are independent given the risk factors. That is, CreditRisk$^+$ allows implicit correlation among defaults via the risk drivers. A linear relationship between the systematic risk factors and the default probability is assumed. The distribution of the default intensity is usually assumed to be a gamma distribution to achieve nice analytical tractability.

Poisson approximationto Bernuolli events of defaults

The probability generating function (pgf) of a discrete non-negative integer-valued random variable K is a function of the auxiliary variable z such that the probability that $K = k$ is given by the coefficient of z^k in the polynomial expansion of the pgf. Formally, we define pgf of K by

$$G_K(z) = E[z^K] = \sum_{k=0}^{\infty} P[K = k]z^k. \tag{5.2}$$

Here, z is complex with $|z| \leq 1$ (technical condition that guarantees convergence of the infinite series). For example, the pgf of the Poisson random variable N with parameter α is

$$G_N(z) = e^{\alpha(z-1)} = e^{-\alpha} \sum_{n=0}^{\infty} \frac{(\alpha z)^n}{n!},$$

which is seen to be consistent with the following probability mass function of N, where

$$P[N = n] = \frac{e^{-\alpha}\alpha^n}{n!}, \quad n = 0, 1, 2, \ldots.$$

As another example, consider the Bernuolli random variable D that is set equal to the indicator function of default event with default probability p, the corresponding pgf is given by

$$G_D(z) = (1 - p) + pz = 1 + p(z - 1).$$

For the random number of defaults in the whole credit portfolio of obligors, the corresponding pgf is

$$G(z) = \sum_{n=0}^{\infty} P[n \text{ defaults}] z^n.$$

Assuming independence between default events, the pgf of the number of defaults in a credit portfolio is given by

$$G(z) = \prod_i [1 + p_i(z - 1)],$$

where p_i is the default probability of obligor i and i runs for all obligors in the credit portfolio. It is known that $\log(1 + z) \approx z$, so when p_i is small, we have

$$\log(1 + p_i(z - 1)) \approx p_i(z - 1),$$

and

$$\log G(z) \approx \sum_i p_i(z - 1).$$

We observe

$$G(z) \approx e^{\sum_i p_i(z-1)} = e^{\mu(z-1)} = \sum_{n=0}^{\infty} \frac{e^{-\mu}\mu^n}{n!} z^n, \tag{5.3}$$

where $\mu = \sum_i p_i$ is the expected number of defaults over the time horizon T for the whole credit portfolio. The above analytic approximation can be interpreted as the Poisson approximation of the number of defaults of the obligors, where $G(z)$ is approximated by the pgf of a Poisson random variable with mean μ. Though multiple defaults of single obligor is possible according to the Poisson assumption of default events, the chance of two or more defaults of obligor i is highly unlikely when p_i is small.

Sector risk factors

The correlation of defaults in CreditRisk$^+$ is assumed to be driven by a vector $\mathbf{X} = (X_1, X_2, \ldots, X_M)$ of M sector risk factors that represent the random fluctuations of macroeconomic conditions with respect to different industries or geographical regions. Conditional on a particular realization of the sector risk factors $\mathbf{X}$, defaults of the obligors within the credit portfolio are assumed to be independent.

The highest level of default correlation is resulted if all obligors are allocated to single factor, while increasing the number of sectors reduces the correlation.

In general, we consider the generalized M-factor sector risk specification, we write $\mathbf{x} = (x_1, x_2, \ldots, x_M)$. The probability of default of obligor i conditional on the realization of the M risk factors, where $(X_1, X_2, \ldots, X_M) = (x_1, x_2, \ldots, x_M)$, is given by

$$p_i(\mathbf{x}) = \bar{p}_i \left(w_{i_0} + \sum_{m=1}^{M} w_{i_m} x_m \right). \tag{5.4}$$

Here, $\bar{p}_i$ is the unconditional default probability of obligor i. To ensure that $E[p_i(\mathbf{x})] = \bar{p}_i$, we require

$$\sum_{m=0}^{M} w_{i_m} = 1$$

for each obligor i. In CreditRisk$^+$, the risk factors $X_m, m = 1, 2, \ldots, M$, are assumed to be independent gamma random variables. The factor loading vector $(w_{i_1}, \ldots, w_{i_m})$ for obligor i measures the sensitivity of default of obligor i to each of the risk factor X_m, $m = 1, 2, \ldots, M$. The constant w_{i_0} represents the loading on the idiosyncratic risk of obligor i. Based on the Poisson approximation, the conditional pgf of default of obligor i is approximated by

$$G_i(z|\mathbf{x}) = 1 + p_i(\mathbf{x})(z-1) \approx \exp(p_i(\mathbf{x})(z-1)),$$

assuming that $p_i(\mathbf{x})$ is small. We are interested in the loss amount from the whole credit portfolio in units of a standardized loss amount (say, ten thousand dollars). Suppose the dollar loss amount as integer multiple of the basic exposure upon default of obligor i is denoted by v_i, the corresponding pgf for units of dollar loss on obligor i is then given by

$$F_i(z|\mathbf{x}) = G_i(z^{v_i}|\mathbf{x}). \tag{5.5}$$

The rounding of loss amounts to integer units does not significantly affect the aggregate risk assessment.

By the assumption of conditional independence of defaults, the pgf of the sum of independent random loss variables is the product of the individual pgfs, so the conditional pgf of the portfolio loss is given by

$$\begin{aligned} F(z|\mathbf{x}) &= \prod_i F_i(z|\mathbf{x}) \approx \exp\left(\sum_i p_i(\mathbf{x})(z^{v_i} - 1) \right) \\ &= \exp\left(\sum_i \bar{p}_i \left(w_{i_0} + \sum_{m=1}^{M} w_{i_m} x_m \right) (z^{v_i} - 1) \right) \end{aligned}$$

$$= \exp\left(\sum_i \bar{p}_i w_{i_0}(z^{v_i} - 1)\right) \exp\left(\sum_{m=1}^{M} x_m \left[\sum_{m=1}^{M} \bar{p}_i w_{i_m}(z^{v_i} - 1)\right]\right)$$

$$= \exp(\mu_0[P_0(z) - 1]) \exp\left(\sum_{m=1}^{M} x_m \mu_m [P_m(z) - 1]\right), \tag{5.6}$$

where $\mu_m = \sum_i w_{i_m} \bar{p}_i$ and $P_m(z) = \frac{\sum_i w_{i_m} \bar{p}_i z^{v_i}}{\mu_m}$, $m = 0, 1, \ldots, M$. To find the unconditional pgf for portfolio loss, one has to integrate over the distribution of the vector risk factor $\mathbf{X}$.

In CreditRisk$^+$, these sector risk factors are assumed to follow the gamma distribution and they are independent. Suppose X_m follows the gamma distribution with parameters α_m and β_m, the corresponding density function is given by

$$P[x_m < X_m < x_m + \mathrm{d}x_m] = f_m(x_m)\,\mathrm{d}x_m = \frac{e^{-x_m/\beta_m} x_m^{\alpha_m - 1}}{\beta_m^{\alpha_m} \gamma(\alpha_m)}\,\mathrm{d}x_m,$$

where the gamma function is defined by

$$\gamma(\alpha_m) = (\alpha_m - 1)! = \int_0^\infty e^{-x} x^{\alpha_m - 1} \mathrm{d}x.$$

The mean ν_m and variance σ_m^2 of X_m are found to be $\alpha_m \beta_m$ and $\alpha_m \beta_m^2$, respectively. These give $\alpha_m = \nu_m^2/\sigma_m^2$ and $\beta_m = \sigma_m^2/\nu_m$. The unconditional pgf of the portfolio loss in units of a standardized loss amount is given by

$$F(z) = \int_0^\infty \cdots \int_0^\infty F(z|\mathbf{x}) f_1(x_1) \cdots f_M(x_M)\,\mathrm{d}x_1 \cdots \mathrm{d}x_M. \tag{5.7}$$

Note that

$$\int_0^\infty e^{x(z-1)} \frac{e^{-x/\beta} x^{\alpha - 1}}{\beta^\alpha \gamma(\alpha)}\,\mathrm{d}x = \left(\frac{1 - Q}{1 - Qz}\right)^\alpha,$$

where

$$Q = \frac{\beta}{1 + \beta},\ 1 - Q = \frac{1}{1 + \beta} \text{ and } 1 - Qz = \frac{1 + \beta(1 - z)}{1 + \beta}.$$

By summing all the results together, the unconditional pgf of portfolio loss is given by

$$F(z) = \exp(\mu_0[P_0(z) - 1]) \prod_{m=1}^{M} \left[\frac{1 - Q_m}{1 - Q_m P_m(z)}\right]^{\alpha_m}, \tag{5.8}$$

where $Q_m = \frac{\sigma_m^2}{\sigma_m^2 + \nu_m}$.

We are interested in calculating the distribution of the loss amount in a credit portfolio. After bucketing of exposures, the aggregate loss amount from the portfolio would be a positive integer multiple of the basic unit. We define

$$F(z) = \sum_{l=0}^{l_{\max}} A_l z^l,$$

where $l_{\max}$ is the maximum number of loss units. The probability that the portfolio loss amount equals ℓ units of the standardized loss amount is given by the coefficient A_l of z^ℓ in the Taylor expansion of $F(z)$. The CreditRisk$^+$ manual (Wilde 1997) outlines the so-called Panjer recursion relation that relates A_l and A_{l+1} for computing these Taylor coefficients, starting from $A_0 = F(0)$. Unfortunately, the Panjer recursion calculation procedure generally suffers from numerical instability due to accumulation of roundoff errors, as exemplified by negative values of the calculated probabilities at the far tail. Giese (2003) proposes an alternative recurrence relation that helps reduce numerical instabilities. An extension of the CreditRisk$^+$ to compound gamma distribution of the sector risk factors can be found in Ebmeyer (2005). In Sect. 5.2, we discuss how to apply the saddlepoint approximation method for an affective evaluation of the distribution of portfolio losses.

