

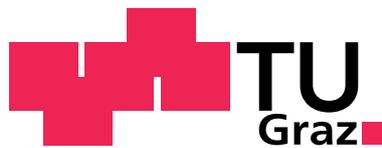
**Birgit SCHARWITZL**

**Varianzreduktion in der  
Finanzmathematik -  
Ein Quasi-Monte Carlo Ansatz**

**MASTERARBEIT**

**zur Erlangung des akademischen Grades einer Diplom-Ingenieurin**

**Masterstudium Finanz- und Versicherungsmathematik**



Graz University of Technology

**Technische Universität Graz**

**Betreuer:**

**O.Univ.-Prof. Dr.phil. Robert Tichy**

**Institut für Analysis und Computational Number Theory (Math A)**

**Graz, im Mai 2014**

## 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)

# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>1</b>
<b>2</b>	<b>Monte Carlo und Quasi-Monte Carlo Verfahren</b>	<b>2</b>
2.1	Monte Carlo Verfahren . . . . .	2
2.1.1	Linearer Kongruenzgenerator . . . . .	6
2.2	Quasi-Monte Carlo Verfahren . . . . .	6
2.2.1	Diskrepanz und Sterndiskrepanz einer Punktmenge . . . . .	8
2.2.2	Folgen kleiner Diskrepanz . . . . .	13
2.2.3	Hybride Verfahren . . . . .	25
2.2.4	Fluch der Dimension . . . . .	27
<b>3</b>	<b>Techniken zur Varianzreduktion</b>	<b>29</b>
3.1	Antithetic Variates . . . . .	29
3.2	Control Variates . . . . .	30
3.3	Importance Sampling . . . . .	31
3.4	Stratified Sampling . . . . .	32
<b>4</b>	<b>Latin Hypercube Sampling</b>	<b>37</b>
4.1	Varianzreduktion im Latin Hypercube Sampling . . . . .	38
4.2	Zentraler Grenzwertsatz für das Latin Hypercube Sampling . . . . .	41
<b>5</b>	<b>Latin Hypercube Sampling with Dependence</b>	<b>44</b>
5.1	Konsistenz des LHSD Schätzers . . . . .	46
5.2	Zentraler Grenzwertsatz für das LHSD und Varianzreduktion . . . . .	51
5.3	Varianzreduktion im LHSD Verfahren . . . . .	55
5.4	Anwendung auf Zufallsvektoren mit beliebiger Randverteilung . . . . .	56
<b>6</b>	<b>Latin Supercube Sampling</b>	<b>58</b>
6.1	Varianzreduktion im Latin Supercube Sampling . . . . .	58
6.2	Effektive Dimension . . . . .	62
<b>7</b>	<b>Anwendungen in der Finanzmathematik</b>	<b>64</b>
7.1	Grundlagen . . . . .	64
7.1.1	Optionen . . . . .	64
7.1.2	Modellannahmen . . . . .	65
7.1.3	Black-Scholes Modell . . . . .	69
7.1.4	Konstruktion von Pfaden einer Brown'schen Bewegung . . . . .	72
7.2	Collateralized Mortgage Obligation . . . . .	79
7.3	Asian Basket Option . . . . .	82
7.3.1	Numerische Ergebnisse . . . . .	85

# 1 Einleitung

Monte Carlo und Quasi-Monte Carlo Verfahren sind entwickelt worden, um komplexe Integrale, wie sie in der Finanzmathematik nicht selten auftreten, numerisch berechnen zu können. Ganz klar haben beide Methoden ihre Vorteile, Monte Carlo mit seiner von der Dimension unabhängigen Konvergenzrate und der bestechend einfachen Implementation und Quasi-Monte Carlo mit seiner präzisen Gleichverteilung im Raum.

Das allgemeine Ziel besteht immer darin, exaktere Ergebnisse zu bekommen und das, wenn möglich, auch noch schneller. Deshalb ist die Varianzreduktion ein starkes Anliegen von vielen Autoren und nicht nur in der Finanzmathematik, sondern auch in Bereichen wie Physik und Medizin sehr gefragt.

Die vier gängigsten Methoden, die vor allem in Verbindung mit dem Monte Carlo Verfahren oft zur Sprache gebracht werden, sind *Antithetic Variates*, *Control Variates*, *Importance Sampling* und *Stratified Sampling*. Diese werden nach einer Einführung in die Techniken des Monte Carlo und des Quasi-Monte Carlo Verfahrens in Kapitel 3 im Detail betrachtet.

Die Weiterentwicklung des *Stratified Sampling* auf mehrere Dimensionen, welche unter dem Namen *Latin Hypercube Sampling* bekannt ist und für das eine maximal gleich große Varianz wie im Monte Carlo Verfahren, sowie ein Zentraler Grenzwertsatz nachgewiesen werden kann, wird in Kapitel 4 erläutert.

Kapitel 5 beschäftigt sich mit einer Variante des *Latin Hypercube Samplings*, die sich von der Unabhängigkeit der Zufallsvariablen trennt und eine Copula bzw. ein Abhängigkeitsverhältnis innerhalb der Zufallsvektoren berücksichtigt. Sie wurde 2008 unter dem Namen *Latin Hypercube Sampling with Dependence* veröffentlicht. Auch für dieses Verfahren existiert ein Zentraler Grenzwertsatz und eine Varianzreduktion gegenüber Monte Carlo kann gezeigt werden.

Die Methode, die unter anderem im Fokus dieser Masterarbeit steht, nennt sich *Latin Supercube Sampling* und wurde 1998 zum ersten Mal präsentiert. Sie ist eine Mischung aus Quasi-Monte Carlo und der Idee des *Latin Hypercube Samplings* und kann durchaus mit den anderen Verfahren im Punkt Varianzreduktion konkurrieren, wie Kapitel 6 beweist.

Um die Performance der einzelnen Verfahren zu testen und miteinander zu vergleichen, werden in Kapitel 7 alle auf das Beispiel einer Asiatischen Basket Option angewandt, deren Payoff sich nicht durch eine geschlossene Formel darstellen lässt und damit für eine Anwendung numerischer Verfahren bestens geeignet ist.

Des Weiteren wird noch das bekannte Beispiel der *Collateralized Mortgage Obligation* gebracht, das die Theorie widerlegt, dass Quasi-Monte Carlo Folgen in Dimensionen höher als 30 keine Verbesserung gegenüber den Zufallsfolgen des Monte Carlo Verfahrens sein können. Warum dies dennoch der Fall ist und warum auch das *Latin Supercube Sampling* eine derart gute Leistung im Beispiel der Basket Option mit Dimension 2500 abliefert, liegt unter anderem an der sogenannten "effektiven" Dimension, die im Zusammenhang mit Quasi-Monte Carlo Folgen eine große Rolle spielt und gegen Ende der Masterarbeit immer wieder zur Sprache kommt.

## 2 Monte Carlo und Quasi-Monte Carlo Verfahren

Erste Ideen zur Monte Carlo Methode gab es schon 1930 von Enrico Fermi. Ausgeführt wurden diese von Stanislaw Ulam und John von Neumann 1946 im Zuge eines geheimen Projektes namens “Manhattan-Project” im Los Alamos Scientific Laboratory zur Entwicklung der Wasserstoffbombe. John von Neumann war auch der Namensgeber der Methode. “Monte Carlo” deshalb, da die Metropole Monte Carlo in Monaco bekannt für ihre vielen Casinos, Glücksspiele und Roulettes war. Roulette ist nichts anderes als ein Zufallsgenerator, der ein faires Spiel garantieren soll.

Eine ähnliche Intention wird bei Anwendung von Monte Carlo Verfahren verfolgt. Mithilfe von “fair”, im Sinne von gleichverteilten Zufallszahlen wird ein Mittelwert berechnet, der nach dem starken Gesetz der großen Zahlen einen konsistenten Schätzer für den Erwartungswert liefert. Quasi-Monte Carlo Verfahren stellen eine deterministische Variante der Monte Carlo Verfahren dar und verwenden statt Zufallszahlen deterministische Punktfolgen, was in vielen Anwendungen einen deutlichen Vorteil mit sich bringt.

Die Definitionen und Ergebnisse in diesem Kapitel folgen in Bezug auf Monte Carlo Verfahren den Erkenntnissen von Glasserman [18] und für Quasi-Monte Carlo Verfahren den Veröffentlichungen von Niederreiter [41].

### 2.1 Monte Carlo Verfahren

Ziel ist die Berechnung des Erwartungswertes  $\mathbb{E}[f(\mathbf{X})]$  der integrierbaren Zufallsvariable  $f(\mathbf{X})$  für eine Funktion  $f : \mathbb{R}^d \rightarrow \mathbb{R}$ . Dabei sei  $\mathbf{X} = (X_1, \dots, X_d)$  ein Zufallsvektor mit  $X_i \stackrel{iid}{\sim} F$  für  $i = 1, \dots, d$ . Dann gilt nach Definition des Erwartungswertes

$$I = \mathbb{E}[f(\mathbf{X})] = \int_{[0,1]^d} f(x_1, \dots, x_d) dF(x_1, \dots, x_d). \quad (2.1)$$

Die Schreibweise der Vektoren in Form von fettgedruckten Buchstaben zieht sich durch die gesamte Masterarbeit um für etwas mehr Übersicht zu sorgen. Der Ausdruck  $X_i \stackrel{iid}{\sim} F$  bedeutet, dass die Zufallsvariablen  $X_i$  identisch und unabhängig voneinander verteilt sind mit Verteilungsfunktion  $F$ .

Der Standard Monte Carlo Schätzer  $\hat{I}_N$  für das Integral  $I$  entsteht durch Auswertung der Funktion  $f$  an  $N$  unabhängigen und identisch verteilten Zufallsvektoren  $\mathbf{X}^{(1)}, \dots, \mathbf{X}^{(N)}$  und Bildung des Mittelwerts

$$\hat{I}_N = \frac{1}{N} \sum_{i=1}^N f(\mathbf{X}^{(i)}). \quad (2.2)$$

Die  $d$ -dimensionale Verteilung  $F(x_1, \dots, x_d)$  in Gleichung (2.1) kann mithilfe von Copulas allgemein über den Satz von Sklar beschrieben werden. Einen guten Einblick in die Theorie der Copulas liefert z.B. Nelsen [38].

**Definition 2.1. (Copula)**

Sei  $C : [0, 1]^d \rightarrow [0, 1]$  eine Funktion mit den Eigenschaften

1.  $C(u_1, u_2, \dots, u_d)$  ist eine monoton steigende Funktion in jeder Variable  $u_i$ ,  $1 \leq i \leq d$ .
2.  $C(1, \dots, 1, u_k, 1, \dots, 1) = u_k$  für  $u_k \in [0, 1]$  und alle  $1 \leq k \leq d$ .
3. Die sogenannte Rechtecksungleichung gilt für alle Vektoren  $(a_1, \dots, a_d)$  und  $(b_1, \dots, b_d) \in [0, 1]^d$ , wobei  $a_k \leq b_k$  für  $k = 1, \dots, d$  und  $u_{j1} = a_j$  sowie  $u_{j2} = b_j$ ,  $j = 1, \dots, d$

$$\sum_{k_1=1}^2 \cdots \sum_{k_d=1}^2 (-1)^{k_1+k_2+\dots+k_d} C(u_{1k_1}, \dots, u_{dk_d}) \geq 0. \quad (2.3)$$

Dann ist  $C$  eine  $d$ -dimensionale Copula.

Kurz kann gesagt werden, dass eine  $d$ -dimensionale Copula eine Verteilungsfunktion auf  $[0, 1]^d$  ist, deren Randverteilungen gleichverteilt sind auf  $[0, 1]$ . Das wichtigste Resultat in Hinblick auf den Zusammenhang zwischen Copulas und Gesamtverteilungsfunktionen ist der Satz von Sklar [58].

**Satz 2.2. (Satz von Sklar)**

Für eine Gesamtverteilungsfunktion  $F : \mathbb{R}^d \rightarrow [0, 1]$  mit eindimensionalen Randverteilungsfunktionen  $F_1, \dots, F_d$  existiert eine Copula  $C$ , sodass

$$F(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d)) \quad (2.4)$$

für alle  $x_1, \dots, x_d \in [-\infty, \infty]$  gilt.

Sind die Randverteilungsfunktionen  $F_1, \dots, F_d$  stetig, so ist die Copula  $C$  eindeutig.

Umgekehrt gilt für eine Copula  $C$  und Verteilungsfunktionen  $F_1, \dots, F_d$ , dass die Funktion  $F$  in (2.4) eine Gesamtverteilungsfunktion mit Randverteilungsfunktionen  $F_1, \dots, F_d$  darstellt.

Mithilfe der verallgemeinerten Inversen kann im Falle von stetigen Randverteilungsfunktionen auch noch eine andere Darstellung der Copula gegeben werden.

**Definition 2.3. (Verallgemeinerte Inverse)**

Ist  $F : \mathbb{R} \rightarrow \mathbb{R}$  eine monoton steigende Funktion, so heißt  $F^{\leftarrow} : \mathbb{R} \rightarrow \mathbb{R}$ , mit der Abbildung  $y \mapsto \inf \{x \in \mathbb{R} : F(x) \geq y\}$  verallgemeinerte inverse Funktion von  $F$ .

Ist  $F$  streng monoton steigend, so gilt  $F^{-1} = F^{\leftarrow}$ .

**Korollar 2.4.**

Für eine Gesamtverteilungsfunktion  $F$  mit stetigen Randverteilungsfunktionen  $F_1, \dots, F_d$  ist die eindeutige Copula  $C$  gegeben durch

$$C(u_1, \dots, u_d) = F(F_1^{\leftarrow}(u_1), \dots, F_d^{\leftarrow}(u_d)), \quad (2.5)$$

für  $u_1, \dots, u_d \in [0, 1]$ .

Mithilfe der Inversionsmethode kann die Erzeugung der Zufallsvektoren  $\mathbf{X}^{(i)}$  mit Verteilung  $F$  auf die Generierung von unabhängigen und im Einheitsintervall gleichverteilten Vektoren zurückgeführt werden.

**Satz 2.5. (Inversionsmethode)**

Sei  $U$  eine gleichverteilte Zufallsvariable ( $U \sim U[0, 1]$ ) und  $F^{\leftarrow}$  die verallgemeinerte inverse Verteilungsfunktion zu  $F$ . Sei  $X = F^{\leftarrow}(U)$  eine Zufallsvariable. Dann besitzt  $X$  die Verteilung  $F$ , da

$$\mathbb{P}(X \leq x) = \mathbb{P}(F^{\leftarrow}(U) \leq x) = \mathbb{P}(U \leq F(x)) = F(x).$$

Durch die Inversionsmethode kann das Problem in (2.1) vereinfacht werden zu

$$I = \mathbb{E}[g(\mathbf{U})] = \int_{[0,1]^d} g(\mathbf{u}) \, d\mathbf{u} = \int_{[0,1]^d} f(F^{\leftarrow}(\mathbf{u})) \, d\mathbf{u} \quad (2.6)$$

wobei  $g(\mathbf{u}) = f(F^{\leftarrow}(\mathbf{u}))$  und  $\mathbf{U} \sim U[0, 1]^d$ .

Der Standard Monte Carlo Schätzer ist dann gegeben durch

$$\hat{I}_N = \frac{1}{N} \sum_{i=1}^N g(\mathbf{U}^{(i)}), \quad \mathbf{U}^{(i)} \stackrel{iid}{\sim} U[0, 1]^d. \quad (2.7)$$

Wie gut aber ist dieser Schätzer, was kann über den Fehler und die Konvergenzrate gesagt werden? Antwort können das Starke Gesetz der großen Zahlen und der Zentrale Grenzwertsatz, die im Folgenden ohne Beweis angeführt sind, geben.

**Satz 2.6. (Starkes Gesetz der großen Zahlen)**

Sei  $(X_n)_{n \in \mathbb{N}}$  eine Folge von iid Zufallsvariablen mit endlichem Erwartungswert  $\mathbb{E}[X_1] < \infty$ . Dann gilt

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{f.s.} \mathbb{E}[X_1]. \quad (2.8)$$

**Satz 2.7. (Zentraler Grenzwertsatz)**

Sei  $(X_n)_{n \in \mathbb{N}}$  eine Folge von iid Zufallsvariablen mit  $\mu := \mathbb{E}[X_1]$  und  $\sigma^2 := \text{Var}[X_1]$ . Der Mittelwert der Stichprobe sei gegeben durch

$$\bar{X}_n = \frac{X_1 + \dots + X_n}{n}.$$

Dann gilt

$$\lim_{n \rightarrow \infty} \mathbb{P} \left( \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \leq z \right) = \Phi(z),$$

d.h. die zentrierte bzw. standardisierte Zufallsvariable

$$Z_n = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$$

konvergiert für  $n \rightarrow \infty$  punktweise gegen die Verteilungsfunktion  $\Phi(z)$  der Standardnormalverteilung  $N(0, 1)$ .

Aus dem starken Gesetz der großen Zahlen folgt für eine über  $[0, 1]^d$  integrierbare Funktion  $g$ , dass  $\hat{I}_N$  ein konsistenter Schätzer für das Integral  $I$  ist, was bedeutet, dass  $\lim_{n \rightarrow \infty} \hat{I}_N = I$  mit Wahrscheinlichkeit 1.

Ist die Funktion  $g$  sogar quadratintegrierbar, so ist der Fehler  $\hat{I}_N - I$  der Monte Carlo Schätzung als Resultat des zentralen Grenzwertsatzes approximativ normalverteilt,

$$\frac{1}{\sqrt{N}} \sum_{i=1}^N \left( g(\mathbf{U}^{(i)}) - \mathbb{E}[g(\mathbf{U})] \right) \xrightarrow{p} N(0, \sigma_g^2),$$

wobei  $\sigma_g^2$  die Varianz von  $g(\mathbf{U})$  bezeichnet und  $\xrightarrow{p}$  Konvergenz in Wahrscheinlichkeit symbolisiert. Die Standardabweichung des Schätzers konvergiert demnach unabhängig von der Dimension  $d$  mit Geschwindigkeit  $N^{-1/2}$  gegen 0.

Die Konvergenzrate  $O(N^{-1/2})$  des Monte Carlo Schätzers, die als Resultat des Zentralen Grenzwertsatzes die bestmögliche Schranke darstellt, hat sowohl Vor- als auch Nachteile. Im eindimensionalen Fall gibt es sehr wohl Verfahren, z.B. die Trapezregel, die bessere Konvergenzraten vorweisen können. Der Vorteil des Monte Carlo Verfahrens liegt darin, dass die Konvergenzrate nicht von der Dimension  $d$  abhängt.

Eine Möglichkeit den Fehler des Monte Carlo Schätzers trotz konstanter Konvergenzrate zu senken, besteht darin die Konstante  $\sigma_g^2$  zu verringern. Dies gelingt durch sogenannte Varianzreduktionsverfahren, wie sie in Kapitel 3 beschrieben werden. Beispiele dafür sind *Antithetic Variates*, *Control Variates*, *Importance Sampling*, *Stratified Sampling*, sowie das darauf aufbauende *Latin Hypercube Sampling*, das gemeinsam mit dem *Latin Supercube Sampling* im Fokus dieser Arbeit liegt.

Die zentrale Aufgabe bei der Anwendung des Monte Carlo Verfahrens besteht in der Generierung von unabhängigen und gleichverteilten Zufallszahlen in  $[0, 1]^d$ . In der Realität werden anstatt zufälliger Stichprobenpunkte sogenannte Pseudozufallszahlen verwendet. Vorteil dieser

scheinbar zufälligen Folgen von Punkten ist zum einen die Einfachheit der Generierung und zum anderen kann bei Bedarf die Simulation jederzeit mit derselben Zahlenfolge immer wieder durchlaufen werden.

### 2.1.1 Linearer Kongruenzgenerator

Einer der bekanntesten und meist verwendeten Pseudozufallszahlengeneratoren ist der Lineare Kongruenzgenerator. Er wurde 1949 von Lehmer [32] eingeführt. Die Berechnung einer Folge  $(z_i)_{i \geq 0}$  erfolgt bezüglich der Vorschrift

$$z_{i+1} = az_i + b \pmod{m}.$$

Die Parameter  $a, b, m$  und  $z_0$ , die Multiplikationsfaktor, Inkrement, Modus und Anfangsglied festlegen, können individuell aus den natürlichen Zahlen gewählt werden. Die Glieder der Folge  $(u_i)_{i > 0}$  mit

$$u_{i+1} = \frac{z_i + 1}{m},$$

stellen schließlich die gewünschten Pseudozufallszahlen dar.

Je nach Wahl der Parameter variieren die Anzahl, sowie die Verteilung der Folgenglieder im Einheitsintervall. Die Wahl von  $m$  als Primzahl,  $a$  als Primitivwurzel von  $m$  und  $b = 0$  garantiert zum Beispiel eine volle Periode, d.h. es werden  $m - 1$  verschiedene Folgenglieder erzeugt, bevor sich die erzeugten Zahlen zu wiederholen beginnen. Im Allgemeinen reicht es einen Modus von  $2^{64}$  zu verwenden. Detaillierte Erläuterungen zur Wahl der Parameter des linearen Kongruenzgenerators können bei L'Ecuyer [31] nachgelesen werden.

Lineare Kongruenzgeneratoren können auch miteinander kombiniert werden, um längere Perioden und damit mehr unterschiedliche Pseudozufallszahlen zu erreichen. Eine Möglichkeit besteht z.B. in der Summation von mehreren Generatoren, wobei dabei verschiedene Methoden angewandt werden. Zwei davon werden von L'Ecuyer [30] und Wichmann and Hill [63] näher erläutert.

## 2.2 Quasi-Monte Carlo Verfahren

Die Quasi-Monte Carlo Methode oder Methode kleiner Diskrepanzen ist eine Alternative zur Monte Carlo Methode mit dem Unterschied, dass Quasi-Monte Carlo keine echte Zufälligkeit erreichen will, sondern mit deterministischen Punktfolgen arbeitet, deren Folgenglieder gleichmäßig innerhalb des gewünschten Bereiches verteilt sind. Dadurch wird die Bildung von Clustern und Löchern der Punkte innerhalb des gewünschten Bereiches, die bei Pseudozufallszahlengeneratoren häufig vorkommen, großteils vermieden.

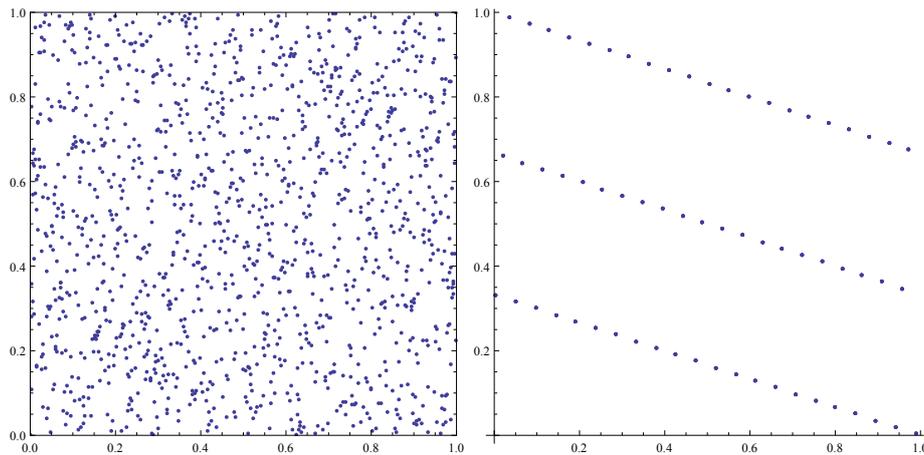


Abbildung 2.1: Links: Linearer Kongruenzgenerator in  $[0, 1]^2$  mit den Parametern  $a_1 = 1896, a_2 = 467, b = 0, m = 1301$ . Rechts: Linearer Zusammenhang der Punkte ( $a = 86, b = 0, m = 256$ ).

Das Ziel von Folgen mit kleiner Diskrepanz ist, durch Generierung von möglichst gleichverteilten Punkten  $x_i, i = 1, \dots, N$ , den Fehler der Schätzung für eine möglichst große Menge an Funktionen  $g$  möglichst klein zu halten. Bekannte Maße zur Bestimmung des Grades der Gleichverteilung von Punktfolgen sind die Diskrepanz  $D_N$  und die eng damit verbundene Stern-diskrepanz  $D_N^*$  einer Folge. Wann aber ist eine Folge überhaupt gleichverteilt?

Zum Nachweis der Gleichverteilung werden  $d$ -dimensionale Intervalle betrachtet und überprüft, wieviele Punkte sich innerhalb dieser Intervalle aufhalten. Das zentrale Ergebnis in der Theorie der Gleichverteilung stammt von Weyl [62] und wird in Satz 2.10 wiedergegeben.

**Definition 2.8. ( $d$ -dimensionales Intervall)**

Seien  $\mathbf{a} = (a_1, \dots, a_d)$  und  $\mathbf{b} = (b_1, \dots, b_d)$  Vektoren mit  $0 \leq a_i, b_i \leq 1$  für  $i = 1, \dots, d$ . Falls  $a_i \leq b_i$  für alle  $i = 1, \dots, d$ , so schreibe  $\mathbf{a} \leq \mathbf{b}$  und bezeichne die Menge aller Punkte  $\mathbf{x} \in [0, 1]^d$ , für die  $\mathbf{a} \leq \mathbf{x} \leq \mathbf{b}$  gilt, als  $d$ -dimensionales Intervall bzw. Hyperrechteck  $[\mathbf{a}, \mathbf{b}]$ .

**Definition 2.9. (Gleichverteilung von Punktfolgen)**

Eine Folge  $(\mathbf{x}_i)_{i \geq 1}$  von Punkten heißt gleichverteilt in  $[0, 1]^d$ , wenn für alle  $d$ -dimensionalen Hyperrechtecke  $[a, b] \subseteq [0, 1]^d$  gilt, dass

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \mathbb{1}_{[a,b]}(\mathbf{x}_i) = \lambda_d([a, b]),$$

wobei  $\lambda_d$  das  $d$ -dimensionale Lebesguemaß darstellt und  $\mathbb{1}_{[a,b]}$  die charakteristische Funktion des Intervalls  $[a, b]$  bezeichnet.

**Satz 2.10. (Satz von Weyl)**

Eine Punktfolge  $(\mathbf{x}_i)_{i \geq 1}$  in  $[0, 1]^d$  ist genau dann gleichverteilt, wenn

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N g(\mathbf{x}_i) = \int_{[0,1]^d} g(\mathbf{x}) \, d\mathbf{x} \quad (2.9)$$

für alle stetigen komplexwertigen Funktionen  $g : [0, 1]^d \rightarrow \mathbb{C}$ .

In Satz 2.10 wird gezeigt, wie mithilfe gleichverteilter Punkte numerisch integriert werden kann. Der Ausdruck auf der linken Seite von Gleichung (2.9) entspricht genau dem Quasi-Monte Carlo Schätzer

$$\bar{I}_N = \frac{1}{N} \sum_{i=1}^N g(\mathbf{x}_i) \quad (2.10)$$

des Integrals in (2.1), wobei  $\mathbf{x}_1, \dots, \mathbf{x}_N$  sorgfältig gewählte Punkte aus dem Einheitshyperwürfel  $[0, 1]^d$  darstellen.

Es gibt auch noch eine einfachere Version von Satz 2.10, die besagt, dass nur trigonometrische Polynome untersucht werden müssen. Auch diese Behauptung wurde von Weyl [62] aufgestellt und bewiesen.

**Satz 2.11. (Weyl'sches Kriterium)**

Eine Punktfolge  $(\mathbf{x}_i)_{i \geq 1}$  in  $[0, 1]^d$  ist genau dann gleichverteilt, wenn

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \exp(2\pi i \mathbf{h} \cdot \mathbf{x}_i) = 0 \quad (2.11)$$

für Vektoren  $\mathbf{h} \in \mathbb{Z}^d \setminus \{0\}$ . Mit “ $\cdot$ ” sei das übliche innere Produkt bezeichnet.

Als erstes Beispiel einer gleichverteilten Folge dient die Kronecker Folge.

Dabei sei  $\{x\} = x - [x]$  der Bruchteil von  $x \in \mathbb{R}$  und für  $\mathbf{x} = (x^1, \dots, x^d) \in \mathbb{R}^d$  sei  $\{\mathbf{x}\} = (\{x^1\}, \dots, \{x^d\}) \in [0, 1]^d$ . Die Folge  $(\{n\alpha\})_{n \geq 0}$  für  $\alpha = (\alpha_1, \dots, \alpha_d) \in \mathbb{R}^d$  wird Kronecker Folge genannt. Sie ist genau dann gleichverteilt, wenn  $1, \alpha_1, \dots, \alpha_d$  linear unabhängige Zahlen aus  $\mathbb{Q}$  sind. Dies folgt unmittelbar aus dem Weyl'schen Kriterium. In Kapitel 2.2.2 wird die Theorie der Gitterpunkte, zu denen die Kronecker Folge gehört, noch etwas näher erläutert.

**2.2.1 Diskrepanz und Sterndiskrepanz einer Punktmenge**

Aus dem Satz von Weyl erfahren wir also etwas über die numerische Integration, nicht ersichtlich aber ist der Integrationsfehler. Aufgrund dessen betrachten wir die Abweichung der Punktfolgen von einer Gleichverteilung unter Zuhilfenahme von Diskrepanz und Sterndiskrepanz, die als weit verbreitete Maße für die Gleichverteilung von Punktfolgen gelten. Anschließend Definitionen, Behauptungen und Beweise können größtenteils in Niederreiter [41] wiedergefunden

werden.

**Definition 2.12. (Diskrepanz und Sterndiskrepanz)**

Sei  $P = (\mathbf{x}_1, \dots, \mathbf{x}_N) \in [0, 1]^d$  eine deterministische Punktfolge,  $\mathbb{1}_A$  die charakteristische Funktion der Menge  $A$ ,  $\lambda_d$  das  $d$ -dimensionale Lebesguemaß. Dann ist die Diskrepanz der Folge  $(\mathbf{x}_1, \dots, \mathbf{x}_N)$  für die Menge  $\mathcal{A}$  aller achsenparallelen Hyperrechtecke definiert als

$$D_N(\mathbf{x}_1, \dots, \mathbf{x}_N) = \sup_{A \in \mathcal{A}} \left| \frac{1}{N} \sum_{i=1}^N \mathbb{1}_A(\mathbf{x}_i) - \lambda_d(A) \right|. \quad (2.12)$$

Die Sterndiskrepanz ist definiert für die Menge  $\mathcal{A}^*$  aller achsenparallelen Hyperrechtecken, wobei eine Ecke jedes solchen im Ursprung liegt, d.h. Hyperrechtecke der Form  $\prod_{i=1}^d [0, b_i]$ ,  $0 \leq b_i \leq 1$ ,

$$D_N^*(\mathbf{x}_1, \dots, \mathbf{x}_N) = \sup_{A \in \mathcal{A}^*} \left| \frac{1}{N} \sum_{i=1}^N \mathbb{1}_A(\mathbf{x}_i) - \lambda_d(A) \right|. \quad (2.13)$$

Als Folge des Kriteriums für Gleichverteilung und der Definition der Diskrepanz resultiert, dass eine Punktfolge  $(\mathbf{x}_n)_{n \geq 1}$  in  $[0, 1]^d$  genau dann gleichverteilt ist, wenn gilt

$$\lim_{N \rightarrow \infty} D_N(\mathbf{x}_1, \dots, \mathbf{x}_N) = 0.$$

**Satz 2.13.**

Sei  $P = (\mathbf{x}_1, \dots, \mathbf{x}_N) \in [0, 1]^d$  eine deterministische Punktfolge. Dann gilt

$$D_N^*(P) \leq D_N(P) \leq 2^d D_N^*(P). \quad (2.14)$$

Satz 2.13 gibt einen Zusammenhang zwischen Diskrepanz und Sterndiskrepanz an. Der Beweis davon kann in Niederreiter [41] gefunden werden.

Aus diesem Satz folgt des Weiteren sofort die Äquivalenz

$$\lim_{N \rightarrow \infty} D_N(P) = 0 \Leftrightarrow \lim_{N \rightarrow \infty} D_N^*(P) = 0.$$

Mithilfe der Sterndiskrepanz kann nun auch eine Abschätzung für den Integrationsfehler angegeben werden. Die nächsten Definitionen bezüglich der Variation von Funktionen gelten als Vorbereitung auf das zentrale Ergebnis der Koksma-Hlawka Ungleichung, mithilfe derer eine solche Abschätzung möglich ist.

**Definition 2.14. (Variation im Sinne von Vitali)**

Sei  $Q$  eine Partition von  $[0, 1]^d$ , d.h. eine Menge von  $d$  sortierten endlichen Folgen  $\zeta_0^{(j)}, \dots, \zeta_{k_j}^{(j)}, 1 \leq$

$j \leq d$  mit  $0 = \zeta_0^{(j)} \leq \dots \leq \zeta_{k_j}^{(j)} = 1$ .  $\mathcal{Q}$  sei die Menge, die alle Partitionen der Form  $Q$  enthält. Sei  $g$  eine Funktion auf  $[0, 1]^d$ , dann definiere für  $0 \leq i < k_j$

$$\begin{aligned} \Delta_j g(x^{(1)}, \dots, x^{(j-1)}, \zeta_i^{(j)}, x^{(j+1)}, \dots, x^{(d)}) &:= \\ g(x^{(1)}, \dots, x^{(j-1)}, \zeta_{i+1}^{(j)}, x^{(j+1)}, \dots, x^{(d)}) - g(x^{(1)}, \dots, x^{(j-1)}, \zeta_i^{(j)}, x^{(j+1)}, \dots, x^{(d)}) \end{aligned}$$

und  $\Delta_{j_1, \dots, j_k} := \Delta_{j_1} \cdots \Delta_{j_k}$ .

Dann ist die Variation von  $g$  nach Vitali definiert als

$$V^{(d)}(g) = \sup_{Q \in \mathcal{Q}} \sum_{i_1=0}^{k_1-1} \cdots \sum_{i_d=0}^{k_d-1} \left| \Delta_{1, \dots, d} g(\zeta_{i_1}^{(1)}, \dots, \zeta_{i_d}^{(d)}) \right|.$$

Eine Funktion  $g$  ist nach Vitali von beschränkter Variation, wenn  $V^{(d)}(g) < \infty$ .

Nach Niederreiter [41] kann die Variation von Vitali auch einfacher berechnet werden durch

$$V^{(d)}(g) = \int_{[0,1]} \cdots \int_{[0,1]} \left| \frac{\partial^d g}{\partial u_1 \cdots \partial u_d} \right| du_1 \cdots du_d,$$

sofern die partiellen Ableitungen der Funktion  $g$  stetig sind.

### Definition 2.15. Variation im Sinne von Hardy und Krause

Sei  $g$  eine Funktion auf  $[0, 1]^d$  und  $V^{(l)}(g; i_1, \dots, i_l)$  die Variation im Sinne von Vitali für  $1 \leq i_1 < i_2 < \cdots < i_l \leq d$  mit der Einschränkung auf die Menge

$$\{(u_1, \dots, u_d) \in [0, 1]^d : u_j = 1, j \neq i_1, \dots, i_l\}.$$

Dann ist die Variation von  $f$  nach Hardy und Krause definiert durch

$$V(g) = \sum_{l=1}^d \sum_{1 \leq i_1 < \cdots < i_k \leq d} V^{(l)}(g; i_1, \dots, i_k).$$

Eine Funktion  $g$  ist nach Hardy und Krause von beschränkter Variation, wenn  $V(g) < \infty$ .

Nun sind schließlich alle Vorbereitungen getroffen um die Koksma-Hlawka Ungleichung [20] zu verstehen.

### Satz 2.16. (Koksma-Hlawka Ungleichung)

Sei  $g$  eine Funktion von beschränkter Variation auf  $[0, 1]^d$  im Sinne von Hardy und Krause,  $(\mathbf{x}_1, \dots, \mathbf{x}_N)$  eine Punktfolge in  $[0, 1]^d$  so gilt

$$\left| \frac{1}{N} \sum_{i=1}^N g(\mathbf{x}_i) - \int_{[0,1]^d} g(\mathbf{x}) d\mathbf{x} \right| \leq V(g) D_N^*(\mathbf{x}_1, \dots, \mathbf{x}_N). \quad (2.15)$$

Die Schranke von Koksma-Hlawka besteht aus dem Produkt der Variation, die nur von der Funktion  $g$  abhängt und der Sterndiskrepanz, die sich wiederum nur nach der Punktfolge richtet. Nachdem die Funktion fix gegeben ist, ist die Punktfolge entscheidend.

Für Sterndiskrepanz und Diskrepanz können zunächst für den eindimensionalen Fall noch andere Darstellungen bzw. Berechnungsmöglichkeiten angegeben werden. Satz 2.17 wurde von Niederreiter [39] gezeigt, Satz 2.18 ist eine Vereinfachung einer Darstellung von de Clerk [12].

**Satz 2.17.**

Für eine Folge  $0 \leq x_1 \leq \dots \leq x_N \leq 1$  gilt

$$D_N^*(x_1, \dots, x_N) = \frac{1}{2N} + \max_{1 \leq n \leq N} \left| x_n - \frac{2n-1}{2N} \right| \quad (2.16)$$

Aus Satz 2.17 folgt unmittelbar als untere Schranke für die Sterndiskrepanz

$$D_N^*(x_1, \dots, x_N) \geq \frac{1}{2N}, \quad (2.17)$$

wobei offensichtlich Gleichheit herrscht für eine Folge  $x_n = \frac{2n-1}{2N}$ ,  $n = 1, \dots, N$ . Setzt man diese Folge in die Quasi-Monte Carlo Approximation ein, erhält man

$$\int_0^1 g(x) dx \approx \frac{1}{N} \sum_{n=1}^N g\left(\frac{2n-1}{2N}\right).$$

Diese Methode ist auch als Mittelpunktregel bekannt und wird oft zur numerischen Integration verwendet.

**Satz 2.18.**

Für  $0 \leq x_1 \leq \dots \leq x_N \leq 1$  gilt

$$D_N(x_1, \dots, x_N) = \frac{1}{N} + \max_{1 \leq n \leq N} \left( \frac{n}{N} - x_n \right) - \min_{1 \leq n \leq N} \left( \frac{n}{N} - x_n \right) \quad (2.18)$$

Wiederum folgt aus dem Satz eine untere Schranke, diesmal für die Diskrepanz,

$$D_N(x_1, \dots, x_N) \geq \frac{1}{N}, \quad (2.19)$$

mit Gleichheit für eine Folge  $x_n = \frac{n}{N}$ . Die beiden Resultate in (2.19) und (2.18) gelten nur dann, wenn die Anzahl  $N$  der Folgenglieder vorher fixiert wird.

