TECHNISCHE UNIVERSITÄT WIEN
Vienna University of Technology

DISSERTATION

Central and Non-Central Limit Theorems

ausgeführt zum Zweck der Erlangung des akademischen Grades eines Doktors der Naturwissenschaften unter der Leitung von

Univ.-Prof. Dipl.-Math. Dr. rer. nat. habil. Uwe Schmock

E 105
Institut für Stochastik und Wirtschaftsmathematik
Forschungsgruppe Finanz- und Versicherungsmathematik

eingereicht an der Technischen Universität Wien
Fakultät für Mathematik und Geoinformation

von

Dipl.-Ing. Piet Nikolaus Porkert

Matrikelnummer: 0326312

Wien, 16. November 2016

Kurzfassung der Dissertation

Diese Dissertation beschäftigt sich mit der Theorie und den Anwendungen Zentraler Grenzwertsätze (ZGWS) und Nicht-Zentraler Grenzwertsätze (NZGWS). Teil I dient als Einleitung. In Teil II beschäftigen wir uns mit dem Grenzverhalten zufälliger Summen und beweisen einen NZGWS und Konvergenzraten. In diesem Teil liegt der Schwerpunkt auf der von uns vorgeschlagenen Beweistechnik. In Teil III analysieren wir das Grenzverhalten von Semimartingalen für kurze Zeiten. Wir beweisen ZGWS und gehen auf Anwendungen in der Finanzmathematik ein. Teil II und Teil III sind voneinander unabhängig.

Teil I, *Vorbereitung*: Kapitel 1 wiederholt klassische ZGWS und ist Grundlage für alles Weitere. In den Kapiteln 2 und 3 beschäftigen wir uns mit der steinschen Methode und der Asymptotik zufälliger Reihen. Kapitel 4 widmet sich dem Konzept der Stichprobenverzerrung. Ziel der Kapitel 2 bis 4 ist es, einen Grundstein für Teil II zu legen. Kapitel 5 motiviert die in Teil II ausgeführte Forschung.

Mit Kapitel 6 wechseln wir das Thema und präsentieren ein heuristisches Argument zugunsten eines ZGWS für eine Klasse stetiger Semimartingale. Kapitel 7 greift diese Fragestellung auf und motiviert die in Teil III ausgeführte Forschung.

Teil II, *Analyse von Zufallssummen poissonscher Mischverteilungen mittels steinscher Methode*: Mithilfe der steinscher Methode geben wir Abschätzungen für die Wasserstein- und die Kolmogorov-Distanz zwischen Zufallssummen poissonscher Mischverteilungen und ihren Grenzverteilungen. Beachtenswert ist wie die steinsche Methode zur Anwendung kommt. Durch stochastisches Bedingen ist es möglich, den Fall einer gaußschen Varianz-Mischverteilung auf den Fall einer Normalverteilung zurückzuführen, wodurch man die Analyse einer unhandlichen Stein-Gleichung umgehen kann.

Teil III, *Zentrale Grenzwertsätze für Semimartingale für kurze Zeiten*: Wir zeigen einen ZGWS sowie einen funktionalen ZGWS für stetige Semimartingale für kurze Zeiten. Wir verallgemeinern diese Resultate für Semimartingale mit Sprüngen. Als Anwendungen in der Finanzmathematik besprechen wir die Bepreisung digitaler Optionen am Geld für kurze Zeiten und geben eine Abschätzung für die Asymptotik erster Ordnung der *implied volatility skew* am Geld.

Abstract

This dissertation is focused on the theory and applications of central limit theorems (CLTs) and non-central limit theorems (NCLTs). Part I has preliminary character. In Part II, we deal with the limit behavior of random sums. We prove a NCLT and rates of convergence. In this part, our primary emphasis is the new method of proof we propose. In Part III, we study the limit behavior of semimartingales for small times. We prove CLTs and extend these to functional CLTs on the process level. Subsequently, we show applications in mathematical finance. Part II and Part III are independent of each other.

Part I, *Preliminaries*: In Chapter 1, which is the foundation of Part II and Part III, we recapitulate classical CLTs. In Chapters 2 and 3, we introduce the reader to the rudiments of Stein's method and review parts of the asymptotic theory of random sums. Chapter 4 deals with the concept of size biasing. The knowledge of Chapters 2 to 4 is essential for Part II. Chapter 5 motivates the research which is carried out in Part II.

In Chapter 6 we change the subject and give a heuristic argument in favor of a small-time CLT for a class of continuous semimartingales. Chapter 7 seizes this idea and further motivates the research in Part III.

Part II, *Analysis of Poisson Mixture Sums via Stein's Method*: By using Stein's method, we study the Wasserstein, as well as the Kolmogorov distances of Poisson mixture sums and their limit distributions. The primary focus is laid on how Stein's method is applied. By stochastic conditioning, it is possible to work with Stein's equation of the Gaussian distribution instead of a more complex Stein equation.

Part III, *Small-Time Central Limit Theorems for Semimartingales*: We prove a CLT, as well as a functional CLT on the process level for continuous semimartingales for small times. These results are extended to semimartingales with jumps. As an application to mathematical finance, we discuss the pricing of at-the-money digital options with short maturities and the asymptotics of at-the-money short time implied volatility skews.

Acknowledgements

I thank Sühan Altay, Peter Eichelsbacher, Stefan Gerhold, Max Kleinert, Uwe Schmock, and Mykhaylo Shkolnikov for collarborations and discussions!

I thank the Christian Doppler Research Association (CDG), Bank Austria, and the Austrian Science Fund (FWF) for financial support!

Contents

Part I

Preliminaries

Chapter 1

Classical Central Limit Theorems

We state classical central limit theorems (CLTs), which we will continue to use freely thereafter.

Since [Pól20], the name central limit theorem is used when the limit distribution of a statistical model is Gaussian. The history of CLTs, portrayed in [Fis11], begins with de Moivre's paper on the normal approximation of the binomial distribution in 1733 [Smi59]. In 1774, de Moivre's work was improved by Laplace [Sti86] and the result became known as the de Moivre–Laplace theorem. Laplace developed the characteristic function as a tool and presented the first general CLT in [Lap95], admittedly with an incomplete proof. According to Kallenberg [Kal02], the first rigorous proof was given in 1901 by Lyapunov [Lya01]. Between 1920 and 1922, Lindeberg proved sufficient conditions for a CLT in [Lin20; Lin22a; Lin22b], which later also turned out to be essentially necessary. In 1927, Bernstein obtained the first extension to higher dimensions [Ber27]. A more quantitative version, known as the Berry–Esseen theorem, was independently established by Berry [Ber41] and Esseen [Ess42] in 1941 and 1942, respectively. It specifies the rate at which the convergence to a Gaussian distribution takes place by giving a bound on the maximum error. In 1972, a whole new chapter of the theory of CLTs was opened by Stein [Ste72]. He introduced a new technique involving a differential operator to deliver explicit estimates of the approximation error. That the heart of this method does not rely on independence has led to a wide range of applications of Stein's method. Consequently, Stein's method was further developed by several mathematicians, see the monographs [BC05; CGS11; NP12] and the references therein.

Let $X = \{X_i\}_{i\in\mathbb{N}}$ be a sequence of independent random variables, with

$$\mathbb{E}[X_i] = 0, \quad 0 < \sigma_i := \sqrt{\mathrm{V}(X_i)} < \infty, \quad i \in \mathbb{N}.$$

For $n \in \mathbb{N}$, we define

$$S_n := \sum_{i=1}^{n} X_i, \quad \Sigma_n := \sqrt{\mathrm{V}(S_n)} = \Big(\sum_{i=1}^{n} \sigma_i^2\Big)^{\frac{1}{2}}, \quad \overline{S}_n := \frac{S_n}{\Sigma_n}.$$

Theorem 1.1 (CLT). *Let $Z \sim \mathrm{N}(0,1)$ be a standard Gaussian random variable on a probability space $(\Omega', \mathscr{F}', \mathbb{P}')$ and let $X = \{X_i\}_{i\in\mathbb{N}}$ be a sequence of independent, square-integrable random variables with zero mean and positive standard deviation on a probability* *space* $(\Omega, \mathscr{F}, \mathbb{P})$*. If the Lindeberg condition*

$$\lim_{n\to\infty} \frac{1}{\Sigma_n^2} \sum_{i=1}^{n} \mathbb{E}\big[X_i^2; |X_i| \geq \varepsilon\Sigma_n\big] = 0, \quad \varepsilon > 0, \tag{1.2}$$

holds, then

$$\lim_{n\to\infty} \mathbb{E}_{\mathbb{P}}[\varphi_n(\overline{S}_n)] = \mathbb{E}_{\mathbb{P}'}[\varphi(Z)],$$

where $\{\varphi_i\}_{i\in\mathbb{N}} \subset \mathrm{C}(\mathbb{R};\mathbb{C})$ *satisfies*

$$\sup_{n\in\mathbb{N}} \sup_{x\in\mathbb{R}} \frac{\varphi_n(x)}{1+|x|^2} < \infty,$$

and converges to φ *uniformly on compacts. Moreover, if* $-\infty \leq a < b \leq \infty$*, then*

$$\lim_{n\to\infty} \mathbb{P}\big[a \leq \overline{S}_n \leq b\big] = \frac{1}{\sqrt{2\pi}} \int_a^b \mathrm{e}^{-\frac{s^2}{2}}\, \mathrm{d}s.$$

Proof. See [Str11, Theorem 2.1.8, p. 64]. □

Remark 1.3. If the random variables $\{X_i\}_{i\in\mathbb{N}}$ are identically distributed, then

$$\lim_{n\to\infty} \frac{1}{\sigma_1^2} \mathbb{E}\big[X_1^2; |X_1| \geq \varepsilon\sqrt{n}\sigma_1\big] = 0, \quad \varepsilon > 0,$$

meaning that the Lindeberg condition (1.2) holds.

Remark 1.4. The Lindeberg condition (1.2) is a sufficient condition for

$$\lim_{n\to\infty} \mathbb{E}_{\mathbb{P}}[\varphi(\overline{S}_n)] = \mathbb{E}_{\mathbb{P}'}[\varphi(Z)] \tag{1.5}$$

to hold for all $\varphi \in \mathrm{C_b}(\mathbb{R};\mathbb{C})$.

Feller proved that (1.5) for all $\varphi \in \mathrm{C_b}(\mathbb{R})$ and

$$\lim_{n\to\infty} \max_{1\leq i\leq n} \frac{\sigma_i}{\Sigma_n} = 0$$

imply that Lindeberg's condition (1.2) holds. These two results combined are known as the Lindeberg–Feller theorem, see [Str11, p. 62].

Theorem 1.6 (Berry–Esseen)**.** *Let* $X = \{X_i\}_{i\in\mathbb{N}}$ *be a sequence of independent random variables on the probability space* $(\Omega, \mathscr{F}, \mathbb{P})$*, with*

$$\mathbb{E}[X_i] = 0, \quad 0 < \mathrm{V}(X_i) < \infty, \quad \xi_i := \big(\mathbb{E}\big[|X_i|^3\big]\big)^{\frac{1}{3}} < \infty, \quad i \in \mathbb{N}.$$

Then, for $n \in \mathbb{N}$*,*

$$\|F_n - \Phi\|_\infty \leq 10 \frac{\sum_{i=1}^n \xi_i^3}{\Sigma_n^3},$$

where

$$F_n(x) := \mathbb{P}\big[\overline{S}_n \leq x\big], \quad x \in \mathbb{R},$$

and Φ *is the cumulative distribution function of a standard Gaussian random variable.*

In particular, if

$$\mathrm{V}(X_i) = 1, \quad i \in \{1, \ldots, n\},$$

then

$$\|F_n - \Phi\|_\infty \leq 10 \frac{\sum_{i=1}^n \xi_i^3}{n^{\frac{3}{2}}} \leq 10 \frac{\max_{i\in\{1,\ldots,n\}} \xi_i^3}{\sqrt{n}}.$$

Proof. See [Str11, Theorem 2.2.17, p. 77 et seqq.] for a proof via Stein's method based on the work by Bolthausen [Bol84]. □

Theorem 1.7 (Multivariate CLT)**.** *For $m \in \mathbb{N}$, let $\{X_i\}_{i\in\mathbb{N}}$ be a sequence of independent, square-integrable, $\mathbb{R}^m$-valued random vectors on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$. Further, assume that for $i \in \mathbb{N}$, $\mathbb{E}[X_i] = 0$ and that the covariance matrix $\mathrm{Cov}(X_i)$ of each X_i is strictly positive definite. Let $n \in \mathbb{N}$, then we define*

$$S_n := \sum_{i=1}^n X_i, \quad C_n := \mathrm{Cov}(S_n) = \sum_{i=1}^n \mathrm{Cov}(X_i), \quad \Sigma_n := (\det(C_n))^{\frac{1}{2m}}, \quad \overline{S}_n := \frac{S_n}{\Sigma_n}.$$

Assume that the limit

$$C := \lim_{n\to\infty} \frac{C_n}{\Sigma_n^2}$$

exists and that the Lindeberg condition

$$\lim_{n\to\infty} \frac{1}{\Sigma_n^2} \sum_{i=1}^n \mathbb{E}\big[|X_i|^2; |X_i| \geq \varepsilon\Sigma_n\big] = 0, \quad \varepsilon > 0,$$

holds. Then, for every sequence $\{\varphi_i\}_{i\in\mathbb{N}} \subset \mathrm{C}(\mathbb{R}^m; \mathbb{C})$ that satisfies

$$\sup_{n\in\mathbb{N}} \sup_{x\in\mathbb{R}^m} \frac{|\varphi_n(x)|}{1 + |x|^2} \leq \infty, \tag{1.8}$$

and converges uniformly on compacts to φ,

$$\lim_{n\to\infty} \mathbb{E}_{\mathbb{P}}[\varphi_n(\overline{S}_n)] = \mathbb{E}_{\mathbb{P}'}[\varphi(Z)],$$

where $Z \sim \mathrm{N}(0, C)$ is a random vector on a probability space $(\Omega', \mathscr{F}', \mathbb{P}')$. In particular, when the random vectors $\{X_i\}_{i\in\mathbb{N}}$ are uniformly square-integrable with zero mean and common covariance C, then

$$\lim_{n\to\infty} \mathbb{E}_{\mathbb{P}}\Big[\varphi_n\Big(\frac{S_n}{\sqrt{n}}\Big)\Big] = \mathbb{E}_{\mathbb{P}'}[\varphi(Z)].$$

Remark 1.9. The case $m = 1$ is consistent with Theorem 1.1.

Proof. See [Str11, Theorem 2.3.8, p. 85 et seqq.]. □

Chapter 2

Introduction to Stein's Method

We illustrate Stein's method on which the results in Part II are built.

Let $\mathbb{P}$ and $\mathbb{Q}$ be probability measures on a measurable space $(\Omega, \mathscr{F})$, then it is natural to ask how close $\mathbb{P}$ and $\mathbb{Q}$ are in the following sense,

$$\mathrm{d}(\mathbb{P}, \mathbb{Q}) := \sup_{h \in \mathscr{H}} \Big| \int_\Omega h \, \mathrm{d}\mathbb{P} - \int_\Omega h \, \mathrm{d}\mathbb{Q} \Big|,$$

where $\mathscr{H}$ is a set of $\mathscr{F}$-measurable, $\mathbb{P}$- and $\mathbb{Q}$-integrable functions. A typical choice is

$$\mathscr{H}_{\mathrm{TV}} := \{1_A : A \in \mathscr{F}\},$$

when

$$\mathrm{d}_{\mathrm{TV}}(\mathbb{P}, \mathbb{Q}) := \sup_{h \in \mathscr{H}_{\mathrm{TV}}} \Big| \int_\Omega h \, \mathrm{d}\mathbb{P} - \int_\Omega h \, \mathrm{d}\mathbb{Q} \Big| = \sup_{A \in \mathscr{F}} \big| \, \mathbb{P}[A] - \mathbb{Q}[A] \, \big|$$

is called *total variation distance* of $\mathbb{P}$ and $\mathbb{Q}$. When $(\Omega, \mathscr{F})$ is given by $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$, the choice

$$\mathscr{H}_{\mathrm{K}} := \{1_{(-\infty, z]} : z \in \mathbb{R}\}$$

defines the *Kolmogorov distance*

$$\mathrm{d}_{\mathrm{K}}(\mathbb{P}, \mathbb{Q}) := \sup_{h \in \mathscr{H}_{\mathrm{K}}} \Big| \int_\mathbb{R} h \, \mathrm{d}\mathbb{P} - \int_\mathbb{R} h \, \mathrm{d}\mathbb{Q} \Big| = \sup_{x \in \mathbb{R}} \big| \, \mathbb{P}[(-\infty, x]] - \mathbb{Q}[(-\infty, x]] \, \big|,$$

whereas

$$\mathscr{H}_{\mathrm{W}} := \{h : \mathbb{R} \to \mathbb{R} : |h(x) - h(y)| \leq |x - y|, \ \forall x, y \in \mathbb{R}\}$$

defines the *Wasserstein distance*

$$\mathrm{d}_{\mathrm{W}}(\mathbb{P}, \mathbb{Q}) := \sup_{h \in \mathscr{H}_{\mathrm{W}}} \Big| \int_\mathbb{R} h \, \mathrm{d}\mathbb{P} - \int_\mathbb{R} h \, \mathrm{d}\mathbb{Q} \Big|,$$

in case the integrals are well-defined[1].

Given two probability measures $\mathbb{P}$ and $\mathbb{Q}$ and a distance d, Stein's method addresses the question of calculating $\mathrm{d}(\mathbb{P}, \mathbb{Q})$. In the special case where the distribution of $\mathbb{P}$ is a standard Gaussian distribution, we make the following observation. Let

$$Z \sim \mathrm{N}(0, 1)$$

[1]For $x \in \mathbb{R}$, let the probability density function with respect to the Lebesgue measure be given by $f(x) := x^{-2} 1_{[1,\infty)}(x)$ and consider the Lipschitz function $h(x) := x$, then $\int_\mathbb{R} h(x) f(x) \, \mathrm{d}x = \int_1^\infty x^{-1} \, \mathrm{d}x = \infty$.

and $f : \mathbb{R} \to \mathbb{R}$ be a continuous and piecewise continuously differentiable function, such that $\mathbb{E}[|f'(Z)|] < \infty$. Then

$$\mathbb{E}\big[f'(Z)\big] - \mathbb{E}[Zf(Z)] = 0, \tag{2.1}$$

since by partial integration

$$\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f'(s)\, \mathrm{e}^{-\frac{s^2}{2}}\, \mathrm{d}s = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} s f(s)\, \mathrm{e}^{-\frac{s^2}{2}}\, \mathrm{d}s.$$

Even more is true, the Gaussian distribution can be characterized by equation (2.1)[2]. Inspired by (2.1), it is plausible that the expression

$$\big|\, \mathbb{E}\big[f'(W)\big] - \mathbb{E}[Wf(W)]\, \big|$$

is *small*, whenever W is a random variable *close* to Z. More concretely, if

$$\big|\, \mathbb{E}[h(W)] - \mathbb{E}[h(Z)]\, \big|$$

is *small* for all $h \in \mathscr{H}$, the expression

$$\big|\, \mathbb{E}\big[f'(W)\big] - \mathbb{E}[Wf(W)]\, \big|$$

should be *small* for all f in a certain $\mathscr{H}'$.

We now describe how $\mathscr{H}$ and $\mathscr{H}'$ are related. For bounded g, the ordinary differential equation (ODE)

$$f'(x) - xf(x) = g(x), \qquad \lim_{x \to -\infty} f(x)\, \mathrm{e}^{-\frac{x^2}{2}} = 0, \tag{2.2}$$

called *Stein's equation*, has the solution[3]

$$f(x) = \mathrm{e}^{\frac{x^2}{2}} \int_{-\infty}^{x} g(s)\, \mathrm{e}^{-\frac{s^2}{2}}\, \mathrm{d}s, \quad x \in \mathbb{R}.$$

Thus, if for a particular bounded $h \in \mathscr{H}$ we define

$$g(x) := h(x) - \mathbb{E}[h(Z)]\,, \quad x \in \mathbb{R} \tag{2.3}$$

then

$$f_h(x) := \mathrm{e}^{\frac{x^2}{2}} \int_{-\infty}^{x} \big(h(s) - \mathbb{E}[h(Z)]\big)\, \mathrm{e}^{-\frac{s^2}{2}}\, \mathrm{d}s, \quad x \in \mathbb{R}, \tag{2.4}$$

solves

$$f_h'(x) - xf_h(x) = h(x) - \mathbb{E}[h(Z)]\,, \qquad \lim_{x \to -\infty} f_h(x)\, \mathrm{e}^{-\frac{x^2}{2}} = 0.$$

Substituting a random variable W for x and integrating leads to

$$\mathbb{E}[h(W)] - \mathbb{E}[h(Z)] = \mathbb{E}\big[f_h'(W)\big] - \mathbb{E}[Wf_h(W)]\,.$$

Thus, instead of analyzing the desired

$$\sup_{h \in \mathscr{H}} \big|\, \mathbb{E}[h(W)] - \mathbb{E}[h(Z)]\, \big| \tag{2.5}$$

[2]See [BC05, Lemma 2.1, p. 9 et seq.]
[3]See [BC05, p. 3 et seq.]

it is possible to analyze

$$\sup_{h\in\mathscr{H}} \big|\, \mathbb{E}\big[f_h'(W)\big] - \mathbb{E}[W f_h(W)] \,\big|, \tag{2.6}$$

although it is not obvious what is gained by that. Notice that (2.5) directly depends on Z, whereas (2.6) depends on Z indirectly via f_h. The following example illustrates how the terms in (2.6) might be analyzed.

Example 2.7. For $n \in \mathbb{N}$, let $X = \{X_i\}_{i\in\{1,\dots,n\}}$ be independent and identically distributed (i.i.d.), with

$$\mathbb{E}[X_1] = 0, \quad \mathbb{E}\big[X_1^2\big] = 1, \quad \mathbb{E}\big[|X_1|^3\big] < \infty,$$

and

$$W := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i.$$

According to Theorem 1.6, the distribution of W is *close* to a standard Gaussian distribution whenever n is large. According to our understanding

$$\mathbb{E}\big[W f(W) - f'(W)\big]$$

should be small for a reasonable large class of functions. For example, let $f \in \mathrm{C}^2(\mathbb{R})$ be such that $\mathbb{E}[W f(W)]$ and $\mathbb{E}[f'(W)]$ exist and

$$W' := \frac{1}{\sqrt{n}} \sum_{i=2}^{n} X_i.$$

Since X is i.i.d. and by Taylor's theorem we get

$$\begin{aligned}
\mathbb{E}[W f(W)] &= n\,\mathbb{E}\Big[\frac{1}{\sqrt{n}} X_1 f(W)\Big] \\
&= \sqrt{n}\,\mathbb{E}\Big[X_1 f\Big(\frac{1}{\sqrt{n}} X_1 + W'\Big)\Big] \\
&= \sqrt{n}\,\mathbb{E}\Big[X_1 \Big(f(W') + \frac{1}{\sqrt{n}} X_1 f'(W')\Big)\Big] + R_1,
\end{aligned}$$

where

$$|R_1| \le \frac{\|f''\|_\infty}{2\sqrt{n}}\, \mathbb{E}\big[|X_1|^3\big]\,.$$

Furthermore, by the same arguments

$$\begin{aligned}
\mathbb{E}\big[f'(W)\big] &= \mathbb{E}\Big[f'\Big(\frac{1}{\sqrt{n}} X_1 + W'\Big)\Big] \\
&= \mathbb{E}\big[f'(W')\big] + R_2,
\end{aligned}$$

where

$$|R_2| \le \frac{\|f''\|_\infty}{\sqrt{n}}\, \mathbb{E}[|X_1|]\,.$$

Since W' and X_1 are independent, $\mathbb{E}[X_1] = 0$ and $\mathbb{E}\big[X_1^2\big] = 1$, we have

$$\big|\, \mathbb{E}\big[W f(W) - f'(W)\big] \,\big| \le \frac{\|f''\|_\infty}{\sqrt{n}} \big(1 + \frac{1}{2}\, \mathbb{E}\big[|X_1|^3\big]\big).$$

As we see, analyzing $\mathbb{E}[Wf(W)]$ and $\mathbb{E}[f'(W)]$ can be rewarding. Thus, to get an upper bound for $|\,\mathbb{E}[Wf(W) - f'(W)]\,|$ in Example 2.7, it is essential to control $\|f''\|_\infty$. If we start with $\sup_{h\in\mathscr{H}} |\,\mathbb{E}[h(W)] - \mathbb{E}[h(Z)]\,|$, we have to analyze $\mathbb{E}[Wf_h(W)]$ and $\mathbb{E}[f_h'(W)]$, where

$$f_h(x) = \mathrm{e}^{\frac{x^2}{2}} \int_{-\infty}^{x} \big(h(s) - \mathbb{E}[h(Z)]\big)\, \mathrm{e}^{-\frac{s^2}{2}}\, \mathrm{d}s, \quad x \in \mathbb{R}.$$

Hence, it is desirable to get estimates for $\|f_h''\|_\infty$ in terms of $h \in \mathscr{H}$. However, f_h'' not necessarily exists for every choice of $\mathscr{H}$, which is the case when the function space $\mathscr{H}_{\mathrm{K}}$ lacks regularity. Under such circumstances, one has to apply a more sophisticated analysis to $\mathbb{E}[Wf(W)]$ and $\mathbb{E}[f'(W)]$ than the one carried out in Example 2.7.

Remark 2.8. This chapter is based on [BC05].

Chapter 3

A Non-Central Limit Theorem

We give examples of random sums similar to those we analyze in Part II and present a non-central limit theorem (NCLT) that describes their asymptotic behavior.

Example 3.1 (Finance)**.** The two standard assumptions for the increments of a stock price process $S = \{S_t\}_{t\in[0,T]}$ are *independence* and *stationarity.* These two assumptions led Bachelier [Bac00] to discover the stochastic process now called Wiener process. If we divide the interval $[0, T]$ into sufficiently many subintervals, the increments $S_t - S_s$, $0 \le s < t \le T$, can be represented as a sum of increments on subintervals. Since we assumed independence, according to CLTs the distribution of $S_t - S_s$ converges to a Gaussian distribution, for example if the Lindeberg condition (1.2) is satisfied. If the subintervals have identical length, the assumption of stationarity implies that the distributions of the subintervals are identical. In this case, the Lindeberg condition is reduced to the requirement of finiteness of variances of the elementary increments. However, statistical analysis shows that a Wiener process seems to be a questionable model for the stock price increments, since in practice they are more leptokurtic.

An attempt to explain the observed leptokurtosis without giving up the assumptions of the model, independence and stationarity of increments, was made by Mandelbrot [Man63; Man69; Man09]. The deviation from the Gaussian distribution means that CLTs are not applicable. The assumption that the variances are not finite forces the use of NCLTs, leading to stable distributions as limit laws for the sums of identically distributed summands. However, the infiniteness assumption about the variances of price increments of arbitrary small time intervals appears to be impossible in practice.

According to [GK96], Clark [Cla70; Cla73] was the first who tried to explain the leptokurtosis of the increment distributions by investigating the heterogeneity of the time of trades. According to Clark, it is not the finiteness of the variances of elementary increments that becomes violated, but rather the assumption of non-randomness of the number of summands. Let the variation of the stock price during the considered time interval be described by $(t_i, S_i)_{i\in\mathbb{N}}$, where t_i is the time of the i^{th} trade and S_i is the price of the i^{th} deal. Hence, $S_t = S_i$, $t_i \le t < t_{i+1}$. Let N_t be the number of deals concluded by time t. Then for $S_0 \ge 0$,

$$S_t - S_0 = \sum_{i=1}^{N_t} (S_i - S_{i-1}), \quad t \in [0, T].$$

For $X_i := S_i - S_{i-1}$, $i \in \mathbb{N}$, we assume that $X = \{X_i\}_{i\in\mathbb{N}}$ are independent random variables and that X and $N = \{N_t\}_{t\in[0,T]}$ are independent. Then the asymptotic theory of random summation tells us that if the number of deals concluded in $[0,T]$ is sufficiently large and the price variations X satisfy the conditions of a CLT, then for some non-negative random variable U

$$\mathbb{P}[S_t - S_0 < x] \approx \mathbb{P}[UZ < x] = \mathbb{E}\big[\Phi(xU^{-1})\big]\,, \quad x \in \mathbb{R},$$

where $\Phi(\frac{x}{u})\big|_{u=0} := 1_{(0,\infty)}(x)$, $Z \sim \mathrm{N}(0,1)$, and Z and U are independent. Thus, the distributions of stock price increments should be searched for among normal variance mixture distributions[1], see Theorem 3.6. Furthermore, normal variance mixture random variables are always more leptokurtic than the Gaussian random variable itself.[2]

Remark 3.2. The example above is given in [GK96, p. 77 et seqq.].

Example 3.3 (Actuarial Science)**.** Let $X = \{X_i\}_{i\in\mathbb{N}}$ be a sequence of i.i.d. random variables, with $\mathbb{E}[X_1] = \mu \in \mathbb{R}$ and $0 < \sigma^2 := V(X_1) < \infty$. Let $N = \{N_t\}_{t\in\mathbb{R}_+}$ be a homogeneous Poisson process with intensity $\lambda > 0$, independent of X. The surplus of an insurance company is typically modeled by

$$S_t := ct - \sum_{i=1}^{N_t} X_i, \quad t \in \mathbb{R}_+, \tag{3.4}$$

where $c > 0$ is the intensity of insurance premiums, N_t is the number of insurance payments during time $(0,t]$ and X are the claims. Then the classical risk process is asymptotically normal

$$\lim_{t\to\infty} \mathbb{P}\left[\frac{S_t - t(c-\mu\lambda)}{\sqrt{t\lambda(\mu^2+\sigma^2)}} < x\right] = \Phi(x), \quad x \in \mathbb{R}.$$

However, when the conditions are weakened, the limit behavior is described by a NCLT.

Remark 3.5. This example is given in [GK96, p. 67 et seqq.].

