Numerische Methoden der Finanzmathematik

WiSe 2009/10

J. Baumeister[1]

15. April 2010

[1]Dies sind Aufzeichnungen, die kritisch zu lesen sind, da sie noch nicht endgültig korrigiert sind, und daher auch nicht zitierfähig sind. Hinweise auf Fehler und Verbesserungsvorschläge an *baumeister@math.uni-frankfurt.de*

Inhaltsverzeichnis

Vorwort **i**

I Teil 1: Elementare Themen und Verfahrensweisen **1**

1 Die Black-Scholes-Preisformel und Berechnung der impliziten Volatilität **1**
1.1 Aktien, Werpapiere und risikofreie Anlagen . . . 1
1.1.1 Aktien . . . 1
1.1.2 Festverzinsliche Wertpapiere . . . 2
1.2 Eine Preisformel . . . 3
1.3 Volatilität . . . 8
1.3.1 Rendite und Risiko . . . 9
1.3.2 Historische Volatilität . . . 9
1.3.3 Implizite Volatilität . . . 9
1.3.4 Lokale Volatilität . . . 10
1.3.5 Stochastische Volatilität . . . 10
1.4 Auswertung der Preisformel . . . 10
1.5 Berechnung der Volatilität . . . 12
1.5.1 Historische Volatilität . . . 13
1.5.2 Implizite Volatilität . . . 13
1.6 Anhang: Newtonverfahren . . . 16
1.6.1 Nullstellensuche nach Newton bei Polynomen . . . 16
1.6.2 Konvergenz im eindimensionalen Fall . . . 19
1.6.3 Der mehrdimensionale Fall . . . 22
1.7 Bibliographische Anmerkungen . . . 23
1.8 Übungen . . . 23

2 Das Binomialmodell **29**
2.1 Marktannahmen . . . 29
2.2 Das Binomialmodell . . . 32
2.2.1 Das Ein-Perioden-Modell . . . 33
2.2.2 Das n-Perioden-Modell . . . 35
2.2.3 Die Parameter u, v, p für das n-Perioden-Modell . . . 37
2.3 Die Binomialmethode . . . 39
2.4 Anmerkungen zu Bermuda-Optionen . . . 40
2.5 Bibliographische Anmerkungen . . . 41
2.6 Übungen . . . 41

3 Zufallszahlen und ihre Erzeugung **45**
3.1 Grundsätzliches . . . 45
3.2 Kongruenzgeneratoren . . . 49
3.3 Ein Zufallsbit-Generator . . . 52
3.4 Fibonacci-Generatoren . . . 53
3.4.1 Additive lagged Fibonacci Generatoren . . . 53
3.4.2 Multiplikative lagged Fibonacci Generatoren . . . 54
3.5 Die Güte von Zufallszahlen . . . 54
3.5.1 Geometrische Korrelation . . . 54
3.5.2 Statistische Tests . . . 55
3.5.3 Der Chi-Quadrat-Test . . . 56
3.6 Anhang: Beweis des Satzes zur Periodenlänge . . . 57
3.7 Bibliographische Anmerkungen . . . 59
3.8 Übungen . . . 59

4 Simulationsgrundlagen **63**
4.1 Erzeugung von Verteilungen: Die Inversionsmethode . . . 63
4.1.1 Die Inversionsmethode . . . 63
4.1.2 Diskrete Verteilung . . . 65
4.1.3 Anwendung: Normalverteilung . . . 66
4.1.4 Ein Transformationssatz für Dichten . . . 67
4.2 Das Box-Muller-Verfahren . . . 68
4.3 Wegwerfmethode/Acceptance-Rejection-Method . . . 70
4.4 Numerische Integration . . . 72
4.4.1 Eindimensionale Integration . . . 72
4.4.2 Mehrdimensionale Integration . . . 74
4.5 Monte Carlo-Methoden . . . 75
4.5.1 Grundidee . . . 75
4.5.2 Konvergenzgeschwindigkeit . . . 77
4.5.3 Monte-Carlo-Integration . . . 78
4.6 Varianzreduktion . . . 80
4.6.1 Abtrennung des Hauptteils . . . 80
4.6.2 Antithetische Variablen . . . 81
4.6.3 Kontroll-Zufallsgrößen . . . 83
4.6.4 Selektives Simulieren (Importance sampling) . . . 84
4.6.5 Varianzreduktion durch Bedingung . . . 86
4.7 Hochdimensionale Integration . . . 88
4.7.1 Gitter-Verfahren/Produktformeln . . . 88
4.7.2 Quasi-Monte Carlo-Integration . . . 88
4.7.3 Dünne Gitter . . . 92
4.8 Anhang: Korreliert verteilte Zufallszahlen . . . 92
4.9 Anhang: Beweis des Dichtetransformationssatzes . . . 93
4.10 Bibliographische Anmerkungen . . . 94
4.11 Übungen . . . 95

5 Bewertung von Optionen mittels Monte Carlo Methoden **101**
5.1 Brownsche Bewegung und stochastische Differentialgleichungen . . . 101
5.1.1 Brownsche Bewegung . . . 101
5.1.2 Itô-Prozesse . . . 103
5.1.3 Das Itô-Lemma . . . 107

5.1.4 Geometrische Brownsche Bewegung und Finanzmathematik 109
5.2 Monte Carlo–Simulation 115
5.2.1 Problemstellung 115
5.2.2 Simulation einer Auszahlungsfunktion 115
5.3 Sensitivitäten 116
5.3.1 Parameterabhängiger Erwartungswert 116
5.3.2 Differenzen-Sensitivitäten 117
5.3.3 Likelihood-Quotienten-Methode 119
5.3.4 Pfadabhängige Überlegungen 119
5.4 Anhang: Der Zoo der Optionen 120
5.4.1 Optionen mit speziellen Ausübungsrechten 120
5.4.2 Kombinierte Handelsstrategien 120
5.4.3 Pfadabhängige Optionen 121
5.4.4 Noch exotischere Optionen 122
5.5 Bibliographische Anmerkungen 122
5.6 Übungen 123

6 Diskretisierung bei der Bewertung von Optionen 127
6.1 Differenzenverfahren: Black-Scholes-Gleichung 127
6.1.1 Explizites Differenzenverfahren/konstante Volatilität 127
6.1.2 Crank-Nicolson–Diskretisierung/Konstante Volatilität 129
6.1.3 Crank-Nicolson-Verfahren/lokaler Volatilität 130
6.2 Die Diskretisierung der Dupire-Gleichung 131
6.3 Diskretisierung einer stochastischen Differentialgleichung 132
6.3.1 Das Euler-Verfahren 132
6.3.2 Das Milstein-Verfahren 134
6.4 Fourier-Transformation und Bewertung von Optionen 135
6.4.1 Eine Preisformel und seine Fouriertransformierte 135
6.4.2 Eine gewichtete Preisformel 137
6.4.3 Die diskrete Fouriertransformation 138
6.5 Anhang: Schnelle Fouriertransformation 142
6.6 Bibliographische Anmerkungen 146
6.7 Übungen 147

II Teil 2: Weiterführende Themen 150

Vorwort

Finanzderivate sind Produkte, mit denen finanzielle Aktivitäten abgesichert werden können. Sie sind im Allgemeinen Produkte, die von einem Basisobjekt, überwiegend Aktien, abgeleitet sind. Die mathematische Theorie dieser Finanzinstrumente wurde zu einer eigenständigen mathematischen Disziplin parallel zur wachsenden Bedeutung des Optionshandels und der weltweiten Vernetzung der Finanzmärkte. Im Mittelpunkt steht die Frage, was ist der „faire“ Preis eines Finanzderivats.

Die geschichtliche Entwicklung rund um die Kapitalmärkte stellt sich so dar:[1]

- 1602: Gründung der ersten Aktiengesellschaft der Welt (Vereinigte Ostindische Kompanie/Niederlande)
- 17. Jahrhundert: Optionshandel für Tulpenzwiebel in den Niederlanden
- 1637: Zusammenbruch des Optionshandels mit Tulpenzwiebeln, da die Verkäufer der Optionen ihre eingegangenen Verpflichtungen nicht nachkommen konnten. Optionen geraten in Verruf
- 1756: Handel der ersten deutschen Aktie, der „Preußischen Kolonialgesellschaft“, in Berlin
- 1792: Gründung der New Yorker Börse als Vorläufers der „New York Stock Exchange“
- 1884: Der erste amerikanische Aktienindex (Railroad Average) wird notiert
- 1900: Bachelier's Dissertation zur „Theorie der Spekulation“ unter Zugrundelegung der Brownschen Bewegung als Marktmodell. Startpunkt für die Theorie der stochastischen Prozesse.
- 1929: Schwarzer Freitag am 25.10.1929
- 1930: Legalisierung des Optionshandels
- 1959: Ausgabe der ersten Volksaktie (PREUSSAG), 1961 folgte die zweite(VW)
- 1963: B. Mandelbrot beschäftigt sich mit der Verteilung der Aktienpreise
- 1973: Entwicklung einer partiellen Differentialgleichung zur Festsetzung von Optionspreisen auf der Basis der geometrischen Brownschen Bewegung durch F. Black, M. Scholes and R. Merton
- 1973: Gründung der Chicago Board Options Exchange (CBOE)

[1] Mehr Zur Geschichte findet man unter `http://www.afajof.org/association/historyfinance.asp`.

- 1981: Martingaltheorie[2] als Werkzeug der mathematischen Theorie des Preisentwicklung für Optionen.
- 1982: Beginn der Entwicklung der so genannten GARCH-Prozesse (R.F. Engle, T. Bollerslev).
- 1983: J.C. Cox und M. Rubinstein: Optionsmärkte
- 1990: Nobelpreis für Wirtschaftswissenschaften an Harry M. Markowitz, Merton H. Miller, William F. Sharpe für bahnbrechende Arbeiten in der Kapitalmarkttheorie
- 1990: Einführung von EUREX (Fortführung der Deutschen Terminbörse)
- Seit ca. 1990: Etablierung von MathFinance/Computational Finance als eigenständige Disziplin
- 1997: Nobelpreis für Wirtschaftswissenschaften an Myron Scholes und Robert Merton[3]
- 2003: Nobelpreis für Wirtschaftswissenschaften an F. Engle und C.W.J. Granger für ihre Forschungen zur Modellierung der Volatilität[4]
- 2008: Die Finanzkrise wird zum Teil der ungenügenden mathematischen Modellierung angelastet. Trotz allem: der durchschnittliche Umsatz pro Tag an der Terminbörse in Frankfurt beträgt ca. 3 Milliarden US$

Der Finanzmarkt teilt sich in den **Kassa-** und den **Terminmarkt** auf. **Derivate**[5] sind Verträge zu Finanzinstrumenten, die das Recht garantieren, ein Gut zu einem fixen Preis zu kaufen oder zu verkaufen. Das zugrundeliegende Gut wird als **Basiswert** (engl. underlying, dengl. Underlying) bezeichnet. Auf den Finanzmärkten werden folgende Klassen von Finanzderivaten gehandelt:

- Optionen (Bedingte Termingeschäfte)
- Forwards und Futures (Unbedingte Termingeschäfte)
- Swaps (Tausch von Zahlungsströmen)

Wir beschäftigen uns im wesentlichen mit der Fragestellung um die „Preisbildung“ bei Optionen herum. Es werden Optionen gehandelt etwa zu den Basiswerten „Aktien (stocks), Rohstoffe (commodities), Anleihen (bonds), Energie (energy assets), Wechselkurse (exchange rates),...“. Hauptargument für den Handel mit Finanzderivaten ist der Wunsch, sich gegen unerwünschte Risiken abzusichern (hedging), von Kursschwankungen eines Basiswertes zu profitieren (Spekulation) und von eventuellen Preisunterschieden auf verschiedenen (geographischen) Märkten zu profitieren (Arbitrage). Gehandelt werden Optionen in der Regel im Interbankenhandel, zum Teil auch an Terminbörsen.

[2] Das Wort stammt aus dem Provenzalischen. Es bezeichnet im Reitsport einen optionalen Teil der Pferdeausrüstung, der das Pferd daran hindern soll, den Kopf nach oben zu reißen und zu steigen. Der Name Martingal bezieht sich auf die französische Stadt Martigues im Departement Bouches du Rhone am Rande der Camargue, wo dieser Hilfszügel erfunden wurde. Seit dem 18. Jahrhundert steht Martingal auch für eine Strategie im Glücksspiel. Von J. Ville, P. Léy wurden die Martingale in der Wahrscheinlichkeitstheorie initiiert, von J.L. Doob wurden sie ab 1953 systematisch untersucht.

[3] Fischer Black, der auch große Verdienste um die Modellierung vorzuweisen hat, konnte in die Auszeichnung nicht mehr einbezogen werden, da der Nobelpreis statutengemäß nicht posthum verliehen werden kann.

[4] Siehe `http://www.nobelprize.org/nobel_prizes/economics/laureates/2003/engle-lecture.html`.

[5] lat. derivare = ableiten

Eine **Option** ist ein Vertrag, der seinen Besitzer (Inhaber der Option) das Recht einräumt, eine bestimmte Menge eines bestimmten Gutes **(Basisobjekt)** zu einem festgelegten Preis, dem **Ausübungspreis (strike)** zu kaufen **(Call, Kaufoption)** bzw. zu veräußern **(Put, Verkaufsoption)**.[6] Für dieses Recht zahlt der Käufer der Option dem Verkäufer eine Prämie, den **Optionspreis**. Ist die Ausübung der Option nur zu einem bestimmten Zeitpunkt möglich, so spricht man von einer **europäischen Option**, kann die Option jederzeit bis zu einem festgelegten Zeitpunkt, dem **Verfallsdatum**, ausgeübt werden, so handelt es sich um eine **amerikanische Option**[7]**.** Der Verkäufer der Option ist verpflichtet, während der festgelegten Frist (Laufzeit) auf Verlangen des Käufers den Basiswert zum vereinbarten Ausübungspreis zu liefern oder abzunehmen.

Call- bzw. Put-Optionen spiegeln die Erwartung wider, dass der Kurs des Basiswertes steigt bzw. fällt. Mit ihrer Hilfe kann sich der Optionshalter gegen Kursschwankungen am Markt absichern bzw. die Risiken am Markt deutlich begrenzen. Optionen haben also wenig mit Spekulation zu tun, sie sind ein Finanzinstrument mit Versicherungscharakter. Zu beachten ist, dass ein Anleger, der Optionen kaufen/verkaufen will, nicht selbständig an der Terminbörse agieren kann, er benötigt zur Durchführung einen entsprechenden Makler.

Um diese Risiken optimal einzugrenzen, gibt es Optionen mit sehr unterschiedlichen Eigenschaften; wir kommen darauf zurück. Die oben beschriebenen Optionstypen (europäisch oder amerikanisch), denen *ein* Basisobjekt zugrundeliegt, sind die Standardoptionen, auch **plain vanilla options** genannt.[8] Die **Bermuda-Option** liegt zwischen der europäischen und amerikanischen Option:[9] sie beinhaltet ein Kündigungsrecht zu fest vereinbarten Zeitpunkten. Aus den Standardoptionen können andere komplexere Optionsformen „geschneidert" werden, man nennt solche Optionstypen **exotische Optionen**. Dazu gehören u.a. *Basket-Optionen, rainbow options, asian options*. Letztendlich können Optionen noch als Wertpapier gestaltet werden (Optionsscheine).

Wir werden uns exemplarisch meist mit Aktienoptionen beschäftigen, also Optionen, denen Aktien als Basisobjekte zugrundeliegen. Bei einer Zinsoption stellt ein Zinssatz/eine Zinskurve das Basisobjekt dar.

FAZ, 25.10.2009

...Dieses Verhalten wird bisher nicht genügend berücksichtigt.
Wie kann man das ändern?
Dafür braucht es komplexere mathematische Verfahren, als wir Ökonomen sie momentan meist verwenden.
Wie bitte, Sie fordern mehr Mathematik? Meistens werden Ökonomen doch dafür kritisiert, dass sie zu mathematisch sind.
Und das ist falsch. In Wahrheit benutzen die meisten Ökonomen nur relativ simple Mathematik. Deshalb lassen ihre Modelle viel von dem außen vor, was im Leben passiert. Die Ökonomen haben nicht genügend Respekt vor den interessanten Ergebnissen, die kompliziertere Mathematik hervorbringen kann. ...

G. Akerlof, amer. Ökonom, Nobelpreisträger 2001

Ein Optionsgeschäft ist also eine asymmetrische Vereinbarung zwischen zwei Vertragspartnern: Der Käufer erwirbt ein Recht, das er nach eigener Entscheidung ausüben kann, der Verkäufer dagegen geht eine Verpflichtung ein. Die mathematische Durchdringung dieses zu umfassender Bedeutung gelangten Finanzinstruments ist wesent-

[6] Im täglichen Sprachgebrauch: „ich habe noch die Option, dieses Angebot anzunehmen".

[7] Die Adjektive "europäisch, amerikanisch" sagen nichts über die geographische Verteilung aus, sondern haben rein historische Gründe.

[8] Der Ursprung sind vermutlich die Südstaaten der USA. Dort scheint Eiscreme mit Vanillegeschmack die Standardsorte zu sein, und so ist „Plain Vanilla", zunächst als Bezeichnung für die Standardausführung von Eiscreme entstanden und dann auf die Standardausführung anderer Dinge übertragen worden.

[9] Bermuda liegt zwischen Europa und Amerika.

licher Bestandteil der **Finanzmathematik**, die in den beiden letzten Jahrzehnten eine rasante Entwicklung genommen hat.

Aufgaben der mathematischen Finanztheorie sind etwa:

1. Modellierung der Finanzmärkte
2. Optionspreisentwicklung
3. Absicherungsstrategien für Derivate
4. Risikomanagement von Portfolios
5. Portfoliooptimierung
6. Kalibrierung der Modelle
7. Statistik der Finanzmärkte

Für die Modellierung von Optionen sind Annahmen über die Eigenheiten des Marktes (Vollständigkeit, Arbitragemöglichkeiten,...) zu vereinbaren; wir sprechen von *qualitativen Annahmen*. Hinzukommen *quantitative Merkmale* wie z.B. Zinshöhen, Besteuerung, Dividenden, ... Es ist u.a. Aufgabe der numerischen Finanzmathematik, diese quantitativen Größen aus den Marktdaten zu errechnen.

Der typische Ansatz bei der Bewertung von Finanzderivaten startet mit einem Modell für die zukünftige Entwicklung der zugrundeliegenden Basiswerte. Hier werden stochastische Differentialgleichungen verwendet, um zufälligen Schwankungen der Kurse Rechnung tragen zu können. Daraus ergeben sich die fairen Optionspreise dann als Erwartungswerte einer Zufallsgröße. Für diese Erwartungswerte sind im Allgemeinen keine geschlossenen oder einfach berechenbare Formeln angebbar. In der Behandlung der mathematischen Probleme im Zusammenhang damit kommen zum Einsatz:

- Numerische Integration und Differentiation
- Modellierung des Zufalls und Monte Carlo–Methoden
- Stochastische Prozesse
- Stochastische Differentialgleichungen und ihre Numerik
- Partielle Differentialgleichungen und ihre Numerik
- Optimierungstheorie in der Kalibrierung

In dieser Vorlesung werden wir Themen besprechen, die in der Finanzmathematik im Zusammenhhang mit Berechnungsverfahren aufkommen. Der Schwerpunkt liegt dabei auf Methoden zur Bewertung von Optionen, für die wir auch die nötigen theoretischen Grundlagen streifen werden. Speziell werden wir die folgenden Verfahren betrachten:

- Das Newtonverfahren als Hilfe bei der Bestimmung der impliziten Volatilität.
- Die Binomialmethode
- Monte-Carlo Simulationen
- Numerische Lösungverfahren für stochastische gewöhnliche Differentialgleichungen
- Finite Differenzen und Finite Elemente Methoden zur Lösung der Black-Scholes-Gleichung

Die Reihenfolge der Themen ist so gewählt, dass die Aufgaben, die modellierungsmäßig/mathematisch weniger anspruchsvoll sind, zuerst behandelt werden. Dabei lernen wir dann parallel und sukzessive auch wichtige Begriffsbildungen der Finanzmathematik kennen. Die Auswahl der Themen ist nicht ausgeprägt auf Numerik ausgerichtet, sondern entspricht eher dem Anspruch „Computational Finance".

Zur Literatur zunächst eine Aufstellung, die allgemeine Hinweise enthält. In den einzelnen Kapiteln kommen wir auch auf spezielle Referenzen zurück, die zwar thematisch nicht vollständig sind, aber die wichtigsten Arbeiten benennen.

Zentral ist Modellierung der Optionspreisberechnung, wie sie von Black, Scoles and Merton begonnen wurde; siehe [4, 28]. Zur Geschichte siehe [17, 15]. Bücher zur Finanzmathematik, die etwas allgemeineren Anspruch haben, sind [1, 5, 7, 13, 20, 21, 22, 26, 25, 30, 34, 35, 41, 42]. Aspekte allgemeiner Modellbildung finden sich in [38, 40]. Als Monographie zum Thema „Computational Finance" kann [37] empfohlen werden. Wir setzen das Basiswissen mathematischer Grundvorlesungen und Kenntnisse der elementaren Numerik und Stochastik voraus. Als Hintergrundliteratur für die Numerische Mathematik führen wir [2, 9, 10, 11, 19, 32, 36] an. Als allgemeine Literatur zur Stochastik/Statistik kann [12, 18, 31, 39] herangezogen werden. In [6, 8, 14, 29, 38] wird die Umsetzung numerischer Verfahren unter Nutzung der Programmpakete MAPLE, MATLAB und MATHEMATICA besprochen. In der Praxis der Finanzproduktanbieter wird fast ausschließlich mit der Programmiersprache C++ gearbeitet; siehe [23, 27] und auch [1]. Wegen der Tatsache, dass viele Aufgaben der Finanzmathematik mit Aspekten inverser Fragestellungen zu tun haben, führen wir dazu [3, 16, 24, 33] an.

Interessante web-sites zur Thematik sind

`http://www.mathematik.de/ger/presse/ausdenmitteilungen/artikel/mdmv-17-3-148.pdf`

`http://www.wiwi.uni-frankfurt.de/en/research/house-of-finance/`

`http://www.mathfinance.de/`

`http://www.math.nyu.edu/financial_mathematics/`

`http://math.gmu.edu/~tsauer/pre/sde.pdf`

`http://www.math.ethz.ch/Finance`

`http://www.finmath.rutgers.edu/`

Der erste Teil dieses Skriptums enthält die elementaren Themen, die in einer Vorlesung im Wintersemester 2009/10 im Wesentlichen behandelt wurden, der zweite Teil wird weiterführende Themen (Lokale Volatilität, Kalibrierung von Modellen, Stochastische Volatilität, Amerikanische Optionen, Finite Elemente-Verfahren, Maliavin-Kalkül, Lévy-Prozesse) enthalten.

Mein Dank gilt Priv.-Doz. Dr. N. Hofmann und Prof. Dr. T. Gerstner, in deren Skripten ich Einsicht nehmen konnte, und Priv.-Doz. Dr. T. Lorenz, Dr. S. Mathew und Dr. C. Sanz, die das Skriptum in Teilen kritisch gelesen haben.

Frankfurt, im März 2010 JOHANN BAUMEISTER

Literaturverzeichnis

[1] Y. Achdou and O. Pironneau. *Computational methods for option pricing.* SIAM, Philadelphia, 2005.

[2] K. Atkinson and W. Han. *Theoretical Numerical Analysis. A Functional Analysis Framework.* Texts in applied mathematics, vol. 39. Springer, New York, 2001.

[3] J. Baumeister. *Stable Solution of Inverse Problems.* Vieweg, Braunschweig, 1987.

[4] F. Black and M. Scholes. The pricing of options and corporate liabilities. *J. of political economy*, 81:637–659, 1973.

[5] M. Bloss, D. Ernst und J. Häcker. *Derivatives - An authoritative guide to derivatives for financial intermediaries and investors.* Oldenbourg, München, 2008.

[6] P. Brandimarte. *Numerical methods in Finance and economics. A MATLAB-based Introduction.* Wiley & Sons, Hoboken, 1995.

[7] R. Cont and P. Tankov. *Financial modelling with jump processes.* Financial Mathematics Series. Chapman & Hall/CRC, 2003.

[8] S. Cyganovski, P. Kloeden and J. Ombach. *From Elementary Probability to Stochastic Differential Equations with MAPLE.* Springer, Berlin, 2001.

[9] P. Deuflhard and F. Bornemann. *Numerical Methods II.* de Gruyter, Berlin, 1997.

[10] P. Deuflhard and A. Hohmann. *Numerical Methods I.* de Gruyter, Berlin, 1995.

[11] B. Flanery, W. Press, S. Teukolsky and W. Vetterling. *Numerical Recipes in C.* Cambridge University Press, Cambridge, 1992.