Schmidt [55] hat für den eindimensionalen Fall ( $d = 1$ ) und eine beliebige Folge  $x_1, x_2, \dots$  aus  $[0, 1]$  gezeigt, dass eine Konstante  $c > 0$  existiert, sodass

$$D_N(x_1, \dots, x_N) \geq \frac{c \log N}{N}.$$

Damit gilt nach Satz 2.13 auch

$$D_N^*(x_1, \dots, x_N) \geq \frac{c' \log N}{N}, \quad (2.20)$$

was bedeutet, dass nichts besseres als  $D_N^*(x_1, \dots, x_N) = O(N^{-1} \log N)$  im eindimensionalen Fall zu erwarten ist. Der bisher beste Wert für  $c$  im Falle der Diskrepanz liegt bei  $c = 0.12$  und wurde von B ejian [7] ermittelt. Aufgrund von Satz 2.13 kann f ur  $c'$  im Falle der Sterndiskrepanz als bester Wert  $c' = 0.06$  angenommen werden.

Allgemein gilt f ur Dimension  $d$  der folgende Satz von Roth [54]:

**Satz 2.19. (Satz von Roth)**

*F ur jede  $N$ -elementige Punktmenge  $\mathbf{x}_1, \dots, \mathbf{x}_N$  aus dem Einheitsintervall  $[0, 1]^d$  gilt*

$$D_N(\mathbf{x}_1, \dots, \mathbf{x}_N) \geq D_N^*(\mathbf{x}_1, \dots, \mathbf{x}_N) \geq c_d \frac{(\log N)^{\frac{d-1}{2}}}{N},$$

wobei die Konstante  $c_d$  nur von der Dimension  $d$  abh angt.

Bilyk, Lacey und Vagharshakyan [8] haben 2008 eine leicht verbesserte untere Schranke f ur die Sterndiskrepanz angegeben, wobei allerdings die Vermutung besteht, dass es noch besser geht. Vielmehr soll die Absch atzung f ur eine Punktmenge mit fixer Gr o e  $N$

$$D_N^*(\mathbf{x}_1, \dots, \mathbf{x}_N) \geq c_d \frac{(\log N)^{d-1}}{N} \quad (2.21)$$

sein. Ob diese Absch atzung tats achlich in jeder Dimensionen f ur  $d \geq 3$  gilt, ist noch nicht bewiesen.

F ur  $d = 1$  stimmt die Ungleichung, wie bereits in (2.17) gezeigt wurde. F ur  $d = 2$  wurde die Aussage von Schmidt [55] bewiesen. Im Fall von  $d = 3$  gibt es bereits erste Fortschritte von Beck [6], der das Resultat

$$D_N^*(\mathbf{x}_1, \dots, \mathbf{x}_N) \geq c'_d \frac{(\log N)(\log \log N)^a}{N}, \quad (2.22)$$

f ur eine absolute Konstante  $a > 0$  bewiesen hat, was aber noch nicht die Schranke in (2.21) best atigt. F ur h ohere Dimensionen sind noch keine Beweise vorhanden.

Wäre Gleichung (2.21) bewiesen, wäre die Abschätzung für die ersten  $N$  Elemente einer beliebigen Folge  $\mathbf{x}_1, \mathbf{x}_2, \dots$ ,

$$D_N^*(\mathbf{x}_1, \dots, \mathbf{x}_N) \geq \hat{c}_d \frac{(\log N)^d}{N},$$

eine leichte Aufgabe, bisher konnte aber auch hier nur das Resultat

$$D_N^*(\mathbf{x}_1, \dots, \mathbf{x}_N) \geq \hat{c}'_d \frac{(\log N)^{\frac{d}{2}}}{N}$$

von Roth [54] gezeigt werden, wobei  $\hat{c}_d$  und  $\hat{c}'_d$  echt positive Konstanten sind, die nur von der Dimension  $d$  abhängen.

Es existieren auch obere Schranken für die Diskrepanz wie z.B. jene, die durch die Erdős-Turán-Koksma Ungleichung [14] gegeben ist.

**Satz 2.20. (Erdős-Turán-Koksma Ungleichung)** Sei  $\{\mathbf{x}_1, \dots, \mathbf{x}_N\}$  eine  $N$ -elementige Punktmenge in  $[0, 1)^d$ , so gilt für die Diskrepanz die Abschätzung

$$D_N(x_1, \dots, x_N) \leq \left(\frac{3}{2}\right)^2 \left( \frac{2}{m+1} + \sum_{0 < \|\mathbf{h}\|_\infty \leq m} \frac{1}{r(\mathbf{h})} \left| \frac{1}{N} \sum_{n=0}^{N-1} e^{2\pi i \mathbf{h} \cdot \mathbf{x}_n} \right| \right), \quad (2.23)$$

mit  $m \in \mathbb{N}$ ,  $\mathbf{h} = (h_1, \dots, h_d) \in \mathbb{Z}^d$  und  $r(\mathbf{h}) = \prod_{i=1}^d \max\{1, |h_i|\}$ .

Der Beweis ist zum Beispiel bei Drmota und Tichy [14] zu finden.

## 2.2.2 Folgen kleiner Diskrepanz

Mit dem Quasi-Monte Carlo Verfahren kann eine bessere asymptotische Konvergenzrate erzielt werden als im klassischen Monte Carlo Verfahren, da im besten Fall eine unendliche Folge, ‘‘Folge kleiner Diskrepanz’’ genannt, erzeugt werden kann, die eine Sterndiskrepanz und aufgrund der Koksma-Hlawka Ungleichung auch einen Approximationsfehler des Integrals von Ordnung  $O(N^{-1}(\log N)^d)$  besitzt.

### Van der Corput Folgen, Halton Folgen

Eine bekannte und vergleichsweise einfache Folge kleiner Diskrepanz ist die eindimensionale Van der Corput Folge. Zur Konstruktion einer solcher muss als erstes eine Basis  $b \geq 2$  gewählt, wobei  $b$  eine natürliche Zahl sein muss. Jeder beliebige Integer  $n$  kann dann eindeutig als lineare Kombination von Potenzen der Basis  $b$  dargestellt werden,

$$n = \sum_{j=0}^{\infty} a_j(n) b^j, \quad (2.24)$$

wobei die Koeffizienten  $a_j(n) \in \{0, \dots, b-1\}$  nur in endlich vielen Fällen ungleich 0 ist. Das  $n$ -te Folgenglied  $x_n$  der Van der Corput Folge ergibt sich aus der Spiegelung dieser Ziffernentwicklung am Dezimalpunkt,

$$\phi_b(n) = \sum_{j=0}^{\infty} \frac{a_j(n)}{b^{j+1}} \quad (2.25)$$

mit  $x_n = \phi_b(n)$  für alle  $n \geq 0$ .

Ein Nachteil der Folge ist, dass die Gleichverteilung vom Umfang der Stichprobe abhängig ist bzw. dass erst ab einer gewissen Anzahl an Stichprobenelementen der Raum mit Sicherheit gleichmäßig ausgefüllt ist. Je höher die Basis gewählt wird, desto mehr Stichprobenelemente sind notwendig. Es kann gezeigt werden, dass jede Van der Corput Folge eine Folge kleiner Diskrepanz ist, für die eine Ordnung von  $O(N^{-1} \log N)$  gilt.

Erweitert man die Methode der Van der Corput Folgen auf mehrere Dimensionen, gelangt man zu den Halton Folgen [19]. Hierfür werden mehrere relativ prime Basen  $b_1, \dots, b_d \geq 2$  benötigt. Üblicherweise werden dafür die ersten  $d$  Primzahlen gewählt, um zu vermeiden, dass Komponenten der einzelnen Folgenglieder übereinstimmen und dadurch Löcher im Integrationsbereich entstehen. Je höher die gewählten Basen sind bzw. je mehr Dimensionen benötigt werden, umso mehr Folgenglieder müssen generiert werden, damit der Raum vollständig gefüllt wird, was in Abbildung 2.2 ersichtlich ist.

Aus den Van der Corput Folgen  $\phi_{b_1}(n), \dots, \phi_{b_d}(n)$  bezüglich der gewählten Basen kann die Halton Folge  $\mathbf{x}_1, \dots, \mathbf{x}_N \in [0, 1]^d$  mit

$$\mathbf{x}_n = (\phi_{b_1}(n), \dots, \phi_{b_d}(n)),$$

für alle  $n \geq 0$  gebildet werden.

Um nun zu zeigen, dass die Halton Folge eine Folge kleiner Diskrepanz ist, muss vorher zuerst noch eine andere Behauptung verifiziert und ein paar grundlegende Notationen eingeführt werden, die sowohl im Lemma als auch im Satz danach zur Anwendung kommen und die Formulierung etwas vereinfachen (vgl. Niederreiter [41]).

- Für fixes  $N > 0$ ,  $J \in [0, 1]^d$  und  $x_1, \dots, x_N$  die ersten  $N$  Elemente der Halton Folge sei die Diskrepanz

$$D(J) = \sum_{n=1}^N \mathbb{1}_J(x_n) - N\lambda_d(J).$$

- Für  $i = 1, \dots, d$ ,  $e \in \mathbb{N}_0$ ,  $a \in \mathbb{Z}$ ,  $0 < a \leq b_i^e$  sei  $\mathcal{E}_i(e)$  eine Familie von Intervallen  $[0, ab_i^{-e})$ .

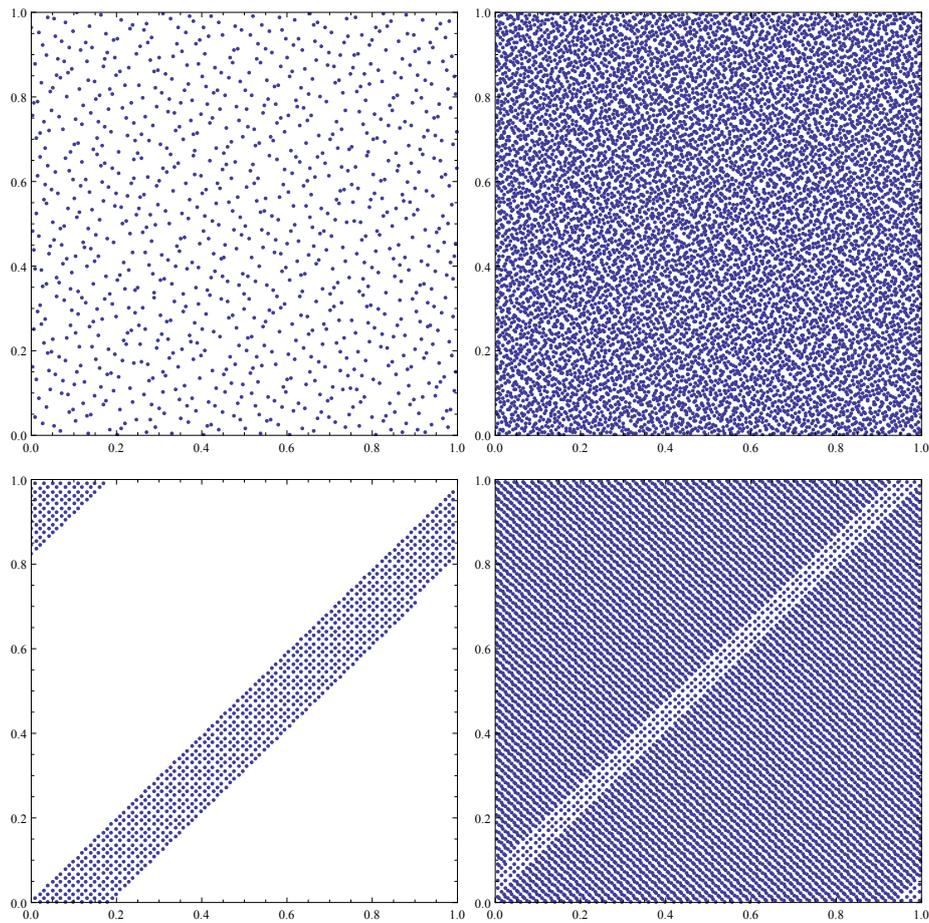


Abbildung 2.2: Oben: Halton Folge in Basen 2 und 3 bzw. Dimensionen 1 und 2. Unten: Halton Folge in Basen 101 und 103 bzw. Dimensionen 26 und 27. Links sind die ersten 1000 Punkte, rechts die ersten 10000 zu sehen.

- Für  $c, f \in \mathbb{Z}, 0 \leq f \leq e, 0 \leq c \leq b_i^f$  sei  $\mathcal{F}_i(e)$  die Familie von Intervallen der Form  $[cb_i^{-f}, (c+1)b_i^{-f})$ .
- $\mathcal{M}(e_1, \dots, e_d)$  bezeichne die Menge aller Intervalle der Form  $E = \prod_{i=1}^d E_i$  mit  $E_i \in \mathcal{E}_i(e_i) \cup \mathcal{F}_i(e_i)$ .

**Lemma 2.21.**

*Es gilt*

$$|D(E)| \leq \prod_{\substack{i=1 \\ E_i \notin \mathcal{F}_i(e_i)}}^d \left( \frac{b_i - 1}{2} e_i + 1 \right), \quad (2.26)$$

für alle  $E \in \mathcal{M}(e_1, \dots, e_d)$  mit relativ primen Basen  $b_1, \dots, b_d$ .

*Beweis.* Das Lemma wird mit Hilfe vollständiger Induktion nach der Anzahl  $k$  der Indizes  $i$ , für welche  $E_i \notin \mathcal{F}_i(e_i)$  gilt, bewiesen.

*Induktionsbasis:*  $k = 0$

Für die Basis gilt es die Aussage  $D(E) \leq 1$  zu beweisen.

Es gilt

$$E = \prod_{i=1}^d [c_i b_i^{-f_i}, (c_i + 1) b_i^{-f_i}]$$

mit  $c_i, f_i \in \mathbb{Z}, 0 \leq f_i \leq e_i, 0 \leq c_i \leq b_i^{f_i}$  für  $i = 1, \dots, d$ , da nur die Familie  $\mathcal{E}_i(e)$  etwas zum Produkt beiträgt.

Ein Element der Halton Folge  $\mathbf{x}_n$  liegt genau dann in  $E$ , wenn für  $i = 1, \dots, d$  die ersten  $f_i$  Koeffizienten  $a_j^i(n), j = 0, \dots, f_i - 1$  von  $\phi_{b_i}(n)$  in Basis  $b_i$  festgelegt sind. Dies ist äquivalent zur Tatsache, dass  $n$  in einer vorgegebenen Restklasse  $\text{mod } b_i^{f_i}, i = 1, \dots, d$  liegt. Aus dem Chinesischen Restsatz folgt aufgrund der relativ primen Basen wiederum, dass  $n$  zu einer vorgegebenen Restklasse  $\text{mod } m = \prod_{i=1}^d b_i^{f_i}$  gehört. Das bedeutet, dass nur ein einziges von  $m$  aufeinander folgenden Gliedern  $\mathbf{x}_n$  in  $E$  liegen kann.

Betrachten wir nun also Gleichung (2.26) mit  $N = mp + r$  und  $r < m$ . Es liegen  $p$  Elemente der Halton Folge von Länge  $N$  in  $E$  und das Lebesguemaß  $\lambda_d(E) = \frac{1}{m}$ . Damit folgt

$$|D(E)| = \left| p - \frac{mp + r}{m} \right| = \left| \frac{r}{m} \right| \quad (2.27)$$

und folglich  $|D(E)| \leq 1$  womit die Induktionsbasis gezeigt ist.

*Induktionsannahme:* Gleichung (2.26) gelte für  $k - 1, k \geq 1$ .

*Induktionsschritt:*  $k - 1 \rightarrow k$

OBdA können die Indizes derart geordnet werden, dass die ersten  $k$  Elemente im Produkt  $E = \prod_{i=1}^d E_i$  nicht in  $\mathcal{F}_i(e_i)$  liegen und für die restlichen Elemente  $E_i \in \mathcal{F}_i(e_i), i = k + 1, \dots, d$ . Unter dieser Annahme gilt  $E_1 = [0, ab_1^{-e_1}), 0 < a < b_1^{e_1}$ . Mithilfe der Ziffernentwicklung von Basis  $b_1$

$$ab_1^{-e_1} = \sum_{j=1}^{e_1} s_j b_1^{-j}$$

kann  $E_1$  in  $s_1$  disjunkte Intervalle  $\mathcal{F}_1(e_1)$  von Länge  $b_1^{-1}$ ,  $s_2$  Intervalle der Länge  $b_1^{-2}$  und so

weiter zerlegt werden mit  $\sum_{l=1}^{e_1} s_l = s$ , d.h. es gilt

$$E_1 = \bigcup_{r=1}^s F_r$$

mit paarweise disjunkten  $F_r \in \mathcal{F}_1(e_1)$ . Damit erhalten wir

$$E = \bigcup_{r=1}^s (F_r \times E_2 \times \cdots \times E_d).$$

Die Menge  $E$  ist nun von der Art, wie sie für die Induktionsannahme benötigt wird und  $D(E)$  kann abgeschätzt werden durch

$$|D(E)| \leq \left| \sum_{r=1}^s D(F_r \times E_2 \times \cdots \times E_d) \right| \leq s \prod_{i=2}^k \left( \frac{b_i - 1}{2} e_i + 1 \right). \quad (2.28)$$

Nun wird die Induktionsannahme noch einmal angewandt mit der Änderung, dass statt der disjunkten Zerlegung von  $E_1$  die Ergänzung  $G = [ab_1^{-e_1}, 1)$  von  $E_1$  auf  $[0, 1)$  betrachtet wird.  $E_1 = [0, 1) \setminus G$ . Also gilt

$$\begin{aligned} |D(E)| &\leq |D([0, 1) \times E_2 \times \cdots \times E_d)| + |D(G \times E_2 \times \cdots \times E_d)| \\ &\leq \prod_{i=2}^k \left( \frac{b_i - 1}{2} e_i + 1 \right) + |D(G \times E_2 \times \cdots \times E_d)|. \end{aligned}$$

$G$  kann auch als disjunkte Vereinigung von  $(b_1 - 1)e_1 - d + 1$  Intervallen aus  $\mathcal{F}_1(e_1)$  geschrieben werden, woraus folgt

$$|D(E)| \leq ((b_1 - 1)e_1 - d + 2) \prod_{i=2}^k \left( \frac{b_i - 1}{2} e_i + 1 \right).$$

Addition von (2.28) und (2.2.2) und anschließende Division durch 2 liefert die Abschätzung (2.26) für den Induktionsschritt.  $\square$

**Satz 2.22. (Sterndiskrepanz der Halton Folge)**

Für eine Halton Folge  $(\mathbf{x}_n)_{n \geq 0}$  mit relativ primen Basen  $b_1, \dots, b_d$  und  $N \geq 1$  gilt

$$D_N^*(\mathbf{x}_1, \dots, \mathbf{x}_N) \leq \frac{d}{N} + \frac{1}{N} \prod_{i=1}^d \left( \frac{b_i - 1}{2 \log b_i} \log N + \frac{b_i + 1}{2} \right). \quad (2.29)$$

*Beweis.* Betrachte ein allgemeines Intervall  $J = \prod_{i=1}^d [0, u_i) \in [0, 1)^d$  und definiere  $e_i$  als kleinsten Integer mit  $b_i^{e_i} \geq N$  sowie  $a_i$  als kleinste Zahl sodass  $a_i b_i^{-e_i} \geq u_i$ . Setze  $E = \prod_{i=1}^d [0, a_i b_i^{-e_i})$ . Damit folgt, dass, wenn  $\mathbf{x}_n \in J$  gilt,  $\mathbf{x}_n$  auch in  $E$  liegt, d.h.

$$\sum_{n=1}^N \mathbb{1}_J(\mathbf{x}_n) = \sum_{n=1}^N \mathbb{1}_E(\mathbf{x}_n).$$

Unter Verwendung von (2.26) aus Lemma 2.21 und der Tatsache  $N b_i^{-e_i} \leq 1$  folgt für die Diskrepanz

$$\begin{aligned} D(J) &= \left| \sum_{n=1}^N \mathbb{1}_J(\mathbf{x}_n) - N \lambda_d(J) \right| \leq N(\lambda_d(E) - \lambda_d(J)) + |D(E)| \\ &\leq |D(E)| + N \sum_{i=1}^d b_i^{-e_i} \leq |D(E)| + d \leq d + \prod_{i=1}^d \left( \frac{b_i - 1}{2} e_i + 1 \right). \end{aligned}$$

Nun werden noch die  $e_i, i = 1, \dots, d$  abgeschätzt. Es gilt  $b_i^{e_i} \geq N$ . Daraus folgt  $e_i \log b_i \geq \log N$  und schließlich  $e_i \leq \frac{\log N}{\log b_i} + 1$ .

Die Sterndiskrepanz  $D_N^*(x_1, \dots, x_N)$  ist gegeben als Supremum von  $\frac{|D(J)|}{N}$  über alle Intervalle  $J$ , woraus

$$\begin{aligned} D_N^*(\mathbf{x}_1, \dots, \mathbf{x}_N) &\leq \frac{d}{N} + \frac{1}{N} \prod_{i=1}^d \left( \frac{b_i - 1}{2} e_i + 1 \right) \leq \frac{d}{N} + \frac{1}{N} \prod_{i=1}^d \left( \frac{b_i - 1}{2} \left( \frac{\log N}{\log b_i} + 1 \right) + 1 \right) \\ &= \frac{d}{N} + \frac{1}{N} \prod_{i=1}^d \left( \frac{b_i - 1}{2 \log b_i} \log N + \frac{b_i + 1}{2} \right). \end{aligned}$$

resultiert. □

Um die Ordnung der Folge besser erkennen zu können, kann (2.29) umgeformt werden zu

$$D_N^*((\mathbf{x}_1, \dots, \mathbf{x}_N)) \leq \left( \prod_{i=1}^d \left( \frac{b_i - 1}{2 \log b_i} \right) \right) \frac{(\log N)^d}{N} + O\left( \frac{(\log N)^{d-1}}{N} \right). \quad (2.30)$$

### Hammersley Punktmengen

Die Idee von Hammersley besteht darin, die Anzahl  $N$  der Punkte schon vorab festzulegen. Dadurch ergibt sich die Hammersley Folge  $(\mathbf{h}_n)_{1 \leq n \leq N}$  mit Basen  $b_1, \dots, b_{d-1}$  als

$$\mathbf{h}_n = \left( \frac{n}{N}, \phi_{b_1}(n), \dots, \phi_{b_{d-1}}(n) \right).$$

Die Sterndiskrepanz kann ähnlich wie bei der Halton Folge abgeschätzt werden mit

$$D_N^*((\mathbf{h}_1, \dots, \mathbf{h}_N)) \leq \left( \prod_{i=1}^{d-1} \left( \frac{b_i - 1}{2 \log(b_i)} \right) \right) \frac{(\log N)^{d-1}}{N} + O\left( \frac{(\log N)^{d-2}}{N} \right). \quad (2.31)$$

Die Halton und Hammersley Punkte weisen also für fixes  $d$  und wachsendes  $N$  eine gute Gleichverteilung auf, was passiert aber mit dem konstanten Faktor  $C_d(b_1, \dots, b_d) = \prod_{i=1}^d \left( \frac{b_i - 1}{2 \log(b_i)} \right)$  in (2.30)? Nach einer Untersuchung von Niederreiter [41] gilt

$$\lim_{d \rightarrow \infty} \frac{\log C_d}{d \log d} = 1, \quad (2.32)$$

woraus aber resultiert, dass die Konstante  $C_d$  superexponentiell wächst für  $d \rightarrow \infty$ . Auch bei Hammersley steigt die führende Konstante superexponentiell mit wachsender Dimension. Dies schwächt die Qualität der beiden Folgen enorm, da die Schranken in hoher Dimension keine brauchbare Aussage mehr über die Folge liefern.

### $(t, m, d)$ -Netze und $(t, d)$ -Folgen

Eine Alternative zu Halton und Hammersley Folgen stellen die sogenannten netzartigen Folgen bzw.  $(t, d)$ -Folgen dar. Sie gehören auch in die Kategorie der Folgen kleiner Diskrepanz und kompensieren das Problem der sehr schnell wachsenden Konstante innerhalb der Schranke für die Sterndiskrepanz. Bevor jedoch auf die Eigenschaften näher eingegangen werden kann, muss zuerst geklärt werden, wodurch besagte Folgen überhaupt definiert sind.

#### **Definition 2.23. (Elementarintervall)**

Sei  $d \geq 1$  die Dimension und  $b \geq 2$  eine Basis.  $d_i$  und  $a_i$  seien ganze Zahlen mit  $s_i \geq 0$  und  $0 \leq a_i < b^{s_i}$  für  $i = 1, \dots, d$ . Dann wird das Intervall

$$E = \prod_{i=1}^d \left[ \frac{a_i}{b^{s_i}}, \frac{a_i + 1}{b^{s_i}} \right) \quad (2.33)$$

als  $d$ -dimensionales Elementarintervall in Basis  $b$  bezeichnet. Es gilt  $E \subseteq [0, 1]^d$ .

#### **Definition 2.24. $((t, m, d)$ -Netz)**

Seien  $t$  und  $m$  natürliche Zahlen mit  $0 \leq t \leq m$ . Dann ist ein  $(t, m, d)$ -Netz in Basis  $b$  definiert als eine endliche Punktmenge  $P$  von  $n = b^m$  Punkten  $\mathbf{x}_1, \dots, \mathbf{x}_n$  des Einheitsintervalls  $[0, 1]^d$ , wobei genau  $b^t$  Punkte in jedes Elementarintervall mit Volumen  $b^{t-m}$  fallen. Das bedeutet, dass der Anteil  $b^t/b^m$  der Punkte, die in einem solchen Intervall liegen genau dem Volumen des Intervalls entsprechen, d.h.

$$\sum_{\mathbf{x} \in P} \mathbb{1}_E(\mathbf{x}) = b^t.$$

**Definition 2.25. (( $t, d$ )-Folge)**

Eine Punktfolge  $(\mathbf{x}_1, \mathbf{x}_2, \dots) \in [0, 1]^d$  wird als  $(t, d)$ -Folge in Basis  $b$  bezeichnet, wenn die Menge

$$P = \{\mathbf{x}_N | kb^m < N < (k+1)b^m\}$$

für alle natürlichen Zahlen  $k$  und  $t < m$  ein  $(t, m, s)$ -Netz in Basis  $b$  darstellt.

Eine wichtige Eigenschaft dieser Netze und Folgen ist die Tatsache, dass für  $t \leq u \leq m$  ein  $(t, m, d)$ -Netz auch ein  $(u, m, d)$ -Netz ist und folglich jede  $(t, d)$ -Folge auch eine  $(u, d)$ -Folge für  $u \geq t$ .

Daher ist ersichtlich, dass kleinere Werte von  $t$  eine bessere Gleichverteilung bzw. überhaupt bessere Eigenschaften der  $(t, d)$ -Folgen hervorrufen. Auch für die Basis  $b$  sind kleine Werte von Vorteil angesichts dessen, dass Gleichverteilung nur dann gegeben ist, wenn  $b^m$  Punkte generiert werden. Je größer also die Basis, desto mehr Punkte müssen erzeugt werden, um die Eigenschaften von  $(t, m, d)$ -Netzen und  $(t, d)$ -Folgen zu erhalten.

**Satz 2.26. (Sterndiskrepanz einer  $(t, d)$ -Folge)**

Für eine  $(t, d)$ -Folge  $(\mathbf{x}_n)_{n \geq 1}$  in Basis  $b$  ist eine Schranke für die Sterndiskrepanz gegeben durch

$$D_N^*(\mathbf{x}_1, \dots, \mathbf{x}_N) \leq C_d(b) b^t \frac{(\log N)^d}{N} + O\left(b^t \frac{(\log N)^{d-1}}{N}\right) \quad (2.34)$$

mit

$$C_d(b) = \begin{cases} \frac{1}{d} \left( \frac{b-1}{2 \log b} \right)^d, & \text{für } d = 2 \text{ oder } b = 2 \text{ und } d = 3, 4 \\ \frac{1}{d!} \frac{b-1}{2 \lfloor b/2 \rfloor} \left( \frac{\lfloor b/2 \rfloor}{\log b} \right)^d, & \text{sonst} \end{cases}$$

Für die Konstante  $C_d$  in dieser Abschätzung gilt  $\lim_{d \rightarrow \infty} C_d(b) = 0$ , was im Vergleich zu den Abschätzungen der Sterndiskrepanzen in (2.30) und (2.31) ein deutlich besseres Ergebnis darstellt. Die verallgemeinerten Behauptungen zu den  $(t, d)$ -Folgen stammen von Niederreiter [41], weshalb sie oft auch mit Niederreiter-Folgen [40] bezeichnet werden. Bei Niederreiter kann des weiteren eine Anleitung zur Konstruktion der  $(t, m, d)$ -Netze und  $(t, d)$ -Folgen gefunden werden.

Ein bekanntes Beispiel für  $(t, d)$ -Folgen ist die bereits angeführte Van der Corput Folge, welche eine  $(0, 1)$ -Folge in Basis  $b$  ist.  $(0, d)$ -Folgen sind ursprünglich von Faure [15] eingeführt worden und zeichnen sich dadurch aus, dass sie den kleinsten bzw. bestmöglichen Wert von  $t$  annehmen, nämlich  $t = 0$ .

Zur Konstruktion einer Faure Folge wird zuerst als Basis die kleinste Primzahl  $b$  gewählt, für die  $b \geq d$  gilt. Anschließend wird wie bei der Halton Folge in Gleichung (2.24) die eindeutige

Zifferndarstellung für  $n$  mit Koeffizienten  $a_j(n) \in \{0, \dots, b-1\}$  ermittelt,

$$n = \sum_{j=0}^{\infty} a_j(n) b^j.$$

Die Faure Folge  $(\mathbf{x}_n)_{n \geq 0}$  mit  $\mathbf{x}_n = (x_n^{(1)}, \dots, x_n^{(d)})$  ist dann definiert durch

$$x_n^{(1)} = \sum_{j=0}^{\infty} a_j^{(1)}(n) b^{-j+1}$$

und für  $d \geq k \geq 1$

$$x_n^{(k)} = \sum_{j=0}^{\infty} y_j^{(k)}(n) b^{-j+1}$$

mit

$$y_j^{(k)}(n) = \sum_{i \geq j} \binom{i}{j} (k-1)^{i-j} a_i(n) \bmod b.$$

Bei Faure [15] ist auch eine Schranke für die Sterndiskrepanz gegeben. Für eine Faure Folge in Dimension  $d \geq 2$  und  $b$  als kleinste Primzahl mit  $b \geq d$  gilt

$$D_N^*(x_1, \dots, x_d) \leq \frac{1}{d!} \left( \frac{b-1}{2 \log b} \right)^d \frac{\log^d N}{N} + O \left( \frac{\log^{(d-1)} N}{N} \right) \quad (2.35)$$

Im Gegensatz zu Halton und Hammersley Punkten konvergiert die Konstante in (2.35) schnell gegen Null für  $d \rightarrow \infty$ , was einen deutlichen Vorteil in der Berechnung von höherdimensionalen Problemen liefert.

Von großer Bedeutung im finanzmathematischen Kontext sind außerdem  $(t, d)$ -Folgen in Basis  $b = 2$ . Diese werden zurückgehend auf die Definitionen von Sobol [59] auch Sobol-Folgen genannt und werden in Abbildung 2.3 veranschaulicht. Wie die Konstruktion und Implementation von derartigen Folgen im Detail funktioniert, kann unter anderem bei Sobol [59], Silva [57] sowie Antonov und Saleev [5] nachgelesen werden und sei im Folgenden nur skizziert.

Wie bereits erwähnt, verwendet eine  $d$ -dimensionale Sobol-Folge in jeder Dimension Basis  $b = 2$ . Die erste Dimension der Folge wird durch die Van der Corput Folge in Basis 2 beschrieben. Alle höheren Dimensionen stellen Permutationen dieser eindimensionalen Folge dar, die auf spezielle Art und Weise durchgeführt werden. Für jede Permutation wird eine Folge

$$v_i = \frac{m_i}{2^i}, \quad i = 1, \dots, w$$

von binären Bruchzahlen, die  $w$  Bits nach dem Komma haben, benötigt. Die Zahlen  $m_i$  mit  $0 < m_i < 2^i$  sind ungerade für  $i = 1, \dots, w$ . Erst wenn alle Glieder dieser Folge bestimmt sind,

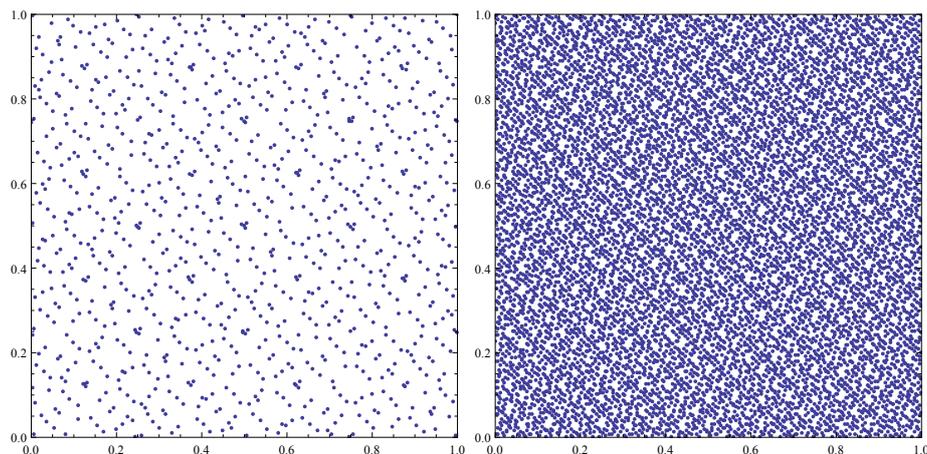


Abbildung 2.3: Zweidimensionale Sobol-Folge mit 1000 Punkten (links) und 10000 Punkten (rechts).

ist die Sobol-Folge eindeutig definiert.

Unter Verwendung der Folge  $v_i, i = 1, \dots, w$  wird eine eindimensionale Sobol-Folge  $(x_n)_{n \geq 0}$  generiert über

$$x_n = a_1 v_1 \oplus a_2 v_2 \oplus \dots \oplus a_w v_w \quad (2.36)$$

mit

$$n = \sum_{i=1}^{w=\lfloor \log n \rfloor + 1} a_i 2^{i-1}$$

als binäre Darstellung der natürlichen Zahl  $n$  mit  $a_i \in \{0, 1\}, i = 1, \dots, w$ . Das Symbol  $\oplus$  bezeichnet das Bit für Bit eingesetzte exklusive Oder, eine Operation, welche zwei Werte addiert und anschließend den Rest modulo 2 ausgibt.

Ein weiteres benötigtes Hilfsmittel ist der sogenannte “*Gray Code*”. Dieser ist für eine natürliche Zahl  $n$  definiert als

$$G(n) = n \oplus \left\lfloor \frac{n}{2} \right\rfloor.$$

Zum Beispiel ist für  $n = 2$   $G(2) = 2 \oplus 1 = (1 \cdot 2^1) \oplus 1 = 10 \oplus 1 = 11 \hat{=} 3$ .

Konkret erfolgt die Konstruktion der Sobol-Folge nach folgendem Schema. Zuerst wird zufällig eine natürliche Zahl  $n$  generiert. Diese Zahl gilt als Startpunkt der Folge. Nach Berechnung des *Grey Codes* von  $x$  und dessen Transformation in eine Binärzahl, werden die Hilfszahlen  $v_i$  für die Permutation entsprechend den Bits von  $G(x)$  über  $\oplus$  miteinander verknüpft. Ist also  $x = 2$  und  $G(x) = 3 \hat{=} 11$ , so werden die beiden ersten Hilfsnummern  $v_1$  und  $v_2$  mit  $\oplus$  aufsummiert. Der letzte Schritt besteht in der Rücktransformation der eben berechneten binären Zahl in eine

Dezimalzahl, die im Intervall  $(0, 1)$  angesiedelt ist. Dazu wird die Ziffer an Stelle  $i$  hinter dem Komma mit  $2^{-i}$  multipliziert und aufsummiert.

Antonov und Saleev [5] zeigen einen Algorithmus, der sich zur Berechnung der Hilfszahlen  $v_i, i = 1, \dots, w$  in (2.36) eines primitiven Polynoms bedient. Sei

$$P(x) = x^n + b_1x^{n-1} + \dots + b_{n-1}x + 1$$

ein primitives Polynom vom Grad  $n$  mit Koeffizienten aus  $\{0, 1\}$ . Dann können die Hilfszahlen  $v_i$  für  $i > n$  berechnet werden durch

$$v_i = b_1v_{i-1} \oplus b_2v_{i-2} \oplus \dots \oplus b_{n-1}v_{i-n+1} \oplus v_{i-n} \oplus \frac{v_{i-n}}{2^n}.$$

Der Vorteil dabei ist, dass in diesem Fall das  $k$ -te Glied der Sobol-Folge direkt aus dem  $k - 1$ -ten berechnet werden kann. Ein Nachteil des Algorithmus ist, dass jede Dimension mit einem anderen primitiven Polynom berechnet werden muss. Das bedeutet für eine Ermittlung einer 200-dimensionalen Sobol-Folge die Kenntnis von 200 verschiedenen primitiven Polynomen. Abhilfe schaffen kann eine von Jäckel [23] zur Verfügung gestellte CD mit über 8 Millionen solcher Polynome.

Alternativ kann auch eine Methode von Silva [57] praktiziert werden um die Hilfsnummern zu berechnen. Dabei werden für jede Dimension  $k$  solange gleichverteilte Zahlen  $u_{ik}$  generiert, sodass für  $0 < i < d_k$

$$m_{ik} = \max[\lfloor u_{ik} \cdot 2^{i-1} \rfloor, 1]$$

eine ungerade Zahl ist und anschließend

$$v_{ik} = \frac{m_{ik}}{2^i}, \quad 0 < i < d_k$$

als Hilfszahlen für Dimension  $k$  ermittelt.

Auch für die Sobol-Folge gilt die in Satz 2.26 gegebene Abschätzung der Sterndiskrepanz. Der Koeffizient  $C_d(2)$  entspricht in diesem Fall

$$C_d(2) = \frac{2^t}{d! \log^d 2},$$

wobei für  $t = t(d)$  die Schranke

$$K \frac{d \log d}{\log \log d} \leq t(d) \leq \frac{d \log d}{\log 2} + O(d \log \log d)$$

gilt (Sobol [59]). Diese Konstante ist zwar nicht annähernd so groß wie jene der Halton Folge, nichtsdestotrotz wächst sie Konstante superexponentiell mit zunehmender Dimension  $d$ .

### Good Lattice Points

Es gibt noch weitere Möglichkeiten, Punktfolgen zu erzeugen, die eine geringe Diskrepanz aufweisen, z.B. über die Methode der “guten” Gitterpunkte.