We now proceed towards a NCLT for random sums. Let $X = \{X_i\}_{i\in\mathbb{N}}$ be a set of independent random variables, $a = \{a_i\}_{i\in\mathbb{N}}$ and $b = \{b_i\}_{i\in\mathbb{N}}$, $b_i > 0$, $i \in \mathbb{N}$, sequences of real numbers, and

$$S_n := \sum_{i=1}^{n} X_i, \quad Y_n := \frac{S_n - a_n}{b_n}, \quad n \in \mathbb{N}.$$

Let $\{N_i\}_{i\in\mathbb{N}}$ be a sequence of $\mathbb{N}_0$-valued random variables, independent of X, $c = \{c_i\}_{i\in\mathbb{N}}$ and $d = \{d_i\}_{i\in\mathbb{N}}$, $d_i > 0$, $i \in \mathbb{N}$, sequences of real numbers, and

$$Z_n := \frac{S_{N_n} - c_n}{d_n}, \quad n \in \mathbb{N}.$$

Theorem 3.6 (NCLT)**.** *Let the sequences a, b, c, d be such that*

$$\lim_{n\to\infty} b_n = \lim_{n\to\infty} d_n = \infty,$$

and

$$Y_n \xrightarrow{\mathrm{d}} Y,$$

[1] For a definition see Definition 1.9 in Part II.

[2] Let X and U be independent random variables with finite fourth moments, with $\mathbb{E}[X] = 0$ and $\mathbb{P}[U > 0] = 1$. Then the excess coefficient kurtosis $\kappa(UX) \geq \kappa(X)$. Furthermore, $\kappa(XU) = \kappa(X)$ if and only if (iff) $\mathbb{P}[U = c] = 1$, $c > 0$. See [GK96, p. 82 et seq.] for a proof.

as $n \to \infty$ with

$$F(x) := \mathbb{P}[Y < x]\,, \quad x \in \mathbb{R}.$$

Furthermore, let

$$\Big(\frac{b_{N_n}}{d_n}, \frac{a_{N_n} - c_n}{d_n}\Big) \xrightarrow{\mathrm{d}} (U, V)$$

as $n \to \infty$, for some random variables U, V. Then

$$\lim_{n\to\infty} \mathbb{P}[Z_n < x] = \mathbb{E}\Big[F\Big(\frac{x - V}{U}\Big)\Big], \quad x \in \mathbb{R},$$

where $F(\frac{x-v}{u})|_{u=0} := 1_{(v,\infty)}(x)$.

Proof. See [GK96, Theorem 3.1.2, p. 47 et seqq.]. □

Remark 3.7. If Z is a random variable with distribution function $\mathbb{E}\big[F\big(\frac{x-V}{U}\big)\big]$, then

$$Z \overset{\mathrm{d}}{=} UY + V,$$

where Y and (U, V) are independent.

Remark 3.8. Theorem 3.6 also holds for arbitrary sequences S_n, $n \in \mathbb{N}$, which are not necessarily cumulative sums of independent random variables, see [Kor92].

Remark 3.9. For necessary and sufficient conditions in the special case $Y \sim \mathrm{N}(0, 1)$, see [GK96, Theorem 3.3.2., p. 64].

Remark 3.10. Let U, V, Y, Z be random variables and

$$\mathscr{V}(Z) := \{(Y, U, V) : Z \overset{\mathrm{d}}{=} UY + V;\, Y \text{ and } (U, V) \text{ are independent}\}.$$

Then the set $\mathscr{V}(Z)$ is not empty, since $(Y, 0, Z) \in \mathscr{V}(Z)$, where Y is an arbitrary random variable independent of Z. Further, the random variables U and V are not determined uniquely by the distribution of Z. For example, let

$$Z \overset{\mathrm{d}}{=} G_1 - G_2,$$

where G_1 and G_2 are independent random variables carrying a gamma distribution[3] $G_1, G_2 \sim \Gamma(\alpha, \beta)$, $\alpha, \beta > 0$. Then $(1, G_1, -G_2)$, $(G_1, 1, -G_2)$, $(Y, 0, G_1 - G_2) \in \mathscr{V}(Z)$. Moreover, also

$$(Y, \sqrt{G}, 0) \in \mathscr{V}(Z),$$

where $Y \sim N(0, 1)$ independent of

$$G \sim \Gamma\Big(\alpha, \frac{1}{2}\beta^2\Big), \quad \alpha, \beta > 0.$$

To see this, we consider the characteristic function of $G_1 - G_2$,

$$\mathbb{E}\big[\mathrm{e}^{\mathrm{i}t(G_1 - G_2)}\big] = \frac{1}{\big(1 - \frac{\mathrm{i}t}{\beta}\big)^{\alpha}} \frac{1}{\big(1 + \frac{\mathrm{i}t}{\beta}\big)^{\alpha}} = \Big(\frac{\beta^2}{\beta^2 + t^2}\Big)^{\alpha}, \quad t \in \mathbb{R}.$$

[3] See Appendix B.

On the other hand, for $\mu = \frac{1}{2}\beta^2$ and $t \in \mathbb{R}$,

$$\begin{aligned}\mathbb{E}\big[\mathrm{e}^{\mathrm{i}t\sqrt{G}Y}\big] =& \mathbb{E}\big[\mathbb{E}\big[\mathrm{e}^{\mathrm{i}t\sqrt{G}Y}\,\big|\, G\big]\big] = \mathbb{E}\big[\varphi_Y(t\sqrt{G})\big] = \frac{\mu^\alpha}{\Gamma(\alpha)}\int_0^\infty \mathrm{e}^{-s(\frac{1}{2}t^2+\mu)}\, s^{\alpha-1}\,\mathrm{d}s \\ =& \frac{\mu^\alpha}{\Gamma(\alpha)(\frac{1}{2}t^2+\mu)^\alpha}\int_0^\infty \mathrm{e}^{-s}\, s^{\alpha-1}\,\mathrm{d}s = \Big(\frac{2\mu}{2\mu+t^2}\Big)^\alpha = \Big(\frac{\beta^2}{\beta^2+t^2}\Big)^\alpha.\end{aligned} \tag{3.11}$$

Note that for $\alpha = 1$, equation (3.11) shows that $\sqrt{G}Y$ carries a Laplace distribution with parameters 0 and $1/\beta$. This remark is based on [GK96, p. 50].

Example 3.12. We now give an application of Theorem 3.6. Let $X = \{X_i\}_{i\in\mathbb{N}}$ be a sequence of independent random variables with $\mathbb{E}[X_i] = 0$ and $\mathbb{E}\big[X_i^2\big] = 1$, $i \in \mathbb{N}$, such that the Lindeberg condition (1.2) holds. Furthermore,

$$\Lambda \sim \Gamma(1,\beta), \quad N_\lambda \sim \mathrm{NB}\big(1, \frac{\lambda}{\beta+\lambda}\big), \quad Z \sim \mathrm{N}(0,1), \quad \beta,\lambda > 0,$$

where NB stands for the negativ binomial distribution[4], and let Λ, N_λ, X, Z be independent, then as $\lambda \to \infty$,

$$Z_\lambda := \frac{1}{\sqrt{\lambda}}\sum_{i=1}^{N_\lambda} X_i \xrightarrow{\mathrm{d}} \sqrt{\Lambda}Z \sim \mathrm{L}\Big(0, \frac{1}{\sqrt{2\beta}}\Big).$$

By Remark 3.10,

$$\sqrt{\Lambda}Z \sim \mathrm{L}\Big(0, \frac{1}{\sqrt{2\beta}}\Big).$$

Since by Theorem 1.1

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i \xrightarrow{\mathrm{d}} Z \quad \text{as} \quad n \to \infty,$$

by Theorem 3.6 we have to show

$$\sqrt{\frac{N_\lambda}{\lambda}} \xrightarrow{\mathrm{d}} \sqrt{\Lambda}, \quad \text{as} \quad \lambda \to \infty. \tag{3.13}$$

The characteristic function of Λ is given by

$$\varphi_\Lambda(t) = \frac{\beta}{\beta - \mathrm{i}t}, \quad t \in \mathbb{R}.$$

On the other hand, the characteristic function of N_λ/λ is given by

$$\varphi_{\frac{N_\lambda}{\lambda}}(t) = \frac{\big(1-\frac{\lambda}{\beta+\lambda}\big)\,\mathrm{e}^{\mathrm{i}t/\lambda}}{1-\frac{\lambda}{\beta+\lambda}\,\mathrm{e}^{\mathrm{i}t/\lambda}} = \frac{\beta\,\mathrm{e}^{\mathrm{i}t/\lambda}}{\beta - \lambda(\mathrm{e}^{\mathrm{i}t/\lambda}-1)} \xrightarrow{\lambda\to\infty} \frac{\beta}{\beta-\mathrm{i}t} = \varphi_\Lambda(t), \quad t \in \mathbb{R}.$$

Thus, by Lévy's continuity theorem (3.13) follows and by Theorem 3.6

$$Z_\lambda \xrightarrow{\mathrm{d}} \sqrt{\Lambda}Z, \quad \text{as} \quad \lambda \to \infty.$$

[4]See Appendix B.

Chapter 4

The Concept of Size Biasing

We introduce the concept of size biasing, which we will use frequently in Part II.

We begin with an illustration given in [AG10]. Imagine the following situation: In a room, 20% of people sit at a table alone, 30% sit in pairs, 30% in groups of three, and 20% in groups of four. Of course this does not mean 20% of occupied tables are occupied by only one person. If there are 100 people and 50 tables in this room, 20 people sit alone, they occupy 20 tables, 30 people sit in pairs, they occupy 15 tables, 30 people sit in pairs of three, they occupy 10 tables, 20 people sit in pairs of 4, they occupy 5 tables. Thus, on 40% of the tables there sits one person, on 30% of the tables there sit two people, on 20% of the tables there sit three people and on 10% of the tables there sit four people. Hence, there is a difference if we *randomly pick a table* and record X, the number of people sitting there, e.g. $\mathbb{P}[X = 1] = 0.4$. Or if we *randomly pick a person* and record X^*, the number of people sitting at her table, e.g. $\mathbb{P}[X^* = 1] = 0.2$. Apparently, the probability $\mathbb{P}[X^* = n]$ is proportional to $n\,\mathbb{P}[X = n]$, that is (i.e.)

$$\mathbb{P}[X^* = n] = cn\,\mathbb{P}[X = n]\,, \quad n \in \{1, 2, 3, 4\},$$

where $c > 0$. Then,

$$1 = \sum_{i=1}^{4} \mathbb{P}[X^* = i] = c \sum_{i=1}^{4} i\,\mathbb{P}[X = i] = c\,\mathbb{E}[X]\,,$$

which implies $c = \frac{1}{\mathbb{E}[X]}$. Hence,

$$\mathbb{P}[X^* = n] = \frac{n\,\mathbb{P}[X = n]}{\mathbb{E}[X]}, \quad n \in \{1, 2, 3, 4\},$$

which leads us to the following definition.

Definition 4.1 (X-size bias distribution)**.** Let X be a non-negative random variable with $0 < \mathbb{E}[X] < \infty$. Then X^* has the X*-size bias* distribution if for all $f : \mathbb{R} \to \mathbb{R}$, such that $\mathbb{E}[Xf(X)]$ exists,

$$\mathbb{E}[Xf(X)] = \mathbb{E}[X]\,\mathbb{E}[f(X^*)]\,.$$

In Example 4.2 we calculate X^* of a Poisson random variable X. The result will be used extensively in Part II.

Example 4.2. Let the random variable X be Poisson distributed,

$$X \sim \mathrm{P}(\lambda), \quad \lambda > 0,$$

then

$$X^* = X + 1,$$

since

$$\begin{aligned}\mathbb{E}[Xf(X)] &= \sum_{i=0}^{\infty} i f(i) \frac{\lambda^i}{i!}\,\mathrm{e}^{-\lambda} = \sum_{i=0}^{\infty} (i+1) f(i+1) \frac{\lambda^{i+1}}{(i+1)!}\,\mathrm{e}^{-\lambda} \\ &= \lambda \sum_{i=0}^{\infty} f(i+1) \frac{\lambda^i}{i!}\,\mathrm{e}^{-\lambda} = \mathbb{E}[X]\,\mathbb{E}[f(X+1)]\,.\end{aligned}$$

In order to sharpen our intuition about size biasing, we finish this chapter with a classic example.

Example 4.3 (Waiting time paradox)**.** Buses arrive at the bus station in accordance with a Poisson process, the expected waiting time between consecutive busses is thus $1/\alpha$, $\alpha > 0$. If we arrive at an arbitrary time $t > 0$, how long do we expect to wait? More concretely, let $X = \{X_i\}_{i \in \mathbb{N}}$ be i.i.d. random variables with $X_1 \sim \mathrm{Exp}(\alpha)$ and $\alpha > 0$. Let $S_0 := 0$ and

$$S_n := \sum_{i=1}^{n} X_i, \quad n \in \mathbb{N}.$$

The time of arrival of the n^{th} bus is modeled by

$$S_n \sim \Gamma(n, \alpha),$$

the successive waiting time between arrivals of the i^{th} and the $(i+1)^{\text{th}}$ bus, $i \in \mathbb{N}_0$, is given by

$$X_{i+1} \sim \mathrm{Exp}(\alpha).$$

For $t > 0$, we define the random variable N_t as the number of indices $i \in \mathbb{N}$, such that $S_i \leq t$. Then $\{N_t = n\}$ iff $S_n \leq t$ and $S_{n+1} > t$. The number of arrivals within $[0, t]$ is thus modeled by

$$N_t \sim \mathrm{P}(\alpha t).$$

We now pick an arbitrary $t > 0$ and ask what is our expected waiting time, if we arrive at the bus stop at time t? Two reasonable *answers* are:

- The lack of memory of the exponential distribution, i.e. $\mathbb{P}[X_i > s + t \,|\, X_i > t] = \mathbb{P}[X_i > s]$, $s > 0$, $i \in \mathbb{N}$, suggests our waiting time does not depend on our arrival. Thus, the average waiting time equals $1/\alpha$.
- Since the average time between arrivals equals $1/\alpha$ and our arrival is arbitrary, by symmetry our expected waiting time is $1/2\alpha$.

To decide which answer is correct, a closer look is necessary. Let $t > 0$ be the time of arrival, then the waiting time is given by

$$W_t := S_i - t,$$

if $S_{i-1} < t \le S_i$, $i \in \mathbb{N}$. For $x > 0$,

$$\begin{aligned}\mathbb{P}[W_t \le x] &= \mathbb{P}[t < S_1 < x + t] + \sum_{i=1}^{\infty} \mathbb{P}[0 < S_i < t, t - S_i < X_{i+1} < t - S_i + x] \\ &= \mathrm{e}^{-\alpha t} - \mathrm{e}^{-\alpha(x+t)} + \sum_{i=1}^{\infty} \int_0^t \int_{t-\lambda}^{t-\lambda+x} \alpha \frac{(\alpha\lambda)^{i-1}}{(i-1)!} \mathrm{e}^{-\alpha\lambda} \, \alpha \, \mathrm{e}^{-\alpha\mu} \, \mathrm{d}\mu \, \mathrm{d}\lambda \\ &= \mathrm{e}^{-\alpha t} - \mathrm{e}^{-\alpha(x+t)} + \sum_{i=1}^{\infty} \int_0^t \alpha \frac{(\alpha\lambda)^{i-1}}{(i-1)!} \mathrm{e}^{-\alpha\lambda} \big(\mathrm{e}^{-\alpha(t-\lambda)} - \mathrm{e}^{-\alpha(t-\lambda+x)}\big) \, \mathrm{d}\lambda \\ &= 1 - \mathrm{e}^{-\alpha x},\end{aligned}$$

which means $W_t \sim \mathrm{Exp}(\alpha)$. Thus, the average waiting time equals $1/\alpha$ and the first answer is correct. However, the symmetry consideration also contains some truth. The mistake in reasoning is just that longer intervals simply have a better chance to cover the point $t > 0$ of our arrival than shorter ones. More concretely, the length of intervals that we arrive at

$$L_t := S_i - S_{i-1},$$

if $S_{i-1} < t \le S_i$, $i \in \mathbb{N}$, is not exponentially distributed. The random variable L_t has the density[1]

$$f_t(x) = \begin{cases} \alpha^2 x \, \mathrm{e}^{-\alpha x} & \text{for } 0 < x \le t, \\ \alpha(1 + \alpha t) \, \mathrm{e}^{-\alpha x} & \text{for } t < x. \end{cases}$$

As it can be veryfied by direct calculation,

$$\mathbb{E}[L_t] = \frac{2 - \mathrm{e}^{-\alpha t}}{\alpha}, \quad t > 0,$$

which converges to $2/\alpha$ for $t \to \infty$. By symmetry the waiting time is thus again approximately $1/\alpha$ for large enough t.

Remark 4.4. This example is given in [Fel71, p. 11 et seqq.]

Remark 4.5. In [AG10], size biasing is discussed extensively.

[1] For a proof see [Fel71, p. 11 et seqq.].

Chapter 5

Motivation for the Research in Part II

We present the research of several authors concerning the limit behavior of random sums and describe how we contribute to it in Part II.

Let $X = \{X_i\}_{i\in\mathbb{N}}$ be a sequence of independent (not necessarily identically distributed) random variables and let the random variable N_p be geometrically distributed[1]

$$N_p \sim \mathrm{Geom}(p), \quad 0 < p < 1,$$

independent of X. Toda [Tod12] studies the limit behavior of the geometric sum

$$Z_p := \sqrt{p}\sum_{i=1}^{N_p} X_i \quad \text{as} \quad p \searrow 0, \tag{5.1}$$

and identifies it to be the Laplace distribution. As we show in Example 3.6 (with different notation), this follows directly from Theorem 3.6. However, Toda proves a NCLT by adapting Lindeberg's method of proving a CLT for a fixed number of independent random variables, see [Tod12; EL14].

Theorem 5.2 (Toda). *Let $X = \{X_i\}_{i\in\mathbb{N}}$ be a sequence of independent random variables, such that*

$$\mathbb{E}[X_i] = 0, \quad \sigma_i := \sqrt{\mathrm{V}(X_i)} < \infty, \quad i \in \mathbb{N}.$$

Let

$$N_p \sim \mathrm{Geom}(p), \quad 0 < p < 1,$$

independent of X and suppose that

1. $\lim_{n\to\infty} n^{-\alpha}\sigma_n^2 = 0$ *for some* $0 < \alpha < 1$ *and* $\sigma^2 := \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n \sigma_i^2 > 0$ *exists,*
2. *for all* $\varepsilon > 0$ *the Lindeberg condition*

$$\lim_{p\to 0}\sum_{i=1}^{\infty}(1-p)^{i-1}p\,\mathbb{E}\big[X_i^2; |X_i| \geq \varepsilon/\sqrt{p}\big] = 0$$

holds.

[1] Note that we distinguish between Geom and G, see Appendix B.

Then, as $p \searrow 0$,

$$\sqrt{p}\sum_{i=1}^{N_p} X_i \xrightarrow{\mathrm{d}} Y \sim \mathrm{L}\Big(0, \frac{\sigma}{\sqrt{2}}\Big). \tag{5.3}$$

Proof. See [Tod12, Theorem 2.1, p. 3 et seqq.]. □

Remark 5.4. In [Tod12], the Theorem 5.2 is stated in slightly more general terms.

As we will see, the limit behavior in (5.3) was further analyzed by several authors. Pike and Ren [PR14] study a Stein operator (5.6) that characterizes the centered Laplace distribution and establish a rate of convergence in the bounded Lipschitz metric as an application, see Theorems 5.5 and 5.9.

Theorem 5.5 (Pike, Ren)**.** *Let $f \in \mathrm{C}(\mathbb{R})$, such that f and f' are locally absolutely continuous and*

$$(\mathscr{A} f)(x) := f(x) - f(0) - \sigma^2 f''(x), \quad x \in \mathbb{R}. \tag{5.6}$$

Let $X \sim \mathrm{L}(0, \sigma)$, then

$$\mathbb{E}[(\mathscr{A} f)(X)] = 0,$$

if $\mathbb{E}[f'(X)]$, $\mathbb{E}[|f''(X)|] < \infty$. Conversely, if X is a random variable with $\mathbb{E}[(\mathscr{A} f)(X)] = 0$ for every $f \in \mathrm{C}^2(\mathbb{R})$, with $\|f\|_\infty$, $\|f'\|_\infty$, $\|f''\|_\infty < \infty$, then $X \sim \mathrm{L}(0, \sigma)$.

Proof. See [PR14, Theorem 1.1, p. 3 et seqq.]. □

Remark 5.7. Compare the second order Stein operator for the Laplace distribution (5.6) with Stein's first order equation for the standard Gaussian distribution (2.2).

Definition 5.8. Let X, Y be random variables, then the *bounded Lipschitz* distance is defined by

$$\mathrm{d_{BL}}(X, Y) := \sup_{h \in \mathscr{H}_{\mathrm{BL}}} |\,\mathbb{E}[h(X)] - \mathbb{E}[h(Y)]\,|,$$

with

$$\mathscr{H}_{\mathrm{BL}} := \{h \in \mathrm{C}(\mathbb{R}) : \|h\|_\infty \leq 1 \text{ and } |h(x) - h(y)| \leq |x - y| \text{ for } x, y \in \mathbb{R}\}.$$

Theorem 5.9 (Pike, Ren)**.** *Let $X = \{X_i\}_{i\in\mathbb{N}}$ be a sequence of independent random variables with, $i \in \mathbb{N}$,*

$$\mathbb{E}[X_i] = 0, \quad 0 < \sigma := \sqrt{\mathrm{V}(X_i)}, \quad \rho := \sup_{i\in\mathbb{N}} \mathbb{E}\big[|X_i|^3\big] < \infty.$$

Let

$$N_p \sim \mathrm{Geom}(p), \quad 0 < p < 1,$$

independent of X and

$$Z \sim \mathrm{L}\Big(0, \frac{\sigma}{\sqrt{2}}\Big).$$

Then,

$$\mathrm{d_{BL}}(Z_p, Z) \leq \sqrt{p}\frac{\sigma + 2\sqrt{2}}{\sigma}\Big(\sigma + \frac{\rho}{3\sigma^2}\Big).$$

Proof. See [PR14, Theorem 1.3, p. 3 et seqq.]. □

Remark 5.10. Instead of analyzing Stein's operator for the Laplace distribution (5.6), Gaunt [Gau14] studies Stein's equation of variance-gamma distributions, which contain normal, gamma and Laplace distributions as special cases.

Döbler [Doe12] provided (5.3) with Berry–Esseen bounds, see Theorem 5.11 and Theorem 5.12 and further continued his work on this topic in [Doe15], see Theorem 5.15 and Theorem 5.17.

Theorem 5.11 (Döbler)**.** *Let* $X = \{X_i\}_{i\in\mathbb{N}}$ *be a sequence of i.i.d. random variables with*

$$\mathbb{E}[X_1] = 0, \quad 0 < \sigma := \sqrt{\mathrm{V}(X_1)} < \infty, \quad \xi := \mathbb{E}\big[|X_1|^3\big] < \infty.$$

Let

$$N_p \sim \mathrm{Geom}(p), \quad 0 < p < 1,$$

independent of X *and*

$$Z \sim \mathrm{L}\Big(0, \frac{\sigma}{\sqrt{2}}\Big),$$

then

$$\mathrm{d_K}(\mathscr{L}(Z_p), \mathscr{L}(Z)) \leq \frac{2C_K\xi}{\sigma^3}\sqrt{p} + 12p,$$

with $0 < C_K \leq 0.56$*.*

Proof. See [Doe12, Theorem 3.4, p. 13.]. □

Theorem 5.12 (Döbler)**.** *Let* $X = \{X_i\}_{i\in\mathbb{N}}$ *be a sequence of independent random variables with,*

$$\mathbb{E}[X_i] = 0, \quad 0 < \sigma_i := \sqrt{\mathrm{V}(X_i)} < \infty, \quad \xi_i := \mathbb{E}\big[|X_i|^3\big] < \infty, \quad i \in \mathbb{N},$$

with

$$\widehat{\xi}_n := \frac{1}{n}\sum_{i=1}^{n} \xi_i, \quad n \in \mathbb{N}.$$

Assume for

$$\widehat{\sigma}_n^2 := \sum_{i=1}^{n} \sigma_i^2, \quad n \in \mathbb{N},$$

that

$$0 < \widehat{\sigma}^2 := \lim_{n\to\infty} \frac{1}{n}\widehat{\sigma}_n^2 < \infty.$$

Let

$$N_p \sim \mathrm{Geom}(p), \quad 0 < p < 1,$$

be independent of X *and let*

$$Z \sim \mathrm{L}\Big(0, \frac{\widehat{\sigma}}{\sqrt{2}}\Big).$$

Then,

$$\mathrm{d_K}(\mathscr{L}(Z_p), \mathscr{L}(Z)) \leq\ 12p + p\sum_{i=1}^{\infty}(1-p)^{i-1}\min\Big(\big|1 - \frac{\widehat{\sigma}^2}{\widehat{\sigma}_i^2}\big|, \big|1 - \frac{\widehat{\sigma}_i^2}{\widehat{\sigma}^2}\big|\Big) + C_K\,\mathbb{E}\Big[\frac{1}{\sqrt{N_p}}\frac{\widehat{\xi}_N}{\widehat{\sigma}_N^3}\Big],$$

with $0 < C_K \leq 0.56$*.*

Proof. See [Doe12, Theorem 3.5, p. 14.]. □

Assumption 5.13. *Let* $X = \{X_i\}_{i\in\mathbb{N}}$ *be a sequence of i.i.d. random variables, with*

$$\mathbb{E}\left[|X_1|^3\right] < \infty.$$

Let N *be a* $\mathbb{N}_0$*-valued random variable, such that*

$$\mathbb{E}\left[N^3\right] < \infty,$$

independent of X*. Furthermore,*

$$\alpha := \mathbb{E}[N]\,, \qquad \beta := \sqrt{\mathbb{E}[N^2]}, \qquad \gamma := \sqrt{\mathrm{V}(N)}, \qquad \delta^3 := \mathbb{E}\left[N^3\right],$$
$$a := \mathbb{E}[X_1]\,, \qquad b := \sqrt{\mathbb{E}\left[X_1^2\right]}, \qquad c := \sqrt{\mathrm{V}(X_1)}, \qquad d^3 := \mathbb{E}\left[|X_1 - \mathbb{E}[X_1]\,|^3\right].$$

Let

$$S := \sum_{i=1}^{N} X_i,$$

then by Wald's equation and the Blackwell–Girshick equation

$$\mu := \mathbb{E}[S] = \alpha a, \quad \sigma^2 := \mathrm{V}(S) = \alpha c^2 + a^2\gamma^2.$$

According to [Rob48], under the assumption that

$$\sigma^2 = \alpha c^2 + a^2\gamma^2 \to \infty,$$

there are three situations in which

$$W := \frac{S-\mu}{\sigma} = \frac{S - \alpha a}{\sqrt{ac^2 + a\gamma^2}} \tag{5.14}$$

is asymptotically normal:

1. $c \neq 0 \neq a$ and $\gamma^2 = \mathrm{o}(\alpha)$ for $\alpha \to \infty$,
2. $a = 0 \neq c$ and $\gamma = \mathrm{o}(\alpha)$ for $\alpha \to \infty$, and
3. N is asymptotically normal and at least a or c is different from zero.

In case $c \neq 0 \neq a$ and $\gamma^2 = \mathrm{o}(\alpha)$, for $\alpha \to \infty$, N tends to infinity in a certain sense, but such that it only fluctuates slightly around its mean α and thus behaves more or less like a constant α tending to infinity. If $c = 0$ and $a \neq 0$, then we have $S = aN$ a.s. and asymptotic normality of S is equivalent to that of N.

Theorem 5.15 (Döbler)**.** *Let* X *and* N *be like in Assumption 5.13 and let* W *be given by* (5.14)*. Let*

$$Z \sim \mathrm{N}(0,1)$$

and let (N, N^*) *be a coupling, where* N^* *has the* N*-size biased distribution, independent of* X*, and* $D := N^* - N$*. Then*

$$\mathrm{d_W}(W,Z) \leq \frac{2c^2b\gamma^2}{\sigma^3} + \frac{3\alpha d^3}{\sigma^3} + \frac{\alpha a^2}{\sigma^2}\sqrt{\frac{2}{\pi}}\sqrt{\mathrm{V}(\mathbb{E}[D\,|\,N])} + \frac{2\alpha a^2 b}{\sigma^3}\,\mathbb{E}\left[1_{\{D<0\}}D^2\right] + \frac{\alpha|a|b^2}{\sigma^3}\,\mathbb{E}\left[D^2\right].$$

Additionally, if $D \geq 0$ then we also have

$$\begin{aligned}\mathrm{d_K}(W,Z) \leq & \frac{(\sqrt{2\pi}+4)bc^2\alpha}{4\sigma^3}\sqrt{\mathbb{E}[D^2]} + \frac{d^3\alpha(3\sqrt{2\pi}+4))}{8\sigma^3} + \frac{c^3\alpha}{\sigma^3} + \Big(\frac{7}{2}\sqrt{2}+2\Big)\frac{\sqrt{\alpha}d^3}{c\sigma^2} \\ & + \frac{c^2\alpha}{\sigma^2}\,\mathbb{P}[N=0] + \frac{d^3\alpha}{c\sigma^2}\,\mathbb{E}\big[N^{-\frac{1}{2}}1_{\{N\geq 1\}}\big] + \frac{\alpha|a|b^2\sqrt{2\pi}}{8\sigma^3}\,\mathbb{E}\big[D^2\big] \\ & + \frac{\alpha a^2}{\sigma^2}\sqrt{\mathrm{V}(\mathbb{E}[D\,|\,N])} + \frac{\alpha|a|b^2}{2\sigma^3}\sqrt{\mathbb{E}\Big[\mathbb{E}[D^2\,|\,N]^2\Big]} + \frac{\alpha|a|b}{\sigma^2}\sqrt{\mathbb{P}[N=0]}\sqrt{\mathbb{E}[D^2]} \\ & + \frac{\alpha|a|b^2}{c\sigma^2\sqrt{2\pi}}\,\mathbb{E}\big[D^2 1_{\{N\geq 1\}}N^{-\frac{1}{2}}\big] + \Big(\frac{d^3\alpha|a|b}{\sigma^2} + \frac{\alpha bc}{\sigma^2\sqrt{2\pi}}\Big)\,\mathbb{E}\big[D 1_{\{N\geq 1\}}N^{-\frac{1}{2}}\big].\end{aligned}$$

Proof. See [Doe15, Theorem 2.5., p. 8 et seqq.]. □

Remark 5.16. Assume that the index N is a positive constant, then

$$\mathrm{d_W}(W,Z) \leq \frac{3d^3}{c^3\sqrt{N}},$$

and

$$\mathrm{d_K}(W,Z) \leq \frac{1}{\sqrt{N}}\Big(1 + \Big(\frac{7}{2}(1+\sqrt{2}) + \frac{3\sqrt{2\pi}}{8}\Big)\frac{d^3}{c^3}\Big).$$

This is the optimal convergence rate for sums of i.i.d. random variables with finite third moments, although with non-optimal constants, see [Doe15, Corollary 2.10, p. 11. et seq.].