[12] J. Franke, W. Härdle and C. Hafner. *Einführung in die Statistik der Finanzmärkte.* Springer, Berlin, 2001.

[13] P. Glasserman. *Monte Carlo Methods in Financial Engineering.* Springer, Baltimore, 2003.

[14] M. Günther and A. Jüngel. *Finanzderivate mit MATLAB.* Vieweg, 2003.

[15] W. Hafner (Editor) and H. Zimmermann (Editor). *Vinzenz Bronzin's Option Pricing Models: Exposition and Appraisal.* Springer, New York, 2009.

[16] H.W. Engl, M. Hanke and A. Neubauer. *Regularization of Inverse Problems.* Kluwer, Dordrecht, 1996.

[17] E.G. Haug. *The Complete Guide to Option Pricing Formulas.* McGraw-Hill, 2007.

[18] N. Henze. *Stochastik für Einsteiger.* Vieweg, Braunschweig, 1997.

[19] G. Hämmerlin and K.-H. Hoffmann. *Numerische Mathematik.* Springer, 1989.

[20] J.C. Hull. *Options, Futures, and other Derivatives.* Prentice Hall, 2003.

[21] A. Irle and C. Prelle. *Übungsbuch Finanzmathematik.* Teubner, Wiebaden, 2007.

[22] M. Jeanblanc, M. Yor and M. Chesnay. *Mathematical Methods for Financial Markets.* Springer, New York, 2009.

[23] M. Joshi. *C++ Design Patterns and Derivative Pricing.* Cambridge University Press, Cambridge, 2008.

[24] A. Kirsch. *An Introduction to the Mathematical Theory of Inverse Problems.* Springer, New York, 1996.

[25] P. Kloeden and E. Platen. *Numerical Solution of SDE through Computer Experiments.* Springer, Berlin, 1994.

[26] R. Korn und E. Korn. *Optionsbewertung und Portfolio–Optimierung.* Vieweg, Braunschweig, 1999.

[27] G. Levy. *Computationals Finance using C and C++.* Academic Press, London, 2008.

[28] R.C. Merton. Theory of rational option pricing. *Bell Journal of Economics and managemen Science*, 4:141–183, 1973.

[29] T. Mikosch. *Elementary stochastic calculus with finance in view.* World Scientific, 1998.

[30] T. Mikosch. *Non-life insurance mathematics. An Introduction with stochastic processes.* Springer, Berlin, 2004.

[31] P.E. Protter. *Stochastic integration and differential equations.* Springer, New York, 2004.

[32] A. Quateroni, R. Sacco and F. Saleri. *Numerische Mathematik 1.* Springer, New York, 2001.

[33] A. Rieder. *Keine Probleme mit Inversen Problemen.* Vieweg, Braunschweig, 2003.

[34] B. Rudolph and K. Schäfer. *Derivative Finanzmarktinstrumente: Eine anwendungsbezogene Einführung.* Springer, Berlin, 2005.

[35] K. Sandmann. *Einführung in die Stochastik der Finanzmärkte.* Springer, Berlin, 2010.

[36] H.R. Schwarz. *Numerische Mathematik.* Teubner, Stuttgart, 1996.

[37] R. Seydel. *Tools for Computational Finance.* Springer, Berlin, 2002.

[38] W. Shaw. *Modelling Financial derivatives with Mathematica.* Cambridge University Press, Cambridge, 1998.

[39] S.E. Shreve. *Stochastic Calculus and Finance.* Carnegie Mellon University, 1996.

[40] T. Sonar. *Angewandte Mathematik, Modellbildung und Informatik.* Vieweg, Braunschweig, 2001.

[41] K. Sydstaetter and P. Hammond. *Essential mathematics for economic analysis.* Englewood Cliffs, Prentice Hall, 2002.

[42] P. Wilmott, S. Howison and J. Dewynne. *The Mathematics of Financial Derivatives.* Cambridge University Press, Cambridge, 1996.

Teil I

Teil 1: Elementare Themen und Verfahrensweisen

Kapitel 1

Die Black-Scholes-Preisformel und Berechnung der impliziten Volatilität

Wir stellen die Black-Scholes-Preisformel vor, die zwar vielfältig verändert und verallgemeinert wurde, sie stellt jedoch bis heute insbesondere in der Bewertung von Aktienoptionen den Standard dar. Die Kennzahl „Volatilität“, die diese Preisformel „dominiert“, bezeichnet das Ausmaß der Schwankungen von Kursen an Finanzmärkten. Zur Bestimmung der Volatilität aus Marktdaten benötigt man ein Verfahren zur Lösung einer nichtlinearen Gleichung. Wir beginnen mit der Behandlung dieser Aufgabestellung, da sie – ohne genauer auf die Modelle einzugehen – schon einige Begriffe und Definitionen der mathematischen Finanztheorie erläutert.

1.1 Aktien, Werpapiere und risikofreie Anlagen

Bevor wir über Optionen oder allgemein über Finanzderivate reden können, sollten wir über die ‚Zutaten“ reden, die hauptsächlich benötigt werden: Aktien, Wertpapiere und Geldanlagen.

1.1.1 Aktien

Eine **Aktie** ist ein Anteil am Grundkapital einer Aktiengesellschaft und eine Urkunde, die dem Inhaber seine Rechte verbrieft.[1] Eine Aktiengesellschaft erhält durch den Verkauf von Anteilen des Unternehmens Kapital in Höhe des Aktienpreises, das im Gegensatz zum Kredit nicht zurückgezahlt werden muss. Als Kompensation erhält der Aktionär/Aktienkäufer Anspruch auf Einbeziehung bei Dividendenzahlungen, spezielle Bezugsrechte und Mitbestimmungsrechte. Aktienausgaben stellen für Unternehmen alternative Quellen zur Fremdfinanzierung am Finanzmarkt dar.

Die Erträge von Aktien sind in der Regel höher als risikolose Geldanlagen, sie sind aber auch mit höheren Risko behaftet. Das Risiko besteht darin, dass Aktien im Preis schwanken, an Wert verlieren können, ja sogar wertlos werden können, wenn die Aktiengesellschaft ihre „Existenz“ verliert. Die Risikoeinschätzung geht ein in die Einteilung nach ihrer Qualität:

Blue Chips Aktien erstklassiger Unternehmen mit sehr guter Marktposition und Erfolg.

Zyklische Aktien Aktien, die sehr stark konjunkturabhängig sind (Autowerte, Bauunternehmen,...).

Nebenwerte Aktien kleinerer und mittlerer Unternehmen mit heftigeren Kursbewegungen.

Penny Stocks Extrem risikoreiche Aktien, deren Wert im „Penny“-Bereich liegen.

[1] actio (lat.)= Handlung, übertragen „einklagbarer Anspruch.

Aktien werden hauptsächlich an den Börsen gehandelt. Ihre Kurswerte sind jedem „Kapitalinvestor“, insbesondere in Internetzeiten, zugänglich. Die Gründe für die Veränderung der Aktienkurse sind zum Teil objektiver Natur, zum Teil aber sehr schwer zu durchschauen. Beobachtet man die Kursentwicklung einer Aktie, so stellt man fest, dass sich meist zwei Effekte überlagern: **langfristig bestimmender Trend** und **kurzfristige Einflüsse**. Diese Erkenntnis wird im Allgemeinen durch jeden Kursverlauf zumindest über bestimmte Zeiträume bestätigt. Es läge nun nahe, nach zufälligen Funktionen (Realisierungen von stochastischen Prozessen) zu suchen, die diese Eigenschaften wiedergeben können. Ein Vergleich mit Realisierungen des Wiener Prozesses/der Brownschen Bewegung zeigt, dass dies möglich ist. Daher spielt dieser Prozess eine überragende Rolle bei der Modellierung von Finanzmärkten.

Bemerkung 1.1 *Investitionsprojekte sind mit Risiken verbunden, wie beispielsweise*

- *Absatzmengen- und Absatzpreisrisiken*
- *Zinsänderungsrisiken*
- *Währungsänderungsrisiken*

Unter **hedging** *versteht man die Eliminierung solcher Risiken durch den Abschluss von Termingeschäften wie Optionen, Forwards, Futures, Swaps.* □

1.1.2 Festverzinsliche Wertpapiere

Ein **Wertpapier** ist eine Urkunde, die ein Vermögensrecht verbrieft, etwa die Miteigentümerschaft an einem Unternehmen. Wertpapiere sind zum Beispiel Aktien, Anleihen, Schecks und Wechsel. Börsenfähige Wertpapiere werden als Effekten bezeichnet.

Unter **festverzinslichen Wertpapieren (Kupon-Anleihen/Bonds)** versteht man Kapitalanlagen, bei denen der Zinssatz über die gesamte Laufzeit völlig gleich bleibt und schon im Vorhinein festgelegt wird, so dass der Anleger hiermit eine völlig risikofreie Geldanlage eingeht. Die Rendite, also das Verhältnis der Auszahlungen zu den Einzahlungen der Anlage, steht dabei also schon vor Ablauf der Vertragszeit genau fest. Der große Vorteil bei dieser Anlageform ist neben der Verlustsicherheit auch die Sicherheit bei der Planung. So weiß der Anleger stets ganz genau, welche Rendite er in welcher Zeit zu erwarten hat.

Als festverzinsliche Wertpapiere werden meistens Anleihen wie zum Beispiel Schuldverschreibungen ausgegeben. Festverzinsliche Wertpapiere dienen Staaten, Institutionen und Unternehmen zur Fremdfinanzierung. Im Gegensatz zu Aktien, durch die der Investor Eigentümer wird, sind die Käufer festverzinslicher Wertpapiere Gläubiger und haben als solche nur Anspruch auf Zinsen und Tilgung des Anleihekapitals. Wir verwenden „festverzinsliche Anleihe“ oder „festverzinsliche Geldanlage“ synonym für festverzinsliches Wertpapier.

Regel 1.2 (Verzinsung) *Der Wert $B(t)$ eines festverzinslichen Wertpapiers vom Betrage $B(0)$ mit einem jährlichen Zinssatz r beträgt nach t Jahren*

- *bei einmaliger Verzinsung pro Jahr:* $B_1(t) = B(0)(1+r)^t$
- *bei m-maliger Verzinsung pro Jahr:* $B_m(t) = B(0)(1+\frac{r}{m})^{tm}$
- *bei kontinuierlicher Verzinsung:* $B_\infty(t) = B(0)e^{rt}$

Die Formel für B_∞ folgt so:

$$B_\infty = \lim_{m\to\infty} B(0)(1+\frac{r}{m})^{tm} = \lim_{ar\to\infty} B(0)(1+\frac{1}{a})^{art} = B(0)(\lim_{a\to\infty}(1+\frac{1}{a})^a)^{rt} = B(0)e^{rt}\,.$$

In der Praxis hängt der (jährliche) Zinssatz von der Laufzeit t ab, denn für feste zeitliche Bindungen gelten in der Regel andere Zinssätze als für täglich fällige Gelder. Der Zinssatz ist also eine Funktion $r(t)$ der Laufzeit t. Diese Funktion bezeichnet man als **Zinskurve (yield curve, term structure of interest rates)**. Diese Zinskurve wird im Wesentlichen durch Angebot und Nachfrage festgelegt. Dabei spielen die Europäische Zentralbank und die Nationalbanken eine Rolle. Weil die Zinsen entlang der Zinskurve im Allgemeinen kein Ausfallrisiko haben, nennt man diese auch die **risikolose Zinskurve**.

Ein **Zinsderivat (Interest Rate Derivative)** ist ein Derivat, dessen Basisobjekt ein Zins oder eine zinsbezogene Größe ist. Sie sind Teil eines aktiven Zinsmanagement der Banken. Solche Derivate haben zu den Optionen, wie wir sie besprechen werden, den Unterschied, dass der Wert des Basisobjekts keine Zahl sondern eine Kurve ist.

Unter **Diskontierung** (Abzinsung) versteht man den zur Verzinsung umgekehrten Vorgang.

Regel 1.3 (Diskontierung) *Der Wert $B(0)$ eines festverzinslichen Wertpapiers vom Betrage $B(t)$ zur Zeit t mit einem jährlichen Zinssatz r beträgt*

- *bei m-maliger Verzinsung pro Jahr:* $B(0) = B(t)(1 + \frac{r}{m})^{-tm}$
- *bei kontinuierlicher Verzinsung:* $B(0) = B(t)e^{-rt}$

1.2 Eine Preisformel

Wir haben im Vorwort die Definition einer Option kennengelernt:

> Eine **Option** ist ein Vertrag, der seinen Besitzer (Inhaber der Option) das Recht einräumt, eine bestimmte Menge eines bestimmten Gutes **(Basisobjekt)** zu einem festgelegten Preis, dem **Ausübungspreis (strike)** zu kaufen **(Call, Kaufoption)** bzw. zu veräußern **(Put, Verkaufsoption)**. Für dieses Recht zahlt der Käufer der Option dem Verkäufer eine Prämie, den **Optionspreis**.

Bemerkung 1.4 *Wir wollen in diesen Aufzeichnungen ausschließlich Optionen (bedingte Termingeschäfte) betrachten. Daher kurz ein Beispiel für* **Forwards**, *also ein unbedingtes Termingeschäft:*

> *Eine Person A vereinbart mit einer Person B, in genau einem Jahr insgesamt X kg Kaffee der Qualität Q zum Preis von W € zu liefern.*

Die Besonderheiten der Forwards sind, dass heute der zukünftige Preis vereinbart wird und Liefermenge, Qualität, Lieferzeit beliebig festgelgt werden können. Diese Vertragsstruktur hat Vorteile (passgenaue Absicherung), aber auch Nachteile (etwa: sehr schwer Vertragspartner zu finden). Der **Future** *will die Nachteile des Forward überwinden:*

- *Liefermenge, Qualität, Lieferzeit werden standardisiert*
- *Handel läuft über ein Clearinghaus, das Sicherheiten verlangt.*

Futures weisen diverse Optionen auf Abweichungen auf (quantity, quality, location, wildcard option). □

Wir betrachten ein Optionsgeschäft für Aktien. Es werde mit V der **Optionspreis**, mit S der Kurs des **Basisobjekts**, mit T die **Laufzeit**, mit K der Ausübungspreis und mit S_T der Kurs der Aktie **(Basiswert)** am **Fälligkeitstag** bezeichnet. Ist $S_T > K$ (die Option ist „in

the money“), so kann der Besitzer der Option die Aktie zum Preis K erwerben und sofort zum höheren Preis S_T am Markt verkaufen. Er erzielt dann eine **Auszahlung (payoff)** in Höhe von $S_T - K$ (unter Vernachlässigung von Transaktionskosten). Ist $S_T < K$ (die Option ist „out of the money“), so lässt der Besitzer der Option sein Recht verfallen, selbst wenn er Interesse am Kauf dieser Aktie hätte. Es ist nämlich dann günstiger, die Aktie am Markt zum Preis S_T zu erwerben. In diesem Fall ist die Auszahlung für die Option gleich Null. Der Fall $S_T = K$ (die Option ist „at the money“), ist eine Situation, die wie der Fall $S_T < K$ zu behandeln ist.
Zusammengefasst ergibt sich für den Besitzer der Option eine „Auszahlung“ zum Zeitpunkt T in Höhe von

$$(S_T - K)^+$$

wobei $h^+ := h$, falls $h \geq 0$, $h^+ := 0$, falls $h < 0$ ist.

Hier haben wir ein Optionsgeschäft beschrieben, das man **europäisch** nennt. Bei einem **amerikanischen Optionsgeschäft** kann man zu jedem Zeitpunkt in $[0, T]$ entscheiden, ob man das Recht ausüben will.

Aus den obigen Ausführungen können wir schließen, dass eine Option ihrem Besitzer eine nichtnegative Auszahlung zusichert, die in ihrer Höhe allerdings unsicher ist. Daher ist es verständlich, dass man für den Erwerb einer Option eine Zahlung, die **Optionsprämie**, leisten muss. Die Auszahlung ist also um den Wert der Optionsprämie zu mindern, genauer um den verzinsten Wert der Optionsprämie, um den Gewinn/Verlust zu ermitteln. Es ist offensichtlich, dass für eine amerikanische Option eine höhere Optionsprämie zu entrichten sein sollte, bietet sie doch mehr Rechte.

Das Problem im (seriösen) Optionshandel ist, die Optionsprämie zu berechnen, d.h. den Preis der Option zum Zeitpunkt $t = 0$ festzusetzen, und, um den Handel mit der Option, solange sie noch nicht ausgeübt ist, zu ermöglichen, zu jedem Zeitpunkt t den Wert der Option zu bestimmen. Die Schwierigkeit besteht darin, dass man den Verlauf des Aktienkurses über den Laufzeitraum nicht kennt.

Wir machen uns die Problematik zunächst an einem einfachen Modell klar, dem sogenannten Ein-Perioden -Binomialmodell; wir kommen darauf in Kapitel 2 zurück. Zur Frage der Festsetzung des Optionspreises wird ein Wertpapierdepot, auch **Portfolio** genannt, gebildet, das folgendermaßen zusammenzusetzen ist:

Aktiendepot der betreffenden Aktie, festverzinsliche Anleihe.

Es ist nicht überraschend, dass nun Anleihen ins Spiel kommen, müssen doch die Aktien bzw. die Optionsprämie finanziert werden.

Wir kaufen also einen Bruchteil[2] Δ der Aktie auf, und finanzieren die Geschäfte durch die Aufnahme eines Kredits B. Zum Zeitpunkt $t = 1$ verfalle die Option, deren Preis wir ermitteln wollen. Diesen Preis setzen wir dann als Wert des Depots zum Zeitpunkt $t = 0$ fest, dessen quantitative Zusammensetzung wir noch nicht kennen, da Δ und B noch unbekannt sind. Man spricht bei diesem Vorgehen von einer **Duplikationsstrategie**. Dabei ist es notwendig, neben den angegebenen Daten die Verzinsung für risikolose Geldaufnahmen und Geldanlagen zu kennen.

Im weiteren wird angenommen, dass der konstante Zinssatz für risikofreie Anlagen für eine Periode am Markt r ist, dass der Aufzinsungsfaktor bei einmaliger Verzinsung also gerade $z := 1 + r$ ist. Offen ist die Kursentwicklung der Aktie. Das einstufige Binomialmodell besteht nun darin, anzunehmen, dass der Kurs der Aktie mit Wahrscheinlichkeit q auf den Wert uS_0 steigt

[2] In der Wirklichkeit erwirbt man ein Paket von Optionen, die Anzahl der aufzukaufenenden Aktien wird dann auch eine ganze Zahl.

Portfoliobewegung	**Wert des Portfolios** $t = 0$	**Wert des Portfolios** $t = 1$	
Aktie kaufen, $t = 0$	ΔS_0	$l\Delta S_0$	$u\Delta S_0$
Anleihe aufnehmen, $t = 0$	$-B$	$-zB$	$-zB$
Summe	$\Delta S_0 - B$	$l\Delta S_0 - zB$	$u\Delta S_0 - zB$
		Auszahlung der Option $t = 1$	
		0	$uS_0 - K$

Abbildung 1.1: Duplikationsstrategie

und mit Wahrscheinlichkeit $1 - q$ auf den Wert lS_0 fällt; also $u > 1, 0 < l \leq 1$. Das Diagramm 1.1 gibt die Entwicklung des Portfolios wieder. Dabei gehen wir davon aus, dass $lS_0 \leq K \leq uS_0$ gilt (um hier anderen Annahmen über den Markt aus dem Wege zu gehen). Die Optionsprämie wird nun so festgesetzt, dass

Endwert des Duplikationsdepots = Auszahlungswert der Option

erfüllt ist. Dies führt auf zwei Gleichungen für die Unbekannten Δ und B :

$$u\Delta S_0 - zB = uS_0 - K \, , \, l\Delta S_0 - zB = 0 \, .$$

Hieraus folgt:

$$\Delta = \frac{uS_0 - K}{(u - l)S_0} \, , \, B = \frac{l(uS_0 - K)}{(u - l)z} \, .$$

Nun ist die Zusammensetzung des äquivalenten Portfolios bekannt und die Optionsprämie C_0 berechenbar:

$$C_0 = \Delta S_0 - B \, .$$

Beachte, dass die Wahrscheinlichkeit q gar nicht eingeht.

Das obige einstufige Modell ist nur von theoretischem Wert. Ersetzt man nun die einmalige Preisänderung der Aktien durch eine endliche Anzahl n von Änderungen im Zeitraum $[0, T]$ kommt man einer kontinuierlicher Preisänderung schon nahe; die Analyse des Modells birgt keine neuen Schwierigkeiten, nur der Aufwand wird größer. Der Übergang vom diskreten Modell zu einem kontinuierlichen Modell gelingt durch die Einbeziehung der **geometrischen Brownschen Bewegung**, einem mathematisch anspruchsvollen Objekt aus dem Bereich der stochastischen Differentialgleichungen. Wir beschreiben diese Zusammenhänge in den folgenden Kapiteln genauer.

Betrachte eine Call-Option, deren Wert über den Laufzeitraum $[0, T]$ zu bestimmen sei. Die Auszahlung zum Zeitpunkt $t = T$ ist

$$(S_T - K)^+ := \begin{cases} S_T - K & \text{falls } S_T \geq K \\ 0 & \text{sonst} \end{cases} \tag{1.1}$$

Hierbei ist S_T der Kurs des Basisobjekts zur Zeit T, K der Ausübungspreis. Der Zinssatz für eine risikolose Anlage über den Zeitraum $[0, T]$ auf dem Finanzmarkt sei r .

Von Black, Scholes und Merton wurde für den kontinuierlichen Fall folgende Formel für den Optionspreis $C(S,t)$ in Abhängigkeit vom Zeitpunkt $t \in [0,T]$ und dem aktuellen Kurs des Basisobjekts S angegeben:

$$C(S,t) = S\mathcal{N}(d_+(\sigma)) - Ke^{-r(T-t)}\mathcal{N}(d_-(\sigma))\,, \tag{1.2}$$

Hierbei ist

$$d_\pm(\sigma) = \frac{\ln(\frac{S}{K}) + (r \pm \frac{\sigma^2}{2})(T-t)}{\sigma\sqrt{T-t}}\,,\; \sigma \geq 0\,,$$

und $\mathcal{N}$ die Verteilungsfunktion der Standardnormalverteilung, also

$$\mathcal{N}(a) := \int\limits_{-\infty}^{a} \frac{1}{\sqrt{2\pi}} \exp(-\frac{s^2}{2})ds\,,\; a \in \mathbb{R}\,.$$

Insbesondere stellt $C_0 := C(S_0, 0)$ den Preis dar, den der Emittent der Option verlangen sollte, wenn S_0 der Basiswert zur Zeit $t = 0$ ist. Auch in diesem kontinuierlichen Fall kann wieder ein äquivalentes Portfolio angegeben werden.

Was sagt uns die Formel? U.a.:

- Der Wert einer Call-Option steigt mit steigenden Kursen des Basisobjekts; darauf kommen wir zurück.
- Ist die Option zur Zeit t tief im Geld, d.h. ist der Kurs des Basisobjekts deutlich höher als der Ausübungspreis, dann ergibt sich nach der Preisformel ein tendenziell sehr großes, positives d_+. Damit gelten approximativ Verteilungswerte nahe bei eins:

  $$\mathcal{N}(d_+) \approx \mathcal{N}(d_-) \approx 1$$

  Der Wert der Option verhält sich also in diesem Fall etwa wie der aktuelle Kurs des Basisobjekts abzüglich des diskontierten Ausübungspreises:

  $$C(S,t) \approx S - Ke^{-r(T-t)}$$

Somit bestimmen fünf Parameter den Optionspreis: der Ausübungspreis K, der Zinssatz r bezogen auf den Zeitraum $[0,T]$, die Laufzeit T, der aktuelle Kurs S, und die Volatilität σ. Der kritische Parameter ist die so genannte **Volatilität** σ. Sie misst die Schwankungsbreite des Kurses des Basiswertes für Kursbewegungen innerhalb eines bestimmten Zeitrahmens; siehe Abschnitt 1.5.2. Sie muss (statistisch) aus Marktdaten geschätzt werden.

Die Parameter S, K, r, und die Restlaufzeit $\tau := T - t$ unterdrücken wir meist. Sie sind bekannte Marktdaten. Daher werden auf den Finanzmärkten meist auch nicht die Preise angegeben, sondern die Volatilität. Damit kann man die Preise dann berechnen.

Die Formel für den Wert V der Option besteht aus zwei Termen. Der erste Term $S\mathcal{N}(d_+(\sigma))$ beschreibt den Wert des zugrundegelegten Basiswertes, den der Besitzer des Call im Falle einer Ausübung seines Kaufrechtes beziehen kann. Der zweite Term $Ke^{-r(T-t)}\mathcal{N}(d_-(\sigma))$ mindert den ersten Term und entspricht dem Wert des Ausübungspreises, den der Besitzer der Option bezahlen muss, wenn er die Option ausübt. Das Verhältnis des Basiswertkurses zum Ausgabekurs spiegelt sich in den beiden Termen durch die Variablen $d_\pm$ wider.