Anstatt unendlichen Folgen wird bei der Gittermethode eine vorher festgelegte Anzahl von Punkten erzeugt. Das kann als Nachteil gesehen werden, vor allem, wenn im Vorhinein nicht bekannt ist, wieviel Punkte für eine adäquate Genauigkeit der Methode benötigt werden.

Die Gitterpunkte  $x_i$  werden in Dimension  $d$  definiert als

$$x_i = \left\{ \frac{k}{N} \mathbf{v} \right\}, \quad k = 0, \dots, N - 1 \quad (2.37)$$

wobei  $\{x\} = x - \lfloor x \rfloor$  wieder den Bruchteil von  $x$  oder das Ergebnis von  $x \bmod 1$  darstellt und  $\mathbf{v}$  ganze Zahlen der Dimension  $d$  repräsentieren. Um wirklich  $N$  verschiedene Gitterpunkte zu erhalten muss gefordert werden, dass die Größen  $N$  und  $\mathbf{v}$  1 als größten gemeinsamen Teiler haben. Gitter von dieser Art werden auch als *Rang-1* Gitter bezeichnet.

Ein einfaches Beispiel von Rang-1 Gitter sind jene von Korobov, deren Erzeugungsvektor  $\mathbf{v}$  die Form  $(1, a, a^2, \dots, a^{d-1})$  besitzt, wobei  $a > 1$  als beliebige natürliche Zahl gewählt wird.

Werden die Anzahl  $N$  der Gitterpunkte, sowie der Erzeugervektor  $\mathbf{v}$  auf bestimmte Art und Weise gewählt, so können Folgen mit möglichst gleichverteilten Punkten generiert werden. Diese werden als “*Good Lattice Points*” bezeichnet und sind vor allem für die Integration periodischer Funktionen mit schnell fallenden Fourierkoeffizienten geeignet. Nähere Informationen zur Wahl von  $\mathbf{v}$  und  $N$  können in [18] gefunden werden.

Bei Niederreiter [41] ist in Kapitel 5 eine genaue Analyse und Herleitung einer Diskrepanzabschätzung von “*Good Lattice Points*” zu finden.

#### Satz 2.27. (Diskrepanz von Good Lattice Points)

Für eine Punktfolge  $P$  wie in (2.37) mit  $\mathbf{v} \in \mathbb{Z}^d$ ,  $d \geq 2$  und  $N \geq 2$  gilt

$$D_N(P) \leq \frac{d}{N} + \frac{1}{2} R(\mathbf{v}, N),$$

wobei

$$R(\mathbf{v}, N) = \sum_{\mathbf{h}} r(\mathbf{h})^{-1}$$

über alle  $\mathbf{h} \in C_d^*(N) = C_d(N) \setminus \{0\}$  für die  $\mathbf{h} \cdot \mathbf{g} = 0 \bmod N$ .  $C_d(N)$  ist das kartesische Produkt von  $d$  Kopien von  $C(N) = (-N/2, N/2] \cap \mathbb{Z}$ .

Die Funktion  $r$  ist definiert als

$$r(\mathbf{h}) = \prod_{i=1}^d \max[1, |h_i|] \quad (2.38)$$

### 2.2.3 Hybride Verfahren

Ein großer Nachteil von Quasi-Monte Carlo und ebenso großer Vorteil der Monte Carlo Methoden ist die Fehlerabschätzung. Während die Berechnung bzw. Schätzung der Hardy Krause Variation in der Koksma-Hlawka Ungleichung (2.16) sich als äußerst schwierig herausstellt, ist der Schätzer

$$\hat{\sigma}_g^2 = \frac{1}{N-1} \sum_{i=1}^N \left( g(U_i) - \hat{I}_N \right)^2$$

der Varianz  $\sigma_g^2$  der Monte Carlo Methode erwartungstreu und hat außerdem den Vorteil, dass Konfidenzintervalle für den Schätzer  $\hat{I}_N$  ohne Probleme angegeben werden können.

Unter anderem ist aus dieser Beobachtung die Idee entstanden die besten Eigenschaften von Monte Carlo und Quasi-Monte Carlo in einem Verfahren zu vereinen, was bedeutet, eine gewisse Zufälligkeit in die Punktfolgen zu bringen ohne aber deren gute Eigenschaften zu zerstören.

Dazu gibt es mehrere unterschiedliche Ansätze von denen ich nur die gängigsten kurz ansprechen möchte. Mehr Details können bei Glasserman [18] oder Owen [45] gefunden werden. Für alle im Folgenden aufgelistete Methoden sei die Punktmenge  $P_N$  beschränkt,

$$P_N = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$$

und die Punkte  $\mathbf{x}_1, \dots, \mathbf{x}_N$  seien Elemente aus dem Bereich  $[0, 1)^d$ .

### Cranley-Patterson Rotations

Die einfachste Methode, vorgestellt von Cranley und Patterson [11], ist es die Punkte aus  $P_N$  um eine zufällige gleichverteilte Größe zu verschieben. Konkret bedeutet dies, dass ein im  $d$ -dimensionalen Einheitshyperwürfel gleichverteilter Zufallsvektor  $\mathbf{U}$  erzeugt wird und anschließend jeder der Punkte um den Vektor  $\mathbf{U}$  modulo 1 geshiftet wird.

Daraus entsteht die Punktmenge

$$P_N(\mathbf{U}) = x_i + \mathbf{U} \bmod 1, \quad i = 1, \dots, n,$$

wobei sich die Anwendung des Operators “mod 1” auf jede Komponente bezieht. Der randomisierte Quasi-Monte Carlo Schätzer ist schließlich gegeben durch

$$\hat{I}_N(\mathbf{U}) = \frac{1}{N} \sum_{i=1}^N g(x_i + \mathbf{U} \bmod 1).$$

### Random Digital Scrambles

Eine weitere Möglichkeit Zufälligkeit in den Punktfolgen zu generieren, besteht in der Anwendung von Permutationen auf die Zähler in der Ziffernentwicklung. Hierbei wird aus einer deterministischen Folge von Punkten  $\mathbf{x}_1, \dots, \mathbf{x}_N$  aus dem Bereich  $[0, 1)^d$  durch zufällige Permutation ihrer Koeffizienten im Zähler eine zufällige Folge von Punkten  $\mathbf{p}_1, \dots, \mathbf{p}_N$ .

Auf einen Punkt  $x \in [0, 1)$ , der die Darstellung

$$x = \sum_{i=1}^{\infty} x_{(i)} b^{-i}$$

besitzt, wird eine Permutation  $\pi$  derart angewandt, dass die Koeffizienten  $x_{\pi(i)} \in \{0, 1, \dots, b-1\}$  des Punktes  $p \in [0, 1)$  die Koeffizienten  $x_{(i)} \in \{0, 1, \dots, b-1\}$  in zufälliger Reihenfolge wiedergeben. Der Punkt  $p$  mit Darstellung

$$p = \sum_{i=1}^{\infty} x_{\pi(i)} b^{-i}$$

ist nun die zufällige Version des Punktes  $x$ .

Die Methode kann auf  $d$  Dimensionen erweitert werden, indem unabhängige Permutationen auf unterschiedliche Koordinaten des Vektors angewandt werden. Innerhalb derselben Koordinate wird immer die gleiche Permutation für alle Punkte verwendet.

Owen [45] zeigt, dass durch die Randomisierung die Eigenschaft der Gleichverteilung der Zufallsvektoren in  $[0, 1)^d$  erhalten bleibt. Zudem werden durch Anwendung von untereinander unabhängigen Permutationen in verschiedenen Durchläufen unterschiedliche Schätzer für das gewünschte Integral erzeugt, was die Bestimmung eines Konfidenzintervalls ermöglicht.

### Scrambled Nets

Die Idee der *Scrambled Nets* ist ähnlich zur Methode der *Random Digital Scrambles* mit dem Unterschied, dass die Permutation der Punkte in Koordinate  $j$  von den ersten  $j-1$  Zählern in

der Ziffernentwicklung abhängig ist. Als erstes wurde diese Überlegung von Owen [45] publiziert.

$(\mathbf{X}_i)_{i \geq 1}$  sei ein  $(t, m, d)$ -Netz oder eine  $(t, d)$ -Folge in Basis  $b$  mit  $\mathbf{X}_i = (X_i^1, \dots, X_i^d)$ . Die Komponenten eines Folgengliedes haben die Darstellung

$$X_i^j = \sum_{k=1}^{\infty} a_{ijk} b^{-k},$$

mit  $a_{ijk} \in \{0, 1, \dots, b-1\}$  für alle  $i, j, k$ . Wie im Abschnitt zuvor, kann nun eine zufällige Version  $(\mathbf{P}_i)_{i \geq 1}$  mit  $\mathbf{P}_i = (P_i^1, \dots, P_i^d)$  der Folge  $(\mathbf{X}_i)_{i \geq 1}$  erzeugt werden über Permutation der Koeffizienten  $a_{ijk}$ . Die Darstellung in Basis  $b$  der permutierten Folge ist dann

$$P_i^j = \sum_{k=1}^{\infty} p_{ijk} b^{-k},$$

wobei die Koeffizienten  $p_{ijk}$  definiert sind als

$$\begin{aligned} p_{ij1} &= \pi_j(a_{ij1}) \\ p_{ij2} &= \pi_{j a_{ij1}}(a_{ij2}) \\ &\vdots \\ p_{ijk} &= \pi_{j a_{ij1} a_{ij2} \dots a_{ijk-1}}(a_{ijk}). \end{aligned}$$

Die Permutation  $\pi_j$  tauscht den ersten Zähler in der Darstellung von  $A_i^j, \forall i, \pi_{j a_{ij1}}$  den zweiten und so weiter. Die Permutation des zweiten Zählers ist allerdings abhängig vom Wert des ersten. So hängt auch die Permutation des  $k$ -ten Zählers in der Darstellung von den ersten  $k-1$  Zählern ab.

Die neu gebildete, zufällige Folge  $(\mathbf{P}_i)_{i \geq 1}$  ist wieder gleichverteilt in  $[0, 1]^d$  und jedes randomisierte  $(t, m, d)$ -Netz ist wieder ein solches und jede randomisierte  $(t, d)$ -Folge ist auch wieder eine  $(t, d)$ -Folge.

#### 2.2.4 Fluch der Dimension

Asymptotisch gesehen hat das Quasi-Monte Carlo Verfahren mit einem Approximationsfehler von  $O(N^{-1}(\log N)^d)$  gegenüber dem Monte Carlo Verfahren mit einem Fehler von  $O(N^{-1/2})$  ein besseres Konvergenzverhalten, hängt aber offensichtlich von der Dimension  $d$  ab. Obwohl die Schranke des Quasi-Monte Carlo Verfahrens deterministisch ist, kann die Dimension durchaus zum Problem werden. Das dies nicht immer der Fall ist, zeigt das Beispiel einer *Collaterized*

*Mortgage Obligation*, aus der ein 360-dimensionales Integral resultiert. Details dazu sind in Kapitel 7 zu finden.

Wie groß muss der Stichprobenumfang sein um eine brauchbare Schätzung zu bekommen bzw. was passiert, wenn der Stichprobenumfang nicht groß genug ist? Aufschluss liefert die Abschätzung in Gleichung (2.21)

$$D_N^*(\mathbf{x}_1, \dots, \mathbf{x}_N) \geq c_d \frac{(\log N)^{d-1}}{N},$$

für ein fix gewähltes Stichprobenvolumen  $N$ .

Die Größe

$$\frac{(\log N)^{d-1}}{N}$$

ist für  $N \leq e^{d-1}$  eine monoton steigende Funktion. Erst danach beginnt sie gegen Null zu fallen. Das bedeutet, dass die Abschätzung der Sterndiskrepanz für  $N \leq e^{d-1}$  nicht brauchbar ist, was ein zunehmendes Problem darstellt, je höher die Dimension  $d$  gewählt wird.

Aus diesem Grund sind Techniken entwickelt worden, um dieser Problematik aus dem Weg zu gehen. Anstatt die zur Simulation benötigte  $N \times d$  Matrix vollständig mit Folgengliedern des Quasi-Monte Carlo Verfahrens auszufüllen, werden nur  $s$  Spalten mit solchen Punkten versehen und die restliche Matrix wird mithilfe von anderen Verfahren “aufgeblasen”. Dazu gibt es mehrere Möglichkeiten.

Die einfachste Methode die restlichen Stellen zu füllen, ist die Verwendung von unabhängigen und in  $[0, 1)$  gleichverteilten Zufallsvariablen. Diese sollten auch unabhängig von eventuell benötigten Permutationsverfahren bei randomisierten Quasi-Monte Carlo Verfahren sein. Eine ausführliche Betrachtung solcher Matrizen und die resultierende Diskrepanz kann bei Spanier [60] und Okten [42] gefunden werden. Die Schwierigkeit bei solch einer Anwendung liegt in der Frage, für welche der  $s$  Variablen Quasi-Monte Carlo verwendet werden soll. Eine Hilfe dabei kann eine Varianzanalyse darstellen, aber dennoch ist einiges an Wissen, Überlegung und Versuchen notwendig, um eine adäquate Auswahl zu treffen.

Sehr oft benötigen Anwendungen Pfade einer Brown’schen Bewegung. Um diese zu generieren, gibt es mehrere Möglichkeiten. Zu den gängigsten Verfahren zählen die Brown’sche Brückenkonstruktion und die *Principal Components Construction*, die beide Einfluss auf die “effektive Dimension” eines Problems haben. Die Definition und Erläuterungen dazu sind in Abschnitt 6.2 nachzulesen. Eine genaue Beschreibung der Verfahren und ihrer Anwendung können in Kapitel 7 gefunden werden.

### 3 Techniken zur Varianzreduktion

In Kapitel 2 wurde bereits erwähnt, wodurch eine Reduktion des Integrationsfehlers in der Monte Carlo Methode erreicht werden kann. Die einzige Option besteht darin, die Varianz  $\sigma_g^2$  der zu integrierbaren Größe  $Y := g(\mathbf{U})$  zu ändern. Zu diesem Zweck sind verschiedenste Verfahren bzw. Techniken entwickelt worden, von denen einige bekannte und oft verwendete im Anschluss beschrieben sind. Ergänzungen dazu können in Glasserman [18] und Korn et al. [29] gefunden werden.

Hauptaugenmerk dieser Arbeit liegt auf der Methode des *Stratified Samplings*, die als Grundlage für das *Latin Hypercube Sampling* und das damit eng verknüpfte *Latin Supercube Sampling* dient. Der Vollständigkeit und Übersicht wegen, werden aber auch noch weitere Varianzreduktionsverfahren betrachtet, schon allein aus dem Grund, dass unter anderem auch die Möglichkeit besteht, die einzelnen Methoden miteinander zu verbinden bzw. nacheinander anzuwenden und daraus eine noch kleine Varianz resultieren kann.

#### 3.1 Antithetic Variates

Das antithetische Verfahren wird als eines der einfachsten Verfahren gesehen. Es verwendet die Idee, dass gegensätzliche Paare sich gegenseitig ausgleichen. So kann ein sehr hoher, nicht unbedingt erwünschter Wert durch dessen sehr kleines bzw. antithetisches Gegenstück abgemildert bzw. aufgehoben werden und dadurch eine geringere Varianz erzielt werden. Es werden also zum einen Teil zufällig gewählte und zum anderen systematisch erzeugte Punkte verwendet.

Zur Berechnung des Erwartungswertes  $\mathbb{E}[Y]$  werden Paare von Zufallsvariablen  $(Y_1, \bar{Y}_1), \dots, (Y_n, \bar{Y}_n)$  benötigt, wobei  $\bar{Y}_i$  antithetisch zu  $Y_i$  ist, aber die gleiche Verteilung wie  $Y_i$  und damit  $Y$  besitzt. Als Schätzer für den gewünschten Erwartungswert resultiert

$$\hat{Y}_{AV} = \frac{1}{2n} \sum_{i=1}^n (Y_i + \bar{Y}_i). \quad (3.1)$$

Wäre die Summe auf der rechten Seite gespalten, wäre jede für sich ein herkömmlicher Monte Carlo Schätzer.

Das einfachste Beispiel für antithetische Zufallsvariablen sind eine gleichverteilte Zufallsvariable  $U$  im Intervall  $[0, 1]$  und deren Ergänzung  $1 - U$ , die als Folge der Gleichverteilung von  $U$  auch eine solche besitzt.  $U$  und  $1 - U$  sind antithetisch zueinander, denn wenn  $U$  groß ist, muss  $1 - U$  dementsprechend klein sein. Offensichtlich resultiert als Mittelwert über antithetische Zufallsvariablen mit Gleichverteilung in  $[0, 1]$  immer der Wert  $1/2$ .

Durch die Inversionsmethode aber können mithilfe der antithetischen gleichverteilten Zufallsvariablen antithetische Zufallsvariablen in jeder beliebigen anderen Verteilung erzeugt werden.

Die Varianz von  $\hat{Y}_{AV}$  ist, in Anbetracht dessen, dass  $Y$  und  $\bar{Y}$  die gleiche Varianz besitzen, gegeben durch

$$\text{Var}[\hat{Y}_{AV}] = \frac{1}{4n} \text{Var}[Y + \bar{Y}] = \frac{1}{4n} (\text{Var}[Y] + \text{Var}[\bar{Y}] + 2\text{Cov}[Y, \bar{Y}]) \quad (3.2)$$

$$= \frac{1}{2n} (\text{Var}[Y] + \text{Cov}[Y, \bar{Y}]). \quad (3.3)$$

Im Vergleich zur Varianz des Standard Monte Carlo Schätzers wird mithilfe des antithetischen Verfahrens genau dann eine Varianzreduktion erreicht, wenn die Kovarianz zwischen den antithetischen Variablen negativ ist. Dadurch ist die Methode für gegensinnig monotone Funktionen am besten geeignet, während bei symmetrischen Funktionen keine Verbesserung erzielt werden kann.

### 3.2 Control Variates

Oft verwendet im Zusammenhang mit Monte Carlo und Varianzreduktion sind auch die sogenannten Kontrollvariablen.

Ziel ist wieder die Berechnung bzw. Schätzung des Erwartungswertes  $\mathbb{E}[Y]$ . In dieser Methode werden zusätzlich zu den  $n$  generierten iid Zufallsvariablen  $Y_1, \dots, Y_n$  im selben Durchlauf weitere  $n$  iid Zufallsvariablen  $X_1, \dots, X_n$  erzeugt, wobei die zugrunde liegenden Zufallsgrößen  $X$  und  $Y$  hoch korreliert sein sollten, um den gewünschten Erfolg zu erzielen. Dies folgt aus den anschließenden Erläuterungen zur Varianz der in der Methode verwendeten Variablen.

Der Erwartungswert  $\mathbb{E}[X]$  der Kontrollvariablen sei bekannt und die Paare  $(X_i, Y_i)$  seien iid für  $i = 1, \dots, n$ . Nun wird die Variable  $Y$  wie folgt mit Hilfe von  $X$  korrigiert,

$$Y'_i = Y_i - b \cdot (X_i - \mathbb{E}[X]), i = 1, \dots, n \quad (3.4)$$

wobei  $b$  als fixer Wert gewählt wird. Mithilfe der Berechnung des arithmetischen Mittels der  $Y'_i, i = 1, \dots, n$  kann ein Schätzer für den Erwartungswert  $\mathbb{E}[Y]$  angegeben werden,

$$\hat{Y}_{CV} = \frac{1}{n} \sum_{i=1}^n Y'_i = \frac{1}{n} \sum_{i=1}^n (Y_i - b \cdot (X_i - \mathbb{E}[X])).$$

Der Schätzer ist unverzerrt und konsistent mit

$$\text{Var}[\hat{Y}_{CV}] = \text{Var}[Y] - 2b\text{Cov}[X, Y] + b^2\text{Var}[X].$$

Diese Varianz wird minimal für

$$b^* = \frac{\text{Cov}[X, Y]}{\text{Var}[X]}, \quad (3.5)$$

was den Ausdruck

$$\text{Var}[\hat{Y}_{CV}] = \text{Var}[Y] - \frac{\text{Cov}^2[X, Y]}{\text{Var}[X]}$$

liefert. Nähere Informationen dazu können bei Glasserman [18] gefunden werden.

Vergleicht man nun die Varianz der eben erklärten Methode mit jener des ursprünglichen Vektors  $Y$ ,  $\text{Var}[Y]$ , so erhält man den Zusammenhang

$$\frac{\text{Var}[Y] - \frac{\text{Cov}^2[X, Y]}{\text{Var}[X]}}{\text{Var}[Y]} = 1 - \frac{\text{Cov}^2[X, Y]}{\text{Var}[X]\text{Var}[Y]} = 1 - \rho_{XY}^2. \quad (3.6)$$

$\rho_{XY}$  beschreibt den Korrelationskoeffizienten zwischen  $X$  und  $Y$ . Diese Relation erklärt, was bereits zu Beginn kurz angesprochen wurde, warum die Variablen  $X$  und  $Y$  hoch korrelieren sollten. Je höher die Korrelation, desto geringer die Varianz unter Verwendung der Kontrollvariablen.

In Realität lässt sich  $b^*$  nicht einfach so berechnen und Standardschätzer müssen für den Erwartungswert und die Varianz sowie Kovarianz Abhilfe schaffen. Damit bleibt der *Control Variates*-Schätzer aber zumindest asymptotisch erwartungstreu.

### 3.3 Importance Sampling

Im *Importance Sampling* wird bei der Berechnung des Erwartungswertes  $\mathbb{E}_f[g(\mathbf{X})]$  die Dichtefunktion  $f$  des Zufallsvektors  $\mathbf{X}$  modifiziert. Die Funktion  $g$  sei dabei immer als bekannt vorausgesetzt.

Gesucht ist eine Wahrscheinlichkeitsdichte  $h$  mit der Eigenschaft, dass  $\forall \mathbf{x} \in \mathbb{R}^d h(\mathbf{x}) > 0$  ist, wenn  $f(\mathbf{x}) > 0$  und

$$\mathbb{E}_f[g(\mathbf{X})] = \int_{\mathbb{R}^d} g(\mathbf{x})f(\mathbf{x}) \, d\mathbf{x} = \int_{\mathbb{R}^d} \frac{g(\mathbf{x})f(\mathbf{x})h(\mathbf{x})}{h(\mathbf{x})} \, d\mathbf{x} = \mathbb{E}_h \left[ g(\mathbf{X}) \frac{f(\mathbf{X})}{h(\mathbf{X})} \right]. \quad (3.7)$$

Die Größe  $f(\mathbf{x})/h(\mathbf{x})$  ist auch unter dem Begriff Likelihood Ratio bekannt.

Statt dem Standard Monte Carlo Schätzer

$$\hat{I}_N = \frac{1}{N} \sum_{i=1}^N h(X_i)$$

wird der *Importance Sampling* Schätzer

$$\hat{I}_{IS} = \frac{1}{N} \sum_{i=1}^N h(X_i) \frac{f(X_i)}{g(X_i)}$$

unter Berücksichtigung der Modifikation und Verwendung der Dichte  $h$  verwendet. Mit dieser Technik wird die Varianz gegenüber jener des Standard MC Schätzers verkleinert, wenn

$$\text{Var}_h \left[ g(\mathbf{X}) \frac{f(\mathbf{X})}{h(\mathbf{X})} \right] < \text{Var}_f[g(\mathbf{X})] \quad (3.8)$$

gilt. Die Varianz von  $g(\mathbf{X})f(\mathbf{X})/h(\mathbf{X})$  muss also unter der neuen Verteilung kleiner sein als jene von  $g(\mathbf{X})$  unter der ursprünglichen Verteilung.

Die optimale *Importance Sampling* Dichte ist gegeben durch

$$h^*(\mathbf{x}) = \frac{f(\mathbf{x})g(\mathbf{x})}{\mathbb{E}_f(g(\mathbf{x}))}, \quad (3.9)$$

wozu allerdings der gesuchte Erwartungswert bekannt sein müsste. Aus dieser Beobachtung folgt, dass *Importance Sampling* vor allem dann effektiv ist, wenn die Dichte  $h$  möglichst proportional zum Produkt der Dichte  $f$  mit der Funktion  $g$  liegt.

### 3.4 Stratified Sampling

Die Idee des *Stratified Samplings* ist, den Stichprobenraum in kleinere Teilräume zu unterteilen mit dem Ziel, dass die Summe der Varianzen in den einzelnen Schichten kleiner bleibt als die Varianz der Gesamtheit. Vor allem bei stufenweise definierten oder partiell in einer Ebene verlaufenden Funktionen kann dies einen großen Unterschied ausmachen, da die Varianz in flachen Regionen verhältnismäßig klein ist.

Im Folgenden soll der Erwartungswert  $\mathbb{E}[Y]$  berechnet werden, wobei  $Y$  eine reellwertige Zufallsvariable sei.  $A_1, \dots, A_K$  seien disjunkte Teilmengen des  $\mathbb{R}^d$  mit  $\mathbb{P}(Y \in \cup_i A_i) = 1$ . Nach der Formel von Bayes gilt

$$\mathbb{E}[Y] = \sum_{i=1}^K p'_i \mathbb{E}[Y|Y \in A_i] \quad (3.10)$$

wobei  $p'_i = \mathbb{P}(Y \in A_i)$  für  $i = 1, \dots, K$ .

Der Unterschied zwischen herkömmlichen Zufallsstichproben und geschichteten Stichproben liegt in der Generierung der Beobachtungen. Während üblicherweise unabhängige Zufallsvariablen  $Y_1, \dots, Y_n$  aus der Verteilung von  $Y$  generiert werden und jene, die in den Bereich  $A_i, i = 1, \dots, K$  fallen, die Wahrscheinlichkeit  $p_i$  für eine hinreichend große Wahl von  $n$  approximieren, wird bei den geschichteten Stichproben im Voraus entschieden, wieviel Elemente aus jeder Schicht  $A_i, i = 1, \dots, K$  genommen werden sollen. Diese besitzen zwangsläufig die bedingte Verteilung von  $Y|Y \in A_i$ .

Im einfachsten Fall ist der Anteil der Stichproben aus Schicht  $A_i$  proportional zur Wahrscheinlichkeit  $p_i$ ,  $i = 1, \dots, K$ . Das bedeutet, dass  $n_i = np'_i$  Exemplare aus jedem Bereich  $A_i$  gezogen werden müssen, wobei  $n$  den gesamten Stichprobenumfang bezeichnet. (Dabei sei einfachheitshalber angenommen, dass  $n_i$ ,  $i = 1, \dots, K$  natürliche Zahlen sind. Ist dies in der Praxis nicht der Fall, muss abgerundet werden,  $n_i = \lfloor np_i \rfloor$ .)

Für unabhängige Stichproben  $Y_{ij}$ ,  $j = 1, \dots, n_i$ ,  $i = 1, \dots, K$  aus der bedingten Verteilung von  $Y$  gegeben  $Y \in A_i$  ist das arithmetische Mittel  $\frac{1}{n_i}(Y_{i1} + \dots + Y_{in_i})$  ein unverzerrter Schätzer des bedingten Erwartungswertes für Schicht  $i$ . Daraus ergibt sich als Schätzer für  $\mathbb{E}[Y]$

$$\hat{Y} = \sum_{i=1}^K p'_i \frac{1}{n_i} \sum_{j=1}^{n_i} Y_{ij} = \frac{1}{n} \sum_{i=1}^K \sum_{j=1}^{n_i} Y_{ij}. \quad (3.11)$$

Um diese Methode zu verallgemeinern, wird eine zweite Variable  $X$ , die sogenannte Schichtungsvariable, eingeführt, welche Werte aus einer beliebigen Menge in  $\mathbb{R}^d$  annehmen kann und deren Verteilung bekannt ist. Die Aufteilung der Schichten sei dabei von  $X$  abhängig. Die Schichten  $A_i$ ,  $i = 1, \dots, K$  seien gegeben als disjunkte Teilmengen des  $\mathbb{R}^d$  mit  $\mathbb{P}(X \in \cup_i A_i) = 1$ . Aus Gleichung (3.10) wird damit

$$\mathbb{E}[Y] = \sum_{i=1}^K p_i \mathbb{E}[Y|X \in A_i] \quad (3.12)$$

mit  $p_i = P(X \in A_i)$ .  $Y$  kann auch eine Funktion in  $X$  sein. Die Aufgabe ist nun, Paare  $(X_{ij}, Y_{ij})$ ,  $i = 1, \dots, K$ ,  $j = 1, \dots, n_i$  mit der bedingten Verteilung von  $(X, Y)$  gegeben  $X \in A_i$  zu generieren.

Die nächste Änderung, die zu einer allgemeineren Methode führt, besteht darin, die Proportionalität aufzuheben. Vielmehr wird eine beliebige, von  $p_i$  unabhängige Wahl der Anzahl  $n_i$  an gezogenen Stichproben in Schicht  $A_i$ ,  $i = 1, \dots, K$  zugelassen. Der Anteil von  $A_i$  an der Gesamtstichprobe wird mit  $q_i = \frac{n_i}{n}$ ,  $i = 1, \dots, K$  bezeichnet. Damit verallgemeinert sich Gleichung (3.11) zu

$$\hat{Y}_{SS} = \sum_{i=1}^K p_i \frac{1}{n_i} \sum_{j=1}^{n_i} Y_{ij} = \frac{1}{n} \sum_{i=1}^K \frac{p_i}{q_i} \sum_{j=1}^{n_i} Y_{ij}. \quad (3.13)$$

Die Hauptaufgaben des *Stratified Samplings* sind die Wahl der Schichtungsvariable  $X$  und der Schichten  $A_1, \dots, A_K$  und die Zuteilung von Stichprobenelementen  $n_1, \dots, n_K$  sowie die Generierung von Stichproben mit der Verteilung von  $(X, Y)$  bedingt auf  $X \in A_i$ .

Die Varianz des Schätzers  $\hat{Y}_{SS}$  berechnet sich als

$$\text{Var}[\hat{Y}_{SS}] = \sum_{i=1}^K p_i^2 \text{Var} \left[ \frac{1}{n_i} \sum_{j=1}^{n_i} Y_{ij} \right] = \sum_{i=1}^K p_i^2 \frac{\sigma_i^2}{n_i} = \frac{\sigma^2}{n}, \quad (3.14)$$

wobei

$$\sigma^2 = \sum_{i=1}^K \frac{p_i^2}{q_i} \sigma_i^2.$$

Bei proportionaler Aufteilung der Stichproben in die einzelnen Schichten vereinfacht sich obiger Ausdruck, da  $p_i = q_i$  zu

$$\sigma^2 = \sum_{i=1}^K p_i \sigma_i^2.$$

Wird keine Schichtung durchgeführt, resultiert eine Varianz von

$$\text{Var}[Y] = \sum_{i=1}^K p_i \sigma_i^2 + \sum_{i=1}^K p_i \mu_i^2 - \left( \sum_{i=1}^K p_i \mu_i \right)^2$$

wobei  $\mu_i = \mathbb{E}[Y|X \in A_i]$  für  $i = 1, \dots, K$ . Die Ungleichung von Jensen besagt

$$\sum_{i=1}^K p_i \mu_i^2 \geq \left( \sum_{i=1}^K p_i \mu_i \right)^2$$

wobei Gleichheit nur dann gilt, wenn alle  $\mu_i$  übereinstimmen. Proportionales Schichten ist bei weitem nicht die beste Lösung, je feiner die Wahl der Aufteilung erfolgt, desto geringer die Varianz. Dennoch folgt schon im proportionalen Fall, dass die Varianz des im *Stratified Sampling* erhaltenen Schätzers nie größer wird als jene im herkömmlichen Monte Carlo Verfahren.

Die größte Varianzreduktion im *Stratified Sampling* kann mithilfe der optimalen Aufteilung

$$n_i^* = \frac{np_i \sigma_i}{\sum_{j=1}^K p_j \sigma_j}$$

erreicht werden. Diese Behauptung kann durch Anwendung von Optimierungsmethoden wie z.B. Lagrangemultiplikatoren unter den Bedingungen  $\sum_{i=1}^K n_i = n$  und  $n_i \geq 0, i = 1, \dots, K$  verifiziert werden.

Detaillierte Ausführungen zur optimalen Wahl der Stichproben können unter anderem auch bei Glasserman [18] gefunden werden.

### Beispiel 3.1. Anwendung auf das Einheitsintervall

Sei  $U \sim U(0, 1)$ , also eine auf dem Einheitsintervall gleichverteilte Zufallsvariable. Ziel sei

nun, den Erwartungswert von  $Y = g(U)$  mithilfe von geschichteten Stichproben zu berechnen bzw. zu schätzen. Dazu wählen wir zunächst eine Aufteilung  $A_1, \dots, A_n$  des Einheitsintervalls. Im einfachsten Fall sind dies  $n$  gleich große, wie auch gleich wahrscheinliche Intervalle der Länge und Wahrscheinlichkeit  $\frac{1}{n}$ ,

$$A_1 = \left(0, \frac{1}{n}\right], A_2 = \left(\frac{1}{n}, \frac{2}{n}\right], \dots, A_n = \left(\frac{n-1}{n}, 1\right),$$

aus denen jeweils eine Stichprobe gezogen wird. Das kann erreicht werden, indem  $n$  unabhängige gleichverteilte Zufallsvariablen  $U_1, \dots, U_n$  im Intervall  $(0, 1)$  generiert werden und anschließend

$$V_i = \frac{i-1}{n} + \frac{U_i}{n}, \quad i = 1, \dots, n, \quad (3.15)$$

gesetzt wird. Dadurch ergeben sich Variablen  $V_i$ , die gleichverteilt in  $\frac{i-1}{n}$  und  $\frac{i}{n}$  sind, also Variablen mit der Wahrscheinlichkeit von  $U$  bedingt auf  $U \in A_i$ .

Schließlich bleibt noch die Berechnung des Erwartungswertes bzw. des in (3.13) gegebenen Schätzers. Nachdem  $Y = g(U)$  gilt, ist der Erwartungswert von  $Y$  gegeben als das Integral von  $g$  über das Einheitsintervall. Dadurch ergibt sich als unverzerrter geschichteter Schätzer für den Erwartungswert

$$\hat{Y}_{SS} = \frac{1}{n} \sum_{i=1}^n g(V_i). \quad (3.16)$$

Mithilfe der Inversionsmethode aus Satz 2.5 kann, wie in Kapitel 2.1 besprochen, aus der Gleichverteilung jede beliebige andere Verteilung erzielt werden.

Theoretisch könnte die Methode des *Stratified Samplings* auch auf mehrere Dimensionen erweitert werden. Wiederum wird aus bekannten Gründen nur der Einheitshyperwürfel betrachtet.

Die einzelnen Schichten resultieren als Multiplikation der Intervalle, in welche jede Koordinate des Würfels aufgeteilt ist. Der Vorteil dabei ist, dass Wahrscheinlichkeiten leicht zu berechnen sind und gleichverteilte Stichproben ähnlich wie im eindimensionalen Fall erzeugt werden können, indem die Transformation (3.15) auf jede einzelne Koordinate angewandt wird.

Wäre aber jede der beispielsweise  $d$  Koordinaten in  $K$  Intervalle geteilt, wäre die Anzahl der Schichten  $K^d$  und genauso viele Stichproben müssten gezogen werden, nämlich eine aus jeder der Schichten. Bereits bei niedrigen Dimensionen bedeutet dies einen sehr hohen Aufwand. Die Option,  $K$  klein zu halten, wäre nicht von großer Sinnhaftigkeit, da in diesem Fall keine große Verbesserung der Varianz zu erwarten ist. Aus diesem Grund wird in Kapitel 4 das *Latin Hypercube* Verfahren vorgestellt, das sich darauf konzentriert, eine geringere Varianz ohne Explosion der Stichprobengröße zu erreichen.

### **Kombinierte Varianzreduktionsverfahren**

Die einzelnen Varianzreduktionstechniken sind untereinander kombinierbar, jedoch ist nicht immer eine weitere Verbesserung zu erwarten. Es kann durchaus sein, dass eine Anwendung der Methode der *Control Variates* nach Verwendung von *Antithetic Variates* eine große Reduktion der Schätzvarianz zur Folge hat, in anderen Fällen allerdings wird ein weiteres Verfahren nicht den erhofften Erfolg bringen. Die Herausforderung liegt darin, die richtige Kombination für den vorliegenden Fall zu finden.

## 4 Latin Hypercube Sampling

Wie bereits in Kapitel 3.4 erwähnt, ist die direkte Anwendung des *Stratified Samplings* auf mehrere Dimensionen mit einem enormen Aufwand verbunden. Deshalb wurde das *Latin Hypercube Sampling* von McKay, Conover und Beckham [35] vorgestellt und von Stein [61] und Owen [43] weiterentwickelt. Es hat den Vorteil, dass die Explosion des Stichprobenumfangs durch alleinige Schichtung der eindimensionalen Ränder der gemeinsamen mehrdimensionalen Verteilung, umgangen wird.

Ziel ist wiederum den Erwartungswert einer Zufallsvariable  $Y = g(\mathbf{U}^1, \dots, \mathbf{U}^d)$  zu schätzen, wobei  $\mathbf{U}^1, \dots, \mathbf{U}^d$  gleichverteilte Zufallsvektoren mit unabhängigen Komponenten darstellen.

Als erstes wird eine Dimension  $d$  und ein Stichprobenumfang  $K$  festgelegt. Im nächsten Schritt wird eine Permutation  $\pi^1, \dots, \pi^d$  der Menge  $\{1, \dots, K\}$  erzeugt, indem Elemente solange ohne Zurücklegen zufällig gezogen werden, bis nur noch die leere Menge überbleibt. Diese Permutation wird zur Generierung der Punkte  $V_i^j, i = 1, \dots, K, j = 1, \dots, d$  benötigt, die ähnlich wie im *Stratified Sampling* berechnet werden.

Hierbei gibt es zwei verschiedene Versionen. Die einfachere der beiden wird als zentriertes *Latin Hypercube Sampling* bezeichnet und wurde erstmals von Patterson [52] vorgestellt. Diese Variante kann aus einer Kombination von  $d$  Mittelpunktregele in permutierter Reihenfolge erzeugt werden. Es gilt

$$V_i^j = \frac{\pi_i^j - 1}{K} + \frac{0.5}{K}, i = 1, \dots, K, j = 1, \dots, d. \quad (4.1)$$

In der 25 Jahre später veröffentlichten Version des *Latin Hypercube Samplings* wird zusätzlich zur Permutation eine zufällige Komponente zur Erzeugung der Punkte  $V_i^j$  verwendet,

$$V_i^j = \frac{\pi_i^j - 1}{K} + \frac{U_i^j}{K}, i = 1, \dots, K, j = 1, \dots, d, \quad (4.2)$$

wobei die  $U_i^j$  für  $i = 1, \dots, K$  und  $j = 1, \dots, d$  unabhängig und gleichverteilt im Intervall  $(0, 1)$  sind.