Theorem 5.17 (Döbler)**.** *Let X and N be like in Assumption 5.13 with $a = \mathbb{E}[X_1] = 0$ and let W be given by* (5.14) *and let*

$$Z \sim \mathrm{N}(0,1).$$

Then,

$$\mathrm{d_W}(W,Z) \leq \frac{2\gamma}{\alpha} + \frac{3d^3}{c^3\sqrt{\alpha}},$$

and

$$\begin{aligned}\mathrm{d_K}(W,Z) \leq & \frac{(\sqrt{2\pi}+4)\gamma}{4\alpha} + \Big(\frac{d^3(3\sqrt{2\pi}+4)}{8c^3} + 1\Big)\frac{1}{\sqrt{\alpha}} + \Big(\frac{7}{2}\sqrt{2}+2\Big)\frac{d^3}{c^3\alpha} \\ & + \mathbb{P}[N=0] + \Big(\frac{d^3}{c^3} + \frac{\gamma}{\sqrt{\alpha}\sqrt{2\pi}}\Big)\sqrt{\mathbb{E}\big[1_{\{N\geq 1\}}N^{-1}\big]}.\end{aligned}$$

Proof. See [Doe15, Theorem 2.7, p. 10 et seqq.]. □

Remark 5.18. According to [Doe15, Remark 2.8 (d), p.11], it is possible to drop the assumption that the summands are identically distributed.

Now we summarize this chapter so far. Several authors have studied the limit behavior of the geometric sum (5.1). However, they essentially differed in their method of attack: Toda [Tod12] adapted Lindeberg's method, Pike and Ren [PR14] analyzed Stein's equation of the Laplace distribution, and Döbler [Doe12] applied conditioning on the value of the index N_p and used known error bounds for sums of a fixed number of independent random variables, like the classical Berry–Esseen theorem. In [Doe15], Döbler studied a slightly different problem and combined Stein's method for normal approximation with coupling constructions and conditional independence to prove his results.

In Part II, we present our analysis of the limit behavior of (5.1). We now give a sketch of our ansatz. Let $\{X_i\}_{i\in\mathbb{N}}$ be a sequence of independent random variables and Λ be an a.s. positive random variable. Let N_λ be a $\mathbb{N}_0$-valued random variable which is conditionally Poisson distributed[2] given Λ, meaning $\mathscr{L}(N_\lambda\,|\,\Lambda) \overset{\text{a.s.}}{=} \mathrm{P}(\lambda\Lambda)$ and assume that (N_λ, Λ) is independent of $\{X_i\}_{i\in\mathbb{N}}$. We apply Stein's method in the spirit of Chapter 2 to prove upper bounds of the Wasserstein and Kolmogorov distances between the random sums

$$Z_\lambda := \frac{1}{\sqrt{\lambda}} \sum_{i=1}^{N_\lambda} X_i, \quad \lambda > 0,$$

and

$$Z := \sqrt{\Lambda} X, \tag{5.19}$$

where $X \sim \mathrm{N}(0,1)$, independent of $(N_\lambda, \Lambda, \{X_i\}_{i\in\mathbb{N}})$. A byproduct is the NCLT

$$\frac{1}{\sqrt{\lambda}} \sum_{i=1}^{N_\lambda} X_i \xrightarrow{\mathrm{d}} \sqrt{\Lambda} X \quad \text{as} \quad \lambda \to \infty. \tag{5.20}$$

For example, when Λ is gamma distributed, N_λ has a negative binomial distribution and in a special case a geometric distribution, see Example 1.5 in Part II. In this case, Z carries the Laplace distribution, see Remark 3.10 as well as Example 1.13 in Part II. Thus, with a change of notation (5.20) resembles (5.3).

However, instead of analyzing a cumbersome Stein operator, like for example (5.6), by conditioning we reduce the problem to the well understood case of (2.2). For a suitable function h and random variable $X \sim \mathrm{N}(0,1)$, the function

$$f_{\sigma^2}(x) := \mathrm{e}^{\frac{x^2}{2\sigma^2}} \int_{-\infty}^{x} h(s) - \mathbb{E}[h(\sigma X)]\, \mathrm{e}^{-\frac{s^2}{2\sigma^2}}\, \mathrm{d}s, \quad x \in \mathbb{R},$$

solves Stein's equation

$$f'_{\sigma^2}(x) - \frac{x}{\sigma^2} f_{\sigma^2}(x) = h(x) - \mathbb{E}[h(\sigma X)]\,, \quad \lim_{x\to-\infty} f_{\sigma^2}(x)\, \mathrm{e}^{-\frac{x^2}{2}} = 0, \quad x \in \mathbb{R}.$$

Substituting a random variable Z_λ for x, a positive random variable Λ for σ^2, and taking the conditional expectation $\mathbb{E}[\,\cdot\,|\Lambda]$, leads to

$$\mathbb{E}\left[f'_\Lambda(Z_\lambda)\,|\,\Lambda\right] - \mathbb{E}\Big[\frac{Z_\lambda}{\Lambda} f_\Lambda(Z_\lambda)\Big|\Lambda\Big] = \mathbb{E}[h(Z_\lambda)\,|\,\Lambda] - \mathbb{E}\Big[h(\sqrt{\Lambda}X)\,|\,\Lambda\Big]\,.$$

Thus, to estimate

$$\sup_{h\in\mathscr{H}} \left(\mathbb{E}[h(Z_\lambda)\,|\,\Lambda] - \mathbb{E}\left[h(\sqrt{\Lambda}X)\,\middle|\,\Lambda\right]\right)$$

for an appropriate set of functions $\mathscr{H}$, it is possible to analyze the terms

$$\mathbb{E}\left[f'_\Lambda(Z_\lambda)\,|\,\Lambda\right] \quad \text{and} \quad \mathbb{E}\Big[\frac{Z_\lambda}{\Lambda} f_\Lambda(Z_\lambda)\Big|\Lambda\Big]. \tag{5.21}$$

[2]For the definition of a Poisson mixture random variable see Definition 1.1 in Part II.

It is well known how to handle the expression (2.6) and the terms therein in the case of the Wasserstein or Kolmogorov distances. As a result, the analysis of (5.21) can be tracked back to the analysis of (2.6), once it is realized that size biasing, e.g. in the form

$$\mathbb{E}[N_\lambda]\,\mathbb{E}[f(Z_\lambda)\,|\,\Lambda] = \mathbb{E}\Big[N_\lambda f\Big(\frac{1}{\sqrt{\lambda}}\sum_{i=1}^{N_\lambda - 1} X_i\Big)\,\Big|\,\Lambda\Big],$$

allows us to connect our ansatz with the existing theory.

Chapter 6

A Heuristic Argument for a Small-Time Central Limit Theorem

We give a heuristic argument for a small-time CLT for a class of Itō diffusions. It raises the question for which class of processes a small-time CLT holds. This is addressed in Part III.

Suppose that $X = \{X_t\}_{t\in\mathbb{R}_+}$ satisfies a one-dimensional stochastic differential equation (SDE)

$$X_t = x_0 + \int_0^t \sigma(X_s)\,\mathrm{d}B_s, \quad t \geq 0,$$

where $x_0 \in \mathbb{R}$, σ is bounded, bounded away from zero, and Lipschitz continuous. Furthermore, for $\delta > 0$ let $X^\delta = \{X^\delta_t\}_{t\in\mathbb{R}_+}$ be the solution of the SDE

$$X^\delta_t = x_0 + \sqrt{\delta}\int_0^t \sigma(X^\delta_s)\,\mathrm{d}W_s, \quad t \geq 0.$$

Then, according to [Øks10, Theorem 8.5.1, p. 148], the distributions of the random variables X_δ and X^δ_1 coincide.

By [DZ10, Remark of Theorem 5.6.7, p. 214], the processes X^δ satisfy a large deviation principle (LDP) in $\mathrm{C}([0,1])$ as $\delta \searrow 0$, with rate function

$$\widetilde{I}(f) := \begin{cases} \frac{1}{2}\int_0^1 \frac{f'(s)^2}{\sigma(f(s))^2}\,\mathrm{d}s & \text{if } f \in \mathrm{H}^1([0,1]), \quad f(0) = x_0, \\ \infty & \text{otherwise,} \end{cases}$$

where $\mathrm{H}^1([0,1])$ denotes the space of absolutely continuous functions on $[0,1]$ with square-integrable derivatives. By the contraction principle [DZ10, Theorem 4.2.1, p. 126]), applied to the evaluation map $\mathrm{C}([0,1]) \to \mathbb{R}$, $f \mapsto f(1)$, we conclude that the random variables X_t satisfy an LDP as $t \searrow 0$ with rate function

$$\begin{aligned} I(x_0+y) &= \inf\{\widetilde{I}(f) : f \in \mathrm{C}([0,1]), \quad f(1) = x_0 + y\} \\ &= \frac{1}{2}\inf_{f\in\mathscr{H}}\int_0^1 \frac{f'(s)^2}{\sigma(f(s))^2}\,\mathrm{d}s, \end{aligned} \tag{6.1}$$

where $\mathscr{H} := \{f \in \mathrm{H}^1([0,1]) : f(0) = x_0 \text{ and } f(1) = x_0 + y\}$ and $y > 0$. In other words, for $y > 0$ we have the asymptotics

$$\mathbb{P}[X_t \geq x_0 + y] \simeq \mathrm{e}^{\frac{-I(x_0+y)}{t}}, \quad t \searrow 0, \tag{6.2}$$

where $\simeq$ stands for exponential equivalence.

Let Σ be an antiderivative of the function $1/\sigma$. By the assumptions on σ, the function $\Sigma(f(\cdot))$ belongs to $\mathrm{H}^1([0,1])$ iff the function f belongs to $\mathrm{H}^1([0,1])$. Hence, the latter infimum in (6.1) can be rewritten as

$$\inf_{v \in \mathscr{H}'} \frac{1}{2} \int_0^1 v'(s)^2 \,\mathrm{d}s,$$

where $\mathscr{H}' := \{v \in \mathrm{H}^1([0,1]) : v(0) = \Sigma(x_0) \text{ and } v(1) = \Sigma(x_0 + y)\}$. Due to Jensen's inequality, the infimum is reached when v is the affine function connecting $\Sigma(x_0)$ and $\Sigma(x_0 + y)$. Thus,

$$I(x_0 + y) = \frac{1}{2}[\Sigma(x_0 + y) - \Sigma(x_0)]^2 = \frac{1}{2}\Big(\int_{x_0}^{x_0+y} \frac{\mathrm{d}s}{\sigma(s)}\Big)^2. \tag{6.3}$$

We now *formally* apply the LDP (6.2) with a time-dependent $y = z\sqrt{t}$, where $z > 0$. Since $I(x_0) = I'(x_0) = 0$, we have

$$I(x_0 + z\sqrt{t}) = \tfrac{1}{2} I''(x_0) z^2 t + \mathrm{o}(t), \quad t \searrow 0,$$

and

$$\mathbb{P}\Big[\frac{X_t - x_0}{\sqrt{t}} \geq z\Big] \simeq \mathrm{e}^{-\frac{z^2}{2\sigma(x_0)^2} + \mathrm{O}(1)}, \quad t \searrow 0,$$

which suggests a Gaussian limit law (the case $z < 0$ is similar).

Remark 6.4. This example is given in [Ger+15, Remark 6, p. 730 et seq.].

Chapter 7

Motivation for the Research in Part III

We describe the research question of Part III and show how the results can be applied.

Limit theorems for finite-dimensional stochastic processes as time goes to infinity have been a classical object of study in probability theory. Many results on the existence and uniqueness of invariant distributions, the convergence of the processes to the latter, and the limiting behavior of the fluctuations around the limiting distributions have been obtained, see e.g. [Has80; JS03; MT92; MT93a; MT93b] and the references therein. More recently, small-time asymptotics of finite-dimensional continuous time stochastic processes have attracted attention. Apart from the theoretical interest, these have become important in various applied fields such as mathematical finance, where the increasingly high frequency of trades in financial markets requires pricing models behaving reasonably both on very short and on long time horizons.

In the works [ALV07; BGM09; BC12; BBF04; FFF10; FLH09; Jac07] and the references therein, the authors study the behavior of the random variables $\mathbb{E}[f(X_{t_0+\delta})|\mathscr{F}^X_{t_0}]$ for small values of $\delta > 0$. In this case, X is a finite-dimensional (jump-) diffusion process, a Lévy process or more generally a semimartingale, $(\mathscr{F}^X_t)_{t\geq 0}$ is the filtration generated by X, and f is taken from a space of suitable real-valued test functions. In [BC12], this program is carried out for general finite-dimensional semimartingales. Under appropriate continuity assumptions on the characteristics of X, as well as smoothness assumptions on the function f, the a.s. limit

$$\lim_{\delta\searrow 0} \delta^{-1}\big(\mathbb{E}[f(X_{t_0+\delta})|\mathscr{F}^X_{t_0}] - f(X_{t_0})\big) \tag{7.1}$$

is determined.

We are interested in small-time CLTs for finite dimensional semimartingales; this means, instead of the a.s. limit (7.1) we are concerned with the limit

$$\lim_{t\searrow 0} t^{-\frac{1}{2}}\big(f(X_t) - f(X_0)\big) \tag{7.2}$$

in distribution. We give sufficient conditions on the semimartingale X under which, for every suitable test function f, the limit (7.2) exists and is given by a centered normal random variable (whose variance depends on the particular choice of the function f). The most closely related result in literature seems to be by Doney and Maller [DM02, Theorem 2.5], which characterizes the Lévy processes that satisfy a small-time CLT.

In addition to the just described CLTs, we prove functional CLTs on the process level and give two applications of our results in the field of mathematical finance: first to the pricing of digital options and second to the asymptotics of implied volatility skews. To outline the first of the two applications, we recall that the price of a digital option with strike K and maturity t on an underlying security with price process X in the presence of a constant interest rate $r > 0$ is given by the formula

$$\mathbb{E}[\mathrm{e}^{-rt}\, 1_{\{X_t > K\}}] = \mathrm{e}^{-rt}\, \mathbb{P}[X_t > K]\,. \tag{7.3}$$

For short maturities, i.e. for $t \searrow 0$, this price tends to zero if $K > X_0$ (out-of-the-money) and to 1 if $K < X_0$ (in-the-money) as soon as X has right-continuous sample paths. The evaluation of the limit in the case $K = X_0$ (at-the-money options) is much trickier and, as we show, the limit can take all values in the interval $[0, 1]$, see Examples 2.13 and 2.17 in Part III. Nevertheless, if a CLT of the type described above holds for the semimartingale X and the limit law is non-singular, then the limit must be given by $1/2$. Moreover, in a special case the price in (7.3) is bounded for any fixed value of $t > 0$ from above and below by explicit functions tending to $1/2$ in the limit $t \searrow 0$. By a well known relation between digital prices and implied volatility skews, we deduce bounds on the latter in certain models with stochastic interest rates.

Remark 7.4. This chapter was taken from [Ger+15].

Part II

Analysis of Poisson Mixture Sums via Stein's Method

Chapter 1

The Setting

We introduce Poisson mixture distributions and normal variance mixture distributions on which the research in Chapter 2 and Chapter 3 is based.

Definition 1.1 (Poisson mixture distribution)**.** Let Λ be a random variable satisfying

$$\mathbb{P}[\Lambda > 0] = 1,$$

and for $\lambda \geq 0$ let N_λ be a $\mathbb{N}_0$-valued random variable which is conditionally Poisson distributed given Λ,

$$\mathbb{P}[N_\lambda = n \,|\, \Lambda] \overset{\text{a.s.}}{=} \frac{(\lambda\Lambda)^n}{n!}\,\mathrm{e}^{-\lambda\Lambda}, \qquad n \in \mathbb{N}_0. \tag{1.2}$$

Then the random variable N_λ is said to have a *Poisson mixture* distribution with mixing variable Λ.

Remark 1.3. Note that if the random variable N_λ carries a Poisson mixture distribution with mixing variable Λ, then $\mathbb{E}[N_\lambda \,|\, \Lambda] \overset{\text{a.s.}}{=} \lambda\Lambda$ and hence $\mathbb{E}[N_\lambda] = \mathbb{E}\big[\mathbb{E}[N_\lambda \,|\, \Lambda]\big] = \lambda\,\mathbb{E}[\Lambda] \in [0,\infty]$.

Remark 1.4. Poisson mixture distributions are applied in actuarial science. In contrast to a Poisson random variable N, where $\mathrm{V}(N) = \mathbb{E}[N]$, Poisson mixture random variables show over-dispersion, i.e. $\mathrm{V}(N) > \mathbb{E}[N]$, if the mixing variable is non-singular. This behavior is often encountered in count data. See [MFE05, Section 10.2.4, p. 482 et seqq.].

Example 1.5 (Gamma distribution for Λ)**.** Suppose that Λ has a gamma distribution

$$\Lambda \sim \Gamma(\alpha,\beta), \quad \alpha,\beta > 0,$$

with density

$$f_{\alpha,\beta}(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1}\,\mathrm{e}^{-\beta x}, \qquad x > 0.$$

For $\gamma \in (-\alpha,\infty)$ and $z \in (-\infty,\beta)$,

$$\begin{aligned}
\mathbb{E}[\Lambda^\gamma\,\mathrm{e}^{z\Lambda}] &= \int_0^\infty x^\gamma\,\mathrm{e}^{xz}\,\frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1}\,\mathrm{e}^{-\beta x}\,\mathrm{d}x \\
&= \frac{\Gamma(\alpha+\gamma)}{\Gamma(\alpha)}\,\frac{\beta^\alpha}{(\beta-z)^{\alpha+\gamma}} \int_0^\infty \frac{(\beta-z)^{\alpha+\gamma}}{\Gamma(\alpha+\gamma)} x^{\alpha+\gamma-1}\,\mathrm{e}^{-(\beta-z)x}\,\mathrm{d}x \\
&= \frac{\Gamma(\alpha+\gamma)}{\beta^\gamma\,\Gamma(\alpha)} (1 - z/\beta)^{-(\alpha+\gamma)} \int_0^\infty f_{\alpha+\gamma,\beta-z}(x)\,\mathrm{d}x \\
&= \frac{\Gamma(\alpha+\gamma)}{\beta^\gamma\,\Gamma(\alpha)} (1 - z/\beta)^{-(\alpha+\gamma)},
\end{aligned} \tag{1.6}$$

which extends to all $z \in \mathbb{C}$ with $\mathrm{Re}(z) < \beta$. For $\gamma = 1$ and $z = 0$ this implies $\mathbb{E}[\Lambda] = \alpha/\beta$ by the functional equation $\Gamma(\alpha+1) = \alpha\,\Gamma(\alpha)$ of the gamma function. Combining (1.2) and (1.6) with $\gamma = n$ and $z = -\lambda$ shows that the unconditional distribution of N_λ is

$$\mathbb{P}[N_\lambda = n] = \mathbb{E}\big[\mathbb{P}[N_\lambda = n|\Lambda]\big] = \frac{\lambda^n}{n!}\,\mathbb{E}\big[\Lambda^n\,\mathrm{e}^{-\lambda\Lambda}\big] = \frac{\Gamma(\alpha+n)}{n!\,\Gamma(\alpha)}\frac{\lambda^n}{\beta^n(1+\lambda/\beta)^{\alpha+n}}$$

for all $n \in \mathbb{N}_0$. Using n times the functional equation of the gamma function and the abbreviation

$$p_\lambda := \frac{\lambda}{\beta+\lambda} \in [0,1)$$

for the success probability yields

$$\mathbb{P}[N_\lambda = n] = \binom{\alpha+n-1}{n}(1-p_\lambda)^\alpha p_\lambda^n, \qquad n \in \mathbb{N}_0, \tag{1.7}$$

thus, N_λ carries the negative binomial distribution

$$N_\lambda \sim \mathrm{NB}(\alpha, p_\lambda).$$

In case $\alpha = 1$, N_λ carries the geometric distribution[1]

$$N_\lambda \sim \mathrm{G}(p_\lambda).$$

Remark 1.8. Similar calculations are carried out in [MFE05, Proposition 10.20, p. 483].

Definition 1.9. [Normal variance mixture distribution] A random variable Z has a *normal variance mixture distribution* with parameters $\mu \in \mathbb{R}$ and $\sigma \in \mathbb{R}$, if there exists a standard Gaussian random variable $X \sim \mathrm{N}(0,1)$ and a random variable $\Lambda \geq 0$, independent of X, such that

$$Z \overset{\mathrm{d}}{=} \mu + \sigma\sqrt{\Lambda}X\,.$$

Remark 1.10. Normal variance mixture distributions are applied in financial modeling, see e.g. Example 3.1 in Part I and [MFE05, Chapter 3.2., p. 73 et seqq.].

Lemma 1.11. *Let Λ be an a.s. positive random variable independent of $X \sim \mathrm{N}(0,\sigma^2)$. Then the characteristic function of $Z := \sqrt{\Lambda}X$ is given by*

$$\mathbb{E}\big[\mathrm{e}^{\mathrm{i}tZ}\big] = \mathbb{E}\big[\mathrm{e}^{-\frac{\sigma^2t^2}{2}\Lambda}\big], \qquad t \in \mathbb{R}. \tag{1.12}$$

Proof. Conditioning on Λ and using the independence of Λ and X yields

$$\mathbb{E}\big[\mathrm{e}^{\mathrm{i}tZ}\big] = \mathbb{E}\big[\mathbb{E}\big[\mathrm{e}^{\mathrm{i}t\sqrt{\Lambda}X}\,\big|\,\Lambda\big]\big] = \mathbb{E}\big[\mathrm{e}^{-\frac{\sigma^2t^2}{2}\Lambda}\big], \quad t \in \mathbb{R}.$$

□

Example 1.13 (Gamma distribution for Λ)**.** Let Λ be a gamma distributed random variable

$$\Lambda \sim \Gamma(\alpha,\beta), \quad \alpha,\beta > 0,$$

[1] See Appendix B.

as in Example 1.5, let $X \sim \mathrm{N}(0, \sigma^2)$, $\sigma > 0$, and let $Z = \sqrt{\Lambda} X$. Then (1.6) with $\gamma = 0$ and $z = -\frac{\sigma^2 t^2}{2}$ and (1.12) imply

$$\mathbb{E}\big[\mathrm{e}^{\mathrm{i}tZ}\big] = \mathbb{E}\Big[\mathrm{e}^{-\frac{\sigma^2 t^2}{2}\Lambda}\Big] = \Big(1 + \frac{\sigma^2 t^2}{2\beta}\Big)^{-\alpha}, \qquad t \in \mathbb{R}.$$

In case $\alpha = 1$, Z carries a centered Laplace distribution

$$Z \sim \mathrm{L}\Big(0, \frac{\sigma}{\sqrt{2\beta}}\Big).$$

Note, this was also established in Remark 3.10 in Part I.

Example 1.14 (Inverse gamma distribution for Λ)**.** Let Λ have an inverse gamma distribution

$$\Lambda \sim \mathrm{Ig}(\alpha, \beta), \quad \alpha, \beta > 0,$$

meaning that $1/\Lambda \sim \Gamma(\alpha, \beta)$ and

$$\mathbb{P}[\,\Lambda \le y\,] = \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^y x^{-\alpha-1}\,\mathrm{e}^{-\beta/x}\,\mathrm{d}x\,, \quad y \ge 0.$$

If $\alpha = \beta = \nu/2$ with $\nu > 0$ and $X \sim \mathrm{N}(0,1)$, then the corresponding normal variance mixture distribution of $\sqrt{\Lambda} Z$ is the Student's t-distribution $t(\nu)$ with $\nu > 0$ degrees of freedom. The special case $\nu = 1$ yields the Cauchy distribution. As a reference, see [MFE05, Example 3.7, p.75 and Section A.2.6, p. 497].

Example 1.15 (Tempered α-stable distribution for Λ)**.** Suppose that Λ has a τ-tempered α-stable distribution with index $\alpha \in (0,1)$, scale parameter $\rho > 0$ and tempering parameter $\tau \ge 0$, which means its Laplace transform is given by

$$\mathbb{E}\big[\mathrm{e}^{-s\Lambda}\big] = \exp(-\gamma_{\alpha,\rho}\big((s+\tau)^\alpha - \tau^\alpha\big)\big), \qquad s \ge -\tau,$$

where $\gamma_{\alpha,\rho} = \frac{\rho^\alpha}{\cos(\alpha\pi/2)}$. Then (1.12) evaluates to

$$\mathbb{E}\big[\mathrm{e}^{\mathrm{i}tZ}\big] = \exp\Big(-\gamma_{\alpha,\rho}\Big(\Big(\frac{\sigma^2 t^2}{2} + \tau\Big)^\alpha - \tau^\alpha\Big)\Big), \qquad t \in \mathbb{R}. \tag{1.16}$$

The special case $\alpha = 1/2$ and $\tau = 0$ is the Lévy distribution with scale parameter $\rho > 0$, for which

$$f(x) = \Big(\frac{\rho}{2\pi x^3}\Big)^{1/2} \mathrm{e}^{-\frac{\rho}{2x}}, \qquad x > 0,$$

is a density and (1.16) simplifies to

$$\mathbb{E}\big[\mathrm{e}^{\mathrm{i}tZ}\big] = \mathrm{e}^{-\sqrt{\rho}\sigma t}, \qquad t \in \mathbb{R}.$$

As a reference, see [GSW10].

Chapter 2

Upper Bounds for the Wasserstein Distance

We prove an upper bound for the Wasserstein distance between Poisson mixture sums and their related normal variance mixture distributions; a NCLT follows as a byproduct.

Let $\{X_i\}_{i\in\mathbb{N}}$ be a sequence of independent (but not necessarily identically distributed), real-valued square-integrable random variables with

$$\mathbb{E}[X_i] = 0, \quad \mathrm{V}(X_i) = 1, \quad \mathbb{E}\left[|X_i|^3\right] < \infty \quad i \in \mathbb{N}.$$

Furthermore, let the random variable N_λ have a Poisson mixture distribution[1]

$$N_\lambda \sim \mathrm{P}(\lambda\Lambda), \quad \lambda > 0,$$

with a.s. positive mixing random variable Λ and assume that (N_λ, Λ) is independent of $\{X_i\}_{i\in\mathbb{N}}$.

We apply Stein's method to prove an upper bound for the Wasserstein distance between the distributions of the random sums

$$Z_\lambda := \frac{1}{\sqrt{\lambda}} \sum_{i=1}^{N_\lambda} X_i, \quad \lambda > 0,$$

and the normal variance mixture random variable[2]

$$Z := \sqrt{\Lambda} X,$$

where $X \sim \mathrm{N}(0,1)$, independent of $(N_\lambda, \Lambda, \{X_i\}_{i\in\mathbb{N}})$. A byproduct is the NCLT

$$Z_\lambda \xrightarrow{\mathrm{d}} Z \quad \text{as} \quad \lambda \to \infty,$$

see Corollary 2.12.

We define

$$\varrho(\lambda) := \mathrm{e}^{-\lambda} \sum_{n=1}^{\infty} \frac{\lambda^{n-1}}{n!} \sum_{i=1}^{n} \mathbb{E}\left[|X_i|^3\right], \quad \lambda > 0, \tag{2.1}$$

[1] See Definition 1.1.

[2] See Definition 1.9.

which is allowed to be infinite and which can be seen as an average of the third absolute moments with Poisson weights. If $P_\lambda \sim \mathrm{P}(\lambda)$, then $\mathbb{E}[P_\lambda] = \lambda$ and

$$\varrho(\lambda) = \frac{1}{\mathbb{E}[P_\lambda]}\,\mathbb{E}\Big[\sum_{i=1}^{P_\lambda} \mathbb{E}\big[|X_i|^3\big]\Big]. \tag{2.2}$$

Note that

$$\varrho(\lambda) \le \sup_{i\in\mathbb{N}} \mathbb{E}[|X_i|^3], \quad \lambda > 0.$$

Theorem 2.3. *Define $Y_\lambda := Z_\lambda/\sqrt{\Lambda}$; on the set $\{\varrho(\lambda\Lambda) < \infty\}$ the conditional expectation $\mathbb{E}\big[|Y_\lambda|\big|\Lambda\big]$ is a.s. finite, the Wasserstein distance of $\mathscr{L}(Y_\lambda\,|\,\Lambda)$ to the standard normal distribution $\mathscr{L}(X) = \mathrm{N}(0,1)$ is a.s. well defined, and*

$$\mathrm{d_W}\big(\mathscr{L}(Y_\lambda\,|\,\Lambda), \mathscr{L}(X)\big) \le \frac{4 + 2\varrho(\lambda\Lambda)}{\sqrt{\lambda\Lambda}} \quad a.s. \tag{2.4}$$

If $\mathbb{E}[\varrho(\lambda\Lambda)] < \infty$, then

$$\sup_{b>0}\sup_{h\in\mathscr{H}_b} \mathbb{E}[h(Z_\lambda) - h(Z)] \le \frac{4 + 2\,\mathbb{E}[\varrho(\lambda\Lambda)]}{\sqrt{\lambda}}, \tag{2.5}$$

where $\mathscr{H}_b$ denotes the set of all absolutely continuous functions $h\colon \mathbb{R} \to \mathbb{R}$ with $\|h\|_\infty \le b$ and Lipschitz constant $\mathrm{Lip}(h) \le 1$.