Der Emittent (auch Stillhalter/Optionsverkäufer genannt) – man sagt, er sei in einer „short position“ – einer Call-Option kann seiner etwaigen Lieferverpflichtung beispielsweise dadurch nachkommen, dass er bereits im Zeitpunkt des Optionsverkaufs das Basisobjekt in sein Portfolio aufnimmt. Er bindet damit Kapital und verzichtet auf mögliche Zinserträge. Mit höherem Zinsniveau wird er daher eine höhere Optionsprämie verlangen.

Der Käufer einer Option – man sagt, er sei in einer „long position“ – ist während der Optionslaufzeit nicht zu einer vergleichbar hohen Kapitalbindung gezwungen und kann bis zur Ausübungszeit entsprechende Mittel auf dem Markt anlegen. Je höher der Zinssatz ist, desto größer wird tendenziell seine Bereitschaft sein, eine höhere Optionsprämie zu akzeptieren.

Bemerkung 1.5 *Im Unterschied zu Optionen beinhalten Forwards und Futures die Pflicht zur Ausübung.* □

Der Preis C einer europäischen Call-Option ist Lösung einer partiellen Differentialgleichung – auf Annahmen, die dies rechtfertigen, wollen wir hier noch nicht eingehen – nämlich der sogenannten **Black–Scholes–Gleichung**:

$$\frac{\partial C}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 C}{\partial S^2} + rS\frac{\partial C}{\partial S} - rC = 0 \;\; (S \in (0,\infty), t \in [0,T)) \tag{1.3}$$

Ferner gelten Randbedingungen

$$C(0,t) = 0\,,\; \lim_{S\to\infty}(C(S,t) - S) = 0\,,\; t \in (0,T)\,. \tag{1.4}$$

und natürlich die Endbedingung

$$C(S,T) = (S-K)^+\,,\; S > 0\,. \tag{1.5}$$

Dass die Lösungsformel (1.2) eine Lösung dieser Anfangs–Randwertaufgabe („Anfang“ wird sich gleich aufklären) darstellt, kann man direkt verifizieren. Den Weg umgekehrt, nämlich die Funktion in (1.2) als Lösung von (1.3), (1.4), (1.6) zu erhalten, wollen wir nun skizzieren.

Transformiert man die Konstanten und Variablen gemäß

$$x = \ln(\frac{S}{K}), \tau = \frac{1}{2}\sigma^2(T-t), Kv(x,\tau) = C(S,t), \rho = \frac{2r}{\sigma^2}\,,$$

– beachte die Zeitumkehr – erhalten wir aus (1.3) die Aufgabe

$$v_\tau - v_{xx} + (1-\rho)v_x + \rho v = 0\,,\; x \in \mathbb{R}, \tau \in (0, T_0 := \sigma^2 T/2]\,.$$

Wegen $(S-K)^+ = K(e^x - 1)^+$ wird die Endbedingung nun zur Anfangsbedingung

$$v(x,0) = (e^x - 1)^+\,.$$

Der Ansatz

$$v(x,\tau) := e^{\alpha x + \beta\tau}u(x,\tau)$$

mit

$$\alpha = -\frac{1}{2}(\rho - 1)\,,\; \beta = -\frac{1}{4}(\rho+1)^2\,,$$

bringt dann die Aufgabe

$$u_\tau - u_{xx} = 0\,,\; x \in \mathbb{R}, \tau \in (0,T_0)\,,$$

für u, also eine einfache Wärmeleitungsgleichung. Die Anfangsbedingung wird zu

$$u(x,0) = (e^{(\rho+1)x/2} - e^{(\rho-1)x/2})^+\,.$$

Die Wärmeleitungsgleichung wird gelöst durch die Schar

$$u(\tau,x) := \frac{1}{2\pi}\int_{-\infty}^{\infty} g(\xi)\exp(-\frac{(x-\xi)^2}{4\tau})d\xi\,, \tag{1.6}$$

wobei $g : \mathbb{R} \longrightarrow \mathbb{R}$ geeignet zu wählen ist. Wie man nun die geeignete Funktion g findet, so dass u auch die Anfangsbedingung erfüllt, dazu verweisen wir auf die Literatur (siehe [12]). Rückwärtssubstitutionen liefern V.

Nun bleibt die Frage im Raume, woher kommt die Anfangs–Randwertaufgabe (1.3), (1.4), (1.6)? Sie wird aus grundsätzlichen Annahmen über den Finanzmarkt, in dem die Option „lebt" abgeleitet, und zwar aus einer stochastischen Differentialgleichung, die wir ohne irgendeine Rechtfertigung hier aufschreiben:[3]

$$dS = \mu S\,dt + \sigma S\,dW_t \tag{1.7}$$

Sie wird ergänzt durch die Annahme über die Geldanlage:

$$dB = rBdt \tag{1.8}$$

Hierbei ist μ ein so genannter **Driftterm**, σ wieder die Volatilität und r der Zinssatz. Die zufällige Entwicklung des Basiswertes wird durch den Wiener Prozess (die Brownsche Bewegung) W „gesteuert". Damit ergibt sich dann als Lösung von (1.7) der stochastische Prozess $(S_t)_{t\geq 0}$ für die preisliche Entwicklung des Basiswertes.

Aus Annahmen über den Markt (Arbitragefreiheit) kann nun mit ziemlich tiefliegenden Rechtfertigungen (Itô-Integral, Itô-Lemma) die Anfangs–Randwertaufgabe (1.3), (1.4), (1.6) abgeleitet werden; dazu später.

30. November 2009, 17:46, NZZ Online

Das Comeback der Volatilität

Schlingernde Börsen nach Dubai-Schock Die Nervosität hält wieder an den Weltbörsen Einzug.
Die Wall Street kämpft weiterhin mit dem Dubai-Schock. Die Börsen reagieren nervös. Die Volatilität ist gestiegen und die europäischen Börsen verloren deutlich an Boden. Die Wall Street hat sich noch nicht für eine Richtung entschieden.

Abbildung 1.2: Volatil und nun?

Damit kennen wir nun alle mathematischen Objekte, die im Rahmen der Optionspreisentwicklung von Bedeutung sind: den stochastischen Ansatz (1.7), (1.8), das deterministische Äquivalent (1.3), (1.4), (1.6), die Lösungsformel für den Optionspreis selbst. Damit sieht es so aus, dass das Problem Optionspreisentwicklung sich auf die Lösung einer bestens untersuchten Gleichung reduzieren lässt. Dies ist ein Trugschluss, da die obige Reduktion nur für die obige sehr spezielle Modellierung durchführbar ist, eine Modellierung der Optionspreisbildung, die mehr realitätsbezogen ist, entzieht sich einer solchen einfachen Reduktion. Hier sind Aufweichungen der Annahmen über den Markt nötig, sind andere stochastische Differentialgleichungen (1.7) zu betrachten, gibt es Variationen der Auszahlungsfunktion. Als Konsequenz steht eine Lösungsformel (1.2) nicht mehr zur Verfügung und es müssen Verfahren zur Lösung von stochastischen Differentialgleichungen entwickelt werden.

1.3 Volatilität

Volatilität ist ein wichtiger Begriff der Finanzmathematik, um den herum sich viele interessante mathematische Fragen stellen.

[3]Wir schreiben in Differentialgleichungen die abhängige Variable „autonom", also nicht in Abhängigkeit von der unabhängigen Variablen, etwa der Zeit, des Ortes,. . . . Gerade bei stochastischen Differentialgleichungen sind wir dabei nicht immer in Übereinstimmung mit der Literatur.

1.3.1 Rendite und Risiko

Rendite bezeichnet den Gesamterfolg einer Kapitalanlage, gemessen als tatsächliche Verzinsung des eingesetzten Kapitals. Sie beruht auf den Ertragseinnahmen (z.B. Zinsen, Dividenden, realisierte Kursgewinne) und den Kursveränderungen. Die Rendite soll erkennbar machen, wie gut sich eine früher angelegte Kapitalanlage entwickelt hat. Rendite wird meist in Prozent und jährlich angegeben.

Mit dem Begriff **Risiko** bezeichnet man in der Finanzwelt die Unsicherheit, mit der die erwarteten Renditen auch wirklich eintreten. Je stärker das Risiko einer Anlageform ist, um so stärker schwankt die Wertentwicklung im Zeitverlauf und umgekehrt.[4] Das Instrument um diese Unregelmäßigkeit oder Flatterhaftigkeit der Renditeentwicklungen zu messen, ist die sogenannte **Volatilität**[5]. Sie misst die Schwankungsbreite des Kurses des Basiswertes für Kursbewegungen innerhalb eines bestimmten Zeitrahmens. Als erste Information (σ ist die Volatilität des Basisobjekts):

$$\sigma^2 = \frac{\mathbb{V}\left(\ln(S_t/S_s)\right)}{t-s}$$

Die Volatilität ist neben der Preisdynamik des Basiswertes der wesentlichste Einflussfaktor für die Optionspreisberechnung.

In einer deterministischen Sichtweise wird die Volatilität als Konstante (Black-Scholes-Modell), in einer Verallgemeinerung auch als Funktion der Zeit, in die Modellgleichungen eingebracht. Die Volatilitätsgröße ist aber keine direkt beobachtbare Größe. Sie ist daher aus Marktdaten zu „schätzen".

Grundsätzlich muss man zwischen **historischer** und **impliziter Volatilität** unterscheiden, solange wir die Volatilität als eine Konstante betrachten. Einen anderen Ansatz stellt die Betrachtung der Volatilität als stochastische Größe dar. Hier wird die Volatilität als Variable einer stochastischen Differentialgleichung „errechnet".

1.3.2 Historische Volatilität

Die historische Volatilität eines Basisobjekts gibt die auf einen Zeitraum bezogene Schwankungsbreite des Kursverlaufs in der Vergangenheit an. Bei der Ermittlung der historischen Volatilität wird auf die Standardabweichung zurückgegriffen, allerdings gehen hier logarithmierte Renditen in die Berechnung ein. Die Rendite wird als Quotient aus dem aktuellen Kurs und seinem Vortagskurs ermittelt. Die logarithmierten Renditen werden verwendet, da diese eher der Normalverteilung folgen. Dann wird das Ergebnis der Volatilitätsberechnung auf ein Jahr annualisiert, da so ein Vergleich besser gelingt.[6] Als Schätzwert für die zukünftige Volatilität geht die historische Volatilität in die Ermittlung des fairen Preises für Optionen ein.

Die auf der Basis der vergangenen 30 und 250 Tage berechneten historischen Volatilitäten der wichtigsten Indizes (beispielsweise DAX, MDAX, SDAX, ...) werden börsentäglich veröffentlicht.

1.3.3 Implizite Volatilität

Die implizite Volatilität entspricht der vom Markt geschätzten Volatilität, welche die erwartete Schwankungsbreite des Basiswertes bis zum Ende der Laufzeit der Option misst. Was drückt diese Art von Volatilität letztendlich aus? Sie ist das Resultat eines theoretischen Modells. In einem

[4] Diese Binsenweisheit wollen nicht alle akzeptieren und reissen damit sich (o.k.) und andere, ja ganze Staaten ins Unglück. Man sollte sich an folgende Weisheit (Andrè Kostelany) halten: *Man sollte wissen, dass hinter den Fassaden großer Finanzinstitute keine Musterknaben sitzen.*

[5] lat. volare: fliegen; volatilis: fliegend, flüchtig

[6] Wenn die Berechnung auf Tagesbasis erfolgte, wird das Ergebnis mit der Wurzel aus 252 multipliziert, bei Wochendaten wird die Wurzel aus 52 und bei Monatsdaten die Wurzel aus zwölf zur Annualisierung verwendet.

klassischen Black-Scholes-Modell geht die historische Volatilität als Konstante ein. Die implizite Volatilität ist dagegen im allgemeinen keine Konstante, sondern eine Funktion der Restlaufzeit und des Ausübungspreises. Sie ergibt sich durch „Abgleich“ der Werte der Preisformel des Black-Scholes-Modells mit den am Markt beobachteten Optionspreisen. Die Berechnung der impliziten Volatilität wird auch als *das* **inverse Problem** der Finanzmathematik bezeichnet. Dass die implizite Volatilität im Allgemeinen keine Konstante ist, könnte auch als wesentliche Schwäche des Black-Scholes-Modells bezeichnet werden.

Bemerkung 1.6 *Die Volatilität hat in den letzten Jahren eine immer größere Beachtung gewonnen. Dies begründet sich hauptsächlich dadurch, dass sich Derivate, also Finanzinstrumente, deren Wert sich vom Kurs eines Basisobjekts ableiten, zunehmender Beliebtheit erfreuen und auch die Volatilität selbst immer häufiger als Anlageklasse (Volatilitätsindizes) entdeckt wird.* □

1.3.4 Lokale Volatilität

Die Einführung der **lokalen Volatilität** ist der Versuch, das Black-Scholes–Modell zu erweitern. Die konstante Volatilität wird ersetzt durch eine Funktion der Zeit und des aktuellen Basiswertes. Mathematisch bedeutet dies, dass die Black-Scholes–Anfangs-Randwertaufgabe dahingehend abzuändern ist. Als Konsequenz haben wir aber, dass eine geschlossene Preisformel nicht mehr herleitbar ist. Die Berechnung lokaler Volatilitäten beschreiben wir in späteren Kapiteln; als Vorgriff siehe [3, 21].

1.3.5 Stochastische Volatilität

Stochastische Volatilität ist eines der Hauptkonzepte zur Behandlung von zeitveränderlichen Volatilitäten in Finanzmärkten. Sie kann als Abhilfe dafür angesehen werden, dass die implizite Volatilität in ihrer Abhängigkeit von der Restlaufzeit und dem Ausübungspreis einen Gegensatz zu den Modellannahmen aufzeigt.

Die stochastische Volatilität wird als stochastischer Prozess dem Prozess für den Basiswert zur Seite gestellt (siehe (1.7)):

$$dS = \mu S\,dt + \sigma S\,dW_t \tag{1.9}$$

$$d\sigma = \lambda_t dt + \xi_t dW_t^v \tag{1.10}$$

Dabei ist nun W_t^v ein weiterer Wiener Prozess/eine weitere Brownsche Bewegung; die Parameter λ_t, ξ_t sind zu wählen. Damit hat man es nun mit einer gekoppelten Dynamik zu tun, um den Endwert S_T zu ermitteln. Ein Konkretisierung ist das so genannte Heston-Modell (siehe [15])

$$dS = \mu S\,dt + \sqrt{v^+}S\,dW_t \tag{1.11}$$

$$dv = \kappa(\theta - v)dt + \sqrt{v^+}dW_t^v \tag{1.12}$$

mit $v = \sigma^2$.

1.4 Auswertung der Preisformel

In nächsten Abschnitt wollen wir die implizite Volatilität σ_{impl} als Nullstelle einer nichtlinearen Gleichung aus den Optionspreisen, gegeben durch (1.2), berechnen. Dazu ist in jedem Iterationsschritt des Verfahrens die Formel (1.2) auszuwerten. Wie kann dies numerisch geschehen? Die wesentliche Aufgabe dabei ist, die Funktion $\mathcal{N}$ auszuwerten. Hierzu gibt es verschiedene Vorgehensweisen: Interpolationsverfahren für den Integranden, Quadraturverfahren, Approximation durch rationale Polynome in Teilgebieten von $[0, \infty)$.

Wir skizzieren nun eine Methode, die diese Ideen verbindet. Wegen $\mathcal{N}(0) = \frac{1}{2}$ und

$$\begin{aligned}\mathcal{N}(x) &= \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \int_0^x \exp(-t^2/2)dt \\ &= \frac{1}{2} + \frac{1}{\sqrt{\pi}} \int_0^{x/\sqrt{2}} \exp(-t^2)dt \\ &= \frac{1}{2}\left(1 + \frac{2}{\sqrt{\pi}} \int_0^{x/\sqrt{2}} \exp(-t^2)dt\right)\end{aligned}$$

reicht es, für die so genannte **Fehlerfunktion** (Gaußsches Fehlerintegral)

$$\mathbf{erf}(x) := \frac{2}{\sqrt{\pi}} \int_0^x \exp(-t^2)dt\,,\ x \geq 0, \tag{1.13}$$

ein Berechnungsverfahren vorzustellen. Programmpakete stellen die Auswertung von **erf** bereit. Wir skizzieren eine Approximationsmethode für **erf**, die unter bescheidenem Aufwand ganz gute Ergebnisse liefert. Sie basiert auf einer besten Approximation im Sinne der Fehlerquadratmethode. Zunächst ein paar Beobachtungen.

$$\mathbf{erf}(0) = 0\,,\ \lim_{x\to\infty} \mathbf{erf}(x) = 1\,,\ \lim_{x\to\infty} \frac{1-\mathbf{erf}(x)}{\mathbf{erf}'(x)} = 0$$

Dabei ergibt sich die letzte Beobachtung aus

$$\frac{1-\mathbf{erf}(x)}{\mathbf{erf}'(x)} = \frac{\frac{2}{\sqrt{\pi}}\int_0^\infty \exp(-t^2)dt - \frac{2}{\sqrt{\pi}}\int_0^x \exp(-t^2)dt}{\frac{2}{\sqrt{\pi}}\exp(-x^2)} = \int_x^\infty \exp(x^2-t^2)dt\,.$$

Die obigen asymptotischen Verhaltensweisen sollten in die Überlegungen eingebaut werden. Unter Verwendung der Variablen $\eta := (1+px)^{-1}, p > 0$, machen wir den Ansatz

$$\frac{1-\mathbf{erf}(x)}{\mathbf{erf}'(x)} := a_1\eta + a_2\eta^2 + a_3\eta^3 + \cdots \tag{1.14}$$

und erhalten

$$\mathbf{erf}(x) = 1 - (a_1\eta + a_2\eta^2 + a_3\eta^3 + \cdots)\mathbf{erf}'(x) \text{ mit } \eta = (1+px)^{-1}\,. \tag{1.15}$$

Beachte, dass die Auswertung von

$$\mathbf{erf}'(x) = \frac{2}{\sqrt{\pi}}\exp(-x^2)$$

keine Probleme bereitet. Als zu bestimmende Parameter haben wir $a_1, a_2, a_3, \ldots, p$. Nun ist ein Vorgehen anzugeben, das

- das Problem auf endlich viele Parameter $a_1, \ldots, a_N, p$ reduziert,
- für die verbleibenden Parameter $a_1, \ldots, a_N, p$ eine Bestimmungsvorschrift angibt,
- den Fehler, der bei der Reduktion auf die endlich vielen Parameter entsteht, beherrschbar macht.

Reduktion auf vier Parameter Gesucht sind die Parameter a_1, a_2, a_3 und der Parameter p.
Wir setzen an: $\mathbf{erf}^*(x) := 1 - (a_1\eta + a_2\eta^2 + a_3\eta^3)\mathbf{erf}'(x)$ mit $\eta = (1+px)^{-1}$.

Exaktheit der Approximation Wir fordern die Exaktheit für $x = 0$. Dies bedeutet

$$1 - (a_1 + a_2 + a_3)\frac{2}{\sqrt{\pi}} = 0 \text{ d.h. } a_3 = \frac{\sqrt{\pi}}{2} - a_1 - a_2 \,.$$

Bestimmung der restlichen Parameter Nun verbleiben p, a_1, a_2. Wir fordern, dass der Fehler

$$e := \int_0^\infty |\mathbf{erf}^*(x) - \mathbf{erf}(x)|^2 dx$$

in einer diskreten Version minimal wird.

Zur Realisierung des letzten Schritts wählen wir Stützstellen $0 = x_0 < x_1 < \cdots < x_n$ und minimieren[7]

$$g(y) := \frac{1}{n+1}|F(y)|^2 \text{ mit } F(y) := (\mathbf{erf}^*(x_i) - \mathbf{erf}(x_i))_{i=0,\ldots,n} \,,\, y := (p, a_1, a_2)\,;$$

beachte die Abhängigkeit von $\mathbf{erf}^*$ von $y := (p, a_1, a_2)$. Die notwendige Bedingung für ein lokales Minimum von g ist offenbar[8]

$$G(y) := DF(y)^t F(y) = \theta \,.$$

Das modifizierte Newtonverfahren angewendet, ergibt die Iteration

$$y^{k+1} := y^k - \lambda_k DG(y^k)^{-1} G(y^k) \,,\, k = 0, 1, \ldots .$$

Wir haben

$$DG(y) = DF(y)^t DF(y) + D^2F(y)^t F(y) \,.$$

Es ist die Idee der Gauß-Newtonverfahren, den Term $D^2F(y)^tF(y)$ zu streichen, um die Bildung zweiter Ableitungen zu vermeiden. Hier lässt sich auch eine inhaltliche Begründung dafür liefern: wir gehen ja davon aus, dass $F(y) \approx \theta$ gilt. Damit erhält man schließlich folgende Iteration:

$$y^{k+1} = y^k - \lambda_k (DF(y^k)^t DF(y^k))^{-1} DF(y^k)^t F(y^k) \,,\, k = 0, 1, \ldots . \tag{1.16}$$

Die Schrittweitensteuerung $(\lambda_k)_{k\in\mathbb{N}}$ nimmt man im allgemeinen so vor, dass ein quantifizierbarer Abstieg im Zielfunktional $y \longmapsto g(y)$ realisiert wird.

Die Idee der Levenberg–Marquardt-Verfahren stabilisiert das Gleichungssystem in (1.16) dadurch, dass $DF(y^k)^tDF(y^k)$ durch $DF(y^k)^tDF(y^k) + \alpha I$ mit einem geeigneten $\alpha > 0$ ersetzt wird.

Bemerkung 1.7 *Die Wahl der Stützstellen $x_1, \ldots, x_n$ haben wir noch offen gelassen. Es reicht im allgemeinen $n = 3$ und die Wahl $x_1, x_2, x_3 \in (0, 1]$ schon aus:*

$$p = 0.3367, a_1 = 0.4361836, a_2 = -0.1201676, a_3 = 0.93729844$$

Der maximale Fehler der resultierenden Approximation ist kleiner als 10^{-5}; siehe [1]. Für größere Genauigkeit siehe [14].

1.5 Berechnung der Volatilität

Wir betrachten die Berechnung der historischen und impliziten Volatilität. Wir gehen nicht auf die Berechnung der lokalen Volatilität ein, dieser Aufgabe widmen wir ein eigenes Kapitel. Dort werden wir auch auf den Aspekt „ill-posedness“ eingehen.

[7] Mit $|\cdot|$ bezeichnen wir den euklidischen Abstand in $\mathbb{R}^d$.

[8] Wir benutzen die Bezeichnung A^t für die Transponierte einer Matrix A. Zur Unterscheidung von einer skalaren Null schreiben wir eine vektorielle Null mit θ.

1.5.1 Historische Volatilität

Die historische Volatilität $\sigma = \sigma_{hist}$ ist durch die Basiswertkurse aus der Vergangenheit gegeben. Mathematisch gesehen ist σ_{hist} die annualisierte Standardabweichung der logarithmischen Kursänderungen. Kennt man die historische Volatilität, lässt sie sich verwenden in einem Modell für die Berechnung von Optionspreisen, das ja die Volatilität in der Zukunft $[0, T]$ benötigt.

Hier ist das Vorgehen:

Algorithm 1.1 Bestimmung der historischen Volatilität

EIN Kurswerte $S_1, S_2, \ldots, S_N$ zu den Zeitpunkten $t_1 < t_2 < \cdots < t_N$.

Schritt 1 Setze $\delta_i := \ln(S_{i+1}) - \ln(S_i), i = 1, \ldots, N-1$.

Schritt 2 Berechnung des Mittelwerts (Erwartungswerts)

$$\overline{\delta} := \frac{1}{N-1} \sum_{i=1}^{N-1} \delta_i \tag{1.17}$$

Schritt 3 Berechnung der historische Volatilität (empirischen Standardabweichung)

$$\sigma_{hist} := \sqrt{N} \left(\frac{1}{N-2} \sum_{i=1}^{N-1} (\delta_i - \overline{\delta})^2 \right)^{\frac{1}{2}} \tag{1.18}$$

AUS „Schätzwert“ σ_{hist} für die historische Volatilität.

Im obigen Algorithmus steht N im Allgemeinen für die (durchschnittliche) Anzahl der Börsentage (252!) im Jahr. Das Vorgehen ist nur eine Möglichkeit von vielen. Beispielsweise läge es nahe aktuellere Basiswerte stärker zu gewichten als ältere (was auf eine gewichtetete l_2-Norm in $\mathbb{R}^{N-1}$ hinausliefe). Festzuhalten ist, dass es allgemeine Ansicht ist, dass die historische Volatilität, berechnet wie auch immer, ein schlechter Schätzer für die zukünftige Volatilität ist. Der Bezeichnung „Schätzer“ kann im Sinne der Statistik eine Bedeutung gegeben werden.