5.1.2 Gaussian Copula Models

Another popular default correlation model is the Gaussian copula model, which captures dependence among the default indicators $D_1, D_2, \ldots, D_n$ through a Gaussian copula function that links the vector of credit variables $(Y_1, Y_2, \ldots, Y_n)$ of the n obligors in the portfolio. The Gaussian copula model was first introduced by Li (2000) and it is seen to be the analytic framework of another industrial default correlation model CreditMetrics. There are two steps in the modeling of the joint distribution of the default times of several obligors in a credit portfolio. Based on the default dynamics of individual obligors, the marginal distributions of default times can be modeled through credit spreads and hazard rates of defaults. The challenge in modeling the dependence structure of defaults is to specify a joint distribution of survival times, based on the corresponding marginal distributions.

A copula function links univariate marginals to their full multivariate distribution. A function $C : [0, 1]^n \to [0, 1]$ is a copula function if

(i) there are random variables $U_1, U_2, \ldots, U_n$ taking values in [0, 1] such that C is their distribution function;
(ii) C has uniform marginal distributions, where

$$C(1, \ldots, 1, u_i, 1, \ldots, 1) = u_i$$

for $i = 1, 2, \ldots, n$ and $u_i \in [0, 1]$.

In the analysis of dependence with a copula function, the joint distribution can be separated into two parts: the marginals of the random variables and the dependence structure between the random variables that is prescribed by the copula function.

We state the following results on the relation between the multivariate distribution functions and copula functions.

1. Given the copula function C and the set of univariate marginal distribution $F_1(x_1), F_2(x_2), \ldots, F_n(x_n)$, the function which is defined by
$$C(F_1(x_1), F_2(x_2), \ldots, F_n(x_n)) = F(x_1, x_2, \ldots, x_n) \tag{5.9}$$
results in a multivariate distribution function with univariate marginal distribution specified as $F_1(x_1), F_2(x_2), \ldots, F_n(x_n)$.
2. As a converse of the above result, any multivariate distribution F can be written in the form of a copula function as stated in (5.9). If each of these marginals $F_i(x_i)$ is continuous, then C is unique. This result is known as the Sklar theorem. This theorem is only an existence proof. The actual finding of the copula function for a given joint multivariate distribution can be very cumbersome.

The Gaussian copula function is the most convenient choice since it is completely specified by a correlation matrix. We let $\Phi(x_1, x_2, \ldots, x_n; \Omega)$ denote the n-dimensional normal distribution function with correlation matrix Ω in which all marginals have zero mean and unit variance (standard normal random variable). The corresponding Gaussian copula is defined by

$$C(\Phi(x_1), \Phi(x_2), \ldots, \Phi(x_n)) = \Phi(x_1, x_2, \ldots, x_n; \Omega) \tag{5.10a}$$

or

$$C(u_1, u_2, \ldots, u_n) = \Phi(\Phi^{-1}(u_1), \Phi^{-1}(u_2), \ldots, \Phi^{-1}(u_n); \Omega), \tag{5.10b}$$

where $u_i \in [0, 1]$, $i = 1, 2, \ldots, n$. Let τ_i be the random default time of obligor i and $F_i(t_i)$ be the corresponding marginal distribution, $i = 1, 2, \ldots, n$, the Gaussian copula model for the joint distribution of the default times is given by

$$P[\tau_1 \le t_1, \tau_2 \le t_2, \ldots, \tau_n \le t_n] = C(F_1(t_1), F_2(t_2), \ldots, F_n(t_n)). \tag{5.11}$$

Single-period default-mode model

We consider the single-period default-mode model, where default loss occurs when an obligor defaults within a fixed time horizon. By following the framework of the Merton model of individual defaults, the default indicator variable D_i and the credit variable Y_i of obligor i are related by

$$D_i = \mathbf{1}_{\{Y_i > y_i\}}, \quad i = 1, 2, \ldots, n. \tag{5.12}$$

Here, the default threshold y_i is determined such that $E[D_i] = p_i$, where p_i is the marginal default probability of the ith obligor. We assume that these individual default probabilities can be inferred from credit ratings, market prices of corporate bonds or credit default swaps. In the Gaussian copula model, we choose Y_i to be a standard normal variable, then $y_i = -\Phi^{-1}(p_i)$, where $\Phi(\cdot)$ is the standard normal distribution function. As a check, we consider

$$P[D_i = 1] = P[Y_i > -\Phi^{-1}(p_i)] = \Phi(\Phi^{-1}(p_i)) = p_i.$$

We assume that the credit variables $(Y_1, Y_2, \ldots, Y_n)$ are driven by a set of random factor X_m, $m = 1, 2, \ldots, M$, and Z_i, where

$$Y_i = \gamma_i Z_i + \sum_{m=1}^{M} w_{im} X_m, \quad i = 1, 2, \ldots, n. \tag{5.13}$$

Here, $X_1, X_2, \ldots, X_M$ are the systematic risk factors and Z_i is the idiosyncratic risk factor of obligor i, and $\gamma_i, w_{i1}, \ldots, w_{iM}$ are the nonnegative factor loadings. In the Gaussian copula model, these random risk factors are assumed to follow the standard normal distribution. In order that Y_i remains to be standard normal, the factor loadings observe the normalization relation

$$\gamma_i^2 + \sum_{m=1}^{M} w_{im}^2 = 1.$$

We let W denote the factor loading matrix whose (i, m)th entry is w_{im}, where W is a $N \times M$ matrix. Therefore, the credit variables $(Y_1, Y_2, \ldots, Y_n)$ are multivariate normally distributed. Let $\boldsymbol{\epsilon}$ denote the vector of M independent standard normal variables, where $E[\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T] = I$. The correlation matrix Ω of the credit variables is given by

$$\Omega = E[W\boldsymbol{\epsilon}(W\boldsymbol{\epsilon})^T] = WE[\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T]W^T = WW^T.$$

Here, the correlation between Y_k and Y_j, $j \neq k$, is given by the (k, j)th entry of WW^T. Given the realization of the risk factors $\mathbf{X} = (X_1, X_2, \ldots, X_M)$ and Z_i, the conditional default probability of the ith obligor is given by

$$\begin{aligned} p_i(\mathbf{X}) &= P[D_i = 1 | \mathbf{X}, Z_i] = P[Y_i > y_i | \mathbf{X}, Z_i] \\ &= P[\gamma_i Z_i + w_{i1}X_1 + w_{i2}X_2 + \cdots + w_{iM}X_M > y_i] \\ &= P[Z_i > \frac{-\Phi^{-1}(p_i) - w_{i1}X_1 - w_{i2}X_2 - \cdots - w_{iM}X_M}{\gamma_i}]. \end{aligned} \tag{5.14}$$

Note that both p_i and $p_i(\mathbf{X})$ appear in (5.14), where p_i is the unconditional default probability while $p_i(\mathbf{X})$ is the conditional default probability of obligor i.

5.2 Risk Measures and Risk Contributions

We first introduce the two most common risk measures for the tail distribution of loss in a credit portfolio, namely, the Value-at-Risk (VaR) and expected shortfall (ES). The major weakness of VaR is that it fails to be a coherent risk measure while ES satisfies the coherent requirements. We discuss the risk contributions of individual obligors to the risk measures and show how to apply the saddlepoint approximation methods to compute the risk contributions of obligors in a credit portfolio. The saddlepoint approximations are shown to provide effective approximate evaluations of VaR and ES under the CreditRisk$^+$ model and Gaussion copula models.

5.2.1 Value-at-Risk and Expected Shortfall

We present the definitions and properties of Value-at-Risk (VaR) and expected shortfall (ES) of a credit portfolio, in particular, the desirable property of subadditivity that is linked to diversification. It is desirable for a risk measure to satisfy a set of so-called coherent properties, like subadditivity. It is observed that VaR is not a coherent risk measure since it fails to satisfy subadditivity [see (5.17)]. On the other hand, ES is seen to be the more desirable risk measure since it is coherent. However, practical evaluation of ES of a credit portfolio may pose various technical challenges since the loss distribution of a credit portfolio beyond a critical threshold is not easily quantified. We present an interesting relation between VaR and ES and the saddlepoint approximation formula for calculating ES.

Value-at-Risk

Given the probability measure P and some chosen confidence level α, VaR of a credit portfolio is defined as the α-quantile of the portfolio loss random variable L, where

$$\mathrm{VaR}_\alpha(L) = \inf\{l \geq 0 | P[L \leq l] \geq \alpha\}. \tag{5.15}$$

In other words, VaR_α is the maximum loss that is not exceeded with a given high probability α. Note that VaR is a risk measure defined on the space of bounded real random variables since portfolio default losses are bounded. Let $F_L(l)$ denote the cumulative distribution function of L. The intuitive interpretation of $\mathrm{VaR}_\alpha(L)$ is given by

$$F_L(\mathrm{VaR}_\alpha(L)) = \alpha. \tag{5.16}$$

Since VaR is a quantile, it satisfies the following properties:

(i) *Monotonicity*

$$\text{For } L \leq Y, \text{ then } \mathrm{VaR}(L) \leq \mathrm{VaR}(Y).$$

A smaller loss variable should have a lower value of VaR.

(ii) *Positive homogeneity*

$$\text{VaR}(\lambda L) = \lambda \text{VaR}(L), \quad \lambda > 0.$$

Multiplying the loss variable by a positive scalar λ changes VaR by the same scalar multiplier.

(iii) *Translation invariance*

$$\text{VaR}(L + x) = \text{VaR}(L) + x$$

Shifting the loss variable L by a fixed amount x shifts the quantile by the same amount accordingly.