Remark 2.6. Note that Z_λ and Z are coupled through their joint dependence on Λ. This makes the bound (2.5) possible even when Z_λ and Z are not integrable. Also note that due to this coupling we do not need a special Stein equation for the limiting normal variance mixture distribution, as it is enough to have the Stein equation for $\mathrm{N}(0,1)$.

Remark 2.7. Note that the assumptions on $\varrho(\lambda\Lambda)$ in Theorem 2.3 and the three corollaries below are trivially satisfied if $\sup_{i\in\mathbb{N}} \mathbb{E}[|X_i|^3] < \infty$. See Example 2.25 below for a more elaborate case with absolute third moments increasing to infinity.

Corollary 2.8. *If $\mathbb{E}[\sqrt{\Lambda}] < \infty$, then Z is integrable. If additionally $\mathbb{E}[\varrho(\lambda\Lambda)] < \infty$ for some $\lambda > 0$, then Z_λ is integrable too, the Wasserstein distance of $\mathscr{L}(Z_\lambda)$ and $\mathscr{L}(Z)$ is well defined, and*

$$\mathrm{d_W}(\mathscr{L}(Z_\lambda), \mathscr{L}(Z)) := \sup_{h\in\mathscr{H}} \mathbb{E}[h(Z_\lambda) - h(Z)] \le \frac{4 + 2\,\mathbb{E}[\varrho(\lambda\Lambda)]}{\sqrt{\lambda}}, \tag{2.9}$$

where $\mathscr{H}$ denotes the set of all $h\colon \mathbb{R} \to \mathbb{R}$ with $\mathrm{Lip}(h) \le 1$.

Proof. Note that $\mathbb{E}[|Z|] = \mathbb{E}[\sqrt{\Lambda}]\,\mathbb{E}[|X|]$ and $\mathbb{E}[|X|] = \sqrt{2/\pi}$. If in addition $\mathbb{E}[\varrho(\lambda\Lambda)] < \infty$, then (2.5) implies that

$$\mathbb{E}[\min(b, |Z_\lambda|)] \le \mathbb{E}[\min(b, |Z|)] + \frac{4 + 2\,\mathbb{E}[\varrho(\lambda\Lambda)]}{\sqrt{\lambda}},$$

and integrability of Z_λ follows by sending $b \to \infty$ and using the monotone convergence theorem. For $h \in \mathscr{H}$ and $b > 0$, the function $h_b(x) := \max(-b, \min(b, h(x)))$ for $x \in \mathbb{R}$ is in $\mathscr{H}_b$ and the upper bound in (2.5) applies. Since

$$|h_b(x)| \le |h(x)| \le |h(0)| + |h(x) - h(0)| \le |h(0)| + |x|, \quad x \in \mathbb{R}, \tag{2.10}$$

$\mathbb{E}[h_b(Z_\lambda) - h_b(Z)]$ converges to $\mathbb{E}[h(Z_\lambda) - h(Z)]$ by the dominated convergence theorem as $b \to \infty$. This proves (2.9). □

Corollary 2.11 (NCLT)**.** *If* $\mathbb{E}[\varrho(\lambda\Lambda)] = o(\sqrt{\lambda})$ *as* $\lambda \to \infty$*, then* Z_λ *converges weakly to* Z*.*

Proof. Consider a bounded $g\colon \mathbb{R} \to \mathbb{R}$ with $\mathrm{Lip}(g) < \infty$. Then there exists a bounded $h\colon \mathbb{R} \to \mathbb{R}$ with $\mathrm{Lip}(h) \le 1$ and $g = \mathrm{Lip}(g)h$. By (2.5),

$$\mathbb{E}[g(Z_\lambda)] \to \mathbb{E}[g(Z)] \quad \text{as} \quad \lambda \to \infty.$$

This implies the corollary, see [EK86, Theorem 3.1, proof of (c) implies (d)]. □

Corollary 2.12. *Suppose that* $\mathbb{E}[\sqrt{\Lambda}] < \infty$*. Let* $\{\lambda_n\}_{n\in\mathbb{N}} \subset (0,\infty)$ *be a sequence with* $\lambda_n \to \infty$ *as* $n \to \infty$*. If* $\mathbb{E}[\varrho(\lambda_n\Lambda)] = \mathrm{o}(\sqrt{\lambda_n})$ *as* $n \to \infty$*, then* $\{Z_{\lambda_n}\}_{n\in\mathbb{N}}$ *is uniformly integrable.*

Proof. For $b > 0$ we define

$$h_b(x) = \max\big(0, |x| - \max(0, b(b - |x|))\big), \quad x \in \mathbb{R}.$$

Then $h_b(x) = |x|$ for $|x| \ge b$, $h_b(x) = 0$ for $|x| \le b^2/(b+1)$ and linear for the remaining $|x|$, in particular $\mathrm{Lip}(h_b) = b + 1$. Furthermore

$$|x| 1_{[b,\infty)}(|x|) \le h_b(x) \le |x|, \quad x \in \mathbb{R},$$

and

$$\lim_{b\to\infty} h_b(x) = 0, \quad x \in \mathbb{R}.$$

Fix $\varepsilon > 0$; since Z is integrable by Corollary 2.8, by dominated convergence there exists $b_\varepsilon > 0$, such that $\mathbb{E}[h_{b_\varepsilon}(Z)] \le \varepsilon/2$. By assumption, there exists $n_\varepsilon \in \mathbb{N}$, such that

$$\mathrm{Lip}(h_{b_\varepsilon})\frac{4 + 2\,\mathbb{E}[\varrho(\lambda_n\Lambda)]}{\sqrt{\lambda_n}} \le \varepsilon/2.$$

Hence, by (2.9)

$$\mathbb{E}[|Z_{\lambda_n}| 1_{\{|Z_{\lambda_n}| > b_\varepsilon\}}] \le \mathbb{E}[h_{b_\varepsilon}(Z_{\lambda_n})] \le \mathbb{E}[h_{b_\varepsilon}(Z)] + \mathrm{Lip}(h_{b_\varepsilon})\frac{4 + 2\,\mathbb{E}[\varrho(\lambda_n\Lambda)]}{\sqrt{\lambda_n}} \le \varepsilon$$

for all $n > n_\varepsilon$. Since $Z_{\lambda_1}, \ldots, Z_{\lambda_{n_\varepsilon}}$ are integrable by Corollary 2.8, by dominated convergence (possibly for a larger b_ε) also $\mathbb{E}[|Z_{\lambda_n}| 1_{\{|Z_{\lambda_n}| > b_\varepsilon\}}] \le \varepsilon$ for every $n \in \{1, \ldots, n_\varepsilon\}$. This proves $\sup_{n\in\mathbb{N}} \mathbb{E}[|Z_{\lambda_n}| 1_{\{|Z_{\lambda_n}| > b_\varepsilon\}}] \le \varepsilon$. □

Proof of Theorem 2.3. Let us first prove that $\mathbb{E}[\varrho(\lambda\Lambda)] < \infty$ and (2.4) imply (2.5). Using the scaling property of the Wasserstein metric and $\mathscr{L}(Z/\sqrt{\Lambda}\,|\,\Lambda) \overset{\text{a.s.}}{=} \mathscr{L}(X)$, it follows for every $b > 0$ and $h \in \mathscr{H}_b$ that

$$\begin{aligned}
\mathbb{E}[h(Z_\lambda) - h(Z)] &= \mathbb{E}\big[\mathbb{E}[h(Z_\lambda) - h(Z)\,|\,\Lambda]\big] \\
&\le \mathbb{E}\big[d_{\mathrm{W}}\big(\mathscr{L}(Z_\lambda\,|\,\Lambda), \mathscr{L}(Z\,|\,\Lambda)\big)\big] \\
&= \mathbb{E}\big[\sqrt{\Lambda}\, d_{\mathrm{W}}\big(\mathscr{L}(Y_\lambda\,|\,\Lambda), \mathscr{L}(X)\big)\big].
\end{aligned}$$

Plugging in (2.4) gives the upper bound (2.5).

To prove (2.4), consider $h \in \mathscr{H}_b$. The function

$$f(x) := \mathrm{e}^{\frac{x^2}{2}} \int_{-\infty}^{x} \big(h(y) - \mathbb{E}[h(X)]\big)\, \mathrm{e}^{-\frac{y^2}{2}}\, \mathrm{d}y, \qquad x \in \mathbb{R},$$

where we dropped the dependence on h in the notation, satisfies the corresponding Stein equation for the centered normal distribution with variance 1, i.e.

$$h(x) - \mathbb{E}[h(X)] = f'(x) - xf(x), \qquad x \in \mathbb{R}.$$

Note that $\|f\|_\infty \leq 2$ and $\|f'\|_\infty \leq \sqrt{2/\pi}$ by [CGS11, (2.13) of Lemma 2.4]. Therefore, $\mathbb{E}[|Y_\lambda f(Y_\lambda)|] \leq \sqrt{2/\pi} + 2b < \infty$, hence $Y_\lambda f(Y_\lambda)$ is integrable and

$$\mathbb{E}[h(Y_\lambda)\,|\,\Lambda] - \mathbb{E}[h(X)] \overset{\text{a.s.}}{=} \mathbb{E}\big[f'(Y_\lambda) - Y_\lambda f(Y_\lambda)\,\big|\,\Lambda\big]. \tag{2.13}$$

To estimate the right-hand side of (2.13), we follow the standard procedure (see [CGS11, Section 1.3]) combined with size biasing of a Poisson random variable. Define

$$Y'_\lambda := \frac{1}{\sqrt{\lambda\Lambda}} \sum_{i=1}^{N_\lambda - 1} X_i \qquad \text{and} \qquad Y_{\lambda,n} := \frac{1}{\sqrt{\lambda\Lambda}} \sum_{i=1, i\neq n}^{N_\lambda} X_i \tag{2.14}$$

and note that

$$Y'_\lambda - Y_{\lambda,n} = \frac{1}{\sqrt{\lambda\Lambda}} (X_n 1_{\{N_\lambda \geq n\}} - X_{N_\lambda}), \qquad n \in \mathbb{N}. \tag{2.15}$$

We first rewrite the term $\mathbb{E}[f'(Y_\lambda)\,|\,\Lambda]$ from (2.13). By size biasing,

$$\mathbb{E}\big[f'(Y_\lambda)\,|\,\Lambda\big] \overset{\text{a.s.}}{=} \mathbb{E}\Big[\frac{N_\lambda}{\lambda\Lambda} f'(Y'_\lambda)\,\Big|\,\Lambda\Big] \overset{\text{a.s.}}{=} \mathbb{E}\Big[\frac{1}{\lambda\Lambda}\sum_{n=1}^{N_\lambda} f'(Y'_\lambda)\,\Big|\,\Lambda\Big]. \tag{2.16}$$

By the trivial identity $f(Y'_\lambda) = \big(f(Y'_\lambda) - f(Y_{\lambda,n})\big) + f(Y_{\lambda,n})$,

$$\mathbb{E}\big[f'(Y_\lambda)\,|\,\Lambda\big] \overset{\text{a.s.}}{=} I_1 + \mathbb{E}\Big[\frac{1}{\lambda\Lambda}\sum_{n=1}^{N_\lambda} f'(Y_{\lambda,n})\,\Big|\,\Lambda\Big]$$

with

$$I_1 := \mathbb{E}\Big[\frac{1}{\lambda\Lambda}\sum_{n=1}^{N_\lambda}\big(f'(Y'_\lambda) - f'(Y_{\lambda,n})\big)\,\Big|\,\Lambda\Big].$$

By the independence of X_n from Λ, N_λ, and $\{X_i\}_{i\in\mathbb{N}\setminus\{n\}}$, it follows that

$$\mathbb{E}\big[g(X_n)\,\big|\,\Lambda, N_\lambda, \{X_i\}_{i\in\mathbb{N}\setminus\{n\}}\big] \overset{\text{a.s.}}{=} \mathbb{E}[g(X_n)] \tag{2.17}$$

for every measurable $g\colon \mathbb{R} \to \mathbb{R}$ with $\mathbb{E}[|g(X_n)|] < \infty$. Using (2.17) for $g(x) = x^2$ and the assumption $\mathbb{E}[X_n^2] = 1$, we see that

$$\mathbb{E}\Big[\frac{X_n^2 1_{\{N_\lambda \geq n\}}}{\lambda\Lambda} f'(Y_{\lambda,n})\,\Big|\,\Lambda\Big] \overset{\text{a.s.}}{=} \mathbb{E}\Big[\frac{1_{\{N_\lambda \geq n\}}}{\lambda\Lambda} f'(Y_{\lambda,n})\,\Big|\,\Lambda\Big], \qquad n \in \mathbb{N},$$

hence

$$\mathbb{E}\big[f'(Y_\lambda)\,|\,\Lambda\big] \overset{\text{a.s.}}{=} I_1 + \mathbb{E}\Big[\frac{1}{\lambda\Lambda}\sum_{n=1}^{N_\lambda} X_n^2 f'(Y_{\lambda,n})\,\Big|\,\Lambda\Big]. \tag{2.18}$$

Next, we rewrite the term $\mathbb{E}[Y_\lambda f(Y_\lambda)\,|\,\Lambda]$ from (2.13). By the fundamental theorem of calculus we get

$$f(Y_\lambda) = f(Y_{\lambda,n}) + (Y_\lambda - Y_{\lambda,n}) \int_0^1 f'(Y_{\lambda,n} + t(Y_\lambda - Y_{\lambda,n}))\,\mathrm{d}t, \qquad n \in \mathbb{N}. \tag{2.19}$$

Using (2.17) for $g(x) = x$ and the assumption $\mathbb{E}[X_n] = 0$, it follows that

$$\mathbb{E}\Big[\frac{X_n 1_{\{N_\lambda \ge n\}}}{\sqrt{\lambda\Lambda}} f(Y_{\lambda,n}) \Big| \Lambda\Big] \overset{\text{a.s.}}{=} 0.$$

Hence, multiplying (2.19) by $Y_\lambda - Y_{\lambda,n} = X_n 1_{\{N_\lambda \ge n\}}/\sqrt{\lambda\Lambda}$ and taking conditional expectations, it follows that

$$\mathbb{E}\Big[\frac{X_n 1_{\{N_\lambda \ge n\}}}{\sqrt{\lambda\Lambda}} f(Y_\lambda) \Big| \Lambda\Big] \overset{\text{a.s.}}{=} \mathbb{E}\Big[\frac{X_n^2 1_{\{N_\lambda \ge n\}}}{\lambda\Lambda} \int_0^1 f'(Y_{\lambda,n} + t(Y_\lambda - Y_{\lambda,n}))\,\mathrm{d}t \Big| \Lambda\Big],$$

and summation over $n \in \mathbb{N}$ shows that

$$\mathbb{E}\big[Y_\lambda f(Y_\lambda)\,\big|\,\Lambda\big] \overset{\text{a.s.}}{=} \mathbb{E}\Big[\frac{1}{\lambda\Lambda}\sum_{n=1}^{N_\lambda} X_n^2 \int_0^1 f'(Y_{\lambda,n} + t(Y_\lambda - Y_{\lambda,n}))\,\mathrm{d}t \Big| \Lambda\Big].$$

Subtracting this equality from (2.18) shows that

$$\mathbb{E}\big[f'(Y_\lambda) - Y_\lambda f(Y_\lambda)\,|\,\Lambda\big] \overset{\text{a.s.}}{=} I_1 + I_2 \tag{2.20}$$

with

$$I_2 := \mathbb{E}\Big[\frac{1}{\lambda\Lambda}\sum_{n=1}^{N_\lambda} X_n^2 \int_0^1 \big(f'(Y_{\lambda,n}) - f'(Y_{\lambda,n} + t(Y_\lambda - Y_{\lambda,n}))\big)\,\mathrm{d}t \Big| \Lambda\Big].$$

To control I_1 and I_2, we use additional information about the derivative of f, namely $\mathrm{Lip}(f') < \infty$. Then by (2.15),

$$\big|f'(Y'_\lambda) - f'(Y_{\lambda,n})\big| \le \mathrm{Lip}(f')\big|Y'_\lambda - Y_{\lambda,n}\big| = \frac{\mathrm{Lip}(f')}{\sqrt{\lambda\Lambda}}\big|X_n 1_{\{N_\lambda \ge n\}} - X_{N_\lambda}\big| \tag{2.21}$$

and

$$\begin{aligned}\big|f'(Y_{\lambda,n}) - f'(Y_{\lambda,n} + t(Y_\lambda - Y_{\lambda,n}))\big| &\le \mathrm{Lip}(f')\,t\,\big|Y_\lambda - Y_{\lambda,n}\big| \\ &\le \frac{\mathrm{Lip}(f')}{\sqrt{\lambda\Lambda}}\,t\,|X_n| 1_{\{N_\lambda \ge n\}}.\end{aligned}$$

for every $n \in \mathbb{N}$ and $t \in [0,1]$. Since $\mathscr{L}(N_\lambda\,|\,\Lambda) \overset{\text{a.s.}}{=} \mathrm{P}(\lambda\Lambda)$,

$$\mathbb{E}\Big[\sum_{n=1}^{N_\lambda} 1_{\{N_\lambda \ge n\}} \Big| \Lambda\Big] \overset{\text{a.s.}}{=} \mathbb{E}[N_\lambda\,|\,\Lambda] \overset{\text{a.s.}}{=} \lambda\Lambda.$$

Hence, using (2.17) for $g(x) = |x|^3$, as well as $\int_0^1 t\,\mathrm{d}t = 1/2$, it follows that

$$\begin{aligned}|I_2| &\le \frac{\mathrm{Lip}(f')}{2\sqrt{\lambda\Lambda}}\,\mathbb{E}\Big[\frac{1}{\lambda\Lambda}\sum_{n=1}^{N_\lambda} |X_n|^3 1_{\{N_\lambda \ge n\}} \Big| \Lambda\Big] \\ &= \frac{\mathrm{Lip}(f')}{2\sqrt{\lambda\Lambda}}\,\mathbb{E}\Big[\frac{1}{\lambda\Lambda}\sum_{n=1}^{N_\lambda} \mathbb{E}\big[|X_n|^3\big] \Big| \Lambda\Big] \le \frac{\mathrm{Lip}(f')}{2\sqrt{\lambda\Lambda}}\,\varrho(\lambda\Lambda) \quad \text{a.s.}\end{aligned} \tag{2.22}$$

Estimating I_1 using (2.21),

$$|I_1| \le \frac{\mathrm{Lip}(f')}{\sqrt{\lambda\Lambda}}\,\mathbb{E}\Big[\frac{1}{\lambda\Lambda}\sum_{n=1}^{N_\lambda} \big(|X_n| + |X_{N_\lambda}|\big) \Big| \Lambda\Big] \quad \text{a.s.}$$

Using (2.17) for $g(x) = |x|$ and recalling that $\mathbb{E}[|X_n|] \le \mathbb{E}[X_n^2] = 1$ for every $n \in \mathbb{N}$ by Jensen's inequality and our assumption,

$$\begin{aligned}\mathbb{E}\Big[\sum_{n=1}^{N_\lambda}\big(|X_n| + |X_{N_\lambda}|\big)\,\Big|\,\Lambda\Big] &= \sum_{m=1}^{\infty}\sum_{n=1}^{m}\mathbb{E}\big[(|X_n| + |X_m|)1_{\{N_\lambda = m\}}\,\big|\,\Lambda\big]\\ &= \sum_{m=1}^{\infty}\mathbb{E}\big[1_{\{N_\lambda=m\}}\,\big|\,\Lambda\big]\sum_{n=1}^{m}(\mathbb{E}[|X_n|] + \mathbb{E}[|X_m|])\\ &\le 2\,\mathbb{E}[N_\lambda\,|\,\Lambda] = 2\lambda\Lambda \quad \text{a.s.}\end{aligned}$$

Hence,

$$|I_1| \le \frac{2\,\mathrm{Lip}(f')}{\sqrt{\lambda\Lambda}} \quad \text{a.s.} \tag{2.23}$$

Note that $\mathrm{Lip}(f') \le 2$ by [CGS11, (2.13) of Lemma 2.4], because $\mathrm{Lip}(h) \le 1$ by assumption. Thus, the combination of (2.13), (2.20), (2.22), and (2.23) shows that

$$\mathbb{E}[h(Y_\lambda)\,|\,\Lambda] - \mathbb{E}[h(X)] \le \frac{4 + 2\varrho(\lambda\Lambda)}{\sqrt{\lambda\Lambda}} \quad \text{a.s.} \tag{2.24}$$

On the set $\{\varrho(\lambda\Lambda) < \infty\}$, for $b > 0$, estimate (2.24) applied to $h(x) := \min(b, |x|)$ for $x \in \mathbb{R}$ implies that

$$\mathbb{E}\big[\min(b, |Y_\lambda|)\,\big|\,\Lambda\big] \le \mathbb{E}[|X|] + \frac{4 + 2\varrho(\lambda\Lambda)}{\sqrt{\lambda\Lambda}} \quad \text{a.s.,}$$

so $\mathbb{E}[|Y_\lambda|\,|\,\Lambda] < \infty$ a.s. by conditional dominated convergence as $b \to \infty$.

For general $h \in \mathscr{H}$ and $b > 0$ define $h_b \in \mathscr{H}_b$ by $h_b(x) = \max(-b, \min(b, h(x)))$ for $x \in \mathbb{R}$. By (2.10) and the conditional dominated convergence theorem, it follows that $\mathbb{E}[h_b(Y_\lambda)\,|\,\Lambda] - \mathbb{E}[h_b(X)]$ converges a.s. on the set $\{\varrho(\lambda\Lambda) < \infty\}$ to $\mathbb{E}[h(Y_\lambda)\,|\,\Lambda] - \mathbb{E}[h(X)]$ as $b \to \infty$, therefore (2.24) holds for h a.s. on the set $\{\varrho(\lambda\Lambda) < \infty\}$, and (2.4) is proven. □

To illustrate the balance between third absolute moments of the sequence $\{X_n\}_{n\in\mathbb{N}}$ increasing to infinity and the corresponding moments of Λ required for Theorem 2.3 and Corollaries 2.8, 2.11, and 2.12 we give the following example.

Example 2.25. Given $\gamma > 0$, consider a sequence of independent random variables $\{X_n\}_{n\in\mathbb{N}}$ with $\mathbb{P}[X_n = \mathrm{e}^{\gamma n}] = \mathbb{P}[X_n = -\mathrm{e}^{\gamma n}] = \mathrm{e}^{-2\gamma n}/2$ and $\mathbb{P}[X_n = 0] = 1 - \mathrm{e}^{-2\gamma n}$ for every $n \in \mathbb{N}$. Then $\mathbb{E}[X_n] = 0$, $\mathbb{E}[X_n^2] = 1$, and $\mathbb{E}[|X_n|^3] = \mathrm{e}^{\gamma n}$ for every $n \in \mathbb{N}$. For $\lambda > 0$ let $P_\lambda \sim \mathrm{P}(\lambda)$. Summing the first P_λ terms of a geometric progression with factor e^{γ} shows that

$$\sum_{n=1}^{P_\lambda}\mathbb{E}[|X_n|^3] = \frac{\mathrm{e}^{\gamma P_\lambda} - 1}{1 - \mathrm{e}^{-\gamma}},$$

hence by the definition of $\varrho(\lambda)$ in (2.2)

$$\varrho(\lambda) = \frac{\mathbb{E}[\mathrm{e}^{\gamma P_\lambda}] - 1}{\lambda(1 - \mathrm{e}^{-\gamma})} = \frac{\exp(\lambda(\mathrm{e}^{\gamma} - 1)) - 1}{\lambda(1 - \mathrm{e}^{-\gamma})} \le \mathrm{e}^{\gamma}\exp(\lambda(\mathrm{e}^{\gamma} - 1)),$$

where the estimate $(\mathrm{e}^x - 1)/x \le \mathrm{e}^x$ for $x > 0$ was used. Therefore, $\varrho(\lambda\Lambda) < \infty$ for every $(0, \infty)$-valued random variable Λ. If Λ has a gamma distribution $\Gamma(\alpha, \beta)$ with shape parameter $\alpha > 0$ and inverse scale parameter $\beta > 0$ as in Example 1.5, then using (1.6) yields

$$\mathbb{E}[\varrho(\lambda\Lambda)] \le \mathrm{e}^{\gamma}\,\mathbb{E}[\exp(\lambda\Lambda(\mathrm{e}^{\gamma}-1))] = \mathrm{e}^{\gamma}\Big(\frac{\beta}{\beta-\lambda(\mathrm{e}^{\gamma}-1)}\Big)^{\alpha}$$

for $\lambda < \beta/(\mathrm{e}^{\gamma}-1)$. Thus, the estimate (2.5) applies for $0 < \lambda < \beta/(\mathrm{e}^{\gamma}-1)$ and also Corollary 2.8, because $\mathbb{E}[\sqrt{\Lambda}] = \Gamma(\alpha + 1/2)/(\sqrt{\beta}\,\Gamma(\alpha)) < \infty$ by (1.6). However, the third absolute moments are increasing too fast to make Corollary 2.11 applicable. This is no surprise, because the expected number of non-zero terms in Z_λ is bounded above, i.e.

$$\mathbb{E}\Big[\sum_{n=1}^{N_\lambda} 1_{\{X_n\neq 0\}}\Big] \le \sum_{n=1}^{\infty}\mathbb{P}[X_n \neq 0] = \frac{1}{\mathrm{e}^{2\gamma}-1}, \qquad \lambda > 0,$$

for every distribution of the $(0,\infty)$-valued Λ.

Remark 2.26. This chapter is based on so far unpublished research by P. Eichelsbacher, P. Porkert, and U. Schmock.

Chapter 3

Upper Bounds for the Kolmogorov Distance

We prove an upper bound for the Kolmogorov distance between Poisson mixture sums and their related normal variance mixture distributions.

As in Chapter 2, let $\{X_i\}_{i\in\mathbb{N}}$ be a sequence of independent (but not necessarily identically distributed), real-valued square-integrable random variables with

$$\mathbb{E}[X_i] = 0, \quad \mathrm{V}(X_i) = 1, \quad \mathbb{E}\big[|X_i|^3\big] < \infty \quad i \in \mathbb{N}.$$

Furthermore, let the random variable N_λ have a Poisson mixture distribution[1]

$$N_\lambda \sim \mathrm{P}(\lambda\Lambda), \quad \lambda > 0,$$

with a.s. positive mixing random variable Λ and assume that (N_λ, Λ) is independent of $\{X_i\}_{i\in\mathbb{N}}$.

We apply Stein's method to prove an upper bound for the Kolmogorov distance between the distributions of the random sums

$$Z_\lambda := \frac{1}{\sqrt{\lambda}} \sum_{i=1}^{N_\lambda} X_i, \quad \lambda > 0,$$

and the normal variance mixture random variable[2]

$$Z := \sqrt{\Lambda} X,$$

where $X \sim \mathrm{N}(0,1)$, independent of $(N_\lambda, \Lambda, \{X_i\}_{i\in\mathbb{N}})$.

In analogy to (2.1), we define

$$\psi(\lambda) := \mathrm{e}^{-\lambda} \sum_{n=1}^{\infty} \frac{\lambda^{n-1}}{n!} \sum_{i=1}^{n+1} \mathbb{E}\big[|X_i|^3\big], \quad \lambda > 0.$$

[1] See Definition 1.1.

[2] See Definition 1.9.

Theorem 3.1. *Under the above conditions an upper bound for the Kolmogorov distance between the conditional distributions is given by*

$$\mathrm{d_K}\big(\mathscr{L}(Z_\lambda|\Lambda),\mathscr{L}(Z|\Lambda)\big) \le \frac{C\psi(\lambda\Lambda)+D}{\sqrt{\lambda\Lambda}} \quad a.s.$$

where

$$C = \frac{4+52\sqrt{6}+\sqrt{2\pi}}{8} \approx 16.7, \quad D = 2+\sqrt{6}+\sqrt{\frac{\pi}{2}} \approx 5.7.$$

Remark 3.2. Let $c := \sup_{i\in\mathbb{N}} \mathbb{E}\big[|X_i|^3\big] < \infty$, then

$$\psi(x) \le c\bigg(1+\frac{1-\mathrm{e}^{-x}}{x}\bigg) \le 2c, \quad x>0.$$

Since

$$F(x) := x\sum_{n=1}^{\infty}\frac{x^n}{n!} = x(\mathrm{e}^x-1), \quad x\in\mathbb{R},$$

and

$$F'(x) = \sum_{n=1}^{\infty}(n+1)\frac{x^n}{n!} = x\,\mathrm{e}^x+\mathrm{e}^x-1, \quad x\in\mathbb{R},$$

for $x>0$,

$$\psi(x) = \mathrm{e}^{-x}\sum_{n=1}^{\infty}\frac{x^{n-1}}{n!}\sum_{i=1}^{n+1}\mathbb{E}\big[|X_i|^3\big] \le \frac{c\,\mathrm{e}^{-x}}{x}F'(x),$$

which proves the claim.

