Markus HOFER

Monte Carlo methods in financial mathematics

PHD THESIS

written to obtain the academic degree of a Doctor of Engineering

Doctoral school "Mathematics and Scientific Computing"

Graz University of Technology

Graz University of Technology

Supervisor:
O.Univ.-Prof. Dr.phil. Robert TICHY

Department of Analysis and Computational Number Theory

Graz, September 2012

Eidesstattliche Erklärung

Ich erkläre an Eides statt, dass ich die vorliegende Arbeit selbständig verfasst, andere als die angegebenen Quellen/Hilfsmittel nicht benutzt, und die den benutzten Quellen wörtlich und inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

Graz, am ....

(Unterschrift)

Statutory Declaration

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources, and that I have explicitely marked all material which has been quotes either literally or by content from the used sources.

....

date (signature)

Preface

Financial mathematics is at least since the publication of the celebrated results of Black and Scholes and Merton in the 1970s a fast growing research field. Researchers and practitioners investigate several different issues, for example risk management, utility theory and asset pricing theory, on which this thesis is mainly focused. A milestone in this context is the famous Black-Scholes model, the structure of which makes it possible to treat many problems in asset pricing theory in a mathematically simple way. In particular, prices of several derivatives can be given in an explicit form. Nevertheless, investigations in the last years have shown that the Black-Scholes model does not replicate all features of the real market properly and therefore several more advanced models were developed with the drawback that closed form solutions are rare.

By the fact that the first fundamental theorem of asset pricing implies that the price of any derivative is given as the discounted expected value of the underlying payoff function under a risk neutral probability measure, we can use numerical integration to find the price of a derivative if an analytic solution is not available. A disadvantage of classic numerical integration methods, like for example the trapezoidal rule, is that their asymptotic convergence rate decreases rapidly when the number of dimensions of the integration domain increases. This is not the case for Monte Carlo (MC) and Quasi Monte Carlo (QMC) integration and since several asset pricing problems involve high-dimensional integration, MC and QMC techniques are frequently applied in practice.

The purpose of this dissertation is to analyse and improve MC and QMC methods and investigate related topics. In the opening chapter we start with a short introduction to the relevant fundamental facts and close with the statement of our new results. In the following three chapters we study the basics of MC and QMC, including criteria for the uniform distribution of special deterministic sequences and probabilistic discrepancy bounds for MC sequences. In Chapter 5 and 6 we analyse two asset pricing techniques which rely on MC and QMC and give theoretical and numerical results which illustrate why these methods are well applicable in practice. The last chapter is dedicated to derivative pricing under an advanced market model where analytic solutions, even for highly complex derivatives, are still available.

I would like to thank some people for their support and help during the last three years. First of all, I thank Astrid for her love and understanding. Furthermore I thank my parents for their

loving support and my brother Hannes for his helpful advice. I am also indebted to Christoph for diverting discussions on mathematics and various other subjects. Moreover I want to thank my colleagues Jochen, Volker, Hermi, Manfred, Christopher and Martin for the nice time and the relaxing coffee breaks.

Special thanks to my supervisor Professor Robert Tichy.

Financial support by the DOC - Doctoral Fellowship Programme of the Austrian Academy of Sciences is gratefully acknowledged.

September 2012,
Markus Hofer

Contents

Chapter 1

Introduction and statement of results

1.1 (Quasi) Monte Carlo integration in option pricing

Starting from the seminal papers of Black and Scholes [14] and Merton [78], financial mathematics has become a very popular and fast growing research field over the last decades. Important topics include among others risk management, utility theory and asset pricing theory, on which this thesis focuses. More precisely we will mainly be concerned with the pricing of special derivatives, so-called options.

Definition 1.1 (Option) *An option is a contract that gives its owner the right, but not the obligation, to buy or to sell an underlying asset at a predefined time and price.*

Before we can define the price of such a derivative, we have to explain some basic properties of financial market models. A detailed introduction into financial market models and asset pricing can be found for example in the book of Cont and Tankov [23].

Definition 1.2 (Stochastic process) *A family of random variables $(X_t)_{t\geq 0}$, indexed by time t, defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is called stochastic process.*

Definition 1.3 (Cadlag function) *A function $f : [0,T] \to \mathbb{R}^d$ is said to be cadlag if it is right-continuous with left limits i.e. for each $t \in [0,T]$ the limits*

$$f(t-) = \lim_{s\to t, s<t} f(s) \qquad f(t+) = \lim_{s\to t, s>t} f(s)$$

exist and $f(t) = f(t+)$.

In the sequel we will use the following financial market model: let $(S_t)_{t\geq 0} = (S_t^1, \ldots, S_t^d)_{t\geq 0}$ be a vector of caglad stochastic processes on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where S_t represents the asset price processes of the underlying assets at time t and let $(S_t^0)_{t\geq 0}$ be defined as a cash account with risk free interest rate r. It is intuitive to assume that the asset price processes are cadlag, since this allows jumps of the asset price processes but these jumps can not be foreseen by an investor.
Note that, due to the evolution of the asset price processes, it follows that at different points in

time there are different amounts of information available. Thus the probability of occurrence of a random event changes with time. To include this flow of information into our market model, we fix $\mathbb{P}$ and introduce the impact of information by conditioning on a so-called filtration $\mathbb{F}$, instead of changing the probability measure $\mathbb{P}$ with time.

Definition 1.4 (Filtration) *A filtration on $(\Omega, \mathcal{F}, \mathbb{P})$ is an increasing family $\mathbb{F} = \{(\mathcal{F}_t)_{t\geq 0} : \forall t \geq s \geq 0, \mathcal{F}_s \subseteq \mathcal{F}_t \subseteq \mathcal{F}\}$ of σ-algebras of $\mathcal{F}$.*

A probability space $(\Omega, \mathcal{F}, \mathbb{P})$ equipped with a filtration is called a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$. We will only use filtrations which satisfy the following three conditions called "usual hypothesis":

- $\mathcal{F}$ is $\mathbb{P}$-complete
- $\mathcal{F}_0$ contains all $\mathbb{P}$-null sets of Ω
- The family $(\mathcal{F}_t)_{t\geq 0}$ is right continuous, $\mathcal{F}_t = \bigcap_{s>t} \mathcal{F}_s$

We define a portfolio $\phi = (\phi^0, \ldots, \phi^d)$ as the amount of cash invested in each asset held by the investor. At time t the value of such a portfolio is

$$V_t(\phi) = \sum_{k=1}^{d} \phi^k S_t^k.$$

Furthermore we define a so-called trading strategy as a dynamic portfolio $(\phi_t)_{t\in[0,T]}$ which is modified at different transaction dates, $T_0 = 0 < T_1 < \ldots < T_{n+1} = T$. In the time interval $[T_i, T_{i+1})$ the current portfolio remains unchanged and it will be denoted by ϕ_i. It is not realistic that an investor knows these trading times in advance, more likely he will change his portfolio, when a specific event has occurred. Hence the transaction dates T_i's become stopping times and the trading strategy becomes a simple predictable process.

Definition 1.5 (Stopping time) *A random time $T > 0$ is called $(\mathbb{F})$-stopping time if $\forall t \geq 0,\ \{T \leq t\} \in \mathcal{F}_t$.*

Definition 1.6 (Adapted process) *A stochastic process $(X_t)_{t\geq 0}$ is said to be adapted to the filtration $\mathbb{F}$ if, for each $t \geq 0$, the value of X_t is revealed at time t, or more precisely, if the random variable X_t is $\mathcal{F}_t$-measurable.*

Definition 1.7 (Predictable process) *The predictable σ-algebra is the σ-algebra $\mathcal{P}$ generated on $[0,T] \times \Omega$ by all adapted left-continuous processes. A mapping $X : [0,T] \times \Omega \to \mathbb{R}^d$ which is measurable with respect to $\mathcal{P}$ is called a predictable process.*

Definition 1.8 (Simple predictable process) *A predictable stochastic process $(\phi)_{t\in[0,T]}$ is called a simple predictable process if it can be represented as*

$$\phi_t = \phi_0 \mathbf{1}_{\{t=0\}} + \sum_{i=0}^{n} \phi_i \mathbf{1}_{\{t\in(T_i, T_{i+1}]\}},$$

where $0 = T_0 < T_1 < \ldots < T_{n+1} = T$ *are stopping times and each* ϕ_i *is a bounded random variable which is* $\mathcal{F}_t$*-measurable.*

Within the class of trading strategies we will be mainly interested in so-called self financing trading strategies, which play an important role in the theory of arbitrage.

Definition 1.9 (Self financing trading strategy) *Let* $(\phi_t)_{t\in[0,T]}$ *denote a* d*-dimensional trading strategy and* $(S_t)_{t\in[0,T]}$ *a* d*-dimensional vector of price processes. If*

$$\phi_t \cdot S_t - \phi_{t-} \cdot S_t = 0,$$

holds for all $t \in (0, T)$*, then the trading strategy is called self financing.*

In other words a trading strategy is called self financing, if after the initial investment there is no money extracted or added.
Now we are able to introduce the crucial concept of arbitrage.

Definition 1.10 (Arbitrage opportunity) *An arbitrage opportunity is a self financing trading strategy with no initial costs, i.e.* $V_0(\phi) = 0$*, where*

$$\mathbb{P}(\forall t \in [0, T] : V_t(\phi) \geq 0) = 1 \text{ and } \mathbb{P}(V_T(\phi) > 0) > 0.$$

Obviously, economically it makes sense to exclude the possibility of arbitrage in asset pricing theory, which is called the no-arbitrage assumption. Note that when trading strategies are not assumed to be predictable, it can easily be shown that the market is not arbitrage free if the asset price processes have jumps.
Having excluded arbitrage opportunities, the next question is how can derivatives be included in the market without introducing arbitrage. An answer to this question is given by the first fundamental theorem, which was first formulated by Harrison and Kreps [50] in a finite state setting and extended to a very general form by Delbaen and Schachermayer [25]. The statement of the theorem is that a financial market model is arbitrage free if and only if there exists a probability measure $\mathbb{Q}$, which is equivalent to the real world probability measure $\mathbb{P}$, such that the discounted asset price processes $(\widetilde{S}^i_t)_{t\geq 0} = (S^i_t/S^0_t)_{t\geq 0},\ 1 \leq i \leq d$ are martingales under $\mathbb{Q}$.

Definition 1.11 (Martingale) *A cadlag process* $(X_t)_{t\geq 0}$ *on* $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ *is called martingale if* $(X_t)_{t\geq 0}$ *is adapted to* $\mathbb{F}$*,* $\mathbb{E}[|X_t|] < \infty$ *for any* $t \geq 0$ *and*

$$\mathbb{E}[X_t|\mathcal{F}_s] = X_s, \quad \forall s > t.$$

Now we will focus on the pricing of derivatives. Options are usually characterised by their so-called payoff function g, which defines the payment to the owner of the option when the contract is executed. The first fundamental theorem of asset pricing implies that the price of an option P is always given as the discounted $\mathbb{Q}$-expected value of its payoff function g, i.e.

$$P = \mathbb{E}_{\mathbb{Q}}[e^{-rT} g(X)],$$

where X denotes a random vector which depends on the vector of price processes $(S_t)_{t\geq 0}$. By transforming the underlying random variables we can write the problem in the following form:

$$P = \mathbb{E}_{\mathbb{Q}}[e^{-rT} f(U)], \tag{1.1}$$

where $U = (U_1, \ldots, U_d)$ is a d-dimensional random vector and all U_i are uniformly distributed on $[0, 1)$.
There are several approaches to calculate P in (1.1) for example by applying analytical methods which use the characteristic function of the asset price process, the numerical solution of a corresponding partial (integro-)differential equation or numerical integration methods. In most cases it depends on the underlying option and the market model whether a specific technique is efficient or not, for more details see e.g. [23].
In this thesis we mainly focus on numerical integration methods. Although classic numerical integration schemes, like for example the trapezoidal rule, have a fast convergence rate for one-dimensional problems, their convergence speed decreases rapidly with increasing d. An alternative is provided by Monte Carlo (MC) integration methods where P is approximated by the Monte Carlo estimator

$$\widehat{P} = \frac{1}{N} e^{-rT} \sum_{n=1}^{N} f(U_n),$$

where $U_1, \ldots, U_N$ are i.i.d. random vectors with the same distribution as U. By the strong law of large numbers it follows that if f is integrable then $\widehat{P}$ is a strongly consistent estimator for P and if f is square-integrable then it follows by the central limit theorem that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} [f(U_i) - \mathbb{E}[f(U)]] \xrightarrow{\mathcal{D}} N(0, \sigma_{MC}^2), \tag{1.2}$$

where $\sigma_{MC}^2 = \mathrm{Var}(f(U))$ and $\xrightarrow{\mathcal{D}}$ denotes the convergence in distribution. In particular this means that the standard deviation of the estimator converges to zero with rate $1/\sqrt{n}$ independently of d.
For applications in financial mathematics the goal is to generate an estimator with a reasonable error in minimal computation time. Since the asymptotic convergence rate is determined by the central limit theorem, there are basically two possibilities to decrease the error of the Monte Carlo estimator: we can increase the number of generated points N or decrease the variance constant σ_{MC}^2 in (1.2). By the fact that the first possibility results in an intense increase of computation time, so-called variance reduction techniques, which aim at decreasing σ_{MC}^2, are very popular. These methods include the use of control variates, antithetic variates, stratified sampling and other concepts. A survey of variance reduction technique can be found in the book of Glassermann [42, Chapter 4].

Nevertheless, the error bounds for the Monte Carlo estimator are of a probabilistic nature and for some applications, especially if we are interested in a worst-case error, deterministic error bounds are required. The idea of Quasi Monte Carlo (QMC) integration is to take an uniformly

distributed but deterministic sequence $(x_n)_{n\geq 1}$ on the d-dimensional unit cube and use

$$\overline{P} = \frac{1}{N}\sum_{n=1}^{N} f(x_n)$$

to approximate P. In the next section we introduce the concept of uniform distribution of a sequence and illustrate that the deterministic bound on the resulting integration error is of order $\mathcal{O}((\log N)^d/N)$ for some special sequences. Although this convergence rate is decreasing for increasing d, it is in any case better than that obtained by Monte Carlo integration.
Note that for practical purposes we are more interested in the error of an estimator which uses exactly N points than in an asymptotic error bound. Thus for high d we need an enormous number of points N to ensure that $(\log N)^d/N \leq 1/\sqrt{N}$. Hence Quasi Monte Carlo integration is an alternative to Monte Carlo methods for moderate d.
Anyway, it is difficult to compare the probabilistic error of MC and the deterministic error of QMC. One approach to avoid this problem is to add the same uniformly distributed random variable to each component of the points of a QMC sequence. This easy example of a so-called randomised Quasi Monte Carlo (RQMC) sequence allows us to compare the resulting probabilistic errors of MC and RQMC. Another idea is to construct a d-dimensional sequence as concatenation of a s-dimensional deterministic QMC sequence and a $(d-s)$-dimensional random MC sequence. In Chapter 5 we show that such a sequence combines positive aspects of MC and QMC. For an overview of different RQMC techniques, see [42, Section 5.4].

1.2 Uniform distribution of sequences and discrepancy

For $x = (x_1, \ldots, x_d) \in [0,1)^d,\ d \geq 1$, let $\mathbf{1}_I(x)$ be the indicator function of the set $I \subseteq [0,1)^d$ and denote by λ_d the d-dimensional Lebesgue measure. For $a = (a_1, \ldots, a_d)$ and $b = (b_1, \ldots, b_d)$ with $0 \leq a_i, b_i \leq 1,\ i = 1, \ldots, d$ we write $a \leq b$ if $a_i \leq b_i$, for $i = 1, \ldots, d$. We call the set of all $x \in [0,1)^d$ with $a \leq x < b$ an axis-parallel box or d-dimensional interval $[a,b)$.

Definition 1.12 (Uniform distribution of sequences) *A sequence $(x_n)_{n\geq 1}$ of vectors in $[0,1)^d$ is said to be uniformly distributed (u.d.) in $[0,1)^d$ if*

$$\lim_{N\to\infty} \frac{\sum_{n=1}^{N} \mathbf{1}_{[a,b)}(x_n)}{N} = \prod_{i=1}^{d}(b_i - a_i),$$

for all d-dimensional intervals $[a,b) \subseteq [0,1)^d$.

The following theorem by Weyl [103] gives a further characterisation of u.d. sequences.

Theorem 1.1 *A sequence $(x_n)_{n\geq 1}$ of vectors in $[0,1)^d$ is said to be u.d. in $[0,1)^d$ if and only if for every continuous complex-valued function f on $[0,1)^d$ the following relation holds:*

$$\lim_{N\to\infty} \frac{\sum_{n=1}^{N} f(x_n)}{N} = \int_{[0,1)^d} f(x)dx.$$

Theorem 1.1 already gives a hint how uniformly distributed sequences can be used for numerical integration. Unfortunately, the uniform distribution property gives no quantitative information on the integration error, therefore we introduce the concept of discrepancy of a sequence. Let $\mathcal{I}$ denote the set of all axis-parallel boxes and let $\mathcal{I}_0$ denote the set of all axis-parallel boxes $[a, b)$ with $a = (0, \ldots, 0)$.

Definition 1.13 (Discrepancy) *Let $x_1, \ldots, x_N$ be a finite sequence in $[0, 1)^d$. The number*

$$D_N = D_N(x_1, \ldots, x_N) = \sup_{I \in \mathcal{I}} \left| \frac{\sum_{n=1}^{N} \mathbf{1}_I(x_n)}{N} - \lambda_d(I) \right|$$

is called discrepancy of the given sequence. Furthermore we denote by

$$D_N^* = D_N^*(x_1, \ldots, x_N) = \sup_{I \in \mathcal{I}_0} \left| \frac{\sum_{n=1}^{N} \mathbf{1}_I(x_n)}{N} - \lambda_d(I) \right|$$

the so-called star-discrepancy of the sequence. For an infinite sequence or a sequence with more than N terms, D_N, D_N^* denote the corresponding quantities of the first N terms of the sequence.

The connection between the discrepancy of a sequence and the uniform distribution property is characterised by the following result of Weyl [103]: a sequence $(x_n)_{n \geq 1}$ is u.d. if and only if $\lim_{N \to N} D_N(x_1, \ldots, x_N) = 0$.
Furthermore it can be shown, see e.g. [66], that

$$D_N^* \leq D_N \leq 2^d D_N^*$$

and thus $\lim_{N \to N} D_N(x_1, \ldots, x_N) = 0$ implies $\lim_{N \to N} D_N^*(x_1, \ldots, x_N) = 0$. Moreover it follows by easy calculations that $1/N \leq D_N \leq 1$ holds for any sequence of N numbers in $[0, 1)^d$.

These bounds on the discrepancy were sharpened in several directions. For d-dimensional infinite sequences the lower bound was improved by Roth [92] to $ND_N^* > C_d^1 (\log N)^{d/2}$ for infinitely many positive integers N, where $C_d^1 > 0$ is an absolute constant only depending on d. Furthermore for a sequence of N points in $[0, 1)^d,\ d \geq 2$ Roth obtained $ND_N^* > C_d^2 (\log N)^{(d-1)/2}$ where C_d^2 is an absolute constant only depending on d.
For case $d = 1$, Schmidt [94] improved the lower bound to $ND_N^* > C \log N$, for infinitely many integers N, where $C > 0$ is an absolute constant. Further improvements for the lower bound of the discrepancy are e.g. due to Beck [11] and Bilyk, Lacey and Vagharshakyan [13], but the precise minimal asymptotic order of the discrepancy is still an open problem. Upper bounds for the discrepancy can be derived by using the inequalities of LeVeque [72] as well as of Erdös and Turán [34, 35].

Let $f(x) = f(x^{(1)}, \ldots, x^{(d)})$ be a function on $[0, 1]^d,\ d \geq 1$. We define a partition P of $[0, 1]^d$ as a set of d finite sequences $\eta_0^{(j)}, \ldots, \eta_{m_j}^{(j)},\ 1 \leq j \leq d$ with $0 = \eta_0^{(j)} \leq \ldots \leq \eta_{m_j}^{(j)} = 1$ and let $\mathcal{P}$ denote the set of all such partitions P. Furthermore we define the operator Δ_j by

$$\Delta_j f(x^{(1)}, \ldots, x^{(j-1)}, \eta_i^{(j)}, x^{(j+1)}, \ldots, x^{(d)})$$

$$=f(x^{(1)},\dots,x^{(j-1)},\eta_{i+1}^{(j)},x^{(j+1)},\dots,x^{(d)})-f(x^{(1)},\dots,x^{(j-1)},\eta_{i}^{(j)},x^{(j+1)},\dots,x^{(d)})$$

for $0 \leq i < m_j$ and we write $\Delta_{j_1,\dots,j_k} = \Delta_{j_1}\cdots\Delta_{j_k}$.

Definition 1.14 (Bounded variation in the sense of Vitali) *Let f be a function on $[0,1]^d$, then the variation in the sense of Vitali is defined as*

$$V^{(d)}(f) = \sup_{P\in\mathcal{P}} \sum_{i_1=0}^{m_1-1}\cdots\sum_{i_d=0}^{m_d-1}\left|\Delta_{1,\dots,d} f(\eta_{i_1}^{(1)},\dots,\eta_{i_d}^{(d)})\right|.$$

A function is said to be of bounded variation in the sense of Vitali if $V^{(d)}(f)$ is finite.

It can easily be seen that if a function f depends on less that d variables than $V^{(d)}(f) = 0$. But since such functions can also have an extremely irregular behavior we extend the notion of variation to the following definition.

Definition 1.15 (Bounded variation in the sense of Hardy and Krause) *Let f be a function on $[0,1]^d$, denote by $V^{(k)}(f;i_1,\dots,i_k)$ the variation of f in the sense of Vitali restricted to the k-dimensional face $F^{(k)}(i_1,\dots,i_k) = \{(u_1,\dots,u_d)\in[0,1]^d : u_j = 1 \text{ for } j\neq i_1,\dots,i_k\}$ and denote by $V(f)$ the variation in the sense of Hardy and Krause. The function f is called of bounded variation in the sense of Hardy and Krause if*

$$V(f) = \sum_{k=1}^{d}\sum_{1\leq i_1<\dots<i_k\leq d} V^{(k)}(f;i_1,\dots,i_k) < \infty.$$

Now we are able to state the Koksma-Hlawka inequality [57]. The one-dimensional analogon is due to Koksma [63].

Theorem 1.2 (Koksma-Hlawka inequality) *Let f be of bounded variation on $[0,1]^d$ in the sense of Hardy and Krause. Then*

$$\left|\frac{1}{N}\sum_{n=1}^{N} f(x_n) - \int_{[0,1]^d} f(x)dx\right| \leq V(f)D_N^*(x_n).$$

Theorem 1.2 gives an deterministic upper bound for the error of Quasi Monte Carlo integration. This bound is a product of the variation of the function f, which is assumed to be finite and the star-discrepancy of $(x_n)_{n\geq 1}$. Therefore it is essential for Quasi Monte Carlo integration to find sequences with a low discrepancy.
A classical example of such a low discrepancy sequence is defined as follows:

Definition 1.16 (Van der Corput sequence) *For $n\in\mathbb{N}_0$ let the function $\phi_b(n)\colon \mathbb{N}_0\to[0,1)$ be given by*

$$\phi_b(n) = \phi_b\left(\sum_{i\geq 0} n_i b^i\right) := \sum_{i\geq 0} n_i b^{-i-1}.$$

The van der Corput sequence in base b is defined as $(\phi_b(n))_{n\geq 0}$.

The van der Corput sequence and its d-dimensional extension the so-called Halton sequence, which is given by $(\phi_{b_1}(n), \ldots, \phi_{b_d}(n))_{n\geq 0}$ for co-prime bases $b_i, 1 \leq i \leq d$, are among those sequences with the best known asymptotic discrepancy of $\mathcal{O}((\log N)^d/N)$, see [49].
For some applications it is possible to use a fixed set of N points with low discrepancy instead of the first N points of an infinite sequence. An example of such a so-called low discrepancy point set is given by the finite Hammersley sequence, defined as
$(n/N, \phi_{b_1}(n), \ldots, \phi_{b_{d-1}}(n))_{n=0,1,\ldots,N-1}$. For such point sets Halton [49] proved that $ND_N^* \leq C_d(\log N)^{d-1}$ holds.
Further examples of low discrepancy sequences and point sets can be found in the book of Niederreiter [81]. More detailed information on uniformly distribution sequences, discrepancy and related topics is provided by Drmota and Tichy [30] and Kuipers and Niederreiter [66].

1.3 Lévy processes

In the classical Black-Scholes (BS) model, for a rigorous definition see [14], the asset price process is modelled as a geometric Brownian motion, which makes it possible that prices of many derivatives are given in an explicit form. However, since several properties of real markets can not be replicated by the BS model, many practitioners and researchers are using more advanced market models. One possible extension is to apply more general exponential Lévy processes as driving stochastic processes which means in the notation of Section 1.1 that $S_t = S_0 e^{X_t}$, where S_0 is a positive d-dimensional vector and $(X_t)_{t\geq 0}$ is a d-dimensional Lévy process. A detailed introduction to the analysis of Lévy processes can be found in the book of Sato [93]. For a survey on Lévy processes in financial modelling, see e.g. the book of Cont and Tankov [23].

Definition 1.17 (Lévy process) *A cadlag stochastic process $(X_t)_{t\geq 0}$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with values in $\mathbb{R}^d$ is called Lévy process, if it possesses the following properties:*

- $X_0 = 0$ *a.s.,*
- *Independent increments: for every increasing sequence of times $t_0, \ldots, t_n$, the random variables $X_{t_0}, X_{t_1} - X_{t_0}, \ldots, X_{t_n} - X_{t_{n-1}}$ are independent.*
- *Stationary increments: the law of $X_{t+h} - X_t$ does not depend on t.*
- *Stochastic continuity: $\forall \epsilon > 0, \lim_{h\to 0} \mathbb{P}(|X_{t+h} - X_t| \geq \epsilon) = 0$.*

Definition 1.18 (Brownian motion) *A stochastic process $(B_t)_{t\geq 0}$ is called Brownian motion on $(\Omega, \mathcal{F}, \mathbb{P})$, if*

- $B_0 = 0$ *a.s.,*
- *for every increasing sequence of times $t_0, \ldots, t_n$, the random variables $B_{t_0}, B_{t_1} - B_{t_0}, \ldots, B_{t_n} - B_{t_{n-1}}$ are independent,*
- *$B_{t+s} - B_t$ has a Gaussian distribution with mean 0 and a positive definite covariance matrix A, $\quad \forall t \in [0, T]$ and*

- *B_t has a.s. continuous sample paths.*

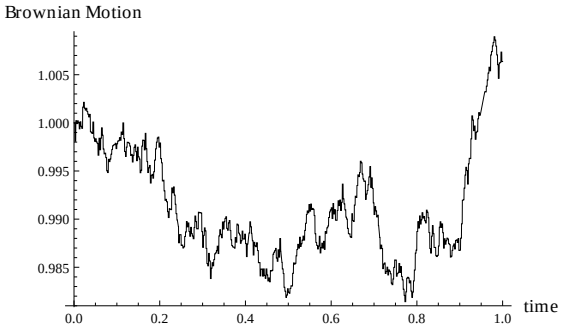

Figure 1.1: Brownian motion

Obviously, the BS model is also an exponential Lévy model where the underlying process is a geometric Brownian motion. Another well known example of a Lévy process is the compound Poisson process.

Definition 1.19 (Poisson process) *Let $(\tau_i)_{i\geq 1}$ be a sequence of i.i.d. exponential random variables with parameter λ and define $T_n = \sum_{i=1}^{n} \tau_i$. The process*

$$N_t = \sum_{n\geq 1} \mathbf{1}_{\{t\geq T_n\}}$$

is called one-dimensional Poisson process with intensity λ.

The Poisson process is a counting process in the following sense: N_T counts the number of random times T_n, which occur in $[0, T]$, where $(T_n - T_{n-1})_{n\geq 1}$ is a sequence of i.i.d. exponential distributed random variables.

Definition 1.20 (Compound Poisson process) *A compound Poisson process on $\mathbb{R}$ with intensity $\lambda > 0$ and jump size distribution f_Y is a stochastic process X_t defined as*

$$X_t = \sum_{i=1}^{N_t} Y_i,$$

where the jump sizes $Y_i \in \mathbb{R}$ are i.i.d. with distribution f_Y and $(N_t)_{t\geq 0}$ is a Poisson process with intensity λ, independent from $(Y_i)_{i\geq 1}$. A compound Poisson process on $\mathbb{R}^d$ is a vector of d compound Poisson processes on $\mathbb{R}$.

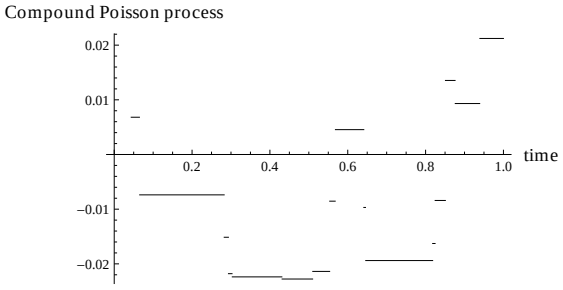

Figure 1.2: Compound Poisson process on $\mathbb{R}$, with intensity = 10 and double-exponentially distributed jump sizes.

Lévy processes which are constructed as the sum of a Brownian motion and a compound Poisson process are frequently used in financial modelling. Because of the advantageous analytical properties of these so-called jump diffusion processes, many option pricing problems can be solved in an effective way.
Proposition 1.1 illustrates the close connection between Lévy processes and infinite divisible distributions, for a proof see [93, Theorem 7.10]. This relationship is a central ingredient to make an explicit calculation of the characteristic function of a Lévy process possible. The closed formula for the characteristic function, which forms the basis of several option pricing techniques, is provided by the Lévy-Khinchin theorem, see [93, Theorem 8.1]. Furthermore the Lévy-Itô theorem [93, Chapter 4] gives detailed information on the structure of Lévy processes.

Definition 1.21 (Infinite divisibility) *A probability distribution F on $\mathbb{R}^d$ is said to be infinitely divisible if for any integer $n \geq 1$, there exist n i.i.d. random variables $Y_1, \ldots, Y_n$ such that $Y_1 + \ldots + Y_n$ has distribution F.*

Proposition 1.1 (Infinite divisible laws and Lévy processes) *Let $(X_t)_{t\geq 0}$ be a Lévy process. Then for every t, the distribution of X_t is infinitely divisible. Conversely, given an infinitely divisible distribution F, there exists a Lévy process (X_t) such that the distribution of X_1 is given by F.*

Definition 1.22 (Poisson random measure) *Let $(\Omega, \mathcal{F}, P)$ be a probability space, $E \subset \mathbb{R}^{d+1}$ and μ a given (positive) Radon measure on a measurable space $(E, \mathcal{E})$. A Poisson random measure on E with intensity measure μ is an integer valued random measure:*

$$\begin{aligned} M : \Omega \times \mathcal{E} &\to \mathbb{N} \\ (\omega, A) &\mapsto M(\omega, A) \end{aligned}$$

such that

1. *For (almost all) $\omega \in \Omega$, $M(\omega, .)$ is an integer-valued Radon measure on E: for any bounded measurable set $A \subset E$, $M(A) < \infty$ is an integer valued random variable.*

2. *For each measurable set $A \subset E$ with $\mu(A) < \infty$, $M(.,A) = M(A)$ is a Poisson random variable with parameter $\mu(A) < \infty$ i.e.*

$$P(M(A) = k) = e^{-\mu(A)} \frac{(\mu(A))^k}{k!}, \quad \forall k \in \mathbb{N}.$$

3. *For disjoint measurable sets $A_1, .., A_n \in \mathcal{E}$, the variables $M(A_1), .., M(A_n)$ are independent.*

We define

$$\widetilde{M}(A) = M(A) - \mu(A)$$

as the compensated Poisson random measure.

Definition 1.23 (Lévy measure) *Let $(X_t)_{t\geq 0}$ be a Lévy process on $\mathbb{R}^d$. The measure ν on $\mathbb{R}^d$ defined by*

$$\nu(A) = \mathbb{E}[\#\{t \in [0,1] : \Delta X_t \neq 0, \Delta X_t \in A\}], \quad A \in \mathcal{B}(\mathbb{R}^d),$$

where $\mathcal{B}(\mathbb{R}^d)$ denotes the Borel σ-algebra, is called Lévy measure of X_t. Note that $\nu(A)$ is exactly the expected number of jumps with sizes in A, per unit time.

Theorem 1.3 (Lévy-Itô decomposition) *Let $(X_t)_{t\geq 0}$ be a Lévy process on $\mathbb{R}^d$ and ν its Lévy measure, given by Definition 1.23. Then the following holds:*

- *ν is a Radon measure on $\mathbb{R}^d \backslash \{0\}$ and verifies:*

$$\int_{|x|\leq 1} |x|^2 \nu(dx) < \infty, \qquad \int_{|x|\geq 1} \nu(dx) < \infty.$$

- *The jump measure of X_t, denoted by J_X, is a Poisson random measure on $[0,\infty) \times \mathbb{R}^d$ with intensity measure $\nu(dx)dt$.*

- *There exists a vector γ and a d-dimensional Brownian motion $(B_t)_{t\geq 0}$ with covariance matrix A such that*

$$\begin{aligned}
X_t &= X_t^1 + X_t^2 + + X_t^3 + X_t^4, \quad \textit{where} \\
X_t^1 &= \gamma t, \\
X_t^2 &= B_t, \\
X_t^3 &= \int_{|x|\geq 1,\ s\in[0,t]} x J_X(ds \times dx) \textit{ and} \\
X_t^4 &= \lim_{\epsilon \searrow 0} \int_{\epsilon \leq |x| \leq 1,\ s\in[0,t]} x(J_X(ds \times dx) - \nu(dx)ds) \\
&= \lim_{\epsilon \searrow 0} \int_{\epsilon \leq |x| \leq 1,\ s\in[0,t]} x \widetilde{J}_X(ds \times dx).
\end{aligned}$$

The terms $X_t^1, X_t^2, X_t^3, X_t^4$ are independent, the convergence in the last term is almost sure and uniform in t on $[0,T]$ and $\widetilde{J}_X$ is the compensated Poisson random measure of J_X.

Basically the Lévy-Itô decomposition states that every Lévy process is a sum of three elementary processes and the limit of the sum of compound Poisson processes. Furthermore the theorem says that only three parameters determine a Lévy process uniquely: the vector γ, the covariance matrix A and the Lévy measure ν. The triplet (γ, A, ν) is called characteristic triplet of a Lévy process.

Definition 1.24 (Characteristic function) *The characteristic function $\phi_X : \mathbb{R}^d \to \mathbb{R}$ of an $\mathbb{R}^d$-valued random variable X is defined by*

$$\Phi_X(z) = \mathbb{E}[\exp(izX)] = \int_{\mathbb{R}^d} e^{izx}\mu_x(dx), \quad \forall z \in \mathbb{R}^d.$$

Theorem 1.4 (Lévy-Khinchin representation) *Let $(X_t)_{t\geq 0}$ be a Lévy process on $\mathbb{R}^d$ with characteristic triplet (γ, A, ν). Then*

$$\mathbb{E}[e^{iz^T X_t}] = e^{t\psi(z)}, \quad z \in \mathbb{R}^d,$$

where the so-called characteristic exponent $\psi(z)$ is given by

$$\psi(z) = -\frac{1}{2}z^T Az + i\gamma^T z + \int_{\mathbb{R}^d}\Big(e^{iz^T x} - 1 - iz^T x 1_{\{|x|\leq 1\}}\nu(dx)\Big).$$

The idea of many option pricing techniques is to calculate the Laplace or Fourier transform of the option price by using the characteristic function of the involved random variables. The inverse transformation can be done very efficiently in many cases, for example by using the Fast Fourier Transform method, see e.g. [21] or [65]. Since the characteristic function of a Lévy process is given by Theorem 1.4, we can apply such Laplace or Fourier transform methods when the underlying process is an exponential Lévy process.
Apart from jump diffusion processes, the so-called pure jump processes are frequently used in financial modelling. Such processes have no diffusion part, but an infinite jump activity. They can be constructed as a so-called subordinated Lévy process, which is for example a Brownian motion where the time progression is not linear but modelled by a so-called subordinator. Theorem 1.5 characterises the characteristic triplet of subordinated Lévy processes, for a proof see e.g. [93, Theorem 30.1].

Definition 1.25 (Subordinator) *A Lévy process $(X_t)_{t\geq 0}$ is called subordinator if its paths are a.s. non-decreasing i.e.*

$$t \geq s \Rightarrow X_t \geq X_s \text{ a.s.}$$

Theorem 1.5 (Subordination of a Lévy process) *Fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let $(X_t)_{t\geq 0}$ be a Lévy process on $\mathbb{R}^d$ with characteristic exponent $\Psi(u)$ and characteristic triplet (γ, A, ν) and let $(S_t)_{t\geq 0}$ be a subordinator with characteristic exponent $l(u)$ and characteristic triplet*

$(b, 0, \rho)$. *Then the process* $(Y_t)_{t\geq 0}$ *defined for each* $\omega \in \Omega$ *by* $Y(t,\omega) = X(S(t,\omega),\omega)$ *is a Lévy process and its characteristic function is given by*

$$\mathbb{E}[e^{iuY_t}] = e^{tl(\Psi(u))},$$

i.e. the characteristic exponent of Y_t *is obtained by composition of the Laplace exponent of* S_t *with the characteristic exponent of* X_t. *The characteristic triplet* (γ^Y, A^Y, ν^Y) *of* Y_t *is given by*

$$\begin{aligned} A^Y &= bA \\ \nu^Y(B) &= b\nu(B) + \int_0^\infty p_s^X(B)\rho(ds), \quad \forall B \in \mathcal{B}(\mathbb{R}^d), \\ \gamma^Y &= b\gamma + \int_0^\infty \rho(ds) \int_{\{|x|\leq 1\}} x p_s^X(dx), \end{aligned}$$

where p_t^X *is the probability distribution of* X_t. *The process* $(Y_t)_{t\geq 0}$ *is said to be subordinate to the process* $(X_t)_{t\geq 0}$.

In the numerical analysis of option pricing problems in the Chapters 6 and 7, we will use two pure jump processes, which are frequently applied for financial modelling: the variance gamma process, which was first introduced by Madan and Seneta [75] and the normal inverse Gaussian process, which is due to Barndorff-Nielsen [10]. An overview of subordinated Lévy processes in financial mathematics can be found in [23].

In the last part of this thesis we consider the pricing of options, where the payoff function depends on the value of the supremum or the infimum of the asset price process during the life time of the option. Hence to apply a pricing technique which uses the Laplace transform we need information on the characteristic function of the extremal processes of the underlying Lévy process.

Theorem 1.6 (Wiener-Hopf factorisation) *Let* $(X_t)_{t\geq 0}$ *be a Lévy process in* $\mathbb{R}$ *and* $(\overline{X}_t)_{t\geq 0}$ *and* $(\underline{X}_t)_{t\geq 0}$ *its supremum and infimum process, respectively. Furthermore, let* θ *be an exponentially distributed random variable with parameter* q. *Then the characteristic function of* $(X_t)_{t\geq 0}$ *at the random time* θ *can be factorised as*

$$\mathbb{E}[e^{izX_\theta}] = \mathbb{E}[e^{iz\overline{X}_\theta}]\,\mathbb{E}[e^{iz\underline{X}_\theta}], \quad \forall z \in \mathbb{R},$$

or equivalently,

$$\frac{q}{q - \log(\phi_X(z))} = \phi_q^+(z)\phi_q^-(z), \quad \forall z \in \mathbb{R},$$

where $\phi(z)$ *denotes the characteristic function of* X_1, $\phi_q^+(z) = \mathbb{E}[e^{iz\overline{X}_\theta}]$ *and* $\phi_q^-(z) = \mathbb{E}[e^{iz\underline{X}_\theta}]$.