One desirable property to be observed for a risk measure γ is *subadditivity*. For a pair of loss variables L and Y, subadditivity means

$$\gamma(L + Y) \leq \gamma(L) + \gamma(Y). \tag{5.17}$$

Note that subadditivity is related to the benefit of diversification, where the risk in the combination of two portfolios should not be greater than the sum of risks of the two portfolios taken separately. Artzner et al. (1999) defines a risk measure to be coherent if it satisfies subadditivity, monotonicity, positive homogeneity and translation invariance. Unfortunately, VaR is not coherent since it does not satisfy subadditivity. The failure of subadditivity weakens the desirability to use VaR as a risk measure. Examples of credit portfolios where VaR_α fails the property of subadditivity can be found in Artzner et al. (1999).

One common use of VaR_α is the determination of the economic capital EC_α at the confidence level α (say, $\alpha = 99\%$) over a specified time horizon (say, 30 days). While credit reserves are set aside to absorb expected losses (EL), we set

$$\text{EC}_\alpha = \text{VaR}_\alpha - \text{EL}, \tag{5.18}$$

where the economic capital EC_α is designed to absorb unexpected losses.

Expected shortfall

A risk manager may be interested to consider the expected loss in the tail of the distribution of loss L of a credit portfolio beyond a critical loss threshold defined by some confidence level α and the corresponding threshold VaR_α. For continuous loss distribution of L, the expected shortfall of a loss variable L for a given confidence level α is defined by

$$ES_\alpha(L) = E[L | L \geq \text{VaR}_\alpha] = \frac{1}{1-\alpha} E\left[L \mathbf{1}_{\{L \geq \text{VaR}_\alpha\}}\right] = \frac{1}{1-\alpha} \int_{\text{VaR}_\alpha}^{\infty} \ell f_L(\ell)\, d\ell, \tag{5.19a}$$

where $f_L(\ell)$ is the density function of L. It is shown in Artzner et al. (1999) that expected shortfall is a coherent risk measure if L has a continuous distribution. The relation between $ES(L)$ and $\mathrm{VaR}(L)$ can be found to be

$$ES_\alpha(L) = \frac{1}{1-\alpha}\int_\alpha^1 \mathrm{VaR}_u(L)\,\mathrm{d}u. \tag{5.19b}$$

The relation between the two integrals in (5.19a) and (5.19b) can be seen if we set $u = F_L(\mathrm{VaR}_u)$ and write ℓ as VaR_u so that $u = \alpha$ when $\ell = \mathrm{VaR}_\alpha$ and $u = 1$ when ℓ goes to ∞. Also, we observe that $\mathrm{VaR}_u(L)\,\mathrm{d}u = \ell f_L(\ell)\,\mathrm{d}\ell$.

Recall from (1.13e) that $E[L\mathbf{1}_{\{L\geq \mathrm{VaR}_\alpha\}}]$ admits the Bromwich integral representation

$$E[L\mathbf{1}_{\{L\geq \mathrm{VaR}_\alpha\}}] = \frac{1}{2\pi i}\int_{\gamma-i\infty}^{\gamma+i\infty} \kappa_L'(z)\frac{e^{\kappa_L(z)-zt}}{z}\,\mathrm{d}z, \quad \gamma \in (0,\alpha_+), \tag{5.20}$$

where t denotes VaR_α. The proof of (5.20) is presented in the Appendix.

Martin (2006) proposes to express the singularity in the integrand in (5.20) as the sum of an integrable singularity and a regular part near the singularity $z = 0$, where

$$\kappa_L'(z)\frac{e^{\kappa_L(z)-zt}}{z} = \left[\frac{\mu_L}{z} + \frac{\kappa_L'(z)-\mu_L}{z}\right]e^{\kappa_L(z)-zt}.$$

The integration of the first term gives

$$\frac{1}{2\pi i}\int_{\gamma-i\infty}^{\gamma+i\infty} \mu_L\frac{e^{\kappa_L(z)-zt}}{z}\,\mathrm{d}z = \mu_L\ P[L>t].$$

For the second term, we transform the Bromwich contour to pass through the saddlepoint $\hat{z}$, where $\kappa_L'(\hat{z}) = t$. We approximate the integration of the second term as follows:

$$\begin{aligned}\frac{1}{2\pi i}\int_{\hat{z}-i\infty}^{\hat{z}+i\infty} \frac{\kappa_L'(z)-\mu_L}{z}e^{\kappa_L(z)-zt}\,\mathrm{d}z &\approx \frac{t-\mu_L}{\hat{z}}\frac{1}{2\pi i}\int_{\hat{z}-i\infty}^{\hat{z}+i\infty} e^{\kappa_L(z)-zt}\,\mathrm{d}z\\ &= \frac{t-\mu_L}{\hat{z}}f_L(t).\end{aligned}$$

Combining the results together, Martin (2006) obtains the following saddlepoint approximation formula for ES:

$$E[L\mathbf{1}_{\{L\geq \mathrm{VaR}_\alpha\}}] \approx \mu_L P[L>t] + \frac{t-\mu_L}{\hat{z}}f_L(t), \quad t = \mathrm{VaR}_\alpha. \tag{5.21}$$

We may use the saddlepoint approximation of various orders to calculate $P[L > t]$ and $f_L(t)$ [Broda and Paolella (2011); Huang and Oosterlee (2011);

Kim and Kim (2017)]. Suppose the first order approximation formula for $P[L > t]$ is used [see (2.15)], we obtain

$$E[L\mathbf{1}_{\{L\geq \mathrm{VaR}_\alpha\}}] \approx \mu_L[1-\Phi(\hat{w})] + \phi(\hat{w})\left[\frac{t}{\hat{u}} - \frac{\mu_L}{\hat{w}}\right], \tag{5.22a}$$

where $\hat{w} = \mathrm{sgn}(\hat{z})\sqrt{2[\hat{z}t - \kappa_L(\hat{z})]}$ and $\hat{u} = \hat{z}\sqrt{\kappa_L''(\hat{z})}$. Taking the next higher order saddlepoint approximation [see (2.1) and (2.17a)], we obtain

$$\begin{aligned} &E[L\mathbf{1}_{\{L\geq \mathrm{VaR}_\alpha\}}] \\ &\approx \mu_L[1-\Phi(\hat{w})] + \phi(\hat{w})\left[\frac{t}{\hat{u}} - \frac{\mu_L}{\hat{w}} + \frac{\mu_L}{\hat{w}^3} - \frac{t}{\hat{u}^3} - \frac{t\hat{\rho}_3}{2\hat{u}^2} + \frac{t}{\hat{u}}\left(\frac{\hat{\rho}_4}{8} - \frac{5\hat{\rho}_3^2}{24}\right) + \frac{1}{\hat{z}\hat{u}}\right], \end{aligned} \tag{5.22b}$$

where $\hat{\rho}_3 = \frac{\kappa_L'''(\hat{z})}{\kappa_L''(\hat{z})^{3/2}}$ and $\hat{\rho}_4 = \frac{\kappa_L''''(\hat{z})}{\kappa_L''(\hat{z})^2}$. This higher order approximation formula (5.22b) for ES involves the third and fourth order derivatives of $\kappa_L(z)$. To derive an alternative saddlepoint approximation formula for ES that does not involve $\kappa_L'''(z)$ and $\kappa_L''''(z)$, we use the saddlepoint approximation formulas for mgf and its logarithmic derivatives derived in Butler and Wood (2004). The resulting saddlepoint approximation for ES is given by

$$\begin{aligned} &E[L\mathbf{1}_{\{L\geq \mathrm{VaR}_\alpha\}}] \\ &\approx \mu_L[1-\Phi(\hat{w})] + \phi(\hat{w})\left[\frac{t}{\hat{u}} - \frac{\mu_L}{\hat{w}} + \frac{\mu_L - t}{\hat{w}^3} + \frac{1}{\hat{z}\hat{u}}\right], \end{aligned} \tag{5.22c}$$

For discrete distribution of L, the generalized expected shortfall is defined by

$$ES_\alpha(L) = \frac{E[L\mathbf{1}_{\{L\geq \mathrm{VaR}_\alpha\}}] + \mathrm{VaR}_\alpha(L)\{\alpha - P[L < \mathrm{VaR}_\alpha(L)]\}}{1-\alpha}, \tag{5.23}$$

where the newly added second term in the numerator is a correction for the mass at the α-quantile of L. Under this definition of $ES_\alpha(L)$, it can be shown that expected shortfall is coherent for discrete distribution of L (Acerbi and Tasche 2002).

5.2.2 Risk Contributions

To make assessment on the contributions of individual obligors to the risk measure, like VaR and ES, we consider the sensitivity of the risk measure with respect to the weights of the obligors in the credit portfolio. A review of risk factor contributions in portfolio credit risk models can be found in Rosen and Saunders (2010).

Let w_i and D_i denote the effective exposure and credit default variable of obligor i, $i = 1, 2, \ldots, n$, where the portfolio loss L is $\sum_{i=1}^{n} w_i D_i$. We compute

$$\frac{\partial \mathrm{VaR}_\alpha}{\partial w_i} \quad \text{and} \quad \frac{\partial ES_\alpha}{\partial w_i}$$

as the sensitivity measures of risk contributions of obligor i to $\mathrm{VaR}_\alpha(L; w_1, \ldots, w_n)$ and $ES_\alpha(L; w_1, \ldots, w_n)$, respectively. It can be shown that both VaR_α and ES_α are linear homogeneous functions of the effective exposures of the obligors. By virtue of the Euler homogeneous function theorem,[1] we have

$$\mathrm{VaR}_\alpha = \sum_{i=1}^{n} w_i \frac{\partial \mathrm{VaR}_\alpha}{\partial w_i} = \sum_{i=1}^{n} \mathrm{RC}_i^{\mathrm{VaR}}(\alpha) \tag{5.24a}$$

$$ES_\alpha = \sum_{i=1}^{n} w_i \frac{\partial ES_\alpha}{\partial w_i} = \sum_{i=1}^{n} \mathrm{RC}_i^{ES}(\alpha), \tag{5.24b}$$

where $\mathrm{RC}_i^{\mathrm{VaR}}(\alpha) = w_i \dfrac{\partial \mathrm{VaR}_\alpha}{\partial w_i}$ and $\mathrm{RC}_i^{ES}(\alpha) = w_i \dfrac{\partial ES_\alpha}{\partial w_i}$, $i = 1, 2, \ldots, n$, are the marginal risk contributions of obligor i to VaR_α and ES_α, respectively.