1.5.2 Implizite Volatilität

Die implizite Volatilität σ_{impl} ist diejenige Volatilität, die bei Unterstellung des Black-Scholes-Modells in einem Marktpreis (einer europäischen) Option zum Ausdruck kommt. Hat man ein Modell für einen Optionspreis, das zu einer (geschlossenen) Formel für den Optionspreis führt, die auch noch die Volatilität σ explizit enthält, dann kann man versuchen, daraus die Volatilität zu berechnen, indem man „die Formel nach σ auflöst“ und so σ_{impl} erhält.[9] Voraussetzung ist, man kennt die Marktpreise der Option.

In Abschnitt 1.2 haben wir die Black-Scholes-Preisformel für eine europäische Call-Option kennengelernt. Wir schreiben sie detailierter nochmals auf:

$$C(S, K, \tau, r, \sigma) := S\mathcal{N}(d_+(\sigma, S, K, \tau, r)) - Ke^{-r\tau}\mathcal{N}(d_-(\sigma, S, K, \tau, r)) \,, \tag{1.19}$$

[9] Die Deutsche Börse AG hat am 5. Dezember 1994 den DAX-Volatilitätsindex (VDAX) als Barometer für implizite Volatilitäten der an der Deutschen Terminbörse gehandelten DAX-Optionen eingeführt. Er wird täglich mit Hilfe des Black-Scholes-Modells aus selektierten Optionsserien zweier Verfallstermine berechnet.

wobei

$$d_\pm(\sigma,S,K,\tau,r) := \frac{\ln(\frac{S}{K}) + (r \pm \frac{\sigma^2}{2})\tau}{\sigma\sqrt{\tau}},$$

$\mathcal{N}$ die Verteilungsfunktion der Standardnormalverteilung und τ die Restlaufzeit $T-t$ bezeichnet. Beachte

$$d_-(\sigma,S,K,\tau,r) = d_+(\sigma,S,K,\tau,r) - \sigma\sqrt{\tau}\,.$$

Also hat man zur Berechnung der impliziten Volatilität in diesem Modell die Gleichung

$$C(S,K,\tau,r,\sigma) = v := C_{Markt} \tag{1.20}$$

nach σ aufzulösen. Dabei haben wir anzunehmen, dass alle anderen Parameter bekannt sind. Wir unterdrücken dann diese Parameter und setzen

$$f(\sigma) := C(S,K,\tau,r,\sigma)\,,\ d_+(\sigma) := d_+(\sigma,S,K,\tau,r) \tag{1.21}$$

Damit haben wir nun die Gleichung

$$f(\sigma) - v = 0 \tag{1.22}$$

zu betrachten. Eine Auflösung wird explizit nicht gelingen, da die Funktion f „hochgradig nichtlinear“ ist. Also greifen wir zur numerischen Lösung der Gleichung und wenden das Newtonverfahren an; siehe Anhang 1.6. Die Newtoniteration sieht so aus

---

Algorithm 1.2 Newtonverfahren zur Bestimmung der impliziten Volatilität

---

EIN Basiskurs S, Ausübungspreis K, Restlaufzeit τ, Zinssatz r; Näherungswert σ^0 für die Volatilität. Iterationsschranke k_{it} .

Schritt 0 Setze $f(\sigma) := C(S,K,\tau,r,\sigma)\,,\ d_+(\sigma) := d_+(\sigma,S,K,\tau,r)$

Schritt 1 Berechne für $k = 0,\dots,k_{it}-1$

$$\sigma^{k+1} = \sigma^k - \frac{f(\sigma^k)-v}{f'(\sigma^k)} = \sigma^k - \frac{f(\sigma^k)-v}{\nu(\sigma^k)}, \tag{1.23}$$

Hierbei ist

$$\nu(\sigma) := S\sqrt{\tau}\mathcal{N}'(d_+(\sigma)) = S\sqrt{\tau}\frac{1}{\sqrt{2\pi}}\exp(-d_+(\sigma)^2/2)\,.$$

AUS Näherungswert $\sigma^{k_{it}}$ für die implizite Volatilität σ_{imp} .

ν im obigen Algorithmus ist eine Kennzahl, die den **Griechen (greeks)**[10] zugerechnet wird; sie ist in der unten angeführten Liste das Vega. Diese Kennzahlen sind die Ableitungen (Sensitivitäten) des Optionspreises C bezüglich der zugrundeliegenden Parameter und Variablen. Sie werden als **Risikokennziffern** verwendet.

Delta $\Delta := \frac{\partial C}{\partial S}$ misst die Sensitivität des Optionspreises bezüglich Änderungen der Basiskurse und wird oft als Hedge-Parameter verwendet. Es ist das Ziel bei der Zusammenstellung eines Portfolios, bestehend aus Basisobjekten und darauf bezogene Optionen, die Abhängigkeit von Variationen des Basiskurses, also Δ, nahe bei Null zu halten.

[10] Sie heißen Griechen, da sie konsequent mit festgelegten griechischen Buchstaben bezeichnet werden. Eine Ausnahme ist das Vega, welches kein griechischer Buchstabe ist.

Gamma $\Gamma := \frac{\partial^2 C}{\partial S^2}$ misst die Sensitivität (zweiter Ordnung) des Optionspreises bezüglich Δ. Damit kann die Änderung eines Portfolios auf große Änderungen des Basiskurses (unter Nutzung einer Taylorentwicklung) hinterfragt werden.

Rho $\mathcal{P} := \frac{\partial C}{\partial r}$ misst die Sensitivität des Optionspreises bezüglich des Zinssatzes.

Vega $\mathcal{V} := \frac{\partial C}{\partial \sigma}$ misst die Sensitivität des Optionspreises bezüglich der Volatilität σ.

Theta $\Theta := -\frac{\partial C}{\partial t}$ misst die Sensitivität des Optionspreises bezüglich der (Restlauf-)Zeit. Dieser Parmeter hebt sich von den anderen ab, da er Veränderung bezüglich einer deterministischen Größe misst. Eine Absicherung gegen solche Veränderungen ist nicht sinnvoll.

Diese Griechen sind unabhängig vom Zustandekommen der Optionspreise definiert. Die Black-Scholes-Gleichung kann mit Hilfe der Griechen und des Optionspreises so aufgeschrieben werden:

$$\Theta + rS\Delta + \frac{1}{2}\sigma^2 S^2 \Gamma = tC$$

Für die Preisformel nach dem Black-Scholes Modell existieren geschlossene Formeln für die Griechen. Für Call-Optionen:

- $\Delta = \mathcal{N}(d_+)$.
- $\Gamma = \mathcal{N}(d_+)/S\sigma\sqrt{T}$.
- $\mathcal{V} = \mathcal{N}(d_+)\sqrt{T}$.

Nun sind die Voraussetzungen des Satzes über die Konvergenz des Newtonverfahrens zu klären. Wir tun dies unter der Voraussetzung

$$S > 0, \tau > 0,$$

die sicherlich nicht problematisch ist.

Differenzierbarkeit Offensichtlich ist die Funktion f unendlich oft differenzierbar; wir haben die erste Ableitung oben schon ausgerechnet:

$$f'(\sigma) = S\sqrt{\tau}\mathcal{N}'(d_+(\sigma)),$$

Nun ist f' positiv, die Durchführbarkeit des Newtonverfahrens ist daher gesichert.

Monotonie Da die erste Ableitung positiv ist, ist f strikt monoton wachsend. Dies bedeutet, dass eine Lösung von (1.22) eindeutig bestimmt ist.

Existenz einer Nullstelle Eine Nullsstelle ist gesichert, wenn wir

$$r_l := \lim_{\sigma \to 0} f(\sigma) - v \leq 0\,,\; r_u := \lim_{\sigma \to \infty} f(\sigma) - v \geq 0 \tag{1.24}$$

nachweisen können, denn wegen der strikten Monotonie von f gilt dann, dass mindestens eine der Ungleichungen $r_l < 0, r_u > 0$ gilt. Auf Grund der Stetigkeit von f gibt es dann eine Lösung in (1.22).

Es gilt offenbar $\lim_{\sigma\to 0} f(\sigma) = (S - Ke^{-r\tau})^+$ und $\lim_{\sigma\to\infty} f(\sigma) = S$. Aus der Monotonie von f folgen die Ungleichungen

$$(S - Ke^{-r\tau})^+ \leq C(S, K, \tau, r, \sigma) \leq S\,.$$

(Später leiten wir die Ungleichungen aus Annahmen über den Markt (Arbitragefreiheit) her.)

Damit tritt (1.24) ein für alle Optionspreise v; siehe folgende Bemerkung 1.8.

Startwert Ein Startwert kann mit einer Bisektionsmethode bestimmt werden: finde ein Intervall $[\sigma_l, \sigma_u]$ mit $f(\sigma_l) - v \leq 0, f(\sigma_u) - v \geq 0$, und wähle $\sigma^0 \in [\sigma_l, \sigma_u]$.

Konvergenzordnung Die Voraussetzungen für die quadratische Konvergenz sind gegeben, wenn der Startwert nahe genug bei der Lösung liegt. Notfalls wende man das modifizierte Newtonverfahren an.

Damit ist nun klar, dass das Newtonverfahren sehr gut anwendbar ist. Es liefert gute Ergebnisse, wie viele Dokumentationen zeigen.

Bei der Berechnung der impliziten Volatilität stellt man eine Abhängigkeit vom Ausübungspreis K der Option (bzw. deren Moneyness) und/oder der Restlaufzeit τ fest. Dies steht im Widerspruch zum Black-Scholes-Modell oder anders ausgedrückt, das Black-Scholes-Modell beschreibt das Marktgeschehen nicht korrekt. Trägt man die implizite Volatilität in Abhängigkeit des Ausübungspreises auf, so erhält man einen Funktionsgraphen, der konvex ist, und das „umso mehr“, je kürzer die Restlaufzeit ist. Man nennt dies den **Smile-Effekt**.

Bemerkung 1.8 *Die Lösbarkeit der Gleichung 1.22 ist im Zweifel, wenn wir unterstellen, dass die Marktpreise nicht dem Black-Scholes Modell entsprechen, was nicht abwegig ist, denn wir haben ja schon eine solche Diskrepanz bei der impliziten Volatilität oben festgehalten. Es kann daher nicht ausgeschlossen werden, dass der Marktpreis außerhalb des Intervalls* $((S-Ke^{-r\tau})^+, S)$ *liegt. Dieser Sachverhalt sollte also ausgeschlossen werden. Numerisch ist aber schon heikel, wenn der beobachtete Marktpreis* v *am Rand dieses Intervalls liegt, denn dann liegt auf Grund des asymptotischen Verhaltens von* f *– beachte, dass die Berührung des Graphen von* f *der Asymptoten* $y = (S - Ke^{-r\tau})^+$ *und* $y = S$ *„schleifend“ ist – eine hohe Instabilität einer Lösung der Gleichung* (1.22) *vor.* □

In Kapitel **??** werden wir eine andere Möglichkeit, die implizite Volatilität zu berechen, kennenlernen. Sie resultiert aus der Tatsache, dass man aus der Black-Scholes-Gleichung eine Gleichung für den Optionspreis in Abhängigkeit von K und T ableiten kann.

1.6 Anhang: Newtonverfahren

Zwar gibt es viele Techniken für die Suche nach Nullstellen, eines der am häufigsten verwendeten Verfahren ist das Newton-Verfahren, denn es bietet im allgemeinen rasche Konvergenz.

1.6.1 Nullstellensuche nach Newton bei Polynomen

Sir Isaac Newton beschreibt[11] ein Rechenverfahren zum Lösen einer polynomialen Gleichung und begründet damit ein Verfahren, das heutzutage als **Newton-Verfahren** bezeichnet wird. Er tut dies am Beispiel des Polynoms $p(x) := x^3 - 2x - 5 = 0$. Eine leicht zu erratende Näherung „0-ter Ordnung“ ist $x^0 = 2$, denn $p(2) = -1$ ist „klein“. Newton machte den Ansatz $x = 2 + u$ mit einem als „klein“ angenommenen u und setzte diesen Ansatz in die Gleichung ein. Es gilt:

$$x^3 = (2+u)^3 = 8 + 12u + 6u^2 + u^3\,,\ 2x = 2(2+u) = 4 + 2u\,.$$

Also folgt

$$x^3 - 2x - 5 \ = \ -1 + 10u + 6u^2 + u^3 \ \stackrel{!}{=}\ 0\,.$$

[11] Isaac Newton, 1643–1727; „Methodus fluxionum et serierum infinitarum“

Da u als „klein" angenommen wurde, können die Terme höherer Ordnung gegen den linearen und konstanten Anteil vernachlässigt werden, womit $10u - 1 = 0$ bzw. $u = 0.1$ übrig bleibt. Als Näherung x^1 1-ter Ordnung resultiert $x^1 = 2.1$.

Wir können nun dieses Vorgehen wiederholen: wir setzen $u = 0.1 + v$ an, betrachten die Gleichung $p(2 + 0.1 + v) = 0$, berücksichtigen wiederum nur den linearen Anteil und erhalten so $v = -0.061/11.23 = -0.0054\ldots$. Als Näherung x^2 2-ter Ordnung resultiert $x^2 = 2.0946$.

Raphson[12] beschrieb diesen Rechenprozess formal und illustrierte den Formalismus an der allgemeinen Gleichung 3. Grades, die abstrakte Form des Verfahrens mit Benutzung von Ableitungen stammt von Thomas Simpson. Zur Simpsonschen Form kommen wir nun.

Sei $f : \mathbb{R} \longrightarrow \mathbb{R}$. Eine Nullstelle wird nach folgendem Vorgehen gesucht:

(1) Man rät eine Näherung x^0. O.E. $f(x^0) \neq 0$.

(2) Man berechnet/zeichnet die Tangente t^0 an den Graphen von f im Punkt $(x^0, f(x^0))$.

(3) Man berechnet/konstruiert die Nullstelle x^1 der Tangente.

(4) Man setzt $x^0 := x^1$ und wiederholt den Vorgang, beginnend bei (1).

Klar, um die Tangente bestimmen zu können, müssen wir voraussetzen, dass diese existiert, was die Differenzierbarkeit von f bedeutet. Dann lautet die Tangentengleichung

$$t^0 \,:\, y = f(x^0) + f'(x^0)(x - x^0) \tag{1.25}$$

und die Berechnung der Nullstelle von t^0 führt zur Formel

$$x^1 = x^0 - f'(x^0)^{-1} f(x^0) \,. \tag{1.26}$$

Hier tritt das Problem auf, dass $f'(x^0) \neq 0$ gelten muss, d.h. dass f in $(x^0, f(x^0))$ keine waagrechte Tangente besitzt. Von der Anschauung her, keine überraschende Forderung, von der Analyse des Verfahrens her eine Forderung, die sukzessive oder a-priori sichergestellt werden muss.

Schreiben wir das Verfahren nun kompakt auf:

$$x^{n+1} := x^n - f'(x^n)^{-1} f(x^n) \,,\; n = 0, \ldots \,. \tag{1.27}$$

Dabei ist die **Startnäherung** x^0 zu wählen. Wir nennen dieses Vorgehen nun **Newton–Verfahren**; siehe Abbildung 1.3.

Das Newton–Verfahren ist ein so genanntes lokal konvergentes Verfahren. Konvergenz der in der Newton–Iteration erzeugten Folge zu einer Nullstelle ist also nur garantiert, wenn der Startwert, d.h. das 0-te Glied der Folge, schon „ausreichend nahe" an der Nullstelle liegt. Ist der Startwert nicht gut genug, so haben wir zu rechnen mit:

- Die Folge divergiert, der Abstand zur Nullstelle wächst über alle Grenzen.
- Die Folge divergiert, bleibt aber beschränkt. Sie kann z.B. periodisch werden, d.h. endlich viele Punkte wechseln sich in immer derselben Reihenfolge ab. Man sagt auch, dass die Folge oszilliert (Bei $f(x) := x^3 - 2x + 2$ ist dies machbar).
- Die Folge konvergiert, falls die Funktion mehrere Nullstellen hat, gegen eine andere als die gewünschte Nullstelle konvergieren; in der Abbildung 1.3 kann man dies erahnen.

[12] Joseph Raphson, 1648–1715; Arbeit „Analysis Aequationum universalis"

Ist der Startwert x^0 so gewählt, dass das Newton–Verfahren konvergiert, so ist die Konvergenz allerdings quadratisch, also mit der Konvergenzordnung 2 (falls die Ableitung an der Nullstelle nicht verschwindet).

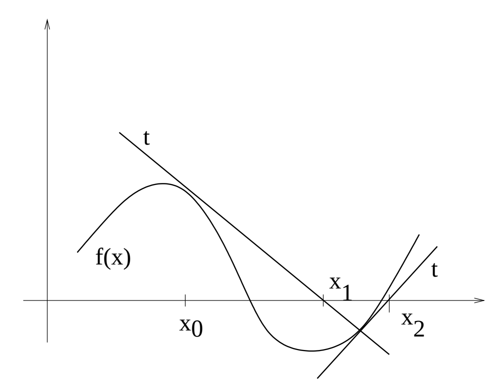

Abbildung 1.3: Newtonverfahren

Bemerkung 1.9 *Wie ordnet sich das Newtonsche Vorgehen hier nun ein? Ausgehend von der Startnäherung $x^0 = 2$ wird ein Newtonschritt auf die Nullstellengleichung $p(x + 2) = 0$ mit $x = 0$ als Startnäherung angewendet:*

$$x^1 := 0 - \frac{p(2)}{p'(2)} = \frac{1}{10}\,.$$

Nun betrachtet man die Nullstellengleichung $p(x + 2.1) = 0$ mit $x = 0$ als Startnäherung und wendet wieder einen Newtonschritt mit Ausgangsnäherung $x = 0$ an:

$$x^2 := 0 - \frac{p(2.1)}{p'(2.1)} = \frac{0.061}{11.23}\,.$$

Und so weiter! □

Viele nichtlineare Gleichungen haben mehrere Lösungen, so hat ein Polynom n-ten Grades bis zu n (reelle) Nullstellen. Will man alle Nullstellen in einem bestimmten Bereich $D \subset \mathbb{R}$ ermitteln, so muss zu jeder Nullstelle ein passender Startwert in D gefunden werden, für den das Newton–Verfahren konvergiert. Ein beliebtes Vorgehen dazu besteht in Einschachtelungsverfahren: zwischen zwei Punkten z^1, z^2, so dass $f(z^1), f(z^2)$ unterschiedliche Vorzeichen besitzen, liegt immer eine Nullstelle von f, da wir ja Differenzierbarkeit von f (und damit Stetigkeit) voraussetzen.

Beispiel 1.10 *Ein Spezialfall des Newtonschen Näherungsverfahrens ist das Babylonische Wurzelziehen, auch bekannt als Heronverfahren nach Heron von Alexandria: Wendet man das Verfahren zur Nullstellenbestimmung auf die Funktion $f(x) := x^2 - a\,(a > 0)$, so erhält man wegen der Ableitungsfunktion $f'(x) = 2x$ für die Lösung $\sqrt{a}$ das Näherungsverfahren*

$$x^{n+1} := x^n - \frac{(x^n)^2 - a}{2x^n} = \frac{1}{2}\left(x^n + \frac{a}{x^n}\right)\,.$$

Dieses Verfahren konvergiert für jedes $a \geq 0$ und für jeden beliebigen Anfangswert $x^0 > 0$. □

Beispiel 1.11 *Die Quadratwurzel einer Zahl $a > 0$ sind die Nullstellen der Funktion $f(x) := 1 - a/x^2$. Diese Funktion hat die Ableitung $f'(x) = 2a/x^3$, die Newton-Iteration erfolgt also nach der Vorschrift*

$$x^{n+1} := x^n - \frac{(x^n)^3}{2a} + \frac{x^n}{2} = \frac{x^n}{2}\left(3 - \frac{(x^n)^2}{a}\right)\,.$$

Der Vorteil dieser Vorschrift gegenüber dem Wurzelziehen nach Heron (siehe Beispiel 1.10) ist, dass es divisionsfrei ist, sobald einmal der Kehrwert von a bestimmt wurde. Als Startwert wurde in der Tabelle $x^0 := (1 + a)/2$ gewählt. Die Iterierten wurden an der ersten ungenauen Stelle *abgeschnitten. Es ist zu erkennen, dass nach wenigen Schritten die Anzahl gültiger Stellen schnell wächst.*

n	x^n *bei* $a = 2$	x^n *bei* $a = 3$	x^n *bei* $a = 5$
0	$1,5$	2	3
1	$1,40$	$1,6$	$1,8$
2	$1,4141$	$1,72$	$2,1$
3	$1,41421355$	$1,73203$	$2,22$
4	$1,41421356237309502$	$1,7320508074$	$2,23601$
5	$1,414213562373095048801688724209697$	$1,73205080756887729351$	$2,236067975$

$\square$

Bei der Bestimmung von Nullstellen von Polynomen ist folgender Hinweis wichtig: hat man eine Nullstelle z^0 gefunden, so kann man diese Nullstelle aus dem Polynom „entfernen durch Polynomdivision durch den Linearfaktor $x - z^0$; man hat so den Grad des Polynoms um eins verkleinert.

Beispiel 1.12 *Betrachte das Polynom*

$$p(x) := x^3 - 3x^2 - x + 3\,.$$

Es hat die Nullstelle $x = 1$, *was man etwa erraten kann. Polynomdivision ergibt*

$$p(x) : (x - 1) = x^2 - 2x - 3$$

und als weitere Nullstellen finden wir $x = 1$ *und* $x = 3$. $\square$

1.6.2 Konvergenz im eindimensionalen Fall

Sei f eine dreimal differenzierbare Funktion mit einer Nullstelle z, in der die erste Ableitung nicht verschwindet, d.h. $f'(z) \neq 0$. Diese Voraussetzung besagt, dass der Graph von f die x-Achse „transversal schneidet“. Wir wissen aus

$$0 = f(z) = f(x) + f'(x)(z - x) + \frac{1}{2} f''(\xi)(z - x)^2$$

und daher

$$x - z = \frac{f(x)}{f'(x)} + \frac{1}{2}\frac{f''(\xi)}{f'(x)}(x - z)^2\,.$$

Nun stellen wir so um, dass wir eine Verbindung mit der Newtoniteration sehen:

$$x - \frac{f(x)}{f'(x)} - z = \frac{1}{2}\frac{f''(\xi)}{f'(x)}(x - z)^2\,.$$

Ist nun I ein Intervall um z, in dem die Ableitung von f nicht verschwindet – dies kann auf Grund der Tatsache, dass $f'(z) \neq 0$ gilt, sichergestellt werden – dann folgt mit

$$m := \inf_{x\in I} |f'(x)|\,,\ M := \max_{x\in I} |f''(x)|\,,\ K := \frac{M}{2m}$$

die Abschätzung

$$|x - \frac{f(x)}{f'(x)} - z| \leq K|x - z|^2\,,\ x \in I\,.$$

Dies hat zur Konsequenz, dass für die Newtoniterierten x^n gilt

$$K|x^n - z| \leq (K|x^0 - z|)^{2^n}\ ,\ n \in \mathbb{N}_0\,,$$

was man mittels vollständiger Induktion beweisen kann. Ist also $K|x^0 - z| < 1$, dann wird $|x^n - z|$ sehr schnell klein. Wir schreiben diese Betrachtungen nun etwas exakter auf.

Satz 1.13 *Sei $f : [a,b] \longrightarrow \mathbb{R}$ zweimal stetig differenzierbar und es gelte*

$$|f'(x)| \geq m\,,\ |f''(x)| \leq M \text{ für alle } x \in [a,b] \tag{1.28}$$

mit $m > 0, M > 0$. Dann gilt:

(a) *f hat in $[a,b]$ höchstens eine Nullstelle.*

(b) *Ist z eine Nullstelle in (a,b), dann ist die Iteration (1.27) definiert für alle $x^0 \in U_r(z) := (z-r, z+r)$ wobei $r := \min(2mM^{-1}, b-z, z-a)$ ist.*
Weiterhin gilt mit $q := M(2m)^{-1}|x^0 - z| < 1$ für alle $n \in \mathbb{N}$:

1. $|z - x^n| \leq \frac{M}{2m}|z - x^{n-1}|^2$ ***(Konvergenzordnung)***
2. $|z - x^n| \leq \frac{2m}{M}q^{2^n}$ ***(a priori Abschätzung)***
3. $|z - x^n| \leq \frac{1}{m}|f(x^n)| \leq \frac{M}{2m}|x^n - x^{n-1}|^2$ ***(a posteriori Abschätzung)***

Beweis:
Seien z^1, z^2 Nullstellen von f in $[a,b]$. Aus

$$0 = |f(z^1) - f(z^2)| = |f'(\eta)||z^1 - z^2|$$

erhalten wir $z^1 = z^2$ und a) ist bewiesen.
Mit der Taylorentwicklung folgt

$$\begin{aligned} 0 &= f(z) = f(x^n) + f'(x^n)(z - x^n) + \frac{1}{2}f''(\eta)(z - x^n)^2\,, \\ 0 &= f(x^n) + f'(x^n)(x^{n+1} - x^n)\,, \end{aligned}$$

und wir erhalten mit Subtraktion

$$0 = (z - x^{n+1})f'(x^n) + \frac{1}{2}f''(\eta)(z - x^n)^2\,;$$

$\eta \in [a,b]$. Dies zeigt

$$|z - x^{n+1}| \leq \frac{M}{2m}|z - x^n|^2.$$

Sei $x^0 \in U_r(z)$. Dann folgt

$$|z - x^1| \leq \frac{M}{2m}|z - x^0|^2 \leq \frac{M}{2m}(\frac{2m}{M})^2 q^2\,.$$

Mittels vollständiger Induktion erhalten wir die a priori Abschätzung.
Es gilt

$$|f(x^{n+1})| = |f(z) - f(x^{n+1})| = |f'(\eta)||z - x^{n+1}| \geq m|z - x^{n+1}|$$

und

$$f(x^{n+1}) = f(x^n - \frac{f(x^n)}{f'(x^n)}) = \frac{1}{2}f''(\xi)(x^{n+1} - x^n)^2$$

was die a posteriori Abschätzung impliziert. ■

Die 1. Abschätzung von (b) in Satz 1.13 besagt, dass die **Konvergenzordnung** der Folge $(x^n)_{n\in\mathbb{N}}$ (mindestens) zwei, also quadratisch ist. Man kann dies so formulieren, dass bei jedem Iterationsschritt sich die Anzahl der signifikanten Stellen der Approximation x^n sich verdoppelt.