In general the calculation of the so-called Wiener-Hopf factors $\phi_q^+(z)$ and $\phi_q^-(z)$ is very complicated and involves a multi-dimensional numerical integration. Nevertheless there are a few classes of Lévy processes for which $\phi_q^+(z)$ and $\phi_q^-(z)$ are given explicitly, for example the Brownian motion or jump diffusions where the jump sizes are exponentially distributed. A detailed discussion of the Wiener-Hopf factorisation including the proof of Theorem 1.6 can be found in [93, Chapter 9]. The Wiener-Hopf factorisation with a view to option pricing is discussed in the book of Kyprianou [67].

1.4 Statement of new results

The structure of the remainder of this thesis is as follows: in the present section we give an overview of our new results. Chapters 2-7 correspond to scientific articles which are submitted for publication or already published, therefore every chapter contains the necessary information and the references to understand its contents. Chapters 2-5 are results of a joint work with Christoph Aistleitner, Chapter 6 is collaboration with Christoph Aistleitner and Robert Tichy and Chapter 7 originates in a cooperation with Philipp Mayer.

The second chapter deals with the distribution properties of the so-called Kakutani's sequence of partitions: For $\alpha \in (0,1)$ and $\pi = \{[t_{i-1}, t_i] : 1 \leq i \leq k\}$ is any partition of $[0,1]$, we denote by $\alpha\pi$ the α-refinement of π which is obtained by subdividing all intervals of π having maximal length in two parts, proportional to α and $1-\alpha$, respectively. Kakutani's sequence of partitions $(\alpha^n\omega)_{n\in\mathbb{N}}$ is obtained as the successive α-refinement of the trivial partition $\omega = \{[0,1]\}$.
Furthermore let $(\pi_n)_{n\in\mathbb{N}}$ be a sequence of partitions of $[0,1]$, with

$$\pi_n = \{[t_{i-1}^n, t_i^n] : 1 \leq i \leq k(n)\}.$$

Then π_n is uniformly distributed (u.d.), if for any continuous function f on $[0,1]$

$$\lim_{n\to\infty} \frac{1}{k(n)} \sum_{i=1}^{k(n)} f(t_i^n) = \int_0^1 f(t)dt.$$

Kakutani [61] shows that $(\alpha^n\omega)_{n\in\mathbb{N}}$ is u.d. for all $\alpha \in (0,1)$. Nevertheless, one can formulate simple examples where $(\alpha^n\pi)_{n\in\mathbb{N}}$ is not u.d. when π is a non-trivial finite partition of $[0,1]$.
We consider a more general splitting rule under which the refinement is defined as subdividing all intervals of maximal length positively homothetically to ρ, where ρ is a finite partition of $[0,1]$. We give conditions on ρ and π, which guarantee that $(\rho^n\pi)_{n\in\mathbb{N}}$ is u.d. and calculate the maximal and minimal limit measures, in the case when $(\rho^n\pi)_{n\in\mathbb{N}}$ is not u.d. The results in Chapter 2 will be published in Annali di Matematica Pura ed Applicata [4].

In the third chapter we focus on the distribution of sequences of vectors of subsequent elements of the van der Corput sequence $(\phi_b(n))_{n\geq 1}$, see Definition 1.16. In [37], Fialová and Strauch calculate the limit distribution of $(\phi_b(n), \phi_b(n+1))_{n\geq 1}$ by using analytic methods. In our approach we use the close connection between the van der Corput sequence and the so-called van Neumann-Kakutani transformation $T_b\colon [0,1) \to [0,1)$ given by

$$T_b(x) = x - 1 + \frac{1}{b^k} + \frac{1}{b^{k+1}},$$

where $k \in \mathbb{N}$ is such that $x \in \left[1 - \frac{1}{b^k}, 1 - \frac{1}{b^{k+1}}\right]$. One can easily show that the orbit of zero under the ergodic van Neumann-Kakutani transformation T_b is exactly the van der Corput sequence in base b, i.e.

$$(T_b^n 0)_{n\geq 0} = (\phi_b(n))_{n\geq 0}, \quad b \geq 2.$$

By using methods from ergodic theory and the properties of the van Neumann-Kakutani transformation we extend the results of Fialová and Strauch to the d-dimensional case, which means that we calculate the limit distribution of $(\phi_b(n), \dots, \phi_b(n+d-1))_{n\geq 1}$. These results are accepted for publication in Uniform Distribution Theory [2].

We define the so-called inverse of the discrepancy $n^*(d,\varepsilon)$ as the smallest possible cardinality of a point set in $[0,1]^d$ having discrepancy bounded by ε. By a profound result of Heinrich, Novak, Wasilkowski and Woźniakowski [52] the inverse of the star-discrepancy $n^*(d,\varepsilon)$ satisfies the upper bound

$$n^*(d,\varepsilon) \leq c^1_{abs} d \varepsilon^{-2},$$

and Hinrichs [55] proved the lower bound

$$n^*(d,\varepsilon) \geq c^2_{abs} d \varepsilon^{-1},$$

where c^1_{abs}, c^2_{abs} are absolute constants. The upper bound is equivalent to the fact that for any N and d there exists a set of N points in $[0,1]^d$ whose star-discrepancy is bounded by $c_{abs} d^{1/2} N^{-1/2}$. The proof is based on the observation that a random point set satisfies the desired discrepancy bound with positive probability.
In Chapter 4 we prove a version of this result, which makes it applicable for computational purposes: for any given number $q \in (0,1)$ there exists an (explicitly stated) number $c(q)$ such that the star-discrepancy of a random set of N points in $[0,1]^d$ is bounded by $c(q) d^{1/2} N^{-1/2}$ with probability at least q, uniformly in N and d. More precisely we prove the following:
For any $s \geq 1,\ N \geq 1$ and $q \in (0,1)$ a randomly generated s-dimensional point set $(z_1, \dots, z_N)$ satisfies

$$D^*_N(z_1, \dots, z_N) \leq 5.70 \sqrt{4.90 + \frac{\log\left((1-q)^{-1}\right)}{d}} \frac{\sqrt{d}}{\sqrt{N}}$$

with probability at least q.
The most interesting fact is that at the moment there exists no construction of a deterministic sequence which satisfies such discrepancy bounds, whereas the above theorem states that a random Monte Carlo sequence has a discrepancy of this form with high probability.

Chapter 5 is dedicated to randomized Quasi Monte Carlo methods, where a d-dimensional sequence is constructed as concatenation of a s-dimensional deterministic sequence and a $d-s$-dimensional random point set. Such constructions, which have been investigated e.g. by Spanier [96], Ökten [82, 83] and Roşca [91], are useful since in moderate dimensions the QMC method typically yields better results, but its performance significantly falls off in quality if the dimension increases. Ökten, Tuffin and Burago [85] proved probabilistic asymptotic bounds for the discrepancy of mixed sequences, which were refined by Gnewuch [45]. Furthermore, Ökten et al. [85] show in numerical examples that this method is very effective for problems in option pricing.
In Chapter 5 we use an interval partitioning technique to obtain improved probabilistic bounds for the discrepancy of mixed sequences. By comparing them with lower bounds we show that

our results are almost optimal. A paper which includes the results of this chapter will be published in Monte Carlo Methods and Applications [3].

We mentioned in Section 1.1 that the convergence of the standard Monte Carlo estimator can be improved by so-called variance reduction techniques. An example of such a method is the so-called stratified sampling method. The idea behind stratified sampling is to generate points conditional uniformly distributed on the elements of a partition of $[0,1)^d$, the so-called strata, instead of generating points uniformly distributed on the whole d-dimensional unit cube. If the strata are chosen properly this technique leads to a reduce of σ^2_{MC} in (1.2).
A modification of stratified sampling is given by the so-called Latin hypercube sampling (LHS) technique, where we divide $[0,1)^d$ into a partition of N^d d-dimensional cubes with the same volume. Exactly one random point is generated uniformly distributed in N of these cubes, where the N cubes are chosen randomly but in accordance with a certain rule. A central limit theorem for the resulting estimator was proved by Owen [86] and Stein [97] shows that the variance of the LHS estimator is never greater than the variance of the standard Monte Carlo estimator, provided the number of sample points is sufficiently large.
In many applications, especially in financial mathematics, one faces a situation where the components of the underlying random vector $U = (U_1, \ldots, U_d)$ are dependent. Usually such dependence structures are modelled by a so-called copula distribution C, which is a distribution function on $[0,1]^d$, where all one-dimensional marginals are uniformly distributed on $[0,1]$. Latin hypercube sampling with dependence (LHSD) is an extension of LHS where the cubes in which points are generated are chosen according to the rank statistics of samples from the copula distribution C. This has the effect that the empirical distribution of the resulting points converges to C, which is in general not the case for LHS.
Packham and Schmidt [88] prove that the d-dimensional LHSD estimator is consistent and that a central limit theorem holds for the two-dimensional LHSD estimator. Furthermore, for the two-dimensional case, they give conditions on the copula C and on the integrand function f which ensure that the variance of the LHSD estimator is never greater than the variance of the standard Monte Carlo estimator. In Chapter 6 we extend these results to the d-dimensional case. The results in this chapter are accepted for publication in the International Journal of Theoretical and Applied Finance [5]

In the last chapter we focus on option pricing in one-dimensional exponential Lévy models, in particular, we consider the pricing of so-called lookback options, where the payoff function is defined for example as $\max(\max_{0\leq t\leq T} S_t - K, 0)$, where $(S_t)_{t\geq 0}$ denotes the asset price process, T denotes the maturity and K denotes the strike price. In our setting the asset price process $(S_t)_{t\geq 0}$ is given by $S_t = S_0 e^{X_t}$, where $S_0 > 0$ and $(X_t)_{t\geq 0}$ is a Lévy process.
Obviously, to price lookback options we need information on the maximum process of $(S_t)_{t\geq 0}$. As mentioned in Section 1.3, we can use the Wiener-Hopf factorisation to obtain the characteristic function of the maximum process, but for general Lévy processes this requires a numerical multi-dimensional integration. Nevertheless there are a few classes of Lévy processes for which the Wiener-Hopf factorisation can be done in closed form, for example when the underlying Lévy process is the sum of a Brownian motion and a compound Poisson process with expo-

nentially distributed jumps. Kou [64] proposed a financial market model (typically called Kou model), in which the logarithmic asset price process is described by a jump diffusion with two-sided exponential jumps, and showed that in this setting the Laplace transform of several exotic derivatives, including lookback options, can be derived analytically (see [65]).
Jeannin and Pistorius [59] consider option pricing under the hyper-exponential jump diffusion (HEJD) model which is defined as jump-diffusion with double sided hyper-exponentially distributed jumps sizes. They present formulae for Laplace transformed price of barrier and digital options together with sensitivities of these prices. Since the class of HEJD lies dense in the class of all Lévy process these results are also useful to approximate prices when der underlying Lévy process is for example a pure jump process like the variance gamma or the normal inverse Gaussian process.
In Chapter 7 we present formulae for the Laplace transformed price of different types of lookback options under the HEJD model. Furthermore, we give formulae for the sensitivities of option prices, which are important for hedging. Moreover we show in numerical examples that the error introduced by the numerical inverse Laplace transformation is insignificant and that prices of lookback options under the normal inverse Gaussian model can be approximated efficiently with this method.

Part I

Uniform distribution of sequences

Chapter 2

Uniform Distribution of generalized Kakutani's sequences of partitions

2.1 Introduction

In this paper we study a generalization of the Kakutani splitting procedure, which was originally introduced in [61].

Definition 2.1 (Kakutani splitting procedure) *If $\alpha \in (0,1)$ and $\pi = \{[t_{i-1}, t_i] : 1 \leq i \leq k\}$ is any partition of $[0,1]$, then $\alpha\pi$ denotes its α -refinement which is obtained by subdividing all intervals of π having maximal length in two parts, proportional to α and $1-\alpha$, respectively. The so-called Kakutani's sequence of partitions $(\alpha^n\omega)_{n\in\mathbb{N}}$ is obtained as the successive α -refinement of the trivial partition $\omega = \{[0,1]\}$.*

Definition 2.2 (Uniform distribution of sequences of partitions) *Let $(\pi_n)_{n\in\mathbb{N}}$ be a sequence of partitions of $[0,1]$, with*

$$\pi_n = \{[t_{i-1}^n, t_i^n] : 1 \leq i \leq k(n)\}.$$

Then π_n is uniformly distributed (u.d.), if for any continuous function f on $[0,1]$

$$\lim_{n\to\infty} \frac{1}{k(n)} \sum_{i=1}^{k(n)} f(t_i^n) = \int_0^1 f(t)dt. \tag{2.1}$$

Remark 2.1 *For a sequence of partitions $(\pi_n)_{n\in\mathbb{N}}$ we define the associated sequence of measures $(\mu_n)_{n\in\mathbb{N}}$ by*

$$\mu_n = \frac{1}{k(n)} \sum_{i=1}^{k(n)} \delta_{t_i^n},$$

where δ_t denotes the Dirac measure concentrated at t. Weak convergence of $(\mu_n)_{n\in\mathbb{N}}$ to the Lebesgue measure on $[0,1]$ is equivalent to condition (2.1)*. In other words, a sequence of parti-*

tions is u.d. if and only if for every interval $[a, b] \subset [0, 1]$

$$\lim_{n\to\infty} \frac{\sum_{i=1}^{k(n)} \mathbf{1}_{[a,b]}(t_i^n)}{k(n)} = b - a.$$

Kakutani [61] proved that for any $\alpha \in (0, 1)$ the sequence of partitions $(\alpha^n\omega)_{n\in\mathbb{N}}$ is uniformly distributed. The properties of the sequence $(\alpha^n\omega)_{n\in\mathbb{N}}$ and related problems have been investigated by many authors. For example, see [15] and [100] for a modification of $(\alpha^n\omega)_{n\in\mathbb{N}}$ where the intervals of maximal length are split at a random position. Carbone and Volčič [19] generalized the splitting procedure for sequences of partitions of $[0, 1]^d$, $d \geq 2$, and derived a generalization of Kakutani's result in higher dimensions. Recently, the following further modification of Kakutani's splitting procedure was presented by Volčič [101].

Definition 2.3 (ρ - refinement) *Let ρ denote a non-trivial finite partition of $[0, 1]$. Then the ρ -refinement of a partition π of $[0, 1]$, denoted by $\rho\pi$, is given by subdividing all intervals of maximal length positively homothetically to ρ.*

Volčič [101] proved, by using arguments from ergodic theory, that the sequence $(\rho^n\omega)_{n\in\mathbb{N}}$ is u.d. for every finite partition ρ. Furthermore, he investigated the behavior of associated uniformly distributed sequences of points. The discrepancy of sequences of partitions constructed as ρ -refinements of ω is discussed in Carbone [18] and Drmota and Infusino [29]. The results of Drmota and Infusino are based on the analysis of a special tree evolution process, namely the Khodak algorithm [62], where the generation of nodes has a similar behavior as the splitting of intervals in the Kakutani splitting sequence.
So far results on the uniform distribution of sequences of partitions were only available in the case when the starting partition π is the trivial partition ω. A simple example shows that there exist starting partitions π for which the sequence $(\rho^n\pi)_{n\in\mathbb{N}}$ is not uniformly distributed. Consider $\pi = \left\{\left[0, \frac{2}{5}\right], \left[\frac{2}{5}, 1\right]\right\}$ and $\rho = \left\{\left[0, \frac{1}{2}\right], \left[\frac{1}{2}, 1\right]\right\}$. In this case the splitting procedure operates alternating on $\left[0, \frac{2}{5}\right]$ and $\left[\frac{2}{5}, 1\right]$ and hence the sequences of associated measures corresponding to the subsequences $(\rho^{2n}\pi)_{n\in\mathbb{N}}$ and $(\rho^{2n+1}\pi)_{n\in\mathbb{N}}$ converge to different measures. Volčič [101] formulated the problem in the following form:

> It is worth noticing that it is necessary to put some restriction on the partition π (even in the simplest case of the Kakutani splitting procedure) if we hope for uniform distribution of $(\rho^n\pi)_{n\in\mathbb{N}}$. It would be interesting to find significant sufficient conditions on π in order to obtain the uniform distribution of $(\rho^n\pi)_{n\in\mathbb{N}}$ even for the case of Kakutani's splitting procedure.

The purpose of the present paper is to present a full solution of this problem.

2.2 The uniform distribution of generalized Kakutani's sequences of partitions

In the sequel we consider a partition ρ of $[0, 1]$ consisting of $m \geq 2$ intervals of lengths $p_1, \ldots, p_m$, and a starting partition π of $[0, 1]$ consisting of $l \geq 2$ intervals of lengths $\alpha_1, \ldots, \alpha_l$.

In the sequel let H denote the entropy of the probability distribution $p_1, \ldots, p_m$, which is defined as

$$\mathrm{H} = p_1 \log\left(\frac{1}{p_1}\right) + \ldots + p_m \log\left(\frac{1}{p_m}\right).$$

Definition 2.4 (Rationally related) *The numbers* $\log\left(\frac{1}{p_1}\right), \ldots, \log\left(\frac{1}{p_m}\right)$ *are called rationally related if there exists a positive real number* Λ *such that*

$$\log\left(\frac{1}{p_j}\right) = \nu_j \Lambda, \quad \nu_j \in \mathbb{Z}, j = 1, \ldots, m.$$

Without loss of generality we choose Λ *as large as possible, which is equivalent to assuming* $\gcd(\nu_1, \ldots, \nu_m) = 1$. *If the numbers* $\log\left(\frac{1}{p_1}\right), \ldots, \log\left(\frac{1}{p_m}\right)$ *are not rationally related, they are called irrationally related.*

Remark 2.2 *Note that the numbers* $\log\left(\frac{1}{p_1}\right), \ldots, \log\left(\frac{1}{p_m}\right)$ *are rationally related if and only if all fractions*

$$\frac{\log p_i}{\log p_j}, \quad i, j = 1, \ldots, m,$$

are rational.

For a fixed real number $\epsilon \in (0, p_{\min})$, where $p_{\min} = \min\{p_1, \ldots, p_m\}$, let $\mathcal{I}_\epsilon$ denote the set of all intervals that appear in the sequence $(\rho^n \omega)_{n\in\mathbb{N}}$ and have length greater than or equal to ϵ. Let $\mathcal{E}_\epsilon$ be the set of intervals which are generated by splitting an interval in $\mathcal{I}_\epsilon$ and which have length l satisfying $p_{\min}\epsilon \leq l < \epsilon$. Denote by $M_\epsilon = |\mathcal{E}_\epsilon|$ the cardinality of $\mathcal{E}_\epsilon$. Note that the set $\mathcal{E}_\epsilon$ changes only for certain values of ϵ, more precisely when ϵ equals the length of at least one interval appearing in $(\rho^n \omega)_{n\in\mathbb{N}}$.

We will use the following result from [29].

Lemma 2.1 *Let* M_ϵ *be defined as above. Then*

1. *if* $\log\left(\frac{1}{p_1}\right), \ldots, \log\left(\frac{1}{p_m}\right)$ *are rationally related, let* Λ *be the largest real number for which* $\log\left(\frac{1}{p_j}\right)$ *is an integer multiple of* Λ, *for* $j = 1, \ldots, m$. *Then there exist a real number* $\eta > 0$ *and an integer* $d \geq 0$ *such that*

$$M_\epsilon = \frac{m-1}{\epsilon H} Q_1\left(\log\left(\frac{1}{\epsilon}\right)\right) + \mathcal{O}\left((\log(\epsilon))^d \epsilon^{-(1-\eta)}\right), \tag{2.2}$$

where

$$Q_1(x) = \frac{\Lambda}{1 - e^{-\Lambda}} e^{-\Lambda\{\frac{x}{\Lambda}\}}$$

and $\{y\}$ *denotes the fractional part of* y.

2. *If* $\log\left(\frac{1}{p_1}\right), \ldots, \log\left(\frac{1}{p_m}\right)$ *are irrationally related, then*

$$M_\epsilon = \frac{m-1}{\epsilon H} + o\left(\frac{1}{\epsilon}\right). \tag{2.3}$$

The following theorem gives sufficient and necessary conditions on π and ρ under which $(\rho^n\pi)_{n\in\mathbb{N}}$ is uniformly distributed.

Theorem 2.1 *Let $\alpha_j, j = 1, \ldots, l$ denote the lengths of the intervals of the starting partition π. Then the sequence $(\rho^n\pi)_{n\in\mathbb{N}}$ is uniformly distributed if and only if one of the following conditions is satisfied:*

(I) the real numbers $\log\left(\frac{1}{p_1}\right), \ldots, \log\left(\frac{1}{p_m}\right)$ *are irrationally related or*

(II) the real numbers $\log\left(\frac{1}{p_1}\right), \ldots, \log\left(\frac{1}{p_m}\right)$ *are rationally related with parameter Λ and the lengths of the intervals of π can be written in the form*

$$\alpha_i = ce^{v_i\Lambda}, \quad c \in \mathbb{R}^+, v_i \in \mathbb{Z}, \tag{2.4}$$

for $i = 1, \ldots, l$.

Remark 2.3 *Condition (II) includes the special case that the starting partition π is a partition consisting of intervals having the same length, and in particular the case when the starting partition is the trivial partition ω.*

For illustration, the next corollary characterizes the starting partitions π for which the original Kakutani's sequence of partitions is u.d. .

Corollary 2.1 *Let the sequence of partitions $(\rho^n\pi)_{n\in\mathbb{N}}$ be defined as a ρ -refinement with $\rho = [[0,p],[p,1]]$ and $\pi = [[0,\alpha],[\alpha,1]]$. Then $(\rho^n\pi)_{n\in\mathbb{N}}$ is u.d. if and only if one of the following conditions is satisfied:*

(i) $\log(p)/\log(1-p)$ *is irrational, or*

(ii) $\log\left(\frac{1}{p}\right)$ *and* $\log\left(\frac{1}{1-p}\right)$ *are rationally related with parameter Λ and $\alpha = \frac{1}{e^{k\Lambda}+1}$ for $k \in \mathbb{Z}$.*

The next theorem describes the asymptotic behavior of the distribution of $(\rho^n\pi)_{n\in\mathbb{N}}$ for those cases which are not covered by Theorem 2.1.

Theorem 2.2 *Assume that neither condition (I) nor condition (II) of Theorem 2.1 is satisfied. Then for any interval $A = [a,b] \subset [0,1]$ which is completely contained in the i-th interval of the starting partition π for some $i, 1 \le i \le l$, we have*

$$\limsup_{n\to\infty} \frac{\sum_{j=1}^{k(n)} \mathbf{1}_{[a,b]}(t_j^n)}{k(n)} = c_1(b-a),$$

$$\liminf_{n\to\infty} \frac{\sum_{j=1}^{k(n)} \mathbf{1}_{[a,b]}(t_j^n)}{k(n)} = c_2(b-a),$$

where

$$c_1 = \left(\sum_{j=1}^{l} \alpha_j \exp\left(-\Lambda \left\{ \frac{\log(\alpha_j) - \log(\alpha_i)}{\Lambda} \right\} \right) \right)^{-1} > 1,$$

$$c_2 = \left(\sum_{j=1}^{l} \alpha_j \exp\left(\Lambda \left\{ \frac{\log(\alpha_i) - \log(\alpha_j)}{\Lambda} \right\} \right) \right)^{-1} < 1$$

are constants depending on i*.*

Remark 2.4 *Observe that only if the conditions (I) and (II) fail to hold,* c_1 *is strictly larger and* c_2 *is strictly smaller than* 1 *and the sequence is not u.d. (cf. Remark 2.1).*

At the end of the introduction we mentioned the example $\pi = \left\{ \left[0, \frac{2}{5}\right], \left[\frac{2}{5}, 1\right] \right\}$ and $\rho = \left\{ \left[0, \frac{1}{2}\right], \left[\frac{1}{2}, 1\right] \right\}$. In this case the theorem indicates that the maximal and minimal asymptotic measure of $[0, \frac{2}{5}]$ is $\frac{1}{2}$ and $\frac{1}{3}$, respectively, and accordingly the maximal and minimal measure of $[\frac{2}{5}, 1]$ is $\frac{2}{3}$ and $\frac{1}{2}$, respectively.

2.3 Proofs

Proof of Theorem 2.1:
Proof:
Denote the l intervals of π by $I_i, i = 1, \ldots, l$. Then I_i has length $\alpha_i, i = 1, \ldots, l$. To show that $(\rho^n \pi)_{n\in\mathbb{N}}$ is uniformly distributed it is sufficient to prove that the relative number of intervals of $(\rho^n \pi)_{n\in\mathbb{N}}$ in I_i converges to α_i, for $i = 1, \ldots, l$, since by [101, Theorem 2.7] the sequences of partitions within the intervals I_i are u.d.
Assume that (I) holds and let $0 < \epsilon \leq (\min_{1\leq j\leq l} \alpha_j)(\min_{1\leq i\leq m} p_i)$. Let $h_\epsilon \in \mathbb{N}$ be the smallest number for which $\rho^{h_\epsilon}\pi$ contains only intervals of length $< \epsilon$. Then the set $\{h_\epsilon : 0 < \epsilon \leq (\min_{1\leq j\leq l} \alpha_j)\ (\min_{1\leq i\leq m} p_i)\}$ is of the form $\{n \in \mathbb{N}, n \geq n_0\}$ for some n_0. Using the notation of Lemma 2.1, the number of intervals of $\rho^{h_\epsilon}\pi$ which are contained in I_i equals M_{ϵ/α_i} for $i = 1, \ldots, l$, where

$$M_{\epsilon/\alpha_i} = \frac{(m-1)\alpha_i}{\epsilon H} + o\left(\frac{1}{\epsilon}\right).$$

For $i = 1, \ldots, l$,

$$\begin{aligned} \lim_{\epsilon\to 0} \frac{M_{\epsilon/\alpha_i}}{\sum_{j=1}^{l} M_{\epsilon/\alpha_j}} &= \lim_{\epsilon\to 0} \frac{\frac{(m-1)\alpha_i}{\epsilon H} + o\left(\frac{1}{\epsilon}\right)}{\sum_{j=1}^{l} \frac{(m-1)\alpha_j}{\epsilon H} + o\left(\frac{1}{\epsilon}\right)} \\ &= \frac{\alpha_i}{\sum_{j=1}^{l} \alpha_j} = \alpha_i, \end{aligned}$$

and thus the sequence $(\rho^n\pi)_{n\in\mathbb{N}}$ is u.d. .
Now assume that condition (I) does not hold. Then the numbers $\log\left(\frac{1}{p_1}\right),\dots,\log\left(\frac{1}{p_m}\right)$ are rationally related with some parameter Λ, and the number of intervals of $\rho^{h_\epsilon}\pi$ which are contained in I_i is M_{ϵ/α_i}, where by Lemma 2.1

$$M_{\epsilon/\alpha_i}=\frac{(m-1)\alpha_i Q_1\left(\log\left(\frac{\alpha_i}{\epsilon}\right)\right)}{\epsilon H}+\mathcal{O}\left((\log(\epsilon))^d\epsilon^{-(1-\eta)}\right). \tag{2.5}$$

Consider

$$\begin{aligned}\frac{M_{\epsilon/\alpha_i}}{\sum_{j=1}^l M_{\epsilon/\alpha_j}}&=\frac{\frac{(m-1)\alpha_i Q_1\left(\log\left(\frac{\alpha_i}{\epsilon}\right)\right)}{\epsilon H}+\mathcal{O}\left((\log(\epsilon))^d\epsilon^{-(1-\eta)}\right)}{\sum_{j=1}^l\frac{(m-1)\alpha_j Q_1\left(\log\left(\frac{\alpha_j}{\epsilon}\right)\right)}{\epsilon H}+\mathcal{O}\left((\log(\epsilon))^d\epsilon^{-(1-\eta)}\right)}\\&=\frac{\alpha_i Q_1\left(\log\left(\frac{\alpha_i}{\epsilon}\right)\right)+\mathcal{O}\left((\log(\epsilon))^d\epsilon^{-(1-\eta)}\right)}{\sum_{j=1}^l\alpha_j Q_1\left(\log\left(\frac{\alpha_j}{\epsilon}\right)\right)+\mathcal{O}\left((\log(\epsilon))^d\epsilon^{-(1-\eta)}\right)}.\end{aligned} \tag{2.6}$$

If (II) holds, then

$$\begin{aligned}\left\{\frac{\log\left(\frac{\alpha_j}{\epsilon}\right)}{\Lambda}\right\}&=\left\{\frac{\log\left(\frac{ce^{v_j\Lambda}}{\epsilon}\right)}{\Lambda}\right\}\\&=\left\{\frac{\log(c)+v_j\Lambda-\log(\epsilon)}{\Lambda}\right\}\\&=\left\{\frac{\log(c)-\log(\epsilon)}{\Lambda}\right\}\end{aligned}$$

and

$$Q_1\left(\log\left(\frac{\alpha_j}{\epsilon}\right)\right)=\frac{\Lambda e^{-\Lambda\left\{\frac{\log(n)-\log(\epsilon)}{\Lambda}\right\}}}{1-e^{-\Lambda}},$$

for all $j=1,\dots,l$. Thus for $i=1,\dots,l$,

$$\lim_{\epsilon\to 0}\frac{M_{\epsilon/\alpha_i}}{\sum_{j=1}^l M_{\epsilon/\alpha_j}}=\frac{\alpha_i}{\sum_{j=1}^l\alpha_j}=\alpha_i,$$

and $(\rho^n\pi)_{n\in\mathbb{N}}$ is u.d. .
Now assume that neither (I) nor (II) holds. Then the numbers $\log\left(\frac{1}{p_1}\right),\dots,\log\left(\frac{1}{p_m}\right)$ are rationally related with some parameter Λ, and the starting partition π has to consist of at least two elements. Furthermore, note that condition (II) is equivalent to assuming

$$\log(\alpha_j)-\log(\alpha_i)=n_{ij}\Lambda,\quad n_{ij}\in\mathbb{Z}, \tag{2.7}$$

for $i,j=1,\dots,l$, so if (II) does not hold there necessarily exist indices i,j for which (2.7) is not satisfied. Fix such i,j. Then

$$\left\{\frac{\log(\alpha_j)-\log(\alpha_i)}{\Lambda}\right\}>0. \tag{2.8}$$

Let the sequence $(\epsilon_k)_{k\in\mathbb{N}}$ be defined by

$$\epsilon_k = \alpha_i e^{-k\Lambda}, \quad k \geq 1.$$

Then for $k \geq 1$ and $n \in \{1, \ldots, l\}$,

$$\begin{aligned}\left\{\frac{\log\left(\frac{\alpha_n}{\epsilon_k}\right)}{\Lambda}\right\} &= \left\{\frac{\log\left(\frac{\alpha_n e^{k\Lambda}}{\alpha_i}\right)}{\Lambda}\right\} \\ &= \left\{\frac{\log(\alpha_n) + k\Lambda - \log(\alpha_i)}{\Lambda}\right\} \\ &= \left\{\frac{\log(\alpha_n) - \log(\alpha_i)}{\Lambda}\right\}. \end{aligned} \tag{2.9}$$

Hence,

$$Q_1\left(\log\left(\frac{\alpha_i}{\epsilon_k}\right)\right) = \frac{\Lambda}{1-e^{-\Lambda}}, \tag{2.10}$$

and

$$Q_1\left(\log\left(\frac{\alpha_j}{\epsilon_k}\right)\right) = \frac{\Lambda e^{-\Lambda\left\{\frac{\log(\alpha_j)-\log(\alpha_i)}{\Lambda}\right\}}}{1-e^{-\Lambda}}. \tag{2.11}$$

By using (2.6), we obtain

$$\lim_{k\to\infty} \frac{\alpha_i Q_1\left(\log\left(\frac{\alpha_i}{\epsilon_k}\right)\right) + \mathcal{O}\left((\log(\epsilon_k))^d \epsilon_k^{-(1-\eta)}\right)}{\sum_{n=1}^{l} \alpha_n Q_1\left(\log\left(\frac{\alpha_n}{\epsilon_k}\right)\right) + \mathcal{O}\left((\log(\epsilon_k))^d \epsilon_k^{-(1-\eta)}\right)} = \frac{\alpha_i}{\sum_{n=1}^{l} \alpha_n e^{-\Lambda\left\{\frac{\log(\alpha_n)-\log(\alpha_i)}{\Lambda}\right\}}}.$$

By (2.8) and $\Lambda > 0$ it follows that

$$e^{-\Lambda\left\{\frac{\log(\alpha_n)-\log(\alpha_i)}{\Lambda}\right\}} \leq 1, \quad n = 1, \ldots, l,$$

and

$$e^{-\Lambda\left\{\frac{\log(\alpha_j)-\log(\alpha_i)}{\Lambda}\right\}} < 1.$$

Thus

$$\sum_{n=1}^{l} \alpha_n e^{-\Lambda\left\{\frac{\log(\alpha_n)-\log(\alpha_i)}{\Lambda}\right\}} < 1$$

and

$$\lim_{k\to\infty} \frac{M_{\epsilon_k/\alpha_i}}{\sum_{j=1}^{l} M_{\epsilon_k/\alpha_j}} \neq \alpha_i.$$

Thus there exists a subsequence along which the relative number of intervals in I_i does not converge to α_i, and hence the sequence $(\rho^n\pi)_{n\in\mathbb{N}}$ cannot be u.d. This proves the theorem. □

Proof of Corollary 2.1:
Proof:
The corollary is a special case of Theorem 2.1. By Remark 2.2, condition (i) is equivalent to condition (I).
Furthermore, condition (ii) is equivalent to (II). Assume that (II) holds, then $\alpha = ce^{r\Lambda}, 1-\alpha = ce^{q\Lambda}$, for $q, r \in \mathbb{Z}, c \in \mathbb{R}^+$, and thus

$$\begin{aligned} 1 &= ce^{r\Lambda} + ce^{q\Lambda} \\ \Leftrightarrow c &= \frac{1}{e^{r\Lambda} + e^{q\Lambda}} \end{aligned}$$

and

$$\alpha = \frac{e^{r\Lambda}}{e^{r\Lambda} + e^{q\Lambda}} = \frac{1}{e^{(q-r)\Lambda} + 1}.$$

□

Proof of Theorem 2.2:
Proof:
Let the i-th interval of π be denoted by I_i and let $h_\epsilon \in \mathbb{N}$ be the smallest number for which $\rho^{h_\epsilon}\pi$ contains only intervals of length $< \epsilon$. Then, following the proof of Theorem 2.1, the number of intervals of $\rho^{h_\epsilon}\pi$ which are contained in I_i is M_{ϵ/α_i}, which is given in (2.5). We denote by $M_A(\epsilon)$ the number of intervals of $\rho^{h_\epsilon}\pi$ which are contained in $A = [a,b] \subseteq I_i$. By [101, Theorem 2.7], the sequences of partitions within I_i are u.d. Hence

$$M_A(\epsilon) = \frac{(b-a)(m-1)Q_1\left(\log\left(\frac{\alpha_i}{\epsilon}\right)\right)}{\epsilon H} + \mathcal{O}\left((\log(\epsilon))^d \epsilon^{-(1-\eta)}\right).$$

Thus the relative number of intervals in A is given by

$$\begin{aligned} \frac{M_A(\epsilon)}{\sum_{j=1}^{l} M_{\epsilon/\alpha_j}} &= \frac{\frac{(b-a)(m-1)Q_1\left(\log\left(\frac{\alpha_i}{\epsilon}\right)\right)}{\epsilon H} + \mathcal{O}\left((\log(\epsilon))^d \epsilon^{-(1-\eta)}\right)}{\sum_{j=1}^{l} \frac{(m-1)\alpha_j Q_1\left(\log\left(\frac{\alpha_j}{\epsilon}\right)\right)}{\epsilon H} + \mathcal{O}\left((\log(\epsilon))^d \epsilon^{-(1-\eta)}\right)} \\ &= \frac{(b-a)Q_1\left(\log\left(\frac{\alpha_i}{\epsilon}\right)\right) + \mathcal{O}\left((\log(\epsilon))^d \epsilon^{-(1-\eta)}\right)}{\sum_{j=1}^{l} \alpha_j Q_1\left(\log\left(\frac{\alpha_j}{\epsilon}\right)\right) + \mathcal{O}\left((\log(\epsilon))^d \epsilon^{-(1-\eta)}\right)}. \end{aligned}$$

Consider

$$\frac{(b-a)Q_1\left(\log\left(\frac{\alpha_i}{\epsilon}\right)\right)}{\sum_{j=1}^{l} \alpha_j Q_1\left(\log\left(\frac{\alpha_j}{\epsilon}\right)\right)} \tag{2.12}$$

$$= \frac{(b-a)\Lambda e^{-\Lambda\{\frac{1}{\Lambda}(\log(\alpha_i)-\log(\epsilon))\}}}{\sum_{j=1}^{l} \alpha_j \Lambda e^{-\Lambda\{\frac{1}{\Lambda}(\log(\alpha_j)-\log(\epsilon))\}}}$$

$$= \frac{b-a}{\sum_{j=1}^{l} \alpha_j \exp\left(-\Lambda\left(\{\frac{1}{\Lambda}(\log(\alpha_j)-\log(\epsilon))\} - \{\frac{1}{\Lambda}(\log(\alpha_i)-\log(\epsilon))\}\right)\right)}. \tag{2.13}$$

For $j \neq i$, one easily sees that the functions

$$f_{i,j}(\epsilon) := \exp\left(-\Lambda\left(\left\{\frac{\log(\alpha_j)-\log(\epsilon)}{\Lambda}\right\}-\left\{\frac{\log(\alpha_i)-\log(\epsilon)}{\Lambda}\right\}\right)\right)$$

are piecewise constant with discontinuities at

$$\epsilon = \alpha_i e^{-k\Lambda} \text{ and } \epsilon = \alpha_j e^{-k\Lambda},$$

for all $k \in \mathbb{Z}$. By

$$f_{i,j}(\alpha_i e^{-k_1\Lambda}) = f_{i,j}(\alpha_i e^{-k_2\Lambda}),$$
$$f_{i,j}(\alpha_j e^{-k_1\Lambda}) = f_{i,j}(\alpha_j e^{-k_2\Lambda}),$$

for all $k_1, k_2 \in \mathbb{Z}$, it follows that $f_{i,j}(\epsilon), 0 < \epsilon < 1$, only takes two different values, which are

$$\exp\left(-\Lambda\left\{\frac{\log(\alpha_j)-\log(\alpha_i)}{\Lambda}\right\}\right) \quad \text{and} \quad \exp\left(\Lambda\left\{\frac{\log(\alpha_i)-\log(\alpha_j)}{\Lambda}\right\}\right).$$

Furthermore, for all $k \in \mathbb{Z}$

$$f_{i,j}(\alpha_i e^{-k\Lambda}) = \exp\left(-\Lambda\left\{\frac{\log(\alpha_j)-\log(\alpha_i)}{\Lambda}\right\}\right)$$

and

$$f_{i,j}(\alpha_j e^{-k\Lambda}) = \exp\left(\Lambda\left\{\frac{\log(\alpha_i)-\log(\alpha_j)}{\Lambda}\right\}\right).$$

By the above arguments it follows that the function

$$\sum_{j=1}^{l} \alpha_j f_{i,j}(\epsilon),$$

where $f_{i,i}(\epsilon) = 1$, can only take at most l different values. Since all the functions $f_{i,j}(\epsilon), 1 \leq j \leq l$, attain their minimal value at the positions $\alpha_i e^{-k\Lambda}, k \in \mathbb{Z}$, it follows that the quotient in equation (2.13) is maximal at these positions and

$$\begin{aligned}\limsup_{n\to\infty} \frac{\sum_{j=1}^{k(n)} \mathbf{1}_{[a,b]}(t_j^n)}{k(n)} &= \limsup_{\epsilon\to 0} \frac{M_A(\epsilon)}{\sum_{j=1}^{l} M_{\epsilon/\alpha_j}} \\ &= \lim_{k\to\infty} \frac{M_A(\alpha_i \exp(-k\Lambda))}{\sum_{j=1}^{l} M_{(\alpha_i \exp(-k\Lambda))/\alpha_j}} \\ &= \frac{b-a}{\sum_{j=1}^{l} \alpha_j \exp\left(-\Lambda\left\{\frac{\log(\alpha_j)-\log(\alpha_i)}{\Lambda}\right\}\right)}.\end{aligned}$$

This proves the upper bound in Theorem 2.2.
To prove the lower bound in Theorem 2.2, we choose $0 < \gamma < 1$ such that

$$\gamma\alpha_i > \max_{1\leq j\leq l} \max_{k\in\mathbb{Z}} \left\{\alpha_j e^{-k\Lambda} \mid \alpha_j e^{-k\Lambda} < \alpha_i\right\}.$$

Then for all $1 \leq j \leq l$ and for all $k \in \mathbb{Z}$ the functions $f_{i,j}$ attain their maximal value at the positions $\gamma\alpha_i e^{-k\Lambda}$, and

$$f_{i,j}(\gamma\alpha_i e^{-k\Lambda}) = f_{i,j}(\alpha_j e^{-k\Lambda}) = \exp\left(\Lambda\left\{\frac{\log(\alpha_i) - \log(\alpha_j)}{\Lambda}\right\}\right).$$

Therefore, the quotient in equation (2.13) attains its minimal possible value at the positions $\gamma\alpha_i e^{-k\Lambda}, k \in \mathbb{Z}$, and

$$\begin{aligned}\liminf_{n\to\infty} \frac{\sum_{j=1}^{k(n)} \mathbf{1}_{[a,b]}(t_j^n)}{k(n)} &= \liminf_{\epsilon\to 0} \frac{M_A(\epsilon)}{\sum_{j=1}^{l} M_{\epsilon/\alpha_j}} \\ &= \lim_{k\to\infty} \frac{M_A(\gamma\alpha_i \exp(-k\Lambda))}{\sum_{j=1}^{l} M_{\left(\gamma\alpha_i e^{-k\Lambda}\right)/\alpha_j}} \\ &= \frac{b-a}{\sum_{j=1}^{l} \alpha_j \exp\left(\Lambda\left\{\frac{\log(\alpha_i)-\log(\alpha_j)}{\Lambda}\right\}\right)}.\end{aligned}$$

This proves the theorem. □

Chapter 3

On the limit distribution of consecutive elements of the van der Corput sequence

3.1 Introduction

In the open problem collection on the web site of *Uniform distribution theory* the following problem is stated:

> *L*et $(\phi_b(n))_{n\geq 0}$ denote the van der Corput sequence in base b. Find the distribution of the sequence $(\phi_b(n), \phi_b(n+1), \ldots, \phi_b(n+s-1))_{n\geq 0}$ in $[0,1)^s$.[1]

The case $s = 2$ has recently been solved by Fialová and Strauch [37]. They showed that every point $(\phi_b(n), \phi_b(n+1))_{n\geq 0}$ lies on the line segment

$$y = x - 1 + \frac{1}{b^k} + \frac{1}{b^{k+1}}, \quad x \in \left[1 - \frac{1}{b^k}, 1 - \frac{1}{b^{k+1}}\right]$$

for $k \geq 0$. Furthermore they could give an explicit formula for the asymptotic distribution function of $(\phi_b(n), \phi_b(n+1))_{n\geq 0}$ to calculate the limit

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} |\phi_b(n) - \phi_b(n+1)| = \frac{2(b-1)}{b^2}$$

previously demonstrated by Pillichshammer and Steinerberger [90]. They also noted that the adf of $(\phi_b(n), \phi_b(n+1))_{n\geq 0}$ is a copula.
In this article we solve the problem for the sequence $(\phi_b(n), \phi_b(n+1), \ldots, \phi_b(n+s-1))_{n\geq 0}$ for $s > 2$. A multi-dimensional extension of the van der Corput sequence $(\phi_b(n))_{n\geq 0}$, is given by the so-called Halton sequence, $(\phi_{b_1}(n), \phi_{b_2}(n), \ldots, \phi_{b_s}(n))_{n\geq 0}$ which is uniformly distributed

[1] Problem 1.12 in the open problem collection as of 11. December 2011 (http://www.boku.ac.at/MATH/udt/unsolvedproblems.pdf)

if and if the bases $b_i 1 \leq i \leq s$ are co-prime (see [53]). These sequences are well-studied objects in discrepancy theory, since they belong to the class of so-called low discrepancy sequences. For classical results in discrepancy theory, on low discrepancy sequences and the van der Corput sequence see e.g. [26], [30] or [66].