Corollary 3.3. *Under the conditions of Theorem 3.1, an upper bound for the Kolmogorov distance between Z_λ and Z is given by*

$$\mathrm{d_K}\big(\mathscr{L}(Z_\lambda),\mathscr{L}(Z)\big) \le \mathbb{E}\Big[\frac{C\psi(\lambda\Lambda)+D}{\sqrt{\lambda\Lambda}}\wedge 1\Big]. \tag{3.4}$$

If the deterministic condition

$$\lim_{x\to\infty}\frac{\psi(x)}{\sqrt{x}} = 0$$

is satisfied, the right-hand side of (3.4) *converges to zero for $\lambda\to\infty$. If*

$$c := \sup_{i\in\mathbb{N}}\mathbb{E}\big[|X_i|^3\big] < \infty,$$

$$\mathrm{d_K}\big(\mathscr{L}(Z_\lambda),\mathscr{L}(Z)\big) \le \frac{2cC+D}{\sqrt{\lambda}}\,\mathbb{E}\big[\Lambda^{-\frac{1}{2}}\big].$$

Example 3.5. Let

$$\Lambda \sim \Gamma(\alpha,\beta), \quad \alpha > 1/2,\ \beta>0,$$

then by equation (1.6) in Example 1.5, with $\gamma = -1/2$ and $z=0$,

$$\mathbb{E}\big[\Lambda^{-\frac{1}{2}}\big] = \frac{\sqrt{\beta}\,\Gamma(\alpha-1/2)}{\Gamma(\alpha)}.$$

Proof of Theorem 3.1. Given $\sigma > 0$ and $z \in \mathbb{R}$ as well as the indicator function

$$h_z := 1_{(-\infty,z]},$$

a solution $f_{\sigma^2,z}$ of the Stein equation

$$h_z(x) - g_z(\sigma^2) = f'(x) - \frac{x}{\sigma^2} f(x), \qquad x \in \mathbb{R}, \tag{3.6}$$

with $g_z(\sigma^2) := \mathbb{E}[h_z(\sigma X)]$ is given by

$$f_{\sigma^2,z}(x) = \sigma\, \mathrm{e}^{\frac{x^2}{2\sigma^2}} \int_{-\infty}^{x/\sigma} \big(h_z(\sigma s) - g_z(\sigma^2)\big)\, \mathrm{e}^{-\frac{s^2}{2}}\, \mathrm{d}s, \qquad x \in \mathbb{R}, \tag{3.7}$$

where we consider f' as the left-sided derivative and

$$\mathbb{P}[Z_\lambda \le z \,|\, \Lambda] - \mathbb{P}\big[\sqrt{\Lambda} X \le z \,\big|\, \Lambda\big] \overset{\text{a.s.}}{=} \mathbb{E}\big[h_z(Z_\lambda) \,\big|\, \Lambda\big] - g_z(\Lambda). \tag{3.8}$$

For every $\lambda > 0$ and $n \in \mathbb{N}$, we define various variants of Z_λ where one or two terms are missing, namely

$$Z'_\lambda := \frac{1}{\sqrt{\lambda}} \sum_{i=1}^{N_\lambda - 1} X_i, \qquad Z_{\lambda,n} := \frac{1}{\sqrt{\lambda}} \sum_{\substack{i=1\\ i\neq n}}^{N_\lambda} X_i \qquad \text{and} \qquad Z'_{\lambda,n} := \frac{1}{\sqrt{\lambda}} \sum_{\substack{i=1\\ i\neq n}}^{N_\lambda - 1} X_i,$$

as well as auxiliary function

$$K_{\lambda,n}(t) := \mathbb{E}\Big[Y_{\lambda,n}\big(1_{\{0\le t\le Y_{\lambda,n}\}} - 1_{\{Y_{\lambda,n}\le t\le 0\}}\big)\Big], \qquad t \in \mathbb{R},$$

where

$$Y_{\lambda,n} := \frac{X_n}{\sqrt{\lambda}}. \tag{3.9}$$

Notice that

$$\int_{\mathbb{R}} K_{\lambda,n}(t)\,\mathrm{d}t = \mathbb{E}\big[Y_{\lambda,n}^2\big] = \frac{1}{\lambda} \qquad \text{and} \qquad \int_{\mathbb{R}} |t| K_{\lambda,n}(t)\,\mathrm{d}t = \frac{1}{2}\,\mathbb{E}\big[|Y_{\lambda,n}|^3\big]\,. \tag{3.10}$$

By size biasing, see (2.16), and (3.10),

$$\begin{aligned}
\mathbb{E}\big[h_z(Z_\lambda) - g_z(\Lambda) \,\big|\, \Lambda\big] &\overset{\text{a.s.}}{=} \mathbb{E}\Big[\frac{N_\lambda}{\lambda\Lambda}\big(h_z(Z'_\lambda) - g_z(\Lambda)\big) \,\Big|\, \Lambda\Big] \\
&\overset{\text{a.s.}}{=} \frac{1}{\Lambda}\,\mathbb{E}\Big[\sum_{n=1}^{N_\lambda} \int_{\mathbb{R}} \big(h_z(Z'_\lambda) - g_z(\Lambda)\big) K_{\lambda,n}(t)\,\mathrm{d}t \,\Big|\, \Lambda\Big] \overset{\text{a.s.}}{=} A + B
\end{aligned}$$

with

$$A := \frac{1}{\Lambda}\,\mathbb{E}\Big[\sum_{n=1}^{N_\lambda} \int_{\mathbb{R}} \big(h_z(Z'_\lambda) - h_z(Z_{\lambda,n} + t)\big) K_{\lambda,n}(t)\,\mathrm{d}t \,\Big|\, \Lambda\Big] \tag{3.11}$$

and

$$B := \frac{1}{\Lambda}\,\mathbb{E}\Big[\sum_{n=1}^{N_\lambda} \int_{\mathbb{R}} \big(h_z(Z_{\lambda,n} + t) - g_z(\Lambda)\big) K_{\lambda,n}(t)\,\mathrm{d}t \,\Big|\, \Lambda\Big]. \tag{3.12}$$

First we treat B from (3.12). Of course, we will use the Stein equation (3.6), but since the (left-hand side) derivative of $f_{\Lambda,z}$ given by (3.7) jumps at z, we do not have a good control for its changes. Therefore, this derivative needs to be removed form the estimates, which explains the introduction of the remainder A given in (3.11).

We keep the slightly cumbersome notation $f_{\Lambda,z}$ to emphasize that the Stein solution is random. Note that

$$Z_\lambda f_{\Lambda,z}(Z_\lambda) = \sum_{n=1}^{N_\lambda} Y_{\lambda,n} f_{\Lambda,z}(Z_\lambda).$$

For every $n \in \mathbb{N}$, by the fundamental theorem of calculus,

$$1_{\{N_\lambda \geq n\}} Y_{\lambda,n} f_{\Lambda,z}(Z_\lambda) = 1_{\{N_\lambda \geq n\}} Y_{\lambda,n} \int_0^{Y_{\lambda,n}} f'_{\Lambda,z}(Z_{\lambda,n}+t)\,\mathrm{d}t + 1_{\{N_\lambda \geq n\}} Y_{\lambda,n} f_{\Lambda,z}(Z_{\lambda,n}), \quad (3.13)$$

where we can rewrite the integral as

$$Y_{\lambda,n} \int_0^{Y_{\lambda,n}} f'_{\Lambda,z}(Z_{\lambda,n}+t)\,\mathrm{d}t = \int_{\mathbb{R}} f'_{\Lambda,z}(Z_{\lambda,n}+t) Y_{\lambda,n} (1_{\{0\leq t\leq Y_{\lambda,n}\}} - 1_{\{Y_{\lambda,n}\leq t\leq 0\}})\,\mathrm{d}t.$$

Note that we use the notion of conditional expectation for random variables, which are σ-integrable with respect to a σ-algebra (see [HWY92, Chapter 4]). This is useful when considering conditional expectations involving $f_{\Lambda,z}$, because by Lemma A.1 the function is bounded on events where Λ is bounded.

Note that since $Y_{\lambda,n}$ and $(\Lambda, N_\lambda, Z_{\lambda,n})$ are independent and $\mathbb{E}[Y_{\lambda,n}] = 0$,

$$\mathbb{E}\big[1_{\{N_\lambda \geq n\}} Y_{\lambda,n} f_{\Lambda,z}(Z_{\lambda,n}) \,\big|\, \Lambda, N_\lambda\big] = 0, \qquad n \in \mathbb{N},$$

for the last term of (3.13). Similarly, for every $\sigma(\Lambda)$-measurable event E where Λ is bounded, $n \in \mathbb{N}$, $t \in \mathbb{R}$,

$$\mathbb{E}\big[1_E\, f'_{\Lambda,z}(Z_{\lambda,n}+t) Y_{\lambda,n} (1_{\{0\leq t\leq Y_{\lambda,n}\}} - 1_{\{Y_{\lambda,n}\leq t\leq 0\}})\big] = \mathbb{E}\big[1_E\, f'_{\Lambda,z}(Z_{\lambda,n}+t)\big]\, K_{\lambda,n}(t).$$

By the Stein equation (3.6) with $x = Z_{\lambda,n} + t$ and $\sigma^2 = \Lambda$,

$$h_z(Z_{\lambda,n}+t) - g_z(\Lambda) = f'_{\Lambda,z}(Z_{\lambda,n}+t) - \frac{Z_{\lambda,n}+t}{\Lambda} f_{\Lambda,z}(Z_{\lambda,n}+t).$$

Substituting this into (3.12) and combining it with the above results shows that

$$B \overset{\text{a.s.}}{=} \frac{1}{\Lambda}\,\mathbb{E}\big[Z_\lambda f_{\Lambda,z}(Z_\lambda)\,\big|\,\Lambda\big] - \frac{1}{\Lambda^2}\,\mathbb{E}\Big[\sum_{n=1}^{N_\lambda} \int_{\mathbb{R}} (Z_{\lambda,n}+t) f_{\Lambda,z}(Z_{\lambda,n}+t) K_{\lambda,n}(t)\,\mathrm{d}t\,\Big|\,\Lambda\Big]. \quad (3.14)$$

By size biasing and (3.10),

$$\begin{aligned}\mathbb{E}\big[Z_\lambda f_{\Lambda,z}(Z_\lambda)\,\big|\,\Lambda\big] &\overset{\text{a.s.}}{=} \frac{1}{\lambda\Lambda}\,\mathbb{E}\big[N_\lambda Z'_\lambda f_{\Lambda,z}(Z'_\lambda)\,\big|\,\Lambda\big] \\ &\overset{\text{a.s.}}{=} \frac{1}{\Lambda}\,\mathbb{E}\Big[\sum_{n=1}^{N_\lambda} \int_{\mathbb{R}} Z'_\lambda f_{\Lambda,z}(Z'_\lambda) K_{\lambda,n}(t)\,\mathrm{d}t\,\Big|\,\Lambda\Big].\end{aligned}$$

Substitution into (3.14) and rearrangement leads to

$$B \overset{\text{a.s.}}{=} \frac{1}{\Lambda^2}\,\mathbb{E}\Big[\sum_{n=1}^{N_\lambda} \int_{\mathbb{R}} \big(Z'_\lambda f_{\Lambda,z}(Z'_\lambda) - (Z_{\lambda,n}+t) f_{\Lambda,z}(Z_{\lambda,n}+t)\big) K_{\lambda,n}(t)\,\mathrm{d}t\,\Big|\,\Lambda\Big].$$

We have the representations

$$Z'_\lambda = Z'_{\lambda,n} + Y_{\lambda,n} 1_{\{N_\lambda \ge n+1\}} \quad \text{and} \quad Z_{\lambda,n} + t = Z'_{\lambda,n} + t + Y_{\lambda,N_\lambda} 1_{\{N_\lambda \ne n\}}.$$

Hence, by the last item of Lemma A.1 we get

$$\bigl|Z'_\lambda f_{\Lambda,z}(Z'_\lambda) - (Z_{\lambda,n}+t) f_{\Lambda,z}(Z_{\lambda,n}+t)\bigr| \le \Bigl(|Z'_{\lambda,n}| + \frac{\sqrt{\pi\Lambda}}{2\sqrt{2}}\Bigr)\bigl(|t| + |Y_{\lambda,n}| 1_{\{N_\lambda \ge n+1\}} + |Y_{\lambda,N_\lambda}| 1_{\{N_\lambda \ne n\}}\bigr).$$

Since $Y_{\lambda,n}$ and $(\Lambda, N_\lambda, Z'_{\lambda,n})$ are independent and since Y_{λ,N_λ} and $(\Lambda, N_\lambda, Z'_{\lambda,n})$ are conditionally independent given N_λ,

$$|B| \le \frac{1}{\Lambda^2}\,\mathbb{E}\Bigl[\sum_{n=1}^{N_\lambda}\Bigl(|Z'_{\lambda,n}| + \frac{\sqrt{\pi\Lambda}}{2\sqrt{2}}\Bigr) C_n \Bigm| \Lambda\Bigr] \quad \text{a.s.}$$

with

$$C_n := \int_{\mathbb{R}} \Bigl(|t| 1_{\{N_\lambda \ge n\}} + \mathbb{E}[|Y_{\lambda,n}|]\, 1_{\{N_\lambda \ge n+1\}} + \sum_{l=n+1}^{\infty} \mathbb{E}[|Y_{\lambda,l}|]\, 1_{\{N_\lambda = l\}}\Bigr) K_{\lambda,n}(t)\,\mathrm{d}t$$

for each $n \in \mathbb{N}$. Using (3.10) and $\mathbb{E}[|Y_{\lambda,l}|] \le \sqrt{\mathbb{E}[Y_{\lambda,l}^2]} = 1/\sqrt{\lambda}$ for all $l \in \mathbb{N}$,

$$C_n \le \frac{1}{2}\,\mathbb{E}\bigl[|Y_{\lambda,n}|^3\bigr]\, 1_{\{N_\lambda \ge n\}} + \frac{2}{\sqrt{\lambda^3}} 1_{\{N_\lambda \ge n+1\}}.$$

By (3.9) and (2.2),

$$\begin{aligned}
|B| &\le \frac{1}{2\Lambda^2\sqrt{\lambda^3}}\,\mathbb{E}\Bigl[\sum_{n=1}^{N_\lambda} |Z'_{\lambda,n}|\,\mathbb{E}\bigl[|X_n|^3\bigr] \Bigm| \Lambda\Bigr] + \frac{\sqrt{\pi}}{4\sqrt{2\lambda^3\Lambda^3}}\,\mathbb{E}\Bigl[\sum_{n=1}^{N_\lambda} \mathbb{E}\bigl[|X_n|^3\bigr] \Bigm| \Lambda\Bigr] \\
&\quad + \frac{2}{\Lambda^2\sqrt{\lambda^3}}\,\mathbb{E}\Bigl[\sum_{n=1}^{N_\lambda - 1} |Z'_{\lambda,n}| \Bigm| \Lambda\Bigr] + \frac{\sqrt{\pi}}{\sqrt{2\lambda^3\Lambda^3}}\,\mathbb{E}\bigl[N_\lambda \bigm| \Lambda\bigr] \\
&= \frac{1}{2\Lambda^2\sqrt{\lambda^3}}\,\mathbb{E}\Bigl[\sum_{n=1}^{N_\lambda} \mathbb{E}\bigl[|Z'_{\lambda,n}| \bigm| \Lambda, N_\lambda\bigr]\,\mathbb{E}\bigl[|X_n|^3\bigr] \Bigm| \Lambda\Bigr] + \frac{\sqrt{\pi}}{4\sqrt{2\lambda\Lambda}}\varrho(\lambda\Lambda) \\
&\quad + \frac{2}{\Lambda^2\sqrt{\lambda^3}}\,\mathbb{E}\Bigl[\sum_{n=1}^{N_\lambda - 1} \mathbb{E}\bigl[|Z'_{\lambda,n}| \bigm| \Lambda, N_\lambda\bigr] \Bigm| \Lambda\Bigr] + \frac{\sqrt{\pi}}{\sqrt{2\lambda\Lambda}} \quad \text{a.s.}
\end{aligned}$$

By the conditional Jensen inequality, for each $n \in \mathbb{N}$,

$$\mathbb{E}\bigl[|Z'_{\lambda,n}| \bigm| \Lambda, N_\lambda\bigr] \overset{\text{a.s.}}{=} \mathbb{E}\bigl[|Z'_{\lambda,n}| \bigm| N_\lambda\bigr] \overset{\text{a.s.}}{=} \sqrt{\mathbb{E}\bigl[(Z'_{\lambda,n})^2 \bigm| N_\lambda\bigr]}$$

and

$$\begin{aligned}
\mathbb{E}\bigl[(Z'_{\lambda,n})^2 \bigm| N_\lambda\bigr] &\overset{\text{a.s.}}{=} \mathrm{V}\bigl(Z'_{\lambda,n} \bigm| N_\lambda\bigr) \overset{\text{a.s.}}{=} \frac{1}{\lambda}\sum_{i=1, i\ne n}^{N_\lambda - 1} \mathrm{V}(X_i) \\
&= \frac{N_\lambda - 1_{\{N_\lambda \ge 1\}} - 1_{\{N_\lambda \ge n+1\}}}{\lambda} \le \frac{N_\lambda}{\lambda},
\end{aligned}$$

hence

$$|B| \leq \frac{\varrho(\lambda\Lambda)}{2\lambda\Lambda}\,\mathbb{E}\Big[\frac{\sqrt{N_\lambda}}{\lambda\Lambda\varrho(\lambda\Lambda)}\sum_{n=1}^{N_\lambda}\mathbb{E}\big[|X_n|^3\big]\,\Big|\,\Lambda\Big] + \sqrt{\pi}\frac{4+\varrho(\lambda\Lambda)}{4\sqrt{2\lambda\Lambda}} \\ + \frac{2}{\Lambda^2\lambda^2}\,\mathbb{E}\big[N_\lambda\sqrt{N_\lambda - 1_{\{N_\lambda\geq 1\}}}\,|\,\Lambda\big] \quad \text{a.s.} \tag{3.15}$$

Note that

$$\mathbb{E}\Big[\frac{1}{\lambda\Lambda\varrho(\lambda\Lambda)}\sum_{n=1}^{N_\lambda}\mathbb{E}\big[|X_n|^3\big]\,\Big|\,\Lambda\Big] \overset{\text{a.s.}}{=} 1$$

by (2.2), hence by the conditional Jensen inequality

$$\mathbb{E}\Big[\frac{\sqrt{N_\lambda}}{\lambda\Lambda\varrho(\lambda\Lambda)}\sum_{n=1}^{N_\lambda}\mathbb{E}\big[|X_n|^3\big]\,\Big|\,\Lambda\Big] \leq \Big(\mathbb{E}\Big[\frac{N_\lambda}{\lambda\Lambda\varrho(\lambda\Lambda)}\sum_{n=1}^{N_\lambda}\mathbb{E}\big[|X_n|^3\big]\,\Big|\,\Lambda\Big]\Big)^{\frac{1}{2}} \quad \text{a.s.}$$

Due to size-biasing and (2.2),

$$\mathbb{E}\Big[\frac{N_\lambda}{\lambda\Lambda\varrho(\lambda\Lambda)}\sum_{n=1}^{N_\lambda}\mathbb{E}\big[|X_n|^3\big]\,\Big|\,\Lambda\Big] \overset{\text{a.s.}}{=} \mathbb{E}\Big[\frac{1}{\varrho(\lambda\Lambda)}\sum_{n=1}^{N_\lambda+1}\mathbb{E}\big[|X_n|^3\big]\,\Big|\,\Lambda\Big] \overset{\text{a.s.}}{=} \frac{\lambda\Lambda\psi(\lambda\Lambda)}{\varrho(\lambda\Lambda)}$$

and

$$\mathbb{E}\Big[\frac{N_\lambda}{\lambda\Lambda}\sqrt{N_\lambda - 1_{\{N_\lambda\geq 1\}}}\,\Big|\,\Lambda\Big] = \mathbb{E}\big[\sqrt{N_\lambda}\,|\,\Lambda\big] \leq \sqrt{\mathbb{E}[N_\lambda\,|\,\Lambda]} = \sqrt{\lambda\Lambda} \quad \text{a.s.}$$

Since $\varrho(\lambda\Lambda) \leq \psi(\lambda\Lambda)$,

$$\begin{aligned} |B| &\leq \frac{1}{2\sqrt{\lambda\Lambda}}\Big(\sqrt{\varrho(\lambda\Lambda)\psi(\lambda\Lambda)} + \sqrt{\pi}\frac{4+\varrho(\lambda\Lambda)}{2\sqrt{2}} + 4\Big) \quad \text{a.s.} \\ &\leq \frac{1}{2\sqrt{\lambda\Lambda}}\Big(\psi(\lambda\Lambda) + \sqrt{\pi}\frac{4+\varrho(\lambda\Lambda)}{2\sqrt{2}} + 4\Big) \quad \text{a.s.} \end{aligned}$$

We now study (3.11). To this end, for $n \in \mathbb{N}$, $k \in \{1,\ldots,n-1\}$, $t \in \mathbb{R}$,

$$E_{k,n}(t) := \Big\{\sqrt{\lambda}z - \max(X_k, X_n + \sqrt{\lambda}t) < \sum_{\substack{j=1\\ j\neq k}}^{n-1} X_j \leq \sqrt{\lambda}z - \min(X_k, X_n + \sqrt{\lambda}t)\Big\},$$

and in case $k = n$,

$$E_{n,n}(t) := \Big\{\sqrt{\lambda}z - \max(0, \sqrt{\lambda}t) < \sum_{j=1}^{n-1} X_j \leq \sqrt{\lambda}z - \min(0, \sqrt{\lambda}t)\Big\}.$$

Then

$$1_{E_{k,N_\lambda}(t)} = \big|1_{\{Z'_\lambda\leq z\}} - 1_{\{Z_{\lambda,k}+t\leq z\}}\big|, \tag{3.16}$$

$$\mathbb{P}[E_{k,N_\lambda}(t)\,|\,N_\lambda = n] = \mathbb{P}[E_{k,n}(t)]\,, \tag{3.17}$$

and

$$\mathbb{P}[E_{k,N_\lambda}(t)\,|\,N_\lambda, \Lambda] = \mathbb{P}[E_{k,N_\lambda}(t)\,|\,N_\lambda]\,, \tag{3.18}$$

where the last equation can be shown by a monotone class argument. By (3.16), the tower property of the conditional expectation, the conditional Fubini Theorem, (3.18), and (3.17),

$$
\begin{aligned}
|A| &\leq \frac{1}{\Lambda}\,\mathbb{E}\Big[\sum_{k=1}^{N_\lambda}\int_{\mathbb{R}}\big|1_{\{Z'_\lambda\leq z\}}-1_{\{Z_{\lambda,k}+t\leq z\}}\big|K_{\lambda,k}(t)\,\mathrm{d}t\,\Big|\,\Lambda\Big]\\
&=\frac{1}{\Lambda}\,\mathbb{E}\Big[\sum_{k=1}^{N_\lambda}\int_{\mathbb{R}}1_{E_{k,N_\lambda}(t)}K_{\lambda,k}(t)\,\mathrm{d}t\,\Big|\,\Lambda\Big]\\
&=\frac{1}{\Lambda}\,\mathbb{E}\Big[\mathbb{E}\Big[\sum_{k=1}^{N_\lambda}\int_{\mathbb{R}}1_{E_{k,N_\lambda}(t)}K_{\lambda,k}(t)\,\mathrm{d}t\,\Big|\,N_\lambda,\Lambda\Big]\,\Big|\,\Lambda\Big] \qquad (3.19)\\
&=\frac{1}{\Lambda}\,\mathbb{E}\Big[\sum_{k=1}^{N_\lambda}\int_{\mathbb{R}}\mathbb{P}[\,E_{k,N_\lambda}(t)\,|\,N_\lambda]\,K_{\lambda,k}(t)\,\mathrm{d}t\,\Big|\,\Lambda\Big]\\
&=\frac{1}{\Lambda}\,\mathbb{E}\Big[\sum_{k=1}^{N_\lambda}\int_{\mathbb{R}}\mathbb{E}[\mathbb{P}[E_{k,N_\lambda}(t)\,|\,X_k,X_{N_\lambda}]\,|\,N_\lambda]\,K_{\lambda,k}(t)\,\mathrm{d}t\,\Big|\,\Lambda\Big].
\end{aligned}
$$

We now analyze

$$\mathbb{E}[\mathbb{P}[E_{k,n}(t)\,|\,X_k,X_n]]\,. \qquad (3.20)$$

For $n\geq 3$ and $k\in\{1,\dots,n-1\}$,

$$(n-2)^{-1}\sum_{\substack{j=1\\j\neq k}}^{n-1}\mathbb{E}\big[X_j^2\big]=1,$$

and we apply the concentration inequality (A.3) to (3.20) for

$$(n-2)^{-\frac{1}{2}}\sum_{\substack{j=1\\j\neq k}}^{n-1}X_j,$$

with

$$\bar{\gamma}_n:=(n-2)^{-\frac{3}{2}}\sum_{\substack{j=1\\j\neq k}}^{n-1}\mathbb{E}\big[|X_j|^3\big]\,.$$

For $n\geq 3$ and $k=n$

$$(n-1)^{-\frac{1}{2}}\sum_{j=1}^{n-1}\mathbb{E}\big[X_j^2\big]=1,$$

and we apply the concentration inequality (A.3) to (3.20) for

$$(n-1)^{-\frac{1}{2}}\sum_{j=1}^{n-1}X_j,$$

with

$$\tilde{\gamma}_n:=(n-1)^{-\frac{3}{2}}\sum_{j=1}^{n-1}\mathbb{E}\big[|X_j|^3\big]\,.$$

Hence, for $n \geq 3$, $k \in \{1, \dots, n\}$,

$$\mathbb{E}[\mathbb{P}[E_{k,n}(t) \,|\, X_k, X_n]] \leq (n-2)^{-\frac{1}{2}} \, \mathbb{E}\big[\sqrt{\lambda}|t| + |X_k - X_n|\big] + 2\gamma_n, \tag{3.21}$$

with

$$\gamma_n := (n-2)^{-\frac{3}{2}} \sum_{j=1}^{n-1} \mathbb{E}\big[|X_j|^3\big] .$$

Plugging (3.21) into (3.19), applying (3.10) and Jensen's inequality we get a.s.

$$\begin{aligned}
|A| \leq & \frac{1}{\Lambda} \, \mathbb{E}\Big[1_{\{N_\lambda=1\}} \int_{\mathbb{R}} K_{\lambda,1}(t) \, \mathrm{d}t \Big| \Lambda\Big] + \frac{1}{\Lambda} \, \mathbb{E}\Big[1_{\{N_\lambda=2\}} \int_{\mathbb{R}} (K_{\lambda,1}(t) + K_{\lambda,2}(t)) \, \mathrm{d}t \Big| \Lambda\Big] \\
& + \frac{1}{\Lambda} \, \mathbb{E}\Big[1_{\{N_\lambda \geq 3\}} (N_\lambda - 2)^{-\frac{1}{2}} \sum_{k=1}^{N_\lambda} \int_{\mathbb{R}} K_{\lambda,k}(t) \Big(\sqrt{\lambda}|t| + \sqrt{2}\Big) \, \mathrm{d}t \Big| \Lambda\Big] \\
& + \frac{2}{\Lambda} \, \mathbb{E}\Big[1_{\{N_\lambda \geq 3\}} \gamma_{N_\lambda} \sum_{k=1}^{N_\lambda} \int_{\mathbb{R}} K_{\lambda,k}(t) \, \mathrm{d}t \Big| \Lambda\Big] \\
\leq & \frac{1}{\lambda\Lambda} \, \mathbb{E}\big[1_{\{N_\lambda=1\}} \,|\, \Lambda\big] + \frac{2}{\lambda\Lambda} \, \mathbb{E}\big[1_{\{N_\lambda=2\}} \,|\, \Lambda\big] \\
& + \frac{\sqrt{3}}{2\lambda\Lambda} \, \mathbb{E}\Big[1_{\{N_\lambda \geq 3\}} \frac{1}{\sqrt{N_\lambda}} \sum_{k=1}^{N_\lambda} \mathbb{E}\big[|X_k|^3\big] \Big| \Lambda\Big] + \frac{\sqrt{6}}{\lambda\Lambda} \, \mathbb{E}\big[\sqrt{N_\lambda} \,|\, \Lambda\big] \\
& + \frac{2}{\lambda\Lambda} \, \mathbb{E}\big[1_{\{N_\lambda \geq 3\}} \gamma_{N_\lambda} N_\lambda \,|\, \Lambda\big] \\
\leq & \frac{13\sqrt{3}}{2\lambda\Lambda} \, \mathbb{E}\Big[1_{\{N_\lambda \geq 1\}} \frac{1}{\sqrt{N_\lambda}} \sum_{k=1}^{N_\lambda} \mathbb{E}\big[|X_k|^3\big] \Big| \Lambda\Big] + \frac{\sqrt{6}}{\sqrt{\lambda\Lambda}}.
\end{aligned}$$

By size biasing,

$$\begin{aligned}
\mathbb{E}\Big[1_{\{N_\lambda \geq 1\}} \frac{1}{\sqrt{N_\lambda}} \sum_{k=1}^{N_\lambda} \mathbb{E}\big[|X_k|^3\big] \Big| \Lambda\Big] &= \mathbb{E}\Big[1_{\{N_\lambda \geq 2\}} \frac{N_\lambda}{\lambda\Lambda\sqrt{N_\lambda - 1}} \sum_{k=1}^{N_\lambda - 1} \mathbb{E}\big[|X_k|^3\big] \Big| \Lambda\Big] \\
&\leq \sqrt{2} \, \mathbb{E}\Big[\frac{\sqrt{N_\lambda}}{\lambda\Lambda} \sum_{k=1}^{N_\lambda} \mathbb{E}\big[|X_k|^3\big] \Big| \Lambda\Big],
\end{aligned}$$

which can be analyzed like the first term on the right hand side in (3.15), thus

$$|A| \leq \frac{13\sqrt{6}}{2\sqrt{\lambda\Lambda}} \psi(\lambda\Lambda) + \frac{\sqrt{6}}{\sqrt{\lambda\Lambda}}.$$

Putting everything together results in

$$|A + B| \leq \frac{1}{\sqrt{\lambda\Lambda}} \Big(\frac{4 + 52\sqrt{6} + \sqrt{2\pi}}{8}\Big) \psi(\lambda\Lambda) + 2 + \sqrt{6} + \sqrt{\frac{\pi}{2}}.$$

□

Remark 3.22. This chapter is based on so far unpublished research by P. Eichelsbacher, P. Porkert, and U. Schmock.

Part III

Small-Time Central Limit Theorems for Semimartingales

Chapter 1

The Setting

We now lay down the setting for the following chapters.