To find $\mathrm{RC}_i^{\mathrm{VaR}}(\alpha)$, we examine the effect of changing w_i infinitesimally while keeping the tail probability $P[L > t]$ constant, where t is identified as VaR_α. We consider the derivative of $P[L > t]$ with respect to w_i and set it be zero. Recall (1.13b) which gives

$$P[L > t] = \frac{1}{2\pi i} \int_{\gamma - i\infty}^{\gamma + i\infty} \frac{e^{\kappa_L(z) - zt}}{z} \, \mathrm{d}z, \quad \gamma \in (0, \alpha_+),$$

so that

$$\frac{\partial}{\partial w_i} P[L > t] = \frac{1}{2\pi i} \int_{\gamma - i\infty}^{\gamma + i\infty} \left[\frac{1}{z} \frac{\partial \kappa_L(z)}{\partial w_i} - \frac{\partial t}{\partial w_i} \right] e^{\kappa_L(z) - zt} \, \mathrm{d}z.$$

[1] A multivariate function $f(w_1, w_2, \ldots, w_n)$ is said to satisfy the first order positive homogeneous property if

$$f(\lambda w_1,\ \lambda w_2,\ \ldots,\ \lambda w_n) = \lambda f(w_1, w_2, \ldots, w_n), \quad \lambda > 0.$$

By differentiating with respect to λ on both sides and setting $\lambda = 1$, we obtain

$$\sum_{i=1}^{n} w_i \frac{\partial f}{\partial w_i} = f.$$

.

Since $P[L > \mathrm{VaR}_\alpha]$ is fixed at $1-\alpha$, we set $\dfrac{\partial}{\partial w_i}P[L > t] = 0$ at $t = \mathrm{VaR}_\alpha$. This gives

$$\frac{\partial \mathrm{VaR}_\alpha}{\partial w_i} = \frac{\frac{1}{2\pi i}\int_{\gamma-i\infty}^{\gamma+i\infty} \frac{1}{z}\frac{\partial \kappa_L(z)}{\partial w_i} e^{\kappa_L(z)-z\mathrm{VaR}_\alpha}\,\mathrm{d}z}{\frac{1}{2\pi i}\int_{\gamma-i\infty}^{\gamma+i\infty} e^{\kappa_L(z)-z\mathrm{VaR}_\alpha}\,\mathrm{d}z}. \tag{5.25}$$

Note that $\kappa_L(z) = E[e^{z\sum_{i=1}^n w_i D_i}]$ and $L = \sum_{i=1}^n w_i D_i$, and so

$$\frac{\partial \kappa_L(z)}{\partial w_i} = \frac{\partial}{\partial w_i} E\left[e^{z\sum_{i=1}^n w_i D_i}\right] = zE[D_i e^{zL}].$$

We then obtain

$$\frac{\partial \mathrm{VaR}_\alpha}{\partial w_i} = \frac{\frac{1}{2\pi i}\int_{\gamma-i\infty}^{\gamma+i\infty} E[D_i e^{\kappa_L(z)+z(L-\mathrm{VaR}_\alpha)}]\,\mathrm{d}z}{f_L(\mathrm{VaR}_\alpha)} = E[D_i|L = \mathrm{VaR}_\alpha], \tag{5.26}$$

where f_L is the density function of L. The sum of the marginal risk contributions of all obligors to VaR_α is simply VaR_α, as revealed by

$$\sum_{i=1}^n w_i E[D_i|L = \mathrm{VaR}_\alpha] = E[\sum_{i=1}^n w_i D_i|L = \mathrm{VaR}_\alpha] = E[L|L = \mathrm{VaR}_\alpha] = \mathrm{VaR}_\alpha. \tag{5.27}$$

Martin et al. (2001) propose to approximate the marginal contribution to VaR_α by obligor i using the saddlepoint approximation method via the approximation of the ratio of the Bromwich integrals in (5.25). We obtain

$$w_i \frac{\partial \mathrm{VaR}_\alpha}{\partial w_i} \approx \frac{w_i}{\hat{z}} \frac{\partial \kappa_L(z)}{\partial w_i}\bigg|_{z=\hat{z}}, \tag{5.28}$$

where $\hat{z}$ is the saddlepoint obtained by solving the algebraic equation:

$$\frac{\partial \kappa_L(z)}{\partial z} = \mathrm{VaR}_\alpha.$$

This crude approximation method may not provide sufficient level of accuracy when the loss distribution is not unimodel.

As a remark, the estimation of the VaR contribution using Monte Carlo simulation may face with numerical instabilities since one has to condition on a precise level of portfolio loss. The strength of an analytic approximation approach is more obvious in calculating the VaR contribution. Numerical examples of calculating loss distributions and VaR contributions using the saddlepoint approximation and Monte Carlo simulation can be found in Martin et al. (2001) and Huang et al. (2007).

By following a similar derivation procedure, the marginal contribution to the expected shortfall can be shown to be

$$\frac{\partial ES_\alpha}{\partial w_i} = E[D_i|L \geq \text{VaR}_\alpha]. \tag{5.29}$$

As a verification to (5.24b), we observe

$$\sum_{i=1}^{n} w_i E[D_i|L \geq \text{VaR}_\alpha] = E[\sum_{i=1}^{n} w_i D_i|L \geq \text{VaR}_\alpha] = E[L|L \geq \text{VaR}_\alpha] = ES_\alpha. \tag{5.30}$$

5.2.3 *Risk Measures Calculations for Default Correlation Models*

We consider the numerical implementation and performance of various saddlepoint approximations in calculating the risk measures of portfolio losses as characterized by VaR and ES, and contributions by obligors under the two popular default correlation models: CreditRisk$^+$ and Gaussian copula models.

CreditRisk$^+$

The main advantage of CreditRisk$^+$ over other competing default correlation models is analytic tractability of the probability generating function (pgf) of the portfolio loss amount in integer multiples of the basic exposure. Once the pgf is known, the cumulant generating function (cgf) of the random loss distribution L is given by

$$\kappa_L(z) = \log(F(e^z)), \tag{5.31}$$

where the pgf of portfolio loss $F(z)$ is given by (5.8). Once $\kappa_L(z)$ is used to replace $F(z)$ in quantifying the loss distribution, it is not necessary to discretize the loss exposures into multiples of basic unit of exposure. Recall that the saddlepoint approximation formulas involve the higher order derivatives of the cgf. By taking advantage of the nice analytic form of $F(z)$, it is relatively straightforward to derive the analytic formulas for the derivatives: $\kappa'_L(z), \ldots, \kappa'''_L(z)$ [see (14a–b) in Gordy (2002)]. Gordy (2002) also shows mathematically that $\kappa_L(z)$ and all its higher order derivatives are positive, continuous, increasing and convex over $(-\infty, z^*)$, where $z^* > 0$. As $z \to z^*$, $\kappa_L(z)$ and all its derivatives tend to infinity. Also, as $z \to -\infty$, $\kappa'_L(z)$ and the higher order derivatives tend to zero. Hence, this guarantees the existence of the saddlepoint $\hat{z}$ that satisfies $\kappa'_L(\hat{z}) = \ell$, for $\ell > 0$.

The calculation of VaR_α at a given level of target solvency probability α is simply the inverse problem of computing the tail probability at a given loss level. By combining the Lugannani-Rice formula (2.15) [denoted by $LR(\text{VaR}_\alpha; \hat{w})$ in (2.15)] and the VaR equation (5.16), we obtain

$$LR(\text{VaR}_\alpha; \hat{w}) = 1 - \Phi(\hat{w}) + \phi(\hat{w}) \left[\frac{1}{\hat{z}\kappa_L''(\hat{z})^{1/2}} - \frac{1}{\hat{w}} \right]$$
$$\approx P[L > \text{VaR}_\alpha] = 1 - F_L(\text{VaR}_\alpha) = 1 - \alpha, \tag{5.32a}$$

where F_L is the cumulative distribution function for the random loss L, $\hat{w} = \text{sgn}(\hat{z})\sqrt{2[\hat{z}\text{VaR}_\alpha - \kappa_L(\hat{z})]}$ and the saddlepoint $\hat{z}$ satisfies

$$\kappa_L'(\hat{z}) = \text{VaR}_\alpha. \tag{5.32b}$$

It is necessary to solve the coupled nonlinear algebraic equations (5.32a, b) simultaneously for $\hat{z}$ and VaR_α since $LR(\text{VaR}_\alpha; \hat{w})$ involves $\hat{z}$. The solution for VaR_α requires an iterative search similar to the standard root-finding procedure. Since $\kappa_L(z)$ and $\kappa_L'(z)$ are smooth and strictly monotonic functions, the search procedure for finding VaR_α for a given α is quite straightforward and efficient. The numerical tests of Gordy (2002) reveal good performance of the LR saddlepoint approximation when compared with the Panger recursive scheme and Monte Carlo simulation, particularly for large credit portfolios where the number of obligors is more than 1000. Indeed, the Panger recursive scheme may fail due to roundoff errors when the number of obligors becomes large while the Monte Carlo simulation procedure may face with numerical instabilities in estimation due to uncertainties in simulating the loss amounts in simulation runs. In additional to VaR_α calculations, Haaf and Tasche (2002) provide details on the calculation of VaR contributions in CreditRisk$^+$.