Mithilfe der Permutation wird eine Gleichverteilung der  $K$  Punkte  $(V_i^1, \dots, V_i^d)$  für  $i = 1, \dots, K$  im Bereich  $[0, 1]^d$  erzielt. Außerdem erhalten wir in dem Verfahren eine Schichtung in jeder Dimension  $j$  durch  $(V_1^j, \dots, V_K^j)$ . Veranschaulicht wird diese Schichtung und Generierung der Zufallspunkte in Abbildung 4.1.

Der konsistente LHS Schätzer für den Erwartungswert  $I = \mathbb{E}[g(U^1, \dots, U^d)]$  ist gegeben als

$$\hat{I}_{LHS} = \frac{1}{K} \sum_{i=1}^K g(V_i^1, \dots, V_i^d) \quad (4.3)$$

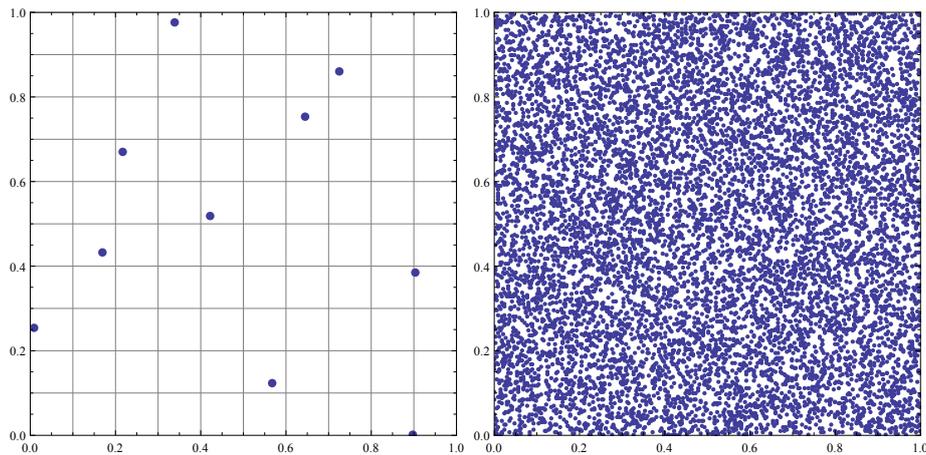


Abbildung 4.1: Links: Schichtung in 10 Bereiche; jede Spalte und jede Zeile beinhaltet einen Punkt. Rechts: 10000 Schichten

für eine integrierbare Funktion  $g : [0, 1]^d \rightarrow \mathbb{R}$ .

#### 4.1 Varianzreduktion im Latin Hypercube Sampling

In Kapitel 3.4 wurde bereits gezeigt, dass bei proportionaler Schichtung die Varianz des *Stratified Sampling* Schätzers immer kleiner oder gleich ist als die des klassischen Monte Carlo Schätzers. Dadurch ist es nicht verwunderlich, dass bei gleichzeitiger Anwendung von  $d$  proportionalen Schichtungen, wie es im *Latin Hypercube Sampling* praktiziert wird, die Varianz gegenüber dem Monte Carlo Schätzers auch dementsprechend reduziert wird.

Das erste Resultat bezüglich Varianzreduktion ist bei McKay, Conover und Beckham [35] im Falle monotoner Funktionen zu finden. Erst Jahre später zeigt Stein [61], dass die asymptotische Varianz des *Latin Hypercube Samplings* kleiner ist als die asymptotische Varianz einer Schätzung mit iid Zufallsvariablen. Der Beweis von Stein sei im Folgenden skizziert.

Eine Eigenschaft, die aus dem *Stratified Sampling* bekannt ist und sich weitgehend auf das *Latin Hypercube Sampling* übertragen lässt, ist die Effizienz des Verfahrens im Hinblick auf stückweise konstante Funktionen, bzw. Funktionen, die sich als Summe eindimensionaler Funktionen zusammensetzen. Um den Integrand  $g$  in eine Summe von einfacheren Funktionen zu zerlegen, können sogenannte *Functional ANOVAs* d.h. funktionale Varianzanalysen angewandt werden, die im Anschluss beschrieben werden.

Bezeichne  $\mathcal{M} = \{1, \dots, d\}$  die Menge der Koordinatenachsen in  $[0, 1]^d$ ,  $u$  eine Teilmenge von  $\mathcal{M}$  mit Kardinalität  $|u|$  und Komplement  $-u$ .  $[0, 1]^u$  sei der Raum, der von den Komponenten  $X^j$  von  $\mathbf{X}$  mit  $j \in u$  aufgespannt wird. Des Weiteren sei  $X^u$  die Koordinatenprojektion vom Punkt  $\mathbf{X}$  auf den Bereich  $[0, 1]^u$  und  $dF^u = \prod_{j \in u} dX^j$  sowie  $dF^{\mathcal{M}-u}$  das Komplement davon.

Unter den Voraussetzungen, dass die Funktion  $g$  messbar ist und

$$\int g(\mathbf{X})^2 dF = \sum_{u \subseteq \mathcal{M}} \int g_u^2(X) dF < \infty$$

gilt, kann  $g(\mathbf{X})$  geschrieben werden als

$$g(\mathbf{X}) = \sum_{u \subseteq \mathcal{M}} g_u(X),$$

eine Summe von  $2^d$  orthogonalen Funktionen, wobei jeder Effekt  $g_u$  nur von den Komponenten in  $X^u$  abhängt.

Die Funktionen  $g_u$  sind rekursiv definiert als

$$g_u = \int \left( g - \sum_{v \subset u} g_v \right) dF^{\mathcal{M}-u}, \quad (4.4)$$

wobei  $v \subset u$  alle echten Teilmengen mit  $v \neq u$  bezeichnet.

Im Falle von  $u = \emptyset$  bedeutet dies

$$g_\emptyset = \int g(\mathbf{X}) dF,$$

was genau den zu schätzenden Erwartungswert darstellt.

Für die einelementigen Mengen  $u = \{j\}$  für  $j = 1, \dots, d$  gilt

$$g_{\{j\}} = \int (g(\mathbf{X}) - g_\emptyset) dF^{\mathcal{M}-\{j\}}, \quad (4.5)$$

und so weiter. Der Ausdruck  $g_{\{j\}}(X^j)$  wird auch als Haupteffekt der  $j$ -ten Variable auf die Funktion  $g(\mathbf{X})$  bezeichnet, die Funktionen über Mengen mit mehr als einem Element beschreiben die Interaktion zwischen den zugehörigen Variablen.

Basierend auf den Überlegungen von Stein [61] ist die beste additive Approximation von  $g$

$$g_{add}(\mathbf{X}) = g_\emptyset + \sum_{j=1}^d g_{\{j\}}. \quad (4.6)$$

Auch die Varianz  $\sigma_g^2$  kann zerlegt werden in

$$\sigma_g^2 = \sum_u \sigma_u^2 = \sum_{|u|>0} \sigma_u^2, \quad (4.7)$$

wobei  $\sigma_u^2$  die Varianz der Funktion  $g_u$  beschreibt, d.h.  $\sigma_u^2 = \int g_u^2 dF^u$  für  $u \neq \emptyset$  und  $\sigma_\emptyset^2 = 0$ .

Mithilfe solch einer Zerlegung der Funktion  $g$  bzw. einer solchen Varianzanalyse kann die Varianz des *Latin Hypercube Sampling* Schätzers nun detaillierter betrachtet werden.

Sei  $g_{add}$  im quadratischen Mittel die beste additive Approximation von  $g$ , sodass

$$\begin{aligned} & \int (g(u_1, \dots, u_d) - g_{add}(u_1, \dots, u_d))^2 du_1 \cdots du_d \\ & \leq \int \left( g(u_1, \dots, u_d) - \sum_{j=1}^d g_{\{j\}}(u_j) \right)^2 du_1 \cdots du_d, \end{aligned}$$

wobei  $g_{\{j\}}$  für  $j = 1, \dots, d$  Funktionen in einer Variable beschreiben. Sei

$$\epsilon = g(U_1, \dots, U_d) - g_{add}(U_1, \dots, U_d)$$

das mit  $g(U_1, \dots, U_d)$  unkorrelierte Residuum, so kann die Varianz  $\sigma_g^2$  von  $g(U_1, \dots, U_d)$  zerlegt werden in

$$\sigma_g^2 = \sigma_{add}^2 + \sigma_\epsilon^2,$$

mit  $\sigma_{add}^2$  als Varianz von  $g_{add}(U_1, \dots, U_d)$  und  $\sigma_\epsilon^2$  als Varianz des Residuums.

Stein [61] zeigt für  $K \rightarrow \infty$ , dass für die Varianz des *Latin Hypercube Samplings* unter der Voraussetzung, dass  $\int g(\mathbf{x})^2 d\mathbf{x} < \infty$

$$\begin{aligned} \text{Var}[\hat{I}_{LHS}] &= \frac{1}{K} \int (g(u_1, \dots, u_d) - g_{add}(u_1, \dots, u_d))^2 du_1 \cdots du_d + o\left(\frac{1}{K}\right) \\ &= \frac{1}{K} \left( \sigma_g^2 - \sum_{j=1}^d \sigma_{\{j\}}^2 \right) + o\left(\frac{1}{K}\right) \\ &= \frac{1}{K} \sigma_\epsilon^2 + o\left(\frac{1}{K}\right) \end{aligned}$$

gilt. Diese Relation zeigt, dass das *Latin Hypercube* Verfahren aus der Varianz des gesamten Schätzers jenen Teil, der dem additiven Teil der Funktion  $g$  entspricht, weitgehend eliminiert. Kann der Integrand  $g$  nicht gut in annähernd eindimensionale Funktionen gesplittet werden, so dürfen keine allzu großen Hoffnungen in eine große Varianzreduktion gesetzt werden, aber

zumindest gilt noch immer die nachstehende Aussage (4.8) von Owen [47], womit eine höhere Varianz als im Monte Carlo Verfahren auszuschließen ist.

Owen [47] zeigt, dass das LHS-Verfahren als *scrambled*  $(0, 1, d)$ -Netz in Basis  $b = K$  gesehen werden kann und verwendet diese Tatsache um folgende Aussage zu beweisen. Unter der Annahme, dass die Funktion  $g$  im gewünschten Erwartungswert  $\mathbb{E}[g(X)]$  eine quadratisch integrierbare Funktionen ist, gilt stets

$$\text{Var}[\hat{I}_{LHS}] \leq \frac{1}{K-1} \text{Var}[\hat{I}_K], \quad (4.8)$$

wobei  $\text{Var}[\hat{I}_K]$  die Varianz des herkömmlichen Monte Carlo Schätzers mit  $K$  Stichprobenelementen darstellt.

## 4.2 Zentraler Grenzwertsatz für das Latin Hypercube Sampling

Bei Owen [43] ist ein zentraler Grenzwertsatz für den *Latin Hypercube* Schätzer  $\hat{I}_{LHS}$  im Falle von beschränkten Funktionen  $g$  zu finden. Es gilt

$$\sqrt{K}(\hat{I}_{LHS} - I) \xrightarrow{p} N(0, \sigma_\epsilon^2),$$

wobei aber beachtet werden muss, dass in Realität die Größe  $\sigma_\epsilon^2$  weder bekannt ist, noch kann sie leicht geschätzt werden.

Loh [33] gibt eine Schranke vom Typ Berry-Esseen an, welche die Konvergenzrate des auf bestimmte Art und Weise standardisierten Fehlers gegen die Standardnormalverteilung ermittelt.

### Satz 4.1. (Satz von Berry Esseen)

Seien  $X_1, \dots, X_K$  unabhängige Zufallsvariablen mit  $\mathbb{E}[X_1] = 0$ ,  $\mathbb{E}[X_1^2] = \sigma^2 > 0$ ,  $\mathbb{E}[|X_1|^3] = \rho < \infty$  so gilt für

$$W = \frac{X_1 + \dots + X_K}{\sqrt{K\sigma^2}}$$

und eine Konstante  $C$  die Abschätzung

$$\sup_{x \in \mathbb{R}} |\mathbb{P}(W \leq x) - \Phi(x)| \leq C \frac{\rho}{\sigma^3 \sqrt{K}}.$$

$\Phi$  stellt die Verteilungsfunktion der Standardnormalverteilung dar.

In der Version von Loh für das *Latin Hypercube Sampling* wird die Schätzgröße folgendermaßen transformiert.

$$W = \sum_{i=1}^K T(\pi_i^1, \dots, \pi_i^d), \quad (4.9)$$

wobei  $\pi_i^j$  für  $i = 1, \dots, K$  und  $j = 1, \dots, d$  die Permutation des Verfahrens symbolisiert.

$$T(i^1, \dots, i^d) = \frac{1}{K\sqrt{\text{Var}[\hat{I}_{LHS}]}} \left( g(V(i^1, \dots, i^d)) - \sum_{k=1}^d \mu_{-k}(i^k) + (d-1)\mu(i^1, \dots, i^d) \right),$$

mit  $V(i^1, \dots, i^d) := (V_{i^j}^1, \dots, V_{i^j}^d)$  für  $j = 1, \dots, d, 1 \leq i^1, \dots, i^d \leq K$  und

$$\mu_{-k}(i^k) = \frac{1}{K^{d-1}} \sum_{j \neq k} \sum_{i^j=1}^K \mu(i^1, \dots, i^d),$$

wobei  $\mu(i^1, \dots, i^d) = \mathbb{E}[g(V(i^1, \dots, i^d))]$ .

**Satz 4.2. (Loh [33])**

Es existiert eine Konstante  $C_d > 0$ , die nur von der Dimension  $d$  abhängt, sodass

$$\sup_{z \in \mathbb{R}} |\mathbb{P}(W \leq z) - \Phi(z)| \leq C_d \beta_3. \quad (4.10)$$

$\Phi$  symbolisiert die Standardnormalverteilung und  $\beta_3$  ist gegeben durch

$$\beta_3 = \frac{1}{K^{d-1}} \sum_{i^1=1}^K \cdots \sum_{i^d=1}^K \mathbb{E}|T(i^1, \dots, i^d)|^3.$$

Als unmittelbare Folgerung daraus gilt

**Korollar 4.3.** Wenn  $\mathbb{E}|g(V)|^3 < \infty$ , so gilt

$$\sup_{z \in \mathbb{R}} |\mathbb{P}(W \leq z) - \Phi(z)| \leq \frac{C_d}{\sqrt{K}}.$$

Im Korollar ist also  $\frac{C_d}{\sqrt{K}}$  als Konvergenzgeschwindigkeit gegeben. Bei Loh [33] ist allerdings kein Wert für die Konstante gegeben. Daraufhin haben sich Neammanee and Rattanawong [37] mit diesem Thema beschäftigt, die unter der strikteren Voraussetzung der Endlichkeit des 4. Moments einen exakten Wert angeben können. Die neuesten Ergebnisse dazu stammen von Rerkruthairat und Neammanee [53], die wie Loh unter anderem die Existenz des 3. Moments voraussetzen und folgende Konstante erhalten.

**Satz 4.4.** Wenn  $\mathbb{E}|g(V(i^1, \dots, i^d))|^3 < \infty$  für  $1 \leq i^1, \dots, i^d \leq K$ , so gilt für  $K \geq 6^d$  und  $\delta_3 \approx K^{-1/2}$

$$\sup_{z \in \mathbb{R}} |\mathbb{P}(W \leq z) - \Phi(z)| \leq \frac{26.76 + 31.08 d}{\sqrt{K}} + \frac{2.92 d}{K} + O(K^{-\frac{13}{24}})$$

Der Ausdruck  $\delta_3$  entspricht im Großen und Ganzen  $\beta_3$  aus Satz 4.2, wobei aber andere Permutationen zur Berechnung von  $T$  verwendet werden.

**Bemerkung**

In Absatz 2.2.4 wurde bereits auf die Möglichkeit hingewiesen, bei Verwendung von Quasi-Monte Carlo Verfahren, die zur Schätzung benötigte Matrix mit Einträgen aus  $[0, 1)$  mithilfe von zufälligen, gleichverteilten Punkten “aufzublasen”, sollte die Dimension zu hoch werden. Anstatt nun das klassische Monte Carlo Verfahren anzuwenden, kann auch das *Latin Hypercube Sampling* für die restlichen  $d - s$  Spalten Abhilfe schaffen. Owen [44] gibt ein Beispiel für solche ein Verfahren an. Wiederum macht es den größten Sinn Monte Carlo gegen *Latin Hypercube* zu tauschen, wenn eine Variable einen rein additiven Einfluss auf die Funktion  $g$  ausübt. In solch einem Fall sollte die Variable auf alle Fälle im Pool der mit *Latin Hypercube* geschätzten Größen sein.

## 5 Latin Hypercube Sampling with Dependence

Für das herkömmliche *Latin Hypercube Sampling* wird die Unabhängigkeit der Stichproben vorausgesetzt. Wird das *Latin Hypercube* Verfahren auf Zufallsvektoren angewandt, deren Komponenten voneinander abhängen, so wird durch Verwendung von untereinander unabhängigen Permutationen diese Abhängigkeit zerstört.

Eine Überlegung wäre, das *Latin Hypercube Sampling* auf unabhängige Vektoren anzuwenden und diese dann so zu transformieren, dass wieder eine Abhängigkeit entsteht. Dies widerspricht allerdings dem eigentlichen Sinn des Verfahrens, nämlich die Schichtung der Grenzen. Auf Grund dessen gibt es die Idee des *Latin Hypercube Samplings With Dependence (LHSD)*.

Das *LHSD* unterscheidet sich vom ursprünglichen Verfahren durch Verwendung einer Rangstatistik anstelle einer zufälligen Permutation, welche die Abhängigkeit zwischen den Vektorkomponenten modellieren und erhalten soll. Die anschließenden Ausführungen und Beweise sind angelehnt an die Ergebnisse von Packham und Schmidt [48].

### Definition 5.1. (Rangstatistik)

Seien  $X_1, \dots, X_n$  unabhängig, identisch verteilte Zufallsvariablen mit stetiger Verteilungsfunktion. Seien  $X^{(1)}, \dots, X^{(n)}$  die geordneten Zufallsvariablen mit  $X^{(1)} < \dots < X^{(n)}$  fast sicher. Dann wird der Index von  $X_i$  innerhalb von  $X^{(1)}, \dots, X^{(n)}$  die  $i$ -te Rangstatistik genannt und ist gegeben durch

$$r_{i,n}(X_1, \dots, X_n) := \sum_{k=1}^n 1_{\{X_k \leq X_i\}}. \quad (5.1)$$

Im Folgenden betrachten wir Zufallsvektoren  $(U^1, \dots, U^d)$  mit gleichverteilten Komponenten  $U^j \sim U(0, 1)$  für  $j = 1, \dots, d$ . Die Abhängigkeit der Komponenten wird mithilfe einer Copula  $C$  beschrieben.  $(U_i^1, \dots, U_i^d)$  für  $i = 1, \dots, n$  seien  $n$  iid Stichproben von  $(U^1, \dots, U^d)$ . Dann hat die Stichprobe im *LHSD* Verfahren mit Abhängigkeiten die Form

$$V_{i,n}^j = \frac{r_{i,n}^j - 1}{n} + \frac{\eta_{i,n}^j}{n}, \quad i = 1, \dots, n, j = 1, \dots, d, \quad (5.2)$$

wobei  $r_{i,n}^j$  für  $i = 1, \dots, n$  und  $j = 1, \dots, d$  die  $i$ -te Rangstatistik des Vektors  $(U_1^j, \dots, U_n^j)$  darstellt und  $\eta_{i,n}^j$  Zufallsvariablen sind, die nur Werte im Intervall  $[0, 1]$  annehmen können.

Wie schon im vorher beschriebenen *Latin Hypercube Sampling* stellen auch hier die einzelnen Vektoren  $(V_1^j, \dots, V_n^j)$  jeder Dimension  $j = 1, \dots, d$  eine Anwendung des *Stratified Samplings* mit proportionaler Schichtung dar.

Wird die Zufallsvariable  $\eta_{i,n}^j$  durch die Komponente  $U_i^j$  ersetzt, so erhalten wir einen ähnlichen Ausdruck wie im *Latin Hypercube Sampling*, aber die Rangstatistik bewirkt, im Gegensatz zur

zufälligen Permutation, dass die Stichprobe nicht länger innerhalb der Schicht gleichverteilt ist. Vielmehr gilt, dass die Verteilung von  $U_i^j$  bedingt auf  $r_{i,n}^j = k$  eine Beta-Verteilung mit Parametern  $k$  und  $n$  ist,  $\mathbb{P}(U_i^j \leq x | r_{i,n}^j = k) = B_k^n(x)$ . Dies folgt aus der Tatsache, dass die Verteilung der  $k$ -ten Rangstatistik von  $n$  unabhängigen Zufallsvariablen eine  $B_k^n$ -Verteilung ist. Dieses Resultat folgt aus den Überlegungen von Feller [16].

Ein oft besseres Ergebnis, vor allem rechnerisch betrachtet, kann erzielt werden, wenn  $\eta_{i,n}^j = 1/2$  gewählt wird, was bedeutet, dass jede Stichprobe in der Mitte der Schicht liegt oder aber  $\eta_{i,n}^j = 1$ , so dass  $V_{i,n}^j$  die empirische Funktion von  $(U_1^j, \dots, U_n^j)$  an der Stelle  $U_i^j$  darstellt.

Der Schätzer des Erwartungswertes  $\mathbb{E}[g(U^1, \dots, U^d)]$  für eine integrierbare Funktion  $g : [0, 1]^d \rightarrow \mathbb{R}$  und einen Zufallsvektor  $(U^1, \dots, U^d)$  mit gleichverteilten Rändern und Copula  $C$  ist gegeben als

$$\hat{I}_{LHSD} = \frac{1}{n} \sum_{i=1}^n g(V_{i,n}^1, \dots, V_{i,n}^d), \quad (5.3)$$

mit  $V_{i,n}^j$  für  $i = 1, \dots, n, j = 1, \dots, d$  wie in Gleichung (5.2).

Wie zu erwarten, bleibt das Abhängigkeitsverhältnis, also die Copula, bei der Transformation von  $(U^1, \dots, U^d)$  auf  $(V_{i,n}^1, \dots, V_{i,n}^d)$  für  $i = 1, \dots, n$  nicht das- bzw. dieselbe. Was aber sehr wohl gilt, ist die Konvergenz der  $V_{i,n}^j$  gegen  $U_i^j$  für  $n \rightarrow \infty$ , wenn der Anteil jener  $V_{k,n}^j$  für  $k = 1, \dots, n$ , welche kleiner als  $V_{i,n}^j$  sind, gegen  $U_i^j$  konvergiert. Diese Eigenschaft folgt direkt aus der Rangstatistik und wird in Lemma 5.5 gezeigt.

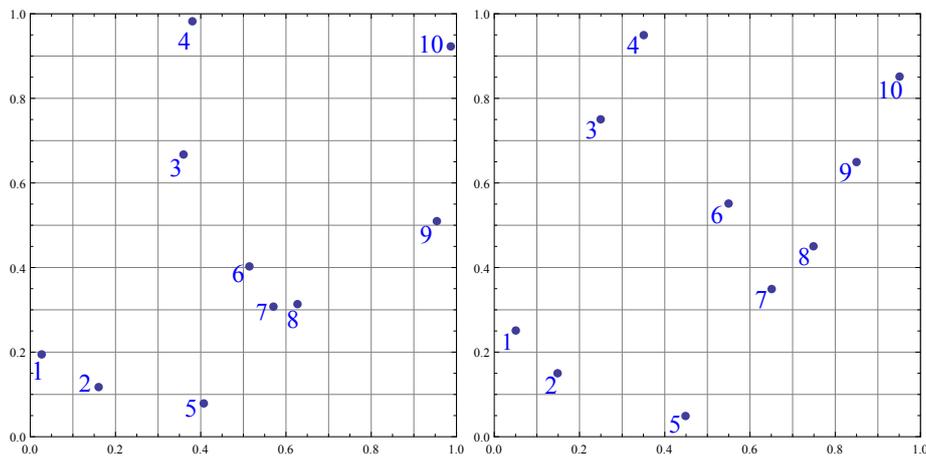


Abbildung 5.1: Links: 10 zufällige Punkte der bivariaten Gauß-Copula; Rechts: Anwendung des LHS auf die selben 10 Punkte

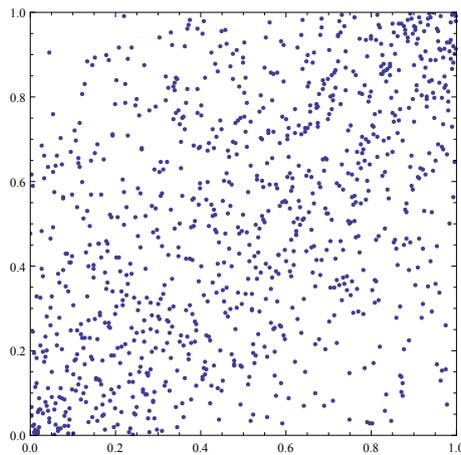


Abbildung 5.2: LHSD Verfahren für 1000 zufällige Punkte der bivariaten Gauß-Copula

Abbildung 5.1 zeigt links zufällig generierte Punkte aus der bivariaten Gauß-Copula. Rechts wurde das *LHSD*-Verfahren auf genau diese Punkte angewandt. Sehr schön zu erkennen ist, dass die Reihenfolge der Punkte gleich bleibt und sich nur die Verteilung innerhalb des zweidimensionalen Einheitsintervalls ändert. Wie schon im *Latin Hypercube Sampling* ist wiederum nur ein Punkt pro Schichtung, hier proportional, erlaubt. Abbildung 5.2 zeigt die Anwendung des Verfahrens auf 1000 Punkte. Deutlich zu erkennen ist die typische Form einer Stichprobe der Gauß-Copula, die im Wesentlichen erhalten bleibt.

## 5.1 Konsistenz des LHSD Schätzers

Aufgrund der Abhängigkeit der einzelnen Komponenten und der Tatsache, dass sich, bei Übergang von  $n$  auf  $n + 1$  Stichprobenelemente, dies auf alle Summanden in (5.3) auswirkt und nicht nur ein weiteres hinzugefügt wird, werden die üblichen Gesetze der großen Zahlen, die unabhängige Zufallsvariablen voraussetzen, außer Kraft gesetzt. Deshalb wird im Folgenden die Konsistenz des *LHSD* Schätzers unter der Annahme, dass die Funktion  $g$  beschränkt ist und diverse Stetigkeitsbedingungen erfüllt, gezeigt. Zusätzlich gibt es noch die Anforderung der Austauschbarkeit an die Zufallsvariablen, die im Folgenden definiert wird.

### Definition 5.2.

Zufallsvariablen  $X_1, \dots, X_n$  heißen *austauschbar*, falls für jede mögliche Permutation  $\pi$  von der Menge  $\{1, \dots, n\}$  die Vektoren  $\mathbf{X} = (X_1, \dots, X_n)$  und  $\mathbf{X}_\pi = (X_{\pi_1}, \dots, X_{\pi_n})$  die gleiche Verteilung besitzen, d.h.

$$(X_1, \dots, X_n) \stackrel{d}{=} (X_{\pi_1}, \dots, X_{\pi_n}),$$

wobei  $\stackrel{d}{=}$  Gleichheit in Verteilung symbolisiert.

**Lemma 5.3.**

Sei  $\nu_{i,n} := (U_i^1, \dots, U_i^d, \eta_{i,n}^1, \dots, \eta_{i,n}^d)$ . Wenn die Vektoren  $\nu_{1,n}, \dots, \nu_{n,n}$  austauschbar sind, so gilt

$$(U_i^1, \dots, U_i^d, V_{i,n}^1, \dots, V_{i,n}^d) \stackrel{d}{=} (U_k^1, \dots, U_k^d, V_{k,n}^1, \dots, V_{k,n}^d), \quad (5.4)$$

für  $i, k \leq n$ .

*Beweis.* Die Aussage folgt für beschränkte stetige Funktionen  $g : [0, 1]^{2d} \rightarrow \mathbb{R}$  direkt aus der Austauschbarkeit der Vektoren  $\nu_{1,n}, \dots, \nu_{n,n}$  nach Einsetzen der  $V_{i,n}^1, \dots, V_{i,n}^d$  in den Erwartungswert.  $\square$

Für die bereits erwähnten Vorschläge zur Bestimmung der  $(\eta_{i,n}^j)_{j=1, \dots, d; i=1, \dots, n}$  gilt einerseits, dass  $\eta_{i,n}^j = \eta_{i,m}^j$  für alle  $m, n$  und außerdem, dass die Vektoren  $(\eta_i^1, \dots, \eta_i^d)$  für  $i = 1, \dots, n$  iid sind und zusätzlich unabhängig von  $(U_i^1, \dots, U_i^d)$  für  $i = 1, \dots, d$ . Daraus folgt sofort, dass auch die  $\nu_{1,n}, \dots, \nu_{n,n}$  iid sind, woraus mit Proposition 5.4 die Austauschbarkeit resultiert.

**Proposition 5.4.**

*Unabhängige und identisch verteilte Zufallsvariablen sind austauschbar.*

*Beweis.*

$\mathbf{X} = (X_1, \dots, X_n)$  sei ein Vektor von iid Zufallsvariablen mit  $X_i \sim F(x)$  für  $i = 1, \dots, n$ .  $\pi$  sei eine Permutation der Menge  $\{1, \dots, n\}$  und  $\mathbf{X}_\pi = (X_{\pi_1}, \dots, X_{\pi_n})$  die permutierte Version von  $\mathbf{X}$ . Wegen der Unabhängigkeit gilt dann

$$G(\mathbf{x}) = \prod_{i=1}^n F(x_i) = G_\pi(\mathbf{x})$$

wobei  $G(\mathbf{x})$  die gemeinsame Verteilungsfunktion von  $\mathbf{X}$  und  $G_\pi(\mathbf{x})$  jene von  $\mathbf{X}_\pi$  darstellt. Aus Definition 5.2 folgt sofort die Austauschbarkeit.  $\square$

Um die Konsistenz des *LHSD* Schätzers zu zeigen, sind noch weitere Vorbereitungen nötig. Als erstes sollte die bereits angesprochene Konvergenz der  $V_{i,n}^j$  gegen  $U_i^j$ ,  $i = 1, \dots, n$ ,  $j = 1, \dots, d$  für  $n \rightarrow \infty$  geklärt werden.

**Lemma 5.5.**

*Es gilt*

$$\sup_{i \in \{1, \dots, n\}} |V_{i,n}^j - U_i^j| \xrightarrow{\mathbb{P}\text{-f.s.}} 0, \quad \text{für } n \rightarrow \infty, \quad (5.5)$$

für alle Dimensionen  $j = 1, \dots, d$ .

*Beweis.*

Der Beweis gilt für jede Dimension im Einzelnen, daher wird im Folgenden, auch aus Gründen

der Übersicht der Index für die Dimension vernachlässigt.

Sei  $i, n$  mit  $i \leq n$  fix gewählt. Dann gilt nach Definition der Rangstatistik

$$V_{i,n} = \frac{r_{i,n}(U_1, \dots, U_i) - 1 + \eta_{i,n}}{n} = \frac{1}{n} \sum_{k=1}^n \mathbb{1}_{\{U_k \leq U_i\}} - \frac{1 - \eta_{i,n}}{n} = F_n(U_i) - \frac{1 - \eta_{i,n}}{n}, \quad (5.6)$$

wobei  $F_n(U_i)$  die empirische Verteilungsfunktion für  $i = 1, \dots, n$  darstellt.

Der Satz von Glivenko-Cantelli besagt für die Gleichverteilung, dass

$$\sup_{u \in [0,1]} |F_n(u) - u| \xrightarrow{\mathbb{P}\text{-f.s.}} 0 \quad \text{für } n \rightarrow \infty.$$

Außerdem gilt stets  $1 - \eta_{i,n} \leq 1$  womit als Ergebnis

$$\sup_{u \in [0,1]} \left| F_n(U_i) - U_i - \frac{1 - \eta_{i,n}}{n} \right| \xrightarrow{\mathbb{P}\text{-f.s.}} 0 \quad \text{für } n \rightarrow \infty$$

folgt, womit der Satz bewiesen ist. □

Als nächstes wird die Konvergenz der empirischen Verteilungsfunktion, der im *LHSD* Verfahren generierten Punkte gegen die ursprüngliche Copula der  $(U^1, \dots, U^d)$  gezeigt, woraus folgt, dass die Randverteilungen der *LHSD* Stichprobe auf dem Einheitsintervall asymptotisch gleichverteilt sind.

**Lemma 5.6.**

Sei  $C_n : [0, 1]^d \rightarrow [0, 1]$  definiert durch

$$C_n(u^1, \dots, u^d) := \frac{1}{n} \sum_{k=1}^n \mathbb{1}_{\{V_{k,n}^1 \leq u^1, \dots, V_{k,n}^d \leq u^d\}}, \quad (5.7)$$

für  $u^1, \dots, u^d \in [0, 1]$ . Dann beschreibt  $C_n$  eine Verteilungsfunktion für die gilt

$$\sup_{(u^1, \dots, u^d) \in [0,1]^d} |C_n(u^1, \dots, u^d) - C(u^1, \dots, u^d)| \xrightarrow{\mathbb{P}\text{-f.s.}} 0 \quad \text{für } n \rightarrow \infty$$

*Beweis.*

Dass  $C_n$  eine Verteilungsfunktion im Bereich  $[0, 1]^d$  ist für  $n \in \mathbb{N}$ , lässt sich leicht verifizieren.

Sei nun  $F_n^j$  die empirische Verteilungsfunktion von  $U_1^j, \dots, U_n^j$  für  $j = 1, \dots, d$  und sei  $\tilde{C}_n : [0, 1]^d \rightarrow [0, 1]$  definiert durch

$$\tilde{C}_n(u^1, \dots, u^d) := \frac{1}{n} \sum_{k=1}^n \mathbb{1}_{\{F_n^1(U_k^1) \leq u^1, \dots, F_n^d(U_k^d) \leq u^d\}}. \quad (5.8)$$

Dann gilt

$$\sup_{(u^1, \dots, u^d) \in [0,1]^d} \left| \tilde{C}_n(u^1, \dots, u^d) - C(u^1, \dots, u^d) \right| \xrightarrow{\mathbb{P}\text{-f.s.}} 0 \quad \text{für } n \rightarrow \infty. \quad (5.9)$$

Diese Behauptung ist eine Folgerung des Satzes 3.1 von Deheuvels [13].  
Offensichtlich gilt

$$F_n^j(U_k^j) = \frac{r_{k,n}^j}{n}$$

woraus für  $(u^1, \dots, u^d) \in [0, 1]^d$

$$\left| C_n(u^1, \dots, u^d) - \tilde{C}_n(u^1, \dots, u^d) \right| \leq \frac{1}{n} \sum_{k=1}^n \mathbb{1} \left\{ u^1 \in \left[ \frac{r_{k,n}^1-1}{n}, \frac{r_{k,n}^1}{n} \right), \dots, u^d \in \left[ \frac{r_{k,n}^d-1}{n}, \frac{r_{k,n}^d}{n} \right) \right\} \leq \frac{1}{n} \quad (5.10)$$

folgt. Durch Zusammensetzung der Gleichungen (5.9) und (5.10) folgt die Aussage.  $\square$

Es folgen noch zwei Behauptungen, die für den Konsistenzbeweis notwendig sind, deren Beweise hier aber nicht erläutert werden.

**Lemma 5.7.**

Wenn  $\mathbf{X}, \mathbf{X}_1, \mathbf{X}_2, \dots$  Zufallsvektoren im  $\mathbb{R}^d$  sind, für die  $\mathbf{X}_n \xrightarrow{\mathbb{P}} \mathbf{X}$  für  $n \rightarrow \infty$  gilt und  $g : \mathbb{R}^d \rightarrow \mathbb{R}$  eine messbare und  $\mathbb{P}$ -f.s. stetige Funktion in  $\mathbf{X}$  ist, so konvergiert  $g(\mathbf{X}_n)$   $\mathbb{P}$ -f.s. gegen  $g(\mathbf{X})$  für  $n \rightarrow \infty$ .

**Lemma 5.8.**

Für  $\mathbf{X} = (X^1, \dots, X^d)$  und  $\mathbf{X}_n = (X_n^1, \dots, X_n^d)$ ,  $n \in \mathbb{N}$  Zufallsvektoren im  $\mathbb{R}^d$  gilt  $\mathbf{X}_n \xrightarrow{\mathbb{P}} \mathbf{X}$  genau dann wenn  $\mathbf{X}_n^j \xrightarrow{\mathbb{P}} \mathbf{X}^j$  für alle  $j = 1, \dots, d$ .

**Proposition 5.9.**

Sei  $g : [0, 1]^d \rightarrow \mathbb{R}$  eine beschränkte stetige Funktion. Dann ist der LHSD Schätzer stark konsistent, d.h. es gilt

$$\frac{1}{n} \sum_{i=1}^n g(V_{i,n}^1, \dots, V_{i,n}^d) \xrightarrow{\mathbb{P}\text{-f.s.}} \mathbb{E}[g(U^1, \dots, U^d)] \quad \text{für } n \rightarrow \infty. \quad (5.11)$$

*Beweis.*

Aus der Definition von  $C_n$  in (5.7) folgt für das Integral

$$\int_{[0,1]^d} g \, dC_n = \frac{1}{n} \sum_{k=1}^n g(V_{k,n}^1, \dots, V_{k,n}^d),$$

was genau den *LHSD* Schätzer aus (5.3) ergibt.

Aus Lemma 5.6 folgt die schwache Konvergenz von  $C_n$  gegen  $C$  für  $\mathbb{P}$ -fast alle  $\omega \in \Omega$  was äquivalent ist zur Konvergenz der Integrale

$$\int_{[0,1]^d} g \, dC_n \rightarrow \int_{[0,1]^d} g \, dC \quad \text{für } \mathbb{P}\text{-fast alle } \omega.$$

Das Integral  $\int g \, dC$  entspricht genau dem Erwartungswert  $\mathbb{E}[g]$ , wobei  $g$  eine beschränkte stetige Funktion von  $[0, 1]^d$  nach  $\mathbb{R}$  ist.  $\square$

**Proposition 5.10.**

Sei  $g : [0, 1]^d \rightarrow \mathbb{R}$  eine beschränkte  $C$ -fast überall stetige Funktion. Dann ist der *LHSD* Schätzer konsistent, d.h. es gilt

$$\frac{1}{n} \sum_{i=1}^n g(V_{i,n}^1, \dots, V_{i,n}^d) \xrightarrow{\mathbb{P}} \mathbb{E}[g(U^1, \dots, U^d)] \quad \text{für } n \rightarrow \infty. \quad (5.12)$$

*Beweis.*

Sei  $i \in \mathbb{N}$  fixiert.