Beispiel 1.14 *Betrachte die Funktion $f(x) := x^2, x \in \mathbb{R}$. Die Nullstelle $z := 0$ von f ist zweifach. Die Newtoniteration mit Startwert $x^0 \neq 0$ ergibt*

$$x^{n+1} = \frac{1}{2}x^n \,,\; n \in \mathbb{N}_0 \,.$$

Also

$$|x^{n+1} - z| = \frac{1}{2}|x^n - z|$$

und die Konvergenzrate ist nur linear. □

Beispiel 1.15 *Die Konvergenz der Newtoniteration ist nur gewährleistet für hinreichend gute Startwerte. Dies zeigt das Beispiel $f(x) := \arctan(x)$. Wegen $f'(x) = 1/(1+x^2)$ hat ein Newtonschritt folgende Form: $x_{new} := x_{old} - \arctan(x_{old})/(1+x_{old}^2)$. Man kann für $|x_{old}| > 1.3917$ zeigen, dass $|x_{new}| > |x_{old}|$ ist. Dies hat die Divergenz der Newtoniteration zur Folge für Startwerte außerhalb von $[-1.3917, 1.3917]$.* □

Bemerkung 1.16 *Newton's Methode kann als eine Fixpunktiteration betrachtet werden. Setze $g(x) := x + h(x)f(x), x \in [a,b]$, mit einer glatten Funktion h. Eine Nullstelle von f ist sicher ein Fixpunkt von g.*

Wir wählen $h(x) := -1/f'(x)$. Wegen $g'(z) = 0$ für jede einfache Nullstelle z von f ist die Kontraktionskonstante von g in einer Nullstelle z von f Null. Dies hat die quadratische Konvergenz der Fixpunktiteration zur Konsequenz. □

Bemerkung 1.17 *Hat f die Nullstelle z mit Vielfachheit p, dann können wir die Iteration*

$$x^{k+1} := x^k - p\frac{f(x^k)}{f'(x^k)}$$

betrachten und man kann beweisen, dass wieder quadratische Konvergenz gegen z gegeben ist. Aber die Iteration ist von wenig praktischem Wert, denn nur selten kennt man die Vielfachheit einer Nullstelle im Vorhinein. □

Bemerkung 1.18 *Man kann die Auswertung der Ableitung in jedem Schritt vermeiden, indem man in jedem Schritt die Ableitung festhält:*

$$x^{n+1} := x^n - \frac{f(x^n)}{f'(x^0)} \,,\; x^0 \text{ gegeben} \,. \tag{1.29}$$

Diese Iteration nennt man das **modifizierte Newtonverfahren.** *Die Konvergenz ist aber nicht mehr quadratisch.*

Ein Kompromiss zwischen Newton– und modifiziertem Newtonverfahren ist das **Sekantenverfahren**. *Hierbei wird die Steigung $f'(x^k)$ der Tangente durch die Steigung $(f(x^k) - f(x^{k-1})(x^k - x^{k-1})^{-1}$ der Sekante ersetzt; das entstehende Verfahren nennt man das* **Sekantenverfahren**. *Damit ist für das modifiziertes Newtonverfahren und das Sekantenverfahren der Aufwand gleich. Dies korrespondiert mit der Tatsache, dass modifiziertes Newtonverfahren und das Sekantenverfahren die gleiche Konvergenzordnung besitzen, nämlich $\frac{1}{2}(\sqrt{5}+1) \approx 1.618$.* □

Bemerkung 1.19 *Die Iterationsfolge nach Newton $(x^n)_{n\in\mathbb{N}}$ hat nicht notwendigerweise die Eigenschaft, dass $|f(x^{n+1})| \leq |f(x^n)|$, $n \in \mathbb{N}$, gilt. Diese Tatsache motiviert die Einführung eines Dämpfungsfaktors:*

$$x^{n+1} := x^n - \lambda_n \frac{f(x^n)}{f'(x^n)} \,,\; x^0 \text{ gegeben} \,. \tag{1.30}$$

Dabei ist λ_n so gewählt, dass $|f(x^{n+1})| \leq |f(x^n)|$ ausfällt. Die Konvergenz ist damit garantiert für jeden Startwert x^0, aber die Konvergenzordnung ist nicht mehr quadratisch (zumindest nicht solange $\lambda \neq 1$). □

Fassen wir zusammen. Das Newton-Verfahren gilt als ein sehr effizientes Verfahren (in den Naturwissenschaften und anderswo). Worin ist dies begründet, obwohl das Problem der guten Startnäherung und die Tatsache, dass eine Ableitung ausgerechnet werden muss, schwer wiegen? Es liegt an vier Beobachtungen, die in der Literatur ausreichend diskutiert wurden und immer noch werden:

(1) Das Verfahren hat eine naheliegende Erweiterung auf Aufgaben in mehreren Variablen; siehe nächster Abschnitt.

(2) Das Verfahren konvergiert unter gut durchschaubaren Voraussetzungen (siehe Satz 1.13) quadratisch.

(3) Das Verfahren kann modifiziert werden, um die Berechnung der Ableitung in jedem Schritt zu vermeiden. Allerdings ist dann die Konvergenzgeschwindigkeit schlechter; siehe Bemerkung 1.18.

(4) Das Verfahren kann globalisiert werden, d.h. man kann Vorkehrungen einbauen, die sicherstellen, dass das so abgeänderte Verfahren auch bei „schlechten“ Startwerten konvergiert; das Stichwort ist Schrittweitensteuerung; siehe Bemerkung 1.19.

1.6.3 Der mehrdimensionale Fall

Betrachte nun eine Abbildung $F : \mathbb{R}^d \longrightarrow \mathbb{R}^d, F(x) = (F_1(x), \dots, F_d(x))$, die (total) differenzierbar ist in einer offenen Menge $U \subset \mathbb{R}^d$; mit $DF(x) \in \mathbb{R}^{d,d}$ bezeichnen wir die Jacobimatrix in $x \in U$, d.h.

$$DF(x) = \left(\frac{\partial F_i}{\partial x_j}(x)\right)_{1\le i,j\le d} .$$

Die Newtoniteration lautet:

$$x^{n+1} := x^n - DF(x^n)^{-1}F(x^n)\,,\ x^0 \in U\,.\ \text{gegeben}\,. \tag{1.31}$$

In der Praxis wird die Iterierte x^{n+1} über die Lösung eines linearen Gleichungssystems berechnet:

$$DF(x^n)u^n = -F(x^n)\,,\ x^{n+1} := x^n + u^n\,. \tag{1.32}$$

Satz 1.20 *Sei $U \subset \mathbb{R}^d$ offen, sei $F : U \ni x \longmapsto F(x) \in \mathbb{R}^d$ stetig differenzierbar, sei $z \in U$ eine Nullstelle von F, und sei $DF(z)$ invertierbar. Es gelte mit Konstanten[13] $r, \beta, L > 0$*

$$B_r(z) \subset U\,,\ \|DF(z)^{-1}\| \le \beta\,,\ \|DF(x) - DF(y)\| \le L|x-y| \text{ für alle } x, y \in U\,.$$

Dann definiert für alle $x^0 \in B_\delta(z)$ mit $\delta := \min\{r, \frac{1}{2\beta L}\}$ die Iteration (1.31) eine Folge $(x^n)_{n\in\mathbb{N}}$ mit

$$|x^{n+1} - z| \le \beta L|x^n - z|^2 \le \frac{1}{2}|x^n - z|\,,\ n = 0, 1, \dots. \tag{1.33}$$

Beweis:
Wir zeigen: Ist $x \in B_\delta(z)$ dann ist $DF(x)$ invertierbar und $\|DF(x)^{-1}\| \le 2\beta$.
Sei $x \in B_\delta(z)$. Dann haben wir

$$\eta := \|DF(z)^{-1}(DF(x) - DF(z))\| \le \|DF(z)^{-1}\|\|DF(x) - DF(z)\| \le \beta L|x-z| \le \beta L\delta \le \frac{1}{2}$$

[13] $|\cdot|$ ist die euklidische Norm in $\mathbb{R}^d$, $\|\cdot\|$ ist eine Matrixnorm in $\mathbb{R}^{d,d}$

und wir sehen, dass $DF(x)$ invertierbar ist; ferner gilt

$$\|DF(x)^{-1}\| \leq (1-\eta)^{-1}\|DF(z)^{-1}\| \leq 2\beta\,.$$

Nun können wir induktiv zeigen, dass $x^n \in B_\delta(z)\,,\ n \in \mathbb{N}$, gilt.
Für $n = 0$ folgt dies aus den Voraussetzungen des Satzes. Sei $x^n \in B_\delta(z)\,$. Wir haben

$$x^{n+1} = x^n - DF(x^n)^{-1}F(x^n) = x^n - DF(x^n)^{-1}(F(x^n) - F(z))$$

und daher

$$x^{n+1} - z = DF(x^n)^{-1}(F(z) - F(x^n) - DF(x^n)(z - x^n))\,.$$

Dies ergibt

$$|x^{n+1} - z| \leq 2\beta\frac{L}{2}|x^n - z|^2 \leq \beta L\delta|x^n - z| \leq \frac{1}{2}|x^n - z|$$

und die Induktion ist abgeschlossen. ■

Bemerkung 1.21 *Beachte, dass die Abschätzung* (1.33) *die quadratische Konvergenz der Folge* $(x^n)_{n\in\mathbb{N}}$ *impliziert.* □

Es gibt eine Reihe von unterschiedlichen Ausführungen der Newtoniteration, um Rechenaufwand zu vermeiden. Eine wesentliche Beobachtung ist, das es oft ausreicht, das lineare Gleichungssystem nicht exakt zu lösen, sondern nur ein paar Iterationsschritte eines iterativen Lösers zu benutzen.

Bemerkung 1.22 *Newton's Methode kann genutzt werden, Extrema einer zu minimierenden Funktion zu bestimmen. Sei* $f : U \longrightarrow \mathbb{R}^d$ *eine zweimal differenzierbare Funktion. Die Extrema von* f *sind Nullstellen des Gradienten* $F := \nabla f$. *Bei der Anwendung sollte in jedem Falle das modifizierte Newtonverfahren Verwendung finden, um sicherzustellen, dass die Folge* $(f(x^n))_{n\in\mathbb{N}}$ *abnimmt.* □

1.7 Bibliographische Anmerkungen

Auf die Grundlagen der Preisformel kommen wir noch zurück. Sie ist etwa zu finden in [9, 10, 20, 24]. Erste systematische Untersuchungen von numerischen Verfahren für stochastische Differentialgleichungen findet man in [18, 19]. Spezielle Verfahren der Berechnung der impliziten Volatilität findet man in [4, 17, 22, 23]. Zur impliziten Volatilität im so genannten Hull-White-Modell siehe [13]. Etwas Grundsätzliches zur Volatilität erfährt man in [8, 11]. Zu den Newtonverfahren gibt es eine Reihe von sehr grundsätzlichen Betrachtungen, siehe etwa [6, 16, 25]. Insbesondere ist das Newtonverfahren auch im unendlichdimensionalen Banachraum wohlstudiert (siehe etwa [2]). Die Varianten, jeden Newtonschritt inexakt zu lösen, ohne die Konvergenz zu verlieren, werden diskutiert in [7].

1.8 Übungen

1.) Betrachte das Problem, eine positive Zahl x zu finden mit $\cos(x) = x^3$.
   - (a) Zeige, dass die Nullstellengleichung $\cos(x) - x^3 = 0$ genau eine Nullstelle besitzt und dass diese in $(0, 1)$ liegt.
   - (b) Formuliere die Newton-Iteration zur Nullstellenaufgabe $\cos(x) - x^3 = 0$.
   - (c) Führe die Newton-Iteration mit dem Startwert $x^0 = 0.5$ (mit Maple) durch.

2.) Betrachte das Sekantenverfahren:

$$x^{k+1} := x^k - \frac{f(x^k)(x^k - x^{k-1})}{(f(x^k) - f(x^{k-1}))}\,,\ x^0, x^1 \text{ gegeben}\,. \tag{1.34}$$

Zeige:
Sei $f : [a,b] \longrightarrow \mathbb{R}$ zweimal stetig differenzierbar und es gelte

$$|f'(x)| \geq m\,,\ |f''(x)| \leq M \text{ für alle } x \in [a,b] \tag{1.35}$$

mit $m > 0, M > 0$. Dann gilt:

(a) f hat in $[a,b]$ höchstens eine Nullstelle.

(b) Ist z eine Nullstelle von f (a,b), dann ist die Iteration (1.34) definiert für alle $x^0, x^1 \in U_r(z) := (z-r, z+r)$ wobei $r := \min(2mM^{-1}, b-z, z-a)$.
Weiterhin haben wir mit $q := \max(M(2m)^{-1}|x^0 - z|, M(2m)^{-1}|x^0 - z|) < 1$ f'ür alle $n \in \mathbb{N}$:
 1. $|z - x^n| \leq \frac{2m}{M} q^{k_n}$ (a priori Abschätzung);
 2. $|z - x^n| \leq \frac{1}{m}|f(x^n)|$ (a posteriori Abschätzung).

Hierbei ist $(k_n)_{n\in\mathbb{N}}$ die Folge der Fibonaccizahlen: $k_0 := k_1 := 1, k_{n+2} := k_{n+1} + k_n, n \in \mathbb{N}_0$.

3.) Betrachte die Dichte f der Normalverteilung: $f(x) := \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}x^2), x \in \mathbb{R}$. Diskutiere die zugehörige Verteilungsfunktion

$$F(x) := \int_{-\infty}^{x} f(t)\,dt\,,\ x \in \mathbb{R}\,,$$

hinsichtlich Infima, Minima, Wendepunkte, Monotoniebereiche, Umkehrbarkeit.

4.) Betrachte die Pareto-Verteilung $P(k,\kappa), k \in \mathbb{N}, \kappa > 0$, mit Verteilungsfunktion

$$F(x) := \begin{cases} \kappa^k x^{-k} & \text{falls } x > \kappa \\ 0 & \text{sonst} \end{cases}\,,\ x \in \mathbb{R}\,.$$

Bestimme die Dichte, falls vorhanden, Erwartungswert und die Varianz.

5.) Betrachte die Cauchy-Verteilung $C(s,t), s,t \in (0,\infty)$, mit der Dichte

$$f(x) := \frac{1}{\pi}\frac{s}{s^2 + (x-t)^2}\,,\ x \in \mathbb{R}\,.$$

Bestimme die Verteilungsfunktion und Erwartungswert, Varianz, falls vorhanden, im Fall $t = 0, s = 1$.

6.) Die charakteristische Funktion ϕ_f einer Verteilung mit Dichte f ist die Fouriertransformierte $\phi_f := \mathcal{F}(f)$:

$$\phi_f(x) = \int_{-\infty}^{\infty} f(t)e^{itx}\,,\ x \in \mathbb{R}\,.$$

Berechne $\phi'(0), \phi''(0)$ und interpretiere (unter geeigneten Annahmen) die erhaltenen Ergebnisse.

7.) Die Verzinsung mit variablem Zinssatz $r = r(t)$ kann mit folgender Differentialgleichung beschrieben werden:

$$z' = r(t)z\,,\ z(0) = B(0)\,. \tag{1.36}$$

Der Wert $B(t)$ der so verzinsten Anlage zum Zeitpunkt t ist dann $B(t) := z(t)$, wobei $z(t)$ der Wert der Lösung von (1.36) ist.

(a) Finde ein Argument dafür, dass (1.36) eine eindeutig bestimmte Lösung besitzt.

(b) Löse die Anfangswertaufgabe (1.36) und finde damit eine explizite Darstellung für den Wert $B(t)$.

8.) Betrachte eine Call-Option, deren Preis nach dem Black-Scholes-Modell berechnet werde. Der Wert des Basisobjekts sei 7133.06, der Zinssatz $r = 0.0487$, die Restlaufzeit $\tau := T - t = 120/365 = 0.328$. Die Preise C_i zu unterschiedlichen Ausübungspreisen K_i sind der folgenden Tabelle zu entnehmen.

K_i	6400	6700	7000	7300	7600	7900	8200	8500	8800
C_i	934.0	690.0	469.0	283.0	145.0	62.0	22.0	7.5	2.1

Berechne die impliziten Volatilitäten $\sigma_{impl}(K_i)$ mit dem Sekantenverfahren. Skizziere dann den Funktionsgraph $K_i \longmapsto \sigma_{impl}(K_i)$.

9.) Die Berechnung der impliziten Volatilitäten kann analytisch auf eine Darstellung der Verteilungsfunktion der Normalverteilung in Form einer Taylorreihe gestützt werden; siehe [5].

(a) Bestimme die Taylorreihe der Normalverteilung:

$$\mathcal{N}(x) = a_0 + a_1 x + \cdots .$$

(b) Berechne die Approximation der Volatilität, die entsteht, wenn man eine lineare Approximation der Normalverteilung nach (a) nutzt.

10.) Eine Call-Option auf Aktien mit einem Ausübungspreis 100 €, einer Restlaufzeit von einem Jahr und einer Volatilität der Aktie von 20 % besitzt bei einem momentanen Aktienkurs von 80 € und einem Zinssatz von 3 % folgende Greeks:

Greeks	Δ	$\mathcal{V}$	Θ	$\mathcal{P}$
Werte	0.91	15.74	−1.60	68.18

Was bedeuten diese Zahlen für die Option?

11.) Betrachte eine Call-Option, deren Preis nach dem Black-Scholes-Modell berechnet werde. Der Wert des Basisobjekts sei 47.16, der Zinssatz $r = 0.03$. Die Preise C_i zu unterschiedlichen Ausübungspreisen K_i und Restlaufzeiten T_i sind der folgenden Tabelle zu entnehmen.

K_i	35.00	40.00	42.50	45.00	45.00	50.00	50.00	50.00	55.00	55.00
T_i	149/365	149/365	23/365	3/365	23/365	3/365	23/365	149/365	23/365	86/365
C_i	14.90	10.80	5.30	2.40	3.40	0.10	0.90	4.70	0.20	1.30

Berechne die zugehörigen impliziten Volatilitäten $\sigma_{impl}(K_i, T_i)$. Skizziere dann den Funktionsgraph $K_i \longmapsto \sigma_{impl}(K_i, T_i)$ zu den Restlaufzeit $T_i = 23/365$.

12.) Die Preisformel nach dem Black-Scholes Modell mit Dividende (δ) lautet:

$$C(S, K, t, T, r, \sigma, \delta) := Se^{-\delta(T-t)}\mathcal{N}(d_+) - Ke^{-r\tau}\mathcal{N}(d_-), \text{ wobei}$$

$$d_+ := \frac{\ln(Se^{(r-\delta)(T-t)}/K)}{\sigma\sqrt{T-t}} + \frac{1}{2}\sigma\sqrt{T-t}, \; d_- = d_+ - \sigma\sqrt{T-t}, \text{ ist}.$$

(a) Finde Variablen x, ν mit

$$C(S, K, t, T, r, \sigma, \delta) = Se^{-\delta(T-t)}c(x, \nu) \text{ mit } c(x, \nu) = \mathcal{N}(x/\nu + \nu/2) - e^{-x}\mathcal{N}(x/\nu - \nu/2).$$

(b) Plotte c im Bereich $-0.5 \leq x \leq 0.5, 0 < \nu \leq 1$.

(c) Wie ist das Verhalten von c in $x = 0$?

13.) Ein europäisches Unternehmen will im Jahr 2010 eine Maschine zum Preis von 1 000 000 US-$ kaufen; der Kaufpreis ist innerhalb des Jahres 2010 zu entrichten. Das Unternehmen kauft, um sich gegen die Wechselkursschwankungen zwischen den beiden Währungen abzusichern, 10 000 Call-Optionen mit folgenden Eigenschaften:

$$\text{Preis } C = 1 \text{ €}, \text{ Verfallszeit } T = 1 \text{ (Jahr)}, \text{ Ausübungspreis } K = 1 \text{ €}$$

(a) Welches Recht erwirbt der Käufer?

(b) Der Wechselkurs beträgt am Ende des Jahres 1.1 € für einen US-$. Welchen Gewinn erzielt das Unternehmen?

(c) Der Wechselkurs beträgt am Ende des Jahres 0.9 € für einen US-$. Welchen Verlust erleidet das Unternehmen?

14.) Bestücke zur Zeit $t = 0$ ein Portfolio folgendermaßen:
Kaufe eine Aktie (Wert S_0), kaufe zwei Puts (Wert P) auf die Aktie zur Ausübungszeit $T = 1$ und zum Ausübungspreis K. Der Zinssatz für festverzinsliche Anleihen sei r.

(a) Bestimme den Geldfluss in $t = 0$.

(b) Bestimme die Nettoauszahlung bei Auflösung des Portfolios in $t = T$, insbesondere für den Fall $K = S_0$.

Literaturverzeichnis

[1] A. Abramowitz und I. Stegun. *Handbook of Mathematical Functions.* Dover Publications, New York, 1972.

[2] K. Atkinson and W. Han. *Theoretical Numerical Analysis. A Functional Analysis Framework*. Texts in applied mathematics, vol. 39. Springer, New York, 2001.

[3] T.F. Coleman, Y. Li and A. Verma. Reconstructing the unknown local volatility function. *J. of Comput. Finance*, 2:77–102, 1986.

[4] R. Cont and F. de Fonseca. Dynamics of implied volatility. *Quantitative Finance*, 2:45–60, 2002.

[5] C.J. Corrado and T.W. Miller. A note on a simple, accurate formula to compute implied standard deviations. *J. of Banking and Finance*, 20:595–603, 1996.

[6] P. Deuflhard and A. Hohmann. *Numerical Methods I.* de Gruyter, Berlin, 1995.

[7] S.C. Eisenstat and H.F. Walker. Choosing the forcing terms in an inexact newton method. *SIAM J. Sci. Computing*, 17:16–32, 1994.

[8] R.F. Engle and A.J. Patton. What good is a volatility model? *Quantitaive Finance*, 1:237–245, 2001.

[9] P. Glasserman. *Gradient Estimation via Perturbation Analysis.* Kluwer, 1991.

[10] P. Glasserman. *Monte Carlo Methods in Financial Engineering.* Springer, Baltimore, 2003.

[11] C.W.J. Granger and S.H. Poon. Forcasting volatility in financial markets: A review. *Journal of Economic Literature*, 41:478–539, 2003.

[12] M. Günther and A. Jüngel. *Finanzderivate mit MATLAB.* Vieweg, 2003.

[13] A. Gulisashivili and E.M. Stein. Implied volatility in the Hull-White model. *Math. Finace*, 19:303–327, 2009.

[14] C. Hastings. *Approximations for digital computers.* Princeton University Press, Princeton, 1955.

[15] S.L. Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies*, 6:327–343, 1993.

[16] G. Hämmerlin and K.-H. Hoffmann. *Numerische Mathematik.* Springer, 1989.

[17] Y. Kagenishi and Y. Shinohara. A note on computation of implied volatility. *Asia-Pacific Financial Markets*, 8:361–368, 2004.

[18] P. Kloeden and E. Platen. *Numerical Solution of SDE through Computer Experiments.* Springer, Berlin, 1994.

[19] P. Kloeden and E. Platen. *Numerical Solutions of Stochastic Differential Equations.* Springer, Berlin, 1999.

[20] R. Korn und E. Korn. *Optionsbewertung und Portfolio–Optimierung.* Vieweg, Braunschweig, 1999.