Gordy (2004) proposes the following saddlepoint approximation procedure to estimate ES that requires two uses of the saddlepoint approximation. Recall the relation

$$ES_\alpha(L) = \frac{1}{1-\alpha} \int_{\text{VaR}_\alpha(L)}^{\infty} \ell f_L(\ell)\, \mathrm{d}\ell = \frac{\mu_L}{1-\alpha} \int_{\text{VaR}_\alpha(L)}^{\infty} \frac{\ell f(\ell)}{\mu_L}\, \mathrm{d}\ell,$$

where $f_L(\ell)$ and μ_L denote the density function and mean of the random loss L, respectively. We define $\widetilde{f}(\ell) = \dfrac{\ell f_L(\ell)}{\mu_L}$ and take $\widetilde{f}$ to be the density function of the new random variable $\widetilde{L}$, then the above equation for $ES_\alpha(L)$ can be rewritten as

$$ES_\alpha(L) = \frac{\mu_L}{1-\alpha} \int_{\text{VaR}_\alpha}^{\infty} \widetilde{f}(\ell)\, \mathrm{d}\ell = \frac{\mu_L}{1-\alpha} P[\widetilde{L} > \text{VaR}_\alpha]. \tag{5.33}$$

The saddlepoint approximation to $P[\widetilde{L} > \text{VaR}_\alpha]$ requires knowledge of the cgf of $\widetilde{L}$. Note that the mgf of $\widetilde{L}$ is given by

$$M_{\widetilde{L}}(z) = \int_0^\infty e^{z\ell} \widetilde{f}(\ell)\, \mathrm{d}\ell = \frac{1}{\mu_L} \int_0^\infty e^{z\ell} \ell f_L(\ell)\, \mathrm{d}\ell = \frac{M_L'(z)}{M_L'(0)}.$$

Recall $\kappa_L'(z) = \dfrac{M_L'(z)}{M_L(z)}$ and $\kappa_L(z) = \log M_L(z)$ so that

$$\log M_L'(z) = \kappa_L(z) + \log(\kappa_L'(z)).$$

Hence, the cgf of $\widetilde{L}$ is related to $\kappa_L(z)$ by

$$\kappa_{\widetilde{L}}(z) = \kappa_L(z) + \log(\kappa_L'(z)) - \log(\kappa_L'(0)). \tag{5.34}$$

The domain of definition of $\kappa_{\widetilde{L}}(z)$ would be the same as that of $\kappa_L(z)$ and the derivatives of $\kappa_{\widetilde{L}}(z)$ can be easily expressed in terms of the derivatives of $\kappa_L(z)$. Indeed, this approach is similar to the change of measure approach in deriving the saddlepoint approximation to tail expectation in Sect. 2.3.1.

In summary, the saddlepoint approximation to $ES_\alpha(L)$ involves two steps: apply the first approximation to find VaR_α using $\kappa_L(z)$ and approximates $ES_\alpha(L) = \dfrac{\mu_L}{1-\alpha} P[\widetilde{L} > \text{VaR}_\alpha]$ using $\kappa_{\widetilde{L}}(z)$. As reported in the numerical tests in Gordy (2004), numerical accuracy and reliability of approximating $\text{VaR}_\alpha(L)$ and $ES_\alpha(L)$ under the CreditRisk$^+$ are quite well acceptable unless the skewness and kurtosis of the loss distribution are significant. Under the extreme cases with very high values of the skewness and kurtosis of the loss distribution, Annaert et al. (2007) demonstrate from their numerical tests that the calculations of $\text{VaR}_\alpha(L)$ and $ES_\alpha(L)$ in these cases may exhibit significant errors using the saddlepoint approximation procedures.

Gaussian copula default model

Under the Gaussian copula default model, the systematic risk factors $\mathbf{X} = (X_1, X_2, \ldots, X_M)$ are joint standard normal random variables with correlation matrix Ω. According to (5.14), the default probability of obligor i conditional on $\mathbf{X} = \mathbf{x} = (x_1, x_2, \ldots, x_M)$ is given by

$$p_i(\mathbf{X}) = P[D_i = 1|\mathbf{X} = \mathbf{x}] = \Phi\left[\frac{\Phi^{-1}(p_i) + \sum_{i=1}^M w_i x_i}{\gamma_i}\right]. \tag{5.35}$$

Given that the n credit obligors are independent conditional on the common systematic risk factor $\mathbf{X}$, the mgf of the loss variable L conditional on $\mathbf{X}$ is given by

$$M_L(z;\mathbf{X}) = \prod_{i=1}^n \left[1 - p_i(\mathbf{X}) + p_i(\mathbf{X})e^{w_i z}\right].$$

The conditional cgf of L and its higher order derivatives are found to be [see (20–24) in Huang et al. (2007)]

$$\kappa_L(z;\mathbf{X}) = \sum_{i=1}^n \log(1 - p_i(\mathbf{X}) + p_i(\mathbf{X})e^{w_i z}), \tag{5.36a}$$

$$\kappa_L'(z;\mathbf{X}) = \sum_{i=1}^n \frac{w_i p_i(\mathbf{X})e^{w_i z}}{1 - p_i(\mathbf{X}) + p_i(\mathbf{X})e^{w_i z}}, \tag{5.36b}$$

$$\kappa_L''(z;\mathbf{X}) = \sum_{i=1}^{n} \frac{[1-p_i(\mathbf{X})]w_i^2 p_i(\mathbf{X})e^{w_i z}}{[1-p_i(\mathbf{X})+p_i(\mathbf{X})e^{w_i z}]^2}, \text{ etc.} \tag{5.36c}$$

The tail distribution of the random loss L under the Gaussian copula default model is given by

$$P[L > \ell] = E_{\mathbf{X}}[P[L > \ell|\mathbf{X}]] = \int P[L > \ell|\mathbf{X}]\,\mathrm{d}P(\mathbf{X}), \tag{5.37}$$

where expectation is taken over the underlying joint Gaussian distribution of $\mathbf{X}$. The corresponding VaR contribution by obligor i is given by [see (5.25)]

$$w_i \frac{\partial \mathrm{VaR}_\alpha}{\partial w_i} = w_i \frac{E_{\mathbf{X}}\left[\frac{1}{2\pi i}\int_{\gamma-i\infty}^{\gamma+i\infty} \frac{p_i(\mathbf{X})e^{w_i z}}{1-p_i(\mathbf{X})+p_i(\mathbf{X})e^{w_i z}} e^{\kappa_L(z;\mathbf{X})-z\mathrm{VaR}_\alpha}\,\mathrm{d}z\right]}{E_{\mathbf{X}}\left[f_L(\mathrm{VaR}_\alpha|\mathbf{X})\right]}. \tag{5.38}$$

Huang et al. (2007) observe that the integrand in the above numerator integral can be rewritten in terms of L given $\mathbf{X}$ and $D_i = 1$, where the corresponding cgf is given by

$$\hat{\kappa}_L^i(z;\mathbf{X}) = \log(p_i(\mathbf{X})e^{w_i z}) + \sum_{j\neq i} \log(1-p_j(\mathbf{X})+p_j(\mathbf{X})e^{w_j z}).$$

In terms of $\hat{\kappa}_L^i(z;\mathbf{X})$, the VaR contribution by obligor i becomes

$$w_i \frac{\partial \mathrm{VaR}_\alpha}{\partial w_i} = w_i \frac{E_{\mathbf{X}}\left[\frac{1}{2\pi i}\int_{\gamma-i\infty}^{\gamma+i\infty} e^{\hat{\kappa}_L^i(z;\mathbf{X})-z\mathrm{VaR}_\alpha}\,\mathrm{d}z\right]}{E_{\mathbf{X}}\left[f_L(\mathrm{VaR}_\alpha|\mathbf{X})\right]}. \tag{5.39}$$

Indeed, (5.39) can be shown to be consistent with (5.26). We let $\hat{L}^i = \sum_{j\neq i} w_j D_j$ and express the VaR contribution by obligor i in terms of the density function of $\hat{L}^i$. From (5.26), we have

$$\begin{aligned} w_i \frac{\partial \mathrm{VaR}_\alpha}{\partial w_i} &= w_i E[D_i|L = \mathrm{VaR}_\alpha] \\ &= w_i \frac{f(L = \mathrm{VaR}_\alpha, D_i = 1)}{f_L(\mathrm{VaR}_\alpha)} \\ &= w_i \frac{E_{\mathbf{X}}[f(\hat{L}^i = \mathrm{VaR} - w_i|\mathbf{X})p_i(\mathbf{X})]}{E_{\mathbf{X}}\left[f_L(\mathrm{VaR}_\alpha|\mathbf{X})\right]}. \end{aligned} \tag{5.40}$$

Both the conditional portfolio loss density excluding the ith obligor in the numerator and the conditional portfolio loss density in the denominator can be calculated using the saddlepoint approximation. The saddlepoint $\hat{z}$ for the numerator integral is found by solving

$$\sum_{j \neq i} \frac{w_j p_j(\mathbf{X}) e^{w_j z}}{1 - p_j(\mathbf{X}) + p_j(\mathbf{X}) e^{w_j z}} = \mathrm{VaR}_\alpha - w_i. \tag{5.41}$$

By following a similar procedure, we derive the formula for calculating the ES contribution of obligor i. As revealed by (5.26) and (5.29) for $\frac{\partial \mathrm{VaR}_\alpha}{\partial w_i}$ and $\frac{\partial \mathrm{ES}_\alpha}{\partial w_i}$, respectively, one uses the substitution of "$L = \mathrm{VaR}_\alpha$" by "$L \geq \mathrm{VaR}_\alpha$" in (5.40). The ES contribution by obligor i is given by (Huang et al. 2007)

$$\begin{aligned} w_i \frac{\partial ES}{\partial w_i} &= w_i E[D_i | L \geq \mathrm{VaR}_\alpha] \\ &= w_i \frac{E_{\mathbf{X}}[p[\hat{L}^i \geq \mathrm{VaR}_\alpha - w_i | \mathbf{X}] p_i(\mathbf{X})]}{E_{\mathbf{X}}[P[L \geq \mathrm{VaR}_\alpha | \mathbf{X}]]}. \end{aligned} \tag{5.42}$$

The conditional tail expectations in both the numerator and denominator can be calculated by the corresponding saddlepoint approximation formulas for tail expectations.