Aus Lemma 5.5 und Lemma 5.8 folgt, dass  $V_{i,n}^j \xrightarrow{\mathbb{P}} U_i^j$  für  $j = 1, \dots, d$  und weiters die Konvergenz des Vektors  $(V_{i,n}^1, \dots, V_{i,n}^d) \xrightarrow{\mathbb{P}} (U_i^1, \dots, U_i^d)$ .

Mit Lemma 5.7 folgt für die  $C$ -fast überall stetige Funktion  $g$  die Konvergenz  $g(V_{i,n}^1, \dots, V_{i,n}^d) \xrightarrow{\mathbb{P}} g(U_i^1, \dots, U_i^d)$ . Durch die Beschränktheit von  $g$  folgt mithilfe des Satzes der dominierten Konvergenz

$$\mathbb{E} [|g(V_{i,n}^1, \dots, V_{i,n}^d) - g(U_i^1, \dots, U_i^d)|] \rightarrow 0, \quad \text{für } n \rightarrow \infty. \quad (5.13)$$

Nachdem das starke Gesetz der großen Zahlen die Konvergenz

$$\frac{1}{n} \sum_{i=1}^n g(U_i^1, \dots, U_i^d) \xrightarrow{\mathbb{P}\text{-f.s.}} \mathbb{E}[g(U^1, \dots, U^d)] \quad \text{für } n \rightarrow \infty$$

garantiert, reicht es, für den Beweis von (5.12)

$$\lim_{n \rightarrow \infty} \mathbb{P} \left( \left| \frac{1}{n} \sum_{i=1}^n g(V_{i,n}^1, \dots, V_{i,n}^d) - \frac{1}{n} \sum_{i=1}^n g(U_i^1, \dots, U_i^d) \right| > \varepsilon \right) = 0 \quad (5.14)$$

für alle  $\varepsilon > 0$  zu zeigen. Diese Gleichung ist äquivalent zu

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[ \left| \frac{1}{n} \sum_{i=1}^n (g(V_{i,n}^1, \dots, V_{i,n}^d) - g(U_i^1, \dots, U_i^d)) \right| \wedge 1 \right] = 0 \quad (5.15)$$

basierend auf den Aussagen von Kallenberg [26].

Für beliebiges  $n$  können folgende Abschätzungen getroffen werden,

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{E} \left[ \left| \frac{1}{n} \sum_{i=1}^n (g(V_{i,n}^1, \dots, V_{i,n}^d) - g(U_i^1, \dots, U_i^d)) \right| \wedge 1 \right] &\leq \mathbb{E} \left[ \left| \frac{1}{n} \sum_{i=1}^n (g(V_{i,n}^1, \dots, V_{i,n}^d) - g(U_i^1, \dots, U_i^d)) \right| \right] \\ &\leq \frac{1}{n} \sum_{i=1}^n \mathbb{E} [|g(V_{i,n}^1, \dots, V_{i,n}^d) - g(U_i^1, \dots, U_i^d)|]. \end{aligned}$$

Aus der Austauschbarkeit und Lemma 5.3 folgt

$$\frac{1}{n} \sum_{i=1}^n \mathbb{E} [|g(V_{i,n}^1, \dots, V_{i,n}^d) - g(U_i^1, \dots, U_i^d)|] = \mathbb{E} [|g(V_{1,n}^1, \dots, V_{1,n}^d) - g(U_1^1, \dots, U_1^d)|],$$

wodurch Gleichung (5.15) mithilfe von (5.13) gezeigt ist und damit auch die Konsistenz des LHSD-Schätzers für  $C$ -fast überall stetige Funktionen bewiesen ist.  $\square$

Während Proposition 5.9 fordert, dass die Funktion  $g$  auf dem ganzen Raum stetig ist, wird diese Voraussetzung in Proposition 5.10 abgeschwächt zu einer Stetigkeit, die nur  $C$ -fast überall erfüllt ist. Aus dieser Aussage folgt sofort mithilfe des Satzes von dominierter Konvergenz das folgende Resultat.

**Korollar 5.11.**

Sei  $g : [0, 1]^d \rightarrow \mathbb{R}$  eine beschränkte  $C$ -fast überall stetige Funktion. Dann ist der LHSD Schätzer asymptotisch erwartungstreu, d.h. es gilt

$$\mathbb{E} \left[ \frac{1}{n} \sum_{i=1}^n g(V_{i,n}^1, \dots, V_{i,n}^d) \right] \rightarrow \mathbb{E}[g(U^1, \dots, U^d)] \quad \text{für } n \rightarrow \infty. \quad (5.16)$$

## 5.2 Zentraler Grenzwertsatz für das LHSD und Varianzreduktion

Im Folgenden wird für den bivariate Fall (Dimension  $d = 2$ ) ein zentraler Grenzwertsatz für den LHSD Schätzer hergeleitet. Die Überlegungen und Beweise sind in einer Veröffentlichung von Packham und Schmidt [48] wiederzufinden. Für den multivariaten Fall bzw. beliebige Dimension haben Aistleitner, Hofer und Tichy [3] einen zentralen Grenzwertsatz inklusive Herleitung und Anwendungen publiziert, der im Anschluss angegeben aber nicht bewiesen wird.

Bevor der zentrale Grenzwertsatz für den bivariaten Fall angegeben werden kann, sind noch ein paar Definitionen und Erklärungen notwendig. Die folgende Definition der Variation im Sinne von Hardy und Krause stellt den zweidimensionalen Fall der Definition 2.15 dar und sei hier zum besseren Verständnis noch einmal explizit angeführt.

**Definition 5.12. (Variation im Sinne von Hardy und Krause für  $d = 2$ )**

Eine Funktion  $g : [0, 1]^2 \rightarrow \mathbb{R}$  ist von beschränkter Variation im Sinne von Hardy und Krause, wenn eine Konstante  $K$  existiert, sodass folgende Punkte gelten.

- Für alle beschränkten Rechtecke  $[a, b] \times [c, d] \subseteq [0, 1]^2$ , für alle  $m, n$  und Punkte  $a = x_0 < \dots < x_m = b$  und  $c = y_0 < \dots < y_n = d$  gilt die Einschränkung

$$\sum_{i=0}^{m-1} \sum_{j=0}^{n-1} |g(x_i, y_j) - g(x_i, y_{j+1}) - g(x_{i+1}, y_j) + g(x_{i+1}, y_{j+1})| \leq K. \quad (5.17)$$

- Für alle  $u \in [0, 1]$  ist die Variation der Funktion  $v \mapsto g(u, v)$  beschränkt durch  $K$ .
- Für alle  $v \in [0, 1]$  ist die Variation der Funktion  $u \mapsto g(u, v)$  beschränkt durch  $K$ .

**Definition 5.13. (Rechtsstetigkeit)**

Eine Funktion  $g : [0, 1]^2 \rightarrow \mathbb{R}$  heißt rechtsstetig, wenn

$$\lim_{n \rightarrow \infty} g(u_n^1, u_n^2) = g(u^1, u^2),$$

für jede Folge  $(u_n^1, u_n^2)_{n \geq 1}$ , wobei  $u_n^j$  von rechts gegen  $u^j$  für  $j = 1, 2$  konvergiert, d.h.  $u_n^j \downarrow u^j$ .

**Lemma 5.14.** Seien  $(X_n)_{n \geq 1}$  und  $(Y_n)_{n \geq 1}$  zwei Folgen von reellwertigen Zufallsvariablen. Wenn  $X_n \xrightarrow{d} X$  und  $|X_n - Y_n| \xrightarrow{\mathbb{P}} 0$ , dann folgt daraus die Konvergenz  $Y_n \xrightarrow{d} X$ .

Der Beweis zu diesem Lemma kann bei Kallenberg [26] gefunden werden.

**Definition 5.15. (Brown'sche Brücke)**

Eine Brown'sche Brücke  $(B_C(u, v))_{(u,v) \in [0,1]^2}$  ist ein Gauss'scher Prozess mit Erwartungswert  $\mathbb{E}[B_C(u, v)] = 0$  und Kovarianzfunktion

$$\mathbb{E}[B_C(u, v), B_C(u', v')] = C(u \wedge u', v \wedge v') - C(u, v)C(u', v'), \quad 0 \leq u, u', v, v' \leq 1.$$

Im Beweis des zentralen Grenzwertsatzes für den LHSD Schätzer wird außerdem noch ein Satz von Fermanian, Radulovic und Wegkamp [17] verwendet, der im Folgenden angegeben, aber nicht bewiesen wird.

**Satz 5.16. (Fermanian, Radulovic, Wegkamp)**

Sei  $g : [0, 1]^2 \rightarrow \mathbb{R}$  eine rechtsstetige Funktion von beschränkter Variation im Sinne von Hardy und Krause und  $C$  die Copula von  $(U, V)$  mit stetigen partiellen Ableitungen, wobei  $U$  und  $V$  gleichverteilte Zufallsvariablen in  $[0, 1]$  sind. Dann gilt

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n (g(F_n^1(U_i^1), F_n^2(U_i^2)) - \mathbb{E}[g(U^1, U^2)]) \xrightarrow{d} \int_{[0,1]^2} \mathbb{G}_C(u, v) dg(u, v), \quad (5.18)$$

wobei  $F_n^j$  die empirische Verteilungsfunktion von  $U_1^j, \dots, U_n^j$  für  $j = 1, 2$  ist und

$$\mathbb{G}_C(u, v) = \left\{ B_C(u, v) - \frac{\partial C(u, v)}{\partial u} B_C(u, 1) - \frac{\partial C(u, v)}{\partial v} C(u, v) B_C(1, v) \right\}, \quad (5.19)$$

wobei  $B_C$  eine Brown'sche Brücke auf  $[0, 1]^2$  darstellt.

Insbesondere ist die Grenzverteilung eine Normalverteilung.

**Satz 5.17. (Zentraler Grenzwertsatz für den LHSD Schätzer in Dimension 2)**

Sei  $g : [0, 1]^2 \rightarrow \mathbb{R}$  eine rechtsstetige Funktion von beschränkter Variation im Sinne von Hardy und Krause. Sei  $C$  die Copula von  $(U, V)$  mit stetigen partiellen Ableitungen, wobei  $U$  und  $V$  gleichverteilte Zufallsvariablen in  $[0, 1]$  sind. Dann gilt

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n (g(V_{i,n}^1, V_{i,n}^2) - \mathbb{E}[g(U^1, U^2)]) \xrightarrow{d} N(0, \sigma_{LHSD}^2). \quad (5.20)$$

Die Varianz  $\sigma_{LHSD}^2$  ist mit der Schreibweise  $\partial_1 C(u, v) = \partial C(u, v)/\partial u$  und  $\partial_2 C(u, v) = \partial C(u, v)/\partial v$  gegeben als

$$\begin{aligned} \sigma_{LHSD}^2 = & \iiint \iiint C(u \wedge u', v \wedge v') dg(u, v) dg(u', v') - \left( \iint C(u, v) dg(u, v) \right)^2 \\ & + \iiint \iiint \{ \partial_1 C(u', v')(C(u, v)u' - C(u \wedge u', v)) \\ & + \partial_1 C(u, v)(C(u', v')u - C(u \wedge u', v')) + \partial_2 C(u', v')(C(u, v)v' - C(u, v \wedge v')) \\ & + \partial_2 C(u, v)(C(u', v')v - C(u', v \wedge v')) + \partial_1 C(u, v)\partial_1 C(u', v')(u \wedge u' - uu') \\ & + \partial_2 C(u, v)\partial_2 C(u', v')(v \wedge v' - vv') + \partial_1 C(u, v)\partial_2 C(u', v')(C(u, v') - uv') \\ & + \partial_1 C(u', v')\partial_2 C(u, v)(C(u', v) - u'v) \} dg(u, v) dg(u', v'), \end{aligned}$$

wobei die Integrale jeweils über den Bereich  $(0, 1]$  laufen.

*Beweis.* Der Ausdruck in (5.20) kann in zwei Summanden aufgespalten werden

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n (g(V_{i,n}^1, V_{i,n}^2) - g(F_n^1(U_i^1), F_n^2(U_i^2))) + \frac{1}{\sqrt{n}} \sum_{i=1}^n (g(F_n^1(U_i^1), F_n^2(U_i^2)) - \mathbb{E}[g(U^1, U^2)]), \quad (5.21)$$

wovon nun der erste Teil durch eine Konstante abgeschätzt werden kann, während die zweite Summe genau die Anforderungen von Satz 5.16 erfüllt und gegen eine Normalverteilung konvergiert.

Aus Gründen der Übersicht, sei

$$v_i^j := V_{i,n}^j = \frac{r_{i,n}^j - 1 + \eta_{i,n}^j}{n}$$

und

$$u_i^j := F_n^j(U_i^j) = \frac{r_{i,n}^j}{n}$$

für  $j = 1, 2$ . Außerdem sei  $n$  fix gewählt und die Variation von  $g$  sei durch  $K$  beschränkt.

Dann gilt laut Definition der Variation von  $g$  die Abschätzung

$$\begin{aligned}
& \left| \sum_{i=1}^n (g(V_{i,n}^1, V_{i,n}^2) - g(F_n^1(U_i^1), F_n^2(U_i^2))) \right| = \left| \sum_{i=1}^n (g(v_i^1, v_i^2) - g(u_i^1, u_i^2)) \right| \\
&= \left| \sum_{i=1}^n (g(v_i^1, v_i^2) + g(u_i^1, u_i^2) - g(v_i^1, u_i^2) - g(u_i^1, v_i^2) - 2g(u_i^1, u_i^2) + g(v_i^1, u_i^2) + g(u_i^1, v_i^2) \right. \\
&\quad \left. - g(v_i^1, 0) + g(u_i^1, 0) + g(0, u_i^2) - g(0, v_i^2) + g(v_i^1, 0) - g(u_i^1, 0) - g(0, u_i^2) + g(0, v_i^2)) \right| \\
&\leq \sum_{i=1}^n |g(v_i^1, v_i^2) + g(u_i^1, u_i^2) - g(v_i^1, u_i^2) - g(u_i^1, v_i^2)| \\
&\quad + \sum_{i=1}^n |g(v_i^1, u_i^2) + g(u_i^1, 0) - g(u_i^1, u_i^2) - g(v_i^1, 0)| + \sum_{i=1}^n |g(u_i^1, v_i^2) + g(0, u_i^2) - g(u_i^1, u_i^2) - g(0, v_i^2)| \\
&\quad + \sum_{i=1}^n |g(v_i^1, 0) - g(u_i^1, 0)| + \sum_{i=1}^n |g(0, v_i^2) - g(0, u_i^2)| \leq 4K.
\end{aligned}$$

Damit ist der Ausdruck nach oben hin beschränkt, womit die Konvergenz

$$\frac{1}{\sqrt{n}} \left| \sum_{i=1}^n (g(V_{i,n}^1, V_{i,n}^2) - g(F_n^1(U_i^1), F_n^2(U_i^2))) \right| \longrightarrow 0, \quad \text{für } n \rightarrow \infty \quad (5.22)$$

gilt.

Anwendung von Lemma 5.14, wobei die Folgen  $(X_n)_{n \geq 1}$  und  $(Y_n)_{n \geq 1}$  durch die beiden Summanden aus (5.21) ersetzt werden und  $X$  eine Normalverteilung repräsentiert, führt zu der im Satz gegebenen Konvergenz gegen eine Normalverteilung.

Es bleibt noch die Parameter der Verteilung festzulegen. Die Erwartung folgt direkt aus Satz 5.16 als Erwartungswert von einer Summe Brown'scher Brücken, deren Erwartungswert jeweils Null ist.

Die Varianz  $\sigma_{LHSD}^2$  ist als zweites Moment der Grenzverteilung gegeben, was soviel bedeutet wie

$$\sigma_{LHSD}^2 = \mathbb{E} \left[ \int_{[0,1]^2} \mathbb{G}_C(u, v) \, dg(u, v) \right]^2. \quad (5.23)$$

Dieser Erwartungswert kann mithilfe des Satzes von Fubini berechnet werden und liefert den gewünschten Ausdruck aus dem Satz. Der Satz von Fubini darf angewendet werden, da  $g$ , als Funktion von beschränkter Variation, die Differenz von zwei monoton wachsenden Funktionen ist und damit als doppeltes Integral bezüglich zwei positiven Maßen geschrieben werden kann. Diese Aussagen wurden von Adams und Clarkson [2] bestätigt.  $\square$

Der zentrale Grenzwertsatz für das *LHSD*-Verfahren wurde von Aistleitner, Hofer und Tichy [3] auf  $d$  Dimensionen verallgemeinert, wobei hier nur das zentrale Ergebnis ohne Beweis präsentiert wird.

**Satz 5.18. (Zentraler Grenzwertsatz für den *LHSD* Schätzer in Dimension  $d$ )**

Sei  $g : [0, 1]^d \rightarrow \mathbb{R}$  eine rechtsstetige Funktion von beschränkter Variation im Sinne von Hardy-Krause. Sei  $C$  die Copula von  $(U^1, \dots, U^d)$  mit stetigen partiellen Ableitungen, mit  $U^1, \dots, U^d$  gleichverteilt auf  $[0, 1]$ . Dann gilt

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n (g(V_{i,n}^1, \dots, V_{i,n}^d) - \mathbb{E}[g(U^1, \dots, U^d)]) \xrightarrow{d} N(0, \sigma_{LHSD}^2), \quad (5.24)$$

mit

$$\sigma_{LHSD}^2 = \int_{[0,1]^{2d}} \mathbb{E} [G_C(u^1, \dots, u^d) G_C(\bar{u}^1, \dots, \bar{u}^d)] d\hat{f}(u^1, \dots, u^d) d\hat{f}(\bar{u}^1, \dots, \bar{u}^d). \quad (5.25)$$

$G_C$  ist die Bezeichnung für ein zufälliges zentriertes Gauß-Feld,

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

mit einer  $d$ -dimensionalen Brown'schen Brücke  $B_C$ . Die Funktion  $\hat{f} : [0, 1]^d \rightarrow \mathbb{R}$  ist definiert als

$$\hat{f}(u^1, \dots, u^d) = \begin{cases} 0 & \text{falls zumindest ein } u^j = 1 \quad \text{für } j = 1, \dots, d \\ f(u^1, \dots, u^d) & \text{sonst.} \end{cases}$$

Der Beweis wird sehr ähnlich jenem im zweidimensionalen Fall geführt und verwendet wiederum den Satz von Fubini sowie ein zu Fermanian äquivalentes Resultat.

### 5.3 Varianzreduktion im *LHSD* Verfahren

Die zentrale Frage, die gestellt werden muss, ist, ob mit dem *LHSD*-Verfahren die Varianz der Schätzung gegenüber dem herkömmlichen Monte Carlo Verfahren tatsächlich reduziert werden kann. Diese kann im Bezug auf Copulas unter Anwendung des ursprünglichen Zentralen Grenzwertsatzes geschrieben werden als

$$\sigma_{MC}^2 = \text{Var}(g(U, V)) = \iint g(u, v)^2 dC(u, v) - \left( \iint g(u, v) dC(u, v) \right)^2.$$

Mithilfe von partieller Integration, diversen Abschätzungen z.B. durch die Jensen Ungleichung [24] und unter einigen Voraussetzungen an die Copula, zeigen Packham und Schmidt [48], dass

die Varianz aus dem *LHSD* Verfahren zumindest gleich gut ist wie jene des Monte Carlo Verfahrens.

Eine dieser Voraussetzungen ist die Eigenschaft “*right-tail increasing*” (Nelsen [38]) die im Folgenden definiert wird.

**Definition 5.19. (Right Tail Increasing, RTI)**

Seien  $X$  und  $Y$  Zufallsvariablen. Dann ist  $Y$  *right-tail increasing in  $X$* , wenn die Abbildung

$$x \mapsto \mathbb{P}(Y > y | X > x) \quad \text{für alle } y$$

nicht monoton fallend ist.

Die Copula  $C$  der Zufallsvariablen  $X$  und  $Y$  heißt *right-tail increasing*, wenn  $X$  *right-tail increasing in  $Y$*  und  $Y$  *right-tail increasing in  $X$*  ist.

Ein Beispiel für eine *RTI* Copula, die zusätzlich auch noch stetige partielle Ableitungen besitzt, ist die bivariate Normalverteilung mit Korrelationsparameter  $\rho \in (0, 1)$ . Der Beweis dieser Eigenschaften kann bei Joe [25] gefunden werden. Das Ergebnis von Packham und Schmidt [48] umfasst nun all jene Copulas, die über solche Qualitäten verfügen und lautet:

**Proposition 5.20.**

Sei die Copula  $C$  *RTI* mit stetigen partiellen Ableitungen und  $g : [0, 1]^2 \rightarrow \mathbb{R}$  eine rechtsstetige Funktion von beschränkter Variation, die in jedem Argument nicht monoton fallend ist. Dann gilt

$$\sigma_{LHSD}^2 \leq \sigma_{MC}^2.$$

## 5.4 Anwendung auf Zufallsvektoren mit beliebiger Randverteilung

Aus gleichverteilten Zufallsvariablen auf  $[0, 1]$  können mithilfe der Inversionsmethode aus Satz 2.5 beliebig verteilte Zufallsvariablen erzeugt werden und als Folgerung daraus gilt für eine messbare Funktion  $f : \mathbb{R} \rightarrow \mathbb{R}$ , dass  $f(X)$ , für  $X = F^{\leftarrow}(U)$  eine beliebige Zufallsvariable und  $g(U)$ , für  $U$  eine gleichverteilte Zufallsvariable, die gleiche Verteilung besitzen unter der Annahme  $g := f \circ F^{\leftarrow}$ .

Eine Copula ist laut Definition 2.1 eine Verteilungsfunktion mit gleichverteilten Rändern auf  $[0, 1]$ , für welche die Eigenschaft

$$C(u_1, \dots, u_d) = F(F_1^{\leftarrow}(u_1), \dots, F_d^{\leftarrow}(u_d)). \quad (5.26)$$

aus Korollar 2.4 für  $(u_1, \dots, u_d) \in [0, 1]^d$ , eine Gesamtverteilungsfunktion  $F$  und stetige Randverteilungsfunktionen  $F_1, \dots, F_d$  gilt (Nelsen [38]).

Im Folgenden sei  $(X_1, \dots, X_d)$  ein Zufallsvektor mit beliebigen Randverteilungen  $F_1, \dots, F_d$  und Gesamtverteilungsfunktion  $F$ . Dann gilt wiederum für eine messbare Funktion  $f : \mathbb{R} \rightarrow \mathbb{R}$

$$g(X_1, \dots, X_d) \stackrel{d}{=} g(F_1^{\leftarrow}(U_1), \dots, F_d^{\leftarrow}(U_d)) := f(U_1, \dots, U_d), \quad (5.27)$$

wobei die Gesamtverteilungsfunktion von  $(U_1, \dots, U_d)$  durch die Copula  $C$  bezüglich den Verteilungsfunktionen  $F, F_1, \dots, F_d$  beschrieben wird.

Zur Anwendung des *Latin Hypercube Samplings with Dependence* müssen nun die Ränge der Zufallsvariablen ermittelt werden. Dabei kann auf folgendes Lemma zurückgegriffen werden, das eine vorherige Transformation der Ränder nicht zwingend erforderlich macht. Der Beweis ist wiederum bei Packham und Schmidt [48] zu finden.

**Lemma 5.21.** *Seien  $X_1, \dots, X_n$  iid Zufallsvariablen mit stetiger Verteilungsfunktion  $F$ . So gilt für alle  $i = 1, \dots, n$*

$$r_{i,n}(X_1, \dots, X_n) = r_{i,n}(F(X_1), \dots, F(X_n)) \quad \mathbb{P} - f.s. \quad (5.28)$$

Das bedeutet, dass keine spezielle Verteilung zur Bestimmung der Ränge notwendig ist, solange sie stetig ist, was eine deutliche Erleichterung ist, vor allem in Hinsicht auf die Rechenzeit, da eine Transformation in eine andere Verteilung durchaus einen großen Aufwand mit sich bringen kann.

Mithilfe der Ränge kann nun wiederum die Modifikation aus Gleichung (5.2) vorgenommen werden, um die entscheidenden Zufallsvariablen  $V_{i,n}^j, i = 1, \dots, n, j = 1, \dots, d$  zu erhalten.

Der *LHSD*-Schätzer von  $\mathbb{E}[f(X_1, \dots, X_d)]$  ist für den Fall einer beliebigen Verteilung gegeben durch

$$\frac{1}{n} \sum_{i=1}^n f(F_1^{\leftarrow}(V_{i,n}^1), \dots, F_d^{\leftarrow}(V_{i,n}^d)). \quad (5.29)$$

## 6 Latin Supercube Sampling

Das *Latin Supercube Sampling* ist eine Mischung aus dem *Latin Hypercube Sampling* und dem Quasi-Monte Carlo Verfahren. Dort, wo im *Latin Hypercube Sampling* die Reihenfolge der zufällig verteilten Punkte im Intervall  $[0, 1]$  geändert wird, werden im *Latin Supercube Sampling* die deterministischen Folgenglieder der Quasi-Monte Carlo Methode permutiert. Die Ideen, Behauptungen und Beweise dazu stammen zum größten Teil von Owen [46].

Der erste Schritt besteht darin, die Simulationsvariablen in  $k$  nicht leere Teilmengen  $\mathcal{A}_r$  mit  $\mathcal{A}_r \cap \mathcal{A}_{r'} = \emptyset$  für  $r \neq r'$  zu unterteilen.  $\mathcal{A} = \sum_{r=1}^k \mathcal{A}_r$ . Die Dimension jeder Teilmenge sei  $s_r = |\mathcal{A}_r| \geq 1$  mit  $\sum_{r=1}^k s_r = d$  und die Punkte  $\mathcal{X}_i^r \in [0, 1]^{\mathcal{A}_r}$ ,  $i = 1, \dots, n$ ,  $r = 1, \dots, k$  seien das Resultat einer Quasi-Monte Carlo Methode in Dimension  $s_r$ .  $\pi_r(i)$ ,  $r = 1, \dots, k$  seien voneinander unabhängige Permutationen der Zahlen  $1, \dots, n$ .

Ein *Latin Supercube Sampling* ist dann von der Form

$$\mathbf{X}_i = (\mathcal{X}_{\pi_1(i)}^1, \dots, \mathcal{X}_{\pi_k(i)}^k), \quad i = 1, \dots, n, \quad (6.1)$$

wobei  $\mathcal{X}_{\pi_r(i)}^r$  die permutierten Quasi-Monte Carlo Punkte  $\mathcal{X}_i^r$  sind, die nicht zwangsweise in der Reihenfolge wie in (6.1) auftreten müssen. Diese Annahme wird nur verwendet, um die Notation möglichst einfach zu halten. Kurz gesagt bedeutet (6.1), dass die ersten  $s_1$  Spalten im *Latin Supercube Sampling* von den zufällig gemischten Punkten  $\mathcal{X}_1^1, \dots, \mathcal{X}_n^1$  gefüllt werden, die nächsten  $s_2$  Spalten von den Punkten  $\mathcal{X}_i^2$ ,  $i = 1, \dots, n$  und so weiter.

Der *Latin Supercube Schätzer* des Integrals  $I$  aus Gleichung (2.6) ist mit den Punkten  $\mathbf{X}_i$  aus (6.1) gegeben als

$$\hat{I}_{LSS} = \frac{1}{n} \sum_{i=1}^n g(\mathbf{X}_i). \quad (6.2)$$

### 6.1 Varianzreduktion im Latin Supercube Sampling

Als erstes stellt sich nun wieder die Frage nach dem Fehler  $\hat{I}_{LSS} - I$ . Wie sehen des Weiteren Erwartungswert und Varianz des Schätzers aus? Kann die Varianz des *Latin Supercube Sampling* mit der aus dem klassischen Monte Carlo Verfahren resultierende Varianz  $\sigma^2$  mithalten?

Zunächst wird der Integrationsfehler in mehrere Teile zerlegt,

$$\hat{I}_{LSS} - I = (\hat{I}_{LSS} - I_G) + (I_G - I) \quad (6.3)$$

mit

$$I_G = \frac{1}{n^k} \sum_{i_1=1}^n \cdots \sum_{i_k=1}^n g(\mathcal{X}_{i_1}^1, \dots, \mathcal{X}_{i_k}^k) \quad (6.4)$$

als Mittelwert von  $g$  über alle  $n^k$  Integrationspunkte die im *Latin Supercube Sampling* erzeugt werden. Weiters seien  $\sigma_G^2$  und  $\sigma_{G(r)}^2$  definiert als

$$\sigma_G^2 = \frac{1}{n^k} \sum_{i_1=1}^n \cdots \sum_{i_k=1}^n (g(\mathcal{X}_{i_1}^1, \dots, \mathcal{X}_{i_k}^k) - I_G)^2 \quad (6.5)$$

und

$$\sigma_{G(r)}^2 = \frac{1}{n} \sum_{i_r=1}^n \left( \frac{1}{n^{k-1}} \sum_{i_1=1}^n \cdots \sum_{i_{r-1}=1}^n \sum_{i_{r+1}=1}^n \cdots \sum_{i_k=1}^n g(\mathcal{X}_{i_1}^1, \dots, \mathcal{X}_{i_k}^k) - I_G \right)^2. \quad (6.6)$$

**Lemma 6.1.**

*Im Latin Supercube Sampling unter Verwendung deterministischer Quasi-Monte Carlo Punktfolgen gilt*

$$\mathbb{E}[\hat{I}_{LSS} - I_G] = 0. \quad (6.7)$$

Außerdem gilt für  $n > 1$

$$\begin{aligned} \text{Var}[\hat{I}_{LSS} - I_G] &\leq \frac{1}{n-1} \left( \sigma_G^2 - \sum_{r=1}^k \sigma_{G(r)}^2 \right) \\ \text{Var}[\hat{I}_{LSS} - I_G] &\geq \frac{1}{n} \left( 1 - \frac{1}{(n-1)^2} \right) \left( \sigma_G^2 - \sum_{r=1}^k \sigma_{G(r)}^2 \right). \end{aligned}$$

Der Beweis verwendet die Tatsache, dass das LHS, das zentrierte LHS und das LSS für jede Permutation  $\pi_1, \dots, \pi_k$  von  $1, \dots, n$  den selben Funktionswert  $g(\mathbf{X}_i) = h(\mathbf{Z}_i)$  liefern, wobei  $h$  eine auf dem Raum  $[0, 1]^k$  definierte Funktion darstellt mit

$$h(z_1, \dots, z_k) = g(\mathcal{X}_{1+\lfloor z_1 n \rfloor}^1, \dots, \mathcal{X}_{1+\lfloor z_k n \rfloor}^k).$$

Nachdem das zentrierte LHS der Funktion  $h$  erwartungstreu ist, ergibt sich daraus die erste Behauptung aus dem Lemma.

Die Aussagen für die Varianz  $\text{Var}[\hat{I}_{LSS} - I_G]$  können ebenso aus der Varianz des zentrierten LHS hergeleitet werden, wobei  $\sigma_G^2$  die Varianz von  $h$  und  $\sigma_{G(r)}^2$  die diskrete Version der Varianz  $\sigma_{\{r\}}$  der Haupteffekte (vgl. Kapitel 4 Gleichung (4.5)) auf die Funktion  $h$  darstellt.

Sei nun

$$\varepsilon_n^r = \varepsilon_n^r(g) = \sup_{\mathbf{X} \in \mathcal{X}^{-A^r}} \left| \int g(\mathbf{X}) d\mathcal{X}^r - \frac{1}{n} \sum_{i=1}^n g(\mathcal{X}^1, \dots, \mathcal{X}_i^r, \dots, \mathcal{X}^k) \right| \quad (6.8)$$

der maximale Fehler, der bei Integration von  $g$  über  $\mathcal{X}^r$  passieren kann, wenn alle anderen Komponenten fixiert sind. Mithilfe dieser Größe kann folgende Aussage getroffen werden.

**Proposition 6.2.**

Es gilt die Abschätzung

$$|I_G - I| \leq \sum_{r=1}^k \varepsilon_n^r,$$

wobei mit  $I$  das Integral aus Gleichung (2.1) bezeichnet wird.

*Beweis.*

Es gilt

$$\begin{aligned} I_G &\leq \frac{1}{n^{k-1}} \sum_{i_1=1}^n \cdots \sum_{i_{k-1}=1}^n \int_{[0,1)^{\mathcal{A}_r}} g(\mathcal{X}_{i_1}^1, \dots, \mathcal{X}_{i_{k-1}}^{k-1}, \mathcal{X}^k) d\mathcal{X}^k + \varepsilon_n^k \\ &= \int_{[0,1)^{\mathcal{A}_r}} \left( \frac{1}{n^{k-1}} \sum_{i_1=1}^n \cdots \sum_{i_{k-1}=1}^n g(\mathcal{X}_{i_1}^1, \dots, \mathcal{X}_{i_{k-1}}^{k-1}, \mathcal{X}^k) \right) d\mathcal{X}^k + \varepsilon_n^k \end{aligned}$$

laut Definition von  $\varepsilon_n^k$ . Wiederholt man dies rekursiv für  $\mathcal{X}^r$ ,  $r = 1, \dots, k$  so resultiert

$$I_G \leq I + \sum_{r=1}^k \varepsilon_n^r.$$

Der Integrationsfehler des Produkts einer mehrdimensionalen Integration ist hier die Summe der Fehler pro Integration. Auf ähnliche Weise kann auch

$$I_G \geq I - \sum_{r=1}^k \varepsilon_n^r$$

gezeigt werden. □

Der maximale Fehler  $\varepsilon_n^r$  darf nicht größer als  $O(n^{-1/2})$  sein, wenn das Quasi-Monte Carlo Verfahren mit dem klassischen Monte Carlo Verfahren vergleichbar sein soll. Genau deshalb ist dies als Voraussetzung im nächsten Satz zu finden.

**Satz 6.3.**

Im Latin Supercube Sampling gilt, falls  $\varepsilon_n^r(g) = O(n^{-1/2})$

$$\mathbb{E}[\hat{I}_{LSS}] = I + O(n^{-1/2}). \quad (6.9)$$

Falls zusätzlich  $\varepsilon_n^r(g_u, g_v) = O(n^{-1/2})$  für alle  $u, v \subseteq \mathcal{A}$ , so gilt

$$\text{Var}[\hat{I}_{LSS}] = \frac{1}{n} \left( \sigma_g^2 - \sum_{r=1}^k \sum_{u \subseteq \mathcal{A}_r} \sigma_u^2 + O(n^{-1/2}) \right). \quad (6.10)$$

*Beweis.*

Das erste Ergebnis des Satzes setzt sich direkt aus Gleichung (6.3), Lemma 6.1 und Proposition 6.2 zusammen.

Im Fall von  $\varepsilon_n^r(g_u g_v) = O(n^{-1/2})$  für alle  $u, v \subseteq \mathcal{A}$ , gilt ebenso  $\varepsilon_n^r(g^2) = O(n^{-1/2})$ , womit aus Gleichung (6.5)

$$\sigma_G^2 = \sigma_g^2 + O(n^{-1/2})$$

folgt. Ähnlich folgt aus Gleichung (6.6) und der Tatsache, dass

$$\int_{[0,1]^{-\mathcal{A}_r}} g(\mathbf{X}) dX^{-\mathcal{A}_r} = \sum_{u \subseteq \mathcal{A}_r} g_u(\mathbf{X}),$$

das Ergebnis

$$\begin{aligned} \sigma_{G(r)}^2 &= \frac{1}{n} \sum_{i_r=1}^n \left( \int_{[0,1]^{-\mathcal{A}_r}} g(\mathcal{X}^1, \dots, \mathcal{X}_{i_r}, \dots, \mathcal{X}^k) dX^{-\mathcal{A}_r} - I_G \right)^2 + O(n^{-1/2}) \\ &= \frac{1}{n} \sum_{i_r=1}^n \left( \sum_{0 < |u|, u \subseteq \mathcal{A}_r} g_u(\mathcal{X}_{i_r}^r) \right)^2 + O(n^{-1/2}) \\ &= \sum_{u \subseteq \mathcal{A}_r} \sigma_u^2 + O(n^{-1/2}). \end{aligned}$$

Diese Größe kann nun in die Varianzabschätzungen von Lemma 6.1 eingesetzt werden. Es resultiert

$$\text{Var}[\hat{I}_{LSS}] = \frac{1}{n} \left( \sigma_g^2 - \sum_{r=1}^k \sum_{u \subseteq \mathcal{A}_r} \sigma_u^2 + O(n^{-1/2}) \right),$$

was gewünscht war. □

Aus der Varianz des *Latin Supercube* Schätzers ist ersichtlich, dass wie schon im *Latin Hypercube Sampling* alle zu den Haupteffekten gehörigen Varianzen, d.h. sämtliche Komponenten  $\sigma_u^2$  mit  $|u| = 1$  gelöscht werden und nicht zur Schätzvarianz beitragen. Zusätzlich werden in der Varianz des *Latin Supercube Samplings* auch noch Komponenten höherer Ordnung gelöscht, nämlich all jene, die  $u \subseteq \mathcal{A}_r$  entsprechen.

Im Paper von Owen [46] sind noch weitere Ausführungen, insbesondere in Hinblick auf Unendlichkeit der Dimension, sowohl im Falle von *Latin Supercube* als auch *Latin Hypercube* zu finden, sowie Überlegungen zu einer ANOVA im Falle von  $d = \infty$ .

Der Vorteil bei Verwendung des *Latin Supercube Samplings* liegt darin, dass die Dimension des Problems in kleinere Dimensionen aufgespalten wird. Wird zum Beispiel eine gleichverteilte

Folge von Dimension 2000 verlangt, so können in diesem Verfahren genauso gut 20 Folgen von Dimension 100 generiert werden und durch Verwendung von Permutationen zu einer 2000-dimensionalen verknüpft werden. Dadurch kann von der Eigenschaft profitiert werden, dass Quasi-Monte Carlo Folgen in niedriger Dimension besser gleichverteilt sind, als höherdimensionale.

Die Frage, die noch bleibt, ist, welche Quasi-Monte Carlo Folgen verwendet werden sollten bzw. ob es Unterschiede diesbezüglich gibt. In der Praxis werden hybride Verfahren den klassischen Quasi-Monte Carlo Verfahren vorgezogen, da, wie bereits in Kapitel 2.2.3 erwähnt, hier mehrere unabhängige Schätzer berechnet werden können, während bei Quasi-Monte Carlo immer von der gleichen Folge ausgegangen wird. Unterschiedliche Payoffs resultieren einzig und allein durch die Verwendung verschiedener Permutationen. Dadurch wird die Varianz verfälscht. Sie ist viel kleiner als bei hybriden Verfahren, wo wirklich unabhängige Folgen generiert werden können. Dadurch kann bei reinen Quasi-Monte Carlo Verfahren über die Genauigkeit des *Latin Supercube Samplings* keine repräsentative Aussage getätigt werden.