Assumption 1.1. *Let $m \in \mathbb{N}$, $T > 0$, $x_0 \in \mathbb{R}^m$. Let $X = (X_t^1, \dots, X_t^m)^\top_{t\in[0,T]}$ be an $\mathbb{R}^m$-valued continuous semimartingale with canonical decomposition*[1]

$$X - x_0 = M + A,$$

where M is a continuous local martingale, A has locally finite variation, and $M_0 = A_0 = 0$. Assume that

1. $X_0 = x_0$ *a.s.,*

2. *there exists an a.s. positive stopping time τ_A such that a.s.*

$$A_t^j = \int_0^t b_s^j \, \mathrm{d}s, \quad t \in [0, \tau_A], \quad j \in \{1, \dots, m\},$$

   for an adapted process b,

3. *there exists a random variable C_b, such that $|b_t^j| \leq C_b < \infty$ for a.e. $t \in [0, \tau_A]$ a.s., $j \in \{1, \dots, m\}$,*

4. *there exists an a.s. positive stopping time τ_M such that the covariation is a.s.*

$$\langle M^j, M^k \rangle_t = \int_0^t \sum_{l=1}^m \sigma_s^{jl} \sigma_s^{kl} \, \mathrm{d}s, \quad t \in [0, \tau_M], \quad j, k \in \{1, \dots, m\},$$

   for a progressive process σ,

5. *there exists a deterministic constant $C_\sigma < \infty$, such that $|\sigma_t^{jk}| \leq C_\sigma$ for a.e. $t \in [0, \tau_M]$ a.s., $j, k \in \{1, \dots, m\}$, and*

6. *as $t \searrow 0$, $\sigma_t \to L$ a.s., where L is a deterministic $m \times m$-matrix.*

---

[1] See e.g. [Kal02, p. 337].

Remark 1.2. Let X be a weak solution of the m-dimensional SDE, $j \in \{0, \dots, m\}$, $d \in \mathbb{N}$,

$$X_t^j = x_0 + \int_0^t b_j(s, X_s)\,\mathrm{d}s + \sum_{k=1}^d \int_0^t \sigma_{jk}(s, X_s)\,\mathrm{d}B_s^k, \quad t \geq 0,$$

where B is a standard d-dimensional Brownian motion, $x_0 \in \mathbb{R}^m$,

$$b : [0, T] \times \mathbb{R}^m \to \mathbb{R}^m,$$

is uniformly bounded in a neighborhood of $(0, x_0)$, and

$$\sigma : [0, T] \times \mathbb{R}^m \to \mathbb{R}^{m \times d},$$

is continuous in $(0, x_0)$. Then X satisfies Assumption 1.1.

Assumption 1.3. *For $m \in \mathbb{N}$, $T > 0$, let $X = (X_t^1, \dots, X_t^m)_{t \in [0,T]}^\top$ be an $\mathbb{R}^m$-valued càdlàg semimartingale with decomposition $X = X^c + J$, such that*

1. *X^c is a continuous semimartingale satisfying Assumption 1.1,*
2. *the process J is given by*

$$J_t = \int_0^t \int_{\boldsymbol{B}_1} \psi(s, z)\,\big(\Pi(\mathrm{d}s, \mathrm{d}z) - \mu(\mathrm{d}s, \mathrm{d}z)\big) + \int_0^t \int_{\mathbb{R}^m \setminus \boldsymbol{B}_1} \varphi(s, z)\,\Pi(\mathrm{d}s, \mathrm{d}z),$$

where $\boldsymbol{B}_1$ denotes the unit ball in $\mathbb{R}^m$, Π is a Poisson random measure on $[0, T] \times \mathbb{R}^m$ with compensator μ; the $\mathbb{R}^m$-valued processes ψ, φ are predictable with respect to the filtration generated by Π, and

$$\mathbb{E}\Big[\int_0^T \int_{\boldsymbol{B}_1} |\psi(s, z)|^2\,\mu(\mathrm{d}s, \mathrm{d}z)\Big] < \infty,$$

3. *there exists an a.s. positive stopping time τ_J, such that*

$$\mathbb{E}\Big[\big|\Pi - \mu\big|\big([0, t \wedge \tau_J] \times \boldsymbol{B}_1\big)\Big] = \mathrm{O}\left(\sqrt{t}\right) \quad \text{as} \quad t \searrow 0. \tag{1.4}$$

Remark 1.5. Note that compound Poisson processes are covered by these assumptions. In this case the left-hand side of (1.4) is $\mathrm{O}(t)$ as $t \searrow 0$.

Chapter 2

Central Limit Theorems for Continuous Semimartingales

We prove a small-time CLT which we support with a theorem on higher order asymptotics, as well as a functional CLT for a class of continuous semimartingales.

Theorem 2.1 (CLT)**.** *Let X satisfy Assumption 1.1. Then for every $f : \mathbb{R}^m \to \mathbb{R}^n$ such that there exists an open neighborhood U of x_0 with $f \in \mathrm{C}^2(U;\mathbb{R}^n)$, we have*

$$\frac{1}{\sqrt{t}}(f(X_t) - f(x_0)) \xrightarrow{\mathrm{d}} N_f \text{ as } t \searrow 0,$$

where N_f is a normal random vector with mean zero and covariance matrix

$$V = (Df)(x_0)L(Df(x_0)L)^\top.$$

Here $(Df)(x_0)$ stands for the Jacobian of f at x_0.

Proof. We first suppose that $m = n$ and $f = \mathrm{id}$; then the general case follows by the delta method, see the end of the proof. Let N_{id} be an $\mathrm{N}(0, LL^\top)$ random vector on some probability space $(\widetilde{\Omega}, \widetilde{\mathscr{A}}, \widetilde{\mathbb{P}})$. We need to show

$$\lim_{t\searrow 0} \mathbb{E}\Big[g\Big(\frac{X_t - x_0}{\sqrt{t}}\Big)\Big] = \mathbb{E}_{\tilde{\mathbb{P}}}\big[g\big(N_{\mathrm{id}}\big)\big], \quad g \in \mathrm{C_b}(\mathbb{R}^n;\mathbb{R}). \tag{2.2}$$

To this end, we fix a function $g \in \mathrm{C_b}(\mathbb{R}^n)$. Then with

$$\tau := \tau_A \wedge \tau_M,$$

we have

$$\begin{aligned}\Big|\mathbb{E}\Big[g\Big(\frac{X_t - x_0}{\sqrt{t}}\Big)\Big] - \mathbb{E}_{\tilde{\mathbb{P}}}[g(N_{\mathrm{id}})]\Big| &\le \Big|\mathbb{E}\Big[g\Big(\frac{X_t - x_0}{\sqrt{t}}\Big) - g\Big(\frac{X_{t\wedge\tau} - x_0}{\sqrt{t}}\Big)\Big]\Big| \\ &\quad + \Big|\mathbb{E}\Big[g\Big(\frac{X_{t\wedge\tau} - x_0}{\sqrt{t}}\Big)\Big] - \mathbb{E}_{\tilde{\mathbb{P}}}[g(N_{\mathrm{id}})]\Big|. \end{aligned} \tag{2.3}$$

Hence, in order to show (2.2) it is sufficient to prove that the two summands in the latter upper bound tend to zero as $t \searrow 0$. Since the event $\{\tau = 0\}$ has probability zero, the first summand converges to zero by the dominated convergence theorem. Moreover, the convergence of the second summand to zero will follow, if we can show

$$\frac{X_{t\wedge\tau} - x_0}{\sqrt{t}} \xrightarrow{\mathrm{d}} N_{\mathrm{id}}, \qquad t \searrow 0. \tag{2.4}$$

In order to prove (2.4), we first note that Doob's Integral Representation Theorem, see e.g. [Kal02, Theorem 18.12, p. 358], in combination with part (*4*) of Assumption 1.1 implies the existence of an m-dimensional Brownian motion B (possibly on an extension of the primary probability space) such that a.s.

$$M^j_{t\wedge\tau} = \sum_{k=1}^m \int_0^{t\wedge\tau} \sigma^{jk}_s \,\mathrm{d}B^k_s, \quad t \in [0,T], \quad j \in \{1,\dots,m\}. \tag{2.5}$$

By part (*2*) of Assumption 1.1 and (2.5) we therefore have a.s.

$$X^j_{t\wedge\tau} = x_0 + \int_0^{t\wedge\tau} b^j_s \,\mathrm{d}s + \sum_{k=1}^m \int_0^{t\wedge\tau} \sigma^{jk}_s \,\mathrm{d}B^k_s, \quad t \in [0,T],\ j \in \{1\dots,m\}. \tag{2.6}$$

In addition, we recall that by the Cramér–Wold Theorem, (2.4) holds iff for every $s = (s_1,\dots,s_m)^\top \in \mathbb{R}^m$

$$\sum_{j=1}^m s_j \,\frac{X^j_{t\wedge\tau} - x^j_0}{\sqrt{t}} \xrightarrow{\mathrm{d}} \sum_{j=1}^m s_j\, N^j_{\mathrm{id}} \tag{2.7}$$

as $t \searrow 0$. To show this, we fix $s = (s_1,\dots,s_m)^\top \in \mathbb{R}^m$. By (2.6), the left-hand side of (2.7) equals

$$\frac{1}{\sqrt{t}} \sum_{j=1}^m s_j \int_0^{t\wedge\tau} b^j_s \,\mathrm{d}s + \frac{1}{\sqrt{t}} \sum_{j=1}^m s_j \sum_{k=1}^m \int_0^{t\wedge\tau} \sigma^{jk}_s \,\mathrm{d}B^k_s. \tag{2.8}$$

By part (*3*) of Assumption 1.1, the terms $t^{-1/2}\int_0^{t\wedge\tau} b^j_s \,\mathrm{d}s$ converge to zero a.s. Before examining the Itō integrals in (2.8) we observe that for every $t \in [0,T]$ the random vector

$$N_t := (N^1_t,\dots,N^m_t) := \Big(\frac{1}{\sqrt{t}} \sum_{k=1}^m L_{jk} B^k_t\Big)_{1\le j\le m}$$

is $\mathrm{N}(0, LL^\top)$ distributed. In particular, the distribution of N_t is independent of t and $N_t \overset{\mathrm{d}}{=} N_{\mathrm{id}}$ for every $t > 0$. We then have for all $h \in \mathrm{C_b}(\mathbb{R})$:

$$\begin{aligned}
&\Big|\mathbb{E}\Big[h\Big(\frac{1}{\sqrt{t}}\sum_{j=1}^n s_j \sum_{k=1}^m L_{jk} B^k_{t\wedge\tau}\Big)\Big] - \mathbb{E}_{\tilde{\mathbb{P}}}\Big[h\Big(\sum_{j=1}^n s_j N^j_{\mathrm{id}}\Big)\Big]\Big| \\
&\quad = \Big|\mathbb{E}\Big[h\Big(\frac{1}{\sqrt{t}}\sum_{j=1}^n s_j \sum_{k=1}^m L_{jk} B^k_{t\wedge\tau}\Big)\Big] - \mathbb{E}\Big[h\Big(\frac{1}{\sqrt{t}}\sum_{j=1}^n s_j \sum_{k=1}^m L_{jk} B^k_t\Big)\Big]\Big| \\
&\quad \le \Big|\mathbb{E}\Big[\Big(h\Big(\frac{1}{\sqrt{t}}\sum_{j=1}^n s_j \sum_{k=1}^m L_{jk} B^k_{t\wedge\tau}\Big) - h\Big(\frac{1}{\sqrt{t}}\sum_{j=1}^n s_j \sum_{k=1}^m L_{jk} B^k_t\Big)\Big)\mathbf{1}_{\{\tau>t\}}\Big]\Big| \\
&\qquad + \Big|\mathbb{E}\Big[\Big(h\Big(\frac{1}{\sqrt{t}}\sum_{j=1}^n s_j \sum_{k=1}^m L_{jk} B^k_{t\wedge\tau}\Big) - h\Big(\frac{1}{\sqrt{t}}\sum_{j=1}^n s_j \sum_{k=1}^m L_{jk} B^k_t\Big)\Big)\mathbf{1}_{\{\tau\le t\}}\Big]\Big| \\
&\quad \le 2\|h\|_\infty \,\mathbb{P}[\tau \le t] \to 0
\end{aligned}$$

as $t \searrow 0$. Therefore, the random variables

$$\frac{1}{\sqrt{t}} \sum_{j=1}^{m} s_j \sum_{k=1}^{m} \int_0^{t\wedge\tau} L_{jk} \,\mathrm{d}B_s^k = \frac{1}{\sqrt{t}} \sum_{j=1}^{m} s_j \sum_{k=1}^{m} L_{jk} \, B_{t\wedge\tau}^k \tag{2.9}$$

converge in distribution to $\sum_{j=1}^m s_j N_{\mathrm{id}}^j$ as $t \searrow 0$. Next, we show that the difference between (2.9) and the second term in (2.8) converges to zero in L^2, i.e. the limit law can be recovered by freezing the integrand at zero. By Itō's isometry and the Cauchy–Schwarz inequality we have

$$\begin{aligned}
&\mathbb{E}\Big[\Big(\sum_{j=1}^{m} s_j \sum_{k=1}^{m} \frac{1}{\sqrt{t}} \int_0^{t\wedge\tau} (\sigma_s^{jk} - L_{jk}) \,\mathrm{d}B_s^k\Big)^2\Big] \\
&= \mathbb{E}\Big[\sum_{k=1}^{m} \frac{1}{t}\Big(\sum_{j=1}^{m} s_j \int_0^{t\wedge\tau} (\sigma_s^{jk} - L_{jk}) \,\mathrm{d}B_s^k\Big)^2\Big] \\
&\leq \Big(\sum_{j=1}^{m} s_j^2\Big) \sum_{k,j=1}^{m} \mathbb{E}\Big[\frac{1}{t}\Big(\int_0^{t\wedge\tau} (\sigma_s^{jk} - L_{jk}) \,\mathrm{d}B_s^k\Big)^2\Big] \\
&= \Big(\sum_{j=1}^{m} s_j^2\Big) \sum_{k,j=1}^{m} \mathbb{E}\Big[\frac{1}{t} \int_0^{t\wedge\tau} (\sigma_s^{jk} - L_{jk})^2 \,\mathrm{d}s\Big] \\
&\leq \Big(\sum_{j=1}^{m} s_j^2\Big) \sum_{k,j=1}^{m} \mathbb{E}\Big[\frac{t\wedge\tau}{t} \max_{s\in[0,t\wedge\tau]} (\sigma_s^{jk} - L_{jk})\Big],
\end{aligned}$$

which indeed converges to zero as $t \searrow 0$ by the dominated convergence theorem. The just established L^2 convergence implies convergence in distribution. By Slutsky's theorem, (2.7) readily follows.

Finally, consider an arbitrary f as in the statement of the theorem. Then the desired CLT essentially follows from a Taylor expansion; this procedure is well known in statistics as the *delta method* [Dav03]. We choose an open ball $\mathscr{B}$ such that $\overline{\mathscr{B}} \subset U$ and define the stopping time

$$\hat{\tau} := \tau_A \wedge \tau_M \wedge \tau_{\overline{\mathscr{B}}^{\mathrm{c}}}, \tag{2.10}$$

where $\tau_{\overline{\mathscr{B}}^{\mathrm{c}}}$ denotes the hitting time of $\overline{\mathscr{B}}^{\mathrm{c}}$ for X. Analogously to the case $f = \mathrm{id}$ (recall (2.3)), it suffices to show convergence in distribution of $t^{-1/2}(f(X_{t\wedge\hat{\tau}}) - f(x_0))$. The Taylor expansion yields

$$\begin{aligned}
\frac{1}{\sqrt{t}}(f(X_{t\wedge\hat{\tau}}) - f(x_0)) &= \frac{1}{\sqrt{t}} Df(x_0)(X_{t\wedge\hat{\tau}} - x_0) \\
&\quad + \frac{1}{2\sqrt{t}} \Big((X_{t\wedge\hat{\tau}} - x_0)^\top H_{f_j}(\xi_t)(X_{t\wedge\hat{\tau}} - x_0)\Big)_{1\leq j\leq n}^\top,
\end{aligned} \tag{2.11}$$

where H_{f_j} denotes the Hessian of f_j and ξ_t is a random vector with $\|\xi_t - x_0\| \leq \|X_{t\wedge\hat{\tau}} - x_0\|$. The first term on the right-hand side of (2.11) converges in distribution to a Gaussian random vector with mean zero and covariance matrix V, by the first part of the proof (viz. (2.4), with τ replaced by $\hat{\tau}$) and Slutsky's theorem. Now consider the second term on the right-hand side of (2.11). The vector $t^{-1/2}(X_{t\wedge\hat{\tau}} - x_0)$ converges in distribution to a normal random vector. By Slutsky's theorem, we are done if we can show that the vector $H_{f_j}(\xi_t)(X_{t\wedge\hat{\tau}} - x_0)$ converges to zero in L^2. But, since

$$\|H_{f_j}(\xi_t)(X_{t\wedge\hat{\tau}} - x_0)\|_2 \leq \|H_{f_j}(\xi_t)\|_{\mathrm{F}} \|X_{t\wedge\hat{\tau}} - x_0\|_2,$$

where $\|\cdot\|_{\mathrm{F}}$ denotes the Frobenius norm, this easily follows from Assumption 1.1 and (2.6). $\square$

If a semimartingale X satisfies Assumption 1.1 and the limit law in Theorem 2.1 is non-singular, we clearly have

$$\lim_{t\searrow 0} \mathbb{P}[X_t > x_0] = 1/2. \tag{2.12}$$

We now give examples where the value of this limit is not $1/2$.

Example 2.13. Let us consider the squared Brownian motion B^2 in one dimension (no confusion with our superindex convention should arise). Then clearly $\lim_{t\searrow 0} \mathbb{P}\big[B_t^2 > 0\big] = 1$, which does not contradict Theorem 2.1. Indeed, the martingale part in the canonical decomposition of B^2 is $B_t^2 - t = 2\int_0^t B_s \,\mathrm{d}B_s$, which leads to

$$\langle B_t^2 - t\rangle = 4\int_0^t B_s^2 \,\mathrm{d}s \to 0 \quad \text{as} \quad t \searrow 0 \quad \text{a.s.} \tag{2.14}$$

Since all items of Assumption 1.1 are satisfied, Theorem 2.1 tells us that $\frac{B_t^2}{\sqrt{t}}$ converges in distribution to a singular Gaussian random variable.

Example 2.15. Denoting by Φ the standard normal cumulative distribution function, we see that for any $p \in (0,1)$ and a standard Brownian motion B, the continuous process $X_t = B_t + \Phi^{-1}(p)\sqrt{t}$ satisfies $\mathbb{P}[X_t > 0] = p$ for all $t \geq 0$. (Although not related to the present topic, we recall that the process $X_t = B_t + \sqrt{t}$ occurs in Delbaen and Schachermayer [DS95, Example 3.4]. They show that when used as the price process of a financial security, X_t (and also $\exp(X_t)$) allows for immediate arbitrage; the arbitrage disappears if proportional transaction costs are introduced [Gua06, Example 4.1]).

Example 2.16. The *non-continuous* martingale $\{t - P_t\}_{t\in\mathbb{R}_+}$, where $\{P_t\}_{t\in\mathbb{R}_+}$ is a Poisson process with parameter 1, satisfies

$$\lim_{t\searrow 0} \mathbb{P}[t - P_t > 0] = 1.$$

Example 2.17. For $\delta \geq 0$, we consider the SDE

$$X_t^\delta = \delta t + 2\int_0^t \sqrt{X_s^\delta}\,\mathrm{d}B_s, \quad t \geq 0. \tag{2.18}$$

The standard existence and uniqueness results for SDEs do not apply, since $x \mapsto 2\sqrt{x}$ is not Lipschitz continuous. However, by the Yamada–Watanabe uniqueness result for one-dimensional SDEs [AS16; WY71], there exists a unique, strong solution $X^\delta = \{X_t^\delta\}_{t\in\mathbb{R}_+}$ of the SDE (2.18), called *squared Bessel process of dimension* δ. Furthermore, the process $\{X_t^\delta - \delta t\}_{t\in\mathbb{R}_+}$ is a continuous martingale. We claim that for every $p \in [0,1)$ there exists a $\delta \in \mathbb{R}_+$, such that

$$\lim_{t\searrow 0} \mathbb{P}\big[X_t^\delta - \delta t > 0\big] = p.$$

The case $p = 0$ is trivial, since for $\delta = 0$, we have $X_t^\delta \equiv 0$, thus

$$\lim_{t\searrow 0} \mathbb{P}\big[X_t^\delta - \delta t > 0\big] = 0. \tag{2.19}$$

We now consider the case $p \in (0, 1/2)$. By the scaling property of squared Bessel processes [RY99, Proposition 1.6, p. 443]

$$\lim_{t \searrow 0} \mathbb{P}\big[X_t^\delta - \delta t > 0\big] = \mathbb{P}\big[X_1^\delta > \delta\big]. \tag{2.20}$$

By [RY99, Corollary 1.4, p. 441] the random variable

$$X_1^\delta \sim \Gamma(\delta/2, 1/2), \quad \delta > 0,$$

in particular $\mathbb{E}\big[X_1^\delta\big] = \delta$ and $\mathrm{V}(X_1^\delta) = 2\delta$. Let $\varepsilon > 0$, then by Chebyshev's inequality

$$\mathbb{P}\big[X_1^\delta > \delta + \varepsilon\big] \le \frac{2\delta}{\varepsilon^2},$$

and so $\lim_{\delta \searrow 0} \mathbb{P}\big[X_1^\delta > \delta + \varepsilon\big] = 0$. Therefore, since for any fixed $0 < \delta \le 2$, the density of X_1^δ is a strictly decreasing function

$$\begin{aligned}
\lim_{\delta \searrow 0} \mathbb{P}\big[X_1^\delta > \delta\big] &= \lim_{\delta \searrow 0} \mathbb{P}\big[\delta + \varepsilon > X_1^\delta > \delta\big] \\
&\le \varepsilon \lim_{\delta \searrow 0} 2^{-\delta/2}\Gamma(\delta/2)^{-1}\delta^{\delta/2-1}\,\mathrm{e}^{-\delta/2} \\
&= \frac{\varepsilon}{2} \lim_{\delta \searrow 0} \delta^{\delta/2}(2\,\mathrm{e})^{-\delta/2}\Gamma(1+\delta/2)^{-1} = \frac{\varepsilon}{2}.
\end{aligned}$$

Thus, by taking the limit $\varepsilon \searrow 0$,

$$\lim_{\delta \searrow 0} \mathbb{P}\big[X_1^\delta > \delta\big] = 0. \tag{2.21}$$

On the other hand, according to [JKB94, p. 340]

$$\frac{X_1^\delta - \delta}{\sqrt{2\delta}} \xrightarrow{\mathrm{d}} X \sim \mathrm{N}(0, 1),$$

for $\delta \to \infty$, which implies

$$\lim_{\delta \to \infty} \mathbb{P}\big[X_1^\delta > \delta\big] = \frac{1}{2}. \tag{2.22}$$

By (2.20), (2.21), (2.22), and the intermediate value theorem for every $p \in (0, 1/2)$ there exists a $\delta \in \mathbb{R}_+$ such that

$$\lim_{t \searrow 0} \mathbb{P}\big[X_t^\delta - \delta t > 0\big] = p.$$

By (2.19) and by considering the martingales $\delta t - X_t^\delta$, $\delta \in \mathbb{R}_+$, we see that all values $p \in [0, 1)$ can be achieved

Thus, there exist continuous martingales $X = \{X_t\}_{t \in \mathbb{R}_+}$ starting in zero, such that

$$\lim_{t \searrow 0} \mathbb{P}[X_t > 0] = p,$$

for every $p \in [0, 1)$.

We now take a look at higher order terms beyond the limit in (2.12). If $X_t = B_t + bt$ is a one-dimensional Brownian motion with drift $b \in \mathbb{R}$, we have $\mathbb{P}[X_t > x_0] = 1/2 + \mathrm{O}\big(\sqrt{t}\big)$. Theorem 2.23, which we prove now, shows that this estimate persists for a larger class of Itō processes.

Theorem 2.23. *Suppose that the process X solves the SDE*

$$\begin{aligned} \mathrm{d}X_t &= b(t,\cdot)\,\mathrm{d}t + \sigma(t)\,\mathrm{d}B_t, \\ X_0 &= x_0, \end{aligned} \tag{2.24}$$

where $b : [0,\infty)\times\Omega \to \mathbb{R}^m$ is a bounded predictable process, $\sigma : [0,\infty) \to \mathbb{R}^{m\times m}$ is a locally square-integrable function taking values in the set of invertible matrices such that the smallest eigenvalue of $\sigma(\cdot)^\top\sigma(\cdot)$ is uniformly bounded away from 0 and B is a standard m-dimensional Brownian motion. Then the bounds

$$\mathrm{e}^{f_1(t)} \le \mathbb{P}\big[X_t^1 > X_0^1\big] \le \mathrm{e}^{f_2(t)}, \quad t > 0 \tag{2.25}$$

apply. Here the functions f_1, f_2 are given by

$$\begin{aligned} f_1(t) &= -\Bigg(1 + \sqrt{\frac{\|\sigma^{-1}b\|_{2,\infty}^2\,t}{2\log 2}}\Bigg)\Bigg(\log 2 + \sqrt{\frac{\|\sigma^{-1}b\|_{2,\infty}^2\,t\,\log 2}{2}}\Bigg), \\ f_2(t) &= -\Bigg(\sqrt{\frac{2\log 2}{\|\sigma^{-1}b\|_{2,\infty}^2\,t}} - 1\Bigg)\Bigg(\sqrt{\frac{\|\sigma^{-1}b\|_{2,\infty}^2\,t\,\log 2}{2}} - \frac{1}{2}\,\|\sigma^{-1}b\|_{2,\infty}^2\,t\Bigg), \end{aligned}$$

where $\|\sigma^{-1}b\|_{2,\infty} = \sup_{t,\omega}\|\sigma^{-1}(t)b(t,\omega)\|_2$. Moreover, in the limit $t \searrow 0$, the functions e^{f_1}, e^{f_2} admit the series expansions

$$\mathrm{e}^{f_1(t)} = \frac{1}{2} - \sqrt{\frac{\log 2}{2}}\|\sigma^{-1}b\|_{2,\infty}\sqrt{t} + \mathrm{O}(t), \tag{2.26}$$

$$\mathrm{e}^{f_2(t)} = \frac{1}{2} + \sqrt{\frac{\log 2}{2}}\|\sigma^{-1}b\|_{2,\infty}\sqrt{t} + \mathrm{O}(t). \tag{2.27}$$

Remark 2.28. The term $\|\sigma^{-1}b\|_{2,\infty}$ in (2.26) and (2.27) has a clear interpretation: A small drift favors symmetry of the law of X_t, whereas a small volatility destroys the symmetry, leading to a larger error term.

Proof. Fix a $t > 0$ and make a change of probability measure according to the Girsanov Theorem, with the corresponding density being given by

$$\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}} := Z_t^{-1} := \mathrm{e}^{-N_t - \frac{1}{2}\langle N\rangle_t} := \mathrm{e}^{-\int_0^t \sigma(s)^{-1}b(s,\cdot)\,\mathrm{d}B_s - \frac{1}{2}\int_0^t \|\sigma(s)^{-1}b(s,\cdot)\|_2^2\,\mathrm{d}s}.$$

Under $\mathbb{Q}$, the process X solves the equation

$$\mathrm{d}X_s = \sigma(s)\,\mathrm{d}B_s^{\mathbb{Q}} \tag{2.29}$$

on $[0,t]$ with initial condition $X_0 = x_0$ and where $B^{\mathbb{Q}}$ is a standard Brownian motion under $\mathbb{Q}$. Thus, $\mathbb{Q}[X_s^1 > X_0^1] = 1/2$ for all $s \in (0,t]$. Moreover,

$$\mathbb{P}\big[X_t^1 > X_0^1\big] = \mathbb{E}_{\mathbb{Q}}\big[Z_t\,1_{\{X_t^1 > X_0^1\}}\big]. \tag{2.30}$$

To obtain upper and lower bounds on the latter expression, we fix numbers $p, q > 1$ such that $p^{-1} + q^{-1} = 1$ and apply Hölder's inequality to deduce

$$\begin{aligned} \mathbb{Q}\big[X_t^1 > X_0^1\big] &= \mathbb{E}_{\mathbb{Q}}\Big[1_{\{X_t^1 > X_0^1\}} Z_t^{\frac{1}{p}} Z_t^{-\frac{1}{p}}\Big] \\ &\le \mathbb{E}_{\mathbb{Q}}\big[1_{\{X_t^1 > X_0^1\}} Z_t\big]^{\frac{1}{p}}\,\mathbb{E}_{\mathbb{Q}}\Big[Z_t^{-\frac{q}{p}}\Big]^{\frac{1}{q}}. \end{aligned}$$

Taking the p-th power and rearranging, we get

$$\begin{aligned}\mathbb{Q}\big[X_t^1 > X_0^1\big]^p\,\mathbb{E}_{\mathbb{Q}}\Big[Z_t^{-\frac{q}{p}}\Big]^{-\frac{p}{q}} &\le \mathbb{E}_{\mathbb{Q}}\big[1_{\{X_t^1>X_0^1\}}Z_t\big]\\ &\le \mathbb{Q}\big[X_t^1 > X_0^1\big]^{\frac{1}{q}}\,\mathbb{E}_{\mathbb{Q}}\big[Z_t^p\big]^{\frac{1}{p}},\end{aligned}$$

where the last upper bound follows again by Hölder's inequality. This can be simplified to

$$2^{-p}\,\mathbb{E}_{\mathbb{Q}}\big[Z_t^{-\frac{q}{p}}\big]^{-\frac{p}{q}} \le \mathbb{P}\big[X_t^1 > X_0^1\big] \le 2^{-\frac{1}{q}}\,\mathbb{E}_{\mathbb{Q}}\big[Z_t^p\big]^{\frac{1}{p}},$$

or

$$2^{-p}\,\mathbb{E}_{\mathbb{P}}\Big[Z_t^{-\frac{q}{p}-1}\Big]^{-\frac{p}{q}} \le \mathbb{P}\big[X_t^1 > X_0^1\big] \le 2^{-\frac{1}{q}}\,\mathbb{E}_{\mathbb{P}}\big[Z_t^{p-1}\big]^{\frac{1}{p}}. \tag{2.31}$$

To estimate the bounds further, we note that

$$\begin{aligned}Z_t^{-\frac{q}{p}-1} &= \mathrm{e}^{-(\frac{q}{p}+1)N_t-\frac{1}{2}(\frac{q}{p}+1)^2\langle N\rangle_t}\,\mathrm{e}^{\frac{1}{2}\left(\frac{q}{p}+1\right)\frac{q}{p}\langle N\rangle_t}\\ &\le \mathrm{e}^{-(\frac{q}{p}+1)N_t-\frac{1}{2}(\frac{q}{p}+1)^2\langle N\rangle_t}\,\mathrm{e}^{\frac{1}{2}\left(\frac{q}{p}+1\right)\frac{q}{p}\,t\,\|\sigma^{-1}b\|^2_{2,\infty}}\end{aligned} \tag{2.32}$$

and

$$\begin{aligned}Z_t^{p-1} &= \mathrm{e}^{(p-1)N_t-\frac{1}{2}(p-1)^2\langle N\rangle_t}\,\mathrm{e}^{\frac{1}{2}(p-1)p\,\langle N\rangle_t}\\ &\le \mathrm{e}^{(p-1)N_t-\frac{1}{2}(p-1)^2\langle N\rangle_t}\,\mathrm{e}^{\frac{1}{2}(p-1)p\,t\,\|\sigma^{-1}b\|^2_{2,\infty}}\,.\end{aligned} \tag{2.33}$$

The first factors in (2.32) respectively (2.33) are $\mathbb{P}$-martingales, since Novikov's condition is satisfied by our assumptions on b and σ. Therefore, by inserting these estimates into (2.31), we obtain

$$\sup_{\substack{\frac{1}{p}+\frac{1}{q}=1\\ p>1}} 2^{-p}\,\mathrm{e}^{-\frac{1}{2}\left(\frac{q}{p}+1\right)t\,\|\sigma^{-1}b\|^2_{2,\infty}} \le \mathbb{P}\big[X_t^1 > X_0^1\big] \le \inf_{\substack{\frac{1}{p}+\frac{1}{q}=1\\ p>1}} 2^{-\frac{1}{q}}\,\mathrm{e}^{\frac{1}{2}(p-1)\,t\,\|\sigma^{-1}b\|^2_{2,\infty}}.$$

It can be easily seen that the lower bound is maximized by

$$p = 1 + \sqrt{\frac{\|\sigma^{-1}b\|^2_{2,\infty}\,t}{2\log 2}}\,,$$

whereas the upper bound is minimized by

$$p = \sqrt{\frac{2\log 2}{\|\sigma^{-1}b\|^2_{2,\infty}\,t}},$$

which together give (2.25). Finally, the expansions given in the statement of the theorem can be computed by Taylor expansions of the explicit functions in the lower and upper bounds. □

Theorem 2.34 (Functional CLT). *Let X satisfy Assumption 1.1. Then for every $f : \mathbb{R}^m \to \mathbb{R}^n$ such that there exists an open neighborhood U of x_0 with $f \in \mathrm{C}^2(U;\mathbb{R}^n)$, the processes*

$$Y^{f,u} := \Big(\frac{f(X_{ut}) - f(x_0)}{\sqrt{u}}\Big)_{t\in[0,T]}, \quad u \in (0,1),$$

converge in law to a Gaussian process with variance-covariance matrix

$$V = (Df)(x_0)L(Df(x_0)L)^\top$$

as $u \searrow 0$.