Huang et al. (2007) performed rather comprehensive numerical tests on the saddlepoint approximation for portfolio with and without exposure concentrations. They used the saddlepoint approximation for calculating the conditional probability densities and tail distributions. The expectation calculation over the systematic risk factors were performed using either the Gauss quadrature for numerical integration in the expectation calculation when the number of risk factors M is small and the Monte Carlo simulation with importance sampling when M is larger than 3. For the calculation of the VaR contributions, they used the zeroth order and $O(\frac{1}{n})$ Daniels' saddlepoint approximation formulas [see (2.1)]. The higher order formula gives sufficiently good accuracy with percentage errors less than 1% while the zeroth order formula may generate percentage error close to 3% in some cases. For the calculation of ES contributions, they use the Lugannani-Rice saddlepoint formula for tail expectations [see (2.15)]. Typically, these calculations give percentage errors that are less than 0.5% in their sample portfolios. Unlike the CreditRisk$^+$ model, their numerical tests reveal no failure of the saddlepoint approximation even when the exposure concentration in the credit portfolio is high.

Numerical examples

Huang et al. (2007) use the less direct approach to compute the ES of a credit portfolio by summing the risk contributions to ES of all the risky obligors in the credit portfolio. It would be more convenient to use the available saddlepoint approximation formulas [see (5.22a, b, c)] to calculate ES directly. We performed numerical tests to assess numerical accuracy of these approximation formulas using two examples of the credit portfolios in Huang et al. (2007). We also compare the performance of the Gordy (2004) approach of approximating $ES_\alpha(L)$ using $\kappa_{\tilde{L}}(z)$ [see (5.33) and (5.34)] under the Vasicek single-factor default model.

In the first sample calculations, we use the same credit portfolio as in Example 4 in Huang et al. (2007). There are $n = 100$ obligors, each has probability of default of 0.01 over a given time horizon. The correlation coefficient ρ in the Vasicek one-factor default model is taken to be 0.5 uniformly. The exposures w_i of the obligors are specified as

$$w_i = \begin{cases} 1 & i = 1, 2, \ldots, 20; \\ 4 & i = 21, 22, \ldots, 40; \\ 9 & i = 41, 42, \ldots, 60; \\ 16 & i = 61, 62, \ldots, 80; \\ 25 & i = 81, 82, \ldots, 100. \end{cases}$$

As the first step to compute $ES_\alpha(L)$ for a given confidence level α, it is necessary to solve for $\mathrm{VaR}_\alpha(L)$ of the credit portfolio using combination of the algebraic equations (5.32a, b). We would like to examine numerical accuracy of the Gordy (2004) approach of expressing $ES_\alpha(L)$ in terms of a modified loss variable $\widetilde{L}$ [see (5.33)], where $\kappa_L(z)$ and $\kappa_{\tilde{L}}(z)$ are related by (5.34). Also, we consider numerical accuracy of the two simplified versions of saddlepoint approximation for calculating $ES_\alpha(L)$ [see (5.22a) and (5.22c)]. The numerical results are compared with the benchmark Monte Carlo simulation calculations using 100,000 simulation paths.

Table 5.1 shows the numerical results of the VaR_α and ES_α calculation for the given credit portfolio using Monte Carlo simulation and various saddlepoint approximation formulas for calculating $ES_\alpha(L)$. Very good agreement between the numerical results using different methods is observed. It is encouraging to observe that even for the simplified low order saddlepoint approximation formulas for $ES_\alpha(L)$ [(5.22a) and (5.22c)], where numerical accuracy is seen to be well within 0.5%.

We repeat similar calculations for $\mathrm{VaR}_\alpha(L)$ and $ES_\alpha(L)$ for another credit portfolio as specified in Example 2 in Huang et al. (2007). Again, there are $n = 100$ obligors, with risk exposures of the obligors ranging from $1, 2, \ldots, 99, 100$. The default probability and correlation coefficient are taken to be 0.1 and 0.2 uniformly. The numerical results of $\mathrm{VaR}_\alpha(L)$ and $ES_\alpha(L)$ for the second credit portfolio using Monte Carlo simulation and various saddlepoint approximation formulas are shown in Table 5.2. We also observe high level of numerical accuracy of the saddlepoint approximation results.

Table 5.1 Numerical results of $\mathrm{VaR}_\alpha(L)$ and $ES_\alpha(L)$ at varying values of confidence level α for the credit portfolio specified in Example 4 in Huang et al. (2007) obtained using Monte Carlo simulation and various saddlepoint approximation formulas. The bracket quantities in the column of Monte Carlo simulation results are the standard derivation of the simulation

α (%)	VaR_α (5.32a, b)	VaR_α Monte Carlo	ES_α (5.33)	ES_α (5.22a)	ES_α (5.22c)	ES_α Monte Carlo
99	194.47	194.37(0.033)	311.82	311.77	311.81	312.34(0.053)
95	60.430	60.328(0.008)	145.50	145.44	145.45	145.75(0.019)
90	27.948	27.826(0.004)	93.504	93.277	93.258	93.316(0.011)

Table 5.2 Numerical results of $\mathrm{VaR}_\alpha(L)$ and $ES_\alpha(L)$ at varying values of confidence level α for the credit portfolio specified in Example 2 in Huang et al. (2007) obtained using Monte Carlo simulation and various saddlepoint approximation formulas. The bracket quantities in the column of Monte Carlo simulation results are the standard derivation of the simulation

α (%)	VaR_α (5.32a, b)	VaR_α Monte Carlo	ES_α (5.33)	ES_α (5.22a)	ES_α (5.22c)	ES_α Monte Carlo
99	2080.75	2079.96(0.120)	2427.47	2427.15	2427.46	2427.55(0.149)
95	1428.64	1428.51(0.059)	1826.93	1826.70	1826.92	1827.26(0.076)
90	1126.03	1125.88(0.042)	1543.95	1543.76	1543.94	1544.08(0.056)

5.3 Pricing of Collateralized Debt Obligations

A collateralized debt obligation (CDO) is an asset backed securities backed by a pool of corporate bonds and/or bank loans. It creates tranches with widely different risk characteristics from a portfolio of risky debt instruments. The CDOs were highly popular and showed phenomenal growth in market size in late 1990s until the financial tsunami in 2008. We discuss the cash flows of the counterparties in CDOs, namely, the originator and the investors on various tranches. Our goal is to show how to use the saddlepoint approximation to compute the fair spread rates for the tranches in CDOs whose payoffs depend on the amount of default losses in a credit basket of debt instruments.

5.3.1 Cashflows in Different Tranches

Consider a CDO with a portfolio of risky debt instruments, and as an illustration, suppose 4 tranches are created. Figure 5.1 shows the allocation of the portfolio debts into 4 tranches. The equity tranche is entitled to 5% of the total notional of the whole debts portfolio. More specifically, the equity tranche investors absorb the credit losses up to 5% of the notional during the life of the CDO. The lower and upper attachment points of the equity tranche is 0 and 0.05, corresponding to the tranche interval $[0, 0.05]$. When the total credit losses reach beyond 5% of the total notional, the losses on the portion beyond the first 5% but up to 15% will be taken by the second tranche investors. Likewise, the second tranche interval is $[0.05, 0.15]$, with 0.05 and 0.15 as the lower and upper attachment points. The credit losses beyond 15% will be taken by the third tranche investors. Similarly, since the third tranche investors pay the credit losses that are above the first 15% and up to 30% of the notional, the third tranche interval is $[0.15, 0.3]$. Lastly, the senior tranche investors compensate the originator for all remaining credit losses that are beyond the first 30% of the notional, so the senior tranche interval is $[0.3, 1]$.

Similar to the credit default swaps, the originator of the CDO is the protection buyer and receives the compensation of the credit losses in his debts portfolio from

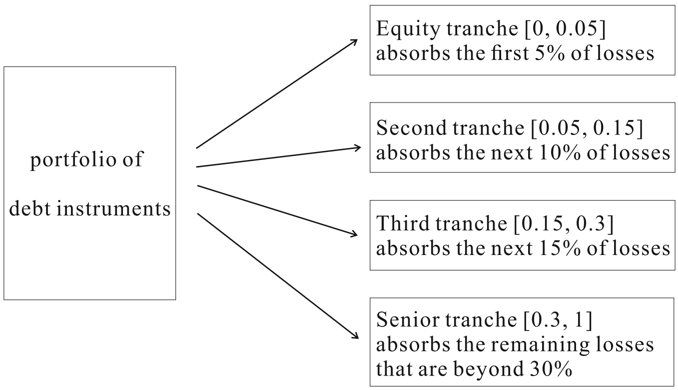


Fig. 5.1 Allocation of absorption of the credit losses in the portfolio of debt instruments into 4 tranches. The equity tranche absorbs the first 5% of credit losses and the senior tranche absorbs the remaining losses that are beyond 30%

the tranche investors. The tranche investors are the protection sellers, so in return they receive periodic protection premium payments from the originator according to the different levels of risks in the respective tranches.