Recently, several authors investigated the ergodic properties of low discrepancy sequences, see e.g. [48] and [84]. In the case of van der Corput sequences this can be done using the so-called von Neumann-Kakutani transformation, which will be discussed in the second section.
The outline of this article is as follows: in the second section we define the van der Corput sequence and the von Neumann-Kakutani transformation and recall their basic properties. In the third section we state our main results on the distribution of $(\phi_b(n), \phi_b(n+1), \ldots, \phi_b(n+s-1))_{n\geq 0}$.

3.2 van der Corput sequence and von Neumann-Kakutani transformation

Let $b \in \mathbb{N}$ and $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$. Then for every $n \in \mathbb{N}_0$, we can write

$$n = \sum_{i\geq 0} n_i b^i$$

where $n_i \in \{0, 1, \ldots, b-1\}, i \geq 0$. The above sum is called b-adic representation of n. The n_i are uniquely determined and at most a finite number of n_i are non-zero. Furthermore, every real $x \in [0, 1)$ has a b-adic representation of the following form

$$x = \sum_{i\geq 0} x_i b^{-i-1} \tag{3.1}$$

where $x_i \in \{0, 1, \ldots, b-1\}, i \geq 0$. We call x a b-adic rational if $x = ab^{-c}$, where a and c are positive integers and $0 \leq a < b^c$. For all b-adic integers there are exactly two representations of the form (3.1), one where $x_i = 0, i \geq i_0$ and one where $x_i = b-1, i \geq i_0$ for sufficiently large $i_0 \in \mathbb{N}$. If we restrict ourselves to representations with $x_i \neq b-1$ for infinitely many i, then the coefficients x_i in (3.1) are uniquely determined for all $x \in [0, 1)$.
For $n \in \mathbb{N}_0$ we define the so-called radical-inverse function or Monna map $\phi_b(n)\colon \mathbb{N}_0 \to [0, 1)$ by

$$\phi_b(n) = \phi_b\left(\sum_{i\geq 0} n_i b^i\right) := \sum_{i\geq 0} n_i b^{-i-1}.$$

Note that $\phi_b(n)$ maps $\mathbb{N}_0$ to the set of b-adic rationals in $[0, 1)$, and therefore the image of $\mathbb{N}_0$ under $\phi_b(n)$ is dense in $[0, 1)$.

Definition 3.1 *The van der Corput sequence in base b is defined as $(\phi_b(n))_{n\geq 0}$.*

It is a classical result that the van der Corput sequence is uniformly distributed in $[0,1)$, see e.g. [66]. Furthermore, its s-dimensional extension, the Halton sequence given by $(\phi_{b_1}(n),\dots,\phi_{b_s}(n))_{n\geq 0}$ for co-prime bases $b_i, 1 \leq i \leq s$, is uniformly distributed on $[0,1)^s$. Properties of the van der Corput and the Halton sequence are very well-understood, since they are so-called low discrepancy sequences, which are central objects in Quasi-Monte Carlo integration.

A second approach to define the van der Corput sequence is by using the von Neumann-Kakutani transformation $T_b\colon [0,1) \to [0,1)$. For any integer $b \geq 2$ the inductive construction of T_b is as follows: at first $[0,1)$ is split into b intervals $I_i^1 = \left[\frac{i}{b}, \frac{i+1}{b}\right)$ for $i = 0,1,\dots b-1$. Then the transformation $T_{1,b}\colon \left[0, \frac{b-1}{b}\right) \mapsto \left[\frac{1}{b}, 1\right)$ is defined as translation of I_i^1 into I_{i+1}^1 for $i = 0,1,\dots,b-1$. The next step is to divide all intervals I_i^1 into b subintervals of the form $I_i^2 = \left[\frac{i}{b^2}, \frac{i+1}{b^2}\right)$ for $i = 0,1,\dots b^2-1$. Transformation $T_{2,b}\colon \left[0, \frac{b^2-1}{b^2}\right) \mapsto \left[\frac{1}{b^2}, 1\right)$ is given as the extension of $T_{1,b}$ which translates $I_{b^2-b+i}^2$ into $I_{b^2-b+i+1}^2$ for $i = 0,1,\dots,b-1$. Such a construction is called splitting-and-stacking-construction and is illustrated in Figure 3.1 for $b = 2$. Finally we define the von Neumann-Kakutani transformation as $T_b = \lim_{n\to\infty} T_{n,b}$. A plot of the transformation T_2 is given in Figure 3.2. By an observation of Lambert [68], [69] (see also Hellekalek [53]) the van der Corput sequence in base b is exactly the orbit of the origin under T_b, which means that

$$(T_b^n 0)_{n\geq 0} = (\phi_b(n))_{n\geq 0}, \quad b \geq 2, \tag{3.2}$$

where $T_b^n x$ denotes the value of x under after n iterations of T_b.

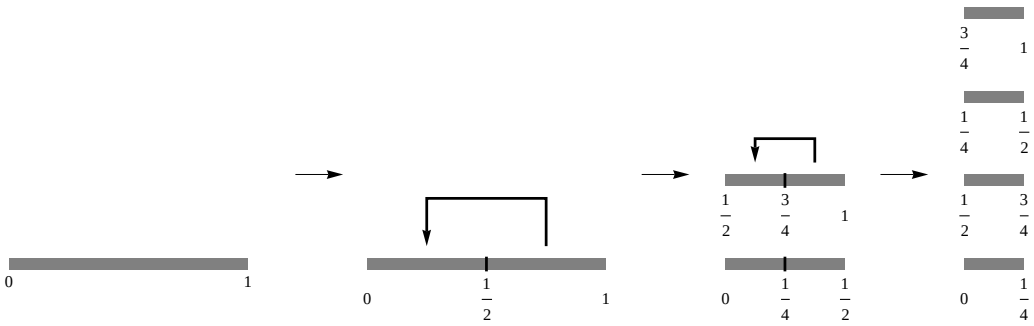

Figure 3.1: The first two steps of a splitting-and-stacking-construction in base $b = 2$.

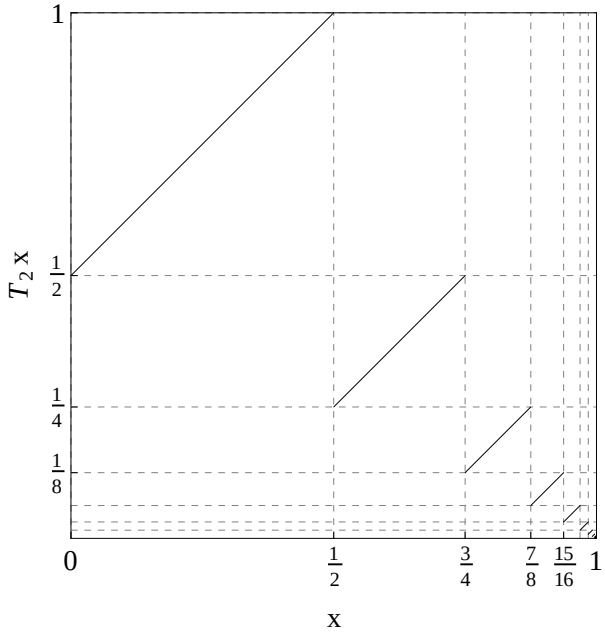

Figure 3.2: The von Neumann-Kakutani transformation in base $b = 2$.

For a proof of the ergodicity and measure-preserving properties of the von Neumann-Kakutani transformation, see e.g. [38] or [39]. It follows from the ergodicity of the von Neumann-Kakutani transformation that $(T_b^n x)_{n\geq 0}$ is uniformly distributed for almost every $x \in [0, 1)$. Furthermore, it can be shown that the von Neumann-Kakutani transformation is uniquely ergodic, which implies that $(T_b^n x)_{n\geq 0}$ is uniformly distributed for every $x \in [0, 1)$, see e.g. [48]. Moreover, Pagés [89] showed that the orbit of the von Neumann-Kakutani transformation starting at an arbitrary point $x \in [0, 1)$ is a low discrepancy sequence. Another possible generalization of the van der Corput sequence is the so-called randomized van der Corput sequence $(T_b^n X)_{n\geq 0}$ where X is uniformly distributed on $[0, 1)$, see [102].

Recently, Fialová and Strauch solved the problem of calculating the limit distribution of the sequence $(\phi_b(n), \phi_b(n + 1))_{n\geq 0}$. They also concluded that the limit distribution is a copula. We consider the multi-dimensional extension of this problem. By (3.2)

$$(\phi_b(n), \phi_b(n + 1))_{n\geq 0} = (T_b^n 0, T_b^{n+1} 0)_{n\geq 0} = (T_b^n 0, T_b(T_b^n 0))_{n\geq 0}.$$

By the fact that $(T_b^n 0)_{n\geq 0}$ is uniformly distributed on $[0, 1)$ one can show that $(\phi_b(n), \phi_b(n + 1))_{n\geq 0}$ is uniformly distributed on

$$\Gamma = \{(x, y) : y = T_b x\}.$$

Note that Γ coincides with the graph of the von Neumann-Kakutani transformation in Figure 3.2. In the next section we use this approach to find the limit distribution of $(\phi_b(n), \phi_b(n + 1), \ldots, \phi_b(n + s - 1))_{n\geq 0}$ for arbitrary $s \geq 2$.

3.3 The limit distribution of consecutive elements of the van der Corput sequence

In the sequel we assume that b, s are fixed. Let T denote the von Neumann-Kakutani transformation in base b as described in Section 3.2. We define a map $\gamma(t)\colon [0,1) \to [0,1)^s$ by setting

$$\gamma(t) := \begin{pmatrix} t \\ Tt \\ T^2t \\ \vdots \\ T^{s-1}t \end{pmatrix}$$

and

$$\Gamma := \{(x_1, x_2, \ldots, x_s) \in [0,1]^s : x_i = T^{i-1}x_1, i = 2, \ldots, s\} = \{\gamma(t) : t \in [0,1)\}\,.$$

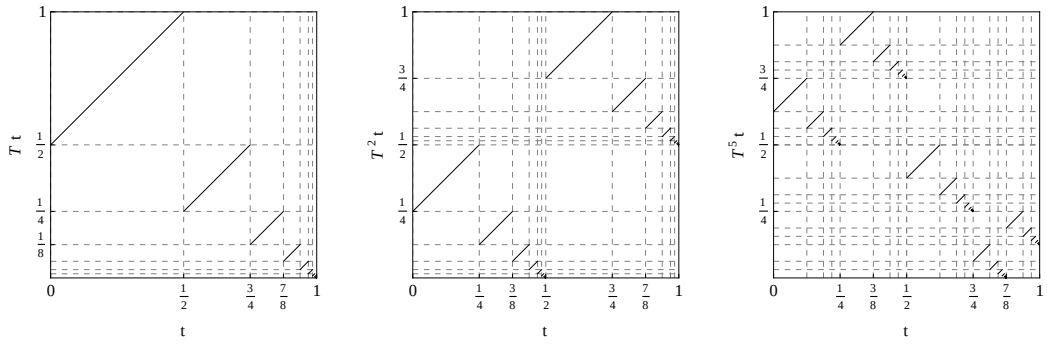

Figure 3.3: Function graphs of Tt, T^2t and T^5t. These curves appear as the two-dimensional projections of Γ for large s.

The Lebesgue measure λ_1 on $[0,1)$ induces a measure ν on Γ by setting

$$\nu(A) = \lambda_1(\{t : \gamma(t) \in A\}), \quad A \subset \Gamma.$$

Furthermore, ν induces a measure μ on $[0,1)^s$ by embedding Γ into $[0,1)^s$. More precisely for every measurable subset $B \subseteq [0,1)^s$ we set

$$\mu(B) = \nu(B \cap \Gamma).$$

Theorem 3.1 *The limit measure of* $(\phi_b(n), \phi_b(n+1), \ldots, \phi_b(n+s-1))_{n\geq 0}$ *is* μ.

Proof:
As mentioned in Section 3.2, we can rewrite

$$(\phi_b(n), \phi_b(n+1), \ldots, \phi_b(n+s-1))_{n\geq 0} = (T^n 0, T^{n+1}0, \ldots, T^{n+s-1}0)_{n\geq 0}$$

$$= (T^n 0, T(T^n 0), \ldots, T^{s-1}(T^n 0))_{n\geq 0}.$$

Since $(T^n 0)_{n\geq 0}$ is uniformly distributed on $[0,1)$ and T is a measure-preserving transformation with respect to λ_1, it follows immediately that $(T^i(T^n 0))_{n\geq 0}$ is uniformly distributed on $[0,1)$ for $i = 1, \ldots, s-1$. Moreover, by construction $(T^n 0, T(T^n 0), \ldots, T^{s-1}(T^n 0))_{n\geq 0} \in \Gamma$ for all $n \geq 0$.
Now consider a measurable set $B \in [0,1)^s$. We define the empirical measure of the first N points of $(T^n 0, \ldots, T^{s-1}(T^n 0))_{n\geq 0}$ as

$$\mu_N(B) = \frac{1}{N}\#\{0 \leq n \leq N : (T^n 0, T(T^n 0), \ldots, T^{s-1}(T^n 0)) \in B\}.$$

We have

$$\begin{aligned}
\lim_{N\to\infty} \mu_N(B) &= \lim_{N\to\infty} \frac{1}{N}\#\{0 \leq n \leq N : (T^n 0, T(T^n 0), \ldots, T^{s-1}(T^n 0)) \in B\} \\
&= \lim_{N\to\infty} \frac{1}{N}\#\{0 \leq n \leq N : (T^n 0, T(T^n 0), \ldots, T^{s-1}(T^n 0)) \in B \cap \Gamma\} \\
&= \lim_{N\to\infty} \frac{1}{N}\#\{0 \leq n \leq N : T^n 0 \in \text{ Projection}_{x_1}(B \cap \Gamma)\} \\
&= \lambda_1(\text{ Projection}_{x_1}(B \cap \Gamma)) \\
&= \nu(B \cap \Gamma) = \mu(B)
\end{aligned}$$

where the fourth equation holds since $(T^n 0)_{n\geq 0}$ is uniformly distributed on $[0,1)$ and since the map $t \to Tt$ is a bijection, and where $\text{Projection}_{x_1}(A)$ denotes the projection of A onto its first coordinate. □

Remark 3.1 *Note that the measure μ is a copula on $[0,1]^s$ for every s since every distribution function of a multi-dimensional sequence $(x_n^1, \ldots, x_n^s)_{n\geq 0}$ is a copula if the sequences $(x_n^1)_{n\geq 0}, \ldots, (x_n^s)_{n\geq 0}$ are uniformly distributed on $[0,1]$.*

Remark 3.2 *The set Γ is a collection of countably many line segments in $[0,1)^s$. Informally speaking Theorem 3.1 means that $(\phi_b(n), \phi_b(n+1), \ldots, \phi_b(n+s-1))_{n\geq 0}$ is uniformly distributed on Γ.*

Remark 3.3 *By the unique ergodicity of T, the conclusion of Theorem 3.1 holds also for the sequence $(T^n x, T(T^n x), \ldots, T^{s-1}(T^n x))_{n\geq 0}$ for arbitrary $x \in [0,1)$.*

Remark 3.4 *Another class of uniformly distributed sequences which can be seen as the orbits of certain points under an ergodic transformation are sequences of the form $(\{n\alpha\})_{n\geq 0}$, where $\{x\}$ denotes the fractional part of x and α is irrational. In this case the corresponding transformation $\widehat{T}$ is simply the rotation $\widehat{T}\colon x \mapsto x + \alpha \mod 1$. It can easily be shown that the limit distribution of consecutive elements $(\{n\alpha\}, \{(n+1)\alpha\}, \ldots, \{(n+s-1)\alpha\})_{n\geq 0}$ is the uniform distribution on the curve $\widehat{\Gamma}$ which is given by*

$$\widehat{\Gamma} := \{(t, \widehat{T}t, \ldots, \widehat{T}^{s-1}t), t \in [0,1)\}.$$

However, since in this case the transformation $\widehat{T}$ has a particularly simple structure, the same result can also be easily obtained using analytic arguments.

Part II

The discrepancy of Monte Carlo point sets

Chapter 4

Probabilistic discrepancy bounds for Monte Carlo point sets

4.1 Introduction and statement of results

The number $n^*(s,\varepsilon)$, which is defined as the smallest possible cardinality of a point set in $[0,1]^s$ having discrepancy bounded by ε, is called the *inverse of the discrepancy*. Heinrich, Novak, Wasilkowski, and Woźniakowski [52] proved the upper bound

$$n^*(s,\varepsilon) \leq c_{abs} s \varepsilon^{-2}, \tag{4.1}$$

which is complemented by the lower bound

$$n^*(s,\varepsilon) \geq c_{abs} s \varepsilon^{-1}$$

due to Hinrichs [55] (throughout the paper, c_{abs} denotes absolute constants, not always the same). Hence the inverse of the star-discrepancy depends linearly on the dimension, while the precise dependence on ε is still unknown. It is easy to see that (4.1) is equivalent to the fact that for any N and s there exists a set $\mathcal{P}_N$ of N points in $[0,1]^s$ such that the star-discrepancy D_N^* of this point set is bounded by

$$D_N^*(\mathcal{P}_N) \leq c_{abs} \frac{\sqrt{s}}{\sqrt{N}} \tag{4.2}$$

(recently we showed that it is possible to choose $c_{abs} = 10$ in (4.2), see [1]). The existence of such a point set directly follows from the surprising observation that a randomly generated point set (that is, a Monte Carlo point set) satisfies the desired discrepancy estimate with positive probability. Of course, for applications such a mere existence result is not of much use, as was remarked by several colleagues at the MCQMC 2012 conference in Sydney. For this reason, in the present paper we prove an applied version of (4.2), which provides estimates for the probability of a random point set satisfying (4.2) (depending on the value of the constant). As our Theorem 4.1 below shows, this probability is extremely large already for moderate values of c, for example for $c = 20$. Additionally, the quality of our estimates for these probabilities *improves* as the dimension s increases (which is somewhat counter-intuitive, and originates from the exponential

inequalities used in the proof, which cause a "concentration of mass" phenomenon).

The fact that the probability of a random point set satisfying (4.2) is very large is in contrast to the fact that no general constructions of point sets satisfying such discrepancy bounds are known. So far, the best results are a component-by-component construction of Doerr, Gnewuch, Kritzer and Pillichshammer [27], a semi-deterministic algorithm based on dependent randomized rounding due to Doerr, Gnewuch, and Wahlström [28], and a construction of Hinrichs of a "structured" set of $N = 1528$ points in dimension $s = 15$ having discrepancy less than 1/4 (by this means solving one instance of an open problem in [33]).

For more information concerning the inverse of the discrepancy and tractability of multidimensional integration we refer to a recent survey article of Gnewuch [46], and to the monograph of Novak and Woźniakowski [32, 33]. A collection of open problems on this topic can be found in [51].

In the present paper, we will prove the following theorem.

Theorem 4.1 *For any $s \geq 1,\ N \geq 1$ and $q \in (0,1)$ a randomly generated s-dimensional point set $(z_1, \ldots, z_N)$ satisfies*

$$D_N^*(z_1, \ldots, z_N) \leq 5.70\sqrt{4.90 + \frac{\log\left((1-q)^{-1}\right)}{s}}\,\frac{\sqrt{s}}{\sqrt{N}} \tag{4.3}$$

with probability at least q.

It is interesting that the quality of the discrepancy estimate in Theorem 4.1 *improves* as the dimension s increases; for example the necessary number $c(q,s)$ to have star-discrepancy bounded by $c(q,s)s^{1/2}N^{-1/2}$ with probability at least 90% is 15.30 in dimension $s = 1$, while it is only 12.65 in dimension $s = 100$. However, neglecting this advantage of large dimensions in order to obtain a result which holds uniformly in s, one immidiately obtains the following corollary.

Corollary 4.1 *For any $s \geq 1,\ N \geq 1$ and $q \in (0,1)$ a randomly generated s-dimensional point set $(z_1, \ldots, z_N)$ satisfies*

$$D_N^*(z_1, \ldots, z_N) \leq 5.70\sqrt{4.90 + \log\left((1-q)^{-1}\right)}\,\frac{\sqrt{s}}{\sqrt{N}} \tag{4.4}$$

with probability at least q.

Theorem 4.1 shows that the probability that a random point set satisfies the discrepancy bound $c(q,s)s^{1/2}N^{-1/2}$ is extremely large already for moderate values of $c(q,s)$. The following table illustrates this fact, for $s = 10$ and $s = 100$.

q	0.01	0.5	0.9	0.99	0.999
c(q,10)	12.62	12.71	12.92	13.20	13.48
c(q,100)	12.62	12.63	12.65	12.68	12.71

As the table shows, the probability that a random point set has "small" discrepancy in the sense that its discrepancy is bounded by $cs^{1/2}N^{-1/2}$ for some moderate c (for example, $c = 20$) is extremely large. This observation is an exciting counterpart of the fact that we do not have the slightest idea of how to construct point sets satisfying such discrepancy bounds, even for moderate N and s. It should also be noted that calculating the star-discrepancy of a given (high-dimensional) point set is computationally very difficult, see [41, 47]. Hence, although our results show that the probability of a random point set having small discrepancy is very large, checking that a concrete point set satisfies such discrepancy bounds is in general (in high dimensions) a computationally intractable problem.

4.2 Preliminaries

Throughout the paper, $s \geq 1$ denotes the dimension and λ denotes the s-dimensional Lebesgue measure. For $x, y \in [0,1]^s$, where $x = (x_1, \ldots, x_s)$ and $y = (y_1, \ldots, y_s)$, we write $x \leq y$ if $x_i \leq y_i, 1 \leq i \leq s$, and for any $x \in [0,1]^s$ we write $[0, x]$ for the set $\{y \in [0,1]^s : \ 0 \leq y \leq x\}$. Furthermore, we write $|A|$ for the number of elements of a set A.

The following Lemma 4.1 of Gnewuch [43, Theorem 1.15] is a central ingredient in the proof of our main result. For convenience we use the notation from [43] and [44]: For any $\delta \in (0,1]$ a set Γ of points in $[0,1]^s$ is called a δ-cover of $[0,1]^s$ if for every $y \in [0,1]^s$ there exist $x, z \in \Gamma \cup \{0\}$ such that $x \leq y \leq z$ and $\lambda([0,z)) - \lambda([0,x)) \leq \delta$. The number $\mathcal{N}(s,\delta)$ denotes the smallest possible cardinality of a δ-cover of $[0,1]^s$.
Similarly, for any $\delta \in (0,1]$ a set Δ of pairs of points from $[0,1]^s$ is called a δ-bracketing cover of $[0,1]^s$, if for every pair $(x,z) \in \Delta$ the estimate $\lambda([0,z)) - \lambda([0,x)) \leq \delta$ holds, and if for every $y \in [0,1]^s$ there exists a pair (x,z) from Δ such that $x \leq y \leq z$. The number $\mathcal{N}_{[\,]}(s,\delta)$ denotes the smallest possible cardinality of a δ-bracketing cover of $[0,1]^s$.

Lemma 4.1 *For any $s \geq 1$ and $\delta \in (0,1]$*

$$\mathcal{N}(s,\delta) \leq (2e)^s(\delta^{-1}+1)^s$$

and

$$\mathcal{N}_{[\,]}(s,\delta) \leq 2^{s-1}e^s(\delta^{-1}+1)^s.$$

By Lemma 4.1 for any $1 \leq k \leq K$ there exists a 2^{-k}-cover of $[0,1]^s$, denoted by Γ_k, such that

$$|\Gamma_k| \leq (2e)^s(2^k+1)^s.$$

Furthermore we denote by Δ_K a 2^{-K}- bracketing cover for which

$$|\Delta_K| \leq 2^{s-1}e^s(2^K+1)^s,$$

which also exists due to Lemma 4.1. Moreover we define Γ_K as

$$\Gamma_K = \{v \in [0,1]^s : \ (v,w) \in \Delta_K \text{ for some } w\}.$$

By definition for every $x \in [0,1]^s$ there exists a pair $(v_K, w_K) = (v_K(x), w_K(x))$ for which $(v_K, w_K) \in \Delta_K$ such that $v_K \le x \le w_K$ and

$$\lambda([0, w_K]) - \lambda([0, v_K]) \le \frac{1}{2^K}.$$

Furthermore for every $k,\ 2 \le k \le K$ and $\gamma \in \Gamma_k$ there exist $v_{k-1} = v_{k-1}(\gamma), w_{k-1} = w_{k-1}(\gamma),\ v_{k-1}, w_{k-1} \in \Gamma_{k-1} \cup \{0\}$, such that $v_{k-1} \le \gamma \le w_{k-1}$ and

$$\lambda([0, w_{k-1}]) - \lambda([0, v_{k-1}]) \le \frac{1}{2^{k-1}}.$$

We define

$$\begin{aligned} p_K(x) &= v_K(x) \\ p_{K-1}(x) &= v_{K-1}(p_K(x)) = v_{K-1}(v_K(x)) \\ p_{K-2}(x) &= v_{K-2}(p_{K-1}(x)) = v_{K-2}(v_{K-1}(v_K(x))) \\ &\vdots \\ p_1(x) &= v_1(p_2(x)), \end{aligned}$$

and

$$p_{K+1}(x) = w_K(x), \qquad p_0(x) = 0.$$

For $x, y \in [0,1]^s$ we set

$$\overline{[x,y]} := \begin{cases} [0,y]\backslash[0,x] & \text{if } x \ne 0, \\ [0,y] & \text{if } x = 0,\ y \ne 0, \\ \emptyset & \text{if } x = y = 0. \end{cases}$$

Then the sets

$$\overline{[p_k(x), p_{k+1}(x)]}, \quad 1 \le k \le K,$$

are disjoint, and we obtain

$$\bigcup_{k=0}^{K-1} \overline{[p_k(x), p_{k+1}(x)]} \subset [0,x] \subset \bigcup_{k=0}^{K} \overline{[p_k(x), p_{k+1}(x)]}, \quad \forall x \in [0,1]^s.$$

Hence for every $x, y \in [0,1]^s$

$$\sum_{k=0}^{K-1} \mathbf{1}_{\overline{[p_k(x), p_{k+1}(x)]}}(y) \le \mathbf{1}_{[0,x]}(y) \le \sum_{k=0}^{K} \mathbf{1}_{\overline{[p_k(x), p_{k+1}(x)]}}(y). \tag{4.5}$$

Moreover, independent of x, we have for $0 \le k \le K$

$$\lambda\left(\overline{[p_k(x), p_{k+1}(x)]}\right) \le \frac{1}{2^k}.$$

For $0 \leq k \leq K$ we define A_k to be the set of all sets of the form

$$\overline{[p_k(x), p_{k+1}(x)]},$$

where $x \in [0,1]^s$. Then for $0 \leq k \leq K$, as a consequence of Lemma 4.1, we can bound the cardinality of A_k by

$$|A_k| \leq (2e)^s \left(2^{k+1}+1\right)^s. \tag{4.6}$$

Note that all elements of A_k, where $0 \leq k \leq K$, have Lebesgue measure bounded by 2^{-k}. This dyadic decomposition method was introduced in [1], where it is described in more detail.

Let $X_1, \ldots, X_N$ be independent, identically distributed (i.i.d.) random variables defined on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$ having uniform distribution on $[0,1]^s$, and let $I \in A_k$ for some $k \geq 0$. Then the random variables $\mathbf{1}_I(X_1), \ldots, \mathbf{1}_I(X_N)$ are i.i.d. random variables, having expected value

$$\lambda(I)$$

and variance

$$\lambda(I) - \lambda(I)^2 \leq \begin{cases} 2^{-k}(1-2^{-k}) & \text{for } k \geq 1, \\ 1/4 & \text{for } k = 0. \end{cases} \tag{4.7}$$

Since the X_n are independent it follows that the random variable

$$\sum_{n=1}^{N} \mathbf{1}_I(X_n)$$

has expected value $N\lambda(I)$ and variance $N(\lambda(I) - \lambda(I)^2)$.

In the proof of our main result we need two well-known results from probability theory, namely Bernstein's and Hoeffding's inequality. Bernstein's inequality states that for $Z_1, \ldots, Z_N$ being i.i.d. random variables, satisfying $\mathbb{E}\, Z_n = 0$ and $|Z_n| \leq C$ a.s. for some $C > 0$,

$$\mathbb{P}\left(\left|\sum_{n=1}^{N} Z_n\right| > t\right) \leq 2\exp\left(-\frac{t^2}{2\left(\sum_{n=1}^{N} \mathbb{E}\, Z_n^2\right) + 2Ct/3}\right).$$

By applying this inequality to the random variables $\mathbf{1}_I(X_n) - \lambda(I)$, we obtain

$$\mathbb{P}\left(\left|\sum_{n=1}^{N} \mathbf{1}_I(X_n) - N\lambda(I)\right| > t\right) \leq 2\exp\left(-\frac{t^2}{2\left(N\lambda(I)\left(1-\lambda(I)\right)\right) + 2t/3}\right)$$

for $t > 0$. Using (4.7) we conclude

$$\mathbb{P}\left(\left|\sum_{n=1}^{N} \mathbf{1}_I(X_n) - N\lambda(I)\right| > t\right) \leq 2\exp\left(-\frac{t^2}{2N2^{-k}(1-2^{-k}) + 2t/3}\right) \quad \text{for} \quad k \geq 2. \tag{4.8}$$

For $k \in \{0,1\}$ we use Hoeffding's inequality, which yields

$$\mathbb{P}\left(\left|\sum_{n=1}^{N} \mathbf{1}_I(X_n) - N\lambda(I)\right| > t\right) \leq 2\exp\left(-\frac{2t^2}{N}\right). \tag{4.9}$$

4.3 Proof of Theorem 4.1

Since the theorem is trivial for $N < 32\left(s + \log\left((1-q)^{-1}\right)\right) < 5.70^2\left(s + \log\left((1-q)^{-1}\right)\right)$ we assume that $N \geq 32\left(s + \log\left((1-q)^{-1}\right)\right)$ and set

$$K = \left\lceil \frac{\log_2 N - \log_2\left(s + \log\left((1-q)^{-1}\right)\right)}{2} \right\rceil.$$

Then $K \geq 3$, and

$$2^{-K} \in \left[\frac{\sqrt{s + \log\left((1-q)^{-1}\right)}}{2\sqrt{N}}, \frac{\sqrt{s + \log\left((1-q)^{-1}\right)}}{\sqrt{N}}\right]. \tag{4.10}$$

Furthermore we have

$$\sqrt{sN} = N\frac{\sqrt{s}}{\sqrt{N}} \leq \frac{2^{-K+1}\sqrt{s}}{\sqrt{s + \log\left((1-q)^{-1}\right)}}N. \tag{4.11}$$

By choosing $t = c\sqrt{sN}$ for some $c > 0$, we conclude from (4.8), (4.9) and (4.11) that for any $c > 0$

$$\begin{aligned} &\mathbb{P}\left(\left|\sum_{n=1}^{N} \mathbf{1}_I(X_n) - N\lambda(I)\right| > c\sqrt{sN}\right) \\ \leq &\begin{cases} 2e^{-2c^2 s} & \text{for } k = 0,1 \\ 2\exp\left(-\frac{c^2 s}{2\cdot 2^{-k}(1-2^{-k}) + \frac{4c2^{-K}\sqrt{s}}{3\sqrt{s+\log\left((1-q)^{-1}\right)}}}\right) & \text{for } 2 \leq k \leq K. \end{cases} \end{aligned} \tag{4.12}$$

Let $B_k, k = 0, \ldots, K$ be given as

$$B_k = \bigcup_{I \in A_k}\left(\left|\sum_{n=1}^{N} \mathbf{1}_I(X_n) - N\lambda(I)\right| > c_k\sqrt{sN}\right). \tag{4.13}$$

The strategy of the proof is to find constants $c_k, k = 0, \ldots, K$ for which

$$\sum_{k=0}^{K} \mathbb{P}\left(B_k\right) < 1 - q$$

holds for any given q.