Das *Latin Hypercube Sampling* stellt einen Spezialfall des *Latin Supercube Samplings* dar. Dazu muss  $k = d$  gewählt werden und die Dimension der einzelnen Teilmengen auf  $s_r = 1$  für alle  $r$  festgesetzt werden. Unter Verwendung der Punkte

$$\mathcal{X}^r = \left( u_{1,r}, \frac{1}{n} + u_{2,r}, \dots, \frac{n-1}{n} + u_{n,r} \right)$$

mit unabhängigen und in  $[0, \frac{1}{n})$  gleichverteilten Zufallsvariablen, ergibt sich daraus das *Latin Hypercube Sampling*.

## 6.2 Effektive Dimension

In Abschnitt 2.2.4 wurden bereits Lösungsvorschläge angesprochen, wie das Dimensionsproblem der Quasi-Monte Carlo Folgen umgangen bzw. gemindert werden kann. In diesem Zusammenhang taucht in der Literatur auch immer wieder der Begriff der “effektiven” Dimension auf.

Die Definition der effektiven Dimension baut auf die Überlegungen der ANOVA-Zerlegung aus Abschnitt 4.1 auf. Die Autoren Caflisch et al. [9] geben zwei verschiedene Begriffe in diesem Kontext an.

### Definition 6.4. (Effektive Dimension im Sinne der Überdeckung)

Die effektive Dimension einer quadratisch integrierbaren Funktion  $g$  im Sinne der Überdeckung (*effective dimension in the superposition sense*) ist gegeben als die kleinste natürliche Zahl  $d_S$ , sodass

$$\sum_{|u| \leq d_S} \sigma_u^2 \geq \alpha \sigma_g^2, \quad (6.11)$$

wobei  $\sigma_u^2$  wie in (4.7) definiert ist und  $\sigma_g^2$  die Varianz des herkömmlichen Monte Carlo Schätzers darstellt. Die Konstante  $\alpha$  wird aus dem Bereich  $(0, 1]$  gewählt.

**Definition 6.5. (Gestutzte Effektive Dimension)**

Die gestutzte effektive Dimension (effective dimension in the truncation sense) einer quadratisch integrierbaren Funktion  $g$  ist gegeben als die kleinste natürliche Zahl  $d_T$ , sodass

$$\sum_{u \subseteq \{1, 2, \dots, d_T\}} \sigma_u^2 \geq \alpha \sigma_g^2. \quad (6.12)$$

Die Konstante  $\alpha$  wird aus dem Bereich  $(0, 1]$  gewählt.

Die Konstante bei Caffisch et al. wird mit  $\alpha = 0.99$  festgesetzt. Diese Wahl kann aber durch jede beliebige Zahl im Bereich  $(0, 1]$  ersetzt werden.

Die erste Definition ist ein Maß dafür, in wieviel Dimensionen die angewandte Methode gut funktionieren sollte bzw. durch Funktionen welcher Dimensionsgröße die Varianz zum größten Teil bestimmt ist. Eine effektive Dimension von  $d_S = 2$  bedeutet zum Beispiel, dass die Variabilität größtenteils durch ein- und zweidimensionale Funktionen bestimmt wird. Die Dimension  $d_S$  hängt nicht von der Ordnung der Argumente ab.

Die gestutzte effektive Dimension gibt an, dass die ersten  $d_T$  Variablen relevant sind und dass das verwendete Verfahren in diesen ersten Komponenten möglichst genau sein sollte, um im Punkt Variabilität Verbesserungen zu erlangen.

Um noch einmal auf die Konstante  $\alpha$  zurückzukommen, kann mithilfe dieser Definitionen auch bestimmt werden, wie “ $d$ -dimensional” die Funktion ist, wie es bei Caffisch et al. ausgedrückt wird. Dazu wird der Quotient

$$\frac{\sum_{0 < |u| \leq d} \sigma_u^2}{\sigma_g^2} \quad \text{oder} \quad \frac{\sum_{u \subseteq \{1, 2, \dots, d\}} \sigma_u^2}{\sigma_g^2}$$

berechnet, wobei zu erwarten ist, dass der Quotient im Sinne der Überdeckung größer ist, als jener im gestutzten Fall.

In Abschnitt 7.3 wird das Verfahren auf das Beispiel der Asiatischen Basket Option angewandt und es kann gezeigt werden, dass es in Kombination mit Methoden, die die effektive Dimension verringern bzw. die mehr Gewicht in die ersten Komponenten legen, eine viel geringere Varianz aufweist. Solche Methoden werden in Abschnitt 7.1.4 genauer erläutert.

## 7 Anwendungen in der Finanzmathematik

### 7.1 Grundlagen

In Hinblick auf die anschließenden Anwendungsbeispiele, möchte ich vorher noch einige vorbereitende Definitionen und Sätze anführen, die für eine Bepreisung finanzmathematischer Produkte unerlässlich sind. Die Werke von Cont und Tankov [10], Karatzas und Shreve [27] sowie Schoutens [56] zählen zur Standardliteratur im Bezug auf finanzmathematische Themen.

#### 7.1.1 Optionen

Betrachtet werden im weiteren Verlauf hauptsächlich Optionen auf Aktien, die dem Käufer das Recht aber nicht die Pflicht geben, eine bestimmte Aktie zu einem vorher festgelegtem Zeitpunkt, dem Ausübungszeitpunkt  $T$  bzw. in einer festgelegten Zeitspanne zu einem fixen Preis, dem Ausübungspreis  $K$ , zu kaufen bzw. zu verkaufen.

Die wohl einfachsten und beliebtesten Typen von Optionen sind die sogenannten *Vanilla Optionen*. Dazu gehören die Europäischen Optionen, deren Ausübung immer zum Zeitpunkt  $T$  erfolgt, und die Amerikanischen Optionen, die während eines fixierten Zeitraumes ausgeübt werden können.

Zu unterscheiden sind des Weiteren Call- und Put-Optionen. Eine Call-Option ist das Recht zu kaufen, die Put-Option das Recht zu verkaufen. Im Falle der Europäischen Option ist der Payoff, d.h. der Wert der Option zum Ausübungszeitpunkt  $T$  für die beiden Typen gegeben als

$$C_T = \max[S(T) - K, 0] = (S(T) - K)^+$$

für die Call-Option und

$$P_T = \max[K - S(T), 0] = (K - S(T))^+$$

für die Put-Option, wobei  $S(T)$  der Preis der Aktie zum Zeitpunkt  $T$  ist, auf welche sich die Option bezieht.

Optionen, die nicht unter den Begriff *Vanilla Option* fallen, werden weitläufig als *Exotische Optionen* bezeichnet. Davon interessieren hier vor allem die Asiatischen Optionen, deren Payoff als Mittelwert aller Aktienpreise einer bestimmten Periode zu berechnen ist und somit eine pfadabhängige Option darstellt. Es werden wiederum zwei verschiedene Typen von Asiatischen Optionen unterschieden, die arithmetische und die geometrische Asiatische Option, wobei diese wiederum unterteilt werden bezüglich diskreter oder kontinuierlicher Mittelwertberechnung.

Der Payoff der arithmetischen Call-Option berechnet sich im diskreten bzw. kontinuierlichen Fall durch

$$V(S, T) = \left( \frac{1}{m} \sum_{i=1}^m S(t_i) - K \right)^+ \quad \text{bzw.} \quad V(S, T) = \left( \frac{1}{T} \int_0^T S(t) dt - K \right)^+. \quad (7.1)$$

Hingegen ist der Payoff der geometrischen Asiatischen Call-Option im diskreten bzw. kontinuierlichen Fall definiert als

$$V(S, T) = \left( \left( \prod_{i=1}^m S(t_i) \right)^{1/m} - K \right)^+ \quad \text{bzw.} \quad V(S, T) = \left( e^{\frac{1}{T} \int_0^T S(t) dt} - K \right)^+.$$

In Abschnitt 7.3 wird der Payoff einer Asiatischen Basket Option berechnet. Bei einer Basket Option dient der Preisfindung nicht ein einzelner Basiswert, sondern ein "Basket", sprich mehrere davon. So ergibt sich der Payoff der Basket Call-Option für  $k$  Basiswerte als

$$V(S, T) = \left( \sum_{i=1}^k a_i S_i(T) - K \right)^+ \quad (7.2)$$

mit Gewichten  $a_i$ .

Als Zusammensetzung aus der arithmetischen Asiatischen und der Basket Option ergibt sich schließlich der Payoff der Asiatischen Basket Call-Option zum Endzeitpunkt  $T$  als

$$V(S, T) = \left( \sum_{i=1}^m \sum_{j=1}^n \omega_{ij} S_i(t_j) - K \right)^+, \quad (7.3)$$

für Zeitpunkte  $t_1 < t_2 < \dots < t_n = T$  und Koeffizienten  $\omega_{ij}$  mit  $\sum_{i,j} \omega_{ij} = 1$ . Optionen solcher Art werden auch gewichtete arithmetische Asiatische Optionen genannt.

### 7.1.2 Modellannahmen

Betrachtet wird ein Markt, wo an jedem Zeitpunkt  $0 \leq t \leq T$  gehandelt werden kann. Er besteht aus  $d + 1$  Finanzgütern wobei  $S_t^j$  den Preis des  $j$ -ten Finanzgutes zum Zeitpunkt  $t$  bezeichnet.

#### Definition 7.1. (Marktmodell in stetiger Zeit)

Ein stetiges Marktmodell setzt sich aus mehreren Komponenten zusammen:

1. Einem Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, \mathbb{P})$ .
2. Einer Filtration  $\mathcal{F} = (\mathcal{F}_t)_{0 \leq t \leq T}$ , welche die gesamte Information enthält.  $\mathcal{F}$  sei immer rechtsstetig und  $\mathcal{F}_0$  enthalte alle  $\mathbb{P}$ -Nullmengen.

3. Einem adaptierten  $(d + 1)$ -dimensionalen Prozess, dem sogenannten Preisprozess  $S = (S_t)_{0 \leq t \leq T}$ , wobei

$$S_t = (S_t^0, \dots, S_t^d).$$

$S^i$  für  $i = 1, \dots, d$  seien Semimartingale,  $S^0 > 0$   $\mathbb{P}$ -f.s..

Um ein Semimartingal definieren zu können, müssen vorher noch die grundlegenden Begriffe der Martingaltheorie erläutert werden.

**Definition 7.2. (Martingal)**

Ein Martingal ist ein reellwertiger, an die Filtration  $\mathcal{F}_t$  adaptierter Prozess  $X = (X_t)_{0 \leq t \leq T}$  mit  $\mathbb{E}|X_t| < \infty$  für  $0 \leq t \leq T$ , für den gilt

$$\mathbb{E}(X_t | \mathcal{F}_s) = X_s \quad \mathbb{P}\text{-f.s.}$$

für  $0 \leq s \leq t \leq T$ .

$X = (X_t)_{0 \leq t \leq T}$  ist ein Supermartingal, wenn  $\mathbb{E}(X_t | \mathcal{F}_s) \leq X_s$   $\mathbb{P}$ -f.s. für  $0 \leq s \leq t \leq T$ .

$X = (X_t)_{0 \leq t \leq T}$  ist ein Submartingal, wenn  $\mathbb{E}(X_t | \mathcal{F}_s) \geq X_s$   $\mathbb{P}$ -f.s. für  $0 \leq s \leq t \leq T$ .

**Definition 7.3. (lokales Martingal und Stoppzeiten)**

Ein reellwertiger stochastischer Prozess  $X$  wird lokales Martingal genannt, wenn eine wachsende Folge von Stoppzeiten  $T_n$  mit  $T_n \uparrow \infty$   $\mathbb{P}$ -f.s. existiert, sodass die gestoppten Prozesse  $X^{T_n}$  Martingale sind.

Eine Stoppzeit ist eine Zufallsvariable  $T : \Omega \rightarrow [0, \infty]$  für die

$$\{T \leq t\} \in \mathcal{F}_t, \quad \forall t \geq 0.$$

**Definition 7.4. (Semimartingal)**

Semimartingale sind reellwertige stochastische Prozesse  $X = (X_t)_{t \geq 0}$  für die gilt

- $X$  ist an die vollständige und rechtsstetige Filtration  $(\mathcal{F}_t)_{t \geq 0}$  adaptiert;
- die Pfade von  $X$  sind càdlàg (rechtsstetig mit linksseitigem Limes);
- $X$  ist die Summe eines lokalen Martingals und eines Prozesses von beschränkter Variation

Ein Prozess von beschränkter Variation ist ein adaptierter Prozess  $(X_t)_{t \geq 0}$  dessen Pfade  $\mathbb{P}$ -f.s. càdlàg sind und von beschränkter Variation auf kompakten Mengen.

**Definition 7.5. (beschränkte Variation)**

Sei  $g : [a, b] \rightarrow \mathbb{R}$  eine Funktion,  $\pi_n : a = t_0 < t_1 < \dots < t_n = b$  eine endliche Zerlegung des Intervalls  $[a, b]$ . Dann ist die Variation von  $g$  auf  $[a, b]$  gegeben als

$$V(g)_{[a,b]} := \sup_{\pi_n} \sum_{k=1}^n |g(t_k) - g(t_{k-1})|.$$

Die Funktion  $g : [0, \infty) \rightarrow \mathbb{R}$  ist von beschränkter Variation, wenn  $V(g)_{[0,t]} < \infty, \forall t \geq 0$ .

Der Preis des ersten Finanzgutes  $S^0$  wird als Vergleichsgröße genutzt, weshalb vorausgesetzt wird, dass  $S_t^0$   $\mathbb{P}$ -f.s. strikt positiv ist. Der diskontierte Preisprozess  $\tilde{S} = (\tilde{S}_t)_{0 \leq t \leq T}$  ist gegeben durch

$$\tilde{S}_t = (1, \tilde{S}_t^1, \dots, \tilde{S}_t^d) \quad \text{mit} \quad \tilde{S}_t^j = \frac{S_t^j}{S_t^0}, \quad j = 1, \dots, d$$

Eine sehr wichtige Voraussetzung an den Markt ist die Arbitragefreiheit. Besteht eine Arbitragemöglichkeit, so kann durch geschicktes Handeln risikoloses Gewinn gemacht werden. Um eine Arbitrage im finanzmathematischen Sinn erklären zu können, sind vorher noch andere Definitionen, wie zum Beispiel die einer Handelsstrategie, nötig.

**Definition 7.6. (Handelsstrategie)**

Eine Handelsstrategie  $X = (X_t)_{0 \leq t \leq T}$  ist ein vorhersagbarer und damit  $\mathcal{F}_{t-1}$  messbarer  $(d+1)$ -dimensionaler Prozess mit

$$X_t = (X_t^0, \dots, X_t^d)$$

und Wertprozess  $V(X) = (V_t(X))_{0 \leq t \leq T}$  wobei  $V_t(X) = X_t \cdot S_t$  mit

$$V_t(X) := X_t \cdot S_t = \sum_{i=0}^d X_t^i S_t^i.$$

Im Zusammenhang mit der Handelsstrategie sind zwei zusätzliche Eigenschaften von großer Wichtigkeit, die im Folgenden erklärt werden. Die Rede ist von Selbstfinanzierung und Zulässigkeit einer Handelsstrategie.

**Definition 7.7. (Selbstfinanzierende Handelsstrategie)**

$X = (X_t)_{0 \leq t \leq T}$  wird als selbstfinanzierende Handelsstrategie bezeichnet, wenn  $X$   $S$ -integrierbar ist und für  $0 \leq t \leq T$

$$\int_0^t X \cdot dS = V_t(X) - V_0(X), \quad (7.4)$$

oder äquivalent in Differentialschreibweise

$$dV(X) = X \cdot dS \quad \text{bzw.} \quad d\tilde{V}(X) = X \cdot d\tilde{S}. \quad (7.5)$$

$$\int X \cdot dS := \sum_{i=0}^d \int X^i dS^i$$

heißt Vermögenszuwachsprozess von  $X$ .

**Definition 7.8. (Zulässige Handelsstrategie)**

$X = (X_t)_{0 \leq t \leq T}$  wird als zulässige Handelsstrategie bezeichnet, wenn  $X$  selbstfinanzierend ist und der diskontierte Wertprozess  $\tilde{V}_t(X)$  für  $0 \leq t \leq T$  nach unten durch eine Konstante  $K$  beschränkt ist, d.h. es gilt  $\mathbb{P}$ -f.s.  $\tilde{V}_t(X) \geq -K$ .

**Definition 7.9. (Arbitrage)**

$X = (X_t)_{0 \leq t \leq T}$  wird als Arbitrage bezeichnet, wenn  $X$  zulässig ist und

$$V_0(X) \leq 0, \quad V_t(X) \geq 0, \quad \mathbb{P}(V_T(X) > 0) > 0 \quad (7.6)$$

gilt. Ist in einem Modell keine Arbitrage zu finden, heißt es arbitragefrei.

Der echte Markt ist nie vollkommen arbitragefrei, allerdings treten Arbitragemöglichkeiten immer nur für sehr kurze Zeit auf, in der sich die Preise wieder einander angleichen. Eng verknüpft mit der Arbitragetheorie ist die Theorie der Martingalmaße.

**Definition 7.10. (äquivalentes Martingalmaß)**

Ist der diskontierte Preisprozess  $\tilde{S}_t = e^{-rt}S_t$  ein Martingal unter dem zu  $\mathbb{P}$  äquivalenten Maß  $\mathbb{Q}$ , d.h. es gilt

$$\mathbb{E}_{\mathbb{Q}}[S_T | \mathcal{F}_t] = \tilde{S}_t$$

für alle  $0 \leq t \leq T$ , so heißt  $\mathbb{Q}$  Martingalmaß.

Zwei Maße  $\mathbb{P}$  und  $\mathbb{Q}$  sind äquivalent zueinander, wenn  $\mathbb{P}$  absolut stetig bezüglich  $\mathbb{Q}$  (d.h.  $\forall N \in \mathcal{A}$  gilt: wenn  $\mathbb{Q}(N) = 0$  so auch  $\mathbb{P}(N) = 0$ ) und  $\mathbb{Q}$  absolut stetig bezüglich  $\mathbb{P}$  ist.

**Satz 7.11. (1. Fundamentalsatz der Preistheorie)**

Ein Marktmodell ist genau dann arbitragefrei, wenn ein zu  $\mathbb{P}$  äquivalentes Wahrscheinlichkeitsmaß  $\mathbb{Q}$  existiert.

Die erste große Anforderung an das Marktmodell ist also die Arbitragefreiheit. Eine weitere Eigenschaft, die wichtig ist zur Bepreisung von Optionen ist die Existenz von genau einem äquivalenten Martingalmaß. Dieses Kriterium fällt unter den Begriff "Vollständigkeit" des Marktes.

**Definition 7.12. (Absicherbare Forderung)**

Eine zum Zeitpunkt  $T$  fällige Forderung ist eine  $\mathcal{F}_T$ -messbare Zufallsvariable  $H_T \geq 0$ .

Eine Forderung heißt absicherbar, wenn eine selbstfinanzierende, zulässige Handelsstrategie  $X$  (Hedge) existiert, sodass  $V_T(X) = H_T$ .

**Definition 7.13. (Vollständiges Marktmodell)**

Ein Marktmodell heißt vollständig, wenn jede Forderung  $H_T$  mit beschränkter diskontierter Auszahlung durch eine Hedge  $X$  absicherbar ist.

**Satz 7.14. (2. Fundamentalsatz der Preistheorie)**

Ein Marktmodell ist genau dann vollständig, wenn es genau ein äquivalentes Martingalmaß bezüglich  $\mathcal{F}_T$  gibt.

Mithilfe des äquivalenten Martingalmaßes kann eine risikoneutrale Bewertungsformel für Forderungen mit Fälligkeit  $T$  angegeben werden.

**Satz 7.15. (Risikoneutraler Preis einer Forderung)**

Sei  $\mathbb{Q}$  ein äquivalentes Martingalmaß in einem arbitragefreien Marktmodell,  $H_T \geq 0$  eine Forderung. Wenn eine perfekte Hedge  $X$  existiert, deren diskontierter Wertprozess  $\tilde{V}(X)$  ein Martingal bezüglich  $\mathbb{Q}$  ist, so gilt als fairer Preis von  $H$  für  $t \leq T$

$$\tilde{\pi}_t(H_T) = \mathbb{E}_{\mathbb{Q}}[\tilde{H}_T | \mathcal{F}_t]. \quad (7.7)$$

**7.1.3 Black-Scholes Modell**

Um den Preis einer Option berechnen zu können, wird ein Modell benötigt, das den Verlauf des Aktienkurses  $(S_t)_{t \geq 0}$  nachbildet. Das geläufigste ist das Black-Scholes Modell, das den Kurs mithilfe einer geometrischen Brown'schen Bewegung modelliert.

**Definition 7.16. (Brown'sche Bewegung)**

Die Brown'sche Bewegung ist ein reellwertiger Prozess  $B = (B_t)_{t \geq 0}$  bezüglich einer Filtration  $(\mathcal{F}_t)_{t \geq 0}$ , der an die Filtration adaptiert ist und für den gilt

- $B_0 = 0$ ;
- für  $0 \leq s < t$  ist der Zuwachs  $B_t - B_s$  unabhängig von  $\mathcal{F}_s$  und hat die Verteilung  $N(0, t - s)$ ;
- die Pfade von  $B$  sind stetig.

Auf welche Weise Pfade einer Brown'schen Bewegung generiert werden können und welche Vorteile die verschiedenen Methoden in Bezug auf die Quasi-Monte Carlo Simulation haben, können in Abschnitt 7.1.4 nachgelesen werden.

Die geometrische Brown'sche Bewegung ist definiert über die stochastische Differentialgleichung

$$dS_t = S_t(\mu dt + \sigma dB_t),$$

deren Lösung durch Anwendung der Itô Formel eindeutig gegeben ist durch

$$S_t = S_0 e^{(\mu - \sigma^2/2)t + \sigma B_t}.$$

Der Parameter  $\mu \in \mathbb{R}$  beschreibt den Drift,  $\sigma > 0$  die Volatilität des Prozesses.

Unter Verwendung des stochastischen Integrals nach Itô ist jedes Integral bezüglich einem (lokalen) Martingal wieder ein lokales Martingal. Dass die Brown'sche Bewegung ein Martingal ist, kann leicht nachgerechnet werden. Damit ist  $S$ , als Summe eines Integrals nach einer Brown'schen Bewegung (=lokales Martingal) und einem Prozess von beschränkter Variation, ein Semimartingal. Nun wäre geklärt, welche Annahmen für den Verlauf des Aktienkurses getätigt werden.

Wichtig ist außerdem noch der Rahmen in dem der Preis berechnet werden soll, d.h. wie der Markt bzw. das Marktmodell konkret aussehen soll. Zusätzlich zur Arbitragefreiheit werden noch weitere Voraussetzungen an den Markt getroffen:

- Es kann zu jedem Zeitpunkt  $0 \leq t \leq T$  gehandelt werden.
- Finanzgüter können unendlich geteilt werden, d.h. auch Teile von Aktien dürfen gehandelt werden.
- Es ist erlaubt Aktien zu verkaufen, die noch gar nicht im eigenen Besitz sind. Dies wird als *Short Selling* bezeichnet.
- Der Gewinn  $\mu$  und die Volatilität  $\sigma$  sowie die risikolose Zinsrate  $r$  sind bekannt und konstant.
- Es gibt keine Transaktionskosten.
- Es werden keine Dividenden- oder sonstige Auszahlungen durchgeführt.

Ein zentraler Punkt in der Preistheorie ist auch die Vollständigkeit des Marktes, die im Falle des Black-Scholes Modells im folgenden Satz formuliert ist.

**Satz 7.17. (Vollständigkeit des Black-Scholes Modells)**

*Das Black-Scholes Modell ist vollständig, d.h. jede Forderung  $H_T \geq 0$ , die bezüglich dem Martingalmaß  $\mathbb{Q}$  integrierbar ist, kann durch eine Hedge  $X$  abgesichert werden.*

Aus der Vollständigkeit folgt auch die Eindeutigkeit des Martingalmaßes  $\mathbb{Q}$  auf  $\mathcal{F}_T$ . Besitzt nämlich jede Forderung  $H_T$  mit  $\tilde{H}_T$  eine Hedge, so ist  $\mathbb{Q}$  eindeutig. Das folgt sofort bei Wahl von  $\tilde{H}_T = \mathbb{1}_A$  für  $A \in \mathcal{F}_T$  aus  $\mathbb{Q}(A) = \mathbb{E}_{\mathbb{Q}}(\mathbb{1}_A) = \pi_0(H_T)$ .

Für europäische Optionen ist eine geschlossene Formel für den Preis der Call- und der Put-Option im Black-Scholes Modell bekannt. Diese resultiert als Lösung der Black-Scholes Differentialgleichung, die erstmalig von Black und Scholes [34] berechnet wurde.

**Satz 7.18. (Black-Scholes Differentialgleichung)**

*Der Wertprozess einer endfälligen Forderung  $f(S_T)$  ist von der Form  $V_t(X) = F(t, S_t)$ . Dabei erfüllt  $F(t, x)$  die partielle Differentialgleichung*

$$\frac{\partial F}{\partial t} + \frac{1}{2}\sigma^2 x^2 \frac{\partial^2 F}{\partial x^2} + rx \frac{\partial F}{\partial x} - rF = 0. \quad (7.8)$$

Die Endbedingung entspricht genau der Forderung,  $F(T, x) = f(x)$ .

**Satz 7.19. (Preis von europäischen Optionen im Black-Scholes Modell)**

Der faire Preis  $C_{BS}(t)$  einer Call-Option im Black-Scholes Modell ist für  $t < T$

$$C_{BS}(t) = S_t \Phi \left( \frac{\ln \frac{S_t}{K} + (r + \frac{\sigma^2}{2})(T-t)}{\sigma \sqrt{T-t}} \right) - K e^{-r(T-t)} \Phi \left( \frac{\ln \frac{S_t}{K} + (r - \frac{\sigma^2}{2})(T-t)}{\sigma \sqrt{T-t}} \right), \quad (7.9)$$

wobei  $\Phi$  die Verteilungsfunktion der Standardnormalverteilung symbolisiert.

Der Preis  $P_{BS}(t)$  einer Put-Option kann über die Put/Call-Parität bestimmt werden,

$$C_{BS}(t) - P_{BS}(t) = S_t - K e^{-r(T-t)}.$$

Wie sieht es nun aber im Falle von Asiatischen Optionen, insbesondere der arithmetischen Version aus (7.1) davon aus? Nach Satz 7.15 kann der Preis der arithmetischen Asiatischen Call-Option dargestellt werden als

$$C_t = e^{-r(T-t)} \mathbb{E}_{\mathbb{Q}} \left[ \left( \frac{1}{m} \sum_{i=1}^m S(t_i) - K \right)^+ \mid \mathcal{F}_t \right]. \quad (7.10)$$

Es kann wiederum, wie im Falle europäischer Optionen eine partielle Differentialgleichung angegeben werden, deren Lösung der Preis einer Asiatischen Option ist. Kemna und Vorst [28] geben eine detaillierte Beschreibung wie die Lösung dieser Differentialgleichung aussieht und warum es keine geschlossene Formel wie im Falle der Black-Scholes Differentialgleichung gibt.

**Satz 7.20. (Differentialgleichung für Asiatische Optionen)**

Für eine Asiatische Call-Option mit Payoff

$$C_T = (A_T - K)^+, \quad A_t = \int_0^t g(S_u) du, \quad (7.11)$$

wobei  $g$  eine zweimal stetig differenzierbare Funktion ist, ist der Preis eindeutig als Lösung der partiellen Differentialgleichung

$$\frac{\partial C}{\partial t} + g(S) \frac{\partial C}{\partial A} + \frac{1}{2} \sigma^2 S^2 \frac{\partial^2 C}{\partial S^2} + rS \frac{\partial C}{\partial S} - rC = 0 \quad (7.12)$$

gegeben.

Kemna und Vorst schlagen eine Lösung von (7.10) mittels eines Monte Carlo Verfahrens vor und berechnen den Preis als Mittelwert über  $n = 10000$  Simulationen von

$$e^{-rT} (A(T) - K)^+ \quad \text{mit} \quad A(T) = \frac{1}{m} \sum_{i=1}^m S(t_i).$$

Genau diese Strategie kommt in Abschnitt 7.3 zur Anwendung, nur dass es sich dort um eine Asiatische Basket Option handelt.

### 7.1.4 Konstruktion von Pfaden einer Brown'schen Bewegung

Die herkömmliche Weise einen Pfad  $(B(t_1), \dots, B(t_n))$  einer Standard-Brown'schen Bewegung für  $0 < t_1 < \dots < t_n$  zu erzeugen, erfolgt direkt unter Verwendung von unabhängigen standardnormalverteilten Zufallsvariablen  $Z_1, \dots, Z_n$  über

$$B(t_{i+1}) = B(t_i) + Z_{i+1} \sqrt{t_{i+1} - t_i}, \quad i = 0, \dots, n-1, \quad (7.13)$$

mit  $t_0 = 0$  und  $B(0) = 0$ .

Der Vektor  $(B(t_1), \dots, B(t_n))$  hat bekannterweise die Verteilung  $N(0, \Sigma)$  mit  $\Sigma_{ij} = \min(t_i, t_j)$ , da die Kovarianz sich für  $t_i < t_j$  durch

$$\begin{aligned} \text{Cov}[B(t_i), B(t_j)] &= \mathbb{E}[B(t_i)B(t_j)] - \mathbb{E}[B(t_i)]\mathbb{E}[B(t_j)] = \mathbb{E}[B(t_i)B(t_j)] \\ &= \mathbb{E}[B(t_i)(B(t_j) + B(t_i) - B(t_i))] = \mathbb{E}[B(t_i)^2] + \mathbb{E}[B(t_i)(B(t_j) - B(t_i))] \\ &= t_i + \mathbb{E}[B(t_j) - B(t_i)]\mathbb{E}[B(t_i)] = t_i \end{aligned}$$

berechnen lässt. Analoges lässt sich für  $t_j < t_i$  zeigen, woraus  $\Sigma_{ij} = \min(t_i, t_j)$  folgt.

Nachdem der Vektor  $(B(t_1), \dots, B(t_n))$  bekannterweise die Verteilung  $N(0, \Sigma)$  mit  $\Sigma_{ij} = \min(t_i, t_j)$  hat, gibt es auch noch andere Möglichkeiten die Brown'sche Bewegung zu generieren. Die wohl am häufigsten angewandte Methode ist die Cholesky-Zerlegung, die eine untere Dreiecksmatrix  $C$  liefert,

$$C = \begin{pmatrix} \sqrt{t_1} & 0 & \dots & 0 \\ \sqrt{t_1} & \sqrt{t_2 - t_1} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \sqrt{t_1} & \sqrt{t_2 - t_1} & \dots & \sqrt{t_n - t_{n-1}} \end{pmatrix}.$$

sodass  $CC^t = \Sigma$ .  $(B(t_1), \dots, B(t_n))$  resultieren dann aus dem Matrix-Vektor Produkt  $CZ$ .

Dass dies nicht alle Wege sind, Pfade einer Brown'schen Bewegung zu konstruieren, zeigen Glasserman [18] sowie Moskowitz und Caflisch [36] und Akesson und Lehoczky [4], aus deren Veröffentlichungen die nachfolgenden Ergebnisse in Bezug auf die Brown'sche Brücke und *Principal Components* zusammengefasst sind.

### Brown'sche Brückenkonstruktion

Während in (7.13) der Pfad sukzessive aufgebaut wird, ist die Überlegung der Brown'schen Brückenkonstruktion, die Elemente des Vektors  $(B(t_1), \dots, B(t_n))$  in mehr oder weniger beliebiger Reihenfolge aufzubauen, wobei aber natürlich zu jedem Zeitpunkt aus der richtigen bedingten Verteilung simuliert werden muss.

Eine günstige Aufteilung ist oft, immer wieder zu halbieren, sprich zuerst den Wert  $B(t_n)$  aus der Verteilung  $N(0, t_n)$  zu generieren, anschließend  $B(t_{\lfloor n/2 \rfloor})$  bedingt auf  $B(t_n)$  und so weiter. Wird eine Brown'sche Bewegung auf ihre Endpunkte bedingt, so entsteht eine sogenannte "Brown'sche Brücke". Der Vorteil, den die Normalverteilung bietet, ist, dass auch die benötigten bedingten Verteilungen immer Normalverteilungen sind, es ändern sich lediglich die Parameter der Verteilung.

Im Falle eines multivariat normalverteilten Vektors  $\mathbf{X} = (X_1, X_2)^t$  mit

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N \left( \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right) \quad (7.14)$$

gilt

$$(X_1 | X_2 = x) \sim N(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x - \mu_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}). \quad (7.15)$$

Für die Konstruktion einer Brown'schen Brücke werden im Folgenden Zeitpunkte  $0 < u < s < t$  betrachtet und  $B(s)$  soll auf  $B(u) = x$  und  $B(t) = y$  bedingt werden. Um obige Formel anzuwenden, werden die Einträge des Vektors und der Kovarianzmatrix derart permutiert, dass  $B(s)$  an erster Stelle steht, d.h. aus

$$\begin{pmatrix} B(u) \\ B(s) \\ B(t) \end{pmatrix} \sim N \left( 0, \begin{pmatrix} u & u & u \\ u & s & s \\ u & s & t \end{pmatrix} \right)$$

wird

$$\begin{pmatrix} B(s) \\ B(u) \\ B(t) \end{pmatrix} \sim N \left( 0, \begin{pmatrix} s & u & s \\ u & u & u \\ s & u & t \end{pmatrix} \right)$$

Der Erwartungswert ergibt sich dann nach der Formel als

$$\mathbb{E}[B(s) | B(u) = x, B(t) = y] = \frac{(t-s)x + (s-u)y}{t-u},$$

die Varianz entspricht

$$\text{Var}[B(s) | B(u) = x, B(t) = y] = \frac{(t-s)(s-u)}{t-u}. \quad (7.16)$$

Allgemein gilt

$$(B(s) | B(s_1) = x_1, \dots, B(s_k) = x_k) \sim N \left( \frac{(s_{i+1} - s)x_i + (s - s_i)x_{i+1}}{s_{i+1} - s_i}, \frac{(s_{i+1} - s)(s - s_i)}{s_{i+1} - s_i} \right), \quad (7.17)$$

wobei die Werte  $B(s_1) = x_1, \dots, B(s_k) = x_k$  zu den Zeitpunkten  $s_1 < s_2 < \dots < s_k$  bereits festgelegt worden sind und ein Wert  $B(s)$  mit  $s_i < s < s_{i+1}$  bedingt auf diese Werte generiert

wird.

Das Ergebnis bezieht sich nur auf die nächstgelegenen Zeitpunkte als Resultat der Markoveigenschaft der Brown'schen Bewegung (Satz 7.21).  $B(s)$  gegeben  $B(s_i)$  und  $B(s_{i+1})$  ist damit unabhängig von allen Zeitpunkten  $t < s_i$  und  $t > s_{i+1}$ .

**Satz 7.21. (Starke Markoveigenschaft der Brown'schen Bewegung)** *Ist  $B$  eine Brown'sche Bewegung und  $T$  eine Stoppzeit, So ist  $\tilde{B}(t) = B(t + T) - B(T)$  wieder eine Brown'sche Bewegung, besitzt die gleiche Verteilung wie  $B$  und ist unabhängig von  $\mathcal{F}_T$ .*

Aus dieser bedingten Verteilung kann nun mithilfe einer von  $B(s_1), \dots, B(s_k)$  unabhängigen Zufallsvariable  $Z \sim N(0,1)$  der Wert  $B(s)$  ermittelt werden als

$$B(s) = \frac{(s_{i+1} - s)x_i + (s - s_i)x_{i+1}}{s_{i+1} - s_i} + Z \sqrt{\frac{(s_{i+1} - s)(s - s_i)}{s_{i+1} - s_i}}. \quad (7.18)$$

In Abbildung 7.1 ist der Aufbau einer Brown'schen Brücke zu sehen. Als erstes stehen die Punkte  $B(0)$  und  $B(t_n)$  fest. Mit jedem weiteren Schritt wird jeweils ein Punkt zwischen zwei bereits fixierten Punkten eingefügt, was in den drei übrigen Bildern ersichtlich ist.

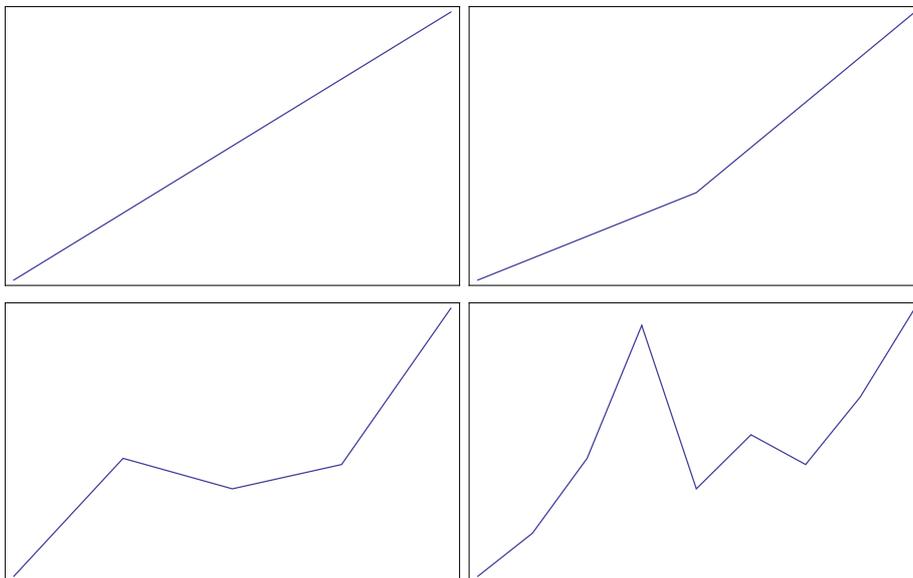


Abbildung 7.1: Die Konstruktion einer Brown'schen Brücke nach 1,2,4 und 8 generierten Punkten.

Für eine Brown'sche Bewegung mit Drift  $\mu$  ändert sich lediglich die Generierung des Wertes  $B(t_n)$ , der nun nicht wie bei der Standard Brown'schen Bewegung aus der Verteilung  $N(0, t_n)$  kommt, sondern eine  $N(\mu t_n, t_n)$ -Verteilung aufweist. Die Vorgangsweise für die übrigen Werte der Brown'schen Brücke ist analog Gleichung (7.18). Hat die Brown'sche Bewegung einen Diffusionskoeffizienten  $\sigma^2$ , so ergänzt sich die bedingte Varianz aus (7.16) um den multiplikativen Faktor  $\sigma^2$ .