We show in details how the originator of the CDO and the tranche investors exchange their cashflows on the M discrete payment dates t_m, $m = 1, 2, \ldots, M$, where $0 = t_0 < t_1 < t_2 < \cdots < t_M = T$, over the life $[0, T]$ of the CDO. Let L_t denote the credit losses from the portfolio up to time t. Consider the jth tranche with interval $[\kappa_{j-1}, \kappa_j]$, $j = 1, 2, \ldots, J$, where $\kappa_0 = 0$ and $\kappa_J = 1$ and let s_j be the spread rate structured for the jth tranche. The periodic payments received by the jth tranche investors from the CDO originator are computed based on the spread rate times the time interval between successive payment dates, then multiplied by the balance of the principal remaining in the tranche after the credit losses have been paid.

At time t_m, suppose $L_{t_m} \leq \kappa_{j-1}$, the remaining principal is the whole tranche width $\kappa_j - \kappa_{j-1}$; the principal becomes zero when $L_{t_m} \geq \kappa_j$. When $\kappa_{j-1} < L_{t_m} < \kappa_j$, the remaining principal is then $\kappa_j - L_{t_m}$. The amount of credit losses up to time t_m that fall within $[\kappa_{j-1}, \kappa_j]$ can be expressed as

$$L_{[\kappa_{j-1},\kappa_j]}(t_m) = (L_{t_m} - \kappa_{j-1})^+ - (L_{t_m} - \kappa_j)^+.$$

The cash amount $C_{j,m}$ received by the jth tranche investor at t_m for the mth time interval $[t_{m-1}, t_m]$ is given by

$$\begin{aligned} C_{j,m} &= \begin{cases} s_j(t_m - t_{m-1})(\kappa_j - \kappa_{j-1}) & L_{t_m} \leq \kappa_{j-1} \\ s_j(t_m - t_{m-1})(\kappa_j - L_{t_m}) & \kappa_{j-1} < L_{t_m} < \kappa_j \\ 0 & L_{t_m} \geq \kappa_j \end{cases} \\ &= s_j(t_m - t_{m-1})[\kappa_j - \kappa_{j-1} - L_{[\kappa_{j-1},\kappa_j]}(t_m)]. \end{aligned} \tag{5.43}$$

The compensated credit losses received by the CDO originator at t_m from the jth tranche investors for the time interval $[t_{m-1}, t_m]$ is given by the credit losses that fall within the jth tranche during the time interval. The expectation of the compensated amount is given by

$$E[L_{[\kappa_{j-1},\kappa_j]}(t_m) - L_{[\kappa_{j-1},\kappa_j]}(t_{m-1})] = E[L_{[\kappa_{j-1},\kappa_j]}(t_m)] - E[L_{[\kappa_{j-1},\kappa_j]}(t_{m-1})].$$

5.3.2 Fair Spread Rates for Tranches

The pricing issue of the CDO is to find the fair spread rate for each tranche. By equating the expected protection premium received by the tranche investors and the expectation of the compensated credit losses paid by the tranche investors, the fair spread rate s_j for tranche j is given by

$$s_j = \frac{\sum_{m=1}^M D(0,t_m)\{E[L_{[\kappa_{j-1},\kappa_j]}(t_m)] - E[L_{[\kappa_{j-1},\kappa_j]}(t_{m-1})]\}}{(t_m - t_{m-1})\sum_{m=1}^M D(0,t_m)\{\kappa_j - \kappa_{j-1} - E[L_{[\kappa_{j-1},\kappa_j]}(t_m)]\}}, \ j = 1, 2, \ldots, J. \tag{5.44}$$

Here, $D(0, t_m)$ is the discount factor over the time interval $[0, t_m]$.

From the above pricing formula for the spread rate, the calculation of s_j is reduced to the expectation calculation of $E[L_{[\kappa_{j-1},\kappa_j]}(t_m)]$ for a number of payment dates and tranche width $[\kappa_{j-1}, \kappa_j]$. The procedure starts with a choice of the default correlation model, like the Gaussian copula model or CreditRisk$^+$ model. With varying time horizons $t_m, m = 1, 2, \ldots, M$, we compute the conditional tail expectations $E[(L_{t_m} - \kappa_{j-1})^+|\mathbf{X}]$ and $E[(L_{t_m} - \kappa_j)^+|\mathbf{X}]$ using the corresponding saddlepoint approximation formula for tail expectation. Subsequently, we perform integration with respect to the underlying joint distribution of the risk factors $\mathbf{X}$. Sample calculations for finding the fair spread rates for the CDO tranches can be found in Yang et al. (2006); Huang and Oosterlee (2011). Their numerical tests reveal high level of accuracy in computing the spread rates of different tranches when the second order saddlepoint approximation formula is used in the tail expectation calculations.

Appendix

Proof of formula (5.20)

The proof follows from Huang and Oosterlee (2011). Let $\mu = E[L]$ and consider

$$E[L\mathbf{1}_{\{L\geq t\}}] = \int L\mathbf{1}_{\{L\geq t\}}\, \mathrm{d}P = \mu \int \frac{L}{\mu}\mathbf{1}_{\{L\geq t\}}\, \mathrm{d}P.$$

Assume that L has a positive lower bound and we define a new measure Q on the same filtered probability space $(\Omega, \mathscr{F})$ such that

$$Q(A) = \int_A \frac{L}{\mu}\,\mathrm{d}P \quad \text{for} \quad A \in \mathscr{F}.$$

The mgf of L under Q is given by

$$M_L^Q(z) = \int e^{zL}\frac{L}{\mu}\,\mathrm{d}P = \frac{M_L'(z)}{\mu} = \frac{1}{\mu}\frac{\mathrm{d}}{\mathrm{d}z}e^{\kappa_L(z)} = \frac{M_L(z)}{\mu}\kappa_L'(z).$$

The cgf of L under the two measures P and Q are related by

$$\kappa_L^Q(z) = \log M_L^Q(z) = \kappa_L(z) + \log\kappa_L'(z) - \ln\mu.$$

For $\gamma \in (0, \alpha_+)$, we recall from (1.13b) that the conditional expectation of X in the tail region is given by

$$\begin{aligned} E[L\mathbf{1}_{\{L\geq t\}}] &= \mu Q[L \geq t] \\ &= \frac{\mu}{2\pi i}\int_{\gamma-i\infty}^{\gamma+i\infty}\frac{e^{\kappa_L^Q(z)-zt}}{z}\,\mathrm{d}z \\ &= \frac{1}{2\pi i}\int_{\gamma-i\infty}^{\gamma+i\infty}\kappa_L'(z)\frac{e^{\kappa_L(z)-zt}}{z}\,\mathrm{d}z. \end{aligned}$$

Next, we show how to extend the result to the scenario where L has a negative lower bound $-B$, where $B > 0$. Define $X = L + B$ so that X has a positive lower bound. We observe

$$\kappa_X(z) = \kappa_L(z) + Bz \quad \text{and} \quad \kappa_X'(z) = \kappa_L'(z) + B$$

so that

$$E\left[L\mathbf{1}_{\{L\geq t\}}\right] = E\left[X\mathbf{1}_{\{X-B\geq t\}}\right] - BP[X - B \geq t].$$

On the other hand, we obtain

$$\begin{aligned} E\left[X\mathbf{1}_{\{X-B\geq t\}}\right] &= \frac{1}{2\pi i}\int_{\gamma-i\infty}^{\gamma+i\infty}[\kappa_L'(z) + B]\frac{\exp(\kappa_L(z) + Bz - z(t+B))}{z}\,\mathrm{d}z \\ &= \frac{1}{2\pi i}\int_{\gamma-i\infty}^{\gamma+i\infty}\kappa_L'(z)\frac{e^{\kappa_L(z)-zt}}{z}\,\mathrm{d}z + BP[X - B \geq t]. \end{aligned}$$

Combining the last two equations, we obtain

$$E\left[L\mathbf{1}_{\{L\geq t\}}\right] = \frac{1}{2\pi i}\int_{\gamma-i\infty}^{\gamma+i\infty} \kappa_L'(z)\frac{e^{\kappa_L(z)-zt}}{z}\,\mathrm{d}z.$$

When L is unbounded, we consider the truncated random variable $L_K = \max(L, K)$ so that L is bounded from below. We can show that $E\left[L\mathbf{1}_{\{L\geq t\}}\right] = E\left[L_K\mathbf{1}_{\{L_K\geq t\}}\right]$ using a similar procedure as above. Lastly, we take $L \to \infty$ and apply the monotone convergence theorem to show that $E\left[L\mathbf{1}_{\{L\geq t\}}\right]$ has the same integral representation as that of bounded L.

References

Aït-Sahalia, Y., Yu, J., "Saddlepoint approximations for continuous-time Markov processes," *Journal of Econometrics*, vol. 134(2) (2006) pp. 507–551.

Acerbi, C., Tasche, D., "On the coherence of Expected Shortfall," *Journal of Banking and Finance*, vol. 26(7) (2002) pp. 1487–1503.

Annaert, J., Garcia, C., Lamoot, J., Lanine, G., "Don't fall from the saddle: The importance of higher moments of credit loss distributions," *Working paper of University of Antwerp* (2007).

Antonov, A., Mechkov, S., Misirpashaev, T., "Analytical techniques for synthetic CDOs and credit default risk measures," *Wilmott Magazine*, November–December issue (2007) pp. 84–105.

Artzner, P., Delbaen, F., Eber, J.M., Hearth, D., "Coherent measures of risk," *Mathematical Finance*, vol. 9 (1999) pp. 203–228.

Barndorff-Nielsen, O., Cox, D.R., "Edgeworth and saddle-point approximations with statistical applications," *Journal of the Royal Statistical Society, Series B (Methodological)*, vol. 41(3) (1979) pp. 279–312.