First we consider the case $k = 0$. By (4.6) we have that

$$|A_0| \leq (6e)^s.$$

We choose

$$\begin{aligned} c_0 &= \sqrt{\frac{1+\log(6)}{2} + \frac{\log\left(8(1-q)^{-1}\right)}{2s}} \\ &\leq \frac{1}{\sqrt{2}}\sqrt{4.88 + \frac{\log\left((1-q)^{-1}\right)}{s}}, \end{aligned} \tag{4.14}$$

thus together with (4.12) and (4.13) it follows that

$$\mathbb{P}(B_0) \leq |A_0| 2e^{-2c_0^2 s} = (6e)^s 2e^{-s(1+\log(6))}\frac{(1-q)}{8} = \frac{1-q}{4}.$$

Furthermore we get by (4.6) that

$$|A_1| \leq (10e)^s$$

and with

$$\begin{aligned} c_1 &= \sqrt{\frac{1+\log(10)}{2} + \frac{\log\left(8(1-q)^{-1}\right)}{2s}} \\ &\leq \frac{1}{\sqrt{2}}\sqrt{5.39 + \frac{\log\left((1-q)^{-1}\right)}{s}}, \end{aligned} \tag{4.15}$$

we obtain that

$$\mathbb{P}(B_1) \leq |A_1| 2e^{-2c_1^2 s} = (10e)^s 2e^{-s(1+\log(10))}\frac{(1-q)}{8} = \frac{1-q}{4}.$$

Next we consider the case $2 \leq k \leq K$. By (4.6), (4.12) and (4.13) we have

$$\mathbb{P}(B_k) \leq |A_k| \cdot 2 \cdot \exp\left(-\frac{c_k^2 s}{2 \cdot 2^{-k}(1-2^{-k}) + \frac{4c_k 2^{-K}\sqrt{s}}{3\sqrt{s+\log((1-q)^{-1})}}}\right). \tag{4.16}$$

We set

$$c_k = \sqrt{1 + \log(2(2^{k+1}+1)) + \frac{\log\left(2^{(k+1)}(1-q)^{-1}\right)}{s}}\sqrt{2 \cdot 2^{-k}(1-2^{-k}) + \frac{2.08 \cdot 4 \cdot 2^{-K}}{3}}$$

hence we get that

$$\begin{aligned} &\left|\frac{c_k\sqrt{s}}{\sqrt{s+\log\left((1-q)^{-1}\right)}}\right| \\ \leq &\left|\frac{\sqrt{1+\log(2(2^{k+1}+1)) + \log(2^{(k+1)}) + \frac{\log((1-q)^{-1})}{s}}\sqrt{2\cdot 2^{-k}(1-2^{-k}) + \frac{2.08\cdot 4\cdot 2^{-K}}{3}}}{\sqrt{1+\frac{\log((1-q)^{-1})}{s}}}\right| \end{aligned}$$

$$\leq \left| \sqrt{1+\log(2(2^{k+1}+1))+\log(2^{(k+1)})} \sqrt{2\cdot 2^{-k}(1-2^{-k})+\frac{2.08\cdot 4\cdot 2^{-K}}{3}} \right|$$
$$\leq 2.08$$

for $2 \leq k \leq K$. Thus by (4.16) we obtain

$$\begin{aligned}
\mathbb{P}(B_k) &\leq |A_k| \cdot 2 \cdot \exp\left(-\frac{c_k^2 s}{2\cdot 2^{-k}(1-2^{-k}) + \frac{4c_k 2^{-K}\sqrt{s}}{3\sqrt{s+\log((1-q)^{-1})}}} \right) \\
&\leq (2e)^s (2^{(k+1)}+1)^s \cdot 2 \cdot \exp\left(-s\left(1+\log(2(2^{(k+1)}+1))\right)\right) \frac{1-q}{2^{(k+1)}} \\
&= \frac{1-q}{2^k}.
\end{aligned}$$

Summing up the estimated probabilities gives

$$\sum_{k=0}^{K} \mathbb{P}(B_k) \leq \left(\frac{3}{4} + \sum_{k=3}^{K} 2^{-k} \right)(1-q) < 1-q.$$

Therefore with at least probability q, a realization $X_1(\omega), \ldots, X_n(\omega)$ is such that

$$\omega \notin \bigcup_{k=0}^{K} B_k.$$

We denote by z_n a point set which is defined by such a realization, i.e.

$$z_n = X_n(\omega), \quad 1 \leq n \leq N, \qquad \text{for some } \omega \notin \bigcup_{k=0}^{K} B_k.$$

Set

$$\lambda_k = \sqrt{2\cdot 2^{-k}(1-2^{-k}) + 2.08\cdot 4\cdot 2^{-K}/3}.$$

Then

$$\begin{aligned}
\sum_{k=2}^{K} c_k &= \sum_{k=2}^{K} \lambda_k \sqrt{1+\log(2(2^{k+1}+1)) + \frac{\log(2^{(k+1)}(1-q)^{-1})}{s}} \\
&\leq \sum_{k=2}^{K} \lambda_k \sqrt{1+\log 2+\log(2^{k+1}) + 0.12 + \frac{\log(2^{(k+1)})}{s} + \frac{\log\left((1-q)^{-1}\right)}{s}} \\
&\leq \sum_{k=2}^{K} \lambda_k \sqrt{1.12 + (2k+3)\log 2 + \frac{\log\left((1-q)^{-1}\right)}{s}} \\
&\leq \sqrt{1.12+7\log 2+\frac{\log\left((1-q)^{-1}\right)}{s}} \frac{1}{\sqrt{2}} \sum_{k=2}^{K} \sqrt{k}\lambda_k
\end{aligned}$$

$$\leq 3.28\sqrt{5.98 + \frac{\log\left((1-q)^{-1}\right)}{s}}. \tag{4.17}$$

Therefore we obtain by using (4.14), (4.15) and (4.17)

$$\begin{aligned}\sum_{k=0}^{K} c_k \leq & \frac{1}{\sqrt{2}}\sqrt{4.88 + \frac{\log\left((1-q)^{-1}\right)}{s}} + \frac{1}{\sqrt{2}}\sqrt{5.39 + \frac{\log\left((1-q)^{-1}\right)}{s}} \\ & + 3.28\sqrt{5.98 + \frac{\log\left((1-q)^{-1}\right)}{s}}\end{aligned} \tag{4.18}$$

Applying (4.5), (4.10), (4.18) and Jensen's inequality we obtain

$$\begin{aligned}\sum_{n=1}^{N} \mathbf{1}_{[0,x]}(z_n) &\leq \sum_{k=0}^{K}\sum_{n=1}^{N} \mathbf{1}_{\overline{[p_k(x),p_{k+1}(x)]}}(z_n) \\ &\leq N\lambda([0, w_K(x)]) + \sqrt{sN}\sum_{k=0}^{K} c_k \\ &\leq N\lambda([0,x]) + N\lambda(\overline{[x, w_K(x)]}) + \sqrt{sN}\sum_{k=0}^{K} c_k \\ &\leq N\lambda([0,x]) + N\frac{\sqrt{s + \log\left((1-q)^{-1}\right)}}{\sqrt{N}} + \sqrt{sN}\sum_{k=0}^{K} c_k \\ &\leq N\lambda([0,x]) + \sqrt{sN}\left(\sum_{k=0}^{K} c_k + \sqrt{1 + \frac{\log\left((1-q)^{-1}\right)}{s}}\right) \\ &\leq N\lambda([0,x]) + 5.70\sqrt{4.90 + \frac{\log\left((1-q)^{-1}\right)}{s}}\sqrt{sN}.\end{aligned}$$

Similarly a lower bound is given by

$$\begin{aligned}\sum_{n=1}^{N} \mathbf{1}_{[0,x]}(z_n) &\geq \sum_{k=0}^{K-1}\sum_{n=1}^{N} \mathbf{1}_{\overline{[p_k(x),p_{k+1}(x)]}}(z_n) \\ &\geq N\lambda([0, p_K(x)]) - \sqrt{sN}\sum_{k=0}^{K-1} c_k \\ &\geq N\lambda([0,x]) - N\lambda\left(\overline{[p_K(x), x]}\right) - \sqrt{sN}\sum_{k=0}^{K-1} c_k \\ &\geq N\lambda([0,x]) - 5.70\sqrt{4.90 + \frac{\log\left((1-q)^{-1}\right)}{s}}\sqrt{sN}.\end{aligned}$$

Combining the above bounds we finally arrive at

$$D_N^*(z_1, \ldots, z_n) \leq 5.70\sqrt{4.90 + \frac{\log\left((1-q)^{-1}\right)}{s}}\frac{\sqrt{s}}{\sqrt{N}}.$$

Part III

Monte Carlo methods in asset pricing

Chapter 5

Probabilistic error bounds for the discrepancy of mixed sequences

5.1 Introduction and statement of results

A common notion to measure the regularity of point distributions is the so-called star discrepancy. Roughly speaking, the star discrepancy compares the relative number of elements of a point set, which are contained in an axis-parallel box, to the volume of this box, and finally takes the maximal deviation over all possible boxes. The Quasi-Monte Carlo method for numerical integration is based on the fact that the difference of the integral of a function and the arithmetic mean of the function values at certain sampling points can be estimated by the product of the variation of this function and the discrepancy of the set of sampling points. Therefore point sets having a small star discrepancy can serve as a tool for numerical integration, a method which is frequently used for the evaluation of high-dimensional integrals in applied mathematics. Many constructions of low-discrepancy point sets only provide good bounds for the discrepancy if the number of points is large (in comparison with the dimension). This led to the development of the so-called randomized Quasi-Monte Carlo method, which tries to combine the advantages of the (deterministic) Quasi-Monte Carlo method and the advantages of the (random) Monte Carlo method. For an introduction to discrepancy theory and its applications in numerical mathematics we refer the reader to the books of Dick and Pillichshammer [26], Drmota and Tichy [30], Kuipers and Niederreiter [66] and Glasserman [42].

To formulate our results in a precise way we need some notation. We write $(x^{(1)}, \ldots, x^{(s)})$ for the coordinates of a point $x \in [0,1]^s$. We write $x \leq y$ if $x^{(i)} \leq y^{(i)}$ for $1 \leq i \leq s$. We write 0 and 1 for the points $(0, \ldots, 0)$ and $(1, \ldots, 1)$ in $[0,1]^s$. For $a \in [0,1]^s$ we define an s-dimensional interval $[0, a]$ as the set $\{x \in [0,1]^s : 0 \leq x \leq a\}$ (which is an s-dimensional axis-parallel box). Let $(x_1, \ldots, x_N)$ be a sequence of points in the s-dimensional unit cube. The star discrepancy D_N^* of $(x_1, \ldots, x_N)$ is defined as

$$D_N^*(x_1, \ldots, x_N) = \sup_{a \in [0,1]^s} \left| \frac{1}{N} \sum_{n=1}^{N} \mathbf{1}_{[0,a]}(x_n) - \lambda([0,a]) \right|.$$

Here and in the sequel λ denotes the Lebesgue measure. For simplicity we write $D_N^*(x_n)$ instead of $D_N^*(x_1, \ldots, x_N)$. If $(x_n)_{n\geq 1}$ is an infinite sequence, we write $D_N^*(x_n)$ for the discrepancy of the first N elements of $(x_n)_{n\geq 1}$.

The importance of discrepancy theory in numerical mathematics is based on the Koksma-Hlawka inequality, which states that for a sequence $(x_1, \ldots, x_N)$ of points in $[0,1]^s$ and a function f having total variation $\operatorname{Var} f$ on $[0,1]^s$ (in the sense of Hardy and Krause)

$$\left| \frac{1}{N} \sum_{n=1}^{N} f(x_n) - \int_0^1 f(x)\, dx \right| \leq D_N^*(x_n) \cdot \operatorname{Var} f.$$

There exist many constructions of so-called low-discrepancy sequences, i.e. sequences $(x_n)_{n\geq 1}$ for which

$$D_N^*(x_n) \ll (\log N)^s N^{-1} \qquad \text{as} \qquad N \to \infty \tag{5.1}$$

(this should be compared with a result of Roth [92] which states that every infinite sequence of points from $[0,1]^s$ has discrepancy $\gg (\log N)^{s/2} N^{-1}$ for infinitely many N; this has been slightly improved by Beck [11] and Bilyk, Lacey and Vagharshakyan [13], but the precise minimal asymptotic order of the discrepancy is still an open problem). Sequences of this type are only of practical use if the number of sampling points N is "large" in comparison with the dimension s; in particular the right-hand side of (5.1) is increasing for $N \leq e^s$. On the other hand, the so-called Monte Carlo method (which uses i.i.d. randomly generated points instead of deterministic points) gives an probabilistic bound of asymptotical order $N^{-1/2}$, independently of the dimension. This led to the development of randomized QMC integration schemes, which try to combine the advantages of (random) MC and (deterministic) QMC. There exist several methods for "randomizing" QMC rules; see for example Hickernell [54], Matoušek [76], Owen [86] and L'Ecuyer and Lemieux [70]. In this paper we consider s-dimensional sequences which are constructed by concatenating the coordinates of a d-dimensional QMC sequence and an $s-d$-dimensional MC sequence. Sequences of this type are called "mixed" sequences, and have been investigated e.g. by Spanier [96], Ökten [82, 83] and Roşca [91]. Extensive numerical experiments have been carried out by Ökten, Tuffin and Burago [85], who showed that the use of mixed sequences can significantly improve the efficienty of the QMC method in applications from financial mathematics.

Let $(q_n)_{n\geq 1}$ be a d-dimensional QMC sequence, and let $(X_n)_{n\geq 1}$ be a sequence of i.i.d. random variables having uniform distribution on $[0,1]^{s-d}$. We write $(x_n)_{n\geq 1}$ for the sequence which consists of the points $x_n = (q_n, X_n)$, i.e. $x_n = (q_n^{(1)}, \ldots, q_n^{(d)}, X_n^{(1)}, \ldots, X_n^{(s-d)})$ for $n \geq 1$. Ökten, Tuffin and Burago [85] showed that for such a sequence, under the additional assumption $D_N^*(q_n) \to 0$, for arbitrary $\varepsilon > 0$

$$\mathbb{P}\left(D_N^*(x_n) \leq D_N^*(q_n) + \varepsilon\right) \geq 1 - 2e^{-\varepsilon^2 N/2} \tag{5.2}$$

for sufficiently large N (in [85, Theorem 5] the exponent $-2\varepsilon^2 N$ appears, but as Gnewuch [45] remarks, the proof only gives $-\varepsilon^2 N/2$). Their paper contains no information on the size of the

values of N for which (5.2) holds. Gnewuch [45] showed that

$$\mathbb{P}\left(D_N^*(x_n) \leq D_N^*(q_n) + \varepsilon\right) \geq 1 - 2\mathcal{N}(s, \varepsilon/2)e^{-\varepsilon^2 N/2}, \tag{5.3}$$

where $\mathcal{N}(d, \delta)$ is defined as the smallest number M for which there exists a set Γ of M points in $[0,1]^s$ such that for all $y \in [0,1]^s$ there exist $x, z \in \Gamma \cup \{0\}$ such that $x \leq y \leq z$ and $\lambda([0,z]) - \lambda([0,x]) \leq \delta$ (the set Γ is called a *δ-cover* of $[0,1]^s$, and the number $\mathcal{N}$ the *covering number*). By [43, Theorem 1.15]

$$\mathcal{N}(s, \delta) \leq (2e)^s(\delta^{-1} + 1)^s,$$

and therefore (5.3) implies

$$\mathbb{P}\left(D_N^*(x_n) \leq D_N^*(q_n) + \varepsilon\right) > 1 - 2(2e)^s(2/\varepsilon + 1)^s e^{-\varepsilon^2 N/2}. \tag{5.4}$$

In dimension $s = 2$ Gnewuch [44] proved a stronger upper bound for covering numbers, and conjectured that in all dimensions

$$\mathcal{N}(s, \delta) \leq 2\delta^{-s} + o_s(\delta^{-s}).$$

(where o_s means that the implied constant may depend on s). This would lead to an improvement of (5.4).

We will also need the notion of *δ-bracketing covers*: Let $\delta \in (0,1]$. A finite set Δ of pairs of points from $[0,1]^s$ is called a δ-bracketing cover of $[0,1]^s$, if for every pair $(x,z) \in \Delta$ the estimate $\lambda([0,z]) - \lambda([0,x]) \leq \delta$ holds, and if for every $y \in [0,1]^s$ there exists a pair (x,z) from Δ such that $x \leq y \leq z$. The number $\mathcal{N}_{[\,]}(s,\delta)$, which is called the *bracketing number*, denotes the smallest cardinality of a δ-bracketing cover of $[0,1]^s$. By [43, Theorem 1.15]

$$\mathcal{N}_{[\,]}(s, \delta) \leq 2^{s-1}e^s(\delta^{-1} + 1)^s.$$

Gnewuch's result (5.3) has the advantage of being valid for all $N \geq 1$. However, (5.2) is asymptotically stronger than (5.3) (as N increases, for fixed ε). On the one hand, the purpose of this paper is to show an improved version of (5.2), which is almost optimal. On the other hand, we want to show that the factor ε^{-s} in (5.3) and (5.4), which essentially comes from the necessity to discretize the discrepancy with respect to a grid of precision ε, is not necessary and can be replaced by γ^s for an appropriate constant γ. This might be surprising at first sight: the impact of the necessity to discretize the discrepancy with respect to a certain (possibly extremely close-meshed) grid does not depend on the accuracy of this grid.

More precisely, we will prove the following theorem:

Theorem 5.1 *Let $(q_n)_{n\geq 1}$ be a d-dimensional sequence, and let $(X_n)_{n\geq 1}$ be a sequence of i.i.d. random variables having uniform distribution on $[0,1]^{s-d}$. Let $(x_n)_{n\geq 1}$ denote the mixed sequence which consists of the points $x_n = (q_n, X_n)$. Then for every $\eta > 0$ there exists a constant $\gamma = \gamma(\eta)$ such that for every $\varepsilon > 0$*

$$\mathbb{P}\left(D_N^*(x_n) \leq 2D_N^*(q_n) + \varepsilon\right) \geq 1 - \gamma^s e^{-2(1-\eta)\varepsilon^2 N}. \tag{5.5}$$

In (5.5) we can choose

$$\gamma = e^{2\cdot\lceil 4\log_2(3/\eta)+2\log_2 7\rceil}. \tag{5.6}$$

As a direct consequence of Theorem 5.1 we obtain the following corollary, which is an improvement of the result of Ökten, Tuffin and Burago (5.2).

Corollary 5.1 *Assume that $D_N^*(q_n) \to 0$, and let $\eta > 0$ be given. Then for arbitrary $\varepsilon > 0$*

$$\mathbb{P}\left(D_N^*(x_n) \le \varepsilon\right) \ge 1 - e^{-2(1-\eta)\varepsilon^2 N}$$

for sufficiently large N.

Proof of Corollary 5.1: Let $\eta > 0$ be given, and let $\hat{\eta}$ be so small that $(1-\hat{\eta})^3 > 1-\eta$. Since $D_N^*(q_n) \to 0$ we have $D_N^*(q_n) \le \hat{\eta}\varepsilon/2$ for sufficiently large N. Thus by Theorem 5.1

$$\begin{aligned}\mathbb{P}\left(D_N^*(x_n) \le \varepsilon\right) &\ge \mathbb{P}\left(D_N^*(x_n) \le 2D_N^*(q_n) + (1-\hat{\eta})\varepsilon\right)\\ &\ge 1 - \gamma(\hat{\eta})^s e^{-2(1-\hat{\eta})((1-\hat{\eta})\varepsilon)^2 N}\\ &\ge 1 - e^{-2(1-\eta)\varepsilon^2 N}\end{aligned}$$

for sufficiently large N. This proves the corollary. □

Remark 5.1 *Theorem 5.1 and Gnewuch's result (5.3) both give probability zero for*

$$\varepsilon \le s^{1/2}N^{-1/2}.$$

It is clear that a result like Theorem 5.1 can not give a positive probability for all possible $d \ge 1,\ s > d$ and $\varepsilon > 0$, since this would imply (by choosing $d = 1$ and $(q_1, \ldots, q_N)$ such that $D_N^*(q_n) = 1/N$) the existence of an s-dimensional sequence $(x_1, \ldots, x_N)$ with discrepancy $\le 2/N + \varepsilon$ for arbitrary s and N, which is in conflict with Roth's result. In fact the bound $s^{1/2}N^{-1/2}$ might be crucial: it is know that for all $N \ge 1$ and $s \ge 1$ there exists an N-element sequence having discrepancy $\le 10s^{1/2}N^{-1/2}$, but it is unknown how far this upper bound is from optimality. For more information we refer to [1], [52] and [55].

Remark 5.2 *Gnewuch [45, Remark 3.4] showed that in every bound of the form*

$$\mathbb{P}\left(D_N^*(x_n) \le D_N^*(q_n) + \varepsilon\right) \ge 1 - f(s,\varepsilon)e^{-\varepsilon^2 N/2}$$

the function $f(s,\varepsilon)$ has to grow at least exponentially in s (this follows from a general result of Heinrich, Novak, Wasilkowski and Woźniakowski [52]). Using exactly the same argument it can be easily shown that every function $f(s)$ replacing the factor γ^s in our Theorem 5.1 (for some fixed η) has to grow at least exponentially in s. Thus the only possible improvement of Theorem 5.1 with respect to s is a reduction of the base γ of the term γ^s.

Remark 5.3 *For any dimensions $d \geq 1$ and $s > d$ it is impossible to find constants $\eta > 0$ and $\gamma > 0$ such that for arbitrary $\varepsilon > 0$*

$$\mathbb{P}\left(D_N^*(x_n) \leq 2D_N^*(q_n) + \varepsilon\right) \geq 1 - \gamma^s e^{-2(1+\eta)\varepsilon^2 N}$$

for sufficiently large N. Thus the exponent $2(1-\eta)\varepsilon^2 N$ in Theorem 5.1 can not be improved to $2(1+\eta)\varepsilon^2 N$ (a proof of this remark will be given at the end of this paper).

Remark 5.4 *Our corollary shows that it is possible to obtain an asymptotic order of $e^{-2(1-\eta)\varepsilon^2 N}$ (for ε fixed, as $N \to \infty$) for arbitrarily small $\eta > 0$. However, as η gets smaller the necessary value of the constant γ in (5.5) and (5.6) increases , and in particular $\gamma \to \infty$ as $\eta \to 0$. We are not able to decide whether it is possible to improve Theorem 5.1 to*

$$\mathbb{P}\left(D_N^*(x_n) \leq 2D_N^*(q_n) + \varepsilon\right) \geq 1 - \gamma^s e^{-2\varepsilon^2 N}$$

for some constant γ. Summarizing these results, we know for every $\eta > 0$ that an asymptotic order of $e^{-2(1-\eta)\varepsilon^2 N}$ is possible and $e^{-2(1+\eta)\varepsilon^2 N}$ is impossible, while the "critical" case $e^{-2\varepsilon^2 N}$ remains open.

Remark 5.5 *There are two differences between Theorem 5.1 and Gnewuch's result (5.3). On the one hand, our bound for the discrepancy is $2D_N^*(q_n) + \varepsilon$ instead of $D_N^*(q_n) + \varepsilon$. The additional term $D_N^*(q_n)$ comes from the interval partitioning method which is used in our proof, and it seems that this extra term can not be avoided. In applications this should not cause problems, since the deterministic sequence $(q_n)_{1 \leq n \leq N}$ is chosen in such a way that $D_N^*(q_n)$ is very small, whereas ε can not be arbitrarily small (see Remark 1). On the other hand, we can avoid the factor ε^{-s} from Gnewuch's result, which can have a significant contribution particularly for large values of s.*

5.2 Preliminaries

We will use Hoeffding's inequality and Bernstein's inequality, two classical inequalities from probability theory.

Hoeffding's inequality: For $Z_1, \ldots, Z_N$ being independent random variables, satisfying $a \leq |Z_n| \leq b$ a.s. for some $a < b,\ b - a \leq 1$,

$$\mathbb{P}\left(\left|\sum_{n=1}^{N}(Z_n - \mathbb{E}\, Z_n)\right| > t\right) \leq 2e^{-2t^2}.$$

Bernstein's inequality: For $Z_1, \ldots, Z_N$ being independent random variables, satisfying $|Z_n - \mathbb{E}\, Z_n| \leq 1$ a.s.,

$$\mathbb{P}\left(\left|\sum_{n=1}^{N}(Z_n - \mathbb{E}\, Z_n)\right| > t\right) \leq 2\exp\left(-\frac{t^2}{2\left(\sum_{n=1}^{N}\mathbb{E}\, Z_n^2\right) + 2t/3}\right).$$

The following lemma will be needed for the proof of Remark 3:

Lemma 5.1 *Let $(Z_n)_{n\geq 1}$ be independent, fair Bernoulli random variables. Let $\eta > 0$ be given. Then there exists an $\varepsilon_0 = \varepsilon_0(\eta)$ such that for every fixed $\varepsilon \in (0, \varepsilon_0)$ for all sufficiently large N*

$$\mathbb{P}\left(\sum_{n=1}^{N} Z_n \geq N/2 + \varepsilon N\right) \geq e^{-2\varepsilon^2(1+\eta)N}.$$

Proof: To simplify notations we assume w.l.o.g. that εN is an integer. Let η be given, and set

$$p = \mathbb{P}\left(\sum_{n=1}^{N} Z_n \geq N/2 + \varepsilon N\right).$$

By Taylor's formula we have for sufficiently small ε

$$\log(1/2 + \varepsilon) \geq \log 1/2 + 2\varepsilon - (1+\eta)\varepsilon^2$$

and

$$\log(1/2 - \varepsilon) \geq \log 1/2 - 2\varepsilon - (1+\eta)\varepsilon^2.$$

By Stirling's formula for sufficiently large N

$$\begin{aligned}
\binom{N}{N/2+\varepsilon N} &\geq \frac{1}{2}\frac{N^N}{\sqrt{2\pi N}(N/2+\varepsilon N)^{(N/2+\varepsilon N)}(N/2-\varepsilon N)^{(N/2-\varepsilon N)}} \\
&\geq \left(e^{-\eta\varepsilon^2}\right)^N \left(\frac{N^N}{(N/2+\varepsilon N)^{(N/2+\varepsilon N)}(N/2-\varepsilon N)^{(N/2-\varepsilon N)}}\right)
\end{aligned}$$

and therefore, also for sufficiently large N,

$$\begin{aligned}
p^{1/N} &= \left(\sum_{k=N/2+\varepsilon N}^{N} \binom{N}{k}\frac{1}{2^N}\right)^{1/N} \\
&\geq \left(\binom{N}{N/2+\varepsilon N}\frac{1}{2^N}\right)^{1/N} \\
&\geq e^{-\eta\varepsilon^2}\left(\frac{N^N}{(N/2+\varepsilon N)^{(N/2+\varepsilon N)}(N/2-\varepsilon N)^{(N/2-\varepsilon N)}2^N}\right)^{1/N} \\
&= e^{-\eta\varepsilon^2}\left(\frac{1}{(1/2+\varepsilon)^{(1/2+\varepsilon)}(1/2-\varepsilon)^{(1/2-\varepsilon)}2}\right) \\
&\geq e^{-\eta\varepsilon^2}\exp\Big(-(1/2+\varepsilon)\log(1/2+\varepsilon) - (1/2-\varepsilon)\log(1/2-\varepsilon) - \log 2\Big) \\
&\geq e^{-\eta\varepsilon^2}\exp\Big(-(1/2+\varepsilon)(\log 1/2 + 2\varepsilon - (1+\eta)\varepsilon^2) \\
&\qquad -(1/2-\varepsilon)(\log 1/2 - 2\varepsilon - (1+\eta)\varepsilon^2) - \log 2\Big) \\
&= e^{-\eta\varepsilon^2}\exp\Big(-(2+\eta)\varepsilon^2\Big).
\end{aligned}$$

Thus for sufficiently large N

$$p \geq \exp\Big(-2(1+\eta)\varepsilon^2 N\Big). \qquad \square$$

5.3 Proof of Theorem 5.1

We use a refined version of the dyadic partitioning technique in [1]. Let $N \geq 1, \varepsilon > 0, \eta > 0$ and a parameter $\mu \geq 10$ be given (μ will be chosen as a function of η, see equation (5.26) below). For simplicity we assume that μ is an integer.
Let $(q_n)_{n\geq 1}$ be a d-dimensional sequence, and write D for the (d-dimensional) star discrepancy of $(q_n)_{1\leq n\leq N}$. Let $X_1, \ldots, X_N$ be i.i.d. random variables defined on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$, having uniform distribution on $[0,1]^{s-d}$, and write $(x_n)_{1\leq n\leq N}$ for the mixed sequence which consists of the s-dimensional points $x_n = (q_n, X_n)$. We will use the estimate

$$(2e)^s(2^k+1)^s \leq e^{(k-1)s}, \tag{5.7}$$

which holds for all $k \geq \mu$ (since we assumed $\mu \geq 10$, and since of course $s \geq 2$).

Assume now that

$$\varepsilon \geq 2^{-\mu}, \tag{5.8}$$

and let Γ be a $2^{-2\mu}$-cover of $[0,1]^s$ for which

$$\#\Gamma \leq (2e)^s(2^{2\mu}+1)^s \leq e^{(2\mu-1)s} \leq \frac{e^{2\mu s}}{2}.$$

Then, using Gnewuch's method from [45] and Hoeffding's inequality we can easily show that

$$\begin{aligned}\mathbb{P}\left(D_N^*(x_n) \leq D + \varepsilon + 2^{-2\mu}\right) &\geq 1 - 2e^{-2\varepsilon^2 N}(\#\Gamma)\\ &\geq 1 - e^{2\mu s}e^{-2\varepsilon^2 N},\end{aligned}$$

which by (5.8) implies

$$\mathbb{P}\left(D_N^*(x_n) \leq D + \varepsilon + \varepsilon 2^{-\mu}\right) \geq 1 - e^{2\mu s}e^{-2\varepsilon^2 N}. \tag{5.9}$$

For the rest of the proof we assume that instead of (5.8)

$$\varepsilon \leq 2^{-\mu} \tag{5.10}$$

holds (which is the much more difficult case). Additionally we assume that

$$\varepsilon > \frac{\sqrt{\mu s}}{\sqrt{2N}}. \tag{5.11}$$

(this additional assumption will be dropped later). Let

$$K = K(\varepsilon) := \min\left\{k \geq 1 :\ 2^{-k/2}k^{-1/2} \leq \varepsilon\right\}.$$

Then

$$2^{-K/2}K^{-1/2} \leq \varepsilon \leq 2 \cdot 2^{-K/2}K^{-1/2}, \tag{5.12}$$

and $\mu \geq 10$ implies

$$K \geq \mu + 15 \geq 25. \tag{5.13}$$

By (5.12) and (5.13) we have $2^{-K} \geq \varepsilon^2$. Thus by (5.10) and (5.11)

$$K \leq \log_2(\varepsilon^{-2}) \leq \log_2(2N/s\mu) \leq \log_2(N/10) \leq 2N^{1/4} \leq \varepsilon^{3/2}N \leq 2^{-\mu/2}\varepsilon N. \tag{5.14}$$

For $\mu \leq k \leq K-1$ let Γ_k be a 2^{-k}-cover of $[0,1]^s$, for which

$$\#\Gamma_k \leq (2e)^s(2^k+1)^s. \tag{5.15}$$

Let Δ_K denote a 2^{-K}-bracketing cover of $[0,1]^s$ for which

$$\#\Delta_K \leq (2e)^s(2^k+1)^s. \tag{5.16}$$

Such sets Γ_k and Δ_K exist by a result of Gnewuch [43, Theorem 1.15]. For notational convenience we also define

$$\Gamma_K = \{x \in [0,1]^s : \ (x,y) \in \Delta_K \text{ for some } y \in [0,1]^s\}$$

and

$$\Gamma_{K+1} = \{y \in [0,1]^s : \ (x,y) \in \Delta_K \text{ for some } x \in [0,1]^s\}$$

For every $x \in [0,1]^s$ there exists a pair $(p_K, p_{K+1}) = (p_K(x), p_{K+1}(x))$ for which $(p_K, p_{K+1}) \in \Delta_K$ such that $p_K \leq x \leq p_{K+1}$ and

$$\lambda([0,p_{K+1}]) - \lambda([0,p_K]) \leq \frac{1}{2^K}. \tag{5.17}$$

For every $x \in [0,1]^s$ and $k = K, K-1, \ldots, \mu+1$ we can recursively determine points $p_{k-1} = p_{k-1}(x) \in \Gamma_{k-1} \cup \{0\}$, such that $p_{k-1}(x) \leq p_k(x)$ and

$$\lambda([0,p_k]) - \lambda([0,p_{k-1}]) \leq \frac{1}{2^{k-1}}.$$

For notational convenience we also define

$$p_{\mu-1} = 0.$$

We define for $x, y \in [0,1]^s$

$$\overline{[x,y]} := \begin{cases} [0,y]\backslash[0,x] & \text{if } x \neq 0 \\ [0,y] & \text{if } x = 0,\ y \neq 0. \\ \emptyset & \text{if } x = y = 0. \end{cases}$$

Then the sets

$$I_k(x) := \overline{[p_k(x), p_{k+1}(x)]}, \quad \mu - 1 \leq k \leq K,$$

are disjoint, we have

$$\bigcup_{k=\mu-1}^{K-1} I_k(x) \subset [0,x] \subset \bigcup_{k=\mu-1}^{K} I_k(x),$$

and for all $x, y \in [0,1]^s$

$$\sum_{k=\mu-1}^{K-1} \mathbf{1}_{I_k(x)}(y) \leq \mathbf{1}_{[0,x]}(y) \leq \sum_{k=\mu-1}^{K} \mathbf{1}_{I_k(x)}(y). \tag{5.18}$$

For $\mu - 1 \leq k \leq K$ we write A_k for the set of all sets of the form $I_k(x)$, where x can take any possible value from $[0,1]^s$. Then by (5.7), (5.15) and (5.16), A_k contains at most

$$\#\Gamma_{k+1} \leq e^{ks} \tag{5.19}$$

elements. All elements of A_k, where $\mu \leq k \leq K$, have Lebesgue measure bounded by 2^{-k}. The elements of $A_{\mu-1}$ can have Lebesgue measure between 0 and 1.

For any $k \in \{\mu, \ldots, K+1\}$ we will represent the numbers $p_k \in \Gamma_k$ in the form (u_k, v_k), where $u_k \in [0,1]^d$ and $v_k \in [0,1]^{s-d}$, such that p_k has the coordinates $(u_k^{(1)}, \ldots, u_k^{(d)}, v_k^{(1)}, \ldots, v_k^{(s-d)})$. We write U_k and V_k for the intervals $[0, u_k]$ and $[0, v_k]$, and (U_k, V_k) for the sets $U_k \times V_k = [0, p_k]$. Every $x \in [0,1]^s$ uniquely determines points $p_k \in \Gamma_k,\ \mu \leq k \leq K+1$, and hence the according values of I_k, u_k, v_k, U_k, V_k are also uniquely defined.

For two sets $I_{k-1} \in A_{k-1}$ and $I_k \in A_k$ we write $I_{k-1} \prec I_k$ if there exists an $x \in [0,1]^s$ such that $I_k = I_k(x)$ and $I_{k-1} = I_{k-1}(x)$. For every $I_k \in A_k,\ \mu \leq k \leq K$ there exists exactly one element I_{k-1} of A_{k-1} for which $I_{k-1} \prec I_k$. Every $I_k \in A_k,\ \mu \leq k \leq K$ uniquely determines sets $I_{\mu-1}, \ldots, I_{k-1}$ such that $I_{\mu-1} \prec \cdots \prec I_{k-1} \prec I_k$. Whenever I_k is fixed we will write $I_{\mu-1} \ldots, I_{k-1}$ for these sets, which are uniquely determined, and $p_l, u_l, v_l, U_l, V_l,\ \mu \leq l \leq k-1$ for the according values, which are also uniquely determined.

Every $I_k \in A_k,\ \mu \leq k \leq K$, is of the form

$$(U_{k+1}, V_{k+1}) \backslash (U_k, V_k) = ((U_{k+1} \backslash U_k) \times V_{k+1}) \cup (U_k \times (V_{k+1} \backslash V_k)).$$

Every $I_{\mu-1} \in A_{\mu-1}$ is of the form $[0, p_\mu] = (U_\mu, V_\mu)$.

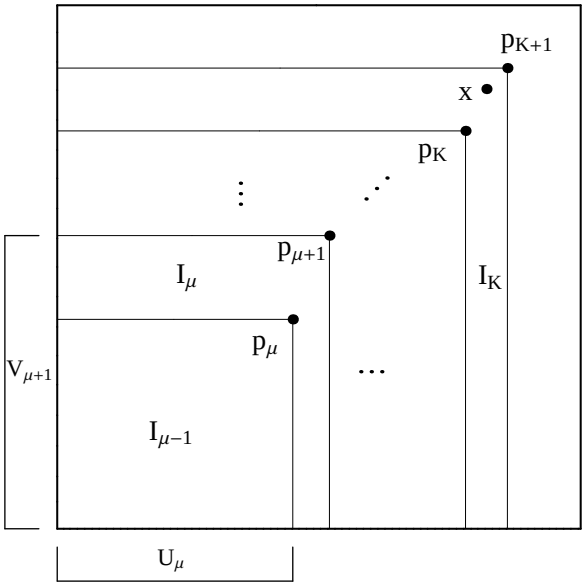

Figure 5.1: An illustration of our construction in the case $d = 1$, $s = 2$. A point $x \in [0,1]^2$ is given and determines points $p_\mu, p_{\mu+1}, \ldots, p_{K+1}$ and sets $I_{\mu-1} \prec I_\mu \prec \cdots \prec I_K$. For exemplification we have also marked the sets U_μ and $V_{\mu+1}$. Every set $I_k,\ \mu \le k \le K$, is of the form $(U_{k+1}, V_{k+1})\backslash(U_k, V_k) = ((U_{k+1}\backslash U_k) \times V_{k+1}) \cup (U_k \times (V_{k+1}\backslash V_k))$.

Step by step we construct a function $S(I)$ for intervals I from $A_{\mu-1}, \ldots, A_K$, such that for every I the function value $S(I)$ is a subset of $\{1, \ldots, N\}$ (we explain the necessity of this function S in the footnote[1]).
Firstly, let $I_{\mu-1} \in A_{\mu-1}$. Then $I_{\mu-1}$ is of the form (U_μ, V_μ), and we can find $\lceil N\lambda(U_\mu) - ND \rceil$ indices n from $\{1, \ldots, N\}$ for which $q_n \in U_\mu$. This is possible since the discrepancy of $(q_n)_{1 \le n \le N}$ is bounded by D, and hence the interval U_μ of Lebesgue measure $\lambda(U_\mu)$ contains at least $\lceil N\lambda(U_\mu) - ND \rceil$ points of $(q_n)_{1 \le n \le N}$. Denote this set of indices by $S(I_\mu)$.
In the next step let I_μ denote an element of A_μ. Then I_μ is of the form $(U_\mu, V_\mu)\backslash I_{\mu-1}$, where $I_{\mu-1} \in A_{\mu-1}$ and $I_{\mu-1} \prec I_\mu$. We can find $\lfloor N\lambda(U_{\mu+1}\backslash U_\mu) \rfloor$ indices n which are not contained in $S(I_{\mu-1})$ but for which $q_n \in U_{\mu+1}$. We write $S(I_{\mu+1})$ for this set of indices.

[1] Our proof is based on the decomposition of the unit cube into parts, and the fact that an arbitrary interval can be written as an union of sets of quickly decreasing Lebesgue measure. However, in our situation this method can only be directly applied if the number of elements of $(q_n)_{1 \le n \le N}$ in a subset U of $[0,1]^d$ is $\approx \lambda(U)N$. Unfortunately, this is not necessarily the case: the sets U we consider can be written in the form $U^+\backslash U^-$ for some axis-parallel boxes U^+ and U^-. Thus, if the discrepancy D of $(q_n)_{1 \le n \le N}$ is large in comparison with $\lambda(U)$, the number of elements of $(q_n)_{1 \le n \le N}$, which are contained in U (which can be any number from $[N\lambda(U) - 2ND, N\lambda(U) + 2ND]$) can be much larger than $N\lambda(U)$ (and this may hold not only for one, but for several of the sets which we need in our decomposition!). To solve this problem, we distribute the indices $\{1, \ldots, N\}$ to the sets in our decomposition in an appropriate regular way, instead of assigning them directly to the sets to which they actually belong.

Generally, assume that the function S is defined for all intervals in A_k for $k = \mu-1, \mu, \ldots, m$ for some m. Let I_{m+1} denote an element of A_{m+1}. Then I_{m+1} is of the form $(U_{m+2}, V_{m+2})\backslash(I_{\mu-1} \cup I_\mu \cup \cdots \cup I_m)$, where $I_k \in A_k$ for $k = \mu-1, \ldots, m$ and $I_{\mu-1} \prec \cdots \prec I_m \prec I_{m+1}$. We can find $\lfloor N\lambda(U_{m+2}\backslash U_{m+1})\rfloor$ indices n which are not contained in $\bigcup_{k=\mu-1}^{m} S(I_k)$, but for which $q_n \in U_{m+2}$. We write $S(I_{k+1})$ for this set of indices.
Proceeding in this way we define the function S for all elements of $A_{\mu-1}, \ldots, A_K$.

Additionally we define for every $I_k \in A_k,\ \mu \le k \le K$,

$$R(I_k) = S(I_{\mu-1}) \cup \cdots \cup S(I_{k-1}),$$

where $I_{\mu-1} \prec \cdots \prec I_k$.