[21] R. Lagnado and S. Osher. A technique for calibrating derivative security pricing models: numerical solution of an inverse problem. *J. of Comput. Finance*, 1:13–25, 1997.

[22] M. Li. You don't have to bother newton for implied volatility . *http://ssrn.com/abstract=952727*, pages 1–26, 2006.

[23] D. Ruppert. *Statistics and Finance.* Springer, Berlin, 2004.

[24] R. Seydel. *Tools for Computational Finance.* Springer, Berlin, 2002.

[25] P. Spellucci. *Numerische Verfahren der nichtlinearen Optimierung.* Birhäuser, Basel, 1993.

Kapitel 2

Das Binomialmodell

Wir beschäftigen uns hier für europäische Optionen mit einer zeitdiskreten Methode zur Bestimmung der Optionspreise; für amerikanische Optionen verweisen wir im Wesentlichen auf die Literatur. Es ist eine Methode, die einfach und intuitiv implementierbar ist und auf viele exotische Optionen erweiterbar ist. Zunächst aber kommen wir überein, in welchen Märkten wir die Methode untersuchen wollen.

2.1 Marktannahmen

Für die folgenden Überlegungen wollen wir Annahmen über den zugrundeliegenden Finanzmarkt anführen:

Regel 2.1 („Geschäftsbedingungen“)

- *Alle Investoren haben den selben Informationsstand und können verzögerungsfrei handeln,*
- *Investoren handeln rational und ziehen ein größeres Vermögen einem kleineren vor,*
- *es werden keine Transaktionskosten und Steuern berücksichtigt,*
- *der Wertpapier– bzw. Optionshandel ist zu jedem Zeitpunkt möglich,*
- *Leerverkäufe (Verkauf eines Basiswertes, den man noch nicht besitzt, aber später liefert) sind möglich,*
- *gewünschte Transaktionen können in beliebigem Umfang ohne Rückwirkungen auf die Kursentwicklung durchgeführt werden,*
- *Wertpapiere sind beliebig teilbar,*
- *Wertpapiere stehen in beliebiger Menge zur Verfügung (Liquidität),*
- *die Verzinsung für festverzinsliche Anleihen und Geldanlagen erfolgt nach demselbem Zinssatz. Zusätzlich vereinbaren wir eine kontinuierliche Verzinsung (eine Vereinbarung, die nicht zwingend wäre, beachte aber, dass daraus quantiatitve Konsequenzen sich ergeben).*

Regel 2.2 (Arbitragefreiheit) *Der Markt lässt keine Arbitragemöglichkeiten zu.*

Regel 2.3 (Modellierbarkeit) *Der Markt, bestehend aus Wertpapieren, Anleihen, Optionen, ... kann durch ein Modell abgebildet werden.*

Die Annahme über die Geschäftsbedingungen werden wir meist ergänzen durch die Annahme, dass keine Dividenden gezahlt werden.

Unter einem **Leerverkauf (short selling)** versteht man eine Handelsstrategie, bei der ein Teilnehmer am Finanzgeschehen (Investor) Objekte, die er nicht besitzt, verkauft und sie später zurückkauft.

Eine **Arbitragemöglichkeit** ist eine Handelsstrategie, die keine Anfangsinvestitionen benötigt und mit positiver Wahrscheinlichkeit einen Gewinn ergibt, ohne das Risiko eines Verlustes zu beinhalten. Etwas formaler:

> Sei $I(t)$ die Entwicklung des Vermögens eines Investors über den Zeitraum $[0, T]$. Man sagt, dass eine **Arbitragemöglichkeit** für den Investor besteht, falls es möglich ist, dass er mit dem Vermögen $I(0)$ startet und für sein Endvermögen $I(T)$
>
> $$I(T) \geq I(0)\,,\ \text{Wahrscheinlichkeit}(\{I(T) > I(0)\}) > 0$$
>
> gilt. Hierbei wird $I(t), t \in [0, T]$, als Zufallsgröße aufgefasst.

Arbitrage und vollkommene kapitalmärkte schließen sich wegen des Fehlens von Marktpreisdifferenzen aus. Aus der bloßen Annahme der Arbitrage-Freiheit können daher allerlei Konsequenzen über die Höhe der Optionspreise gezogen werden. Wir werden konkrete Anwendungen dieses Prinzips kennenlernen.

Bemerkung 2.4 *Betrachte den Fall, dass eine bestimmte Aktie in Frankfurt mit* 60 €*, in New York zum selben Zeitpunkt mit* $102\ US - \$$ *notiert. Der Wechselkurs sei* $1\ US - \$ = 0.6$ €*. Mit welcher Handelsstrategie lässt sich ein Gewinn erzielen? Natürlich: Kaufe eine Aktie in Frankfurt, verkaufe sie sofort wieder in New York. Gewinn* 1.2 € *(*$102 \cdot 0.6 - 60 = 1.2$*)*

Bei den Modellen unterscheidet man zwischen deterministischen und stochastischen Modellen. Hier „zähmen“ wir den Zufall, indem wir ein sehr einfaches Modell verwenden.

Eine erste Anwendung zu den getroffenen Regeln ist die Ableitung einer Identität, die Call- und Put-Option verbindet.

Wir betrachten folgende Aktion auf dem Finanzmarkt: Wir kaufen einen Basiswert S, einen europäischen Put P mit dem Ausübungspreis K und dem Verfallstag T und verkaufen (Leerverkauf) einen europäischen Call C mit dem Ausübungspreis K und dem Verfallstag T. Damit haben wir ein Portfolio zusammengestellt, das den Wert $S + P - C$ besitzt. Wie groß ist der Wert zu einem Zeitpunkt $t \in [0, T]$? Wir behaupten $Ke^{-r(T-t)}$, und wir werden dies beweisen. Dies korrespondiert mit folgender Beobachtung: Legen wir zum Zeitpunkt t den Betrag $Ke^{-r(T-t)}$ an, dann erhalten wir zur Zeit T den Betrag K zurück.

Satz 2.5 (Put-Call-Parität) *Betrachte ein Portfolio mit einem gekauften Basiswert* S*, einem gekauften europäischen Put* P*, einem verkauften europäischen Call* C*, beide mit Ausübungspreis* K *und dem Verfallstag* T*.* P_t *und* C_t *seien Abkürzungen für den Wert der Put- bzw. Call-Option zur Zeit* t*,* Π_t *sei der Wert des Portfolios zur Zeit* $t \in [0, T]$*. Dann gilt*

$$\Pi_t = S_t + P_t - C_t = Ke^{-r(T-t)}\,,\ t \in [0, T]\,.$$

Beweis:
Offenbar gilt $\Pi_T = S_T + (K - S_T)^+ - (S_T - K)^+ = K$. Sei nun $t \in [0, T)$.
Annahme: $\Pi_t < Ke^{-r(T-t)}$.
Handelsstrategie: Leihe den Betrag $Ke^{-r(T-t)}$, kaufe ein wie oben zusammengestelltes Portfolio, lege den Betrag $Ke^{-r(T-t)} - \Pi_t$ zur Seite.

Zur Zeit T liefert das Portfolio den Betrag K, den wir der Bank zur Ablösung des Kredits geben. Dies bedeutet, dass wir zur Zeit t einen sofortigen Gewinn $Ke^{-r(T-t)} - \Pi_t > 0$ erzielt haben, ohne zum Zeitpunkt T ein Risiko einzugehen. Das ist ein Widerspruch zur Arbitragefreiheit!

Annahme: $\Pi_t > Ke^{-r(T-t)}$.
Verkaufe ein wie oben zusammengestelltes Portfolio, d.h. verkaufe den Basiswert und den Put, kaufe einen Call, lege den Betrag $Ke^{-r(T-t)}$ an und lege die Differenz $\Pi_t - Ke^{-r(T-t)} > 0$ beiseite. Von der Bank erhalten wir zur Zeit T den Betrag K zurück, damit kaufen wir das Portfolio zum Preis $\Pi_T = K$. Wir haben zur Zeit t einen sofortigen, risikofreien Gewinn $\Pi_t - Ke^{-r(T-t)} > 0$ erzielt. Das ist ein Widerspruch zur Arbitragefreiheit! ■

Satz 2.6 (Schranken für Optionspreise/europäisch) *Betrachte eine europäische Put- bzw. Call-Option zum Basiswert S, Verfallszeit T, Ausübungspreis K; der Zinssatz für festverzinsliche Anleihen sei r. P_t und C_t seien Abkürzungen für den Wert einer Put- bzw. Call-Option zur Zeit t. Dann gelten die folgenden Schranken:*

(a) $(S_t - Ke^{-r(T-t)})^+ \le C_t \le S_t$, $t \in [0,T]$.

(b) $(Ke^{-r(T-t)} - S_t)^+ \le P_t \le Ke^{-r(T-t)}$, $t \in [0,T]$.

Beweis:
Wir gehen nur auf (a) ein, (b) überlassen wir dem Leser. Sei $t \in [0,T]$.
Die Schranke $C_t \ge 0$ ist offensichtlich, denn anderenfalls ergäbe ein Kauf einer solchen Option – der Preis ist negativ! – einen sofortigen Gewinn, während zur Zeit T jedenfalls keine negative Auszahlung droht.
Es gilt $C_t \le S_t$, denn anderenfalls verkaufe einen Call, kaufe einen Basiswert, verkaufe eventuell den Basiswert zur Zeit T, und realisiere zur Zeit t den sofortigen risikofreien Gewinn $C_t - S_t > 0$, denn zur Zeit T ist der Erlös sicher nichtnegativ. Also haben wir einen Widerspruch zur Arbitragefreiheit. Den Beweis der Schranke $(S_t - Ke^{-r(T-t)})^+ \le C_t$ entnehme man der Tabelle 2.1. Beachte, dass die dort angegebene Handelsstrategie zur Zeit t und T einen Geldfluss erzeugt, der das Negative des Portfoliowertes beträgt. ■

Annahme $C_t < S_t - Ke^{-r(T-t)}$			
Handelsstrategie	**Geldfluss zur Zeit** t	**Geldfluss zur Zeit** T	
		$S_T \le K$	$S_T > K$
Verkaufe Basiswert S_t	S_t	$-S_T$	$-S_T$
Kaufe Call C_t	$-C_t$	0	$S_T - K$
Anlage von $Ke^{-r(T-t)}$	$Ke^{-r(T-t)}$	K	K
Summe	$-C_t + S_t + Ke^{-r(T-t)} > 0$	$K - S_T \ge 0$	0
Risikofreier Gewinn zur Zeit t: $-C_t + S_t + Ke^{-r(T-t)} > 0$			

Abbildung 2.1: Zum Beweis von Satz 2.6

Wie ist diese untere Schranke in (a) von Satz 2.6 zu verstehen? Bei einer amerikanischen Call-Option lautet die untere Schranke $(S_t - K)^+$ (innerer Wert), die der Auszahlung entspricht, die der Besitzer bei sofortiger Ausübung realisieren kann. Für eine europäische Call-Option ist dies nicht so einfach, denn eine sofortige Ausübung ist ja nicht möglich, sie kann nur über ein Termingeschäft realisiert werden. Hierbei ist zu beachten, dass der Ausübungspreis abgezinst werden muss: K ist also in der unteren Schranke durch $Ke^{-r(T-t)}$ zu ersetzen.

Die Sachlage ist bei amerikanischen Optionen etwas verschieden. Von vorneherein ist allerdings klar, dass der Preis einer amerikanischen Call-Option nie unter dem Preis einer europäischen Call-Option liegen kann (unter Zugrundelegung unserer Verabredung).

Satz 2.7 (Schranken für Optionspreise/amerikanisch) *Betrachte eine amerikanische Put- bzw. Call-Option zum Basiswert S, Verfallszeit T, Ausübungspreis K; der Zinssatz für festverzinsliche Anleihen sei r . P_t^a und C_t^a seien Abkürzungen für den Wert einer amerikanischen Put- bzw. Call-Option zur Zeit t , C_t^e ist der Wert einer europäischen Call-Option zur Zeit t , Dann gelten die folgenden Schranken:*

(a) $C_t^a = C_t^e\,,\ t \in [0,T]\,.$

(b) $Ke^{-r(T-t)} \le S_t + P_t^a - C_t^a \le K\,,\ t \in [0,T]\,.$

(c) $(Ke^{-r(T-t)} - S_t)^+ \le P_t^a \le K\,,\ t \in [0,T]\,.$

Beweis:
Wir gehen nur auf die doch möglicherweise überraschende Gleichheit in (a) ein, für $(b), (c)$ verweisen wir auf die Literatur.

Für $t = T$ ist die Aussage klar.
Annahme: Die amerikanische Call-Option wird zu einem Zeitpunkt $t < T$ frühzeitig ausgeübt. Dann erhalten wir die positive Auszahlung $S_t - K$. Wegen

$$C_t^a \ge C_t^e \ge (S_t - Ke^{-r(t-T)})^+ = S_t - Ke^{-r(t-T)} > S_t - K$$

ist es sinnvoller, die Option zu verkaufen, als auszuüben. Die frühzeitige Ausübung ist daher nicht angeraten. ■

Bemerkung 2.8 *Es ist also nicht sinnvoll, einen amerikanischen Call vor seinem Verfalldatum auszuüben,* **wenn auf die zugrundeliegende Aktie während der Optionslaufzeit keine Dividende gezahlt wird.** *Bei amerikanischen Puts ist die Situation anders, auch wenn keine Dividenden gezahlt werden.*

Bezieht man Dividendenzahlungen ein, so verändert sich die Sachlage bei amerikanischen Optionen grundsätzlich. Insbesondere die Zeit vor der Dividendenausschüttung ist dann „kritisch“. □

Wegen Aussage (a) in Satz 2.7 ist nun auch eine Preisformel für amerikanische Call-Optionen bekannt; siehe Abschnitt 1.2.

Bemerkung 2.9 *Mit Hilfe der Put-Call-Parität sieht man ein, dass die Black-Scholes-Gleichung unverändert auch für die Optionspreise für europäische Put-Optionen Verwendung finden kann. Ist C der Preis für die Call-Option und P der Preis für die Put-Option, so gilt nach Satz 2.6*

$$C(S,t) + Ke^{-r(T-t)} = P(S,t) + S\,,\ S \in [0,\infty), t \in [0,T]\,.$$

□

2.2 Das Binomialmodell

Das Binomialmodell wurde 1979 von Cox, Ross und Rubinstein entwickelt. Es modelliert das Problem der Optionsbewertung durch eine diskrete Irrfahrt.

2.2.1 Das Ein-Perioden-Modell

Die Regeln für den Finanzmarkt sind die Geschäftsbedingungen und die Arbitragefreiheit; die Modellierbarkeit wird beschrieben.
Das Ein-Perioden-Modell lässt nur zwei Handelszeitpunkte zu: $t = 0$, $t = \Delta t$. Als Wertpapiere sind zugelassen: Aktie (Basiswert) mit Kurswert S_t, Bond mit Kurswert B_t, Call-Option mit Wert C_t. Als Aktionen auf dem Markt sind möglich:

- Ankauf/Verkauf von Anleihen vom Einheitswert $B_0 = 1$ zum Zinssatz r.
- Ankauf/Verkauf von Aktien zum Kurswert S_0.
- Ankauf/Verkauf eines Call mit Ausübungspreis K und Verfallszeit Δt.
- Kursverlauf der Aktie:

$$S_{\Delta t} = \begin{cases} uS_0 & \text{mit Wahrscheinlichkeit } q \\ dS_0 & \text{mit Wahrscheinlichkeit } 1-q \end{cases}$$

  mit $q \in [0,1]$. Dabei sind u (up-Faktor) und d (down-Faktor) Zahlen mit $u > d > 0$.
- Auszahlung zur Zeit Δt:

$$C_{\Delta t} = \begin{cases} C^u := (uS_0 - K)^+ & \text{mit Wahrscheinlichkeit } q \\ C^d := (dS_0 - K)^+ & \text{mit Wahrscheinlichkeit } 1-q \end{cases}$$

Da der Markt als arbitragefrei angenommen wird, muss die folgende Ungleichung gelten:

$$d \leq e^{r\Delta t} \leq u \tag{2.1}$$

denn:
1. Wenn $e^{r\Delta t} > u$ ist, dann kann man durch den Kauf von Anleihen, der durch Aktienverkäufe finanziert ist, einen Gewinn ohne Risiko erzielen.
2. Wenn $e^{r\Delta t} < d$ ist, dann kann man durch den kreditfinanzierten Kauf von Aktien einen Gewinn ohne Risiko erzielen.
Zu 1. etwas genauer:
Betrachte die folgende Handelsstrategie: eine Aktie verkaufen, damit Anleihen kaufen, keine Calls kaufen.
Geldfluss zur Zeit $t = 0$:

$$1 \cdot S_0 - S_0 \cdot B_0 - 0 \cdot C_0 = 0\,.$$

Geldfluss zur Zeit $t = \Delta t$:

$$-1 \cdot S_{\Delta t} + S_0 \cdot B_{\Delta t} + 0 \cdot C_{\Delta t} = -S_{\Delta t} + S_0 e^{r\Delta t} = \begin{cases} (e^{r\Delta t} - u)S_0 & \text{mit Wahrscheinlichkeit } q \\ (e^{r\Delta t} - d)S_0 & \text{mit Wahrscheinlichkeit } 1-q \end{cases}$$

Damit ist wegen $e^{r\Delta t} - d > e^{r\Delta t} - u > 0$ klar, dass es zur Zeit $t = \Delta t$ einen risikofreien Gewinn gibt. Widerspruch!
Beachte: Die obige Ungleichung hängt von unserer Verzinsungsvereinbarung ab.

Nun wenden wir uns der Aufgabe zu, den Preis C_0 der Option zu ermitteln. Wir tun dies wieder mit einer **Duplikationsstrategie**. Dass wir dies tun können, haben wir sichergestellt durch die „Geschäftsbedingungen“.[1] Die Option wird also durch den Kauf bzw. Verkauf von c_1 Anteilen des Bonds und c_2 Anteilen der Aktie nachgebildet. Gesucht ist also ein zur Call-Option äquivalentes Portfolio derart, dass gilt:

$$c_1 \cdot B_t + c_2 \cdot S_t = C_t \, , \; t \in \{0, \Delta t\} \, . \tag{2.2}$$

Die Unbekannten c_1, c_2 können aus (2.2) für $t = \Delta t$ bestimmte werden:

$$c_1 e^{r\Delta t} + c_2 u S_0 = C^u \, , \; c_1 e^{r\Delta t} + c_2 d S_0 = C^d \, . \tag{2.3}$$

Dies ergibt

$$c_1 = \frac{uC^d - dC^u}{(u-d)e^{r\Delta t}} \, , \; c_2 = \frac{C^u - C^d}{(u-d)S_0} \, . \tag{2.4}$$

Daraus bestimmt sich nun der Optionspreis C_0 folgendermaßen:

$$\begin{aligned} C_0 &= c_1 \cdot 1 + c_2 \cdot S_0 \\ &= \frac{uC^d - dC^u}{(u-d)e^{r\Delta t}} + \frac{C^u - C^d}{u-d} \\ &= e^{-r\Delta t} \left(\frac{uC^d - dC^u + e^{r\Delta t}(C^u - C^d)}{u-d} \right) \\ &= e^{-r\Delta t} \left(\frac{e^{r\Delta t} - d}{u-d} C^u + \frac{u - d - (e^{r\Delta t} - d)}{u-d} C^d \right) \end{aligned}$$

Also haben wir

$$\begin{aligned} C_0 &= e^{-r\Delta t} \left(pC^u + (1-p)C^d \right) \\ &= e^{-r\Delta t} \left(p(uS_0 - K)^+ + (1-p)(dS_0 - K)^+ \right) \text{ mit } p = \frac{e^{r\Delta t} - d}{u-d} \, . \end{aligned}$$

Beachte: Der Optionspreis hängt nicht von q ab.

Ist dies nicht verwunderlich? Angenommen, die obige Wahrscheinlichkeit q wäre Null, d.h. die Wahrscheinlichkeit eines Kursabfalls ist eins. Dann wird doch sicher kein Investor eine Option kaufen wollen. Aber ist eine solche Sicherheit denn verträglich mit der Arbitragefreiheit? Nein, denn wir können dann eine Aktie verkaufen, und nach der Zeit Δt zurückkaufen, wobei wir einen sicheren risikofreien Gewinn machen. Es ist also so, dass der Markt so sich darstellt, dass die eben beschriebene Arbitragemöglichkeit nicht besteht.

Aus (2.1) folgt $0 \le p \le 1$. Damit kann p als eine Wahrscheinlichleit interpretiert werden. Bezüglich dieser Wahrscheinlichkeit hat der Wert der Aktie zur Zeit $t = \Delta t$ den Erwartungswert

$$\mathbb{E}_p(S_{\Delta t}) = puS_0 + (1-p)dS_0 = e^{r\Delta t} S_0 \tag{2.5}$$

und der Optionspreis stellt sich als Erwartungswert

$$C_0 = e^{-r\Delta t} \mathbb{E}_p((S_{\Delta t} - K)^+) \tag{2.6}$$

dar, d.h. als abgezinster Erwartungswert der Auszahlungsfunktion bezüglich der Wahrscheinlichkeit p. Das Ergebnis

$$C_0 = e^{-r\Delta t}(pC^u + (1-p)C^d) \, , \tag{2.7}$$

[1] Ein Markt heißt **vollständig**, wenn sich für jede Option ein duplizierendes Portfolio konstruieren lässt.

das wir oben schon gesehen haben, kann man derart interpretieren, dass der heutige Wert der Option dem abgezinsten Erwartungswert der zukünftigen Optionswertes entspräche, wenn sich die Werte des Basisobjekts (und damit auch der Option) mit der Wahrscheinlichkeit p anstelle von q entwickeln würden. Die Wahrscheinlichkeit p heisst daher auch **risikoneutrale Wahrscheinlichkeit**. Sie ist eine synthetische Wahrscheinlichkeit, ohne Bezug zu einer realen Wahrscheinlichkeit. Die Berechnung des fairen Preises einer Option ist leicht, weil sich unter dieser Wahrscheinlichkeit der Marktpreis einfach als diskontierter Zahlungsstrom ergibt, sagt (2.5) doch gerade, dass der Erwartungswert des Aktienkurses gleich dem Erlös aus einem risikofreien Bond entspricht.

Die Verallgemeinerung zu einem n-Perioden-Modells wird **Cox-Ross-Rubinstein-Modell (1979)** genannt. Dazu kommen wir nun.

2.2.2 Das n-Perioden-Modell

Das n-Perioden-Modell lässt nun $n+1$ Handelszeitpunkte zu: $t = 0, t = \Delta t, \ldots, T := n\Delta t$. Als Wertpapiere sind wieder zugelassen: Aktie (Basiswert) mit Kurswert S_t, Bond mit Kurswert B_t, Call-Option mit Wert C_t. Als Aktionen auf dem Markt sind wieder die in Abschnitt 2.2.1 zugelassenen möglich:

- Ankauf/Verkauf von Anleihen vom Wert $B_0 = 1$ zum Zinssatz r.
- Ankauf/Verkauf von Aktien zum Kurswert S_0.
- Ankauf/Verkauf eines Call mit Ausübungspreis K und Verfallszeit T.
- Der Aktienkurs kann sich in jeder Periode $[i\Delta t, (i+1)\Delta t], i = 0, \ldots, n-1$, mit Wahrscheinlichkeit q um den Faktor u und mit Wahrscheinlichkeit $1-q$ um den Faktor d verändern. Dabei sind u, d Zahlen mit $u > d > 0$. Der Wert der Aktie $S^{n,k}$ zur Zeit T bei k u–Veränderungen und bei $n-k$ d–Veränderungen ist offenbar
$$S^{n,k} = u^k d^{n-k} S_0 .$$
- Der Aktienkurs verhält sich also wie ein Baum, der aus einzelnen Ein-Perioden-Modellen zusammengesetzt ist; man spricht daher von **Baummodellen**. Er kann zum Endzeitpunkt T offenbar $n+1$ verschiedene Werte annehmen.
- $(n = 2)$ Wert des Call zur Zeit $T = 2\Delta t$:
$$C_T = \begin{cases} C^{uu} := (u^2 S_0 - K)^+ & \text{mit Wahrscheinlichkeit } q^2 \\ C^{ud} := (udS_0 - K)^+ & \text{mit Wahrscheinlichkeit } 2q(1-q) \\ C^{dd} := (d^2 S_0 - K)^+ & \text{mit Wahrscheinlichkeit } (1-q)^2 \end{cases}$$
Wir erhalten damit wieder wie oben eine Preisformel für die Option:
$$C_0 = e^{-2r\Delta t}(p^2 C^{uu} + 2p(1-p)C^{ud} + (1-p)^2 C^{dd}) = e^{-rT}\mathbb{E}_p((S_T - K)^+) .$$
Dabei ist wieder
$$p = \frac{e^{r\Delta t} - d}{u - d} \in [0, 1] .$$

Nun schreiben wir den allgemeinen Fall auf. Auf die Herleitung können wir verzichten, es wäre reine Schreibarbeit. Die Preisformel für die Option sieht so aus:

$$C_0 = e^{-rn\Delta t} \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} (S^{n,k} - K)^+ \,. \tag{2.8}$$

Der Name **Binomialmodell** erklärt sich nun dadurch, dass die Anzahl X_n der Aufwärtsbewegungen im n-Perioden-Modell einer Binomialverteilung mit Parameter n und p genügt; wir schreiben $X_n \sim \mathcal{B}(n,p)$. Als Konsequenz haben wir, dass

$$\binom{n}{k} p^k (1-p)^{n-k}$$

gerade die Wahrscheinlichkeit ist, dass der Aktienkurs zur Zeit T den Wert $u^k d^{n-k} S_0$ annimmt. Jeder Pfad im Baum, der zu diesem Wert führt, kann in Verbindung mit einer Münzwurffolge gebracht werden, wobei jeder Wurf den Ausgang „up“ mit Wahrscheinlichkeit q bzw. „down“ mit Wahrscheinlichkeit $1-q$ haben kann.