Proof. Let $\widetilde{B}$ be a Brownian motion with variance-covariance matrix V and let $(u_l)_{l\in\mathbb{N}}$ be a sequence with elements in $(0,1)$ such that $u_l \searrow 0$ as $l \to \infty$. It is sufficient to verify the convergence of the finite-dimensional distributions

$$(Y^{f,u_l}_{t_1}, \dots, Y^{f,u_l}_{t_w}) \xrightarrow{\mathrm{d}} (\widetilde{B}_{t_1}, \dots, \widetilde{B}_{t_w}), \quad t_1,\dots,t_w \in [0,T], \quad w \in \mathbb{N}, \tag{2.35}$$

and the tightness condition

$$\lim_{\delta\searrow 0} \varlimsup_{l\to\infty} \mathbb{P}\Big[\sup_{|s-t|\le\delta} |Y^{f,u_l}_s - Y^{f,u_l}_t| > \varepsilon\Big] = 0, \quad \varepsilon > 0. \tag{2.36}$$

By [SV06, Theorem 1.3.2], condition (2.36) implies the tightness of the laws of Y^{f,u_l}, $l \in \mathbb{N}$. Moreover, the convergence (2.35) allows to identify the limit points with the law of $\widetilde{B}$.

First, we focus on (2.35). Fix $t_1, \dots, t_w \in [0,T]$ for some $w \in \mathbb{N}$; then by the Cramér–Wold theorem it suffices to show

$$\sum_{d=1}^{w}\sum_{j=1}^{n} s_{dj} (Y^{f,u_l}_{t_d})^j \xrightarrow{\mathrm{d}} \sum_{d=1}^{w}\sum_{j=1}^{n} s_{dj} \widetilde{B}^j_{t_d}$$

for all $s \in \mathbb{R}^{w\times n}$ as $l \to \infty$. Let τ be defined as $\hat{\tau}$ in (2.10). Arguing as in the proof of Theorem 2.1, we see that it is enough to demonstrate

$$\sum_{d=1}^{w}\sum_{j=1}^{n} s_{dj} (Y^{f,u_l}_{t_d\wedge\tau})^j \xrightarrow{\mathrm{d}} \sum_{d=1}^{w}\sum_{j=1}^{n} s_{dj} \widetilde{B}^j_{t_d}$$

as $l \to \infty$. However, this can be proved analogously to (2.7).

To show (2.36), note that we may work with the stopped processes $Y^{f,u_l}_{t\wedge\tau}$, $l \in \mathbb{N}$. Indeed, since τ is a.s. positive, we have

$$\varlimsup_{l\to\infty} \mathbb{P}\Big[\sup_{t\in[0,T]} |Y^{f,u_l}_t - Y^{f,u_l}_{t\wedge\tau}| > \varepsilon\Big] = 0, \quad \varepsilon > 0.$$

By the triangle inequality, equation (2.36) is implied by

$$\lim_{\delta\searrow 0} \varlimsup_{l\to\infty} \mathbb{P}\Big[\sup_{|s-t|\le\delta} |(Y^{f,u_l}_{s\wedge\tau})^j - (Y^{f,u_l}_{t\wedge\tau})^j| > \varepsilon\Big] = 0, \quad \varepsilon > 0,\ j \in \{1,\dots,n\}. \tag{2.37}$$

By Itō's formula, we have

$$f_j(X_{t\wedge\tau}) - f_j(x_0) = \int_0^{t\wedge\tau} (\mathscr{L}_s f_j)(X_s)\,\mathrm{d}s + \sum_{k,l=1}^m \int_0^{t\wedge\tau} \frac{\partial f_j}{\partial x_l}(X_s)\,\sigma_s^{lk}\,\mathrm{d}B_s^k, \quad t \geq 0, \tag{2.38}$$

where

$$(\mathscr{L}_s f_j)(u) = \frac{1}{2}\sum_{k,l=1}^m \psi_s^{kl}\,\frac{\partial^2 f_j}{\partial x_k \partial x_l}(u) + \sum_{k=1}^m b_s^k\,\frac{\partial f_j}{\partial x_k}(u), \quad u \in U, \quad s \in [0,\tau],$$

with $\Psi_t = (\psi_t^{jk})_{1\leq j,k\leq m} := \sigma_t\sigma_t^\top$ for all $j \in \{1,\ldots,n\}$. From (2.38) we get

$$\begin{aligned}\mathbb{P}\Big[\sup_{|s-t|\leq\delta} |(Y_{s\wedge\tau}^{f,u_l})^j - (Y_{t\wedge\tau}^{f,u_l})^j| > \varepsilon\Big] &\leq \mathbb{P}\Big[\sup_{|s-t|\leq\delta} \frac{1}{\sqrt{u_l}}\int_{u_l(s\wedge\tau)}^{u_l(t\wedge\tau)} |(\mathscr{L}_r f_j)(X_r)|\,\mathrm{d}r > \frac{\varepsilon}{2}\Big] \\ &\quad + \mathbb{P}\Big[\sup_{|s-t|\leq\delta}\Big|\sum_{k,v=1}^m \int_{u_l(s\wedge\tau)}^{u_l(t\wedge\tau)} \frac{\partial f_j}{\partial x_v}(X_r)\sigma_r^{vk}\,\mathrm{d}B_r^k\Big| > \frac{\varepsilon\sqrt{u_l}}{2}\Big].\end{aligned} \tag{2.39}$$

By Part 3 and Part 5 of Assumption 1.1 and the choice of $\mathscr{B}$ there exists a random variable $C < \infty$ a.s. such that $\sup_{u\in\mathscr{B}} |(\mathscr{L}_s f_j)(u)| \leq C$ a.s. for $s \in [0,\tau]$, $j \in \{1,\ldots,n\}$. Thus, we have

$$\mathbb{P}\Big[\sup_{|s-t|<\delta} \frac{1}{\sqrt{u_l}}\int_{u_l(s\wedge\tau)}^{u_l(t\wedge\tau)} |(\mathscr{L}_r f_j)(X_r)|\,\mathrm{d}r > \frac{\varepsilon}{2}\Big] \leq \mathbb{P}\Big[C\sqrt{u_l}\,\delta > \frac{\varepsilon}{2}\Big] \xrightarrow{l\to\infty} 0, \quad \varepsilon > 0.$$

We now investigate the second term on the right-hand side of (2.39). After fixing δ, j and l, we define the process

$$F_t := \sum_{k,v=1}^m \int_0^{u_l(t\wedge\tau)} \frac{\partial f_j}{\partial x_v}(X_r)\sigma_r^{vk}\,\mathrm{d}B_r^k, \quad t \in [0,T].$$

In addition, we introduce the processes

$$G_t^i := F_{i\delta+t} - F_{i\delta}, \quad t \in I_i := [0,\delta], \quad i \in \{0,\ldots,\lfloor T/\delta\rfloor - 1\},$$

and for $i = \lfloor T/\delta\rfloor$,

$$G_t^{\lfloor T/\delta\rfloor} := F_{\lfloor T/\delta\rfloor\delta+t} - F_{\lfloor T/\delta\rfloor\delta}, \quad t \in I_{\lfloor T/\delta\rfloor} := [0, T - \lfloor T/\delta\rfloor].$$

These are continuous local martingales and thus, each of them can be represented as a time changed Brownian motion (see e.g. [Kal02, Theorem 18.4, p. 352]): $G_t^i = W_{\langle G^i\rangle_t}^i$. Moreover, the quadratic variation of G^i can be bounded according to

$$\langle G^i\rangle_t \leq \gamma\, C_\sigma^2\, u_l\,\delta, \quad t \in I_i, \quad i \in \{0,\ldots,\lfloor T/\delta\rfloor\},$$

where $0 < \gamma < \infty$ only depends on m and the Jacobian of f on the ball $\overline{\mathscr{B}}$ (see the paragraph preceding (2.10) for the definition of the latter). Now, consider the event $\{\sup_{|t-s|<\delta} |F_t - F_s| > \frac{\varepsilon\sqrt{u_l}}{2}\}$. Clearly, on this event there exist s_0, $t_0 \in [0,T]$ such that $|s_0 - t_0| \leq \delta$ and $|F_{t_0} - F_{s_0}| > \frac{\varepsilon\sqrt{u_l}}{2}$. Without loss of generality we may assume that $0 \leq s_0 < \delta \leq t_0 < 2\delta$

(the other cases can be dealt with in the same manner). Then, either $|F_\delta - F_{s_0}| > \frac{\varepsilon\sqrt{u_l}}{4}$, or $|F_{t_0} - F_\delta| > \frac{\varepsilon\sqrt{u_l}}{4}$. In the first case we get

$$\frac{\varepsilon\sqrt{u_l}}{4} < |F_\delta - F_{s_0}| \le |F_{s_0} - F_0| + |F_\delta - F_0| \le 2 \sup_{r\in[0,\delta]} |F_r - F_0|. \tag{2.40}$$

In the second case we have

$$\frac{\varepsilon\sqrt{u_l}}{4} < |F_{t_0} - F_\delta| \le \sup_{r\in[0,\delta]} |F_{\delta+r} - F_\delta|. \tag{2.41}$$

These considerations show that on the event $\{\sup_{|t-s|<\delta} |F_t - F_s| > \frac{\varepsilon\sqrt{u_l}}{2}\}$ there exists an index $i \in \{0, \dots \lfloor T/\delta\rfloor\}$ such that $\sup_{t\in I_i} |G_t^i| > \frac{\varepsilon\sqrt{u_l}}{8}$. Putting everything together we obtain

$$\begin{aligned}
\mathbb{P}\Big[\sup_{|s-t|<\delta}\Big|\sum_{k,v=1}^{m}\int_{u_l(s\wedge\tau)}^{u_l(t\wedge\tau)} \frac{\partial f_j}{\partial x_v}(X_r)\sigma_r^{vk}\,\mathrm{d}B_r^k\Big| > \frac{\varepsilon\sqrt{u_l}}{2}\Big] &\le \mathbb{P}\Big[\sup_{t\in I_i}|G_t^i| > \frac{\varepsilon\sqrt{u_l}}{8} \text{ for at least one } i\Big] \\
&\le \sum_{i=0}^{\lfloor T/\delta\rfloor} \mathbb{P}\Big[\sup_{t\in I_i}|G_t^i| > \frac{\varepsilon\sqrt{u_l}}{8}\Big] \\
&\le \Big(\frac{T}{\delta}+1\Big)\mathbb{P}\Big[\sup_{0\le r\le\gamma\, C_\sigma^2\, u_l\,\delta} |W_r^i| > \frac{\varepsilon\sqrt{u_l}}{8}\Big] \\
&\le \Big(\frac{T}{\delta}+1\Big)\,\mathrm{e}^{-\frac{\varepsilon^2}{128\,\gamma\, C_\sigma^2\,\delta}} \xrightarrow{\delta\to 0} 0.
\end{aligned}$$

The last estimate follows from Bernstein's inequality [RY99, Exercise 3.16, p. 153]. We have established (2.37), which finishes the proof. □

Remark 2.42. This chapter was taken from [Ger+15].

Chapter 3

Central Limit Theorems for Semimartingales with Jumps

This chapter is devoted to extensions of Theorems 2.1, 2.23, and 2.34 to semimartingales with jumps.

Theorem 3.1 (CLT with jumps)**.** *Let X satisfy Assumption 1.3. Then for every $f : \mathbb{R}^m \to \mathbb{R}^n$ such that there exists an open neighborhood U of x_0 with $f \in \mathrm{C}^2(U;\mathbb{R}^n)$, we have*

$$\frac{1}{\sqrt{t}}(f(X_t) - f(x_0)) \xrightarrow{\mathrm{d}} N_f \qquad \text{as } t \searrow 0,$$

where N_f is a normal random vector with mean 0 and covariance matrix

$$V = (Df)(x_0)L(Df(x_0)L)^\top.$$

Proof. Let $r > 0$ be such that the closed ball $\overline{\mathscr{B}}_r(x_0)$ with radius r around x_0 is contained in U. Further, we denote by $\overline{\mathscr{B}}_{r/2}(x_0)$ the closed ball with radius $r/2$ around x_0 and define the hitting time $\overline{\tau} := \tau_{\overline{\mathscr{B}}_{r/2}(x_0)^c}$ for X, which is a stopping time by [Pro05, Theorem 3, p.4]. Finally, we introduce the stopping time

$$\tau := \overline{\tau} \wedge \tau_A \wedge \tau_M \wedge \tau_J \tag{3.2}$$

and notice that τ a.s. positive. Therefore, by the same argument as in the proof of Theorem 2.1, it suffices to show

$$\frac{1}{\sqrt{t}}(f(X_{t\wedge\tau}) - f(x_0)) \xrightarrow{\mathrm{d}} N_f \quad \text{as} \quad t \searrow 0.$$

By Itō's formula in the form of [CT04, Proposition 8.19], we have for all $j \in \{1, \ldots, n\}$ and $t \in [0, T]$:

$$f_j(X_{t\wedge\tau}) - f_j(x_0) = \int_0^{t\wedge\tau} (\mathscr{L}_s f_j)(X_s)\,\mathrm{d}s + \sum_{k,l=1}^m \int_0^{t\wedge\tau} \frac{\partial f_j}{\partial x_l}(X_s)\sigma_s^{lk}\,\mathrm{d}B_s^k \tag{3.3}$$

$$+ \int_0^{t\wedge\tau} \int_{\mathscr{B}_1} \big(f_j(X_{s-} + \psi(s,z)) - f_j(X_{s-})\big)\big(\Pi(\mathrm{d}s,\mathrm{d}z) - \mu(\mathrm{d}s,\mathrm{d}z)\big) \tag{3.4}$$

$$+ \int_0^{t\wedge\tau} \int_{\mathbb{R}^m\setminus\mathscr{B}_1} \big(f_j(X_{s-} + \varphi(s,z)) - f_j(X_{s-})\big)\,\Pi(\mathrm{d}s,\mathrm{d}z). \tag{3.5}$$

Arguing similarly to the proof of Theorem 2.1, we see that the vector of terms on the right-hand side of (3.3), rescaled by $1/\sqrt{t}$, converges in distribution to N_f as $t \searrow 0$. Thus, the theorem will follow if we can show that the terms (3.4) and (3.5), rescaled by $1/\sqrt{t}$, converge to zero in probability as $t \searrow 0$.

The term (3.4), rescaled by $1/\sqrt{t}$, can be decomposed into a sum $T_t^1 + T_t^2$ of the following two terms:

$$\frac{1}{\sqrt{t}} \int_0^{t\wedge\tau} \int_{\mathscr{B}_1} \big(f_j(X_{s-} + \psi(s,z)) - f_j(X_{s-})\big)\, 1_{\{|\psi(s,z)|<r/2\}} \big(\Pi(\mathrm{d}s,\mathrm{d}z) - \mu(\mathrm{d}s,\mathrm{d}z)\big),$$
$$\frac{1}{\sqrt{t}} \int_0^{t\wedge\tau} \int_{\mathscr{B}_1} \big(f_j(X_{s-} + \psi(s,z)) - f_j(X_{s-})\big)\, 1_{\{|\psi(s,z)|\geq r/2\}} \big(\Pi(\mathrm{d}s,\mathrm{d}z) - \mu(\mathrm{d}s,\mathrm{d}z)\big).$$

Then:

$$\mathbb{E}\big[|T_t^1|\big] \leq \frac{2\|f|_{\overline{\mathscr{B}}_r(x_0)}\|_\infty}{\sqrt{t}}\, \mathbb{E}\Big[\big|\Pi - \mu\big|\big([0, t\wedge\tau_J] \times \mathscr{B}_1\big)\Big],$$

which converges to zero as $t \searrow 0$ by part (3) of Assumption 1.3. Moreover, since J a.s. has only finitely many jumps of absolute size greater than $r/2$ on every finite time interval, T_t^2 converges to 0 a.s. as $t \searrow 0$.

Lastly, the term (3.5), rescaled by $1/\sqrt{t}$, converges to zero a.s. as $t \searrow 0$, since J a.s. has only finitely many jumps of absolute size greater than 1 on every finite time interval. □

As in the case of continuous semimartingales, the CLT can be strengthened to a Functional CLT, which in the presence of jumps reads as follows.

Theorem 3.6 (Functional CLT with jumps)**.** *Let X satisfy Assumption 1.3. Then for every $f : \mathbb{R}^m \to \mathbb{R}^n$, such that there exists an open neighborhood U of x_0 with $f \in \mathrm{C}^2(U;\mathbb{R}^n)$, the processes*

$$Y^{f,u} := \Big(\frac{f(X_{ut}) - f(x_0)}{\sqrt{u}}\Big)_{t\in[0,T]}, \quad u \in (0,1),$$

converge in law to a Brownian motion with variance-covariance matrix given by

$$V = (Df)(x_0)L(Df(x_0)L)^\top$$

as $u \searrow 0$.

Proof. For each f and u as in the statement of the theorem, we write $Q^{f,u}$ for the law of the process $Y^{f,u}$ on $\mathrm{D}([0,T];\mathbb{R}^n)$, the space of right-continuous functions on $[0,T]$ having left limits; we denote by $Q_c^{f,u}$ the law of the continuous part of $Y^{f,u}$ on $\mathrm{C}([0,T];\mathbb{R}^n)$. We claim first that the family $(Q^{f,u})_{u\in(0,1)}$ is tight on $\mathrm{D}([0,T],\mathbb{R}^n)$ if and only if the family $(Q_c^{f,u})_{u\in(0,1)}$ is tight on $\mathrm{C}([0,T];\mathbb{R}^n)$ and moreover that the limit points of the two families are the same.

To prove the claim, it suffices to show that for every $\varepsilon > 0$ and $j \in \{1,\dots,n\}$:

$$\mathbb{P}\Big[\sup_{t\in[0,T]} |(J_t^{f,u})^j| > \varepsilon\Big] \to 0 \quad \text{as} \quad u \searrow 0, \tag{3.7}$$

where $J^{f,u}$ denotes the jump part of $Y^{f,u}$. Indeed, if this is the case, then every converging subsequence of $(Q^{f,u})_{u\in(0,1)}$ in $\mathrm{D}([0,T],\mathbb{R}^n)$ corresponds to a converging subsequence of

$(Q_c^{f,u})_{u\in(0,1)}$ in $\mathrm{C}([0,T];\mathbb{R}^n)$ and the limits of the two subsequences have to coincide. Now, since the stopping time defined in (3.2) is a.s. positive, (3.7) is implied by

$$\mathbb{P}\Big[\sup_{t\in[0,T]} |(J_t^{f,u})^j| > \varepsilon,\ \tau > uT\Big] \to 0 \quad \text{as} \quad u \searrow 0. \tag{3.8}$$

Furthermore, by Itō's formula in the form of [CT04, Proposition 8.19], we have on the event $\{\tau > uT\}$:

$$(J_t^{f,u})^j = \frac{1}{\sqrt{u}} \int_0^{ut} \int_{\mathscr{B}_1} \big(f_j(X_{s-} + \psi(s,z)) - f_j(X_{s-})\big)\big(\Pi(\mathrm{d}s,\mathrm{d}z) - \mu(\mathrm{d}s,\mathrm{d}z)\big) \tag{3.9}$$

$$+ \frac{1}{\sqrt{u}} \int_0^{ut} \int_{\mathbb{R}^m\setminus\mathscr{B}_1} \big(f_j(X_{s-} + \varphi(s,z)) - f_j(X_{s-})\big)\,\Pi(\mathrm{d}s,\mathrm{d}z). \tag{3.10}$$

As in the proof of Theorem 3.1, we decompose the integral on the right-hand side of (3.9) according to whether $|\psi(s,z)| < r/2$, or $|\psi(s,z)| \geq r/2$ and call the two resulting processes $(J^{f,u,1})^j$ and $(J^{f,u,2})^j$. Since the process $(J^{f,u,1})^j$ is obtained by integrating a predictable process with respect to a compensated Poisson random measure, it is a square-integrable martingale. Thus, by Doob's maximal inequality, we have

$$\mathbb{P}\Big[\sup_{t\in[0,T]} |(J_t^{f,u,1})^j| > \frac{\varepsilon}{2},\ \tau > uT\Big]$$
$$\leq \frac{2}{\varepsilon\sqrt{u}}\mathbb{E}\Big[\Big|\int_0^{(uT)\wedge\tau} \int_{\mathscr{B}_1} \big(f_j(X_{s-} + \psi(s,z)) - f_j(X_{s-})\big)1_{\{|\psi(s,z)|<r/2\}}\overline{\Pi}(\mathrm{d}s,\mathrm{d}z)\Big|\Big],$$

where we write $\overline{\Pi}$ for $\Pi - \mu$. Moreover, the same argument as in the proof of Theorem 3.1 shows that the latter upper bound tends to zero as $u \searrow 0$ (by virtue of part (3) of Assumption 1.3). Finally, since a.s. the process $J^{f,u}$ has finitely many jumps of size greater than $r/2$ on every finite time interval, the random variables $\sup_{t\in[0,T]} |(J_t^{f,u,2})^j|$ converge to zero a.s. as $u \searrow 0$. In addition, by the same reasoning, the supremum over $t \in [0,T]$ of (3.10) tends to zero a.s. as $u \searrow 0$ as well. Putting everything together, we end up with (3.8), finishing the proof of the claim.

Lastly, one can proceed as in the proof of Theorem 2.34 to first show the tightness of the family $(Q_c^{f,u})_{u\in(0,1)}$ on $\mathrm{C}([0,T];\mathbb{R}^n)$ and to subsequently identify each of its limit points with the law of a Brownian motion with variance-covariance matrix V. In view of the claim above, this finishes the proof. ☐

We conclude this section by stating and proving the analogue of Theorem 2.23 in the presence of jumps.

Theorem 3.11. *Suppose that the process X solves the SDE*

$$\mathrm{d}X_t = b(t,\cdot)\,\mathrm{d}t + \sigma(t)\,\mathrm{d}B_t + \int_{\mathbb{R}^m} \psi(t,y)\,\Pi(\mathrm{d}t,\mathrm{d}y), \tag{3.12}$$
$$X_0 = x_0,$$

where $b : [0,\infty)\times\Omega \to \mathbb{R}^m$ is a bounded predictable process with respect to the filtration of the standard m-dimensional Brownian motion B, $\sigma : [0,\infty) \to \mathbb{R}^{m\times m}$ is a locally square-integrable function taking values in the set of invertible matrices such that the smallest eigenvalue of $\sigma(\cdot)^\top\sigma(\cdot)$ is uniformly bounded away from 0, and ψ is a predictable process *with respect to the filtration of the Poisson random measure* Π. *Suppose further that* Π *is symmetric with respect to* y *(so that, in particular, its compensator vanishes) and that* $\psi_1(t,y) = -\psi_1(t,y)$ *for all* $t \geq 0$ *and* $y \in \mathbb{R}^m$ *with probability* 1. *Then the bounds*

$$\mathrm{e}^{f_1(t)} \leq \mathbb{P}\big[X_t^1 > X_0^1\big] \leq \mathrm{e}^{f_2(t)}, \quad t > 0, \tag{3.13}$$

of Theorem 2.23 apply with the same functions f_1, f_2 *as there.*

Proof. We start by fixing a $t > 0$ and changing the underlying probability measure $\mathbb{P}$ to an equivalent probability measure $\mathbb{Q}$ according to

$$\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}} = \mathrm{e}^{-\int_0^t \sigma(s)^{-1}b(s,\cdot)\,\mathrm{d}B_s - \frac{1}{2}\int_0^t \|\sigma(s)^{-1}b(s,\cdot)\|_2^2\,\mathrm{d}s}. \tag{3.14}$$

Then in view of the independence of the continuous and the jump parts of X under $\mathbb{P}$ and the Girsanov Theorem, the process X solves the SDE

$$\mathrm{d}X_s = \sigma(s)\,\mathrm{d}B_s^{\mathbb{Q}} + \int_{\mathbb{R}^m} \psi(s,y)\,\Pi(\mathrm{d}s,\mathrm{d}y), \quad s \in [0,t] \tag{3.15}$$

with a standard Brownian motion $B^{\mathbb{Q}}$ under $\mathbb{Q}$ and initial condition $X_0 = x_0$. Moreover, the random variables

$$U_t^{(1)} := \Big(\int_0^t \sigma(s)\,\mathrm{d}B_s^{\mathbb{Q}}\Big)_1 \quad \text{and} \quad U_t^{(2)} := \int_0^t \int_{\mathbb{R}^m} \psi_1(s,y)\,\Pi(\mathrm{d}s,\mathrm{d}y)$$

are independent under $\mathbb{Q}$ and their distributions $\eta^{(1)}$ and $\eta^{(2)}$ are symmetric. Hence,

$$\begin{aligned}
\mathbb{Q}[X_t^1 > X_0^1] &= \mathbb{Q}\big[U_t^{(1)} + U_t^{(2)} > 0\big] \\
&= \int_0^\infty \mathbb{Q}\big[U_t^{(1)} > -c\big]\,\eta^{(2)}(\mathrm{d}c) + \int_{-\infty}^0 \mathbb{Q}\big[U_t^{(1)} > -c\big]\,\eta^{(2)}(\mathrm{d}c) \\
&= \int_0^\infty \mathbb{Q}\big[U_t^{(1)} > -c\big]\,\eta^{(2)}(\mathrm{d}c) + \int_{-\infty}^0 \big(1 - \mathbb{Q}\big[U_t^{(1)} > c\big]\big)\,\eta^{(2)}(\mathrm{d}c) \\
&= \frac{1}{2}.
\end{aligned}$$

From now on, one can follow the lines of the proof of Theorem 2.23 to finish the proof. □

Remark 3.16. This chapter was taken from [Ger+15].

Chapter 4

Applications to Digital Options and the Implied Volatility Skew

We give applications to mathematical finance, concretely to the pricing of at-the-money digital options with short maturities and the asymptotics of at-the-money short time volatility skews.

Digital Options

Suppose that the one-dimensional, positive process S models the price of a financial asset and that $\mathbb{P}$ is the pricing measure. The riskless rate is $r > 0$. The holder of a digital call option with maturity T and strike K receives the payoff $1_{\{S_T > K\}}$ at maturity. Digital options are peculiar in that the owner receives the full payoff as soon as they are only slightly in the money, as opposed to call options which kick in gradually. By the risk-neutral pricing formula, the value of the digital call at time zero is

$$D(K,T) := \mathrm{e}^{-rT}\ \mathbb{E}[1_{\{S_T > K\}}] = \mathrm{e}^{-rT}\ \mathbb{P}[S_T > K]\,.$$

There exists considerable literature on short-maturity approximations for option prices. For out-of-the-money ($S_0 < K$) or in-the-money ($S_0 > K$) digitals, the first order approximation is clear: As soon as the underlying S is a.s. right-continuous at $t = 0$, the dominated convergence theorem yields

$$\lim_{T \searrow 0} D(K,T) = \begin{cases} 0 & \text{if } S_0 < K, \\ 1 & \text{if } S_0 > K. \end{cases}$$

Finer information on the out-of-the-money decay (which trivially also covers the in-the-money behavior) comes from small-time large deviations principles for the underlying. See e.g. Forde and Jacquier [FJ09] for the case of the Heston model and references about other diffusion processes. Our CLT-type results are useful in the at-the-money case ($S_0 = K$). As an immediate consequence of our limit theorems, we get:

Theorem 4.1. *If the process S satisfies the assumptions of Theorem 3.1 (in particular, if it satisfies those of Theorem 2.1 or Remark 1.2) and the limit law is non-singular, then the limiting price of an at-the-money digital call is $1/2$:*

$$\lim_{T \searrow 0} D(S_0,T) = \frac{1}{2}. \tag{4.2}$$

This (intuitive) result captures virtually all diffusion-based models that have been considered (e.g. Black-Scholes, constant elasticity of variance, Heston, Stein-Stein). Although it seems to be new in its generality, in particular for jump processes, some special cases can be inferred from literature.