Then

$$\begin{aligned}
\#\bigcup_{k=\mu-1}^{K-1} S(I_k) &\ge \lceil N\lambda(U_\mu) - ND\rceil + \sum_{k=\mu}^{K-1} \lfloor N\lambda(U_{k+1}\backslash U_k)\rfloor \\
&\ge N\lambda\left(\bigcup_{k=\mu}^{K-1} U_{k+1}\backslash U_k\right) - ND - (K-\mu) \\
&= N\lambda(U_K) - ND - (K-\mu) \\
&\ge \sum_{n=1}^{N} \mathbf{1}_{U_K}(q_n) - 2ND - (K-\mu),
\end{aligned}$$

and accordingly

$$\#\bigcup_{k=\mu-1}^{K} S(I_k) \ge \sum_{n=1}^{N} \mathbf{1}_{U_{K+1}}(q_n) - 2ND - (K+1-\mu).$$

Thus

$$\begin{aligned}
&\sum_{n=1}^{N} \mathbf{1}_{[0,x]}(x_n) \\
\ge\ &\sum_{n=1}^{N} \mathbf{1}_{[0,p_K]}(x_n) \\
=\ &\sum_{n=1}^{N} \mathbf{1}_{U_K}(q_n)\cdot \mathbf{1}_{V_K}(X_n) \\
=\ &\sum_{n=1}^{N} \mathbf{1}_{U_\mu}(q_n)\cdot \mathbf{1}_{V_\mu}(X_n) + \sum_{k=\mu}^{K-1}\sum_{n=1}^{N} \left(\mathbf{1}_{U_{k+1}\backslash U_k}(q_n)\cdot \mathbf{1}_{V_{k+1}}(X_n) + \mathbf{1}_{U_k}(q_n)\cdot \mathbf{1}_{V_{k+1}\backslash V_k}(X_n)\right) \\
\ge\ &\sum_{n\in S(I_{\mu-1})} \mathbf{1}_{V_\mu}(X_n) + \sum_{k=\mu}^{K-1}\left(\sum_{n\in S(I_k)} \mathbf{1}_{V_{k+1}}(X_n) + \sum_{n\in R(I_k)} \mathbf{1}_{V_{k+1}\backslash V_k}(X_n)\right) \qquad (5.20)
\end{aligned}$$

and

$$\begin{aligned}
&\sum_{n=1}^{N} \mathbf{1}_{[0,x]}(x_n) \\
\leq\ & \sum_{n=1}^{N} \mathbf{1}_{U_{K+1}}(q_n) \cdot \mathbf{1}_{V_{K+1}}(X_n) \\
\leq\ & \left(\sum_{n \in \bigcup_{k=\mu-1}^{K} S(I_k)} \mathbf{1}_{[0,p_{K+1}]}(x_n) \right) + 2ND + (K+1-\mu) \\
\leq\ & \sum_{n \in S(I_{\mu-1})} \mathbf{1}_{V_\mu}(X_n) + \sum_{k=\mu}^{K} \left(\sum_{n \in S(I_k)} \mathbf{1}_{V_{k+1}}(X_n) + \sum_{n \in R(I_k)} \mathbf{1}_{V_{k+1} \setminus V_k}(X_n) \right) \\
& \quad +2ND + (K+1-\mu). \qquad (5.21)
\end{aligned}$$

Let $I_{\mu-1} \in A_{\mu-1}$ and define

$$Z = Z(I_{\mu-1}) = \sum_{n \in S(I_{\mu-1})} \mathbf{1}_{V_\mu}(X_n).$$

Then by Hoeffding's inequality

$$\mathbb{P}\left(|Z - \mathbb{E}\, Z| > \varepsilon N\right) \leq 2e^{-2\varepsilon^2 N}. \qquad (5.22)$$

Now assume that $I_k \in A_k$ for some $k,\ \mu \leq k \leq K$. Then the random variable

$$Z = Z(I_k) \quad = \quad \sum_{n \in S(I_k)} \mathbf{1}_{V_{k+1}}(X_n) + \sum_{n \in R(I_k)} \mathbf{1}_{V_{k+1} \setminus V_k}(X_n)$$

(which is a sum of independent random variables) has expected value

$$\sum_{n \in S(I_k)} \lambda(V_{k+1}) + \sum_{n \in R(I_k)} \lambda(V_{k+1} \setminus V_k)$$

and variance

$$\begin{aligned}
&\sum_{n \in S(I_k)} \lambda(V_{k+1})(1 - \lambda(V_{k+1})) + \sum_{n \in R(I_k)} \lambda(V_{k+1} \setminus V_k)(1 - \lambda(V_{k+1} \setminus V_k)) \\
\leq\ & \sum_{n \in S(I_k)} \lambda(V_{k+1}) + \sum_{n \in R(I_k)} \lambda(V_{k+1} \setminus V_k) \\
\leq\ & \lambda(V_{k+1}) \cdot \#S(I_k) + \lambda(V_{k+1} \setminus V_k) \cdot \#R(I_k) \\
\leq\ & N\lambda(U_{k+1} \setminus U_k)\lambda(V_{k+1}) + N\lambda(U_k)\lambda(V_{k+1} \setminus V_k) \\
=\ & N\lambda(I_k) \\
\leq\ & N2^{-k}.
\end{aligned}$$

We apply Bernstein's inequality and obtain for $t > 0$

$$\mathbb{P}\left(|Z - \mathbb{E}\, Z| > t\right) \leq 2\exp\left(-\frac{t^2}{2^{-k+1}N + 2t/3}\right). \tag{5.23}$$

If we let

$$t = \frac{6k^{1/2}\varepsilon N}{5 \cdot 2^{k/2}},$$

then by (5.12) we have

$$2t/3 \leq \frac{24N}{15 \cdot 2^k},$$

and therefore

$$\begin{aligned} \mathbb{P}\left(|Z - \mathbb{E}\, Z| > t\right) &\leq 2\exp\left(-\frac{36k\varepsilon^2 N}{25\,(2 + 24/15)}\right) \\ &= 2e^{-2k\varepsilon^2 N/5}. \end{aligned} \tag{5.24}$$

Let

$$B_{\mu-1} = \bigcup_{I \in A_{\mu-1}} \left(|Z(I_{\mu-1}) - \mathbb{E}\, Z(I_{\mu-1}| > \varepsilon N\right)$$

Then by (5.19) and (5.22) we have

$$\mathbb{P}(B_{\mu-1}) \leq 2e^{-2\varepsilon^2 N/5} e^{\mu s}.$$

For $\mu \leq k \leq K$ define

$$B_k = \bigcup_{I_k \in A_k} \left(|Z(I_k) - \mathbb{E}\, Z(I_k)| > \frac{6k^{1/2}\varepsilon N}{5 \cdot 2^{k/2}}\right).$$

Then by (5.19) and (5.24), and since $\varepsilon^2 N > \mu s/2 \geq 5s$,

$$\sum_{k=\mu}^{K} \mathbb{P}(B_k) \leq \sum_{k=\mu}^{K} 2e^{-2k\varepsilon^2 N/5} e^{ks} \leq \sum_{k=\mu}^{K} 2e^{-k\varepsilon^2 N/5} \leq 3e^{-\mu\varepsilon^2 N/5} \leq 3e^{-2\varepsilon^2 N}.$$

Overall we have

$$\mathbb{P}\left(\bigcup_{k=\mu-1}^{K} B_k\right) \leq 3e^{-2\varepsilon^2 N} + 2e^{-2\varepsilon^2 N} e^{3\mu s} \leq 3e^{-2\varepsilon^2 N} e^{\mu s}.$$

Thus by (5.12), (5.14), (5.17) and (5.21) we have with probability at least $1 - 3e^{-2\varepsilon^2 N} e^{\mu s}$ for all $x \in [0,1]^s$

$$\sum_{n=1}^{N} \mathbf{1}_{[0,x]}(x_n)$$

$$
\begin{aligned}
&\leq \sum_{n\in S(I_{\mu-1})} \mathbf{1}_{V_\mu}(X_n) + \sum_{k=\mu}^{K} \left(\sum_{n\in S(I_k)} \mathbf{1}_{V_{k+1}}(X_n) + \sum_{n\in R(I_k)} \mathbf{1}_{V_{k+1}\setminus V_k}(X_n) \right) \\
&\qquad +2ND + (K+1-\mu) \\
&\leq \mathbb{E}\left(\sum_{n\in S(I_{\mu-1})} \mathbf{1}_{V_\mu}(X_n) + \sum_{k=\mu}^{K} \left(\sum_{n\in S(I_k)} \mathbf{1}_{V_{k+1}}(X_n) + \sum_{n\in R(I_k)} \mathbf{1}_{V_{k+1}\setminus V_k}(X_n) \right) \right) \\
&\qquad +\varepsilon N \left(1 + \sum_{k=\mu}^{K} \frac{6k^{1/2}}{5\cdot 2^{k/2}} \right) + 2ND + 2^{-\mu/2}\varepsilon N \\
&= \mathbb{E}\left(\sum_{n\in \bigcup_{k=\mu-1}^{K} S(I_k)} \mathbf{1}_{V_K}(X_n) \right) + \varepsilon N \left(1 + 2^{-\mu/2} + \sum_{k=\mu}^{K} 4k^{1/2}2^{-k/2} \right) + 2ND \\
&\leq N\lambda([0,p_{K+1}]) + \varepsilon N \left(1 + 2^{-\mu/2} + \sum_{k=\mu}^{K} \frac{6k^{1/2}}{5\cdot 2^{k/2}} \right) + 2ND + (K+1-\mu) \\
&\leq N\lambda([0,x]) + N2^{-K} + \varepsilon N \left(1 + 2^{-\mu/2} + \sum_{k=\mu}^{K} \frac{6k^{1/2}}{5\cdot 2^{k/2}} \right) + 2ND) \\
&\leq N\lambda([0,x]) + \varepsilon N \left(1 + K^{1/2}2^{-K/2} + 2^{-\mu/2} + \underbrace{\sum_{k=\mu}^{K} \frac{6k^{1/2}}{5\cdot 2^{k/2}}}_{\leq 5\mu^{1/2}2^{-\mu/2}} \right) + 2ND \\
&\leq N\lambda([0,x]) + \varepsilon N \left(1 + 7\mu^{1/2}2^{-\mu/2} \right) + 2ND
\end{aligned}
$$

Similarly by (5.12), (5.14), (5.17) and (5.20) we have with probability at least $1 - 3e^{-2\varepsilon^2 N}e^{\mu s}$ for all $x \in [0,1]$

$$
\sum_{n=1}^{N} \mathbf{1}_{[0,x]}(x_n) \quad \geq \quad N\lambda([0,x]) - \varepsilon N \left(1 + 7\mu^{1/2}2^{-\mu/2} \right) - 2ND.
$$

Therefore we have, with probability at least $1 - 3e^{-2\varepsilon^2 N}e^{\mu s}$,

$$
D_N^*(x_n) \quad \leq \quad 2D + \varepsilon \left(1 + 7\mu^{1/2}2^{-\mu/2} \right). \tag{5.25}
$$

This holds under assumptions (5.10) and (5.11). Now it is easy to see that (5.25) also holds without assuming (5.11), since in this case $1 - 3e^{-2\varepsilon^2 N}e^{\mu s} \leq 0$ (cf. Remark 1). Comparing this result with (5.9), which holds under assumption (5.8) we see that (5.25) holds with probability greater than or equal to

$$
1 - e^{-2\varepsilon^2 N}e^{2\mu s}.
$$

Now let η be given. Set

$$\mu = \lceil 4\log_2(3/\eta) + 2\log_2 7 \rceil \tag{5.26}$$

and

$$\gamma = \gamma(\eta) = e^{2\mu} = e^{2\cdot\lceil 4\log_2(3/\eta)+2\log_2 7\rceil}.$$

Then $\mu \geq 10$. Some calculations show that for $y \in (0,1]$

$$\sqrt{4\log_2(3/y) + 2\log_2 7} \leq \frac{4}{y}$$

and consequently

$$\begin{aligned} & \left(1 + 7\sqrt{4\log_2(3/y) + 2\log_2 7}\cdot 2^{-(4\log_2(3/y)+2\log_2 7)/2}\right)^{-2} \\ \geq\ & \left(1 + 7\cdot 4\cdot y\cdot y^2/9\cdot 7^{-1}\right)^{-2} \\ \geq\ & (1+y^2/2)^{-2} \\ \geq\ & 1-y. \end{aligned} \tag{5.27}$$

Thus by (5.25) and (5.27) for $\varepsilon > 0$

$$\begin{aligned} \mathbb{P}\left(D_N^*(x_n) > D + \varepsilon\right) &\leq e^{2\mu s}\exp\left(-2\varepsilon^2 N\left(1+7\mu^{1/2}2^{-\mu/2}\right)^{-2}\right) \\ &\leq \gamma^s e^{-2(1-\eta)\varepsilon^2 N}, \end{aligned}$$

which proves the theorem. □

In conclusion we prove Remark 3 on the asymptotic optimality of the probability estimate

$$1 - \gamma(\eta)^s e^{-2(1-\eta)\varepsilon^2 N}.$$

We show that this lower bound can not be replaced by

$$1 - \gamma(\eta)^s e^{-2(1+\eta)\varepsilon^2 N}$$

for any positive η, no matter how large the constant $\gamma(\eta)$ is chosen. More precisely, let $d \geq 1$, $s > d$ and $\eta > 0$ be given, and assume that it is possible to find a constant γ such that for every sequence $(q_n)_{n\geq 1}$ and every $\varepsilon > 0$ for sufficiently large N

$$\mathbb{P}\left(D_N^*(x_n) \leq 2D_N^*(q_n) + \varepsilon\right) \geq 1 - \gamma^s e^{-2\varepsilon^2 N(1+\eta)}. \tag{5.28}$$

Chose $\hat{\eta}$ so small that

$$(1+\hat{\eta})^3 \leq (1+\eta), \tag{5.29}$$

and let $(q_n)_{n\geq 1}$ be a d-dimensional sequence for which $D_N^*(q_n) \to 0$. Write I for the indicator of the s-dimensional box of the form $[0,1]^d \times [0, 2^{1/(s-d)}]$. Then I has Lebesgue measure 1/2. Let $X_n,\ n \geq 1$ be i.i.d. random variables having uniform distribution on $[0,1]^{s-d}$, and write

$(x_n)_{n\geq 1}$ for the mixed sequence. Then $x_n \in I$ if and only if $X_n \in [0, 2^{1/(s-d)}]$. The random variables

$$\mathbf{1}_I(x_n) = \mathbf{1}_{[0,2^{1/(s-d)}]}(X_n)$$

are independent, fair Bernoulli random variables. Thus, if ε is chosen appropriately small, we have by Lemma 5.1

$$\mathbb{P}\left(\sum_{k=1}^{N} \mathbf{1}_I(x_n) \geq \frac{N}{2} + (1+\hat{\eta})\varepsilon N\right) \geq e^{-2\varepsilon^2 N(1+\hat{\eta})^3},$$

for sufficiently large N. Since $D_N^*(q_n) \to 0$, this implies

$$\mathbb{P}\left(D_N^*(x_n) \geq 2D_N^*(q_n) + \varepsilon\right) \quad \geq \quad e^{-2\varepsilon^2 N(1+\hat{\eta})^3} \tag{5.30}$$

for sufficiently large N. By (5.29)

$$\frac{e^{-2\varepsilon^2 N(1+\hat{\eta})^3}}{e^{-2\varepsilon^2 N(1+\eta)}} \to 0 \qquad \text{as} \qquad N \to \infty,$$

and hence (5.30) implies

$$\mathbb{P}\left(D_N^*(x_n) \leq 2D_N^*(q_n) + \varepsilon\right) < 1 - \gamma^s e^{-2\varepsilon^2 N(1+\eta)}$$

for sufficiently large N, which is a contradiction to (5.28).

Chapter 6

A central limit theorem for Latin hypercube sampling with dependence

6.1 Introduction

In this article we consider the problem of reducing the variance of a Monte Carlo (MC) estimator for special functionals of a random vector with dependent components. Several different techniques can be used for this kind of problem, with different advantages and shortcomings (for a detailed comparison, see [42, Section 4]). A well-known technique is *Latin hypercube sampling* (LHS), which is a multi-dimensional version of the *stratified sampling* method and has been introduced by [77]. Although this method is well applicable to many different types of problems, it cannot deal with dependence structures among the components of random vectors. Therefore, we consider *Latin hypercube sampling with dependence* (LHSD), which was introduced by [97] and provides variance reduction for many problems, especially in financial mathematics.
Consider the problem of estimating $\mathbb{E}[f(U^1, \dots, U^d)]$ for a Borel-measurable and C-integrable function $f : [0,1]^d \to \mathbb{R}$, where $(U^1, \dots, U^d)$ is a random vector with uniformly distributed marginals and copula C. Let $(U_i^1, \dots, U_i^d),\ 1 \leq i \leq n$, denote an i.i.d. sample from this distribution. The standard Monte Carlo estimator, which is given by $1/n \sum_{i=1}^n f(U_i^1, \dots, U_i^d)$, is strongly consistent, and by the central limit theorem for sums of independent random variables the distribution of the scaled estimator converges to a normal distribution, ie:

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} [f(U_i^1, \dots, U_i^d) - \mathbb{E}[f(U^1, \dots, U^d)]] \xrightarrow{\mathcal{D}} N(0, \sigma_{MC}^2),$$

where $\sigma_{MC}^2 = \mathrm{Var}(f(U^1, \dots, U^d))$. In particular this means that the standard deviation of the estimator converges to zero with rate $\frac{1}{\sqrt{n}}$.
The aim of this paper is to establish a similar result for the LHSD estimator, under some additional conditions on the copula C and the function f. This has already been done in the bivariate case by [88] by using a result of [36]. [88, Proposition 5.9] also showed that under more restrictive conditions on the copula function C, the variance of the bivariate LHSD estimator does not exceed the variance of the standard Monte Carlo estimator.

An important application of Monte Carlo integration techniques lies in the field of financial mathematics. Many problems in finance result in the numerical computation of high-dimensional integrals, for which MC methods provide an efficient solution. Two examples are the pricing of Asian and discrete lookback options on several possibly correlated assets. We will investigate these special derivatives in numerical examples in the last section.
This paper is organised as follows: in the second section we introduce the main ideas of LHSD and recall some important results. Our main results are presented in the third section, where we state a central limit theorem and show under which conditions a reduction of variance, compared to the standard Monte Carlo method, is possible. The last section is dedicated to a comparison of the effectiveness of LHSD and MC in numerical examples.

6.2 Preliminaries

In this section, we recall the concept of stratified sampling and its extensions to Latin hypercube sampling and Latin hypercube sampling with dependence. We also state a consistency result, which was proved by [88].

6.2.1 Stratified sampling and LHS

Suppose that we want to estimate $\mathbb{E}(f(U))$, where U is an uniformly distributed random variable on the interval $[0,1]$ (from now on denoted by $U([0,1])$), and where $f : [0,1] \to \mathbb{R}$ is a Borel-measurable and integrable function. By the simple fact that

$$\mathbb{E}(f(U)) = \sum_{i=1}^{n} \mathbb{E}(f(U)|U \in A_i)\, \mathbb{P}(U \in A_i),$$

where the intervals $A_1, \dots, A_n$ (the so-called *strata*) form a partition of $[0,1]$, we get an estimator for $\mathbb{E}(f(U))$ by sampling U conditionally on the events $\{U \in A_i\}, i = 1, \dots, n$. Choosing strata of the form $A_i = [\frac{i-1}{n}, \frac{i}{n})$ we can simply transform independent samples $U^1, \dots, U^n$ from $U([0,1])$ by setting

$$V_i := \frac{i-1}{n} + \frac{U_i}{n}, \qquad i = 1, \dots, n,$$

which implies $V_i \in A_i, i = 1, \dots, n$. The resulting estimator for $\mathbb{E}(f(U))$ given by $\frac{1}{n}\sum_{i=1}^{n} f(V_i)$ is consistent, and by the central limit theorem for sums of independent random variables the limit variance is smaller than the limit variance of a standard Monte Carlo estimator. For a more detailed analysis of stratified sampling techniques, see [42, Section 4.3.1].
This approach can be extended to the multivariate case in different ways. If we require that there has to be exactly one sample in every stratum, we need to draw n^d samples, which is not feasible for a high number of dimensions d. One way to avoid this problem is Latin hypercube sampling. Assume we want to estimate $\mathbb{E}(f(U^1, \dots, U^d))$, where $f : [0,1]^d \to \mathbb{R}$ is a Borel-measurable and integrable function. For fixed n we generate n independent samples denoted by $(U_i^1, \dots, U_i^d), i = 1, \dots, n$, where the $U_i^j, j = 1, \dots, d$ are uniformly distributed on $[0,1]$.

Additionally, we generate d independent permutations of $\{1,\dots,n\}$, denoted by $\pi_1,\dots,\pi_d$, drawn from a discrete uniform distribution on the set of all possible permutations. Denote by π_i^j the value to which i is mapped by the j-th permutation. Then the j-th component of a Latin hypercube sample is given by

$$V_i^j := \frac{\pi_i^j - 1}{n} + \frac{U_i^j}{n}, \qquad j = 1,\dots,d;\ i = 1,\dots,n.$$

By fixing a dimension j, the components $(V_1^j,\dots,V_n^j)$ form a stratified sample with strata of equal length. It can be shown that the resulting estimator for $\mathbb{E}(f(U))$ is consistent, and by assuming that $f(U^1,\dots,U^d)$ has a finite second moment it follows that the variance of the LHS estimator

$$\frac{1}{n}\sum_{i=1}^{n} f(V_i^1,\dots,V_i^d)$$

is smaller than the variance of the standard MC estimator, provided the number of sample points is sufficiently large, see [97]. If f is bounded a central limit theorem for the LHS estimator can be shown, see [86]. Berry-Esseen-type bounds are also known, see [73]. A detailed discussion of LHS is given in [42, Section 4.4].
This technique is not suitable for dealing with random vectors with dependent components since the random variables $V_i^j, j = 1,\dots,d$, are independent. One way to extend the LHS method to random vectors with dependent components is to apply LHS to independent components and then introduce dependencies through a transformation of the LHS points. Such a procedure is tedious in general, and we will not pursue this approach any further.

6.2.2 Latin hypercube sampling with dependence

In this subsection, we introduce Latin hypercube sampling with dependence. The main difference to the LHS method is that instead of random permutations π_i we use rank statistics, which are defined as follows:

Definition 6.1 (Rank statistics) *Let $X_1,\dots,X_n$ be i.i.d. random variables with a continuous distribution function. Denote the ordered random variables by $X_{(1)} < \cdots < X_{(n)}$, $\mathbb{P}$-a.s. We call the index of X_i within $X_{(1)} < \cdots < X_{(n)}$ the i-th rank statistic, given by*

$$r_{i,n} = r_{i,n}(X_1,\dots,X_n) := \sum_{k=1}^{n} \mathbf{1}_{\{X_k \le X_i\}}. \tag{6.1}$$

Consider a random vector $U = (U^1,\dots,U^d)$, where every component U^j is uniformly distributed on $[0,1]$ and the dependence structure of U is modeled by a copula C.
Let $(U_i^1,\dots,U_i^d), i = 1,\dots,n$ denote a sequence of independent samples of $(U^1,\dots,U^d)$, and let $r_{i,n}^j$ be the i-th rank statistic of $(U_1^j,\dots,U_n^j)$ for $i = 1,\dots,n$ and $j = 1,\dots,d$. Then a LHSD is given by

$$V_{i,n}^j := \frac{r_{i,n}^j - 1}{n} + \frac{\eta_{i,n}^j}{n}, \qquad i = 1,\dots,n, \forall j = 1,\dots,d, \tag{6.2}$$

where $\eta_{i,n}^j$ are random variables in $[0,1]$. It is clear that $(V_{1,n}^j, \ldots, V_{n,n}^j)$ forms a stratified sampling in every dimension j, where every stratum has equal length.
[88] consider different choices for $\eta_{i,n}^j$ to obtain special properties. For example, by choosing all $\eta_{i,n}^j$ uniformly distributed on $[0,1]$ and independent of U_i^j, the distribution of the $V_{i,n}^j$ within their strata is uniform. This choice has the disadvantage of necessitating the generation of $2n$ random variables instead of only n. An effective choice in terms of computation time is $\eta_{i,n}^j = 1/2$, which means that every $V_{i,n}^j$ is located exactly in the centre of its stratum. In the remainder of this section, we briefly recall a result of [88] concerning the consistency of the LHSD estimator for $\mathbb{E}(f(U))$, which is defined by

$$\frac{1}{n}\sum_{i=1}^{n} f(V_{i,n}^1, \ldots, V_{i,n}^d). \tag{6.3}$$

The usual law of large numbers for sums of independent random variables does not apply in this case for two reasons: firstly in each dimension the samples fail to be independent because of the application of the rank statistic, and secondly, increasing the samples size n by one changes every term of the sum instead of just adding one. Nevertheless, it can be shown that the following consistency result holds, see [88, Proposition 4.1]:

Proposition 6.1 *Let $f : [0,1]^d \to \mathbb{R}$ be bounded and continuous C-a.e. . Then the LHSD estimator* (6.3) *is strongly consistent, ie :*

$$\frac{1}{n}\sum_{i=1}^{n} f(V_{i,n}^1, \ldots, V_{i,n}^d) \xrightarrow{\mathbb{P}\,a.s.} \mathbb{E}(f(U^1, \ldots, U^d)), \qquad \text{as } n \to \infty.$$

6.3 Central limit theorem and variance reduction

In this section we investigate the speed of convergence of the LHSD estimator and discuss situations in which the use of LHSD results in a reduction of variance. This has already been done for the bivariate case by [88]. They have also guessed the higher-dimensional version of the main theorem, but no rigorous proof was given. Because of the fact that most problems in finance for which Monte Carlo techniques are suitable are high-dimensional integration problems, it is reasonable to investigate the speed of convergence and the (asymptotic) value of the variance also in the multivariate case.
In the sequel, let $\overline{C}_n$ denote the empirical distribution of the LHSD sample given by

$$\overline{C}_n(u^1, \ldots, u^d) := \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\{V_{i,n}^1 \le u^1, \ldots, V_{i,n}^d \le u^d\}},$$

which is a distribution function. Furthermore, we define C_n as

$$C_n(u^1, \ldots, u^d) := \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\{F_n^1(U_i^1) \le u^1, \ldots, F_n^d(U_i^d) \le u^d\}}, \tag{6.4}$$

where

$$F_n^j(u) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\{U_i^j \le u\}}, \qquad u \in [0,1],$$

are the one-dimensional empirical distribution functions based on $U_1^j, \ldots, U_n^j$ for $j = 1, \ldots, d$. To formulate a central limit theorem we will need some regularity conditions on the integrand f and the copula C.

Definition 6.2 (Hardy-Krause bounded variation) *A function $f : [0,1]^d \to \mathbb{R}$ is of bounded variation (in the sense of Hardy-Krause) if $V(f) < \infty$ with*

$$V(f) = \sum_{k=1}^{d} \sum_{1 \le i_1 < \ldots < i_k \le d} V^{(k)}(f; i_1, \ldots, i_k).$$

Here, the functional $V^{(k)}(f)$ denotes the variation in the sense of Vitali of f restricted to the k - dimensional face $F^{(k)}(i_1, \ldots, i_k) = \{(u_1, \ldots, u_d) \in [0,1]^d : u_j = 1 \text{ for } j \ne i_1, \ldots, i_k\}$. The variation of a function f in the sense of Vitali is defined by

$$V^{(k)}(f; i_1, \ldots, i_k) = \sup_{\mathcal{P}} \sum_{J \in \mathcal{P}(i_1, \ldots, i_k)} |\Delta(f; J)|,$$

where the supremum is extended over all partitions $\mathcal{P}(i_1, \ldots, i_k)$ of $F^{(k)}(i_1, \ldots, i_k)$ into subintervals J and $\Delta(f; J)$ denotes the alternating sum of the values of f at the vertices of J. For more information on this topic, see [87].

Definition 6.3 *A function $f : [0,1]^d \to \mathbb{R}$ is right continuous if for any sequence $(u_n^1, u_n^2, \ldots, u_n^d)_{n \in \mathbb{N}}$ with $u_n^j \downarrow u^j, j = 1, \ldots, d$,*

$$\lim_{n \to \infty} f(u_n^1, u_n^2, \ldots, u_n^d) = f(u^1, u^2, \ldots, u^d).$$

The next statement concerning the convergence of random sequences will be used to prove Proposition 6.2 and Theorem 6.2. For more details see eg [58, Theorem 18.8].

Lemma 6.1 *Let $(X_n)_{n \ge 1}$ and $(Y_n)_{n \ge 1}$ be sequences of $\mathbb{R}$-valued random variables, with $X_n \xrightarrow{\mathcal{D}} X$ and $|X_n - Y_n| \xrightarrow{\mathbb{P}} 0$. Then $Y_n \xrightarrow{\mathcal{D}} X$.*

The following proposition of [99] is a generalization of earlier results of [98] and [36]. It is the essential ingredient in proofs of our main theorems.

Proposition 6.2 *Assume that C is differentiable with continuous partial derivatives $\partial_j C(u^1, \ldots, u^d) = \frac{\partial C(u^1, \ldots, u^d)}{\partial u^j}$ for $j = 1, \ldots, d$. Then*

$$\sqrt{n}\Big(\widetilde{C}_n(u^1, \ldots, u^d) - C(u^1, \ldots, u^d)\Big) \xrightarrow{\mathcal{D}} G_C(u^1, \ldots, u^d),$$

where

$$\widetilde{C}_n(u^1,\dots,u^d) = \frac{1}{n}\sum_{k=1}^{n} \mathbf{1}_{\{U_k^1 \le F_n^{1-}(u^1),\dots,U_k^d \le F_n^{d-}(u^d)\}},$$

denotes the empirical copula function and F_n^{j-} denote the generalised quantile functions of F_n^j for $j = 1,\dots,d$, defined by

$$F_n^{j-}(u) = \inf\{x \in \mathbb{R} | F_n^j(x) \ge u\}.$$

Furthermore, G_C is a centred Gaussian random field given by

$$G_C(u^1,\dots,u^d) = B_C(u^1,\dots,u^d) - \sum_{j=1}^{d} \partial_j C(u^1,\dots,u^d) B_C(1,\dots,1,u^j,1,\dots,1), \quad (6.5)$$

B_C is a d-dimensional pinned Brownian sheet on $[0,1]^d$ with covariance function

$$\begin{aligned} &\mathbb{E}[B_C(u^1,\dots,u^d)\cdot B_C(\overline{u}^1,\dots,\overline{u}^d)] \\ =&C((u^1,\dots,u^d)\wedge(\overline{u}^1,\dots,\overline{u}^d)) - C(u^1,\dots,u^d)C(\overline{u}^1,\dots,\overline{u}^d), \end{aligned} \quad (6.6)$$

where $(u^1,\dots,u^d)\wedge(\overline{u}^1,\dots,\overline{u}^d)$ denotes the componentwise minimum.

We can formulate a similar result for the sequence C_n.

Proposition 6.3 *Under the conditions of Proposition 6.2,*

$$\sqrt{n}\Big(C_n(u^1,\dots,u^d) - C(u^1,\dots,u^d)\Big) \xrightarrow{\mathcal{D}} G_C(u^1,\dots,u^d) \quad (6.7)$$

holds, where all definitions are as in Proposition 6.2 and $C_n(u^1,\dots,u^d)$ is given in (6.4)*.*

Proof:
We only have to show that the supremum of the difference of C_n and $\widetilde{C}_n$ vanishes for $n \to \infty$ to apply Lemma 6.1, which completes the proof. Note that C_n and $\widetilde{C}_n$ coincide on the grid $\{(i_1/n,\dots,i_d/n), 1 \le i_1,\dots,i_d \le n\}$. It follows that

$$\begin{aligned} &\sup_{u^1,\dots,u^d} |\widetilde{C}_n(u^1,\dots,u^d) - C_n(u^1,\dots,u^d)| \\ \le &\max_{1\le i^1,\dots,i^d\le n} \Big|\widetilde{C}_n\Big(\frac{i_1}{n},\dots,\frac{i_d}{n}\Big) - \widetilde{C}_n\Big(\frac{i_1-1}{n},\dots,\frac{i_d-1}{n}\Big)\Big| \le \frac{d}{n}. \end{aligned}$$

Thus, $\sup_{u^1,\dots,u^d} |\widetilde{C}_n(u^1,\dots,u^d) - C_n(u^1,\dots,u^d)| \to 0$ for $n \to \infty$ and (6.7) follows. □

In the sequel, all $U^i, i = 1,\dots,d$ are uniformly distributed random variables on $[0,1]$ and all integrals have to be understood in the sense of Lebesgue-Stieltjes. Note that the next theorem is an extension of [36, Theorem 6] from the case of bivariate to the case of multi-variate random vectors $U = (U^1,\dots,U^d)$.

Theorem 6.1 *Let the copula C of $(U^1, \ldots, U^d)$ have continuous partial derivatives and let $f : [0,1]^d \to \mathbb{R}$ be a right-continuous function of bounded variation in the sense of Hardy-Krause. Then*

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Big(f(F_n^1(U_i^1), \ldots, F_n^d(U_i^d)) - \mathbb{E}[f(U^1, \ldots, U^d)] \Big) \xrightarrow{\mathcal{D}} \int_{[0,1]^d} G_C(u^1, \ldots, u^d) d\widehat{f}(u^1, \ldots, u^d),$$

where the function $\widehat{f} : [0,1]^d \to \mathbb{R}$ is defined by:

$$\widehat{f}(u^1, \ldots, u^d) = \begin{cases} 0 & \text{if at least one } u^j = 1, \text{ for } j = 1, \ldots, d, \\ f(u^1, \ldots, u^d) & \text{otherwise.} \end{cases} \tag{6.8}$$

Furthermore, the limit distribution is Gaussian.

Proof:
By definition $\widehat{f}$ is right-continuous and of bounded variation in the sense of Hardy-Krause. Furthermore, it follows that almost surely

$$\begin{aligned} &\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Big(f(F_n^1(U_i^1), \ldots, F_n^d(U_i^d)) - \mathbb{E}[f(U^1, \ldots, U^d)] \Big) \\ &\quad = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Big(\widehat{f}(F_n^1(U_i^1), \ldots, F_n^d(U_i^d)) - \mathbb{E}[\widehat{f}(U^1, \ldots, U^d)] \Big), \end{aligned}$$

by the fact that C is continuous on $[0,1]^d$.
We use a multidimensional integration-by-parts technique proposed by [104, Proposition 2]. Using the notation of [104] we get

$$\begin{aligned} &\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Big(\widehat{f}(F_n^1(U_i^1), \ldots, F_n^d(U_i^d)) - \mathbb{E}[\widehat{f}(U^1, \ldots, U^d)] \Big) \\ &\quad = \sqrt{n} \int_{[0,1]^d} \widehat{f}(u^1, \ldots, u^d) d(C_n - C)(u^1, \ldots, u^d) \\ &\quad = \sqrt{n} \sum_{k=0}^{d} (-1)^k \sum_{1,\ldots,d;k} \Delta^*_{j_{k+1},\ldots,j_d} \int_{[0,1]^k} (C_n - C)(u^1, \ldots, u^d) d_{j_1,\ldots,j_k} \widehat{f}(u^1, \ldots, u^d). \end{aligned} \tag{6.9}$$

Here $\sum_{1,\ldots,d;k}$ denotes the sum over all possible partitions of the set $\{j_1, \ldots, j_d\}$ into two subsets $\{j_1, \ldots, j_k\}$ and $\{j_{k+1}, \ldots, j_d\}$ of k respectively $d-k$ elements, where each partition is taken exactly once. In the cases $k = 0$ and $k = d$, the sum is interpreted as being reduced to one term.
Furthermore, the operator $d_{j_1,\ldots,j_k}$ indicates that the integral only applies to the variables $j_1, \ldots, j_k$. Note that after the application of the integral with respect to $d_{j_1,\ldots,j_k} \widehat{f}(u^1, \ldots, u^d)$, the integrated

function is a function in $d-k$ variables. Furthermore for a function g of $d-k$ variables, the operator $\Delta^*_{j_{k+1},\dots,j_d}$ is given by

$$\Delta^*_{j_{k+1},\dots,j_d} g(j_{k+1},\dots,j_d) = \sum_{\{i_1,\dots,i_{d-k}\}\in\{0,1\}^{d-k}} (-1)^m g(i_1,\dots,i_{d-k}),$$

where m denotes the number of zeros in $\{i_1,\dots,i_{d-k}\}$. This means that, for $j \notin \{j_1,\dots,j_k\}$

$$\begin{aligned}
&\Delta^*_j \int_{[0,1]^{d-k}} (C_n - C)(u^1,\dots,u^d) d_{j_1,\dots,j_k}\widehat{f}(u^1,\dots,u^d) \\
&\quad = \int_{[0,1]^{d-k}} (C_n - C)(u^1,\dots,u^{j-1},1,u^{j+1},\dots,u^d) d_{j_1,\dots,j_k}\widehat{f}(u^1,\dots,u^{j-1},1,u^{j+1},\dots,u^d) \\
&\quad - \int_{[0,1]^{d-k}} (C_n - C)(u^1,\dots,u^{j-1},0,u^{j+1},\dots,u^d) d_{j_1,\dots,j_k}\widehat{f}(u^1,\dots,u^{j-1},0,u^{j+1},\dots,u^d)
\end{aligned}$$

and

$$\Delta^*_{j_{k+1},\dots,j_d} = \Delta^*_{j_{k+1}} \dots \Delta^*_{j_d}.$$

Thus

$$\begin{aligned}
\sqrt{n}\sum_{k=0}^{d}(-1)^k \sum_{1,\dots,d;k} \Delta^*_{j_{k+1},\dots,j_d} \int_{[0,1]^k} (C_n - C)(u^1,\dots,u^d) d_{j_1,\dots,j_k}\widehat{f}(u^1,\dots,u^d) \\
= \sqrt{n}\sum_{k=0}^{d-1}(-1)^k \sum_{1,\dots,d;k} \Delta^*_{j_{k+1},\dots,j_d} \int_{[0,1]^k} (C_n - C)(u^1,\dots,u^d) d_{j_1,\dots,j_k}\widehat{f}(u^1,\dots,u^d) \\
+ \sqrt{n}(-1)^d \int_{[0,1]^d} (C_n - C)(u^1,\dots,u^d) d\widehat{f}(u^1,\dots,u^d) \\
= \sqrt{n}(-1)^d \int_{[0,1]^d} (C_n - C)(u^1,\dots,u^d) d\widehat{f}(u^1,\dots,u^d).
\end{aligned}$$

The term

$$\sqrt{n}\sum_{k=0}^{d-1}(-1)^k \sum_{1,\dots,d;k} \Delta^*_{j_{k+1},\dots,j_d} \int_{[0,1]^k} (C_n - C)(u^1,\dots,u^d) d_{j_1,\dots,j_k}\widehat{f}(u^1,\dots,u^d)$$

vanishes because each of its terms is equal to zero due to at least one of the following two reasons: firstly, at least one $u^j, j = 1,\dots,d$ is equal to one and therefore $\widehat{f}(u^1,\dots,u^d) = 0$ by definition, or, secondly, at least one $u^j, j = 1,\dots,d$ is equal to zero, hence $C_n(u^1,\dots,u^d) = C(u^1,\dots,u^d) = 0$.
Thus, by the continuous mapping theorem and (6.7), it follows that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big(f(F_n^1(U_i^1),\dots,F_n^d(U_i^d)) - \mathbb{E}[f(U^1,\dots,U^d)]\Big)$$

$$= (-1)^d \sqrt{n} \int_{[0,1]^d} (C_n - C)(u^1, \ldots, u^d) d\widehat{f}(u^1, \ldots, u^d)$$
$$\xrightarrow{\mathcal{D}} \int_{[0,1]^d} G_C(u^1, \ldots, u^d) d\widehat{f}(u^1, \ldots, u^d).$$