(2.8) lässt wieder folgende Interpretation zu:

$$C_0 = e^{-rT} \mathbb{E}((S_T - K)^+) \ \text{ (bezüglich der Wahrscheinlichkeit } p) \,. \tag{2.9}$$

Bestimme $0 \le m \le n$ mit $S^{n,m-1} < K \le S^{n,m}$. Dann gilt (beachte $u > d$)

$$(S^{n,k} - K)^+ = 0, 0 \le k \le m-1 \,,\ (S^{n,k} - K)^+ = S^{n,k} - K, m \le k \le n \,.$$

Damit erhalten wir aus (2.8)

$$\begin{aligned} C_0 &= e^{-rT} \sum_{k=m}^{n} \binom{n}{k} p^k (1-p)^{n-k} (u^k d^{n-k} S_0 - K)^+ \\ &= S_0 \sum_{k=m}^{n} \binom{n}{k} (pue^{-r\Delta t})^k ((1-p)de^{-r\Delta t})^{n-k} - Ke^{-rT} \sum_{k=m}^{n} \binom{n}{k} p^k (1-p)^{n-k} \end{aligned}$$

Sei $\overline{p} := pue^{-r\Delta t}$. Dann folgt

$$\begin{aligned} 1 - \overline{p} &= 1 - pue^{-r\Delta t} \\ &= \frac{u - d - (e^{r\Delta t} - d)ue^{-r\Delta t}}{u-d} \\ &= \frac{u - d - (u - due^{-r\Delta t})}{u-d} \\ &= \frac{(u - d - (e^{r\Delta t} - d))de^{-r\Delta t}}{u-d} \\ &= (1-p)de^{r\Delta t} \end{aligned}$$

und daher

$$C_0 = S_0 \sum_{k=m}^{n} \binom{n}{k} \overline{p}^k (1-\overline{p})^{n-k} - Ke^{-rT} \sum_{k=m}^{n} \binom{n}{k} p^k (1-p)^{n-k} \,.$$

Dies können wir so hinschreiben:

$$C_0 = S_0 \mathcal{N}(m, \overline{p}) - Ke^{-rT} \mathcal{N}(m,p) \text{ mit } \mathcal{N}(m,s) := \sum_{k=m}^{n} \binom{n}{k} s^k (1-s)^{n-k} \tag{2.10}$$

Die Darstellung in (2.10) wird **diskrete Black-Scholes-Formel** für den Optionspreis genannt. Die Ähnlichkeit zu (1.2) ist unverkennbar.

Bemerkung 2.10 *Analog zum n-Perioden-Binomialmodell kann ein n-***Perioden-Trinomialmodell** *entwickelt werden. Der Unterschied besteht darin, dass zu jedem Zeitpunkt 3 Kursentwicklungen möglich sind.* □

Wir werden später ein zeitkontinuierliches Modell betrachten, das sich aus einer stochastischen Differentialgleichung heraus entwickelt. Dann stellt sich die Frage, inwiefern das diskrete Vorgehen des n-Perioden-Binomialmodells eine Approximation davon darstellt (qualitative Frage) und wenn dies zutrifft, welche Parameter u, v, p hierzu dann heranzuziehen sind. Dieser quantitativen Frage widmen wir den nächsten Abschnitt. Die qualitative Frage der Konverenz des n-Perioden-Modells gegen das (kontinuierliche) Black-Scholes Model für $n \to \infty$ bei fest gehaltenem T, lässt sich dann für geeignet gewählte Parameter u, v, p beweisen. Hauptargument ist der *Zentrale Grenzwertsatz*. Wir kommen später darauf zurück.

2.2.3 Die Parameter u, v, p für das n-Perioden-Modell

Die Parameter u, d, p für ein n-Perioden-Modell sollen sich aus den folgenden drei Gleichungen errechnen:

$$
\begin{aligned}
e^{r\Delta t} &= pu + (1-p)d && (2.11)\\
e^{(2r+\sigma^2)\Delta t} &= pu^2 + (1-p)d^2 && (2.12)\\
ud &= 1 && (2.13)
\end{aligned}
$$

Wir werden unten zeigen, dass diese drei nichtlinearen Gleichungen eine eindeutige Lösung besitzen.

Wir wollen hier die Argumente für die Herleitung der Formeln nicht geben, wir wollen aber skizzieren, wo sie herkommen und was sie bedeuten.

Die Formeln (2.11),(2.12) beschreiben den Sachverhalt, dass Erwartungswert und Varianz in einem diskreten und kontinuierlichen Modell für die Aktienkursentwicklung gleich sein sollen: (2.11) beschreibt dies für den Erwartungswert, (2.12) tut dies für die Varianz; der Parameter σ^2, der ja für die Varianz der Basiswertkurse steht, kommt so ins Spiel. Die Gleichung (2.13) wurde hinzugenommen, um die Bestimmtheit der drei Parameter zu erzwingen. Sie beschreibt eine gewisse Symmetrie bei Kursanstieg und Kursabfall: wenn sich der Kurs in zwei direkt aufeinanderfolgenden Schritten nach oben und dann nach unten bewegt, kehrt er wieder in den Ausgangspunkt zurück.

Nun zur Lösung der drei Gleichungen. Wir listen Konsequenzen auf, die sich sukzessive ergeben:

$$
\begin{aligned}
1 &= e^{-r\Delta t}(pu + (1-p)d)\\
ud &= e^{-r\Delta t}p(u-d) + de^{-r\Delta t}\\
p &= \frac{e^{r\Delta t} - d}{u-d}, \; 1-p \;=\; \frac{u - e^{r\Delta t}}{u-d}\\
pu^2 + (1-p)d^2 &= e^{r\Delta t}(u+d) - 1\\
e^{(2r+\sigma^2)\Delta t} &= e^{r\Delta t}(u+d) - 1\\
u &= e^{(r+\sigma^2)\Delta t} + e^{-r\Delta t} - d\\
1 &= d(e^{(r+\sigma^2)\Delta t} + e^{-r\Delta t} - d)\\
0 &= d^2 - (e^{(r+\sigma^2)\Delta t} + e^{-r\Delta t})d + 1
\end{aligned}
$$

Daraus schließt man:

$$d_{\pm} \;=\; \beta \pm \sqrt{\beta^2 - 1}\;, \text{ wobei } \beta = \frac{1}{2}(e^{-r\Delta t} + e^{(r+\sigma^2)\Delta t})\,. \tag{2.14}$$

Wir haben (wegen $u > d > 0$) das negative Vorzeichen zu wählen und erhalten

Lemma 2.11 *Die Gleichungen* (2.11),(2.12),(2.13) *besitzen die Lösung*

$$u \;=\; \beta + \sqrt{\beta^2 - 1} \tag{2.15}$$

$$d \;=\; \beta - \sqrt{\beta^2 - 1} \tag{2.16}$$

$$p \;=\; \frac{e^{r\Delta t} - d}{u - d} \tag{2.17}$$

wobei $\beta = \frac{1}{2}(e^{-r\Delta t} + e^{(r+\sigma^2)\Delta t})$ *ist.*

Im Ein-Perioden-Modell haben wir die Ungleichung $d \leq e^{r\Delta t} \leq u$ aus der Arbitragefreiheit des Marktes gefolgert. Diese Ungleichung ist für die oben berechneten Werte für u, d gültig, wie folgende Rechenschritte zeigen:

$$\begin{aligned} \beta - \sqrt{\beta^2 - 1} \leq e^{r\Delta t} \leq \beta + \sqrt{\beta^2 - 1} \quad &\Longleftrightarrow \quad |e^{r\Delta t} - \beta| \leq \sqrt{\beta^2 - 1} \\ &\Longleftrightarrow \quad 2\beta e^{r\Delta t} \geq e^{2r\Delta t} + 1 \\ &\Longleftrightarrow \quad \beta \geq \frac{1}{2}(e^{-r\Delta t} + e^{r\Delta t}) \end{aligned}$$

Die letzte Ungleichung ist nun richtig für alle $r \geq 0, \sigma \geq 0, \Delta \geq 0$, da $\beta = \frac{1}{2}(e^{-r\Delta t} + e^{(r+\sigma^2)\Delta t})$ ist.

Aus der Darstellung $u = \beta + \sqrt{\beta^2 - 1}$ erhält man mit der Darstellung $\beta = \frac{1}{2}(e^{-r\Delta t} + e^{(r+\sigma^2)\Delta t})$ die Taylorentwicklung

$$u = e^{\sigma\sqrt{\Delta t}} + O((\Delta t)^{\frac{3}{2}}) \text{ für } \Delta t \to 0\,.$$

Mit der Wahl der Parameter

$$u = e^{\sigma\sqrt{\Delta t}}\,,\; d = e^{-\sigma\sqrt{\Delta t}}\,,\; p = \frac{e^{r\Delta t} - d}{u - d}$$

erhalten wir eine approximative Binomialmethode. Die Eigenschaft $d \leq e^{r\Delta t} \leq u$, die notwendig ist für die Arbitragefreiheit des Marktes, ist genau dann erfüllt, wenn

$$\Delta t \leq \frac{\sigma^2}{r^2} \tag{2.18}$$

gilt. Dies bedeutet, dass der Zeitschritt nicht zu groß sein darf.

Bemerkung 2.12 *Eine alternative Binomialmethode entsteht, wenn man die Gleichung* (2.13) *durch*

$$p \;=\; \frac{1}{2}$$

ersetzt. Andere Wahlen von p führen zu weiteren Varianten des Binomialmodells. □

Bemerkung 2.13 *Ist* C_0 *der Call-Preis zum Zeitpunkt* $t = 0$ *nach dem* n*-Periodenmodell mit Schrittweite* Δ, *Verfallszeit* $T = n\Delta t$, *Ausübungspreis* K, *up-Faktor* $u = e^{\sigma\sqrt{\Delta t}}$, *down-Faktor* $v = e^{-\sigma\sqrt{\Delta t}}$, *dann gilt*

$$d_{\pm}(\sigma) = \frac{\ln(S/K) + (r \pm \sigma^2/2)T}{\sigma\sqrt{T}}$$

die Grenzwertbeziehung

$$\lim_{\Delta t \to 0} C_0 = S\mathcal{N}(d_+) - Ke^{-rT}\mathcal{N}(d_-)\,,$$

also die Formel, die wir im letzten Kapitel schon benutzt haben. □

2.3 Die Binomialmethode

Die Idee des Binomialverfahrens ist die Verwendung das Binomialmodells als Approximation des Black-Scholes Modells zur Optionpreisberechnung. Das Binomialverfahren besteht aus zwei Phasen, der **Vorwärtsphase** und der **Rückwärtsphase**.

Vorwärtsphase: Initialisierung der zukünftigen Wertpapierkurse.

- Berechne die Parameter u, d, p.
- Dann erzeugt man einen Vektor, in dem die möglichen Ausgänge des Binomialbaumes abgespeichert werden:
$$S^{n,j} := u^j d^{n-j} S_0\,,\ j = 0, \ldots, n\,.$$
- Berechnung der Optionswerte zum Ausübungszeitpunkt T :
$$V^{n,j} := (S^{n,j} - K)^+\,,\ j = 0, \ldots, n\,.$$

Rückwärtsphase: Berechnung der Optionspreise zu früheren Zeitpunkten

- $$V^{i,j} := e^{-r\Delta t}(pV^{i+1,j+1} + (1-p)V^{i+1,j})\,,\ i = 0, \ldots, j, j = 0, \ldots, n$$
- Ergebnis: $V^{0,0} = C_0$

Wir haben nur noch die Rückwärtsphase zu erläutern. Sie ist die Umsetzung der Darstellung des Optionspreises mittels der risikoneutralen Wahrscheinlichkeit p. Aus

$$e^{r\Delta t} = pu + (1-p)d$$

folgt

$$e^{r\Delta t} S^{i,j} = puS^{i,j} + (1-p)dS^{i,j} = pS^{i+1,j+1} + (1-p)S^{i+1,j}$$

und

$$e^{r\Delta t}\mathbb{E}_p(S^{i,j}) = p\mathbb{E}_p(S^{i+1,j+1}) + (1-p)\mathbb{E}_p(S^{i+1,j})\,.$$

Daraus folgt die obige Darstellung für die Preise $V^{i,j}$.

Algorithm 2.1 Binomialverfahren/Europäische Call-Option

EIN Basiskurs S_0, Ausübungszeitpunkt T, Ausübungspreis K, Baumtiefe n, „Volatilität“ σ.

Schritt 0 $\Delta t := T/n$, $u := e^{\sigma\sqrt{\Delta t}}$, $d = e^{-\sigma\sqrt{\Delta t}}$, $p = \dfrac{e^{r\Delta t} - d}{u - d}$

Schritt 1 Für $j = 0, \ldots, n$ berechne $S^{j,n} := u^j d^{n-j} S_0$.

Schritt 2 Für $j = 0, \ldots, n$ berechne $V^{j,n} := (S^{j,n} - K)^+$.

Schritt 3 Für $i = n-1, \ldots, 0, j = 0, \ldots, i$ berechne $V^{j,i} := e^{-r\Delta t}(pV^{j+1,i+1} + (1-p)V^{j,i+1})$.

AUS Optionspreis $C_0 := V^{0,0}$.

Der Rechenaufwand bei der Realisierung des obigen Algorithmus wächst wie $O(n^2)$. Bei der Implementierung des Verfahrens hat man Schleifen zu verwenden, oder die Vektorisierung zu nutzen, wie MATLAB etwa sie bietet.

Bemerkung 2.14 *Bei der Berechnung amerikanischer Call-Optionen muss in Schritt 0 ein zweidimensionales Feld berechnet werden:* $S^{j,i} := u^j d^{i-j} S_0, i = 1, \dots, n, j = 0, \dots, i$. *Ferner muss in Schritt 3 geprüft werden, ob eine vorzeitige Ausübung erfolgt. Dazu muss zu jedem Zeitpunkt, an dem eine vorzeitige Ausübung erlaubt ist, der „europäisch" berechnete Wert der Option verglichen werden mit dem Erlös, der bei sofortiger Ausübung bezüglich des zu diesem Zeitpunkt berechneten Preises des Basisobjekts erzielt werden könnte. Dies geschieht so:*

$$\tilde{V}^{j,i} := e^{-r\Delta t}(pV^{j+1,i+1} + (1-p)V^{j,i+1})$$

und

$$V^{j,i} := \max\{(S^{i,j} - K)^+, \tilde{V}^{j,i}\}.$$

Die „control variate-Technik" – wir kommen auf diese Technik zurück unter dem Stichwort „Varianzreduktion" in Kapitel 4 – macht sich zunutze, dass im Gegensatz zur amerikanischen Option die europäische Option mit Hilfe der Black-Scholes-Gleichung ausgewertet werden kann. Daraus folgt für die amerikanische Option eine Korrekturmöglichkeit:

$$C^a := C^a_{Binom} + C^e_{BS} - C^e_{Binom}.$$

Dabei ist, zum jeweiligen Zeitpunkt, C^a *der korrigierte Werte der amerikanischen Option,* C^e_{BS} *der Wert der europäischen Option nach der Black-Scholes-Formel,* C^a_{Binom}, C^e_{Binom} *die Näherungswerte, die sich aus dem Binomialverfahren für die amerikanische bzw. europäische Option ergeben. Siehe hierzu [8].* □

Bei der rechnerischen Realisierung des Binomialverfahrens können auch Risikokennzahlen miterrechnet werden, etwa die Griechen. Zum Beispiel ergibt sich zum Zeitpunkt $t_1 = \Delta t$ als Approximation für das Delta:

$$\Delta \approx \frac{V^{11} - V^{10}}{S^{11} - S^{10}}.$$

Diese Kennzahl wird eventuell verwendet, um das gewählte Portfolio so umzuschichten, dass es „Δ-neutral wird, d.h. es insensitiv wird gegenüber Veränderungen des Basispreises.

Will man das Vega $\mathcal{V}$ schätzen, hat man zu zwei verschiedenen Werten σ, σ^* der Volatilität das Binomialverfahren durchzuführen. Dann hat man mit den errechneten Werten C, C^* als Vega

$$\mathcal{V} \approx \frac{C - C^*}{\sigma - \sigma^*}.$$

Bemerkung 2.15 *Das Binomialmodell von Cox, Ross und Rubistein konvergiert für* $\Delta t \to 0$ *gegen das kontinuierliche Black-Scholes Modell. Die Konvergenz verhält sich jedoch nicht monoton, sondern oszillierend. Leisen und Reimer (siehe [11]) erreichten eine Verbesserung der Konvergenzgeschwindigkeit durch Abänderungen der Parameter* u, d, p *im Binomialmodell.* □

2.4 Anmerkungen zu Bermuda-Optionen

Bermudsche Optionen dürfen im Gegensatz zu den amerikanischen Optionen nur an bestimmten Zeitpunkten während der Laufzeit vorzeitig ausgeübt werden. Sie können beispielsweise täglich, monatlich oder wöchentlich ausgeübt werden. Der Fälligkeitszeitpunkt T ist dabei in der Menge der Ausübungszeitpunkten enthalten. Die angenommene äquidistante Schrittweite unter den Ausübungszeitpunkten sei mit

$$\Delta a := \frac{T}{m}$$

bezeichnet. Hieraus ergeben sich also die Ausübungszeitpunkte

$$a_l := l\Delta a\,,\; l = 0, \ldots, m\,.$$

Nun stellt sich die Frage, wie das Binomialverfahren angepasst werden kann, um den Wert einer Bermudschen Option mit vorzeitigen Ausübungszeitpunkten zu bestimmen. Drei Lösungsmöglichkeiten bieten sich an:

Anpassung Die einfachste Lösung ist die Anpassung der Diskretisierung im Binomialverfahren an das Ausübungsgitter, also $n = sm, s \in \mathbb{N}$. Dann können wir das Binomialverfahren analog zum Verfahren bei amerikanischen Optionen realisieren.

Interpolation Die Werte des Basisobjekts werden interpolativ aus dem Ausübungsgitter bestimmt. Dies kann auf vielfältige Weisen geschehen: konstant, linear, exponentiell. Dabei meint „exponentiell“:

$$S_{a_l} = S_{t_i} \exp\left(\frac{a_l - t_i}{t_{i+1} - t_i} \cdot \ln(S_{t_{i+1}}/S_{t_i})\right)\,.$$

Diese Art der Interplation nimmt die „geometrische Brownsche Bewegung“ auf und vorweg.

Extrapolation Es liegt nahe, wie die konstante, lineare Interpolation und exponentielle Interpolation umgedeutet werden kann, dass sie zur Approximation von Werten außerhalb von $[t_i, t_{i+1}]$ verwendet werden können.

2.5 Bibliographische Anmerkungen

Die Put-Call-Parität wurde erstmals um 1900 beschrieben (Nelson, Deutsch, Bronzin), allerdings blieben diese Arbeiten lange unentdeckt. Das Binomialmodell wurde 1979 von Cox, Ross und Rubinstein entwickelt; siehe [2, 3]. Über die Verbesserungen von Leisen-Reimer hinaus wurden weitere Verbesserungen/Abänderungen vorgenommen; siehe etwa [1]. Zur Verallgemeinerung „Trinomialbäume“ findet man etwas in [10]. Zum Thema der impliziten Binomialmodelle siehe [4, 5, 9]. Vorschläge zur Realisierung allgemeinerer Binomialverfahren findet man in [6]. Die Bewertung Bermudscher Optionen mit dem Binomialverfahren wird detailliert diskutiert in [12]. Binomialverfahren wurden auch entwickelt zu Optionen mit einer Zinskurve als Basisobjekt (siehe [7]).

2.6 Übungen

1.) Beweise für eine Put-Option P_t die folgenden Schranken:

$$(Ke^{-r(t-T)} - S_t)^+ \le P_t \le Ke^{-r(t-T)},$$

wobei K, r, T, S_t die üblichen Bedeutungen haben.
Hinweis: Put-Call-Parität und Schranken für eine Call-Option.

2.) Erkläre, warum der Wert einer amerikanischen Put-Option

(a) nicht negativ sein kann,

(b) nicht größer als der Kurs des Basisobjekts sein kann,

(c) sich bei der Ausübung aus der Differenz des Ausübungspreises und dem Kurs des Basisobjekts ergibt.

3.) Leite aus der Call-Put-Parität und der Black-Scholes-Preisformel die folgende Optionsprämie P für eine Put-Option ab:

$$P = -S\mathcal{N}(-d_+) + e^{-rT}\mathcal{N}(-d_-)$$

4.) Betrachte

$$u := 1 + \sigma\sqrt{\Delta t}\,,\ d := 1 - \sigma\sqrt{\Delta t}\,,\ p := \frac{1}{2} + \frac{b\sqrt{\Delta t}}{2\sigma} \text{ mit } r, \sigma > 0\,.$$

Unter welchen Bedingungen ist die Bedingung

$$d \le e^{r\Delta t} \le u$$

erfüllt?

5.) Betrachte Zahlen $u, d, p, \sigma, \Delta t, T, n$ mit

$$u = e^{\sigma\sqrt{\Delta t}}\,,\ ud = 1\,,\ T = n\Delta t,\ ,\ p = \frac{e^{r\Delta t} - d}{u - d}\,.$$

Zeige (man kann auch Maple dazu nutzen):

(a) $\lim_{\Delta t \to 0} np(1-p)(\ln(u/d))^2 = \sigma^2 T\,.$

(b) $\lim_{\Delta t \to 0} n(p\ln(u/d) + \ln d) = (r - \sigma^2/2)T\,.$

(c) $\lim_{\Delta t \to 0} \dfrac{2p-1}{\sqrt{\Delta t}} = \dfrac{2r - \sigma^2}{2\sigma}\,.$

6.) Betrachte einen Markt, auf dem die folgenden Call-Optionen auf dasselbe Basisobjekt vorhanden sind:

$$\begin{array}{rrr} C_0 = 5.2 & K = 47.7 & T = 1 \\ C_0 = 12.4 & K = 40 & T = 1 \end{array}$$

Der Zinssatz betrage $r = 0.1$. Konstruiere eine Arbitragemöglichkeit.

7.) Für eine amerikanische Put-Option gibt es immer einen Aktienkurs $S > 0$, so dass ein vorzeitiges Ausüben der Option optimal ist.

(a) Formuliere diesen Sachverhalt etwas genauer.

(b) Beweise den Sachverhalt.

8.) Betrachte zwei europäische Put-Optionen zum gleichen Basisobjekt mit Laufzeit $T = 1$, aber unterschiedlichen Ausübungspreisen. Die Put-Option P_1 mit Ausübungspreis $K =$ 300 € ist zu 17 € notiert, die Put-Option P_2 mit Ausübungspreis $K = 320$ € wird mit 35 € gehandelt. Der Zinssatz ist $r = 25\%$. Finde eine Zusammensetzung eines Portfolios, das zeigt, dass die Fehlbewertung der Optionen in eine Arbitragestrategie umgesetzt werden kann. Ergänze dazu die folgende Tabelle.

Strategie	**Zahlungsstrom zur Zeit** $t = 0$	**Zahlungsstrom zur Zeit** $t = 1$		
		$S_1 \le 300$	$300 < S_1 < 320$	$S_1 \ge 320$
Kaufe ... Verkaufe ... Geldanleihe ...				
Summe				

9.) Ein Investor habe das Vermögen $x > 0$. Er kann zur Zeit $t = 0$ sein Vermögen aufteilen und Aktien kaufen oder verkaufen, Geld anlegen oder Geld leihen; der Znssatz sei r. Die Handelsstrategie sei $x = k + aS_0$, wobei k der angelegte/geliehene Geldbetrag ist, a die

Anzahl der gehaltenen Aktien beschreibt. Er lässt dann seine Wertpapierzusammenstellung bis zum Zeitpunkt $T > 0$ unverändert. Wir nehmen nun an, dass der Aktienkurs zur Zeit T mit Wahrscheinlichkeit q auf uS_0 gestiegen und mit Wahrscheinlichkeit $1 - q$ auf dS_0 gefallen ist ($u > 1 > d > 0$).