The jump processes used in financial modeling are often Lévy processes. It is clear that a compensated compound Poisson process will yield an (unrealistic) at-the-money digital price limit of either zero or one, see Example 2.16. As for the infinite activity case, limit laws are not the appropriate way to get a result like (4.2). Doney and Maller [DM02] have determined all Lévy processes that admit a short-time CLT, with a criterion involving the tail of the Lévy measure. While there do exist infinite activity Lévy processes that satisfy a CLT [DM02, Remark 9], the Lévy processes that have been considered in mathematical finance are typically *not* of this kind. For instance, it is easy to see from the characteristic function that the variance gamma process [MCC98] does not admit *any* non-singular limit law for $t \searrow 0$ for any normalization. These issues are further discussed in [GGP16].

Implied Volatility Skew

Finally, we discuss the implied volatility skew. Suppose that the underlying S generates the call price surface

$$C(K,T) = \mathrm{e}^{-rT}\, \mathbb{E}[(S_T - K)^+], \quad K, T > 0.$$

Then the implied volatility for strike K and maturity T is the volatility $\sigma_{\mathrm{imp}}(K,T)$ that makes the Black-Scholes call price equal to $C(K,T)$,

$$C_{\mathrm{BS}}(K, \sigma_{\mathrm{imp}}, T) = C(K,T),$$

see e.g. [Lee05]. The map $K \searrow \sigma_{\mathrm{imp}}(K,T)$ is called the *volatility smile* for maturity T. It is also called the volatility skew, because it is often monotone instead of smile-shaped, but we will reserve the term *volatility skew* for the derivative $\partial_K \sigma_{\mathrm{imp}}(K,T)$. If $C(K,T)$ is smooth in K, it equals

$$\partial_K \sigma_{\mathrm{imp}} = -\frac{\partial_K C_{\mathrm{BS}} - \partial_K C}{\partial_\sigma C_{\mathrm{BS}}}.$$

Under mild assumptions (e.g., if the law of S_T is absolutely continuous), we have

$$\partial_K C = -\,\mathrm{e}^{-rT}\, \mathbb{P}[S_T \geq K] = -D(K,T), \tag{4.3}$$

from which we deduce the well-known connection between the volatility skew and the price of a digital call

$$\partial_K \sigma_{\mathrm{imp}} = -\frac{D(K,T) + \partial_K C_{\mathrm{BS}}}{\partial_\sigma C_{\mathrm{BS}}},$$

see e.g. [Gat06]. Inserting the explicit Black-Scholes vega and digital price, we obtain

$$\partial_K \sigma_{\mathrm{imp}} = \frac{-D(K,T) + \Phi(-\sigma_{\mathrm{imp}}\sqrt{T}/2)}{K\sqrt{T}\, n(\sigma_{\mathrm{imp}}\sqrt{T}/2)},$$

see e.g. [MR05], with Φ and n denoting the standard Gaussian cumulativ distribution function and density, respectively. For $T \searrow 0$, we have $\sigma_{\mathrm{imp}}\sqrt{T} = \mathrm{o}(1)$ under the following mild assumptions [RR09, Proposition 4.1]:

$$(S_0 - K)^+ \le C(K,T) \le S_0 \quad \text{(no arbitrage bounds)}, \tag{4.4}$$

$$\lim_{T \searrow 0} C(K,T) = (S_0 - K)^+, \tag{4.5}$$

$$T \mapsto C(K,T) \quad \text{is non-decreasing.} \tag{4.6}$$

Therefore,

$$\partial_K \sigma_{\text{imp}} \sim \frac{\sqrt{2\pi}}{K\sqrt{T}}\Big(\frac{1}{2} - D(K,T) - \frac{\sigma_{\text{imp}}\sqrt{T}}{2\sqrt{2\pi}} + \mathrm{O}((\sigma_{\text{imp}}\sqrt{T})^3)\Big), \quad T \searrow 0. \tag{4.7}$$

We notice that the small-time behavior of the skew is related to that of the digital price. At the money, the latter will typically tend to $1/2$ for continuous models (see Theorem 4.1) and so higher order estimates are needed to get the first order asymptotics of the at-the-money skew $\partial_K \sigma_{\text{imp}}|_{K=S_0}$. To this end, we apply Theorem 2.23 and compare our findings with the standard model free slope bounds [FPS00, p. 36]

$$-\frac{\sqrt{2\pi}}{S_0\sqrt{T}}(1 - \Phi(d_2))\,\mathrm{e}^{-rT + \frac{d_1^2}{2}} \le \frac{\partial \sigma_{\text{imp}}}{\partial K} \le \frac{\sqrt{2\pi}}{S_0\sqrt{T}}\Phi(d_2)\,\mathrm{e}^{-rT + \frac{d_1^2}{2}}, \tag{4.8}$$

where

$$d_1 = \frac{\log(S_0/K) + (r + \frac{1}{2}\sigma_{\text{imp}}^2)T}{\sigma_{\text{imp}}\sqrt{T}}, \quad d_2 = d_1 - \sigma_{\text{imp}}\sqrt{T}.$$

Such bounds can give guidance on model choice; recall that the market slope seems to grow like $T^{-1/2}$ for short maturities [Alò+08]. Note that the following result accommodates stochastic interest rates and remember that we assume in this section that the dimension is $m = 1$. Under stochastic interest rates, the digital call price is

$$D(K,T) = \mathbb{E}\big[\mathrm{e}^{-\int_0^T r(s)\,\mathrm{d}s}\, 1_{\{S_T > K\}}\big]. \tag{4.9}$$

To calculate the implied volatility, a deterministic rate r has to be chosen, e.g. by $\mathrm{e}^{-rT} = \mathbb{E}\big[\mathrm{e}^{-\int_0^T r(s)\,\mathrm{d}s}\big]$. However, this choice is irrelevant for Theorem 4.10.

Theorem 4.10. *Assume that the price process satisfies the SDE*

$$\frac{\mathrm{d}S_t}{S_t} = r(t)\,\mathrm{d}t + \sigma(t)\,\mathrm{d}B_t$$

with the stochastic short rate process $(r(t))_{t\ge 0}$ *and that the log-price* $X = \log S$*, whose drift is* $b(t) = r(t) - 1/2\sigma^2(t)$*, satisfies the assumptions of Theorem 2.23. Further assume that* $\partial_K C(K,T) = -D(K,T)$ *holds (see* (4.3)*) and that* (4.5) *and* (4.6) *are satisfied. Then we have the at-the-money slope bounds*

$$\partial_K \sigma_{\text{imp}}|_{K=S_0} \ge \frac{\sqrt{2\pi}}{K\sqrt{T}}\Big(- C\sqrt{T} - \frac{\sigma_{\text{imp}}\sqrt{T}}{2\sqrt{2\pi}} + \mathrm{O}(T) + \mathrm{O}\left(\left(\sigma_{\text{imp}}\sqrt{T}\right)^3\right)\Big),$$

$$\partial_K \sigma_{\text{imp}}|_{K=S_0} \le \frac{\sqrt{2\pi}}{K\sqrt{T}}\Big(C\sqrt{T} - \frac{\sigma_{\text{imp}}\sqrt{T}}{2\sqrt{2\pi}} + \mathrm{O}(T) + \mathrm{O}\left(\left(\sigma_{\text{imp}}\sqrt{T}\right)^3\right)\Big),$$

where

$$C = \sqrt{\frac{\log 2}{2}}\|\sigma^{-1} b\|_{2,\infty}.$$

Proof. According to (4.9), the at-the-money digital price equals

$$D(S_0, T) = \mathbb{E}\big[\, \mathrm{e}^{-\int_0^T r(s)\,\mathrm{d}s}\ 1_{\{X_T > x_0\}}\big].$$

The discount factor is $1 + \mathrm{O}(T)$ for $T \searrow 0$, so we can apply Theorem 2.23 to conclude

$$\frac{1}{2} - C\sqrt{T} + \mathrm{O}(T) \le D(S_0, T) \le \frac{1}{2} + C\sqrt{T} + \mathrm{O}(T).$$

Now, the result follows from (4.7). Note that $\sigma_{\mathrm{imp}}\sqrt{T} = \mathrm{o}(1)$ by [RR09, Proposition 4.1], since we assume (4.5), (4.6), and (4.4) is satisfied in our setup. ☐

The bounds in Theorem 4.10 are asymptotically stronger than the general estimate (4.8), which is of order $\mathrm{O}(T^{-1/2})$, since $\sigma_{\mathrm{imp}}\sqrt{T} = \mathrm{O}(1)$. If the Berestycki–Busca–Florent formula [BBF02] holds, then implied volatility tends to a constant. Therefore, our bounds are considerably stronger than (4.8) in this case, namely of order $\mathrm{O}(1)$. Thus, the models covered by Theorem 4.10 do *not* match the empirical slope behavior $T^{-1/2}$, similarly to stochastic volatility models [Lew00], whose slope also behaves like $\mathrm{O}(1)$.

To conclude our discussion of at-the-money digitals and the implied volatility skew, note that for some diffusion processes the result in Theorem 4.1 is implicitly in literature. To wit, by (4.7) a non-exploding at-the-money slope requires a limit price of $1/2$ of the digital. See Durrleman [Dur04, p. 59] for a general expression for the implied volatility slope that shows that it does not explode, e.g., in the Heston model.

Remark 4.11. This chapter was taken from [Ger+15].

Appendices

Appendix A

Auxiliary Results

Lemma A.1. *Given $\sigma > 0$ and $z \in \mathbb{R}$, the Stein solution*

$$f_{\sigma^2,z}(x) = \sigma\, \mathrm{e}^{\frac{x^2}{2\sigma^2}} \int_{-\infty}^{x/\sigma} \big(h_z(\sigma s) - g_z(\sigma^2)\big)\, \mathrm{e}^{-\frac{s^2}{2}}\, \mathrm{d}s, \qquad x \in \mathbb{R},$$

has the following properties for $u, v, w \in \mathbb{R}$:

1. $x \mapsto x f_{\sigma^2,z}(x)$, $x \in \mathbb{R}$ *is increasing,*
2. $|w f_{\sigma^2,z}(w)| \leq \sigma^2$ *and* $|w f_{\sigma^2,z}(w) - u f_{\sigma^2,z}(u)| \leq \sigma^2$,
3. $|f'_{\sigma^2,z}(w)| \leq 1$ *and* $|f'_{\sigma^2,z}(w) - f'_{\sigma^2,z}(v)| \leq 1$,
4. $0 < f_{\sigma^2,z}(w) \leq \min(\sqrt{2\pi\sigma^2}/4, \sigma^2/|z|)$,
5. $|(w+u) f_{\sigma^2,z}(w+u) - (w+v) f_{\sigma^2,z}(w+v)| \leq (|w| + \sqrt{2\pi\sigma^2}/4)(|u| + |v|)$;

Proof. The proof is analog to [BC05, Proof of Lemma 2.2, p. 54]. □

Lemma A.2 (Concentration inequality)**.** *For $n \in \mathbb{N}$, let $X_1, X_2, \ldots, X_n$ be independent square-integrable random variables with zero means such that $\sum_{i=1}^n \mathbb{E}[X_i^2] = 1$. Then, for all real $a < b$,*

$$\mathbb{P}\Big[a \leq \sum_{i=1}^n X_i \leq b\Big] \leq b - a + 2 \sum_{i=1}^n \mathbb{E}\big[|X_i|^3\big]\,, \tag{A.3}$$

Proof. See [CGS11, Proposition 3.1, p. 57]. □

Appendix B

Symbols and Notation

a.s.	Almost sure
e.g.	For example
i.e.	Id est
iff	If and only if
i.i.d.	Independent and identically distributed
CLT	Central limit theorem
LDP	Large deviations principle
NCLT	Non-central limit theorem
ODE	Ordinary differential equation
SDE	Stochastic differential equation
$\mathbb{N}$	$\{1, 2, 3, \dots\}$
$\mathbb{N}_0$	$\{0, 1, 2, \dots\}$
$\mathbb{R}_+$	$[0, \infty)$
$\sum_{i=1}^{0} X_i$	Zero per definition
$\mathrm{C}(A; B)$	Continuous functions $f : A \to B$
$\mathrm{C}(A)$	Continuous functions $f : A \to \mathbb{R}$
$\mathrm{C}^n(A)$	n times continuously differentiable functions $f : A \to B$
$\mathrm{C_b}(A)$	Set of bounded, continuous functions $f : A \to B$
$\mathrm{D}(A)$	Set of right-continuous functions on [0,T] having left limits
$\mathrm{H}^1(A)$	Absolutely continuous functions on A with a square-integrable derivative
$\mathrm{L}^2(A)$	Set of square-integrable functions on A
$\\|\varphi\\|_\infty$	$\sup_{x \in A} \|\varphi(x)\|$
$\\|x\\|_2$	Euclidean distance
$\\|\varphi\\|_{2,\infty}$	$\sup_{x \in A} \\|\varphi(x)\\|_2$
$\\|x\\|_F$	Frobenius norm
$\mathbb{E}[X]$	Expectation of X
$\mathbb{E}[X; A]$	$\int_A X \, \mathrm{d}\mathbb{P}$
$\mathbb{E}[X \mid \mathscr{F}]$	Conditional expectation of X with respect to $\mathscr{F}$
$\mathrm{V}(X)$	Variance of X
$\mathscr{B}(\mathbb{R})$	Borel sets of $\mathbb{R}$
$\xrightarrow{\mathrm{d}}$	Convergence in distribution
$\overset{\mathrm{d}}{=}$	Equal in distribution

$\mathscr{L}(X)$	Distribution of X
$\mathrm{B}(n,p)$	Binomial distribution
$\mathrm{Exp}(\lambda)$	Exponential distribution
$\mathrm{G}(p)$	Geometric distribution, $\mathbb{P}[X=n]=(1-p)p^n, \quad n\in\mathbb{N}_0$
$\mathrm{Geom}(p)$	Geometric distribution, $\mathbb{P}[X=n]=(1-p)^{n-1}p, \quad n\in\mathbb{N}$
$\mathrm{L}(\mu,\sigma)$	Laplace distribution
$\mathrm{N}(\mu,C)$	Gaussian distribution
$\mathrm{NB}(\alpha,p)$	Negative binomial distribution, $\mathbb{P}[X=n]=\binom{\alpha+n-1}{n}(1-p)^\alpha p^n, \quad n\in\mathbb{N}_0$
$\mathrm{P}(\lambda)$	Poisson distribution
$\Gamma(\alpha,\beta)$	Gamma distribution with density $f_{\alpha,\beta}(x)=\frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}\,\mathrm{e}^{-\beta x}, \quad x>0$
$\mathrm{d_{BL}}$	Bounded Lipschitz distance
$\mathrm{d_K}$	Kolmogorov distance
$\mathrm{d_W}$	Wasserstein distance
$\Phi(x)$	Standard Gaussian cumulative distribution function
$\varphi_X(t)$	Characteristic function of the random variable X
$\langle M,N\rangle_t$	Covariaton of the stochastic processes M and N
$\langle M\rangle_t$	Quadratic variation of the process M

Bibliography

[AG10] R. Arratia and L. Goldstein. "Size bias, sampling, the waiting time paradox, and infinite divisivility". arXiv:1007.3910v1. 2010.

[ALV07] E. Alòs, J. A. León, and J. Vives. "On the short-time behavior of the implied volatility for jump-diffusion models with stochastic volatility". In: *Finance Stoch.* 11.4 (2007), pp. 571–589.

[Alò+08] E. Alòs, J. A. León, M. Pontier, and J. Vives. "A Hull and White formula for a general stochastic volatility jump-diffusion model with applications to the study of the short-time behavior of the implied volatility". In: *J. Appl. Math. Stoch. Anal.* (2008), p. 17.

[AS16] S. Altay and U. Schmock. "Lecture notes on the Yamada–Watanabe condition for the pathwise uniqueness of solutions of certain stochastic differential equations". 2016.

[Bac00] L. Bachelier. "Théorie de la spéculation". In: *Ann. Sci. École Norm. Sup. (3)* 17 (1900), pp. 21–86.

[BBF02] H. Berestycki, J. Busca, and I. Florent. "Asymptotics and calibration of local volatility models". In: *Quant. Finance* 2.1 (2002). Special issue on volatility modelling, pp. 61–69.

[BBF04] H. Berestycki, J. Busca, and I. Florent. "Computing the implied volatility in stochastic volatility models". In: *Comm. Pure Appl. Math.* 57.10 (2004), pp. 1352–1373.

[BC05] A. D. Barbour and L. H. Y. Chen. *An Introduction to Stein's Method.* Singapore University Press, 2005.

[BC12] A. Bentata and R. Cont. "Short-time asymptotics for marginal distributions of semimartingales". arXiv:1202.1302. 2012.

[Ber27] S. Bernstein. "Sur l'extension du théoréme limite du calcul des probabilités aux sommes de quantités dépendantes". In: *Math. Ann.* 97.1 (1927), pp. 1–59.

[Ber41] A. C. Berry. "The accuracy of the Gaussian approximation to the sum of independent variates". In: *Trans. Amer. Math. Soc.* 49 (1941), pp. 122–136.

[BGM09] E. Benhamou, E. Gobet, and M. Miri. "Smart expansion and fast calibration for jump diffusions". In: *Finance Stoch.* 13.4 (2009), pp. 563–589.

[Bol84] E. Bolthausen. "An estimate of the remainder in a combinatorial central limit theorem". In: *Z. Wahrsch. Verw. Gebiete* 66.3 (1984), pp. 379–386.

[CGS11] L. H. Y. Chen, L. Goldstein, and Q.-M. Shao. *Normal Approximation by Stein's Method.* Springer, 2011.

[Cla70] P. Clark. "A subordinated stochastic process model of cotton futures prices". PhD thesis. Harvard University, 1970.

[Cla73] P. Clark. "A subordinated stochastic process model with finite variance for speculative prices". In: *Econometrica* 41.1 (1973), pp. 135–155.

[CT04] R. Cont and P. Tankov. *Financial Modelling with Jump Processes.* Chapman & Hall, 2004.

[Dav03] A. C. Davison. *Statistical Models.* Vol. 11. Cambridge University Press, 2003.

[DM02] R. A. Doney and R. A. Maller. "Stability and attraction to normality for Lévy processes at zero and at infinity". In: *J. Theoret. Probab.* 15.3 (2002), pp. 751–792.

[Doe12] C. Doebler. "On rates of convergence and Berry–Esseen bounds for random sums of centered random variables with finite 3rd moments". arXiv:1212.5401v1. 2012.

[Doe15] C. Doebler. "New Berry–Esseen and Wasserstein bounds in the central limit theorem for non-randomly centered random sums by probabilistic methods". arXiv:1504.05938v3. 2015.

[DS95] F. Delbaen and W. Schachermayer. "The existence of absolutely continuous local martingale measures". In: *Ann. Appl. Probab.* 5.4 (1995), pp. 926–945.

[Dur04] V. Durrleman. "From implied to spot volatilities". PhD thesis. Princeton University, 2004.

[DZ10] A. Dembo and O. Zeitouni. *Large Deviations Techniques and Applications.* Corrected reprint of the second edition. Springer, 2010.

[EK86] S. N. Ethier and T. G. Kurtz. *Markov Processes.* Characterization and Convergence. Wiley, 1986.

[EL14] P. Eichelsbacher and M. Löwe. "90 Jahre Lindeberg-Methode". In: *Mathematische Semesterberichte* 61.1 (2014), pp. 7–34.

[Ess42] C.-G. Esseen. "On the Liapounoff limit of error in the theory of probability". In: *Ark. Mat. Astr. Fys.* 28A.9 (1942), p. 19.

[Fel71] W. Feller. *An Introduction to Probability Theory and its Applications. Vol. II.* Second edition. Wiley, 1971.

[FFF10] J. Feng, M. Forde, and J.-P. Fouque. "Short-maturity asymptotics for a fast mean-reverting Heston stochastic volatility model". In: *SIAM J. Financial Math.* 1 (2010), pp. 126–141.

[Fis11] H. Fischer. *A History of the Central Limit Theorem.* From Classical to Modern Probability Theory. Springer, 2011.

[FJ09] M. Forde and A. Jacquier. "Small-time asymptotics for implied volatility under the Heston model". In: *Int. J. Theor. Appl. Finance* 12.6 (2009), pp. 861–876.

[FLH09] J. E. Figueroa-López and C. Houdré. "Small-time expansions for the transition distributions of Lévy processes". In: *Stochastic Process. Appl.* 119.11 (2009), pp. 3862–3889.

[FPS00] J.-P. Fouque, G. Papanicolaou, and K. R. Sircar. *Derivatives in Financial Markets with Stochastic Volatility.* Cambridge University Press, 2000.

[Gat06] J. Gatheral. *The Volatility Surface, a Practitioner's Guide*. Wiley, 2006.

[Gau14] R. E. Gaunt. "Variance-gamma approximation via Stein's method". In: *Electron. J. Probab.* 19 (2014), pp. 1–33.

[Ger+15] S. Gerhold, M. Kleinert, P. Porkert, and M. Shkolnikov. "Small time central limit theorems for semimartingales with applications". In: *Stochastics* 87.5 (2015), pp. 723–746.

[GGP16] S. Gerhold, I. C. Gülüm, and A. Pinter. "Small-maturity asymptotics for the at-the-money implied volatility slope in Lévy models". In: *Appl. Math. Finance* 23.2 (2016), pp. 135–157.

[GK96] B. V. Gnedenko and V. Y. Korolev. *Random Summation*. Limit Theorems and Applications. CRC Press, 1996.

[GSW10] S. Gerhold, U. Schmock, and R. Warnung. "A generalization of Panjer's recursion and numerically stable risk aggregation". In: *Finance Stoch.* 14.1 (2010), pp. 81–128.

[Gua06] P. Guasoni. "No arbitrage under transaction costs, with fractional Brownian motion and beyond". In: *Math. Finance* 16.3 (2006), pp. 569–582.

[Has80] R. Z. Hasminskii. *Stochastic Stability of Differential Equations*. Sijthoff & Noordhoff, 1980.

[HWY92] S. W. He, J. G. Wang, and J. A. Yan. *Semimartingale Theory and Stochastic Calculus*. CRC Press, 1992.

[Jac07] J. Jacod. "Asymptotic properties of power variations of Lévy processes". In: *ESAIM Probab. Stat.* 11 (2007), pp. 173–196.

[JKB94] N. L. Johnson, S. Kotz, and N. Balakrishnan. *Continuous Univariate Distributions. Vol. 1*. Second edition. Wiley, 1994.

[JS03] J. Jacod and A. N. Shiryaev. *Limit Theorems for Stochastic Processes*. Second edition. Springer, 2003.

[Kal02] O. Kallenberg. *Foundations of Modern Probability*. Second edition. Springer, 2002.

[Kor92] V. Y. Korolev. "Limit distributions of randomly indexed random sequences". In: *Teor. Veroyatnost. i Primenen.* 37.3 (1992), pp. 564–570.

[Lap95] P.-S. Laplace. *Théorie Analytique des Probabilités. Vol. I*. Introduction: Essai philosophique sur les probabilités. Du calcul des fonctions génératrices. Reprint of the 1819 fourth edition (Introduction) and the 1820 third edition (Book I). Éditions Jacques Gabay, Paris, 1995.

[Lee05] R. W. Lee. "Implied volatility: statics, dynamics, and probabilistic interpretation". In: *Recent advances in applied probability*. Springer, 2005, pp. 241–268.

[Lew00] A. L. Lewis. *Option Valuation Under Stochastic Volatility*. With Mathematica code. Finance Press, 2000.

[Lin20] J. W. Lindeberg. "Über das Exponentialgesetz in der Wahrscheinlichkeitsrechnung". In: *Annales academiae scientiarum Fennicae* 16 (1920), pp. 1–23.

[Lin22a] J. W. Lindeberg. "Neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung". In: *Math. Z.* 15.1 (1922), pp. 211–225.

[Lin22b] J. W. Lindeberg. “Über das Gauss’sche Fehlergesetz”. In: *Skandinavisk Aktuarietidskrift* 5 (1922), pp. 217–234.

[Lya01] A. M. Lyapunov. *Nouvelle forme du théorème sur la limite des probabilités.* Vol. 12. Mem. Acad. Sci. St. Petersbourg. 1901.

[Man09] B. Mandelbrot. “New methods of statistical economics revisited: Short versus long tails and Gaussian versus power-law distributions”. In: *Complexity* 14.3 (2009), pp. 55–65.

[Man63] B. Mandelbrot. “The variation of certain speculative prices”. In: *The Journal of Business* 36 (1963), pp. 394–419.

[Man69] B. Mandelbrot. “Long-run linearity, locally Gaussian process, H-spectra and infinite variances”. In: *International Economic Review* 10.1 (1969), pp. 82–111.

[MCC98] D. Madan, P. Carr, and E. Chang. “The variance gamma process and option pricing”. In: *European Finance Review* 2 (1998), pp. 79–105.

[MFE05] A. J. McNeil, R. Frey, and P. Embrechts. *Quantitative Risk Management.* Concepts, Techniques and Tools. Princeton University Press, 2005.

[MR05] M. Musiela and M. Rutkowski. *Martingale Methods in Financial Modelling.* Second edition. Stochastic Modelling and Applied Probability. Springer, 2005.

[MT92] S. P. Meyn and R. L. Tweedie. “Stability of Markovian processes. Criteria for discrete-time chains”. In: *Adv. in Appl. Probab.* 24.3 (1992), pp. 542–574.

[MT93a] S. P. Meyn and R. L. Tweedie. “Stability of Markovian processes. Continuous-time processes and sampled chains”. In: *Adv. in Appl. Probab.* 25.3 (1993), pp. 487–517.

[MT93b] S. P. Meyn and R. L. Tweedie. “Stability of Markovian processes. Foster-Lyapunov criteria for continuous-time processes”. In: *Adv. in Appl. Probab.* 25.3 (1993), pp. 518–548.

[NP12] I. Nourdin and G. Peccati. *Normal Approximations with Malliavin Calculus: From Stein’s Method to Universality.* Cambridge University Press, 2012.

[PR14] J. Pike and H. Ren. “Stein’s method and the Laplace distribution”. In: *ALEA Lat. Am. J. Probab. Math. Stat.* 11.1 (2014), pp. 571–587.

[Pro05] P. E. Protter. *Stochastic Integration and Differential Equations.* Second edition. Version 2.1, Corrected third printing. Springer, 2005.

[Pól20] G. Pólya. “Über den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung und das Momentenproblem”. In: *Math. Z.* 8.3-4 (1920), pp. 171–181.

[Rob48] H. Robbins. “The asymptotic distribution of the sum of a random number of random variables”. In: *Bull. Amer. Math. Soc.* 54 (1948), pp. 1151–1161.

[RR09] M. Roper and M. Rutkowski. “On the relationship between the call price surface and the implied volatility surface close to expiry”. In: *Int. J. Theor. Appl. Finance* 12.4 (2009), pp. 427–441.

[RY99] D. Revuz and M. Yor. *Continuous Martingales and Brownian Motion.* Third edition. Springer, 1999.

[Smi59] D. E. Smith. *A Source Book in Mathematics.* Dover Publications, 1959.

[Ste72] C. Stein. "A bound for the error in the normal approximation to the distribution of a sum of dependent random variables". In: *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Vol. II: Probability theory*. Univ. California Press, 1972, pp. 583–602.

[Sti86] S. M. Stigler. "Laplace's 1774 memoir on inverse probability". In: *Statist. Sci.* 1.3 (1986), pp. 359–378.

[Str11] D. W. Stroock. *Probability Theory*. Second edition. An analytic view. Cambridge University Press, 2011.

[SV06] D. W. Stroock and S. R. S. Varadhan. *Multidimensional Diffusion Processes*. Reprint of the 1997 edition. Springer, 2006.

[Tod12] A. A. Toda. "Weak limit of the geometric sum of independent but not identically distributed random variables". arXiv:1111.1786v2. 2012.

[WY71] S. Watanabe and T. Yamada. "On the uniqueness of solutions of stochastic differential equations". In: *J. Math. Kyoto Univ.* 11 (1971), pp. 553–563.

[Øks10] B. Øksendal. *Stochastic Differential Equations: An Introduction with Applications*. Sixth edition. Springer, 2010.

Piet Nikolaus Porkert

Brünnlbadgasse 4/5
1090 Vienna
Austria

Date of Birth: September 23, 1983
Place of Birth: Vöcklabruck, Austria
Nationality: Austrian

Education

2011	Master's Degree in Mathematics Mathematics in Natural Sciences Vienna University of Technology
2002 – 2003	Military Service, Pionierbataillon Salzburg
2002	Matura, Bundesgymnasium Vöcklabruck

Employment History

2011 – present	Assistant Researcher Research Unit of Financial and Actuarial Mathematics Vienna University of Technology

Teaching Experience

2014 – 2015	Life Insurance Mathematics (Exercise Session) Vienna University of Technology
2013 – 2014	Risk and Ruin Theory (Exercise Session) Vienna University of Technology
2013	Stochastic Analysis 2 (Exercise Session) Vienna University of Technology
2012 – 2013	Stochastic Analysis 1 (Exercise Session) Vienna University of Technology
2011	Stochastic Analysis 2 (Exercise Session) Vienna University of Technology