Since $\int_{[0,1]^d} G_C(u^1, \ldots, u^d) d\widehat{f}(u^1, \ldots, u^d)$ is a continuous, linear transformation of a tight Gaussian process, it follows that the limiting distribution is Gaussian. □

Remark 6.1 *The reason for using the function $\widehat{f}$ instead of f is that the integrals of dimension $k = 2, \ldots, d-1$ in* (6.9) *are in general not vanishing. The one-dimensional integrals are zero for every right-continuous function of bounded variation f because of special properties of the function C_n, for more details see [36]. In particular, this means that in the two-dimensional case it is sufficient to assume*

$$\widehat{f}(x) = f(x), \quad x \in \mathbb{R}^2.$$

With this assumption instead of (6.8) *and $d = 2$, Theorem 6.1 is equivalent to [36, Theorem 6]. We use the function $\widehat{f}$ to get a more convenient representation for the limit variance of the LHSD technique, which we state in the next theorem.*

Theorem 6.2 *Under the assumptions and notations of Theorem 6.1, we have*

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Big(f(V_{i,n}^1, \ldots, V_{i,n}^d) - \mathbb{E}[f(U^1, \ldots, U^d)] \Big) \xrightarrow{\mathcal{D}} N(0, \sigma_{LHSD}^2), \tag{6.10}$$

where

$$\sigma_{LHSD}^2 = \int_{[0,1]^{2d}} \mathbb{E}\Big[G_C(u^1, \ldots, u^d) G_C(\overline{u}^1, \ldots, \overline{u}^d) \Big] d\widehat{f}(u^1, \ldots, u^d) d\widehat{f}(\overline{u}^1, \ldots, \overline{u}^d). \tag{6.11}$$

Proof:
We want to apply Theorem 6.1 together with Lemma 6.1, so we have to show that

$$\frac{1}{\sqrt{n}} \left| \sum_{i=1}^{n} \Big[f(V_{i,n}^1, \ldots, V_{i,n}^d) - f(F_n^1(U_i^1), \ldots, F_n^d(U_i^d)) \Big] \right| \to 0, \qquad \text{as } n \to \infty.$$

By [71, Corollary 1]

$$\left| \sum_{i=1}^{n} \Big[f(V_{i,n}^1, \ldots, V_{i,n}^d) - f(F_n^1(U_i^1), \ldots, F_n^d(U_i^d)) \Big] \right| \leq V(f) < \infty,$$

where $V(f)$ is the Hardy-Krause variation of f. Hence

$$\frac{1}{\sqrt{n}} \left| \sum_{i=1}^{n} \Big[f(V_{i,n}^1, \ldots, V_{i,n}^d) - f(F_n^1(U_i^1), \ldots, F_n^d(U_i^d)) \Big] \right| \to 0, \qquad \text{as } n \to \infty,$$

which, together with Lemma 6.1 and Theorem 6.1, proves equation (6.10).
To derive equation (6.11) we apply Fubini's theorem to $\mathbb{E}[(\int_{[0,1]^d} G_C(u^1,\dots,u^d)d\widehat{f}(u^1,\dots,u^d))^2]$. By [71, Theorem 3] a function of bounded variation $\widehat{f}$ can always be written as the difference of two completely monotone functions g,h and therefore an integral with respect to $\widehat{f}$ can be written as a difference of two integrals with respect to positive measures g,h. Thus

$$
\begin{aligned}
&\mathbb{E}\Big[\Big(\int_{[0,1]^d} G_C(u^1,\dots,u^d)d\widehat{f}(u^1,\dots,u^d)\Big)^2\Big] = \\
&\quad= \mathbb{E}\Big[\Big(\int_{[0,1]^d} G_C(u^1,\dots,u^d)d\widehat{f}(u^1,\dots,u^d)\Big)\cdot\Big(\int_{[0,1]^d} G_C(\overline{u}^1,\dots,\overline{u}^d)d\widehat{f}(\overline{u}^1,\dots,\overline{u}^d)\Big)\Big] \\
&\quad= \mathbb{E}\Big[\Big(\int_{[0,1]^d} G_C(u^1,\dots,u^d)dg(u^1,\dots,u^d) - \int_{[0,1]^d} G_C(u^1,\dots,u^d)dh(u^1,\dots,u^d)\Big) \\
&\quad\cdot\Big(\int_{[0,1]^d} G_C(\overline{u}^1,\dots,\overline{u}^d)dg(\overline{u}^1,\dots,\overline{u}^d) - \int_{[0,1]^d} G_C(\overline{u}^1,\dots,\overline{u}^d)dh(\overline{u}^1,\dots,\overline{u}^d)\Big)\Big] \\
&\quad= \mathbb{E}\Big[\Big(\int_{[0,1]^{2d}} G_C(u^1,\dots,u^d)G_C(\overline{u}^1,\dots,\overline{u}^d)dg(u^1,\dots,u^d)dg(\overline{u}^1,\dots,\overline{u}^d) \\
&\quad- \int_{[0,1]^d} G_C(u^1,\dots,u^d)G_C(\overline{u}^1,\dots,\overline{u}^d)dh(u^1,\dots,u^d)dg(\overline{u}^1,\dots,\overline{u}^d) \\
&\quad- \int_{[0,1]^d} G_C(u^1,\dots,u^d)G_C(\overline{u}^1,\dots,\overline{u}^d)dg(u^1,\dots,u^d)dh(\overline{u}^1,\dots,\overline{u}^d) \\
&\quad+ \int_{[0,1]^d} G_C(u^1,\dots,u^d)G_C(\overline{u}^1,\dots,\overline{u}^d)dh(u^1,\dots,u^d)dh(\overline{u}^1,\dots,\overline{u}^d)\Big)\Big] \\
&\quad= \int_{[0,1]^{2d}} \mathbb{E}\Big[G_C(u^1,\dots,u^d)G_C(\overline{u}^1,\dots,\overline{u}^d)\Big]dg(u^1,\dots,u^d)dg(\overline{u}^1,\dots,\overline{u}^d) \\
&\quad- \int_{[0,1]^d} \mathbb{E}\Big[G_C(u^1,\dots,u^d)G_C(\overline{u}^1,\dots,\overline{u}^d)\Big]dh(u^1,\dots,u^d)dg(\overline{u}^1,\dots,\overline{u}^d) \\
&\quad- \int_{[0,1]^d} \mathbb{E}\Big[G_C(u^1,\dots,u^d)G_C(\overline{u}^1,\dots,\overline{u}^d)\Big]dg(u^1,\dots,u^d)dh(\overline{u}^1,\dots,\overline{u}^d) \\
&\quad+ \int_{[0,1]^d} \mathbb{E}\Big[G_C(u^1,\dots,u^d)G_C(\overline{u}^1,\dots,\overline{u}^d)\Big]dh(u^1,\dots,u^d)dh(\overline{u}^1,\dots,\overline{u}^d) \\
&\quad= \int_{[0,1]^{2d}} \mathbb{E}\Big[G_C(u^1,\dots,u^d)G_C(\overline{u}^1,\dots,\overline{u}^d)\Big]d\widehat{f}(u^1,\dots,u^d)d\widehat{f}(\overline{u}^1,\dots,\overline{u}^d),
\end{aligned}
$$

where the use of Fubini's theorem is justified since $\widehat{f}$ is bounded and $\mathbb{E}[XY] < \infty$ for two jointly normal random variables X and Y. □

Remark 6.2 *Note that by* (6.5) *and* (6.6) *the expression for* σ^2_{LHSD} *in equation* (6.11) *can be represented in terms of* C. *Additionally, further simplifications can be given for the following*

terms:

$$\begin{aligned}
&\mathbb{E}[B_C(u^1,\dots,u^d)\cdot B_C(1,\dots,1,\overline{u}^j,1,\dots,1)]\\
&\qquad = C((u^1,\dots,u^{j-1},u^j\wedge\overline{u}^j,u^{j+1}\dots,u^d)) - C(u^1,\dots,u^d)\overline{u}^j,\\
&\mathbb{E}[B_C(1,\dots,1,u^i,1,\dots,1)\cdot B_C(1,\dots,1,\overline{u}^j,1,\dots,1)]\\
&\qquad = C((1,\dots,1,u^i,1,\dots,1,\overline{u}^j,1,\dots,1)) - u^i\overline{u}^j,\\
&\mathbb{E}[B_C(1,\dots,1,u^j,1,\dots,1)\cdot B_C(1,\dots,1,\overline{u}^j,1,\dots,1)] = u^j\wedge\overline{u}^j - u^j\overline{u}^j,
\end{aligned}$$

since $C(1,\dots,1,u^j,1,\dots,1) = u^j$ *for all* $j = 1,\dots,d$.

It is important to know if the LHSD estimator has a smaller variance than the Monte Carlo estimator. The variance of a standard Monte Carlo estimator is given by

$$\sigma^2_{MC} = \int_{[0,1]^d} f(u^1,\dots,u^d)^2 dC(u^1,\dots,u^d) - \Big(\int_{[0,1]^d} f(u^1,\dots,u^d)dC(u^1,\dots,u^d)\Big)^2.$$

We use this fact to establish a relation between σ^2_{MC} and σ^2_{LHSD}.

Proposition 6.4 *Let the copula* C *of* $(U^1,\dots,U^d)$ *have continuous partial derivatives, let* $f : [0,1]^d \to \mathbb{R}$ *be a right-continuous function of bounded variation in the sense of Hardy-Krause and let* $\widehat{f}$ *be as defined in Theorem 6.1. Set* $\partial_j C(u^1,\dots,u^d) = \frac{\partial C(u^1,\dots,u^d)}{\partial u^j}$ *and*

$$C_{i,j}(u^i,\overline{u}^j) = \begin{cases} C(1,\dots,1,u^i,1,\dots,1,\overline{u}^j,1,\dots,1), & i\neq j\\ u^i\wedge\overline{u}^j, & i=j. \end{cases}$$

Then

$$\begin{aligned}
&\sigma^2_{LHSD} = \sigma^2_{MC}\\
&+\int_{[0,1]^{2d}} 2\sum_{j=1}^d \partial_j C(u^1,\dots,u^d)\Big(C(\overline{u}^1,\dots,\overline{u}^d)u^j - C(\overline{u}^1,\dots,\overline{u}^{j-1},\overline{u}^j\wedge u^j,\overline{u}^{j+1},\dots,\overline{u}^d)\Big)\\
&+\sum_{j=1}^d\sum_{i=1}^d \partial_j C(\overline{u}^1,\dots,\overline{u}^d)\partial_i C(u^1,\dots,u^d)\Big(C_{i,j}(u^i,\overline{u}^j) - u^i\overline{u}^j\Big)d\widehat{f}(u^1,\dots,u^d)d\widehat{f}(\overline{u}^1,\dots,\overline{u}^d).
\end{aligned} \tag{6.12}$$

Proof:
Note that

$$\int_{[0,1]^d} f(u^1,\dots,u^d)^2 dC(u^1,\dots,u^d) = \int_{[0,1]^{2d}} f(u^1,\dots,u^d)f(\overline{u}^1,\dots,\overline{u}^d)dC(u^1\wedge\overline{u}^1,\dots,u^d\wedge\overline{u}^d),$$

and that the function $C(u^1\wedge\overline{u}^1,\dots,u^d\wedge\overline{u}^d)$ is also a copula, which follows by observing that

$$C(u^1\wedge\overline{u}^1,\dots,u^d\wedge\overline{u}^d) = \mathbb{P}(U^1\leq u^1\wedge\overline{u}^1,\dots,U^d\leq u^d\wedge\overline{u}^d)$$

$$= \mathbb{P}(U^1 \leq u^1, U^1 \leq \overline{u}^1, \ldots, U^d \leq u^d, U^d \leq \overline{u}^d)$$

is a joint probability distribution with uniform marginals.
By integration-by-parts like in Theorem 6.1 it follows for the variance of the Monte Carlo estimator that

$$\begin{aligned}
\sigma^2_{MC} &= \int_{[0,1]^d} f(u^1,\ldots,u^d)^2 dC(u^1,\ldots,u^d) - \Big(\int_{[0,1]^d} f(u^1,\ldots,u^d) dC(u^1,\ldots,u^d)\Big)^2 \\
&= \int_{[0,1]^{2d}} f(u^1,\ldots,u^d) f(\overline{u}^1,\ldots,\overline{u}^d) dC\Big((u^1,\ldots,u^d)\wedge(\overline{u}^1,\ldots,\overline{u}^d)\Big) \\
&- \int_{[0,1]^{2d}} f(u^1,\ldots,u^d) f(\overline{u}^1,\ldots,\overline{u}^d) dC(u^1,\ldots,u^d) dC(\overline{u}^1,\ldots,\overline{u}^d) \\
&= \int_{[0,1]^{2d}} C\Big((u^1,\ldots,u^d)\wedge(\overline{u}^1,\ldots,\overline{u}^d)\Big) d\widehat{f}(u^1,\ldots,u^d) d\widehat{f}(\overline{u}^1,\ldots,\overline{u}^d) \\
&- \int_{[0,1]^{2d}} C(u^1,\ldots,u^d) C(\overline{u}^1,\ldots,\overline{u}^d) d\widehat{f}(u^1,\ldots,u^d) d\widehat{f}(\overline{u}^1,\ldots,\overline{u}^d).
\end{aligned}$$

The proof is completed by using equations (6.5), (6.6), (6.11) and Remark 6.2. □

Theorem 6.3 *Let C and f satisfy the assumptions in Theorem 6.1 and let $\widehat{f}$ be defined as in Theorem 6.1. Furthermore let the function f be monotone non-decreasing in each argument and $\max_{x\in[0,1]^d}(f(x)) \leq 0$. Moreover assume that C satisfies the following conditions:*

$$\frac{C(u^1,\ldots,u^d)}{u^j} \geq \partial_j C(u^1,\ldots,u^d), \quad j \in \{1,\ldots,d\}, \tag{6.13}$$

$$\sum_{i=1,i\neq j}^{d} \frac{C_{i,j}(u^j,\overline{u}^i)}{\overline{u}^j} \leq (d-2)u^j + \frac{C(\overline{u}^1,\ldots,\overline{u}^{j-1},\overline{u}^j\wedge u^j,\overline{u}^{j+1},\ldots,\overline{u}^d)}{C(\overline{u}^1,\ldots,\overline{u}^d)}, \tag{6.14}$$

where $u^j \in [0,1], (\overline{u}^1,\ldots,\overline{u}^d), (u^1,\ldots,u^d) \in [0,1]^d$.
Then $\sigma^2_{LHSD} \leq \sigma^2_{MC}$.

Proof:
By the assumptions on f it follows that $\widehat{f}$ is right-continuous, of bounded variation in the sense of Hardy-Kraus and monotone non-decresing in each argument. Thus by (6.12) it is sufficient to show that

$$\begin{aligned}
&2\sum_{j=1}^{d} \partial_j C(u^1,\ldots,u^d)\Big(C(\overline{u}^1,\ldots,\overline{u}^d)u^j - C(\overline{u}^1,\ldots,\overline{u}^{j-1},\overline{u}^j\wedge u^j,\overline{u}^{j+1},\ldots,\overline{u}^d)\Big) \\
&+\sum_{j=1}^{d}\sum_{i=1}^{d} \partial_i C(\overline{u}^1,\ldots,\overline{u}^d)\partial_j C(u^1,\ldots,u^d)\Big(C_{i,j}(u^j,\overline{u}^i) - u^j\overline{u}^i\Big) \leq 0
\end{aligned}$$

for all $(u^1,\dots,u^d),(\overline{u}^1,\dots,\overline{u}^d)\in[0,1]^d$.
This is true if

$$\begin{aligned}&2\left(C(\overline{u}^1,\dots,\overline{u}^d)u^j-C(\overline{u}^1,\dots,\overline{u}^{j-1},\overline{u}^j\wedge u^j,\overline{u}^{j+1},\dots,\overline{u}^d)\right)\\&\leq\sum_{i=1}^d\partial_iC(\overline{u}^1,\dots,\overline{u}^d)\left(u^j\overline{u}^i-C_{i,j}(u^j,\overline{u}^i)\right)\end{aligned}$$

holds for every $j\in\{1,\dots,d\}$ and all $u^j\in[0,1],(\overline{u}^1,\dots,\overline{u}^d)\in[0,1]^d$.
First we show that

$$C(\overline{u}^1,\dots,\overline{u}^d)u^j-C(\overline{u}^1,\dots,\overline{u}^{j-1},\overline{u}^j\wedge u^j,\overline{u}^{j+1},\dots,\overline{u}^d)\leq\partial_jC(\overline{u}^1,\dots,\overline{u}^d)\left(u^j\overline{u}^j-u^j\wedge\overline{u}^j\right).$$

Note that this is always true if $u^j\wedge\overline{u}^j\in\{0,1\}$. Now assume that $0<\overline{u}^j\leq u^j<1$, then

$$\begin{aligned}C(\overline{u}^1,\dots,\overline{u}^d)u^j-C(\overline{u}^1,\dots,\overline{u}^d)&\leq\partial_jC(\overline{u}^1,\dots,\overline{u}^d)\left(u^j\overline{u}^j-\overline{u}^j\right)\\C(\overline{u}^1,\dots,\overline{u}^d)(u^j-1)&\leq\partial_jC(\overline{u}^1,\dots,\overline{u}^d)\overline{u}^j(u^j-1)\\\frac{C(\overline{u}^1,\dots,\overline{u}^d)}{\overline{u}^j}&\geq\partial_jC(\overline{u}^1,\dots,\overline{u}^d)\end{aligned}$$

which is true by assumption (6.13). Next assume that $0<u^j<\overline{u}^j<1$, then

$$\begin{aligned}C(\overline{u}^1,\dots,\overline{u}^d)u^j-C(\overline{u}^1,\dots,\overline{u}^{j-1},u^j,\overline{u}^{j+1},\dots,\overline{u}^d)&\leq\partial_jC(\overline{u}^1,\dots,\overline{u}^d)\left(u^j\overline{u}^j-u^j\right)\\C(\overline{u}^1,\dots,\overline{u}^d)u^j-C(\overline{u}^1,\dots,\overline{u}^{j-1},u^j,\overline{u}^{j+1},\dots,\overline{u}^d)&\leq\partial_jC(\overline{u}^1,\dots,\overline{u}^d)u^j\left(\overline{u}^j-1\right)\\C(\overline{u}^1,\dots,\overline{u}^d)-\frac{C(\overline{u}^1,\dots,\overline{u}^{j-1},u^j,\overline{u}^{j+1},\dots,\overline{u}^d)}{u^j}&\leq\partial_jC(\overline{u}^1,\dots,\overline{u}^d)\left(\overline{u}^j-1\right)\\C(\overline{u}^1,\dots,\overline{u}^d)-\frac{C(\overline{u}^1,\dots,\overline{u}^{j-1},u^j,\overline{u}^{j+1},\dots,\overline{u}^d)}{u^j}&\leq\frac{C(\overline{u}^1,\dots,\overline{u}^d)}{\overline{u}^j}\left(\overline{u}^j-1\right)\\\frac{C(\overline{u}^1,\dots,\overline{u}^{j-1},u^j,\overline{u}^{j+1},\dots,\overline{u}^d)}{u^j}&\geq\frac{C(\overline{u}^1,\dots,\overline{u}^d)}{\overline{u}^j},\end{aligned}$$

which holds since assumption (6.13) implies that $\frac{C(u^1,\dots,u^d)}{u^j}$ is non-increasing in u^j for all $u^j\in[0,1]$, $(u^1,\dots,u^d)\in[0,1]^d$.
Let $C(\overline{u}^1,\dots,\overline{u}^d)>0$ then

$$\begin{aligned}&C(\overline{u}^1,\dots,\overline{u}^d)u^j-C(\overline{u}^1,\dots,\overline{u}^{j-1},\overline{u}^j\wedge u^j,\overline{u}^{j+1},\dots,\overline{u}^d)\\&\qquad\leq\sum_{\substack{i=1\\i\neq j}}^d\partial_iC(\overline{u}^1,\dots,\overline{u}^d)\left(u^j\overline{u}^i-C_{i,j}(u^j,\overline{u}^i)\right)\\&C(\overline{u}^1,\dots,\overline{u}^d)u^j-C(\overline{u}^1,\dots,\overline{u}^{j-1},\overline{u}^j\wedge u^j,\overline{u}^{j+1},\dots,\overline{u}^d)\\&\qquad\leq\sum_{\substack{i=1\\i\neq j}}^d\frac{C(\overline{u}^1,\dots,\overline{u}^d)}{\overline{u}^i}\left(u^j\overline{u}^i-C_{i,j}(u^j,\overline{u}^i)\right)\end{aligned}$$

$$(d-2)u^j + \frac{C(\overline{u}^1,\ldots,\overline{u}^{j-1},\overline{u}^j \wedge u^j,\overline{u}^{j+1},\ldots,\overline{u}^d)}{C(\overline{u}^1,\ldots,\overline{u}^d)} \geq \sum_{\substack{i=1\\ i\neq j}}^{d} \frac{C_{i,j}(u^j,\overline{u}^i)}{\overline{u}^i}$$

which is true by assumption (6.14). The case $C(\overline{u}^1,\ldots,\overline{u}^d) = 0$ follows by the fact that $\frac{C(\overline{u}^1,\ldots,\overline{u}^d)}{\overline{u}^i} \leq 1$ for all $(\overline{u}^1,\ldots,\overline{u}^d) \in [0,1]^d$. □

Remark 6.3 *Note that in the two-dimensional case, assumption* (6.13) *is equivalent to the left tail increasing property which implies a positive quadrant dependence of the copula C. Loosely speaking this means that the components of C are more likely to be simultaneously small or simulatneously large than in the independent case. More information on different dependence properties can be found in [60] and [80].*

In the following two remarks we give examples of copula distributions which satisfy the assumptions of Theorem 6.3.

Remark 6.4 *Consider a multi-dimensional, one-parametric extension of the Farlie-Gumbel-Morgenstern (FGM) copula given by*

$$C(u^1,\ldots,u^d) = \left(\prod_{i=1}^{d} u^i\right)\left(\alpha\prod_{i=1}^{d}(1-u^i)+1\right)$$

where $\alpha \in [-1,1]$. Simple calculations show that the assumption (6.13) *is true if $\alpha \in [0,1]$. Now consider the right hand-side of* (6.14)

$$\begin{aligned}\sum_{i=1,i\neq j}^{d} \frac{C_{i,j}(u^j,\overline{u}^i)}{\overline{u}^i} &= \sum_{i=1,i\neq j}^{d} \frac{u^j\overline{u}^i}{\overline{u}^i}\\ &= (d-1)u^j.\end{aligned}$$

Finally assumption (6.14) *holds since*

$$\begin{aligned}&\frac{C(\overline{u}^1,\ldots,\overline{u}^{j-1},\overline{u}^j \wedge u^j,\overline{u}^{j+1},\ldots,\overline{u}^d)}{C(\overline{u}^1,\ldots,\overline{u}^d)}\\ &= \min\left(1, \frac{C(\overline{u}^1,\ldots,\overline{u}^{j-1},u^j,\overline{u}^{j+1},\ldots,\overline{u}^d)}{C(\overline{u}^1,\ldots,\overline{u}^d)}\right)\\ &= \min\left(1, \frac{\left(\prod_{i=1,i\neq j}^{d}\overline{u}^i\right)u^j\left(\alpha\prod_{i=1,i\neq j}^{d}(1-\overline{u}^i)(1-u^j)+1\right)}{\left(\prod_{i=1}^{d}\overline{u}^i\right)\left(\alpha\prod_{i=1}^{d}(1-\overline{u}^i)+1\right)}\right)\\ &= \min\left(1, u^j\frac{\left(\alpha\prod_{i=1,i\neq j}^{d}(1-\overline{u}^i)(1-u^j)+1\right)}{\overline{u}^j\left(\alpha\prod_{i=1}^{d}(1-\overline{u}^i)+1\right)}\right)\end{aligned}$$

$$\geq u^j$$

for $\alpha \in [0,1]$.
Note that the independence copula $C(u^1,\ldots,u^d) = \prod_{i=1}^d u^i$ *is the special case of the FGM copula with* $\alpha = 0$*, therefore Theorem 6.3 holds also for the independence copula.*

Remark 6.5 *A multi-dimension version of the Ali-Mikhail-Haq (AMH) copula is given by*

$$C(u^1,\ldots,u^d) = \frac{\prod_{i=1}^d u^i}{1-\alpha\prod_{i=1}^d(1-u^i)}$$

where $\alpha \in [-1,1]$*. As in the previous example it is easy to see that* (6.13) *is fullfilled if* $\alpha \in [0,1]$*. To prove* (6.14) *consider again the term on the right hand-side*

$$\begin{aligned}\sum_{i=1,i\neq j}^d \frac{C_{i,j}(u^j,\overline{u}^i)}{\overline{u}^i} &= \sum_{i=1,i\neq j}^d \frac{u^j\overline{u}^i}{\overline{u}^i}\\ &= (d-1)u^j.\end{aligned}$$

Furthermore Theorem 6.3 applies since

$$\begin{aligned}&\frac{C(\overline{u}^1,\ldots,\overline{u}^{j-1},\overline{u}^j\wedge u^j,\overline{u}^{j+1},\ldots,\overline{u}^d)}{C(\overline{u}^1,\ldots,\overline{u}^d)}\\ &= \min\left(1,\frac{C(\overline{u}^1,\ldots,\overline{u}^{j-1},u^j,\overline{u}^{j+1},\ldots,\overline{u}^d)}{C(\overline{u}^1,\ldots,\overline{u}^d)}\right)\\ &= \min\left(1,u^j\frac{\left(\prod_{i=1,i\neq j}^d\overline{u}^i\right)\left(1-\alpha\prod_{i=1}^d(1-\overline{u}^i)\right)}{\left(\prod_{i=1}^d\overline{u}^i\right)\left(1-\alpha\prod_{i=1}^d(1-\overline{u}^i)(1-u_j)\right)}\right)\\ &\geq u^j\end{aligned}$$

6.4 Application to option pricing

In this section we illustrate the effectiveness of Latin hypercube sampling with dependence in basket option pricing problems. The derivatives which we consider are Asian and lookback basket options. Let $(S_t)_{t\geq 0}$ be a d-dimensional vector of asset price processes and let $(S_t^j)_{t\geq 0}$ denote its j-th component. Then the price of an Asian basket call option is given by

$$\text{ABC} = \mathbb{E}\Big[e^{-rT}\Big(\frac{1}{m}\sum_{j=1}^m\frac{1}{d}\sum_{i=1}^d S_{t_j}^i - K\Big)^+\Big],$$

where $K>0$ denotes the fixed strike price, d is the number of underlying assets, $0=t_0<t_1<t_2<\ldots<t_m=T$ denote the observation points, T is the maturity of the option and r denotes

the risk free interest rate. Similarly, the price of a discrete lookback basket call option is given by

$$\text{DLC} = \mathbb{E}\Big[e^{-rT}\Big(\max_{j=1,\dots,m}\frac{1}{d}\sum_{i=1}^{d}S^i_{t_j} - K\Big)^+\Big].$$

As a model for the asset price process $(S^j_t)_{t\geq 0}$ of each asset $j = 1,\dots,d$, we use

$$S^j_t = S^j_0 e^{(w^j - r)t + X^j_t}, \qquad j = 1,\dots,d, t \geq 0,$$

where $w^j \in \mathbb{R}$ are constants, $S^j_0 > 0$ denote the constant initial asset values and X^j_t are variance gamma (VG) processes for $j = 1,\dots,d$. The VG process $(X^j_t)_{t\geq 0}$ with parameters $(\theta^j, \sigma^j, c^j)$, which was first introduced by [75], is defined as a subordinated Brownian motion by

$$X^j_t = X^j_t(\theta^j, \sigma^j, c^j) = B^j_{G^j_t(c^j,1)}(\theta^j, \sigma^j), \qquad j = 1,\dots,d,\ t \geq 0, \tag{6.15}$$

where $B^j_t(\theta^j, \sigma^j)$ are independent Brownian motions with drift parameters θ^j and volatility parameters $\sigma^j, j = 1,\dots,d$, and $G^j_t(c^j, 1)$ are independent gamma processes independent of $B^j, j = 1,\dots,d$ with drift equal to one and volatility $c^j > 0$. To ensure that the discounted value of a portfolio invested in the asset is a martingale, we choose

$$w^j = \log(1 - \mu^j c^j - (\sigma^j)^2 c^j/2)/c^j, \qquad j = 1,\dots,d.$$

By [74] a VG process can also be represented as the difference of two independent gamma processes, ie $X^j_t = G^{+,j}_t - G^{-,j}_t, j = 1,\dots,d$. Let (μ^j_+, ν^j_+) and (μ^j_-, ν^j_-) denote the parameters of the gamma processes $G^{+,j}, G^{-,j}$, respectively. These pairs of parameters can be easily calculated from the parameters in equation (6.15) through

$$\mu^j_\pm = (\sqrt{(\theta^j)^2 + 2(\sigma^j)^2/c^j} \pm \theta^j)/2, \qquad \nu^j_\pm = (\mu^j_\pm)^2 c^j, \qquad j = 1,\dots,d.$$

Due to the fact that a gamma process has non-decreasing paths, $G^{+,j}_t$ corresponds to the positive movements of X^j_t and $G^{-,j}_t$ corresponds to the negative movements of X^j_t. Our assumption is that all positive movements of components of $X_t = (X^1_t,\dots,X^d_t)$ are dependent and all negative movements of components of X_t are dependent, but positive (negative) movements of the j-th component are independent of negative (positive) movements of all other components, for all $j = 1,\dots,d$. The dependence structure between positive and negative movements will be modelled by copulae $C^\pm$, respectively. Summarising, the increment of the d-dimensional gamma processes in the interval $[t_{i-1}, t_i]$ given by $(G^{\pm,1}_{t_i} - G^{\pm,1}_{t_{i-1}},\dots, G^{\pm,d}_{t_i} - G^{\pm,d}_{t_{i-1}})$ has cumulative distribution function $C^\pm(F^{-1}_{1,\pm},\dots,F^{-1}_{d,\pm})$, where $F^{-1}_{j,\pm}$ is the inverse cumulative distribution function of a gamma distribution with the specific parameters of the j-th asset.

6.4.1 Numerical results

In this subsection, we compare the performance of LHSD with a standard Monte Carlo method in option pricing problems.

Parameters of the numerical examples

VG parameters:	
$\mu_j, j = 1, \ldots, d$	-0.2859
$\sigma_j, j = 1, \ldots, d$	0.1927
$c_j, j = 1, \ldots, d$	0.2505
Option parameters:	
number of assets d	10
maturity T	1
initial asset price $S_0^j, j = 1, \ldots, d$	100
risk free interest rate r	0.05
number of monitoring points k	4
time between monitoring points $t_i - t_{i-1}, i = 1, \ldots, k$	0.25
Simulation parameters:	
number of simulated option prices per estimator n	8000
number of simulations of the estimators m	100
choice of parameters $\eta_{i,n}^j, j = 1, \ldots, d, i = 1, \ldots, n$	0.5

Table 6.1: Parameters sets for the VG processes, the options and the simulations.

The parameters of the underlying VG processes are stated in Table 6.1 and are the same for all components of $(S_t)_{t\geq 0}$. The parameter values are taken from a calibration of the VG process against options on the S&P 500 index by [56]. We observed in price valuations, which we do not state here in detail, that the computation of one LHSD estimator took about 1.4 times of the computation time of a corresponding Monte Carlo estimator. Nevertheless in our concrete implementation the most time consuming part was the transformation of uniformly distributed random variables into gamma distributed random variables. This has to be done only once for all LHSD estimations since by (6.2) where $\eta_{i,n}^j = 1/2, j = 1, \ldots, d, i = 1, \ldots, n$ one only needs fixed quantiles of the gamma distribution. Therefore computation of 4000 LHSD estimators was about five times faster than the computation of 4000 Monte Carlo estimators. One the other hand for the Monte Carlo estimator, one has to perform the transformation dn times for each estimator.

Using the parameters of Table 6.1, the evaluation of each of the option values included the computation of an 80-dimensional integral. Standard deviation and variance were computed based on the $m = 100$ runs of the LHSD and MC estimators. The ratios in columns 6 and 7 of each table were computed as the quotient of MC value and LHSD value.

It is obvious that the effectiveness of LHSD compared to MC decreases with increasing strike price K. The same phenomenon was also observed by [88] in a multi-dimensional Black-Scholes model for the LHSD estimator and by [42] for the standard LHS estimator.

Prices of Asian basket call options with varying strike price K

α	K	LHSD	MC	SD. LHSD	SD. MC	SD. ratio	Var. ratio
0.5	80	22.0542	22.0448	0.00071	0.00748	10.419	108.575
0.5	90	12.5511	12.5419	0.00080	0.00748	9.270	85.944
0.5	100	3.79294	3.78732	0.00241	0.00621	2.577	6.642
0.5	110	0.17227	0.17210	0.00119	0.00140	1.174	1.379
0.5	120	0.00024	0.00024	0.000040	0.000041	1.009	1.018

Table 6.2: Prices of Asian basket call options, where the dependence structure of positive and negative movements are modelled by a FGM copula with parameter α.

Prices of Lookback basket call options with varying strike price K

α	K	LHSD	MC	SD. LHSD	SD. MC	SD. ratio	Var. ratio
0.5	80	25.662	25.658	0.00294	0.00839	2.850	8.125
0.5	90	16.151	16.147	0.00294	0.00839	2.850	8.125
0.5	100	6.893	6.890	0.00322	0.00760	2.356	5.553
0.5	110	1.192	1.192	0.00305	0.00406	1.332	1.775
0.5	120	0.060	0.060	0.00086	0.00089	1.029	1.060

Table 6.3: Prices of Lookback basket call options, where the dependence structure of positive and negative movements are modelled by a FGM copula with parameter α.

Part IV

Lookback options in the HEJD model

Chapter 7

Pricing and hedging of lookback options in the HEJD model

7.1 Introduction

It has been known for many years that the classic Black-Scholes model suffers from many shortcomings and is not capable of explaining many important stylised facts of financial markets, like skewed and heavy tailed return distributions, or the thereby introduced volatility smile/skew. Thus, despite the superior analytical tractability of the geometric Brownian motion model, many authors proposed the more general class of Lévy processes as underlying model for prices of financial quantities. Most definitely we cannot do justice to the vast literature in this field and we limit ourselves to cite just three classics related to our work, namely [10], [20] or [64], and refer the reader to those and references therein for more details on the use of Lévy processes in finance. However, the extra flexibility of Lévy driven financial models often comes at the cost of more complicated pricing algorithms for exotic path-dependent options. The purpose of this article is thus to contribute to the development of more efficient pricing algorithms for certain popular exotic derivatives. More precisely, we will calculate the (time-)Laplace transformed price of different kinds of lookback options and propose and test an efficient inversion algorithm for this transform.

Loosely speaking, there are three approaches for pricing derivatives related to the maximum or minimum of the asset price: Monte Carlo methods, Partial (integro)-differential equations (PIDE) schemes, and Laplace transform based methods, where the latter ones, if applicable, are in general preferable in terms of performance. Focusing on the Laplace transform approach we would like to mention the very nice theoretical discussion regarding this kind of methods for general Lévy processes by Eberlein et al. [31], where very general formulae for the (multi)-Laplace transformed prices of many different option types were derived. For general Lévy processes these formulae have the drawback that the inversion of the Laplace transform is typically quite involved and for a numerical evaluation several numerical integrations need to be performed. However, for some particular Lévy processes these formulae simplify significantly and option prices can be calculated by applying just a standard one-dimensional inversion. For example, Kou [64] proposed a financial market model (typically called Kou model), in which the logarith-

mic asset price process is described by a jump diffusion with two-sided exponential jumps and showed that in this setting, the Laplace transform of several exotic options, including lookback options, can be given in an analytic way (see [65]). Notably, for the same class of processes, Sepp [95] presents a PIDE approach for the pricing of lookback options.
The Kou model also sets the basis for the more general hyper-exponential jump diffusion model (HEJD), where the up- and downward jumps are not modeled by a single exponential random variable, but by a mixture of several exponential random variables with different parameters. Apart from the obvious advantage of more flexibility the main motivation for considering this kind of models was established by Jeannin and Pistorius [59], who showed that many frequently used Lévy based financial models can be approximated arbitrarily well by HEJD processes. More precisely, any process in the class of the so-called general hyper-exponential Lévy processes, that includes e.g. the normal inverse Gaussian (NIG) [10], or the CGMY process [20] to name only two, can be represented as a limit of a sequence of HEJD processes. Moreover, Jeannin and Pistorius also derived the time-Laplace transforms prices of barrier and digital options, and some sensitivities, within the framework of HEJD models. Pricing of double barrier options in HEJD models was discussed by N. Cai et al. [17] where also formulae for the first passage time and related identities of HEJD processes are given. Those two papers also form the basis of this work, where we slightly extend the existing results to apply them to the problem of pricing lookback options.
Apart from applications in asset pricing jump diffusion models, in particular models with exponentially distributed jump sizes, are frequently used in ruin theory, see e.g. Albrecher et al. [6]. A detailed overview on this topic can be found in the book of Asmussen and Albrecher [7].
The rest of the paper is organised as follows: in Section 2 we give a brief introduction to HEJD processes and the Wiener-Hopf factorisation for HEJD processes. In the third section, we derive prices and sensitivities for different types of lookback options and in Section 4 we justify the approximation of lookback option price under a NIG process by corresponding prices under a HEJD process. A numerical analysis of our methods concludes the paper in Section 5.

7.2 Introduction to HEJD models and preliminary results

We will consider lookback options and similar derivatives on an underlying asset, the price process of which, $(S_t)_{t\geq 0}$, is given as $S_t = S_0 e^{X_t}$, where $S_0 > 0$. We assume $\mathbb{E}[e^{X_t}] = e^{rt}$ for all $t \geq 0$, where r denotes the risk free interest rate and $(X_t)_{t\geq 0}$ to be a Lévy process with $X_0 = 0$ a.s.
To valuate a lookback option we have to analyse the supremum and infimum process of the asset price process. Let us hence define

$$\overline{X}_t = \sup_{0\leq s\leq t} X_s, \qquad \underline{X}_t = \inf_{0\leq s\leq t} X_s$$

and recall the well-known Wiener-Hopf factorisation.

Theorem 7.1 (Wiener-Hopf factorisation) *Let $(X_t)_{t\geq 0}$ be a Lévy process in $\mathbb{R}$ and $(\overline{X}_t)_{t\geq 0}$ and $(\underline{X}_t)_{t\geq 0}$ its supremum and infimum process, respectively. Furthermore, let θ be an exponentially distributed random variable with parameter q. Then the characteristic function of $(X_t)_{t\geq 0}$*

at the random time θ can be factorised in the following way:

$$\mathbb{E}[e^{izX_\theta}] = \mathbb{E}[e^{iz\overline{X}_\theta}]\,\mathbb{E}[e^{iz\underline{X}_\theta}], \qquad \forall z \in \mathbb{R},$$

or equivalently

$$\frac{q}{q - \log(\phi_X(z))} = \phi_q^+(z)\phi_q^-(z), \qquad \forall z \in \mathbb{R},$$

where $\phi(z)$ is the characteristic function of X_1, $\phi_q^+(z) = \mathbb{E}[e^{iz\overline{X}_\theta}]$ and $\phi_q^-(z) = \mathbb{E}[e^{iz\underline{X}_\theta}]$.