Ob nun ein *Random Walk* wie in (7.13) berechnet oder gemäß (7.18) ermittelt wird, macht, zumindest was den Rechenaufwand betrifft, keinen relevanten Unterschied. Interessant wird so eine Konstruktion erst in Verbindung mit diversen Varianzreduktionsverfahren und Quasi-Monte Carlo Verfahren. Eine der ersten Veröffentlichungen solch einer Kombination stammt von Moskowitz und Caflisch [36]. Die Überlegung besteht darin, dass der entscheidende Wert zur Bepreisung von Optionen meist jener Wert ist, der zum Fälligkeitszeitpunkt berechnet wird. Während in (7.13)  $n$  Zufallsvariablen benötigt werden, um den Endwert zu generieren, wird in (7.18) nur eine einzige gebraucht. Dadurch kann ein großer Teil der Variabilität im Payoff der Funktion beseitigt werden, betrachtet man nämlich die Varianz aus (7.17), so ist erkennbar, dass die Varianz, je kleiner das Intervall wird, immer mehr zunimmt und für  $s_{i+1} - s_i \rightarrow 0$  gegen unendlich strebt.

In Bezug auf Quasi-Monte Carlo Verfahren birgt die Idee einer Kombination mit Brown'schen Brücken noch einen anderen Vorteil. Zur Ermittlung von Pfaden einer Brown'schen Bewegung wird jeweils eine Koordinate der Quasi-Monte Carlo Punkte mithilfe der Inversionsmethode in eine normalverteilte Stichprobe transformiert. Diese wird anschließend skaliert und für die Berechnung des *Random Walks* verwendet. Werden nun die einzelnen Koordinaten einer Folge kleiner Diskrepanzen betrachtet, so sind die ersten Koordinaten besser gleichverteilt, als jene mit einem höheren Index. Aus diesem Grund ist es sinnvoll eine Ordnung zu schaffen, wo durch jene Variablen, die den ersten Koordinaten entsprechen, ein großer Teil der Varianz erklärt ist. Genau dies passiert mithilfe der Brown'schen Brückenkonstruktion.

### Principal Components Construction

In der vorangestellten Methode werden die Hauptpunkte des Pfades der Brown'schen Bewegung über die ersten Zufallsvariablen festgelegt. Die Frage ist, ob dies nicht noch ausbaufähig ist und genau festgelegt werden kann, welche Koordinaten den größten Anteil an Varianz tragen sollen, um diese schlussendlich noch weiter reduzieren zu können.

Die *Principal Components Construction* maximiert die Variabilität des Pfades, der durch die normalverteilten Zufallsvariablen  $Z_1, \dots, Z_k$  für alle  $k = 1, \dots, n$  erklärt ist. Bevor näher auf diese Konstruktion eingegangen wird, ist noch eine kurze Erklärung nötig, was überhaupt unter einem *Principal Component* bzw. einer Hauptkomponente zu verstehen ist.

Im Gegensatz zur Cholesky-Zerlegung der Varianz-Kovarianzmatrix  $\Sigma$ , wird hier eine Zerlegung mithilfe der Eigenwerte und Eigenvektoren angestrebt. Als symmetrische und positiv (semi-)definite  $d \times d$  Varianz-Kovarianzmatrix besitzt  $\Sigma$   $d$  reelle, nichtnegative Eigenwerte  $\lambda_1, \dots, \lambda_d$  mit  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_d$ , sowie eine dazugehörige Menge an orthonormalen Eigenvektoren  $v_1, \dots, v_d$ , sodass für  $i, j = 1, \dots, d$  und  $i \neq j$

$$v_i^t v_i = 1, \quad v_i^t v_j = 0, \quad \text{und} \quad \Sigma v_i = \lambda_i v_i.$$

Damit existiert eine Zerlegung  $\Sigma = V \Lambda V^t$ , wobei die Spalten von  $V$  genau den Eigenvektoren  $v_1, \dots, v_d$  entsprechen, und damit  $V$  eine orthogonale Matrix mit  $V V^t = I$  darstellt, und  $\Lambda$  eine Diagonalmatrix mit Einträgen  $\lambda_1, \dots, \lambda_d$  ist. Mit  $A = V \sqrt{\Lambda}$  folgt  $A A^t = V \Lambda V^t = \Sigma$ .

Ein Vektor  $X \sim N(0, \Sigma)$  kann generiert werden als

$$X = a_1 Z_1 + a_2 Z_2 + \dots + a_d Z_d$$

mit  $Z \sim N(0, I)$  und  $a_j$  als  $j$ -te Spalte von  $A$  für  $j = 1, \dots, d$ . Je nach Rang  $k$  der Matrix  $\Sigma$ , wird  $X$  von den Variablen  $Z_1, \dots, Z_k$ ,  $k = 1, \dots, d$  repräsentiert. Hat  $\Sigma$  vollen Rang, so hat auch  $A$  einen vollen Rang und es gilt

$$Z = A^{-1} X = \sqrt{\Lambda} V^t.$$

Aus dieser Darstellung folgt unmittelbar, dass die  $Z_j$ ,  $j = 1, \dots, d$  Linearkombinationen der  $X_i$ ,  $i = 1, \dots, d$  sind und außerdem proportional zu  $v_j^t X$ . Genau diese Linearkombinationen  $v_j^t X$  werden *Principal Components* genannt.

Das Optimierungsproblem besteht darin, den mittleren quadratischen Fehler

$$\mathbb{E} \left[ \left\| X - \sum_{i=1}^k a_i Z_i \right\|^2 \right], \quad (7.19)$$

wobei  $\|x\|^2 = x^t x$  ist, in Bezug auf die Zufallsvariablen  $Z_1, \dots, Z_k$  und die Vektoren  $a_1, \dots, a_k$  zu minimieren. Dies passiert für jedes  $k = 1, \dots, d$ , wenn  $a_1, \dots, a_k$  die Spalten der Matrix  $A = V \sqrt{\Lambda}$  sind und

$$Z_j = \frac{v_j^t X}{\sqrt{\lambda_j}},$$

für  $j = 1, \dots, k$  (proportional zu  $v_j^t X$ ) gewählt werden. Im Grunde kann durch die *Principal Components* eine Approximation des Zufallsvektors erreicht werden, welche aber über eine geringere Dimension verfügt. Die durch die ersten  $k$  *Principal Components* erklärte Varianz wird beschrieben über

$$\frac{\lambda_1 + \dots + \lambda_k}{\lambda_1 + \dots + \lambda_d}. \quad (7.20)$$

Der erste Hauptkomponent ist immer jener, der am meisten der Variabilität in sich trägt.

Um zur Generierung von Pfaden einer Brown'schen Bewegung zurückzukehren, sei nun

$$\begin{pmatrix} B(t_1) \\ B(t_2) \\ \vdots \\ B(t_n) \end{pmatrix} = \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{n1} \end{pmatrix} Z_1 + \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{n2} \end{pmatrix} Z_2 + \cdots + \begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{nn} \end{pmatrix} Z_n \quad (7.21)$$

mit  $a_i = (a_{1i}, \dots, a_{ni})^t, i = 1, \dots, n$  als Spalten der  $n \times n$  Matrix  $A$ , wobei  $AA^t = \Sigma$ , die Varianz-Kovarianzmatrix von  $(B(t_1), \dots, B(t_n))$  ergibt. Der Vorgang um den mittleren quadratischen Fehler

$$\mathbb{E} \left[ \left\| B - \sum_{j=1}^k a_j Z_j \right\|^2 \right]$$

zu minimieren, ist analog dem eben gezeigten, mit  $a_i = \sqrt{\lambda_i} v_i, i = 1, \dots, n$  wobei  $\lambda_1 > \dots > \lambda_n > 0$  die Eigenwerte und  $v_i, i = 1, \dots, n$  die zugehörigen Eigenvektoren von  $\Sigma$  sind.

Interessant ist das Verhalten der Eigenvektoren. Werden die Einträge der normalisierten Eigenvektoren geplottet, so kann eine Sinusschwingung erkannt werden. Sind die Vektoren nach der Größe ihrer zugehörigen Eigenwerte geordnet, so nimmt die Frequenz der Schwingung ab, je größer der Eigenwert ist. Dass der Verlauf der Eigenvektoren wirklich einer Sinusschwingung folgt, zeigen Akesson und Lehoczky [4]. Für einen Pfad der Länge  $n$  mit äquidistanten Zeitpunkten  $t_{i+1} - t_i = \Delta t$  können die Eigenvektoren  $v_i, i = 1, \dots, n$  und Eigenwerte  $\lambda_i, i = 1, \dots, n$  über die Formeln

$$v_i(j) = \frac{2}{\sqrt{2n+1}} \sin \left( \frac{2i-1}{2n+1} j\pi \right), \quad j = 1, \dots, n$$

und

$$\lambda_i = \frac{\Delta t}{4} \sin^{-2} \left( \frac{2i-1}{2n+1} \frac{\pi}{2} \right)$$

berechnet werden.

Abbildung 7.2 zeigt den Verlauf der einzelnen Eigenwerte für  $n = 64$  mit  $\Delta t = 1/64$ . Sehr gut zu sehen ist der rapide Abfall der Kurve. Die größten 8 Eigenwerte liegen bei 26.3464, 2.92853, 1.05511, 0.538959, 0.326553, 0.219035, 0.157197, 0.1184. Eingesetzt in Gleichung (7.20) bedeutet dies eine Variabilität von 81%, 90%, 93%, ... für  $k = 1, 2, 3, \dots$  bis hin zu einer Variabilität von über 99% für  $k = 19$ . In Abbildung 7.3 sind die zugehörigen Eigenvektoren zu den vier größten Eigenwerten der  $64 \times 64$  großen Kovarianzmatrix einer Brown'schen Bewegung zu sehen.

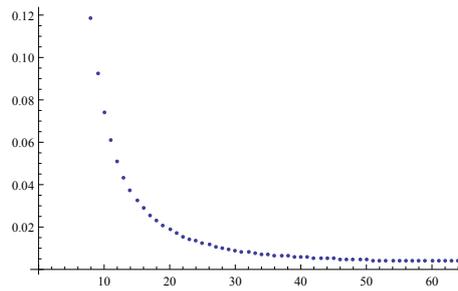


Abbildung 7.2: Eigenwerte der  $64 \times 64$  Kovarianzmatrix von einem Pfad der Brown'schen Bewegung

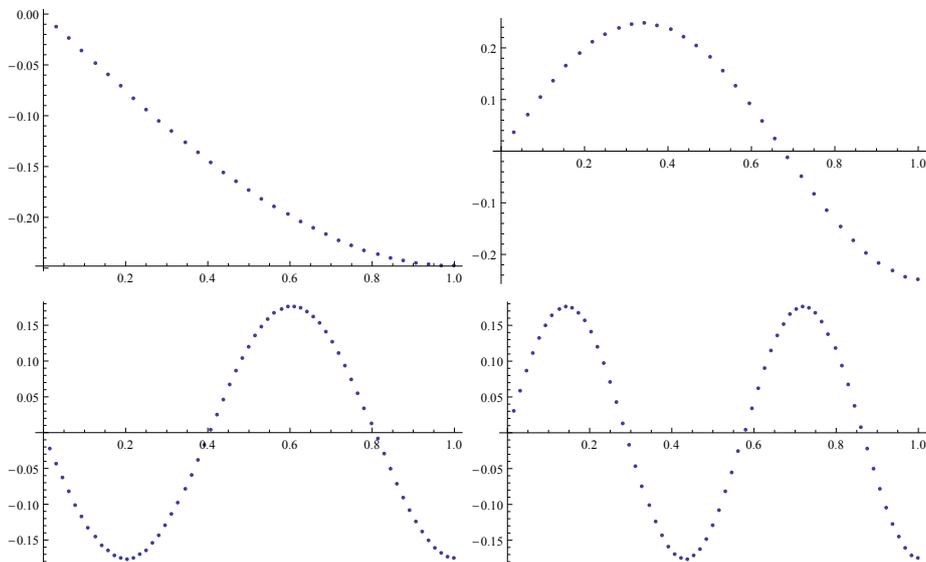


Abbildung 7.3: Veranschaulichung der Eigenvektoren zu den vier größten Eigenwerten der  $64 \times 64$  Kovarianzmatrix von einem Pfad der Brown'schen Bewegung

Im Vergleich zur Konstruktion der Brown'schen Brücke ist die *Principal Components Construction* optimal bezüglich der erklärten Variabilität. Nachteile der letzteren sind, dass zum einen  $O(n^2)$  anstatt den  $O(n)$  Operationen des *Random Walks* notwendig sind, zum anderen kann erst etwas über einen Wert  $B(t_i)$ ,  $i = 1, \dots, n$  ausgesagt werden, nachdem alle Zufallsvariablen  $Z_i$ ,  $i = 1, \dots, n$  generiert worden sind. Bei der Brown'schen Brücke ist mit jeder einzelnen Zufallsvariable ein weiterer Wert des Pfades fixiert.

## 7.2 Collateralized Mortgage Obligation

Bevor nun das bereits angesprochene Beispiel der Asiatischen Basket Option zur Sprache kommt, soll hier noch das Beispiel der *Collateralized Mortgage Obligation* angeführt werden, da die Veröffentlichungen von Paskov [49], [50] und Paskov und Traub [51], in Bezug auf die Berechnung des Preises einer solchen, für große Aufregung in der Theorie der Quasi-Monte Carlo Verfahren gesorgt haben.

Die Preisberechnung beinhaltet die Auswertung eines 360-dimensionalen Integrales, was, begründet durch die Abhängigkeit der asymptotischen Fehlerschranke  $O(N^{-1} \ln^d N)$  von der Dimension  $d$ , ein Problem in der Anwendung des Quasi-Monte Carlo Verfahren sein müsste. Interessanterweise ist dies hier nicht der Fall und die Quasi-Monte Carlo Simulation, insbesondere bei Verwendung von Sobol-Folgen, sticht sogar das Monte Carlo Verfahren mit seiner von der Dimension unabhängigen Schranke  $O(N^{-1/2})$  aus.

Bevor näher auf die genaue Struktur der Anwendung eingegangen wird und erklärt wird, wie die gute Performance der Quasi-Monte Carlo Methode zustande kommen kann, ist im Folgenden erst einmal eine kurze Beschreibung des Problems zu finden.

**Definition 7.22.** (Mortgage Backed Security, MBS)

*Mortgage Backed Securities sind durch Hypotheken gesicherte Wertpapiere. Ist ein Investor mit einem gewissen Prozentsatz am Hypothekenpool beteiligt, so erhält er eben diesen an zurückgezahltem Grundkapital und an der Zinszahlung der Hypotheken an den Pool.*

Das Risiko für die Investoren besteht in einer Vorauszahlungsoption der Hypotheken, d.h. die Hypothek oder Teile davon können schon vorher zum Nominalwert zurückgezahlt werden. Der Wert der MBS wird durch den Erwartungswert der Vorauszahlung zum Ablaufzeitpunkt der MBS festgelegt.

**Definition 7.23.** (Collateralized Mortgage Obligation, CMO)

*Bei einer CMO werden mehrere Investoren der MBS zu Tranchen (Klassen) zusammengefasst und die Rückzahlungen je nach Tranche aufgeteilt. Der Wert der CMO entspricht dem Erwartungswert der summierten Barwerte an zukünftigen Zahlungen pro Tranche.*

Wir betrachten eine CMO mit 10 Tranchen. Der Hypothekenpool habe eine Fälligkeit von 30 Jahren bei 360 monatlichen Zahlungen in Höhe  $C$ . Sei  $i_k$  die Zinsrate in Monat  $k$ ,  $\omega_k$  der Prozentsatz jener, die Vorauszahlungen leisten und  $a_{360-k+1}$  der Barwert jener Zahlungen, die nach dem  $k$ -ten Monat noch ausstehend sind,  $k = 1, \dots, 360$ .

Der Barwert der übrigen Zahlungen nach Monat  $k$  ist bzgl. dem fixen Anfangszinssatz  $i_0$  gegeben als

$$a_k = C \left( 1 + v_0 + \dots + v_0^{k-1} \right), \quad v_0 = \frac{1}{1 + i_0}.$$

Damit gelten  $C$  und  $a_{360-k+1}$  als konstant und es bleibt noch, das Zinsmodell zu besprechen und die Berechnungsregeln für die stochastischen Variablen  $i_k$  und  $\omega_k$  anzugeben.

Das Zinsmodell ist von der Form

$$i_k = K_0 e^{\xi_k} i_{k-1} = K_0^k i_0 e^{\xi_1 + \dots + \xi_k}$$

mit Konstante  $K_0$  und unabhängig und normalverteilten Zufallsvariablen  $\xi_1, \dots, \xi_{360}$ , die Erwartungswert 0 und Varianz  $\sigma^2$  aufweisen. In den Ausführungen von Paskov und Traub [49] ist für  $\sigma^2$  der Wert 0.0004 gewählt worden.

Das Modell für die Vorauszahlungen  $\omega_k$  ist als Funktion des Zinssatzes gegeben als

$$\begin{aligned} \omega_k &= \omega_k(\xi_1, \dots, \xi_k) = K_1 + K_2 \arctan(K_3 i_k + K_4) \\ &= K_1 + K_2 \arctan(K_3 K_0^k i_0 e^{\xi_1 + \dots + \xi_k} + K_4), \end{aligned}$$

mit gegebenen Konstanten  $K_1, K_2, K_3, K_4$ . Daraus können die Zahlungen in den Pool pro Monat  $k$  berechnet werden durch

$$\begin{aligned} M_k &= M_k(\xi_1, \dots, \xi_k) \\ &= C(1 - \omega_1(\xi_1)) \cdots (1 - \omega_{k-1}(\xi_1, \dots, \xi_{k-1})) \cdot (1 + \omega_k(\xi_1, \dots, \xi_k)(a_{360-k+1} - 1)). \end{aligned}$$

Sei der Anteil von  $M_k$  der Tranche  $T$  für den Monat  $k$  mit  $G_{k;T}(\xi_1, \dots, \xi_k)$  bezeichnet. Im Allgemeinen ist diese Funktion stetig als Zusammensetzung von glatten Funktionen und solchen, die sich aus dem Minimum von Funktionen ergeben. Um daraus den Wert zum jetzigen Zeitpunkt zu erhalten, muss diese Größe mit dem Diskontierungsfaktor

$$u_k(\xi_1, \dots, \xi_{k-1}) = v_0 v_1(\xi_1) \cdots v_{k-1}(\xi_1, \dots, \xi_{k-1}), \quad k = 1, \dots, 360$$

wobei

$$v_j(\xi_1, \dots, \xi_j) = \frac{1}{1 + i_j(\xi_1, \dots, \xi_j)} = \frac{1}{1 + K_0^j i_0 e^{\xi_1 + \dots + \xi_j}}, \quad j = 1, \dots, 359$$

multipliziert werden.

Der Barwert  $PV_T$  der Tranche  $T$  ist demnach als Summe der Zahlungen pro Monat  $k$  für  $k = 1, \dots, 360$  gegeben durch

$$PV_T(\xi_1, \dots, \xi_{360}) = \sum_{k=1}^{360} G_{k;T}(\xi_1, \dots, \xi_k) u_k(\xi_1, \dots, \xi_{k-1}). \quad (7.22)$$

Nach einer Transformation der Variablen ergibt sich der Erwartungswert des Barwerts als

$$\mathbb{E}[PV_T] = \int_{[0,1]^{360}} PV_T(y_1(x_1), \dots, y_{360}(x_{360})) dx_1 \cdots dx_{360} \quad (7.23)$$

mit  $y_i = y_i(x_i)$  implizit gegeben durch

$$x_i = \frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^{y_i} e^{-\frac{t^2}{2\sigma}} dt.$$

Der Wert der CMO ergibt sich demnach aus der Berechnung von 10 Integralen über den 360-dimensionalen Einheitswürfel, eines für jede der 10 Tranchen.

Im Allgemeinen besteht die Vermutung, dass Folgen kleiner Diskrepanzen nur bis zu einer Dimension von  $d \leq 30$  einen Vorteil in der Anwendung haben. Wie kann also gerechtfertigt werden, dass Quasi-Monte Carlo in Dimension  $d = 360$  zu bevorzugen ist?

Wird das Problem im Detail untersucht, kann als erste Verbesserung erkannt werden, dass nicht jede der Tranchen von Dimension 360 ist. Der Integrand  $PV_T$  von Tranche  $T$  hängt offensichtlich von den einzelnen Cash-Flows  $G_{k;T}$  ab. Tranche  $T$  kann aber schon vorzeitig im Monat  $k_T$  ausbezahlt worden sein, was bedeutet, dass keine Cash-Flows  $G_{k;T}$  für die restlichen Monate  $k > k_T$  mehr eingehen, d.h.

$$G_{k;T}(\xi_1, \dots, \xi_k) = 0 \quad \forall k > k_T = k_T(\xi_1, \dots, \xi_{k_T}).$$

Die Untersuchungen von Paskov [49] zeigen, dass in Realität nur eine einzige der Tranchen Dimension 360 aufweist. Eine wirkliche Erleichterung bringt dies aber nicht, denn, auch wenn nicht alle von maximaler Dimension sind, so übersteigt diese aber dennoch in jeder Tranche die Marke von  $d = 30$  bei weitem. Deshalb besteht die nächste Überlegung darin, wie groß die "effektive Dimension" der einzelnen Tranchen sein könnte. Dadurch erfolgt wiederum eine Reduktion und die Anwendung des Verfahrens wird immer plausibler.

## Ergebnisse

Paskov [49] präsentiert den Vergleich von Monte Carlo Simulation, Antithetic Variates, Halton Folgen und Sobol Folgen. Deutlicher Sieger in Hinblick auf Konvergenz und Fehleranalyse ist die Sobol Folge. Zusammengefasst kann gesagt werden:

- The Sobol algorithm converges significantly faster than the Monte Carlo algorithm;
- The convergence of the Sobol algorithm is smoother than the convergence of the Monte Carlo algorithm - This makes automatic termination easier for the Sobol algorithm;
- (...) the Sobol algorithm terminates 2 to 5 times faster than the Monte Carlo algorithm often with smaller error;

- The Monte Carlo algorithm is sensitive to the initial seed<sup>1</sup>.

Das Quasi-Monte Carlo Verfahren kann also durchaus auch in höherdimensionalem Kontext von Vorteil sein. Entscheidend ist immer das Umfeld, in dem es angewandt wird und wie die Funktion bzw. der Integrand beschaffen ist. Wichtig ist vor allem die effektive Dimension des Problems. Dies haben auch andere Autoren erkannt wie z.B. Caflisch, Morokoff und Owen [9], die zur Berechnung der Mortgage Backed Securities eine Brown'sche Brücke verwenden. Wie wichtig die effektive Dimension sein kann, ist auch in folgendem Beispiel zu sehen.

### 7.3 Asian Basket Option

Um die einzelnen Methoden zur Generation von Zufallszahlen an einem Beispiel zu demonstrieren, wird der Preis einer Asiatischen Basket Option ermittelt. Die Berechnung des Payoffs einer solchen pfadabhängigen Option basiert auf mehreren zugrundeliegenden Assets, die zu verschiedenen Zeitpunkten über die gesamte Laufzeit hin betrachtet werden.

Im Folgenden wird ein Basket mit  $m$  Assets betrachtet, die zum Zeitpunkt  $t$  Preise  $S_t^i$  für  $i = 1, \dots, m$  aufweisen. Die Fälligkeit  $T$  sei fix gegeben, ebenso der Strike  $K$ , sowie die  $n$  beobachteten Zeitpunkte  $0 < t_1 < \dots < t_n = T$  und die Gewichte  $\omega_{ij}$  für  $i = 1, \dots, m$ ,  $j = 1, \dots, n$  mit  $\sum_{i,j} \omega_{ij} = 1$ . Die Preise  $S_t^i$  der Assets seien als Lösung der Stochastischen Differentialgleichung

$$dS_t^i = rS_t^i dt + \sigma^i S_t^i dW_t^i, \quad i = 1, \dots, m \quad (7.24)$$

gegeben, was bedeutet, dass sie einer geometrischen Brown'schen Bewegung mit risikolosem Zins  $r$  und Volatilitäten  $\sigma^i$ ,  $i = 1, \dots, m$  folgen.  $(W^1, \dots, W^m)$  bezeichnet eine  $m$ -dimensionale Brown'sche Bewegung mit Korrelation  $\rho_{ik}$  zwischen den Komponenten  $W^i$  und  $W^k$  für  $i, k = 1, \dots, m$ .

Die explizite Darstellung der Lösung von (7.24) lässt sich schreiben als

$$S_t^i = S_0^i \exp^{(r - (\sigma^i)^2)/2)t + \sigma^i W_t^i}, \quad i = 1, \dots, m. \quad (7.25)$$

Unter all diesen Voraussetzungen ist der Payoff der Asiatischen Basket Call Option festgelegt als

$$\left( \sum_{i=1}^m \sum_{j=1}^n \omega_{ij} S_{t_j}^i - K \right)^+ \quad (7.26)$$

und der Preis der Option zum Zeitpunkt  $t = 0$  ergibt sich aus 7.15 als

$$\mathbb{E}_{\mathbb{Q}}[\tilde{H}_T | \mathcal{F}_t] = e^{-rT} \mathbb{E}_{\mathbb{Q}} \left[ \left( \sum_{i=1}^m \sum_{j=1}^n \omega_{ij} S_{t_j}^i - K \right)^+ \right]. \quad (7.27)$$

<sup>1</sup>Spassimir H. Paskov: *New Methodologies for Valuing Derivatives* (Seite 5)

Nachdem es nun aber, wie bereits in Abschnitt 7.1.3 erwähnt, keine geschlossene Formel für die Berechnung dieses Preises gibt und die Verwendung von numerischen Verfahren einen großen Aufwand mit sich bringt, wird im Folgenden, wie schon von Kemna und Vorst [28] vorgeschlagen, unter anderem auf das Monte Carlo Simulationsverfahren für eine Approximation des Preises zurückgegriffen.

Eine der Hauptaufgaben bei der Berechnung des Preises der Option ist die Simulation der Pfade  $W_{t_j}^i$  für  $j = 1, \dots, n, i = 1, \dots, m$  zu den gegebenen Zeitpunkten  $t_1, \dots, t_n$ . Der Einfachheit halber sei angenommen, dass der Abstand zwischen den Zeitpunkten immer gleich bleibt, d.h.  $\Delta t = T/n$  mit  $t_j = j\Delta t, j = 1, \dots, n$ . Es werden zwei verschiedene Verfahren an diesem Beispiel zur Ermittlung solcher Pfade ausprobiert. Dazu gehören die Cholesky-Zerlegung und die *Principal Components Construction*. Die Konstruktion einer Brown'schen Brücke findet hier keine praktische Anwendung, nachdem die *Principal Components Construction* eine optimierte Version der Brown'schen Brücke darstellt und somit keine weitere Varianzreduktion zu erwarten ist.

Hat die Simulation der Pfade einer Brown'schen Bewegung stattgefunden, so kann der Payoff der Asiatischen Basket Option zum Fälligkeitszeitpunkt  $T$  ermittelt werden als

$$\max \left[ g(\tilde{W}) - K, 0 \right], \quad g(\tilde{W}) = \sum_{k=1}^{mn} e^{\mu^k + \tilde{W}_k} \quad (7.28)$$

mit

$$\mu^k = \ln(\omega_{k_1 k_2} S_{k_1}(0)) + (r - (\sigma^{k_1})^2/2)t_{k_2},$$

und

$$k_1 = ((k-1) \bmod m) + 1$$

$$k_2 = \left\lfloor \frac{k-1}{m} \right\rfloor + 1, \quad k = 1, \dots, mn.$$

### Die Standardvariante - Cholesky-Zerlegung

Die erste Variante  $W_{t_j}^i$  für  $j = 1, \dots, n, i = 1, \dots, m$  zu simulieren, gelingt über eine Transformation mithilfe einer Cholesky-Zerlegung.

Ziel ist, den Vektor

$$\tilde{W} = (\sigma^1 W_{t_1}^1, \dots, \sigma^m W_{t_1}^m, \sigma^1 W_{t_2}^1, \dots, \sigma^m W_{t_2}^m, \dots, \sigma^m W_{t_m}^1, \dots, \sigma^m W_{t_m}^m)^t \quad (7.29)$$

aus Gleichung (7.25) zu ermitteln. Dazu kann die Transformation

$$\tilde{W} = \tilde{C}Z,$$

verwendet werden, wobei  $Z$  ein Vektor von  $mn$  unabhängigen  $N(0, 1)$ -verteilten Zufallsvariablen ist und  $\tilde{C}$  die Cholesky Zerlegung der Matrix  $\tilde{\Sigma}$ , deren Aufbau im Anschluss erklärt wird.

Zuerst wird die Kovarianzmatrix  $\Sigma$  aus den Korrelationen und Assetvolatilitäten aufgebaut über

$$\Sigma_{ik} = \rho_{ik}\sigma^i\sigma^k\Delta t, \quad i, k = 1, \dots, m.$$

Daraus wird die  $mn \times mn$  Matrix  $\tilde{\Sigma}$  berechnet als

$$\tilde{\Sigma} = \begin{pmatrix} \Sigma & \Sigma & \cdots & \Sigma \\ \Sigma & 2\Sigma & \cdots & 2\Sigma \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma & 2\Sigma & \cdots & n\Sigma \end{pmatrix}. \quad (7.30)$$

Von dieser Matrix wird die Cholesky-Zerlegung ermittelt, d.h. es wird eine untere Dreiecksmatrix  $\tilde{C}$  gesucht, welche die Eigenschaft  $\tilde{C}\tilde{C}^t = \tilde{\Sigma}$  erfüllt. Diese Dreiecksmatrix kann nun zur Berechnung von (7.29) verwendet werden.

### Principal Components Construction

In Abschnitt 7.1.4 wird die eindimensionale Variante der *Principal Components Construction* beschrieben. Anders als bei der Konstruktion einer Brown'schen Brücke, ist diese Methode nicht getrennt für jede Dimension anwendbar.

Es wird wieder eine Zerlegung der Kovarianzmatrix  $\tilde{\Sigma}$  aus Gleichung (7.30) des Vektors (7.29) benötigt. Dieses Mal kommen allerdings die Eigenwerte und Eigenvektoren von  $\tilde{\Sigma}$  zum Einsatz.

Setzt sich eine Matrix aus dem Kroneckerprodukt von zwei Matrizen zusammen,

$$A \otimes \Sigma = \begin{pmatrix} A_{11}\Sigma & A_{12}\Sigma & \cdots & A_{1n}\Sigma \\ A_{21}\Sigma & A_{22}\Sigma & \cdots & A_{2n}\Sigma \\ \vdots & \vdots & \ddots & \vdots \\ A_{n1}\Sigma & A_{n2}\Sigma & \cdots & A_{nn}\Sigma \end{pmatrix}$$

wobei  $A$  Eigenvektoren  $v_1, \dots, v_n$  und Eigenwerte  $\lambda_1 \geq \dots \geq \lambda_n$  besitzt und  $\Sigma$  Eigenvektoren  $w_1, \dots, w_m$  und Eigenwerte  $\eta_1 \geq \dots \geq \eta_m$ , so sind die Eigenvektoren von  $(A \otimes \Sigma)$  gegeben als  $(v_i \otimes w_j)$ , die Eigenwerte als  $\lambda_i\eta_j$  für  $i = 1, \dots, n, j = 1, \dots, m$ . Das bedeutet bei dementsprechend großen Matrizen eine wesentliche Reduktion der Rechenzeit. Der Beweis zu dieser

Aussage ist bei Horn und Johnson [21] zu finden.

Der Vektor  $\tilde{W}$  aus (7.29) kann nun wiederum über die Matrixzerlegung unter Zuhilfenahme eines normalverteilten Zufallsvektors  $Z$  berechnet werden. Es gilt

$$\tilde{W} = \tilde{B}Z,$$

wobei  $\tilde{B} = V \cdot \sqrt{\lambda}$  mit  $\tilde{B}\tilde{B}^t = \tilde{\Sigma}$ .

### 7.3.1 Numerische Ergebnisse

Die Ermittlung des Preises der Asiatischen Basket Option beinhaltet mehrere Schritte.

1. Generierung von Zufallszahlen mithilfe von Monte Carlo, Monte Carlo mit Kontrollvariablen, Quasi-Monte Carlo, Latin Hypercube, Latin Hypercube with Dependence und Latin Supercube.
2. Die Simulation der Pfade erfolgt über die Cholesky Zerlegung bzw. über die *Principal Components Construction*
3. Es wird über 10 Berechnungen von 1024, 2048, 4096 und 8192 Simulationen gemittelt. (Da die Sobol-Folge ein  $(t, m, d)$ -Netz in Basis 2 ist, liefert sie nach Niederreiter [41] die beste Performance bei Stichprobengrößen, die Potenzen von 2 sind.)
4. Das Beispiel wird einmal mit und einmal ohne Korrelation der Komponenten der Brown'schen Bewegung betrachtet und für drei verschiedene Strikes (at the money, in the money, out of the money) durchgerechnet.

Die Wahl der Parameter, die in Tabelle 1 ersichtlich ist, geht zurück auf Imai und Tan [22], die sich in ihrer Arbeit rein auf das *Latin Supercube Sampling* beziehen. Dabei wird eine Asiatische Option auf einen Basket mit  $m = 10$  zugrundeliegenden Assets an  $n = 250$  Zeitpunkten betrachtet. Die Dimension des gesamten Problems ist also von der Größe  $m \cdot n = 2500$ . Dementsprechend groß ist damit auch die Kovarianzmatrix  $\tilde{\Sigma}$ .

Im Folgenden werden nun die Ergebnisse der einzelnen Verfahren präsentiert, sowie Ideen zur Implementation derselben gegeben. Es gibt wie bereits angesprochen immer vier Durchläufe, wobei die Ergebnisse jeweils für die Stichprobengrößen  $q = 2048$  und  $q = 8192$  genau präsentiert werden und die übrigen Werte nur zur Veranschaulichung der Konvergenz dienen. Es werden jeweils 10 Simulationen durchgeführt, von denen der Mittelwert genommen wird. Die Ausdrücke in den Klammern sind die gegebenen Standardabweichungen gemessen an den 10 Simulationen.

Parameter	Wert des Parameters
Laufzeit	$T = 1$ Jahr
Zinsrate	$r = 0.04$
Anfangswert der Aktie	$S_0^i = 100, i = 1, \dots, m$
Gewichte	$\omega_{ij} = \frac{1}{nm}, i, j = 1, \dots, m$
Korrelation	$\rho_{ij} = 0.4$ bzw $\rho_{ij} = 0, 1 \leq i < j \leq m$
Volatilität	$\sigma_i = 0.1 + \frac{i-1}{m-1}0.4, i = 1, \dots, m$
Strike	$K = 90, 100, 110$

Tabelle 1: Parameter für die Asiatische Basket Option

Zusätzlich werden die Pfade einmal mit der Cholesky-Zerlegung und einmal mit der *Principal Components Construction* berechnet, wobei, wie erwartet, aber nur im Falle des *Latin Hypercube Sampling* und des *Latin Supercube Samplings* eine wirkliche Varianzreduktion zu erkennen ist. Für die Monte Carlo Simulation und das *Latin Supercube Sampling* werden beide Zerlegungen zum Vergleich angegeben, bei den übrigen wird jeweils nur die bessere gezeigt. Abschließend wird noch der Vergleich über alle angegebenen Verfahren gezogen.

### Monte Carlo Simulation

Die Implementation der klassischen Monte Carlo Simulation erfolgt über die Generierung von  $q$  voneinander unabhängigen ident  $N(0, 1)$ -verteilten Zufallsvariablen der Dimension  $n \cdot m$ . Für jeden der Zufallsvektoren wird der Payoff bzw. der Preis der Asiatischen Basket Option berechnet und anschließend über all diese Werte gemittelt.

Nachdem hier willkürlich gewählte Punkte verwendet werden, ist die Erwartungshaltung gegenüber einer verhältnismäßig kleinen Standardabweichung eher gering. Den Beweis liefert der Vergleich der Werte in Tabelle 2 mit denen der übrigen Verfahren.

### Monte Carlo Simulation und Control Variates

Um eine höhere Genauigkeit der Zufallsvektoren zu erzielen, wird die Technik der *Control Variates* in Verbindung mit einer Monte Carlo Simulation angewandt. Als Kontrollvariable dient der Preis einer geometrischen Asiatischen Basket Option. Diese Kombination ist deshalb sehr vielversprechend, da der Zusammenhang zwischen geometrischen und arithmetischen Asiatischen Optionen sehr hoch ist bei Verwendung der gleichen Parameter.

Stichprobengröße	Methode	Strike	$\rho = 0$	$\rho = 0.4$
2048	Cholesky	$K = 90$	11.5891 (0.12274)	12.2334 (0.25799)
		$K = 100$	3.4309 (0.10493)	5.6773 (0.20071)
		$K = 110$	0.3652 (0.02943)	2.0463 (0.12312)
2048	PCC	$K = 90$	11.6043 (0.12956)	12.3503 (0.18893)
		$K = 100$	3.4578 (0.10147)	5.6960 (0.14372)
		$K = 110$	0.3754 (0.03648)	2.0653 (0.11755)
8192	Cholesky	$K = 90$	11.5859 (0.03720)	12.2930 (0.11666)
		$K = 100$	3.4428 (0.02848)	5.6268 (0.08366)
		$K = 110$	0.3760 (0.01356)	2.0133 (0.02936)
8192	PCC	$K = 90$	11.5981 (0.05861)	12.3328 (0.07782)
		$K = 100$	3.4587 (0.04231)	5.6653 (0.07137)
		$K = 110$	0.3733 (0.01156)	2.0470 (0.05643)

Tabelle 2: Ergebnisse des Monte Carlo Verfahrens

Die Zufallsvariablen  $Y_i, i = 1, \dots, q$  aus Gleichung (3.4) werden von den simulierten Payoffs der arithmetischen Basket Option verkörpert,

$$Y_i = e^{-rT} \left( \frac{1}{mn} \sum_{k=1}^{mn} \bar{S}_{t_k}^i - K \right)^+, \quad (7.31)$$

wobei  $\bar{S}_{t_k}^i$  dem Ausdruck  $g(\tilde{W})$  aus Gleichung (7.28) entspricht und nur die Gewichtung, die für alle  $k$  immer  $\omega = \frac{1}{mn}$  ist, aus dem Exponenten vor die Summe gezogen wird.