Barndorff-Nielsen, O., "Modified signed log likelihood ratio," *Biometrika*, vol. 78 (1991) pp. 557–563.

Bertoin, J., 1998, *Lévy processes*, Cambridge University Press, Cambridge (1998).

Booth, J.G., Wood, A.T., "An example in which the Lugannani-Rice saddlepoint formula fails," *Statistics and Probability Letters*, vol. 23(1) (1995) pp. 53–61.

Broda, S.A., Paolella, M.S., "Expected shortfall for distributions in finance," *Statistical Tools for Finance and Insurance*, edited by Čižek, P., Härdle, W.K., Weron, R., (2011) pp. 57–99, Springer-Verlag, Berlin.

Broda, S.A., Paolella, H.S., "Saddlepoint approximations: A review and some new applications," *Handbook of Computational Statistics*, Springer (2012) pp. 953–983.

Butler, R.W., Wood, A.T., "Saddlepoint approximation for moment generating functions of truncated random variables," *Annals of Statistics*, (2004) pp. 2712–2730.

Butler, R.W., *Saddlepoint approximations with applications*, Cambridge University Press, Cambridge (2007).

Carr, P., Madan, D., "Option valuation using the fast Fourier transform," *Journal of Computational Finance*, vol. 2(4) (1999) pp. 61–73.

Carr, P., Madan, D., "Saddlepoint methods for option pricing," *Journal of Computational Finance*, vol. 13 (2009) pp. 49–61.

Carr, P., Geman, H., Madan, M., Yor, M., "Pricing options on realized variance," *Finance and Stochastics*, vol. 9 (2005) pp. 453–475.

CBOE, "The CBOE volatility index—VIX," Chicago Board Options Exchange (2003), available at: http://www.cboe.com/micro/vix/vixwhite.pdf.

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

Y. K. Kwok and W. Zheng, *Saddlepoint Approximation Methods in Financial Engineering*, SpringerBriefs in Quantitative Finance, https://doi.org/10.1007/978-3-319-74101-7

Daniels, H.E., "Exact suddlepoint approximations," *Biometrika*, vol. 67(1) (1980) pp. 59–63.

Daniels, H.E., "Tail probability approximations," *International Statistical Review*, vol. 55(1) (1987) pp. 37–48.

Duffie, D., Pan, J., Singleton, K., "Transform analysis and asset pricing for affine jump-diffusions," *Econometrica*, vol. 68(6) (2000) pp. 1343–1376.

Ebmeyer, D., "An asymptotic analysis of the compound gamma CreditRisk$^+$ model," *Working paper of DZ Bank AG* (2005).

García-Céspedes, R., "Three essays on computational finance and credit risk measurement and management," PhD thesis, Universidad de Castilla La Mancha (2014).

Giese, G., "Enhancing CreditRisk$^+$," *Risk Magazine*, vol. 16(4) (2003) pp. 73–77.

Giese, G., "A saddle for complex credit portfolio models," *Risk Magazine*, July issue (2006) pp. 84–89.

Glasserman, P., Kim, K.K., "Saddlepoint approximations for affine jump-diffusion models," *Journal of Economic Dynamics and Control*, vol. 33 (2009) pp. 15–36.

Goard, J., Mazur, M., "Stochastic volatility models and the pricing of VIX options," *Mathematical Finance*, vol. 23(3) (2013) pp. 439–458.

Gordy, M.B., "Saddlepoint approximation of CreditRisk$^+$," *Journal of Banking and Finance*, vol. 26(7) (2002) pp. 1335–1353.

Gordy, M.B., "Saddlepoint approximation," *CreditRisk$^+$ in the Banking Industry*, edited by Gundlach, V.M. and Lehrbass, F.B., (2004) pp. 91–110, Springer-Verlag, Berlin.

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

Gundlach, M., Lehrbass, F. (editors), *CreditRisk$^+$ in the banking industry*, Springer (2004).

Haaf, H., Tasche, D., "Calculating Value-at-Risk contributions in CreditRisk$^+$," *Working paper of Commerzbank AG and ETH Zürich* (2002).

Huang, X.Z., "Credit portfolio losses," PhD thesis, Technische Universiteit Delft (2009).

Huang, X.Z., Oosterlee, C.W., "Saddlepoint approximations for expectations and an application to CDO pricing," *SIAM Journal on Financial Mathematics*, vol. 2 (2011) pp. 692–714.

Huang, X.Z., Oosterlee, C.W., van der Weide, H., "Higher-order saddlepoint approximations in the Vasicek portfolio credit loss model," *Journal of Computational Finance*, vol. 11(1) (2007) pp. 93–113.

Jensen, J.L., *Saddlepoint approximations*, Oxford University Press, Oxford (1995).

Kallsen, J., Muhle-Karbe, J., Voβ, M., "Pricing options on variance in affine stochastic volatility models," *Mathematical Finance*, vol. 21(4) (2011) pp. 627–641.

Keller-Ressel, M., Muhle-Karbe, J., "Asymptotic and exact pricing of options on variance," *Finance and Stochastic*, vol. 17 (2013) pp. 107–133.

Kim, S.J., Kim, K.K., "Saddlepoint methods for conditional epectations with applications to risk management," to appear in *Bernoulli* (2017).

Kou, S., "A jump-diffusion model for option pricing," *Management Science*, vol. 48(8) (2002) pp. 1086–1101.

Kwok, Y.K., *Applied complex variables for scientists and engineers*, second edition, Cambridge University Press, Cambridge (2010).

Kwok, Y.K., Leung, K.S., Wong, H.Y., "Efficient option pricing using the Fast Fourier transform," *Handbook of Computational Finance*, Springer (2012) pp. 579–603.

Li, D., "On default correlation: A copula function approach," *Journal of Fixed Income*, vol. 9 (2000) pp. 43–54.

Lian, G.H., Zhu, S.P., "Pricing VIX options with stochastic volatility and random jumps," *Decisions in Economics and Finance*, vol. 36(1) (2013) pp. 71–88.

Lieberman, O., "On the approximation of saddlepoint expansions in statistics," *Econometric Theory*, vol. 10(5) (1994) pp. 900–916.

Lin, Y.N., Chang, C.H., "VIX option pricing," *Journal of Futures Markets*, vol. 29(6) (2009) pp. 523–543.

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

Martin, R., Thompson, K., Browne, C., "VAR: who contributes and how much?" *Risk Magazine*, August issue (2001) pp. 99–102.

Martin, R., "The saddlepoint method and portfolio optionalities," *Risk Magazine*, December issue (2006) pp. 93–95.

Martin, R., "Shortfall: who contributes and how much?" *Risk Magazine*, October issue (2009) pp. 84–89.

Martin, R., "Saddlepoint methods in portfolio theory," *Handbook of Credit Derivatives*, Oxford University Press (2013) pp. 533–572.

Nayfeh, A.H., *Introduction to perturbation techniques* (Chapter 3), John Wiley and Sons, New York, USA (1981).

Owen, A., McLeod, A., Thompson, K., "The hybrid saddlepoint method for credit portfolios," *Risk Magazine*, August issue (2009) pp. 82–86.

Reid, N., "Saddlepoint methods and statistical inference," *Statistical Science*, vol. 3(2) (1988) pp. 213–238.

Rogers, L.C.G., Zane, O., "Saddlepoint approximations to option prices," *Annals of Applied Probabilities*, vol. 9 (1999) pp. 493–503.

Rosen, D., Saunders, D., "Risk factor contributions in portfolio credit risk models," *Journal of Banking and Finance*, vol. 34(2) (2010) pp. 336–349.

Sepp, A., "VIX option pricing in a jump-diffusion model," *Risk Magazine*, April issue (2008a) pp. 84–89.

Sepp, A., "Pricing options on realized variance in the Heston model with jumps in returns and volatility," *Journal of Computational Finance*, vol. 11 (2008b) pp. 33–70.

Strawderman, R.L., "Higher-order asymptotic approximation: Laplace, saddlepoint, and related methods," *Journal of the American Statistical Association*, vol. 95 (2000) pp. 1358–1364.

Studer, M., "Stochastic Taylor expansions and saddlepoint approximations for risk management," PhD thesis, ETH Zürich (2001).

Wilde, T., "CreditRisk^{+}: A credit risk management framework," *Credit Suisse Financial Products* (1997).

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

Xiong, J., Wong, A., Salopek, D., "Saddlepoint approximations to option price in a general equilibrium model," *Statistics and Probability Letters*, vol. 71 (2005) pp. 361–369.

Yang, J., Hurd, T.R., Zhang, X., "Saddlepoint approximation method for pricing CDOs," *Journal of Computational Finance*, vol. 10(1) (2006) pp. 1–20.

Zhang, M.Z., Chan, L.L., "Saddlepoint approximations to option price in a regime-switching model," *Annals of Finance*, vol. 12 (2016) pp. 55–69.

Zhang, Y.T., Kwok, Y.K., "Saddlepoint approximation formulas for calculating tail expectation and expected shortfall under non-Gaussian base distribution," Working paper of Hong Kong University of Science and Technology (2018).

Zheng, W.D., "Analytic methods and numerical algorithms for pricing and hedging discretely sampled variance products and volatility derivatives," PhD thesis, Hong Kong University of Science and Technology (2012).

Zheng, W.D., Kwok, Y.K., "Saddlepoint approximation methods for pricing derivatives on discrete realized variance," *Applied Mathematical Finance*, vol. 21(1) (2014) pp. 1–31.

Zhu, S.P., Lian, G.H., "An analytical formula for VIX futures and its applications," *Journal of Futures Markets*, vol. 32(2) (2012) pp. 166–190.

Index

Y. K. Kwok and W. Zheng, *Saddlepoint Approximation Methods in Financial Engineering*, SpringerBriefs in Quantitative Finance, https://doi.org/10.1007/978-3-319-74101-7