Additionally, formulae for the Wiener-Hopf factors ϕ_q^- and ϕ_q^+ can be given (see e.g. Sato [93]). For general Lévy processes, however, the actual computation of the factors affords numerical evaluations of high-dimensional numerical integrals. Of course, for some particular types of Lévy processes it is possible to give explicit formulae for ϕ_q^- and ϕ_q^+ (see e.g. Kyprianou [67, Chapter 6]).
A class of Lévy processes, which is well suited for asset price models and allows for considerably simplified formulae for the Wiener-Hopf factors and other identities are jump diffusions with phase-type distributed jumps (cf. Asmussen et al. [8]). Here, at least in the first sections, we concentrate on a special kind of the this last category of Lévy processes, more precisely on so called hyper-exponential jump-diffusions.

Definition 7.1 (Hyper-exponential jump-diffusion) *Let X_t be a Lévy process with $X_0 = 0$ a.s. , then X_t is called hyper-exponential jump-diffusion (HEJD), if it has the following representation*

$$X_t = \mu t + \sigma W_t + \sum_{i=1}^{N_+} Y_i^+ + \sum_{j=1}^{N_-} Y_j^-, \qquad t \geq 0,$$

where W is a Wiener process, N_+, N_- are Poisson processes with parameters $\lambda_+ > 0$ and $\lambda_- > 0$, respectively and $(Y_j^+), (Y_j^-)$ are i.i.d. families of mixed exponential random variables, i.e.

$$Y_j^+ = \sum_{i=1}^{n^+} p_i^+ Z_i^+, \qquad Y_j^- = \sum_{i=1}^{n^-} p_i^- Z_i^-,$$

where $\sum_{i=1}^{n^+} p_i^+ = \sum_{j=1}^{n^-} p_j^- = 1$, $p_i^+ > 0, i = 1, \ldots, n^+$, $p_j^- > 0, j = 1, \ldots, n^-$ and Z_i^+, Z_i^- are exponentially distributed with means $\alpha_i^+ > 0$ and $\alpha_i^- > 0$, respectively. Moreover, all random variables and processes are assumed to be independent.

By the Lévy-Khinchin formula (see e.g. Sato [93]), the characteristic exponent of a HEJD can be written as

$$\begin{aligned}\phi(u) = \log(E[e^{iuX_1}]) = ui\mu - \frac{\sigma^2}{2}u^2 + \lambda^+ \sum_{k=1}^{n^+} p_k^+ \Big(\frac{\alpha_k^+}{\alpha_k^+ - ui} - 1\Big) \\ + \lambda^- \sum_{j=1}^{n^-} p_j^- \Big(\frac{\alpha_j^-}{\alpha_j^- + ui} - 1\Big). \qquad (7.1)\end{aligned}$$

The function $\phi(u)$ can be extended analytically (cf. e.g. Sato [93, Chapter 9]) to the whole complex plane except for the finite sets $\{-\mathrm{i}\alpha_i^+, \text{ for } i = 1, \dots, n^+\}, \{-\mathrm{i}\alpha_i^-, \text{ for } i = 1, \dots, n^-\}$ and we will denote the roots of the Cramér-Lundberg equation $-q + \phi(-is) = 0$ with positive and negative real part, by $\rho_i^+(q), i = 1, \dots, n^+ + 1$ and $\rho_i^-(q), i = 1, \dots, n^- + 1$, respectively. Applying the formulae for general two-sided phase-type distributed jumps on HEJD processes, we find

$$\phi_q^+(u) = \frac{\prod_{k=1}^{n^+}\left(1 - \frac{ui}{\alpha_k^+}\right)}{\prod_{k=1}^{m^+}\left(1 - \frac{ui}{\rho_k^+(q)}\right)} \qquad \text{and} \qquad \phi_q^-(u) = \frac{\prod_{k=1}^{n^-}\left(1 + \frac{ui}{\alpha_k^-}\right)}{\prod_{k=1}^{m^-}\left(1 - \frac{ui}{\rho_k^-(q)}\right)}.$$

Moreover, the time-Laplace transforms of the distributions of $\overline{X}_t$ and $\underline{X}_t$ can be calculated explicitly (cf. Mordecki [79])

$$\int_0^\infty e^{-qt} P(\overline{X}_t \le z)dt = \frac{1}{q}\Big(1 - \sum_{k=1}^{m^+} A_k^+(q) e^{-\rho_k^+(q)z}\Big), \qquad z \ge 0 \tag{7.2}$$

$$\int_0^\infty e^{-qt} P(-\underline{X}_t \le z)dt = \frac{1}{q}\Big(1 - \sum_{k=1}^{m^-} A_k^-(q) e^{\rho_k^-(q)z}\Big), \qquad z \ge 0 \tag{7.3}$$

where the coefficients $A_k^+(q)$ and $A_k^-(q)$ are given by

$$A_k^+(q) = \frac{\prod_{v=1}^{n^+}\left(1 - \frac{\rho_k^+(q)}{\alpha_v^+}\right)}{\prod_{v=1, v\neq k}^{m^+}\left(1 - \frac{\rho_k^+(q)}{\rho_v^+(q)}\right)}, \tag{7.4}$$

$$A_k^-(q) = \frac{\prod_{v=1}^{n^-}\left(1 + \frac{\rho_k^-(q)}{\alpha_v^-}\right)}{\prod_{v=1, v\neq k}^{m^-}\left(1 - \frac{\rho_k^-(q)}{\rho_v^-(q)}\right)}. \tag{7.5}$$

Let us shortly note here, that another way to understand the above formula is that for any exponentially distributed random variable θ, $\overline{X}_\theta$ and $\underline{X}_\theta$ are hyper-exponential distributed random variables.

With the notable exception of the Kou model (for which $n^+ = n^- = 1$) the roots of the Cramér Lundberg equation cannot be calculated analytically. However, due to favorable structural properties of the Cramér Lundberg equation the numerical computation of the roots is not difficult and can be efficiently implemented. The following Lemma 2.1, which is a slight extension of [16, Lemma 1], states the precise result.

Lemma 7.1 (Characterisation of the moment generating function of X_t) *The function $\phi(-is)$ is a convex function for $s \in (-\alpha_1^-, \alpha_1^+)$. Furthermore:*

- *If $\sigma > 0$, the equation $-q + \phi(-is) = 0$ for $q \in \mathbb{R}^+$ has roots $\rho_k^+, k = 1, \dots, n^+ + 1 = m^+$ and $\rho_j^-, j = 1, \dots, n^- + 1 = m^-$, which satisfy the condition*

$$-\infty < -\rho_{n^-+1}^-(q) < -\alpha_{n^-}^- < -\rho_{n^-}^-(q) < \ldots < -\rho_2^-(q) < -\alpha_1^- < -\rho_1^-(q) < 0,$$
$$0 < \rho_1^+(q) < \alpha_1^+ < \rho_2^+(q) < \ldots < \rho_{n^+}^+(q) < \alpha_{n^+}^+ < \rho_{n^++1}^+(q) < \infty.$$

- *If $\sigma = 0$ and $\mu > 0$, the equation $-q + \phi(-is) = 0$ for $q \in \mathbb{R}^+$ has roots $\rho_k^+, k = 1, \ldots, n^+ + 1 = m^+$ and $\rho_j^-, j = 1, \ldots, n^- = m^-$, which satisfy the condition*

$$-\infty < -\alpha_{n^-}^- < -\rho_{n^-}^-(q) < \ldots < -\rho_2^-(q) < -\alpha_1^- < -\rho_1^-(q) < 0,$$
$$0 < \rho_1^+(q) < \alpha_1^+ < \rho_2^+(q) < \ldots < \rho_{n^+}^+(q) < \alpha_{n^+}^+ < \rho_{n^++1}^+(q) < \infty$$

- *if $\sigma = 0$ and $\mu < 0$, the equation $-q + \phi(-is) = 0$ for $q \in \mathbb{R}^+$ has roots $\rho_k^+, k = 1, \ldots, n^+ = m^+$ and $\rho_j^-, j = 1, \ldots, n^- + 1 = m^-$, which satisfy the condition*

$$-\infty < -\rho_{n^-+1}^-(q) < -\alpha_{n^-}^- < -\rho_{n^-}^-(q) < \ldots < -\rho_2^-(q) < -\alpha_1^- < -\rho_1^-(q) < 0,$$
$$0 < \rho_1^+(q) < \alpha_1^+ < \rho_2^+(q) < \ldots < \rho_{n^+}^+(q) < \alpha_{n^+}^+ < \infty.$$

- *if $\sigma = 0$ and $\mu = 0$, the equation $-q + \phi(-is) = 0$ for $q \in \mathbb{R}^+$ has roots $\rho_k^+, k = 1, \ldots, n^+ = m^+$ and $\rho_j^-, j = 1, \ldots, n^- = m^-$, which satisfy the condition*

$$-\infty < -\alpha_{n^-}^- < -\rho_{n^-}^-(q) < \ldots < -\rho_2^-(q) < -\alpha_1^- < -\rho_1^-(q) < 0,$$
$$0 < \rho_1^+(q) < \alpha_1^+ < \rho_2^+(q) < \ldots < \rho_{n^+}^+(q) < \alpha_{n^+}^+ < \infty.$$

Proof:
For simplicity of notation, we set $\psi(s) = \phi(-is)$. Note that in every case, $\psi(s)$ is a convex function on $(-\alpha_1^-, \alpha_1^+)$, because it is a sum of convex functions on this interval.
Furthermore, $\psi(s)$ has poles on on the sets $\{\alpha_i^+, \text{ for } i = 1, \ldots, n^+\}, \{\alpha_i^-, \text{ for } i = 1, \ldots, n^-\}$. For a positive pole α_i^+ it follows $\psi(\alpha_i^+ -) = +\infty$ and $\psi(\alpha_i^+ +) = -\infty$ and for a negative pole α_i^- it follows $\psi(\alpha_i^- -) = -\infty$ and $\psi(\alpha_i^- +) = +\infty$. Furthermore $\psi(s)$ is continuous between two poles, so that there is always at least one root of the equation $-q + \phi(-is) = 0$ between two such poles. From the fact that $\psi(0) = 0$ and the convexity of ψ in $(-\alpha_1^-, \alpha_1^+)$, we conclude that there is exactly one root on each of the intervals $(-\alpha_1^-, 0)$ and $(0, \alpha_1^+)$. While all of the observations so far hold in every of the four cases, we will now consider different combinations of σ and μ separately.
If $\sigma > 0$, it follows by simple transformations that the equation $-q + \psi(s) = 0$ has two more roots than $\psi(s)$ has poles and that $\lim_{s \to +\infty} \psi(s) = \lim_{s \to -\infty} \psi(s) = +\infty$. Because of these facts, there is exactly one root in $(-\infty, \alpha_{n^-}^-)$ and $(\alpha_{n^+}^+, +\infty)$. Hence there is exactly one root in each of the intervals $(\alpha_i^+, \alpha_{i+1}^+)$ for $i = 1, \ldots, m^+ - 1$, $(\alpha_{i+1}^-, \alpha_i^-)$ for $i = 1, \ldots, m^- - 1$, $(\alpha_{n^+}^+, +\infty)$ and $(-\infty, \alpha_{n^-}^-)$.
The argumentation is similar in the three remaining cases, where $\sigma = 0$. If $\mu \neq 0$, then $-q + \psi(s) = 0$ has one more root than $\psi(s)$ has poles. Because of $\lim_{s \to +\infty} \psi(s) = +\infty$ if $\mu > 0$ and $\lim_{s \to -\infty} \psi(s) = +\infty$ if $\mu < 0$, there must be a root on $(\alpha_{n^+}^+, +\infty)$ and $(-\infty, \alpha_{n^-}^-)$, respectively. The case $\mu = 0$ and $\sigma = 0$ follows directly from the above considerations. $\square$

7.3 Prices and Greeks of lookback options

In this section we will give pricing formulae for different lookback options on an underlying asset, that is modeled by the exponential of a HEJD. More precisely, we will assume the asset price S to be given as:

$$S_t = S_0 e^{X_t},$$

where X_t is a HEJD process with $\alpha_1 > 1$. This last assumption guarantees that the expectation of the stock price is finite.
We consider two classes of lookback options, namely floating and fixed strike lookback options. Denoting the maturity by T and the strike price by K, the payoff of fixed strike calls and puts are defined by $(\max_{0\leq t\leq T} S_t - K)^+$ with $K \geq S_0$ and $(K - \min_{0\leq t\leq T} S_t)^+$ with $0 < K \leq S_0$, respectively. The prices of these options are given by

$$\mathrm{LC}_{\text{fixed}}(T, S_0, K) = \mathbb{E}[e^{-rT}(\max_{0\leq t\leq T} S_t - K)^+], \quad K \geq S_0, \tag{7.6}$$

and

$$\mathrm{LP}_{\text{fixed}}(T, S_0, K) = \mathbb{E}[e^{-rT}(K - \min_{0\leq t\leq T} S_t)^+], \quad 0 < K \leq S_0,$$

respectively. In the same manner the prices of puts and calls of floating strike lookback options are defined as expectations of their payoffs $(\max\{M, \max_{0\leq t\leq T} S_t\} - S_T)$ and $(S_T - \min\{N, \min_{0\leq t\leq T} S_t\})$, respectively, where $M \geq S_0 \geq N$. Thus

$$\begin{aligned}
\mathrm{LP}_{\text{float}}(T, S_0, M) &= \mathbb{E}[e^{-rT}(\max\{M, \max_{0\leq t\leq T} S_t\} - S_T)] \\
&= \mathbb{E}[e^{-rT}(\max\{M, \max_{0\leq t\leq T} S_t\})] - S_0 \\
&= \mathbb{E}[e^{-rT}(\max_{0\leq t\leq T} S_t - M)^+] + e^{-rT}M - S_0 \\
&= \mathrm{LC}_{\text{fixed}}(T, S_0, M) + e^{-rT}M - S_0
\end{aligned} \tag{7.7}$$

and

$$\begin{aligned}
\mathrm{LC}_{\text{float}}(T, S_0, N) &= \mathbb{E}[e^{-rT}(S_T - \min\{N, \min_{0\leq t\leq T} S_t\})] \\
&= S_0 - \mathbb{E}[e^{-rT}(\min\{N, \min_{0\leq t\leq T} S_t\})], \\
&= S_0 - e^{-rT}N + \mathbb{E}[e^{-rT}(N - \min_{0\leq t\leq T} S_t)^+] \\
&= S_0 - e^{-rT}N + \mathrm{LP}_{\text{fixed}}(T, S_0, N)
\end{aligned} \tag{7.8}$$

It follows by (7.7) and (7.8) that the price of a floating strike lookback put option is just the sum of the price of a fixed strike lookback call option and a constant with respect to X_t. An analogous statement applies to floating strike lookback call options. We will use these facts frequently in the proofs of the following theorems and corollaries.

7.3.1 Prices of lookback options

As mentioned before the aim is to calculate the Laplace transform of prices of lookback options and the following lemma will prove useful for this.

Lemma 7.2 *Let X_t be a HEJD. Then*

$$\lim_{y\to\infty} e^y\,\mathbb{P}[\overline{X}_T \geq y] = 0, \qquad \textit{and} \qquad \lim_{y\to-\infty} e^y\,\mathbb{P}[\underline{X}_T \leq y] = 0, \qquad \forall T \geq 0.$$

Proof:
Observe that $(e^{\theta X_t - \phi(-i\theta)t})_{t\geq 0}$ is a martingale for any $\theta \in (-\alpha_1^-, \alpha_1^+)$. Since $\alpha_1^+ > 1$, ϕ is continuous and $\phi(-i) = r > 0$, there exists some $\beta \in (1, \alpha_1^+)$ such that $\phi(-i\theta) > 0$. Hence

$$e^y\,\mathbb{P}[\overline{X}_T \geq y] = e^{(1-\beta)y} e^{\beta y}\,\mathbb{P}[\tau_y \leq T],$$

where τ_y denotes the first passage time of the process X over a level y. By the optimal sampling theorem the second term can be dominated by

$$e^{\beta y}\,\mathbb{P}[\tau_y \leq T] \leq \mathbb{E}[e^{\beta X_{(\tau_y\wedge T)}}] \leq e^{\phi(-i\beta)T}\,\mathbb{E}[e^{\beta X_{(\tau_y\wedge T)} - \phi(-i\beta)(\tau_y\wedge T)}] = e^{\phi(-i\beta)T}, \qquad (7.9)$$

where the second inequality follows from the fact that $\mathbb{E}[e^{\phi(-i\beta)(T-(\tau_y\wedge T))}] > 1$ and the required result follows since $\beta > 1$. The second limit result follows by applying the same arguments on the dual reflecting process $-\overline{X}_t$. □

Theorem 7.2 *Let $A_k^+(q)$ and $A_k^-(q)$ be given as in* (7.4) *and* (7.5) *and let the negative and positive roots of the equation $\phi(-is) - q = 0$ be given by $\rho_k^-(q), k = 1, \ldots, m^-$ and $\rho_k^+(q), k = 1, \ldots, m^+$, respectively. Then the Laplace transform of the price of a fixed strike lookback call is given by*

$$\int_0^\infty e^{-\alpha T}\,\mathrm{LC}_{\textit{fixed}}(T, S_0, K)dT = S_0 \frac{1}{\alpha + r} \sum_{k=1}^{m^+} A_k^+(\alpha{+}r) \frac{e^{-\log(K/S_0)(\rho_k^+(\alpha+r)-1)}}{\rho_k^+(\alpha + r) - 1} \qquad K \geq S_0,$$

while for a fixed strike lookback put option we have

$$\int_0^\infty e^{-\alpha T}\,\mathrm{LP}_{\textit{fixed}}(T, S_0, K)dT = S_0 \frac{1}{\alpha + r} \sum_{k=1}^{m^-} A_k^-(\alpha{+}r) \frac{e^{-\log(K/S_0)(\rho_k^-(\alpha+r)-1)}}{1 - \rho_k^-(\alpha + r)} \qquad K \leq S_0.$$

Proof:
We need to calculate the Laplace transform of $\mathbb{E}[e^{-rT}(S_0 e^{\overline{X}_T} - K)^+], K \geq S_0$. Defining

$$z = \log(K/S_0) \geq 0,$$

we have

$$\mathbb{E}[e^{-rT}(S_0 e^{\overline{X}_T} - K)^+] = S_0\,\mathbb{E}[e^{-rT}(e^{\overline{X}_T} - e^z)\mathbf{1}_{\{\overline{X}_T \geq z\}}]. \qquad (7.10)$$

Applying integration-by-parts and Lemma 7.2 yields

$$\begin{aligned}\mathbb{E}[e^{-rT}e^{\overline{X}_T}\mathbf{1}_{\{\overline{X}_T\geq z\}}] &= -e^{-rT}\int_z^\infty e^y d\,\mathbb{P}[\overline{X}_T\geq y]\\ &= -e^{-rT}\Big(-e^z\,\mathbb{P}[\overline{X}_T\geq z]-\int_z^\infty e^y\,\mathbb{P}[\overline{X}_T\geq y]dy\Big)\\ &= \mathbb{E}[e^{-rT}e^z\mathbf{1}_{\{\overline{X}_T\geq z\}}]+e^{-rT}\int_z^\infty e^y\,\mathbb{P}[\overline{X}_T\geq y]dy.\end{aligned}$$

Hence

$$S_0\,\mathbb{E}[e^{-rT}(e^{\overline{X}_T}-e^z)\mathbf{1}_{\{\overline{X}_T\geq z\}}] = S_0e^{-rT}\int_z^\infty e^y\,\mathbb{P}[\overline{X}_T\geq y]dy$$

and for all $\alpha>0$

$$\begin{aligned}\int_0^\infty e^{-\alpha T}S_0\,\mathbb{E}[e^{-rT}(e^{\overline{X}_T}-e^z)\mathbf{1}_{\{\overline{X}_T\geq z\}}]dT &= S_0\int_0^\infty e^{-\alpha T}e^{-rT}\int_z^\infty e^y\,\mathbb{P}[\overline{X}_T\geq y]dydT\\ &= S_0\int_z^\infty e^y\int_0^\infty e^{-(\alpha+r)T}\,\mathbb{P}[\overline{X}_T\geq y]dTdy,\end{aligned}$$

where changing the order of integration in the last step is justified by Tonelli's theorem.
Note that the inner integral in the above is exactly the Laplace transform of the distribution of the supremum process $\overline{X}$ and is hence given by equation (7.2), i.e. we have

$$\int_0^\infty e^{-(\alpha+r)T}\,\mathbb{P}[\overline{X}_T\geq y]dT = \frac{1}{\alpha+r}\sum_{k=1}^{m^+}A_k^+(\alpha+r)e^{-\rho_k^+(\alpha+r)y}.$$

By Lemma 7.1 and $\phi(-i)=r$ we have $\min_k\rho_k^+(\alpha+r)>\min_k\rho_k^+(r)=1$ for $\alpha>0$ and therefore,

$$\begin{aligned}\int_0^\infty e^{-\alpha T}S_0\,\mathbb{E}[e^{-rT}(e^{\overline{X}_T}-e^z)\mathbf{1}_{\{\overline{X}_T\geq z\}}]dT &= S_0\int_z^\infty\frac{1}{\alpha+r}\sum_{k=1}^{m^+}A_k^+(\alpha+r)e^{y(1-\rho_k^+(\alpha+r))}dy\\ &= S_0\frac{1}{\alpha+r}\sum_{k=1}^{m^+}A_k^+(\alpha+r)\frac{e^{-\log(K/S_0)(\rho_k^+(\alpha+r)-1)}}{\rho_k^+(\alpha+r)-1},\end{aligned}$$

which proves the first statement.
The second result follows from similar reasoning. □

Corollary 7.1 *Let $0<N\leq S_0\leq M$ and let A_k^+, A_k^-, ρ_k^+ and ρ_k^- be as in Theorem 7.2, and denote the maturity by T. Then we have*

$$\int_0^\infty e^{-\alpha T}\,\mathrm{LP}_{\mathit{float}}(T,S_0,M)dT = S_0\frac{1}{\alpha+r}\sum_{k=1}^{m^+}A_k^+(\alpha+r)\frac{e^{-\log(M/S_0)(\rho_k^+(\alpha+r)-1)}}{\rho_k^+(\alpha+r)-1}+\frac{M}{\alpha+r}-\frac{S_0}{\alpha},\tag{7.11}$$

$$\int_0^\infty e^{-\alpha T}\,\mathrm{LC}_{\textit{float}}(T,S_0,N)dT = \frac{S_0}{\alpha} + S_0\frac{1}{\alpha+r}\sum_{k=1}^{m^-} A_k^-(\alpha+r)\frac{e^{-\log(N/S_0)(\rho_k^-(\alpha+r)-1)}}{1-\rho_k^-(\alpha+r)} - \frac{N}{\alpha+r}. \tag{7.12}$$

Proof:
The proof follows immediately from Theorem 7.2, (7.7) and (7.8). □

7.3.2 Greeks of lookback options

In this subsection, we use the results of the previous subsections to give formulae for the Laplace transforms of sensitivities of lookback options. We derive expressions for Θ_V, Δ_V and Γ_V, which are defined by

$$\Delta_V = \frac{\partial V}{\partial S_0}, \qquad \Gamma_V = \frac{\partial^2 V}{\partial S_0^2}, \qquad \Theta_V = \frac{\partial V}{\partial T},$$

where S_0 denotes the initial price of the underlying asset, T is the maturity of the option and V is the price of an option on the underlying asset.

Theorem 7.3 *Suppose X_t is a HEJD process with $\sigma > 0$ and let $\alpha > 0$. Then the Laplace transforms of $\Delta_{\mathrm{LC}_{\textit{fixed}}}$ and $\Gamma_{\mathrm{LC}_{\textit{fixed}}}$ are given by*

$$\widehat{\Delta}_{\mathrm{LC}_{\textit{fixed}}}(\alpha) = \frac{1}{\alpha+r}\sum_{k=1}^{m^+} A_k^+(\alpha+r)\rho_k^+(\alpha+r)\frac{e^{-\log(K/S_0)(\rho_k^+(\alpha+r)-1)}}{\rho_k^+(\alpha+r)-1}, \qquad 0 < S_0 \le K, \tag{7.13}$$

$$\widehat{\Gamma}_{\mathrm{LC}_{\textit{fixed}}}(\alpha) = \frac{1}{\alpha+r}\frac{1}{S_0}\sum_{k=1}^{m^+} A_k^+(\alpha+r)\rho_k^+(\alpha+r)e^{-\log(K/S_0)(\rho_k^+(\alpha+r)-1)}. \qquad 0 < S_0 \le K. \tag{7.14}$$

The Greeks of fixed strike lookback put option are given by

$$\widehat{\Delta}_{\mathrm{LP}_{\textit{fixed}}}(\alpha) = \frac{1}{\alpha+r}\sum_{k=1}^{m^-} A_k^-(\alpha+r)\rho_k^-(\alpha+r)\frac{e^{-\log(K/S_0)(\rho_k^-(\alpha+r)-1)}}{1-\rho_k^-(\alpha+r)}, \qquad 0 \le K \le S_0, \tag{7.15}$$

$$\widehat{\Gamma}_{\mathrm{LP}_{\textit{fixed}}}(\alpha) = -\frac{1}{\alpha+r}\frac{1}{S_0}\sum_{k=1}^{m^-} A_k^-(\alpha+r)\rho_k^-(\alpha+r)e^{-\log(K/S_0)(\rho_k^-(\alpha+r)-1)}, \qquad 0 \le K \le S_0. \tag{7.16}$$

Proof:
First note that Δ_V and Γ_V exist and are continuous, since $\mathrm{LP}_{\text{fixed}}$ and $\mathrm{LP}_{\text{fixed}}$ can be understood

(viewed as function of S_0) as convolution of the continuous density of $\overline{X}_T$ and the function $f(x) = (x-c)^+$, the second derivative of which in the sense of distributions is given by the Dirac-Delta measure.
Formulae (7.13) – (7.16) all directly follow by interchanging differentiation and the Laplace transform. So we only have to show that changing the order is in fact justified. To this end note that

$$\begin{aligned}&\frac{\partial}{\partial S_0}\int_0^\infty e^{-\alpha T}\,\mathrm{LC}_{\text{fixed}}(T,S_0,K)dT\\ &=\lim_{\epsilon\to 0}\frac{1}{\epsilon}\int_0^\infty e^{-\alpha T}\Big(\mathrm{LC}_{\text{fixed}}(T,S_0+\epsilon,K)-\mathrm{LC}_{\text{fixed}}(T,S_0,K)\Big)dT.\end{aligned}$$

Now the aim is to apply the dominated convergence theorem on the difference quotient. Observe that

$$\begin{aligned}&\frac{1}{\epsilon}e^{-\alpha T}\Big|\mathrm{LC}_{\text{fixed}}(T,S_0+\epsilon,K)-\mathrm{LC}_{\text{fixed}}(T,S_0,K)\Big|\\ &\quad=\frac{1}{\epsilon}e^{-(\alpha+r)T}\,\mathbb{E}\Big[\big((S_0+\epsilon)e^{\overline{X}_T}-K\big)\mathbf{1}_{\{e^{\overline{X}_T}\geq K/(S_0+\epsilon)\}}-\big(S_0e^{\overline{X}_T}-K\big)\mathbf{1}_{\{e^{\overline{X}_T}\geq K/S_0\}}\Big]\\ &\quad=\frac{1}{\epsilon}e^{-(\alpha+r)T}\,\mathbb{E}\Big[\epsilon e^{\overline{X}_T}\mathbf{1}_{\{e^{\overline{X}_T}\geq K/S_0\}}+\big((S_0+\epsilon)e^{\overline{X}_T}-K\big)\mathbf{1}_{\{K/(S_0+\epsilon)\leq e^{\overline{X}_T}\leq K/S_0\}}\Big]\\ &\quad\leq e^{-(\alpha+r)T}\Big(\mathbb{E}[e^{\overline{X}_T}\mathbf{1}_{\{e^{\overline{X}_T}\geq K/S_0\}}]+\frac{K}{S_0}\Big)\\ &\quad\leq e^{-(\alpha+r)T}\Big(\mathbb{E}\big[\max(e^{\overline{X}_T},K/S_0)\big]+\frac{K}{S_0}\Big).\end{aligned}$$

Furthermore we have that

$$\begin{aligned}&\int_0^\infty e^{-(\alpha+r)T}\Big(\mathbb{E}\big[\max(e^{\overline{X}_T},K/S_0)\big]+\frac{K}{S_0}\Big)dT\\ &=\int_0^\infty e^{-(\alpha+r)T}\,\mathbb{E}\big[\max(e^{\overline{X}_T},K/S_0)\big]dT+\frac{K}{(\alpha+r)S_0},\end{aligned}$$

where the first term on the right hand-side was already calculated and shown to be finite in the proof of Theorem 7.2. Thus the dominated convergence theorem can be applied to justify the interchange of integration and differentiation and we have

$$\begin{aligned}\frac{\partial}{\partial S_0}\int_0^\infty e^{-\alpha T}\,\mathrm{LC}_{\text{fixed}}(T,S_0,K)dT&=\int_0^\infty e^{-\alpha T}\frac{\partial}{\partial S_0}\,\mathrm{LC}_{\text{fixed}}(T,S_0,K)dT\\ &=\int_0^\infty e^{-\alpha T}\Delta_{\mathrm{LC}_{\text{fixed}}}dT=\widehat{\Delta}_{\mathrm{LC}_{\text{fixed}}}(\alpha).\end{aligned}$$

The argumentation in the case of $\Gamma_{\mathrm{LC}_{\text{fixed}}}$ is similar. Again we consider the differentiation quotient and again we want to apply the dominated convergence theorem. First note that

$$\frac{1}{\epsilon^2}e^{-\alpha T}\big|(\,\mathrm{LC}_{\text{fixed}}(T,S_0+\epsilon,K)-2\,\mathrm{LC}_{\text{fixed}}(T,S_0,K)+\mathrm{LC}_{\text{fixed}}(T,S_0-\epsilon,K)\big|$$

$$
\begin{aligned}
&= \frac{1}{\epsilon^2} e^{-(\alpha+r)T} \Big| \mathbb{E}\Big[\epsilon e^{\overline{X}_T} \mathbf{1}_{\{e^{\overline{X}_T} \geq K/S_0\}} + ((S_0+\epsilon)e^{\overline{X}_T} - K)\mathbf{1}_{\{K/(S_0+\epsilon) \leq e^{\overline{X}_T} \leq K/S_0\}}\Big] \\
&\qquad - \mathbb{E}\Big[\epsilon e^{\overline{X}_T} \mathbf{1}_{\{e^{\overline{X}_T} \geq K/(S_0-\epsilon)\}} + (S_0 e^{\overline{X}_T} - K)\mathbf{1}_{\{K/S_0 \leq e^{\overline{X}_T} \leq K/(S_0-\epsilon)\}}\Big]\Big| \\
&\leq \frac{1}{\epsilon^2} e^{-(\alpha+r)T}\, \mathbb{E}\Big[\epsilon e^{\overline{X}_T} |\mathbf{1}_{\{e^{\overline{X}_T} \geq K/S_0\}} - \mathbf{1}_{\{e^{\overline{X}_T} \geq K/(S_0-\epsilon)\}}|\Big] \\
&\quad + \frac{1}{\epsilon^2} e^{-(\alpha+r)T}\, \mathbb{E}\Big[\epsilon e^{\overline{X}_T} \mathbf{1}_{\{K/(S_0+\epsilon) \leq e^{\overline{X}_T} \leq K/S_0\}}\Big] \\
&\quad + \frac{1}{\epsilon^2} e^{-(\alpha+r)T}\, \mathbb{E}\Big[\big|(S_0 e^{\overline{X}_T} - K)(\mathbf{1}_{\{e^{\overline{X}_T} \geq K/S_0\}} - \mathbf{1}_{\{e^{\overline{X}_T} \geq K/(S_0-\epsilon)\}})\big|\Big] \\
&\leq \frac{1}{\epsilon} e^{-(\alpha+r)T}\, \mathbb{E}\Big[e^{\overline{X}_T} \mathbf{1}_{\{K/(S_0+\epsilon) \leq e^{\overline{X}_T} \leq K/(S_0-\epsilon)\}}\Big] \\
&\quad + \frac{1}{\epsilon^2} e^{-(\alpha+r)T}\, \mathbb{E}\Big[\big|S_0 e^{\overline{X}_T} - K\big| \mathbf{1}_{\{K/(S_0+\epsilon) \leq e^{\overline{X}_T} \leq K/(S_0-\epsilon)\}}\Big] \\
&\leq e^{-(\alpha+r)T} \frac{K+1}{S_0-\epsilon} \frac{\mathbb{P}\big[K/(S_0+\epsilon) \leq e^{\overline{X}_T} \leq K/(S_0-\epsilon)\big]}{\epsilon}.
\end{aligned}
$$

Hence the dominated convergence theorem can be applied, if

$$\int_0^\infty e^{-(\alpha+r)T} \frac{\mathbb{P}\big[K/(S_0+\epsilon) \leq e^{\overline{X}_T} \leq K/(S_0-\epsilon)\big]}{\epsilon} dT < \infty,$$

for any sufficiently small $\epsilon > 0$. In fact, this is easily seen to be the case, if the distribution of $\overline{X}_T$ admits a density.

Hence, finally, we have to argue why the density of the distribution of $\overline{X}_T$ exists. For this purpose, we use a result of Chaumont [22, Theorem 2], who states that $\overline{X}_T$ is absolutely continuous for $T > 0$ with respect to the Lebesgue measure on $\mathbb{R}^+$ if and only if the potential measure of X_T is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^+$ and 0 is a regular point for $(-\infty, 0)$ and $(0, \infty)$. Since $\sigma^2 > 0$, 0 is a regular point for both intervals in our case and following Bertoin [12, Theorem II.16] we get that the absolute continuity of the potential measure is equivalent to

$$\int_{\mathbb{R}} \Re\Big(\frac{1}{q + \phi(x)}\Big) dx < \infty, \tag{7.17}$$

where $\phi(x)$ is the characteristic exponent of X_T given in (7.1). Using Lemma 7.1, we conclude that all singularities and roots of $q + \log(\phi(x))$ have non-zero imaginary part and since $\sigma^2 > 0$, $\phi(x)$ is a polynomial of degree $n^+ + n^- + 2$ divided by a polynomial of degree $n^+ + n^-$, thus it follows that integral in (7.17) is finite.

The equations (7.15) and (7.16) follow by similar arguments. □

Corollary 7.2 *Under the assumptions of Theorem 7.3, the first and second order derivatives of prices of floating strike lookback options with respect to S_0 are given by*

$$\widehat{\Delta}_{\mathrm{LP}_{\textit{float}}}(\alpha) = \frac{1}{\alpha + r} \sum_{k=1}^{m^+} A_k^+(\alpha+r)\rho_k^+(\alpha+r) \frac{e^{-\log(M/S_0)(\rho_k^+(\alpha+r)-1)}}{\rho_k^+(\alpha+r) - 1} - \frac{1}{\alpha}, \qquad 0 < S_0 \leq M,$$

$$\widehat{\Gamma}_{\mathrm{LP}_{\textit{float}}}(\alpha) = \frac{1}{\alpha+r}\frac{1}{S_0}\sum_{k=1}^{m^+} A_k^+(\alpha+r)\rho_k^+(\alpha+r)e^{-\log(M/S_0)(\rho_k^+(\alpha+r)-1)}. \qquad 0 < S_0 \leq M,$$

$$\widehat{\Delta}_{\mathrm{LC}_{\textit{float}}}(\alpha) = \frac{1}{\alpha} + \frac{1}{\alpha+r}\sum_{k=1}^{m^-} A_k^-(\alpha+r)\rho_k^-(\alpha+r)\frac{e^{-\log(N/S_0)(\rho_k^-(\alpha+r)-1)}}{1-\rho_k^-(\alpha+r)}, \qquad 0 \leq N \leq S_0,$$

$$\widehat{\Gamma}_{\mathrm{LC}_{\textit{float}}}(\alpha) = -\frac{1}{\alpha+r}\frac{1}{S_0}\sum_{k=1}^{m^-} A_k^-(\alpha+r)\rho_k^-(\alpha+r)e^{-\log(N/S_0)(\rho_k^-(\alpha+r)-1)}, \qquad 0 \leq N \leq S_0.$$

Proof:
The corollary follows directly by Corollary 7.1, Theorem 7.3, (7.7) and (7.8). □

Theorem 7.4 *Suppose X_t is a HEJD process with σ and let $\alpha > 0$. Then the Laplace transforms of the sensitivities of fixed strike lookback options with respect to the maturity T are given by*

$$\widehat{\Theta}_{\mathrm{LC}_{\textit{fixed}}}(\alpha) = \alpha\widehat{\mathrm{LC}}_{\textit{fixed}}(\alpha) = S_0\frac{\alpha}{\alpha+r}\sum_{k=1}^{m^+} A_k^+(\alpha+r)\frac{e^{-\log(K/S_0)(\rho_k^+(\alpha+r)-1)}}{\rho_k^+(\alpha+r)-1}, \qquad K \geq S_0, \tag{7.18}$$

$$\widehat{\Theta}_{\mathrm{LP}_{\textit{fixed}}}(\alpha) = \alpha\widehat{\mathrm{LP}}_{\textit{fixed}}(\alpha) = S_0\frac{\alpha}{\alpha+r}\sum_{k=1}^{m^-} A_k^-(\alpha+r)\frac{e^{-\log(K/S_0)(\rho_k^-(\alpha+r)-1)}}{1-\rho_k^-(\alpha+r)}, \qquad K \leq S_0. \tag{7.19}$$

Proof:
Note that the first equations in (7.18) and (7.19) are classic for Laplace transforms, given that $\mathrm{LC}_{\text{fixed}}$ and $\mathrm{LP}_{\text{fixed}}$ are differentiable (with respect to T). Here we will show that both are Lipschitz continuous and thus almost everywhere differentiable, which is sufficient for the beforementioned results to apply.
Thus let us turn to the proof of the Lipschitz continuity and note that

$$\begin{aligned}
&|\mathrm{LC}_{\text{fixed}}(T+\epsilon, S_0, K) - \mathrm{LC}_{\text{fixed}}(T, S_0, K)| \\
=&\left|\mathbb{E}\left[e^{-r(T+\epsilon)}(S_0e^{\overline{X}_{T+\epsilon}} - K)^+\right] - \mathbb{E}\left[e^{-rT}(S_0e^{\overline{X}_T} - K)^+\right]\right| \\
\leq&\left|\mathbb{E}\left[e^{-r(T+\epsilon)}S_0e^{\overline{X}_{T+\epsilon}}\right] - \mathbb{E}\left[e^{-rT}S_0e^{\overline{X}_T}\right]\right| \\
=&\left|\mathbb{E}\left[e^{-rT}S_0(e^{\overline{X}_{T+\epsilon}} - e^{\overline{X}_T})\right] + \mathbb{E}\left[S_0e^{\overline{X}_{T+\epsilon}}(e^{-r(T+\epsilon)} - e^{-rT})\right]\right| \\
=&\left|e^{-rT}S_0\Big(\mathbb{E}\left[e^{\overline{X}_T}(e^{\overline{X}_{T+\epsilon}-\overline{X}_T} - 1)\right] + (e^{-r\epsilon}-1)\,\mathbb{E}\left[e^{\overline{X}_{T+\epsilon}}\right]\Big)\right| \\
\leq&c_1\left|\mathbb{E}\left[e^{\overline{X}_T}(e^{\overline{X}_{T+\epsilon}-\overline{X}_T} - 1)\right]\right| + c_2(e^{-r\epsilon}-1) \\
\leq&c_1\left|\mathbb{E}\left[e^{\overline{X}_T}\right]\mathbb{E}\left[(e^{\overline{X}_\epsilon} - 1)\right]\right| + c_3\epsilon \\
\leq&c_4\left|\mathbb{E}\left[(e^{\overline{X}_\epsilon} - 1)\right]\right| + c_3\epsilon,
\end{aligned}$$

where the c_i's denote some constants and we used the independence of the increments of the Lèvy process X, the fact that $\mathbb{E}[e^{\overline{X}_T}] < \infty$, and the local Lipschitz continuity of the exponential function.
Furthermore for any $1 < \beta < \alpha_1$ we have

$$\begin{aligned}\left|\mathbb{E}\left[(e^{\overline{X}_\epsilon} - 1)\right]\right| &\leq \left(\mathbb{E}\left[(e^{\overline{X}_\epsilon} - 1)^\beta\right]\right)^{1/\beta} \\ &\leq \frac{\beta}{\beta - 1}\left(\mathbb{E}\left[(e^{X_\epsilon} - 1)^\beta\right]\right)^{1/\beta} \\ &= \frac{\beta}{\beta - 1}\left(e^{\epsilon\phi(-i\beta)} - 1\right) < c_5\epsilon\end{aligned}$$

where we applied Jensen's inequality, Doob's martingale inequality, and again the local Lipschitz continuity of the exponential function.
This completes the proof of the Lipschitz continuity of $\mathrm{LC}_{\text{fixed}}$. The second result in (7.18) follows by similar arguments. □

Corollary 7.3 *Let $0 < N \leq S_0 \leq M$, $\alpha > 0$ and let the assumptions of Theorem 7.4 be satisfied. Then the Laplace transforms of $\Theta_{\mathrm{LP}_{\textit{float}}}$ and $\Theta_{\mathrm{LC}_{\textit{float}}}$ are given by*

$$\widehat{\Theta}_{\mathrm{LP}_{\textit{float}}}(\alpha) = \alpha\widehat{\mathrm{LP}}_{\textit{float}}(\alpha) = S_0\left(\frac{\alpha}{\alpha + r}\sum_{k=1}^{m^+} A_k^+(\alpha + r)\frac{e^{-\log(M/S_0)(\rho_k^+(\alpha+r)-1)}}{\rho_k^+(\alpha + r) - 1} - 1\right) + \frac{\alpha M}{\alpha + r}, \tag{7.20}$$

$$\widehat{\Theta}_{\mathrm{LC}_{\textit{float}}}(\alpha) = \alpha\widehat{\mathrm{LC}}_{\textit{float}}(\alpha) = S_0\left(1 + \frac{\alpha}{\alpha + r}\sum_{k=1}^{m^-} A_k^-(\alpha + r)\frac{e^{-\log(N/S_0)(\rho_k^-(\alpha+r)-1)}}{1 - \rho_k^-(\alpha + r)}\right) - \frac{\alpha N}{\alpha + r}, \tag{7.21}$$

respectively.