Die kontrollierenden Zufallsvariablen  $X_i, i = 1, \dots, q$  sind von der Beschaffenheit

$$X_i = e^{-rT} \left( \left( \prod_{k=1}^{mn} \bar{S}_{t_k}^i \right)^{\frac{1}{mn}} - K \right)^+.$$

Die Konstante  $b$ , die für die Berechnung der Stichprobenwerte

$$Y_i - b \cdot (X_i - \mathbb{E}[X]) \quad (7.32)$$

benötigt wird und deren optimaler Wert  $b^*$  aus Gleichung (3.5) nicht exakt bestimmbar ist, wird geschätzt durch

$$\hat{b} = \frac{\sum_{i=1}^q (Y_i - \bar{Y})(X_i - \bar{X})}{\sum_{i=1}^q (X_i - \bar{X})^2}.$$

Als letztes Element fehlt noch der Erwartungswert der Kontrollvariable. Der Vorteil gegenüber der arithmetischen Asiatischen Call-Option ist, dass der Preis  $C_{GA}$  der geometrischen Version analytisch mithilfe der Black-Scholes Formel, ähnlich wie in Gleichung (7.9) bestimmbar ist.

Unter der Annahme von  $T = 1$  kann der Preis zum Zeitpunkt  $t_0 = 0$  berechnet werden mit

$$C_{GA}(0) = e^{-r} \left( S_0 e^{\tilde{\mu} + \frac{\tilde{\sigma}^2}{2}} \cdot \Phi \left( \frac{\ln \frac{S_0}{K} + \tilde{\mu}}{\tilde{\sigma}} + \tilde{\sigma} \right) - K \cdot \Phi \left( \frac{\ln \frac{S_0}{K} + \tilde{\mu}}{\tilde{\sigma}} \right) \right), \quad (7.33)$$

die Parameter  $\tilde{\mu}$  und  $\tilde{\sigma}$  sind durch

$$\tilde{\mu} = \frac{1}{mn} \sum_{k=1}^{mn} \frac{r - \frac{(\sigma^k)^2}{2}}{2} \quad (7.34)$$

und

$$\tilde{\sigma} = \sqrt{\frac{2mn + 1}{6(mn + 1)}} \sqrt{\frac{1}{(mn)^2} \sum_{i=1}^{mn} \sum_{k=1}^{mn} \sigma^k \sigma^i \rho_{ik}} \quad (7.35)$$

gegeben. Eine detaillierte Herleitung der Black Scholes Formel für geometrische Asiatische Optionen kann bei Zhang [64] gefunden werden.

Stichprobengröße	Methode	Strike	$\rho = 0$	$\rho = 0.4$
2048	Cholesky	$K = 90$	11.5909 (0.02432)	12.3088 (0.12866)
		$K = 100$	3.4375 (0.02626)	5.6304 (0.01649)
		$K = 110$	0.3675 (0.01402)	2.0141 (0.01442)
8192	Cholesky	$K = 90$	11.5939 (0.01063)	12.3195 (0.00927)
		$K = 100$	3.4368 (0.01180)	5.6417 (0.00716)
		$K = 110$	0.3704 (0.00091)	2.0223 (0.00617)

Tabelle 3: Ergebnisse des Monte Carlo Verfahrens mit Kontrollvariablen

### Quasi-Monte Carlo Simulation

Bei der klassischen Quasi-Monte Carlo Simulation macht es wenig Sinn für jede Stichprobengröße  $n$  mehrere Durchläufe zu rechnen, da die dahinter liegende Punktfolge, hier die Halton-Folge, immer dieselbe ist und dadurch auch immer dasselbe Ergebnis resultiert. Aus diesem Grund sind die Werte schlecht vergleichbar mit den übrigen Verfahren, da keine Standardabweichung berechnet werden kann bzw. diese immer Null ergeben würde. Betrachtet man aber dennoch die Payoffs aus Tabelle 4 und vergleicht sie mit denen der anderen Verfahren, so ist eine Annäherung bei wachsender Stichprobengröße durchaus erkennbar, auch wenn dies nur sehr langsam erfolgt.

Stichprobengröße	Methode	Strike	$\rho = 0$	$\rho = 0.4$
2048	PCC	$K = 90$	11.6508	12.3063
		$K = 100$	3.4951	5.6364
		$K = 110$	0.3488	2.0302
8192	PCC	$K = 90$	11.6081	12.3154
		$K = 100$	3.4593	5.6505
		$K = 110$	0.3646	2.0321

Tabelle 4: Ergebnisse der Quasi-Monte Carlo Simulation

Stichprobengröße	Methode	Strike	$\rho = 0$	$\rho = 0.4$
2048	PCC	$K = 90$	11.6481 (0.11117)	12.3971 (0.21167)
		$K = 100$	3.4986 (0.07032)	5.7038 (0.17044)
		$K = 110$	0.3828 (0.02403)	2.1020 (0.14355)
8192	PCC	$K = 90$	11.6043 (0.06482)	12.3140 (0.09861)
		$K = 100$	3.4540 (0.04529)	5.6506 (0.08244)
		$K = 110$	0.3728 (0.01310)	2.0635 (0.07426)

Tabelle 5: Ergebnisse der randomisierten Quasi-Monte Carlo Simulation

### Randomisierte Quasi-Monte Carlo Simulation

Um das Ergebnis bei Verwendung von Quasi-Monte Carlo Punktfolgen repräsentativer zu machen, wird nun ein Hybrides Verfahren zur Anwendung gebracht. Die deterministische Halton-Folge wird mithilfe der Cranley-Patterson Rotation, die in Kapitel 2.2.3 vorgestellt wurde, in eine Zufallszahlenfolge umgewandelt. Davon können wieder 10 verschiedene voneinander unabhängige Durchläufe simuliert werden, von denen die Stichprobenstandardabweichung berechnet werden kann. Die Ergebnisse sind in Tabelle 5 zu sehen.

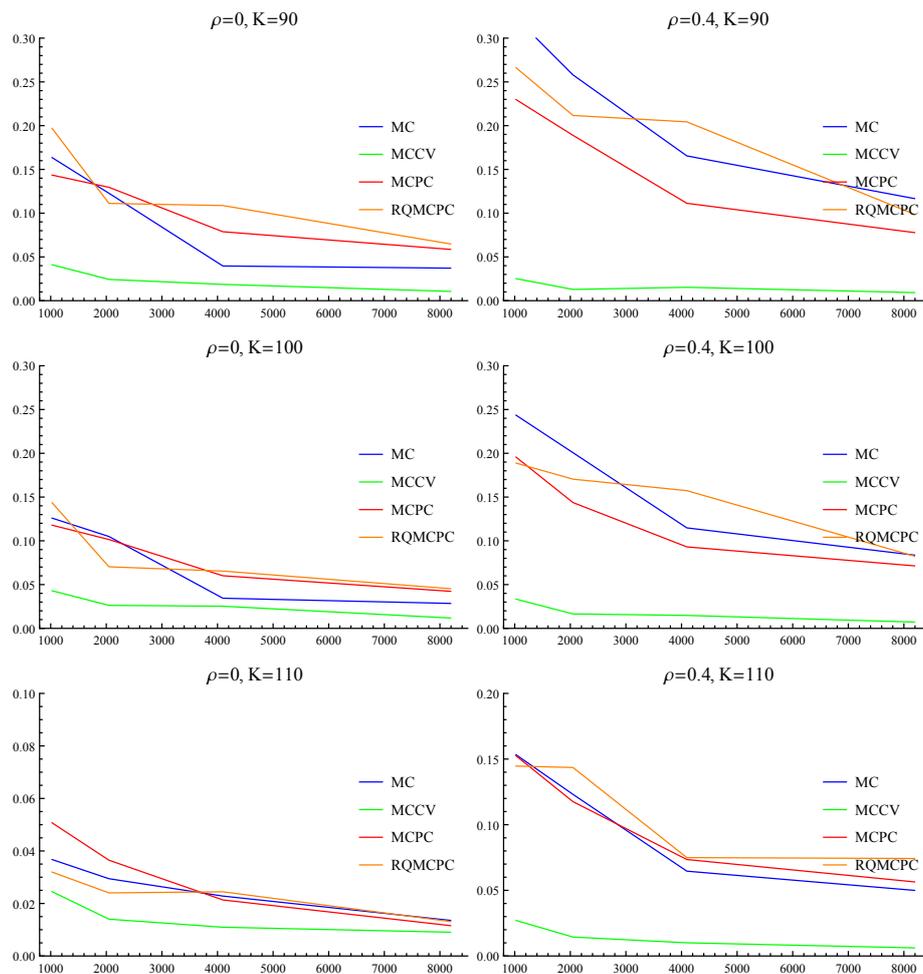


Abbildung 7.4: Vergleich der Verfahren: Monte Carlo mit Cholesky-Zerlegung (MC), Monte Carlo mit Principal Components Construction (MCPC) und randomisiertes Quasi-Monte Carlo Verfahren (RQMC).

Abbildung 7.4 zeigt den Vergleich der Monte Carlo basierten Verfahren und dem randomisierten Quasi-Monte Carlo Verfahren gemessen an der Standardabweichung. Deutlich zu erkennen ist die Varianzreduktion bei Verwendung von Kontrollvariablen und, dass die *Principal Components Construction* im Zusammenhang mit Monte Carlo keine wesentliche Verbesserung gegenüber der Cholesky Zerlegung darstellt. Trotz der hohen Dimension liegt das Hybride Verfahren bei Verwendung der Eigenwert-Eigenvektor-Zerlegung nicht allzu weit hinter dem Monte Carlo Verfahren zurück. Im Folgenden wird aus Gründen der Übersicht nur noch die Monte Carlo Simulation mit Kontrollvariablen als bestes dieser vier Varianten zum Vergleich herangezogen.

### Latin Hypercube Sampling

Im *Latin Hypercube Sampling* werden zuerst  $q$  unabhängige und im Bereich  $(0, 1)$  gleichverteilte Zufallsvektoren der Dimension  $n \cdot m$  generiert. Anschließend wird auf jede Dimension der Permutationsschritt aus (4.2) angewandt. Mithilfe der Inversionsmethode aus (2.5) können die Einträge unter Anwendung der Inversen Verteilungsfunktion auf eine Standardnormalverteilung transformiert werden und haben somit wieder die gewünschte Form zur Berechnung der Preise, die in Tabelle 6 ersichtlich sind.

Stichprobengröße	Methode	Strike	$\rho = 0$	$\rho = 0.4$
2048	PCC	$K = 90$	11.5945 (0.00591)	12.3216 (0.01160)
		$K = 100$	3.4430 (0.03616)	5.6644 (0.02510)
		$K = 110$	0.3559 (0.01753)	2.0411 (0.01992)
8192	PCC	$K = 90$	11.5915 (0.00361)	12.3318 (0.00596)
		$K = 100$	3.4443 (0.01321)	5.6575 (0.00776)
		$K = 110$	0.3703 (0.00866)	2.0333 (0.00612)

Tabelle 6: Ergebnisse des Latin Hypercube Samplings

Da die Berechnung der Inversen Standardnormalverteilung einen recht zeitintensiven Prozess darstellen kann, gibt es bereits gute Approximationen mittels Polynomen, die in den meisten Fällen ausreichend sind, was die Genauigkeit betrifft und wovon eine im Folgenden angeführt wird. Diese können unter anderem im Handbuch von Abramowitz und Stegun [1] gefunden werden.

Die Inverse Standardnormalverteilung kann für  $x \in (0, 1)$  approximiert werden durch

$$\Phi^{-1}(p) \approx \begin{cases} -p \left( \sqrt{\ln \frac{1}{x^2}} \right), & x \in (0, \frac{1}{2}] \\ p \left( \sqrt{\ln \frac{1}{(1-x)^2}} \right), & x \in (\frac{1}{2}, 1) \end{cases}$$

wobei

$$p(t) = t - \frac{c_0 + c_1 t + c_2 t^2}{1 + d_1 t + d_2 t^2 + d_3 t^3} + \epsilon, \quad \text{mit } |\epsilon| < 4.5 * 10^{-4}.$$

### Latin Hypercube Sampling with Dependence

Analog zum Monte Carlo Verfahren werden  $q$  voneinander unabhängige ident  $N(0, 1)$ -verteilten Zufallsvariablen der Dimension  $n \cdot m$  erzeugt und mit der Matrix  $A$ , die aus der Zerlegung  $AA^t = \tilde{\Sigma}$  resultiert, multipliziert. Danach erfolgt der Schritt, in dem die dadurch entstandenen Vektoren Dimension für Dimension sortiert und mithilfe von (5.2) transformiert werden.

Im Anschluss müssen die Einträge wie bei der Anwendung des *Latin Hypercube Samplings* wieder auf Normalverteilung transformiert werden. Diese ist aber hier keine Standardnormalverteilung, sondern vielmehr eine Normalverteilung mit einer Varianz von  $j^T \sigma_i^2$  für  $j = 1, \dots, n, i = 1, \dots, m$ , die zu berücksichtigen ist. Diese Transformation ist mitunter einer der Prozesse, die am meisten Rechenaufwand haben und viel Zeit beanspruchen. Dafür kann sich das Ergebnis in Tabelle 7 durchaus sehen lassen, übertrifft es im Punkt Standardabweichung sowohl das Monte Carlo als auch das herkömmliche *Latin Hypercube Sampling*.

Stichprobengröße	Methode	Strike	$\rho = 0$	$\rho = 0.4$
2048	Cholesky	$K = 90$	11.5901 (0.00562)	12.3376 (0.02537)
		$K = 100$	3.4460 (0.02710)	5.6759 (0.04047)
		$K = 110$	0.3612 (0.02213)	2.0413 (0.04097)
8192	Cholesky	$K = 90$	11.5910 (0.00228)	12.3274 (0.01386)
		$K = 100$	3.4417 (0.01378)	5.6498 (0.02226)
		$K = 110$	0.3644 (0.01079)	2.0217 (0.01887)

Tabelle 7: Ergebnisse des Latin Hypercube Samplings with Dependence

### Latin Supercube Sampling

Anstatt  $q$  Quasi-Monte Carlo Folgen von Dimension  $n \cdot m = 2500$  zu erzeugen, wie es bei einer klassischen Quasi-Monte Carlo Simulation vonnöten wäre, werden hier 100 mal  $q$  Punkte der Dimension 25 generiert, die nach Permutation das *Latin Supercube Sampling* wie in (6.1) bilden.

Als erstes kommt die Sobol-Folge zur Anwendung, da sie sich auch schon in vielen anderen finanzmathematischen Anwendungen wie zum Beispiel bei der Berechnung der *Collateralized Mortgage Obligation* aus Abschnitt 7.2 bewährt hat.

Stichprobengröße	Methode	Strike	$\rho = 0$	$\rho = 0.4$
2048	Cholesky	$K = 90$	11.6505 (0.02523)	12.1492 (0.08775)
		$K = 100$	3.5158 (0.03600)	5.6852 (0.09153)
		$K = 110$	0.3875 (0.02607)	2.0620 (0.08846)
2048	PCC	$K = 90$	11.6036 (0.00064)	12.2390 (0.03401)
		$K = 100$	3.4508 (0.00170)	5.6921 (0.00153)
		$K = 110$	0.3673 (0.00104)	2.0604 (0.00140)
8192	Cholesky	$K = 90$	11.5429 (0.01733)	12.3350 (0.03789)
		$K = 100$	3.4200 (0.02961)	5.6055 (0.04534)
		$K = 110$	0.3695 (0.00813)	2.0198 (0.04269)
8192	PCC	$K = 90$	11.5906 (0.00025)	12.3255 (0.00031)
		$K = 100$	3.4457 (0.00124)	5.6526 (0.00074)
		$K = 110$	0.3602 (0.00074)	2.0320 (0.00051)

Tabelle 8: Ergebnisse des Latin Supercube Samplings unter Verwendung der Sobol-Folge

In Tabelle 8 ist die deutliche Reduktion der Standardabweichung des Verfahrens unter Verwendung der *Principal Components Construction* im Gegensatz zur Cholesky-Zerlegung zu sehen. Das liegt an der effektiven Dimension des Problems. Mithilfe der folgenden zwei Gleichungen kann berechnet werden, wieviel Anteil der Varianz durch die ersten  $p = 1, \dots, 10$  Komponenten erklärt werden. Sie folgen direkt als Varianz von  $\tilde{g} = g - K$  in Gleichung (7.28).

$$\sigma^2(\tilde{g}) = \sum_{i=1}^{mn} \sum_{j=1}^{mn} \exp \left[ \mu_i + \mu_j + \frac{1}{2} \sum_{l=1}^{mn} (a_{il}^2 + a_{jl}^2) \right] \left( \exp \left[ \sum_{l=1}^{mn} a_{il} a_{jl} \right] - 1 \right) \quad (7.36)$$

beschreibt die gesamte Varianz von  $\tilde{g}$ , während

$$\sigma_p^2(\tilde{g}) = \sum_{i=1}^{mn} \sum_{j=1}^{mn} \exp \left[ \mu_i + \mu_j + \frac{1}{2} \sum_{l=1}^p (a_{il}^2 + a_{jl}^2) \right] \left( \exp \left[ \sum_{l=1}^p a_{il} a_{jl} \right] - 1 \right) \quad (7.37)$$

den Beitrag der ersten  $p$  Dimensionen zur Varianz beschreibt. Der Anteil der ersten  $p$  Komponenten resultiert als Quotient dieser beiden Größen. Die Einträge  $a_{ij}$  für  $i, j = 1, \dots, mn$  sind die Einträge der Matrix, die aus der jeweiligen Zerlegung  $AA^t = \tilde{\Sigma}$  entsteht.

Während bei der klassischen Methode gerade einmal ein Prozent der Varianz durch die ersten 10 Dimensionen erklärt ist, sind es bei der *Principal Components Construction* schon über 95 Prozent im unkorrelierten Fall und sogar fast 99 Prozent bei einer Korrelation von  $\rho = 0.4$ .

Zusätzlich wurde das *Latin Supercube Sampling* noch einmal zum Vergleich mit der Halton-Folge durchgeführt, wobei ein ähnlich gutes Ergebnis resultiert.

Stichprobengröße	Methode	Strike	$\rho = 0$	$\rho = 0.4$
2048	PCC	$K = 90$	11.5890 (0.00534)	12.3308 (0.01612)
		$K = 100$	3.4934 (0.00155)	5.6314 (0.00384)
		$K = 110$	0.3480 (0.00105)	2.0305 (0.00106)
8192	PCC	$K = 90$	11.5903 (0.00365)	12.3286 (0.00587)
		$K = 100$	3.4596 (0.00111)	5.6501 (0.00083)
		$K = 110$	0.3647 (0.00062)	2.0323 (0.00054)

Tabelle 9: Ergebnisse des Latin Supercube Samplings unter Verwendung der Halton Folge

## Fazit

Beide Matrixzerlegungen hängen von der Beschaffenheit der Kovarianzmatrix  $\tilde{\Sigma}$  und demzufolge auch von der Korrelation  $\rho$  ab. Das Ergebnis im Fall von  $\rho = 0.4$  weist im Bezug auf die Cholesky Zerlegung generell höhere Werte in der Standardabweichung als im unkorrelierten Fall auf. Bei Zerlegung in Eigenwerte und Eigenvektoren ist diese Eigenschaft nicht mehr in allen Verfahren so eindeutig gegeben, im Falle des *Latin Supercube Samplings* hat sich das Blatt sogar gewendet.

Abgesehen vom *Latin Supercube Sampling* ist nur noch im *Latin Hypercube Sampling* eine Verbesserung der Varianz unter Verwendung der *Principal Components Construction* bemerkbar. Dies lässt sich darauf zurückführen, dass das *Latin Hypercube* Verfahren, ein  $(t, m, d)$ -Netz

darstellt und damit zu den Quasi-Monte Carlo Verfahren zählt. Wie bereits oft angesprochen, hängt die Performance bei solchen sehr wohl davon ab, wieviel Variabilität von den ersten paar Dimensionen getragen wird, während dies die übrigen Verfahren, die mit rein zufälligen Zahlen arbeiten, nicht unbedingt beeindruckt.

Der Vergleich der Ergebnisse bezüglich der Anzahl der simulierten Pfade ist genau wie erwartet. Im Falle des Monte Carlo Verfahrens ist im Schnitt durchaus eine Reduktion der Standardabweichung bei Erhöhung der Stichprobenanzahl ersichtlich, egal welche Matrixzerlegung gewählt wird. In den übrigen Verfahren, welche die Punkte gleichmäßiger auf den Raum verteilen, kann es passieren, dass die Standardabweichung, vor allem bei Simulation von einer niedrigen Anzahl an Stichprobenelementen, wie es hier der Fall ist, zwischendurch ansteigt. Das kann unter anderem daran liegen, dass der gesetzte *Seed*, der die Zufallszahlengeneration initiiert, nicht optimal gewählt ist und, wie bereits erwähnt, die Stichprobengröße noch nicht ausreichend ist.

Was aber deutlich im Vergleich der Standardabweichungen aus Abbildung 7.5 zu erkennen ist, ist die herausragende Performance des *Latin Supercube Samplings*. Während die anderen Verfahren mehr oder weniger langsam konvergieren, ist das *Latin Supercube Sampling* mit der *Principal Components Construction* schon bei 1024 Punkten bei einer Genauigkeit, welche alle übrigen noch nicht bei über 8000 Punkten erreichen. Bestätigt werden diese Ergebnisse auch von den veröffentlichten Werten von Imai und Tan [22], die zusätzlich zur Zerlegung in Eigenwerte und Eigenvektoren noch weitere sinnvolle Matrixzerlegungen präsentieren.

Aus Implementationssicht ist das klassische Monte Carlo Verfahren in seiner Einfachheit den anderen Methoden voraus. Dass es aber im Bezug auf die Varianz nicht immer auch die beste Lösung darstellt, zeigt der Vergleich mit den anderen Verfahren. Hier kann das Monte Carlo Verfahren mit Kontrollvariablen schon eher mithalten. Auch das *Latin Hypercube Sampling* ist relativ einfach zu implementieren und liegt varianztechnisch durch eine bessere Verteilung der Punkte im Raum voraus. Dass das nicht immer der Fall sein muss, hängt unter anderem von der Beschaffenheit der Zielfunktion ab und wurde bereits in Kapitel 4 ausgiebig diskutiert.

Komplexer ist die Implementierung des *Latin Hypercube Samplings with Dependence*, da die Abhängigkeitsstruktur erhalten bleiben soll und anstatt unabhängigen Zufallsvektoren aus der Standardnormalverteilung, speziell korrelierte Einträge zu berechnen sind. In diesem Beispiel liefert es kein ausgezeichnetes Ergebnis und gerade dort, wo eine Korrelation der Zufallsvariablen vorgesehen ist, sind die Werte im Vergleich zu den anderen Verfahren eher enttäuschend, auch wenn sie trotz alledem besser als jene der Monte Carlo Simulation sind. Beachtet man allerdings den Umstand, dass, im Falle von abhängigen Zufallsvariablen, die nicht gerade mit einer Gauß'schen Copula oder multivariabel normalverteilt sind, in alle Verfahren noch diese Abhängigkeitsstruktur gebracht werden muss, so entspricht das einem zusätzlichen Aufwand, wodurch sich vor allem in Bezug auf die Laufzeit ein deutlicher Vorteil für das *Latin Hypercube*

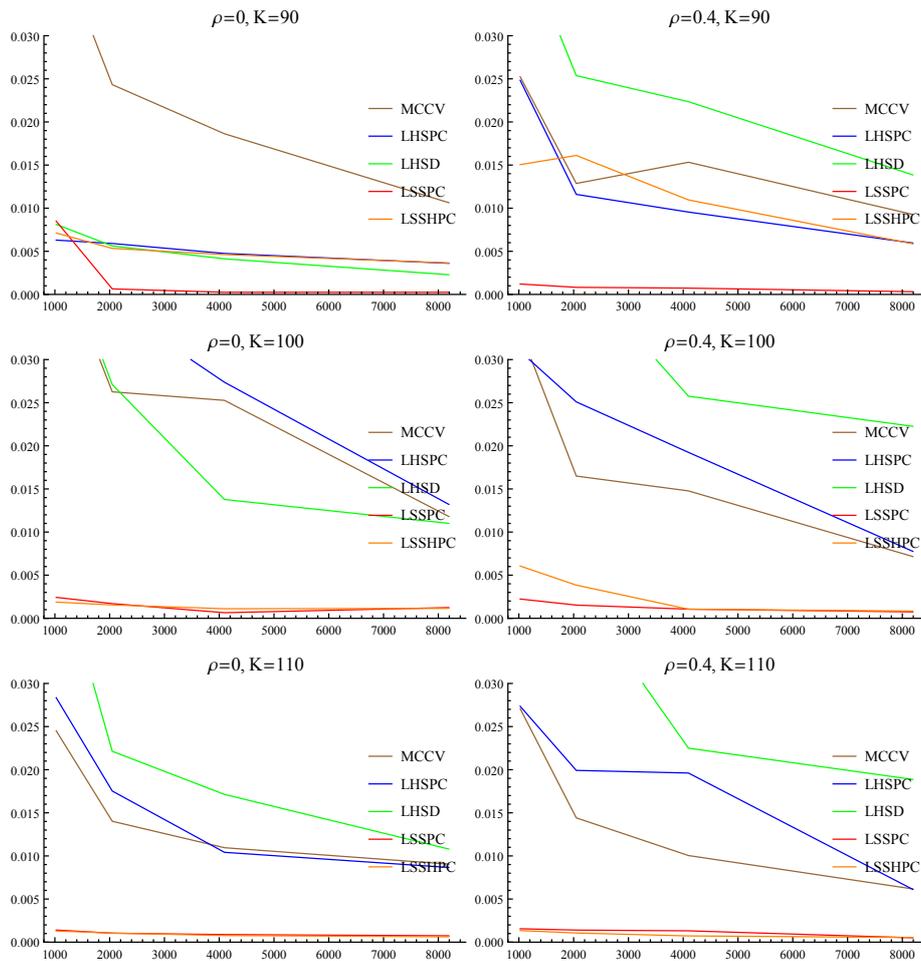


Abbildung 7.5: Vergleich der Verfahren: Monte Carlo mit Kontrollvariablen (MCCV), Latin Hypercube (LHSPC), Latin Hypercube with Dependence (LHSD), Latin Supercube mit Sobol-Folge (LSSPC), Latin Supercube mit Halton Folge (LSSHPC)

*Samplings with Dependence* ergibt.

Das *Latin Supercube Sampling* überzeugt vor allem in Kombination mit Methoden, wie der *Principal Components Construction*, die die effektive Dimension des Problems reduzieren oder dafür sorgen, dass die Dimensionen, in denen die Quasi-Monte Carlo Punkte die größte Regelmäßigkeit aufweisen, bevorzugt für die Erklärung der Varianz zuständig sind. Die Implementierung des Verfahren selbst benötigt wenige Zeilen, die Schwierigkeit liegt in der Simulation der mehrdimensionalen Quasi-Monte Carlo Punkte, von denen allerdings die gängigsten schon in den meisten mathematischen Programmen implementiert sind und verwendet werden können.

Hier wurden sowohl Sobol- als auch Halton-Folgen verwendet, die beide im Schnitt ein deutlich besseres Ergebnis als alle anderen Simulationsmethoden liefern. Im Vergleich untereinander hat aber, wie vor allem im Fall von  $K = 90$  ersichtlich, die Sobol-Folge die Nase vorne.

	Strike	MC	RQMC	MCCV	LHS	LHSD	LSS
$\rho = 0$	$K = 90$	0.03720	0.06482	0.01063	0.00361	0.00228	0.00025
	$K = 100$	0.02848	0.04529	0.01180	0.01321	0.01378	0.00124
	$K = 110$	0.01356	0.01310	0.00091	0.00866	0.01079	0.00074
Laufzeit		33.17	37.9	45.33	37.83	57.83	37.33
$\rho = 0.4$	$K = 90$	0.11666	0.09861	0.00927	0.00596	0.01386	0.00031
	$K = 100$	0.08366	0.08244	0.00716	0.00776	0.02226	0.00074
	$K = 110$	0.02936	0.07426	0.00617	0.00612	0.01887	0.00051

Tabelle 10: Standardabweichungen und Laufzeiten (in Minuten) für die Stichprobengröße  $n = 8192$ .

Um eine endgültige Aussage über die Effektivität der Verfahren treffen zu können werden noch einmal die Standardabweichungen in Verbindung mit den Laufzeiten gebracht und verglichen. Tabelle 10 zeigt die Standardabweichungen bei  $n = 8192$  und die Laufzeit bei dieser Punktezahlanzahl in Minuten. Alle Simulationen wurden mit dem Programm *Wolfram Mathematica 9.0* durchgeführt. Am besten schneidet wiederum das *Latin Supercube Sampling*, gefolgt vom *Latin Hypercube Sampling*, ab. Beide Methoden benötigen kaum mehr Zeitaufwand als das reine Monte Carlo Verfahren. Das *Latin Hypercube Sampling with Dependence* ist mit fast doppelt so langer Laufzeit das langsamste von allen. Am schlechtesten sind, wie erwartet, die Werte des randomisierten Quasi-Monte Carlo Verfahren, was auch die durchschnittliche Laufzeit nicht ändert.

Diese Auswertung zeigt keinesfalls, dass für jedes Problem das *Latin Supercube Sampling* als optimales Verfahren gewählt werden sollte. Alle vorgestellten Methoden haben ihre Vor- und Nachteile und sind meist abhängig von der zu integrierenden Funktion, der effektiven Dimension, diversen Matrixzerlegungen oder sonstigen Parametern. Ein universelles Rezept, für das beste Verfahren gibt es nicht und es sind oft mehrere Tests und Erfahrung notwendig um die optimale Varianzreduktionstechnik zu finden.

## Literatur

- [1] M. Abramowitz and I. A. Stegun. *Handbook of Mathematical Functions*. United States Department of Commerce, 1964.
- [2] C. Adams and J. Clarkson. Properties of functions  $f(x,y)$  of bounded variation. *Transactions of the American Mathematical Society*, 36(4):711–730, 1934.
- [3] 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*, 15(7):20, 2012.
- [4] F. Akesson and J. Lehocky. Discrete eigenfunction expansion of multidimensional brownian motion and the ornstein-uhlenbeck process. Technical report, Department of Statistics, Carnegie Mellon University, Pittsburgh, 1998.
- [5] I. A. Antonov and V. M. Saleev. An economic method of computing lp sequences. *USSR Computational Mathematics and Mathematical Physics*, 19(1):252–256, 1979.
- [6] J. Beck. A two-dimensional van aardenne-ehrenfest theorem in irrregularities of distribution. *Compositio Mathematica*, 72:269–339, 1989.
- [7] R. BÉjjan. Mintoation de la discrÉpance d’une suite quelconque sur t. *Acta Arithmetica*, 41:185–202, 1982.
- [8] D. Bilyk, M. T. Lacey, and A. Vagharshakyan. On the small ball inequality in all dimensions. *Journal of Functional Analysis*, 254:2470–2502, 2008.
- [9] R. E. Caflisch, W. Morokoff, and A. Owen. Valuation of mortgage backed securities using brownian bridges to reduce effective dimension. *Computational Finance*, 1:27–46, 1997.
- [10] R. Cont and P. Tankov. *Financial Modelling with Jump Processes*. Chapman and Hall/CRC, 2004.
- [11] R. Cranley and T. N. L. Patterson. Randomization of the number theoretic methods for numerical integration. *SIAM Journal of Numerical Analysis*, 13:904–914, 1976.
- [12] L. de Clerck. *De exacte berekening van de sterdiscrepantie van de rijen van Hammersley in 2 dimensies*. PhD thesis, Katholieke Universiteit Leuven, 1984.
- [13] P. Deheuvels. La fonction de dépendance empirique et ses propriétés. un test non paramétrique d’indépendance. *Bulletin de la Classe des Sciences, V.Série, Académie Royale de Belgique*, 65:274–292, 1979.

- [14] M. Drmota and R. F. Tichy. *Sequences, Discrepancies and Applications, Lecture Notes in Mathematics 1651*. Springer, 1997.
- [15] H. Faure. Discrepancies de suites associées à un système de numération (en dimension  $s$ ). *Acta Arithmetica*, 41(4):337–351, 1982.
- [16] W. Feller. *An introduction to probability theory and its applications*. John Wiley and Sons, 1971.
- [17] J. D. Fermanian, D. Radulovic, and M. Wegkamp. Weak convergence of empirical copula processes. *Journal of the Bernoulli Society*, 10(5):847–860, 2004.
- [18] P. Glasserman. *Monte Carlo Methods in Financial Engineering*. Springer, 2004.
- [19] J. Halton. On the efficiency of certain quasi-random points in evaluating multidimensional integrals. *Numerische Mathematik*, 2:84–90, 1960.
- [20] E. Hlawka. Funktionen von beschränkter Variation in der Theorie der Gleichverteilung. *Annali di Matematica Pura ed Applicata*, 54:325–333, 1961.
- [21] R. A. Horn and C. R. Johnson. *Topics in Matrix Analysis*. Cambridge University Press, 1994.
- [22] J. Imai and K. S. Tan. A general dimension reduction technique for derivative pricing. *Journal of Computational Finance*, 10(2):129–155, 2006.
- [23] P. Jaeckel. *Monte Carlo Methods In Finance*. Wiley, 2002.
- [24] J. L. W. V. Jensen. Sur les fonctions convexes et les inégalités entre les valeurs moyennes. *Acta Mathematica*, 30:175–193, 1906.
- [25] H. Joe. *Multivariate Models and Dependence Concepts*. CRC Press, 1997.
- [26] O. Kallenberg. *Foundations of Modern Probability*. Springer, 2001.
- [27] I. Karatzas and S. E. Shreve. *Methods of Mathematical Finance*. Springer, 1998.
- [28] A. G. L. Kemna and V. A. C. F. A pricing method for options based on average asset values. *Journal of Banking and Finance*, 14:113–129, 1990.
- [29] R. Korn, E. Korn, and G. Kroisandt. *Monte Carlo Methods and Models in Finance and Insurance*. CRC Press, 2010.
- [30] P. L'Ecuyer. Efficient and portable combined random number generators. *Communications of the ACM*, 31(6):742–751, 1988.
- [31] P. L'Ecuyer. Tables of linear congruential generators of different sizes and good lattice structure. *Mathematics of Computation*, 68:249–260, 1999.

- [32] D. H. Lehmer. Mathematical methods in large-scale computing units. *Proceedings of the Second Symposium on Large Scale Digital Computing Machinery, Harvard University Press*, pages 141–146, 1951.
- [33] W. L. Loh. On latin hypercube sampling. *The Annals of Statistics*, 24(5):2058–2080, 1996.
- [34] B. M and S. F. The pricing of options and corporate liabilities. *Journal of Political Economy*, 81(3):637–654, 1973.
- [35] M. D. McKay, R. J. Beckman, and W. J. Conover. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21(2):239–245, 1979.
- [36] B. Moskowitz and R. E. Caflisch. Smoothness and dimension reduction in quasi-monte calo methods. *Mathematical and Computer Modelling*, 23(8/9):37–54, 1996.
- [37] K. Neammanee and P. Rattanawong. A uniform bound on the generalization of a combinatorial central limit theorem. *International Mathematical Forum*, 2008.
- [38] R. B. Nelsen. *An Introduction to Copulas*. Springer, 1999.
- [39] H. Niederreiter. Discrepancy and convex programming. *Annali di Matematica Pura ed Applicata*, 1972.
- [40] H. Niederreiter. Point sets and sequences with small discrepancy. *Monatshefte für Mathematik*, 104:273–337, 1987.
- [41] H. Niederreiter. *Random Number Generation and Quasi-Monte Carlo Methods*. CBMS-NFS regional conference series in applied mathematics 63, SIAM Philadelphia, 1992.
- [42] G. Okten. A probabilistic result on the discrepancy of a hybrid-monte carlo sequence and applications. Technical report, Mathematics Department, Claremont Graduate School, 1996.
- [43] A. B. Owen. A central limit theorem for latin hypercube sampling. *Journal of the Royal Statistical Society*, 54(2):541–551, 1992.
- [44] A. B. Owen. Lattice sampling revisited: Monte carlo variance of means over randomized orthogonal arrays. *The Annals of Statistics*, 22:930–945, 1994.
- [45] A. B. Owen. Randomly permuted  $(t, m, s)$ -nets and  $(t, s)$ -sequences. *Monte Carlo Methods in Scientific Computing*, 106:299–317, 1995.
- [46] A. B. Owen. Latin supercube sampling for very high dimensional simulation. *ACM Transactions on Modeling and Computer Simulation*, 8(1):71–102, 1997.

- [47] A. B. Owen. Monte carlo variance of scrambled equidistribution quadrature. *SIAM Journal of Numerical Analysis*, 34(5):1884–1910, 1997.
- [48] N. Packham and W. Schmidt. Latin hypercube sampling with dependence and applications in finance. *Frankfurt School of Finance & Management*, 15:1–23, 2008.
- [49] S. H. Paskov. Computing high dimensional integrals with applications to finance. Technical report, Joint Summer Research Conference on Continuous Algorithms and Complexity, Mount, 1994.
- [50] S. H. Paskov. New methodologies for valuing derivatives. Technical report, Department of Computer Science, Columbia University, 1996.
- [51] S. H. Paskov and J. F. Traub. Faster valuation of financial derivatives. *Journal of Portfolio Management*, 22(1):113–123, 1995.
- [52] H. D. Patterson. The errors of lattice sampling. *Journal of the Royal Statistical Society*, 16:140–149, 1954.
- [53] N. Rerkruthairat and K. Neammanee. Uniform bound on normal approximation of latin hypercube sampling. *Journal of Mathematics Research*, 3(2):98–106, 2011.
- [54] K. F. Roth. On irregularities of distribution. *Mathematika*, 1:73–79, 1954.
- [55] W. M. Schmidt. Irregularities of distribution iv. *Acta Arithmetica*, 21:45–50, 1972.
- [56] W. Schoutens. *L $\tilde{A}$ vy Processes in Finance*. John Wiley and Sons Ltd, 2003.
- [57] M. E. Silva and B. T. Quasi-monte carlo in finance: extending for problems of high effective dimension. *Economia Aplicada*, 9(4):577–594, 2005.
- [58] A. Sklar. Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de statistique de l'Université de Paris*, 8:229–231, 1959.
- [59] I. M. Sobol. On the distribution of points in a cube and the approximate evaluation of integrals. *USSR Journal of Computational Mathematics and Mathematical Physics*, 7(4):86–112, 1967.
- [60] J. Spanier. Quasi-monte carlo methods for particle transport problems. *Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing*, 106:121–148, 1995.
- [61] M. Stein. Large sample properties of simulations using latin hypercube sampling. *Technometrics*, 29(2):13–151, 1987.
- [62] H. Weyl. Über die Gleichverteilung von Zahlen mod. Eins. *Mathematische Annalen*, 77:313–352, 1916.

- [63] B. A. Wichmann and I. D. Hill. An efficient and portable pseudorandom number generator. *Applied Statistics*, 31:188–190, 1916.
- [64] H. Zhang. Pricing asian options using monte carlo methods. Master's thesis, Uppsala Universitet, 2009.