Proof:
By (7.7) it follows that the price of a floating strike lookback put option is the sum of the price of a fixed strike lookback call option and an exponential function with respect to T. Thus by the proof of the previous theorem the price of a floating strike lookback put option is the sum of two Lipschitz continuous functions and therefore Lipschitz continuous, which proves (7.20). The second statement (7.21) follows analogously. □

7.4 Estimation of infinite activity processes via HEJD processes

Having seen that lookback options can be priced efficiently in HEJD-model markets the goal is now to apply these results to more general Lévy processes. While a direct generalisation is

due to the lack of explicit formulae for the Laplace transforms of the supremum and infimum processes typically not possible, for so-called generalised hyper-exponential processes, there is another possibility, which we will discuss now.

Definition 7.2 (Generalised hyper-exponential Lévy process) *A Lévy process is called generalised hyper - exponential Lèvy process (GHE), if its Lévy measure admits a density k of the form $k(x) = k_+(x)\mathbf{1}_{\{x>0\}} + k_-(-x)\mathbf{1}_{\{x<0\}}$, where k_+, k_- are completely monotone functions on $(0, \infty)$.*

Obviously the class of hyper-exponential jump diffusions is a subclass of the GHE processes, since their Lévy density can be written as

$$k_{HEJD}(x) = \lambda^+ \sum_{i=1}^{n^+} p_i^+ \alpha_i^+ e^{-\alpha_i^+ x}\mathbf{1}_{\{x>0\}} + \lambda^- \sum_{j=1}^{n^-} p_j^- \alpha_j^- e^{-\alpha_j^- x}\mathbf{1}_{\{x<0\}}. \tag{7.22}$$

Another well-known member of the GHE class is the NIG process which has the following representations of its Lévy densities:

$$k_{NIG}(x) = \frac{\delta\alpha}{\pi} e^{\beta x}\frac{K_1(\alpha x)}{x}\mathbf{1}_{\{x>0\}} + \frac{\delta\alpha}{\pi} e^{\beta x}\frac{K_1(-\alpha x)}{-x}\mathbf{1}_{\{x<0\}}, \tag{7.23}$$

where $\alpha > |\beta| > 0, \delta > 0$ and K_1 is the McDonald function

$$K_1(x) = x\int_1^\infty e^{-vx}(v^2-1)^{1/2}dv.$$

Jeannin and Pistorius [59] show, that for every process X in GHE, a sequence of HEJD processes $(X^n)_{n\geq 0}$ can be constructed which converges weakly to X in the Skorokhod topology on the space of real-valued cadlag functions on $\mathbb{R}_+$. They also show that the sequence of maximum processes $(\overline{X}^n)_{n\geq 0}$ converges in distribution to the maximum process $\overline{X}$. The next theorem states that also the sequence of lookback option prices converges in distribution to the lookback option price under X.

Theorem 7.5 *Let X be a GHE process, which is not a compound Poisson process, let the price process be given as $S_t = S_0 e^{X_t}$ and let $\mathrm{LC}_{\textit{float}}(S_0, K, T)$ be the pricing function of a floating strike lookback call option. Let $(X^n)_{n\geq 0}$ be a sequence of HEJD processes, with $X^n \to X$ for $n \to \infty$. Then the sequence of floating strike lookback put option prices $\mathrm{LC}^n_{\textit{float}}$ under the approximated processes X^n converges to $\mathrm{LC}_{\textit{float}}$.*

Proof:
Following the proof of Theorem 7.2, by using equation (7.10) it is sufficient to show

$$\lim_{n\to\infty} \mathbb{E}[e^{-rT}(e^{\overline{X}^n_T} - e^z)\mathbf{1}_{\{\overline{X}^n_T \geq z\}}] = \mathbb{E}[e^{-rT}(e^{\overline{X}_T} - e^z)\mathbf{1}_{\{\overline{X}_T \geq z\}}],$$

which is, by the arguments used in the proof of Theorem 7.2, equivalent to

$$\lim_{n\to\infty}\int_z^\infty e^y\,\mathbb{P}[\overline{X}^n_T \geq y]dy = \int_z^\infty e^y\,\mathbb{P}[\overline{X}_T \geq y]dy.$$

By using inequality (7.9) from Lemma 7.2, we can dominate $e^y \, \mathbb{P}[\overline{X}_T^n \geq y]$ and apply the dominated convergence theorem, thus the proof is complete. □

Remark 7.1 *The convergence of prices of fixed strike lookback options and floating strike lookback put options follows by similar arguments.*

7.5 Numerical results

In this last section we give numerical values of prices and Greeks of lookback options, which result by applying the Gaver-Stehfest algorithm for numerical Laplace transform inversion (see e.g. [40]) to the formulae given in the Theorems 3.1 - 3.3 and Corollaries 3.1 - 3.3. These results are compared to corresponding values derived via Monte Carlo integration. The main advantage of our method is that computing the numerical Laplace inversion of prices and Greeks takes only 1 second per option while the Monte Carlo simulation values takes several minutes.
The numerical analysis is divided into three subsections: in the first subsection we analyse the numerical error of the Gaver-Stehfest algorithm, by comparing results from the presented Laplace inversion method with a Monte Carlo (MC) simulation of a HEJD process. In the second subsection we give prices of lookback options under a HEJD process which is fitted to a NIG process and compare them to a Monte Carlo simulation of the original NIG process. In the last section, we compare sensitivities resulting from our technique with the corresponding simulated values of a NIG process. All computations were done in Mathematica.
The problem of fitting a HEJD process to a NIG process is considered in the articles of Crosby, LeSaux and Mijatovic [24] and Jeannin and Pistorius [59]. In both papers, a HEJD process is fitted to a NIG process with parameters $\alpha = 8.858, \beta = -5.808, \delta = 0.176$. All methods use a mixture of seven exponentially distributed upward jump variables Z_i^+ and a mixture of seven exponentially distributed downward jump variables Z_i^- to model the jumps of the logarithmic price process. In [59] the parameters $\alpha_i^{\pm}, i = 1, \ldots, 7$ are fixed in the beginning and the remaining parameters $\lambda^{\pm}$, σ and $p_i^{\pm}, i = 1, \ldots, 7$ are derived by a least squares approximation. The parameter μ follows from no-arbitrage considerations. Crosby, LeSaux and Mijatovic present several fitting methods which use more complicated optimisation techniques. In the following numerical examples we use the parameter set corresponding to method c) in [24] because it showed the best performance. These parameters are given in Table 7.1.

Parameter set CLM

Parameter	Value
$\sigma; \lambda_+; \lambda_-$	0.04062; 3.09468; 4.55662
p^+	{0.07858, 0.15033, 0.20017, 0.22039, 0.20704, 0.14327, 0.00022}
p^-	{0.05004, 0.12865, 0.22579, 0.21569, 0.18166, 0.13097, 0.06717}
α^+	{70.53135, 64.58179, 54.96035, 43.32801, 31.69567, 22.07423, 16.12466}
α^-	{4.58662, 10.85414, 20.98976, 33.24374, 45.49773, 55.63335, 61.90087}

Table 7.1: Parameters of the calibrated HEJD process fitted to a NIG process with parameters $\alpha = 8.858, \beta = -5.808, \delta = 0.176$ (method by Crosby, LeSaux and Mijatovic [24]).

7.5.1 Error of the Gaver-Stehfest algorithm

As a benchmark for our analysis, we use an unbiased Monte Carlo simulation method for the HEJD, which applies to general jump diffusion processes. The number of simulated paths is 100.000 and the computation time for one price is about 5 minutes. The results for prices of fixed strike lookback call and put options are given in the Tables 7.2 and 7.3, respectively. The numerical errors in the approximation of sensitivities of fixed strike lookback put options are given in Table 7.4.

Prices of fixed strike lookback call options in the HEJD model

S_0	MC Price	95%-conf.int.	CLM
70	0.00116	(0.00048; 0.00185)	0.00088
75	0.00387	(0.00227; 0.00546)	0.00322
80	0.00953	(0.00722; 0.01185)	0.01071
85	0.03632	(0.03091; 0.04174)	0.03253
90	0.09638	(0.08842; 0.10435)	0.09083
95	0.24363	(0.23116; 0.25609)	0.23522
100	0.56749	(0.54821; 0.58677)	0.56626
105	1.25536	(1.22719; 1.28353)	1.26059
110	2.53137	(2.49160; 2.57114)	2.56846
115	4.78483	(4.73138; 4.83828)	4.74117
120	7.93318	(7.86722; 7.99914)	7.90993
122.5	9.83583	(9.76352; 9.90814)	9.86508
125	12.03170	(11.95470; 12.10870)	12.0539
127.5	14.44220	(14.36186; 14.52254)	14.4661
128	14.98910	(14.90728; 15.07092)	14.975
128.5	15.57960	(15.49714; 15.66206)	15.4928
129	16.00890	(15.92693; 16.09087)	16.0197
129.5	16.55780	(16.47508; 16.64052)	16.556
130	17.01520	(16.93203; 17.09837)	17.102

Table 7.2: Prices of fixed strike lookback call options with varying initial asset price S_0, strike price $K = 130$ and maturity $T = 1$.

Prices of fixed strike lookback put options in the HEJD model

S_0	MC Price	95%-conf.int.	CLM
70	6.52566	(6.47190; 6.57942)	6.53762
70.5	6.13835	(6.08455; 6.19215)	6.17411
71	5.85091	(5.79742; 5.90440)	5.88262
71.5	5.63913	(5.58548; 5.69278)	5.62039
72	5.39465	(5.34158; 5.44772)	5.37709
72.5	5.12923	(5.07689; 5.18157)	5.14933
75	4.16099	(4.11118; 4.21080)	4.19454
77.5	3.43517	(3.38830; 3.48204)	3.46899
80	2.92254	(2.87822; 2.96686)	2.90436
85	2.03532	(1.99668; 2.07396)	2.09797
90	1.57349	(1.53832; 1.60866)	1.56554
95	1.19229	(1.16122; 1.22336)	1.19899
100	0.95105	(0.92272; 0.97937)	0.93796
105	0.72813	(0.70269; 0.75356)	0.74684
110	0.56268	(0.53989; 0.58548)	0.60358
115	0.47751	(0.45603; 0.49899)	0.49407
120	0.39429	(0.37399; 0.41458)	0.40891
125	0.33014	(0.31157; 0.34870)	0.34171
130	0.26002	(0.24218; 0.27786)	0.28801

Table 7.3: Prices of fixed strike lookback put options with varying initial asset price S_0, strike price $K = 70$ and maturity $T = 1$.

The prices derived by our method are located in almost all cases in the 95%-confidence interval of the Monte Carlo estimator. Therefore, we conclude that numerical error resulting from the Gaver-Stehfest algorithm is very small, especially when the difference between the initial asset price and the strike price is not too large.
The Monte Carlo sensitivities in Table 7.4 are estimated by unbiased central finite difference estimators as described in Glasserman [42, Chapter 7]. To derive unbiased MC estimators for the sensitivities, we use the same set of random paths for each price computation, therefore a comparison of the MC prices with prices resulting from our method is omitted.

Greeks of fixed strike lookback put options in the HEJD model

S_0	MC Δ	MC Γ	CLM Δ	CLM Γ	Δ diff. in %	Γ diff. in %
70	-0.86616	0.43778	-0.87705	0.90464	1.26%	106.64%
72.5	-0.44270	0.05355	-0.44126	0.05496	-0.33%	2.64%
75	-0.32967	0.04057	-0.33031	0.03604	0.20%	-11.16%
77.5	-0.25327	0.02434	-0.25452	0.02544	0.49%	4.52%
80	-0.19952	0.02035	-0.20002	0.01862	0.25%	-8.50%
82.5	-0.15964	0.01346	-0.15965	0.01396	0.00%	3.68%
85	-0.12936	0.00842	-0.12911	0.01065	-0.19%	26.58%
87.5	-0.10584	0.00736	-0.10563	0.00826	-0.20%	12.15%
90	-0.08711	0.00741	-0.08731	0.00649	0.23%	-12.43%
92.5	-0.07259	0.00568	-0.07283	0.00516	0.33%	-9.23%
95	-0.06109	0.00363	-0.06126	0.00414	0.28%	14.25%
97.5	-0.05159	0.00346	-0.05192	0.00336	0.64%	-2.87%
100	-0.04420	0.00268	-0.04430	0.00275	0.23%	2.80%

Table 7.4: Prices of fixed strike lookback put options with varying initial asset price S_0, strike price $K = 70$ and maturity $T = 1$.

The relative differences in the last two columns are calculated using the following formulae:

$$\Delta\text{diff.} = \frac{CLM\ \Delta - MC\ \Delta}{MC\ \Delta}, \quad \Gamma\text{diff.} = \frac{CLM\ \Gamma - MC\ \Gamma}{MC\ \Gamma}. \tag{7.24}$$

The numerical error in the computation of the sensitivities is relatively small, although the values of the error of the second derivative vary quite a lot. Especially, when S_0 is close to the strike price the Monte Carlo estimator and the Laplace inversion values differ.

7.5.2 Error of the parameter fit

The next step is to compare prices derived by our numerical Laplace inversion method with a Monte Carlo simulation of the corresponding NIG process. The paths of the NIG process were simulated on an equidistant grid, which of course introduces a bias, but numerical experiments show that a simulation of 100.000 simulated paths with 1.000 grid points provides a reasonable accuracy.
The computation times are about one hour for the Monte Carlo method and 1 second for the computation of one price together with the corresponding sensitivities using the presented Laplace inversion method. Our first example (see Table 7.5) is a fixed strike lookback call option. In Table 7.6, prices of floating strike lookback put options are compared.

Prices of fixed strike lookback call options in the NIG model

S_0	MC Price	95%-conf.Int.	CLM
70	0.00116	(0.00051; 0.00182)	0.00088
75	0.00438	(0.00273; 0.00603)	0.00322
80	0.01202	(0.00931; 0.01474)	0.01071
85	0.03254	(0.02776; 0.03733)	0.03253
90	0.08844	(0.08092; 0.09595)	0.09083
95	0.21810	(0.20634; 0.22987)	0.23522
100	0.54534	(0.52699; 0.56369)	0.56626
105	1.23040	(1.20301; 1.25779)	1.26059
110	2.52913	(2.49046; 2.56780)	2.56846
115	4.60783	(4.55722; 4.65844)	4.74117
120	7.78941	(7.72576; 7.85306)	7.90993
122.5	9.64449	(9.57600; 9.71298)	9.86508
125	11.9014	(11.8273; 11.9754)	12.0539
127.5	14.3110	(14.2333; 14.3886)	14.4661
128	14.83000	(14.7516; 14.9083)	14.975
128.5	15.3338	(15.2547; 15.4128)	15.4928
129	15.8319	(15.7524; 15.9113)	16.01972
129.5	16.3457	(16.2658; 16.4255)	16.556
130	16.9719	(16.8910; 17.0527)	17.102

Table 7.5: Prices of fixed strike lookback call options with varying initial asset price S_0, strike price $K = 130$ and maturity $T = 1$.

Prices of floating strike lookback put options in the NIG model

S_0	MC Price	95%-conf.Int.	CLM
70	56.1569	(56.0765; 56.2372)	56.1588
75	51.1523	(51.0661; 51.2384)	51.1611
80	46.1821	(46.0902; 46.2739)	46.1686
85	41.1983	(41.1016; 41.2949)	41.1906
90	36.1890	(36.0873; 36.2906)	36.2488
95	31.3931	(31.2864; 31.4997)	31.3932
100	26.7064	(26.5971; 26.8156)	26.7242
105	22.3340	(22.2236; 22.4443)	22.4185
110	18.6218	(18.5126; 18.7309)	18.7265
115	15.8606	(15.7538; 15.9673)	15.8990
120	13.9640	(13.8601; 14.0678)	14.0679
122.5	13.3252	(13.2226; 13.4277)	13.5231
125	13.0204	(12.9178; 13.1229)	13.2119
127.5	12.9782	(12.8746; 13.0817)	13.1239
128	12.9709	(12.8670; 13.0747)	13.1330
128.5	12.8558	(12.7526; 12.9589)	13.1509
129	13.0367	(12.9317; 13.1416)	13.1778
129.5	12.9958	(12.8911; 13.1004)	13.2141
130	13.0514	(12.9466; 13.1561)	13.2599

Table 7.6: Prices of floating strike lookback put options with varying initial asset price S_0, initial maximum $M = 130$ and maturity $T = 1$.

The fitting procedure for HEJD processes is accurate and robust in the case of vanilla options, see [24]. Nevertheless, especially for values of S_0 near K and M, respectively, there is a remarkable difference between the corresponding prices. A possible improvement could be to consider a fitting method which concentrates more on the tail behavior of the distribution of the increments of the underlying process. See [9], for a fitting method which takes that into account in the case of fitting a HEJD to a CGMY process.

7.5.3 Overall error of the sensitivity estimators

The purpose of this subsection is to compare sensitivities of prices of fixed strike lookback options computed with a Monte Carlo method with our method, using the parameter set CLM. The last two columns in every of the following tables are calculated via (7.24).

Greeks of fixed strike lookback call options in the NIG model

S_0	MC Δ	MC Γ	CLM Δ	CLM Γ	Δ Diff. in %	Γ Diff. in %
70	0.02012	0.00471	0.0215133	0.005021	6.92%	6.61%
72.5	0.03534	0.00781	0.0379496	0.008369	7.38%	7.16%
75	0.06035	0.01230	0.0647412	0.013370	7.28%	8.71%
77.5	0.10287	0.01968	0.10642	0.020297	3.45%	3.14%
80	0.16421	0.02750	0.1677	0.028963	2.13%	5.32%
82.5	0.24754	0.03792	0.251891	0.038380	1.76%	1.21%
85	0.35310	0.04624	0.358668	0.046664	1.58%	0.92%
87.5	0.47680	0.05184	0.482484	0.051717	1.19%	-0.24%
90	0.60902	0.05424	0.613796	0.0527	0.78%	-2.84%
92.5	0.73850	0.05136	0.743544	0.050765	0.68%	-1.16%
95	0.86106	0.04752	0.866602	0.047605	0.64%	0.18%
97.5	0.97660	0.04320	0.982458	0.045664	0.60%	5.70%
100	1.09440	0.04800	1.10199	0.052875	0.69%	10.16%

Table 7.7: Prices of fixed strike lookback call options with varying initial asset price S_0, strike price $K = 100$ and maturity $T = 1$.

Greeks of fixed strike lookback put options in the NIG model

S_0	MC Δ	MC Γ	CLM Δ	CLM Γ	Δ Diff. in %	Γ Diff. in %
70	-0.87740	0.31746	-0.87704	0.90464	-0.04%	184.96%
72.5	-0.44139	0.05647	-0.44126	0.05496	-0.03%	-2.67%
75	-0.32881	0.03777	-0.33031	0.03604	0.46%	-4.58%
77.5	-0.25198	0.02701	-0.25452	0.02544	1.01%	-5.81%
80	-0.19713	0.01994	-0.20001	0.01861	1.46%	-6.63%
82.5	-0.15679	0.01551	-0.15964	0.01395	1.82%	-10.02%
85	-0.12624	0.01051	-0.12911	0.01065	2.28%	1.36%
87.5	-0.10313	0.00858	-0.10562	0.00825	2.42%	-3.75%
90	-0.08494	0.00425	-0.08730	0.00648	2.78%	52.64%
92.5	-0.07081	0.00436	-0.07282	0.00515	2.85%	18.28%
95	-0.05947	0.00340	-0.06125	0.00414	3.01%	21.87%
97.5	-0.05034	0.00291	-0.05191	0.00336	3.13%	15.54%
100	-0.04280	0.00293	-0.04430	0.00275	3.51%	-6.06%

Table 7.8: Prices of fixed strike lookback call options with varying initial asset price S_0, strike price $K = 70$ and maturity $T = 1$.

Note that the computation of the Greeks (Δ, Γ) means almost no additional computational effort as one can see for example by comparing the formulae in Theorem 3.1 and 3.2. As in Subsection 7.5.1, the error of the Γ values is relatively high, especially near $S_0 = K$.

7.6 Conclusion

In this paper, we present explicit formulae for the Laplace transforms of prices and sensitivities of lookback options in a hyper-exponential jump diffusion model. Since a wide class of exponential Lévy processes can be approximated arbitrarily close by HEJD processes, these results give the possibility to efficiently approximate prices of lookback options for a vast class of processes used in financial modelling. The effectiveness of the inversion of the Laplace transformed values was illustrated in several numerical examples.

Bibliography

[1] C. Aistleitner. Covering numbers, dyadic chaining and discrepancy. *Journal of Complexity*, 27(6):531–540, 2011.

[2] C. Aistleitner and M. Hofer. On the limit distribution of consecutive elements of the van der Corput sequence. *Uniform Distribution Theory: to appear*.

[3] C. Aistleitner and M. Hofer. Probabilistic error bounds for the discrepancy of mixed sequences. *Monte Carlo Methods and Applications: to appear*.

[4] C. Aistleitner and M. Hofer. Uniform Distribution of generalized Kakutani's sequences of partitions. *Annali di Matematica Pura ed Applicata: to appear*.

[5] C. Aistleitner, M. Hofer, and R. Tichy. A central limit theorem for Latin hypercube sampling with dependence and application to exotic basket option pricing. *International Journal of Theoretical and Applied Finance: to appear*.

[6] H. Albrecher, C. Constantinescu, G. Pirsic, G. Regensburger, and M. Rosenkranz. An Algebraic Operator Approach to the Analysis of Gerber-Shiu Functions. *Insurance: Mathematics and Economics*, 46(1):42–51, 2010.

[7] S. Asmussen and H. Albrecher. *Ruin Probabilities, 2nd eds.* World Scientific, Singapore, 2010.

[8] S. Asmussen, F. Avram, and M. Pistorius. Russian and American put options under exponential phase-type Lévy models. *Stochastic Processes and Applications*, 109:79–111, 2004.

[9] S. Asmussen, D. Madan, and M. Pistorius. Pricing Equity Default Swaps under an approximation to the CGMY Lévy Model. *Journal of Computational Finance*, 11:79–93, 2007.

[10] O. Barndorff-Nielsen. Processes of Normal Inverse Gaussian type. *Finance and Stochastics*, 2:41–68, 1997.

[11] J. Beck. A two-dimensional van Aardenne-Ehrenfest theorem in irregularities of distribution. *Compositio Mathematica*, 72(3):269–339, 1989.

[12] J. Bertoin. *Lévy processes*. Cambridge University Press, Cambridge, 1996.

[13] D. Bilyk, M. Lacey, and A.Vagharshakyan. On the small ball inequality in all dimensions. *Journal of Functional Analysis*, 254(9):2470–2502, 2008.

[14] F. Black and M.Scholes. The Pricing of Options and Corporate Liabilities. *Journal of Political Economy*, 81(3):637–654, 1973.

[15] M. Brennan and R. Durrett. Splitting intervals. *Annals of Probability*, 14(3):1024–1036, 1986.

[16] N. Cai. On the first passage time of a hyper exponential jump diffusion process. *Operation Research Letters*, 37(2):127–134, 2009.

[17] N. Cai, N. Chen, and X. Wan. Pricing double-barrier options under a flexible jump diffusion model. *Operation Research Letters*, 37(3):163–167, 2009.

[18] I. Carbone. Discrepancy of LS-sequences of partitions. *Annali di Matematica Pura ed Applicata*, 2011.

[19] I. Carbone and A. Volčič. Kakutani's splitting procedure in higher dimension. *Rend. Ist. Mathem. Univ. Trieste*, XXXIX:1–8, 2007.

[20] P. Carr, P. Geman, D. Madan, and M. Yor. The fine structure of asset returns: an empirical investigation. *Journal of Business*, 75:305–332, 2002.

[21] P. Carr and D. Madan. Option Valuation using the Fast Fourier Transform. *Journal of Computational Finance*, 2:61–73, 1998.

[22] L. Chaumont. On the law of the supremum of Lévy processes. *arXiv:1011.4151v1*, 2010.

[23] R. Cont and P. Tankov. *Financial Modelling With Jump Processes*. Chapman & Hall / CRC, London, 2000.

[24] J. Crosby, N. LeSaux, and A. Mijatovic. Approximating Lévy processes with a view to option pricing. *Interantional Journal of Theoretical and Applied Finance*, 13(1):63–91, 2009.

[25] F. Delbaen and W. Schachermayer. *The Mathematics of Arbitrage*. Springer Verlag, Berlin, 2006.

[26] J. Dick and F. Pillichshammer. *Digital nets and sequences. Discrepancy and quasi-Monte Carlo integration*. Cambridge University Press, 2010.

[27] B. Doerr, M. Gnewuch, P. Kritzer, and F. Pillichshammer. Component-by-component construction of low-discrepancy point sets of small size. *Monte Carlo Methods Applications*, 14(2):129–149, 2008.

[28] B. Doerr, M. Gnewuch, and M. Wahlström. Algorithmic construction of low-discrepancy point sets via dependent randomized rounding. *Journal of Complexity*, 26(5):490–507, 2010.

[29] M. Drmota and M. Infusino. On the discrepancy of some generalized Kakutani's sequences of partitions. *Uniform Distribution Theory: to appear*.

[30] M. Drmota and R. Tichy. *Sequences, Discrepancies and Applications*. Springer Verlag, New York, 1997.

[31] E. Eberlein, K. Glau, and A. Papapantoleon. Analyticity of the Wiener-Hopf factors and valuation of exotic options in Lévy models. *In: Advanced Mathematical Methods for Finance, G. Di Nunno, B. Øksendal (eds.), Springer*, 2011.

[32] E.Novak and H.Woźniakowski. *Tractability of multivariate problems. Vol. I: Linear information, volume 6 of* EMS Tracts in Mathematics. European Mathematical Society (EMS), Z 2008.

[33] E.Novak and H.Woźniakowski. *Tractability of multivariate problems. Vol. II: Linear information, volume 6 of* EMS Tracts in Mathematics. European Mathematical Society (EMS), Z 2010.

[34] P. Erdös and P. Turán. On a problem in the theory of uniform distribution I. *Indagationes Mathematicae*, 10:370–378, 1948.

[35] P. Erdös and P. Turán. On a problem in the theory of uniform distribution II. *Indagationes Mathematicae*, 10:406–413, 1948.

[36] J.-D. Fermanian, D. Radulovic, and M. Wegkamp. Weak convergence of empirical copula processes. *Journal of the Bernoulli Society*, 10(5):847–860, 2004.

[37] J. Fialová and O. Strauch. On two-dimensional sequences composed by one-dimensional uniformly distributed sequences. *Uniform Distribution Theory*, 6(1):101–125, 2011.

[38] N. Friedman. *Introduction to Ergodic Theory*. Van Nostrand Reinhold, New York, 1970.

[39] N. Friedman. Replication and Stacking in Ergodic Theory. *The American Mathematical Monthly*, 99(1):31–41, 1992.

[40] D. P. Gaver. Observing stochastic processes and approximate transform inversion. *Operations Research*, 14:444–489, 1966.

[41] P. Giannopoulos, C. Knauer, M. Wahlström, and D. Werner. Hardness of discrepancy computation and -net verification in high dimension. *Journal of Complexity*, 28(2):162–176, 2012.

[42] P. Glasserman. *Monte Carlo Methods in Financial Engineering*. Springer Verlag, New York, 2004.

[43] M. Gnewuch. Bracketing numbers for axis-parallel boxes and applications to geometric discrepancy. *Journal of Complexity*, 24(2):154–172, 2008.

[44] M. Gnewuch. Construction of minimal bracketing covers for rectangles. *Electronic Journal of Combinatorics*, 15(1):Research Paper 95, 20, 2008.

[45] M. Gnewuch. On probabilistic results for the discrepancy of a hybrid-Monte Carlo sequence. *Journal of Complexity*, 25(4):312–317, 2009.

[46] M. Gnewuch. Entropy, randomization, derandomization, and discrepancy. *In* Monte Carlo and quasi-Monte Carlo methods 2010, 2010.

[47] M. Gnewuch, A. Srivastav, and C. Winzen. Finding optimal volume subintervals with k-points and calculating the star discrepancy are NP-hard problems. *Journal of Complexity*, 25(2):115–127, 2009.

[48] P. Grabner, P. Hellekalek, and P. Liardet. The dynamical point of view of low-discrepancy sequences. *Uniform Distribution Theory*, 7(1):11–70, 2012.

[49] J. Halton. On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. *Numerische Mathematik*, 2:84–90, 1960.

[50] J. M. Harrison and D. M. Kreps. Martingales and arbitrage in multiperiod securities markets. *Journal of Economic Theory*, 20:381–408, 1979.

[51] S. Heinrich. Some open problems concerning the star-discrepancy. *Journal of Complexity*, 19(3):416 – 419, 2003.

[52] S. Heinrich, E. Novak, G. Wasilkowski, and H. Woźniakowski. The inverse of the star-discrepancy depends linearly on the dimension. *Acta Arithmetica*, 96(3):279 – 302, 2001.

[53] P. Hellekalek and H. Niederreiter. Constructions of uniformly distributed sequences unsing the b-adic method. *Uniform Distribution Theory*, 6(1):185–200, 2011.

[54] F. Hickernell. The mean square discrepancy of randomized nets. *ACM Trans. Model. Comput. Simul.*, 6:274 – 296, 1996.

[55] A. Hinrichs. Covering numbers, Vapnik-Červonenkis classes and bounds for the star-discrepancy. *Journal of Complexity*, 20(4):477 – 483, 2004.

[56] A. Hirsa and D.B.Madan. Pricing American options under variance gamma. *Journal of Computational Finance*, 7(2):63–80, 2004.

[57] E. Hlawka. Funktionen von beschränkter Variation in der Theorie der Gleichverteilung. *Annali di Mathematica Pura ed Applicata*, (54):325–333, 1961.

[58] J. Jacod and P. Protter. *Probability Essentials, 2nd edn.* Springer, Berlin, 2003.

[59] M. Jeannin and M. Pistorius. A transform approach to calculate prices and Greeks of barrier options driven by a class of Lévy processes. *Quantitative Finance*, 10(6), 2010.

[60] H. Joe. *Multivariate Models and Dependence Concepts*. Chapmann and Hall, London, 1997.

[61] S. Kakutani. A problem of equidistribution on the unit interval [0,1]. *Proceedings of Oberwolfach Conference on Measure Theory. Lecture Notes in Mathematics*, 541:369–375, 1976.

[62] G. Khodak. Connection Between Redundancy and Average Delay for Fixed-Length Coding. *All-Union Conference on Problems of Theoretical Cybernetics (in Russian)*, 12, 1969.

[63] J. Koksma. Een algemeene stelling uit de theorie der gelijkmatige verdeeling modulo 1. *Mathematica B*, 11:7 – 11, 1943.

[64] S. Kou. A Jump-Diffusion Model for Option Pricing. *Management Science*, 48:1086–1101, 2002.

[65] S. Kou, G. Petrella, and H. Wang. Pricing Path-Dependent Options with Jump Risk via Laplace Transforms. *The Kyoto Economic Review*, 74(9):1–23, 2005.

[66] L. Kuipers and H. Niederreiter. *Uniform Distribution of Sequences*. John Wiley & Sons, New York, 1974.

[67] A. Kyprianou. *Introductory Lectures on Fluctuations of Lévy Processes with Applications*. Springer, Berlin, 2006.

[68] J. Lambert. *Some developments in Optimal and Quasi-Monte Carlo Quadrature, and a New Outlook on a Classical Chebyshev Problem*. Ph.D. Dissertation, The Claremont Graduate School, 1982.

[69] J. Lambert. Quasi-Monte Carlo, low discrepancy sequences, and ergodic transformations. *Journal of Computational and Applied Mathematics*, 12:419 – 423, 1985.

[70] C. Lemieux and P. L'Ecuyer. Variance reduction via lattice rules. *Management Science*, 46:1214–1235, 2000.

[71] A. Leonov. On the total variation for functions of several variables and a multidimensional analog of Helly's selection principle. *Mathematical Notes*, 63(1):61–71, 1998.

[72] W. LeVeque. An inequality connected with Weyl's criterion for uniform distribution. *Proceedings of Symposia in Pure Mathematics*, 8:22–30, 1963.

[73] W.-L. Loh. On Latin hypercube sampling. *The annals of statistics*, 24(5):2058–2080, 1996.

[74] D. Madan, P. Carr, and E. Chang. The variance gamma process and option pricing. *European Finance Review*, 1:39–55, 1998.

[75] D. Madan and E. Seneta. The variance gamma (v.g.) model for share market returns. *The Journal of Business*, 63:511–524, 1990.

[76] J. Matoušek. On the L_2-discrepancy for anchored boxes. *Journal of Complexity*, 14(4):527–556, 1998.

[77] M. McKay, R. Beckman, and W. Conover. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21:239–245, 1979.

[78] R. Merton. The Theory of Rational Option Pricing. *Bell Journal of Economics and Management Science*, 4(1):141–183, 1973.

[79] E. Mordecki. The distribution of the maximum of a Lévy process with positive jumps of phase-type. *Theory of Stochastic Processes*, 8:309–316, 2002.

[80] R. Nelsen. *An Introduction to Copulas*. Springer, New York, 1999.

[81] H. Niederreiter. *Random Number Generation and Quasi-Monte Carlo Methods*. SIAM, Philadelphia, 1992.

[82] G. Ökten. A probabilistic result on the discrepancy of a hybrid-Monte Carlo sequence and applications. *Monte Carlo Methods Applications*, 2(4):255–270, 1996.

[83] G. Ökten. High dimensional simulation. *Mathematics and Computers in Simulation*, 55:215–222, 2001.

[84] G. Ökten. Generalized von Neumann-Kakutani transformation and random-start scrambled Halton sequences. *Journal of Complexity*, 25:318–331, 2009.

[85] G. Ökten, B. Tuffin, and V. Burago. A central limit theorem and improved error bounds for a hybrid-Monte Carlo sequence with applications in computational finance. *Journal of Complexity*, 22(4):435–458, 2006.

[86] A. Owen. A central limit theorem for Latin hypercube sampling. *Journal of the Royal Statistical Society Series B 54*, 13:541–551, 1992.

[87] A. Owen. Multidimensional variation for quasi-Monte Carlo. *In J. Fan and G. Li (Eds.), International Conference on Statistics in honour of Professor Kai-Tai Fangs 65th birthday*, 2005.

[88] N. Packham and W. Schmidt. Latin hypercube sampling with dependence and application in finance. *Journal of Computational Finance*, 13(3):81–111, 2010.

[89] G. Pagés. Van der Corput sequences, Kakutani transforms and one-dimensional numerical integration. *Journal of Computational and Applied Mathematics*, 44:21 – 39, 1992.

[90] F. Pillichshammer and S. Steinerberger. Average distance between consecutive points of uniformly distributed sequences. *Uniform Distribution Theory*, 4(1):51 – 67, 2009.

[91] A. Roşca. A mixed Monte Carlo and quasi-Monte Carlo sequence for multidimensional integral estimation. *Acta Univ. Apulensis Math. Inform.*, 14:141 – 160, 2007.

[92] K. Roth. On irregularities of distribution. *Mathematika*, 1:73 – 79, 1954.

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

[94] W. Schmidt. Irregularities of distribution VII. *Acta Arithmetica*, 21:45 – 50, 1972.

[95] A. Sepp. Analytical Valuation of Lookback and Russian Options in a Double-Exponential Jump-Diffusion Model. *Working paper*, 2005.

[96] J. Spanier. Quasi-Monte Carlo methods for particle transport problems. *In* Monte Carlo and quasi-Monte Carlo methods in scientific computing (Las Vegas, NV, 1994), volume 106 of *Lecture Notes in Statistics*, pages 121 – 148, 1995.

[97] M. Stein. Large sample properties of simulations using Latin hypercube sampling. *Technometrics*, 29(2):143–151, 1987.

[98] W. Stute. The oscillation behaviour of empirical processes: the multivariate case. *The Annals of Probability*, 12:361–379, 1984.

[99] H. Tsukahara. Semiparametric estimation in copula models. *The Canadian Journal of Statistics*, 33:357–375, 2005.

[100] W. van Zwet. A proof of Kakutani's conjecture on random partitions. *Annals of Probability*, 6(1):133–137, 1978.

[101] A. Volčič. A generalization of Kakutani's splitting procedure. *Annali di Matematica Pura ed Applicata*, 190:45–54, 2011.

[102] X. Wang and F. Hickernell. Randomized Halton sequences. *Mathematical and Computer Modelling*, 32:887–889, 2000.

[103] H. Weyl. Über die Gleichverteilung von Zahlen mod. Eins. *Mathematische Annalen*, 77:313–352, 1916.

[104] S. Zaremba. Some applications of multidimensional integration by parts. *Ann. Polon. Math.*, 21:85–96, 1968.