(a) Welche Werte nimmt das Endvermögen $X(T)$ an?

(b) Berechne den Erwartungswert $\mathbb{E}(X(T))$.

(c) Berechne die Varianz $\mathbb{V}(X(T))$.

10.) Eine europäische Call-Option zum Basisobjekt Aktie notiert zu 5.75 € bei einem Ausübungspreis $K = 90$ €. Die Laufzeit ist ein Jahr, der Zinssatz $r = 10\%$. Der aktuelle Kurs der Aktie betrage 87.50 €.
Bestimme über die Put-Call-Parität den Wert einer Put-Option gleichen Typs.

11.) Bestimme das asymptotische Verhalten des Wertes C einer europäischen Call-Option nach der Black-Scholes-Preisformel bei der jeweiligen grenzwertigen Veränderung einer Einflussgröße und bei gleichzeitiger Konstanz der restlichen Einflussgrößen:

(a) $\lim_{S\to 0} C(S, K, r, T, \sigma)$

(b) $\lim_{S\to \infty} C(S, K, r, T, \sigma)$

(c) $\lim_{K\to 0} C(S, K, r, T, \sigma)$

(d) $\lim_{S\to \infty} C(S, K, r, T, \sigma)$

12.) Berechne mit dem Binomialverfahren den Preis der Option für die „Diskretisierung“ $n = 10$ zu folgenden Ausgangsdaten:

$$K = 100, S_0 = 103, r = 0.04, \sigma = 03, T = 1\,.$$

13.) Berechne mit dem alternativen Binomialverfahren ($p = \frac{1}{2}$) den Preis der Option für die „Diskretisierung“ $n = 10$ zu folgenden Ausgangsdaten:

$$K = 100, S_0 = 103, r = 0.04, \sigma = 03, T = 1\,.$$

Literaturverzeichnis

[1] B.H Camp. Approximation to the point binomial. *Annals of Mathematical Statistics*, 22:130–131, 1951.

[2] J.C. Cox, J.E. Ingersoll and S.A. Ross. A theory of the term structure of interest rates. *Econometrica*, 53:385–407, 1985.

[3] J.C. Cox, S.A. Ross and M. Rubinstein. Option pricing: a simplified approach. *J. of Financial Economics*, 7:229–263, 1979.

[4] E. Derman and I. Kani. The volatility smile and its implied tree. *http://www.gs.com/qs/*, xx, 1994.

[5] E. Derman, I. Kani and N. Chriss. Implied trinomial trees of volatility smile, 1996. Goldman & Sachs Quantitative Research Notes.

[6] D.J. Higham. Nine ways to implement the binomial method for option evaluation in MATLAB. *SIAM Review*, 44:661–677, 2002.

[7] T. Ho and S. Lee. Term structure movements and pricing interest rate contingent claims. *J. of Finance*, 41:1011–1029, 1986.

[8] J.C. Hull and A. White. The use of the control variate technique in option pricing. *J. of Financial and quanitative analysis*, 23:237–251, 1988.

[9] J.C. Jackwerth. Option-implied risk-neutral distribution and implied binomial trees: a literature review. *J. of Derivatives*, 7:66–82, 1999.

[10] Y. Kwok. *Mathematical Models of Financial Derivatives.* Springer, Singapur, 1998.

[11] D. Leisen and M. Reimer. Binomial methods for option valuation – examining and improving convergence. *Appl. Math. Finance*, 3:319–346, 1996.

[12] C. Warawko. Numerische Verfahren zur Bewertung Bermudscher Optionen. Master's thesis, Universität Bonn, 2006.

Kapitel 3

Zufallszahlen und ihre Erzeugung

Gott würfelt nicht!
Albert Einstein

Finanzderivate werden modelliert, indem man im Allgemeinen den Zufall, der in der Kursentwicklung der Basisobjekte liegt, modelliert. Optionspreise stellen sich dann, wie wir schon gesehen haben, als Erwartungswerte der Auszahlungsfunktion dar. Zur Simulation dieser Erwartungswerte werden Zufallszahlen unterschiedlicher Verteilungscharakteristiken benötigt. Wir beschreiben hier die Erzeugung von Zahlen, die als Zufallszahlen Verwendung finden können.[1] Wir nennen sie Pseudozufallszahlen und ihre Generatoren Pseudozufallszahlen-Generatoren.

In diesem Kapitel betrachten wir ausschließlich gleichverteilte Zufallszahlen, im nächsten Kapitel erzeugen wir daraus Zahlen, die eine vorgegebene Verteilung besitzen. und wenden sie auf die Berechnung von Erwartungswerten an.

3.1 Grundsätzliches

Der Wunsch, zufällige Ereignisse zu generieren, ist zwar keine ausschließliche Erscheinung des Computerzeitalters, er wurde aber durch die Rechenmöglichkeiten doch stark in den Vordergrund gerückt.[2] Zur Geschichte:

- 1938: Kendall und Babington-Smith erzeugen mit einer schnell drehenden Scheibe 100 000 zufällige Ziffern.
- Seit 1940/50 werden numerische und arithmetische Verfahren verwendet, um Zufallszahlen zu generieren.
- 1957: Das 1. ERNIE-Projekt (Electronic Random Number Indicator Equipment) wurde durch Sidney Broadhurst, Tommy Flowers and Harry Fensom realisiert. Es wurden mit Hilfe von Vakuumröhren bis zu 50 Zufallsziffern pro Sekunde erzeugt.
- 1955: Die Rand-Corporation veröffentlicht ein Buch mit ca. 1 Million Zufallsziffern.
- 1983: Miyatake baut eine Vorrichtung, um durch das Zählen von Gammastahlen zufällige Ergebnisse zu generieren.

[1] Anyone who considers arithmetical methods of producing random digits is, of course, in a state of sin. *John von Neumann*

[2] Aus der Bibel: …*Nachdem sie ihn gekreuzigt hatten, warfen sie das Los und verteilten seine Kleider unter sich.*… Matthäus 27:35

- 1995 Marsaglia produziert eine CD-ROM, auf der ca. 4.8 Milliarden Zufallszahlen gespeichert sind.[3]

Was ist die Motivation für das Bestreben, Zufallszahlen zu erzeugen? Warum zu Generatoren von Pseudozufallszahlen greifen, also zu Generatoren, die nicht den "echten Zufall" verwenden? Es sind die vielfältigen Anwendungsbereiche, die nach Zufallszahlen fragen: *Erzeugung zufälliger Ereignisse entsprechend statistischer Vorgaben, Kryptographische Anwendungen (Erzeugung von Schlüsselzahlen,...), Monte Carlo Simulation (insbesondere in Computational Finance), Simulation von Abläufen der realen Welt (Ampelschaltungen), Globale Optimierung, Spiele (wo taucht der Bösewicht auf?),...*.

Echten Zufall könnte man etwa bei **physikalischen Generatoren** vorfinden: Zerfall radioaktiven Materials, Schwankungen in elektrischen Netzen, Messung thermischen Rauschens, ... Diesen so genannten echten Zufall herzustellen ist sehr aufwendig. Man nennt solche physikalische Effekte nutzende Generatoren **Hardware-Generatoren**. Hier steht natürlich die Frage im Raume, ob das physikalische Phänomen wirklich den Zufall simuliert, wie es etwa die Quantenmechanik voraussagt.

Das Problem jeden Verfahrens, Zufallszahlen (auf dem Computer) mittels eines Algorithmus zu erzeugen, ist offenbar: „erzeugen" und „Zufall" ist ein Widerspruch in sich! Da der Determinismus schon per Definition eine Eigenschaft des Algorithmus ist, steht jeder Nachfolger einer Zufallszahl deterministisch fest. Für die oben genannten Anwendungsgebiete genügen jedoch diese „Pseudozufallszahlen". Wir lassen das Präfix „Pseudo" meist weg.

Pseudozufallszahlen sollen Zahlenfolgen sein, die „zufällig" sind, d.h. die Eigenschaften besitzen, die dem echten Zufall nahe kommen. Also ist man gezwungen, den Zufall deterministisch möglichst gut nachzustellen. Um die umständliche Verwendung von Tabellen zu vermeiden, werden Zufallszahlen verwendet, die im Allgemeinen durch Iterationen ad hoc hergestellt werden. Die so erzeugten Zufallszahlen haben den Vorteil, dass sie konstruierbar sind, und haben den Nachteil, dass sie deterministischen Charakter besitzen. Alles, was wir hier zunächst zur Sprechweise „Zufallszahl" sagen können, ist, dass jedenfalls kein Muster, keine Struktur in der Folge erkennbar sein soll. Die Wahrscheinlichkeitstheorie und Statistik stellt Hilfsmittel bereit, solche Folgen auf ihre Zufälligkeit zu testen; siehe unten.

In diesem Kapitel wollen wir Zufallszahlen konstruieren, die gleichverteilt sind im Einheitsintervall. Im nächsten Kapitel erzeugen wir dann Zahlenfolgen mit vorgegebener Verteilung.

Damit können wir nun Forderungen formulieren, die an einen Zufallszahlen-Generator zu stellen sind. Er soll Zahlen erzeugen, die folgende Eigenschaften haben:

Gleichverteilung Die Zufallsfolge genügt der Gleichverteilung. Diese Forderung ergibt sich aus unserem Ziel. Daraus leitet sich die Forderung ab, dass eine Zufallszahl immer an einen Bereich gebunden ist. Unten ziehen wir uns immer auf den Bereich $[0, 1]$ zurück.

Unvorhersagbarkeit Hat man eine Zahl erzeugt, sollte die nächste nicht vorhersagbar sein. Dies bedeutet, dass der Konstruktionsmechanismus komplex genug ist, um zu vermeiden, dass das Konstruktionsprinzip abgelesen werden kann.

Reproduzierbarkeit Um die Fehlersuche zu erleichtern und verschiedene Simulationen einfacher miteinander vergleichen zu können, ist es wichtig, dass eine einmal erzeugte Zufallsfolge immer wieder reproduziert werden kann.

Bei Spielen (der Bösewicht soll nicht immer zur selben Zeit auf der Bildfläche erscheinen) und in der Kryptologie steht dem gegenüber die Forderung nach Irreproduzierbarkeit. In der Kryptographie ist diese Forderung „unverzichtbar".

[3] `random.org` brüstet sich, seit 1998 857 Milliarden Zufallsbits erzeugt zu haben.

Dazu kommen die Forderungen, dass der Generator schnell ist und möglichst wenig Speicherplatz auf dem Computer belegt.

Definition 3.1 *Eine Zufallsvariable X heißt* **gleichverteilt** *im Intervall $[a,b]$, in Zeichen $X \sim \mathcal{U}[a,b]$, wenn die Verteilungsfunktion von X eine Dichte f besitzt mit*

$$f(x) = \begin{cases} \frac{1}{b-a} & , \textit{falls } x \in [a,b] \\ 0 & , \textit{sonst} \end{cases} .$$

□

Was bedeutet „Gleichverteilung" im Zusammenhang mit Pseudozufallszahlen? Die Gleichverteilung einer Zufallsvariable kommt nicht direkt in Frage, denn man kann ja beim „Ziehen" einer Zufallszahl nach obigem Vorgehen nicht von einem Zufallsexperiment sprechen. Wir müssen auf die „**empirische Gleichverteilung**" zurückgreifen und dann die Frage stellen, inwieweit die empirische Verteilungsfunktion in Übereinstimmung mit einer Gleichverteilung steht.

Definition 3.2 *Die* **empirische Verteilungsfunktion** *einer Stichprobe $x^1, \ldots, x^n$ ist die Funktion*

$$H(x) := \frac{1}{n} \sum_{i=1}^{n} \chi_{(-\infty,x]}(x^i) \, , \; x \in \mathbb{R} \, .$$

□

Eine erste Realisierung der Pseudozufallserzeugung bestand in der Nutzung der Dezimalziffern transzendenter Zahlen. Die Zahl π wurde 1873 mit 703, 1960 mit 100 000 und 1986 mit 10^7 Dezimalstellen berechnet. Die statistische Analyse ergab, dass keine signifikanten Abweichungen von der Gleichverteilung auftraten. Da die Algorithmen zur Berechnung transzendenter Zahlen in der Regel sehr kompliziert sind, werden in der Praxis meist andere Algorithmen benutzt.

i	z_i	$u_i := 0.z_i$	z_i^2
0	7182	$---$	51 5811 24
1	5811	0.5811	33 7677 21
2	7677	0.7677	58 9363 29
3	9363	0.9363	87 6657 69
...	...	...	...
12	0012	0.0012	00 0001 44
13	0001	0.0001	00 0000 01
14	0000	0.0000	00 0000 00
...	...	...	...

Abbildung 3.1: Middle-Square

Einer der ältesten Generatoren ist die **Middle-Square-Methode**, die 1940 von *von Neumann* und *Metropolis* entwickelt wurde und die wie folgt durchgeführt wird.

> Wähle eine 4-stellige Zahl, quadriere sie, man erhält eine höchstens 8-stellige Zahl. Ist das Ergebnis nicht 8-stellig, füllt man sie links mit Nullen auf 8 Stellen auf. Die mittleren 4 Ziffern wählt man nun als erste Zufallszahl und als neuen Startwert für das Vorgehen.

Schneide aus einer mehrstelligen Zahl eine Ziffernfolge heraus, quadriere sie, schneide wieder heraus, ... Ist man unvorsichtigt, bekommt man eine nicht sehr brauchbare Folge. Etwa erhält man mit dem Startwert 8441:

$$8441, 2504, 2700, 2900, 4100, 8100, 6100, 2100, 4100, 8100, \ldots \, .$$

Es ist sogar noch „schlimmer", wie das Beispiel in obiger Tabelle (siehe 3.1) andeutet. Die ersten Schritte des Middle-square-Algorithmus scheinen brauchbare Zufallszahlen zu liefern, die

Fortsetzung bei $i = 12$ zeigt aber, dass die Iteration bei der „Zufallszahl“ Null endet. In der Tat tendiert der Algorithmus in vielen Fällen dazu, bei Null zu enden. Also scheint der Algorithmus unbrauchbar zu sein, Zufallszahlen zu erzeugen. Anderenfalls ist das obige kurze Stück $8100, 6100, \ldots, 8100$ das periodische Stück einer doch recht langen nichtperiodischen Zahlensequenz, die mit dem Startwert 6239 beginnt.

Gehen wir die Generation von Zufallszahlen etwas grundsätzlicher an. Dazu einige allgemeine Bemerkungen. Sei M eine endliche Menge. Pseudozufallszahlen, deren Konstruktionsmethode wir hier besprechen wollen, ergeben sich als Iterierte einer Funktion

$$f : M \longrightarrow M$$

in folgender Weise:

$$x^{n+1} := f(x^n)\,,\ n \in \mathbb{N}_0\,. \tag{3.1}$$

Die Abbildung f heißt der **Generator** der Folge und der Startwert x^0 **Samen/seed** der Pseudozufallsfolge $(x^n)_{n\in\mathbb{N}}$. Die Folge selbst wird oft auch **Orbit** genannt.

Bemerkung 3.3 *Die Iteration* (3.1) *kann verallgemeinert werden zu*

$$x^{n+1} := R(x^n, x^{n-1}, \ldots, x^{n-k})\,,\ n \in \mathbb{N}_0\,, \tag{3.2}$$

mit einer Funktion $R : M^k \longrightarrow M$. Wir werden dies in Spezialfällen kennenlernen. Es ist klar, dass wir dann ein „Anfangsstück“ $x^0, \ldots, x^k$ benötigen. □

Da die Menge M endlich ist, können nicht alle Folgenglieder x_n verschieden sein. Es gibt also Indizes k, l mit $x^k = x^l$; o. E. $k > l$. Seien k, l die ersten Indizes, für die dies eintritt. Sei damit $r := k - l$. Da $x^k = x^l$ gilt, folgt $x^{n+r} = x^n$ für alle $n \geq l$. Also wird der **Orbit** $(x^n)_{n\in\mathbb{N}}$ periodisch mit Periode r; wir haben einen Zyklus der Länge r. Verlangt man, dass jedes Element der Menge M die Chance hat im Orbit aufzutauchen, muss der Zyklus ganz M umfassen. Daraus folgt, dass die Abbildung f surjektiv sein muss. Da M endlich ist, hat f also sogar bijektiv zu sein. Wir werden unten sehen, dass die Bijektivität keineswegs dafür schon ausreicht, ein guter Generator zu sein.

Bemerkung 3.4 *Die obigen allgemeinen Überlegungen zu den Iterationen sind völlig parallel zur Betrachtung von dynamischen Systemen. Es ist daher nicht überraschend, dass gerade chaotische dynamische Systeme auch genutzt werden, um Zufallsfolgen zu erzeugen.* □

Die Folge gemäß (3.1) ist durch die Wahl von f und x^0 vollständig bestimmt. Durch geschickte Wahl von f – gewünscht wird eine gute Durchmischung von M – kann man erreichen, dass sich die Folge für viele Anwendungen wie eine Zufallsfolge verhält.

Bemerkung 3.5 *Die Wahl des Samens (Startwert der Iteration) kann für Anwendungen durchaus kritisch sein. Will man etwa ein Internet-Pockerspiel mit zufälligen Kartenzuordnungen erzeugen, so hat man 2^{226} verschiedene Möglichkeiten zu bedenken. Ist der Samen aber grundsätzlich nur eine 32-Bit-Zahl, so wird diese Anzahl auf 2^{32} Möglichkeiten reduziert. Die Ermittlung des Samens mittels Synchronisation mit der Server-Uhr etwa schränkt diese Anzahl der Möglichkeiten noch viel drastischer ein.* □

3.2 Kongruenzgeneratoren

Eine der ältesten Methode, Pseudozufallszahlen nach obiger Vorgehensweise zu konstruieren, ist es, die (affin-)linearen Kongruenz-Generatoren (LCG) zu nutzen. Sie gehen im wesentlichen zurück auf Lehmer (1948). Zunächst jedoch ein kleiner Ausflug in die modulare Arithmetik.

Wie wir wissen, lassen sich die ganzen Zahlen in Restklassen modulo eines Moduls $m \in \mathbb{N}, m \geq 2$, einteilen. Dies geschieht durch Division mit Rest:

> Für alle $a \in \mathbb{Z}, m \in \mathbb{N}$ gibt es eindeutig bestimmte Zahlen $q, r \in \mathbb{Z}$ mit $a = mq + r$ und $0 \leq r < m$.
> Dafür schreiben wir auch kurz: $a \equiv r \mod m$.

Damit teilt man die ganzen Zahlen nach den Resten den Restklassen $\mathbb{Z}_m := \{[0], [1], \ldots, [m-1]\}$ zu:

$$[i] := \{z \in \mathbb{Z} | m \text{ teilt } z - i\}.$$

Wir verzichten nun auf die algebraische Schreibweise und schreiben die Restklassen kurz so hin: $0, 1, \ldots, m-1$. Wichtig ist nun, dass wir vereinbaren, wie nun addiert und multipliziert werden soll:

$$i + j := k \in \{0, 1, \ldots, m-1\} \text{ falls } i + j - k \text{ von } m \text{ geteilt wird}$$

$$i \cdot j := ij := k \in \{0, 1, \ldots, m-1\} \text{ falls } ij - k \text{ von } m \text{ geteilt wird}$$

Dahinter versteckt sich nun die modulare Arithmetik, die auf jedem Rechner implementiert ist. Sie kann auf vielfältige Weisen genutzt werden. Beliebt sind die Module $m = p^e,\ e \in \mathbb{N}, p$ Primzahl, oder $m = 2^e, e \in \mathbb{N}$.

Kehren wir nun zurück zur Erzeugung von Zufallszahlenfolgen. Wir betrachten folgende Konkretisierung:

$$M := \{0, 1, \ldots, m-1\},\ f : M \ni x \longmapsto ax + b \mod m \in M, \tag{3.3}$$

mit einem Modul m. Hier sind $a, b \in \mathbb{Z}$; natürlich ist $a \neq 0$.

Warum könnte diese Konstruktionsweise zu guten Zufallszahlen führen? Wir vermuten, dass die modulo-Rechnung Spuren der Erzeugung verschleiert. Dabei ist der Modul m oft von der Registergröße der „Rechenmaschine". Die Qualität der Generatoren hängt ab von der Wahl von a, b, m; siehe unten.

Algorithm 3.1 Kongruenz-Generator nach Lehmer

EIN Parameter a, b, m. Startwert $x^0 \in [0, m-1)$. Iterationsanzahl $k_{it} \leq m - 1$.

Schritt 1 Für $k = 0, \ldots, k_{it} - 1$ berechne $x^{k+1} := ax^k + b \mod m$.

AUS „Zufallszahlen" $x^1, \ldots, x^{k_{it}}$.

Bemerkung 3.6 *Durch die Generatoren in* (3.13) *werden Zufallszahlen in* $M := \{0, 1, \ldots, m-1\}$ *erzeugt. Aus einer Zufallszahl* $y \in \{0, \ldots, m-1\}$ *ergibt sich eine Zufallszahl* z *in* $[0, 1]$ *ganz einfach so:* $z := \dfrac{y}{m-1}$. □

Damit die Abbildung f aus (3.13) bijektiv wird, muss a zu m teilerfremd sein. Als Vorbereitung auf die Frage, unter welchen Bedingungen dieser Typ von Generatoren einen Zyklus maximaler Länge erzeugt, schauen wir uns Beispiele an.

a	b	m	**Periode**	**Quelle**
24298	99991	199017		Texas Instruments (TI-59)
65539	0	2^{31}	nicht maximal	RANDU
137	187	2^{16}	maximal	Knuth
193	73	2^{16}	maximal	
7^5	0	$2^{31}-1$	maximal	Park-Miller
131	0	2^{35}	maximal	Neave
16333	25887	2^{15}	maximal	Oakenfull
3432	6789	9973	???	Oakenfull
1711	0	30269	???	Wichman-Hill

Abbildung 3.2: Gebräuchliche Generatoren

Beispiel 3.7 *Betrachte die spezielle Wahl* $m = 10, a = b = 7$. *Hier ist der erzeugte Zyklus*

$$7, 6, 9, 0, 7, 6, 9, 0, \ldots$$

ziemlich kurz, obwohl natürlich $a = 7$ *zu* m *teilerfremd, d.h.* a *ein invertierbares Element in* $\mathbb{Z}_{10}$ *ist.* □

Beispiel 3.8 *Betrachte die spezielle Wahl* $m = 2^{31}, a = 65539, b = 0$. *Dies ist der Zufallsgenerator RANDU, wie er von IBM in den Computern in den 60er Jahren verwendet wurde. Die maximal erreichbare Zykluslänge* r *ist hier nicht ganz maximal, aber mit* $r = 2^{29}$ *nahezu maximal. Wir kommen später auf die Güte dieses Generators noch zu sprechen.*[4] □

Hier ist das Hauptergebnis über affine Kongruenzgeneratoren.[5]

Satz 3.9 (Knuth) *Mit* $m, a, b \in \mathbb{Z}$, $m \geq 2$ *betrachte die Abbildung*

$$f : \{0, \ldots, m-1\} \ni x \longmapsto ax + b \mod m \in \{0, \ldots, m-1\}. \tag{3.4}$$

Für beliebiges $x^0 \in \{0, \ldots, m-1\}$ *sei die Folge* $(x^n)_{n \in \mathbb{N}}$ *definiert durch*

$$x^{n+1} := f(x^n), \, n \in \mathbb{N}_0 .$$

Genau dann ist diese Folge periodisch mit der maximalen Periodenlänge m, *wenn folgende Bedingungen erfüllt sind:*

a) b *und* m *sind teilerfremd.*

b) $p|(a-1)$ *für alle Primteiler* p *von* m;

c) $4|(a-1)$ *falls* $4|m$.

Beweis:
Für den Beweis verweisen wir auf den Anhang 3.6.

Satz 3.22 nennt uns die Bedingungen für einen linearen Kongruenz–Generator, damit er der Minimalforderung, einen Zyklus maximaler Länge zu erzeugen, genügt. Jedoch garantieren diese Bedingungen noch lange keinen guten Zufallsgenerator, wie nachfolgendes Beispiel zeigt.

[4] 2008: Microsoft confirms that XP contains random number generator bug.
[5] Random number generators should not be chosen at random. *D. Knuth*

Beispiel 3.10 *Betrachte für einen beliebigen Modul m den Generator $f(x) := x + 1 \mod m$. Kein Zweifel, die Zykluslänge ist maximal, nämlich m, aber die erzeugte Folge $0, 1, 2, \ldots, m - 1, 0, 1 \ldots$ kann sicherlich nicht den Anspruch einer Zufallsfolge erheben.* □

In der Praxis wird häufig ein Modul der Form $m = 2^k$ verwendet (und dazu in der Regel der Multiplikator a im Bereich $\sqrt{m} < a < m - \sqrt{m}$). In diesem Fall bedeuten die Bedingungen des Satzes 3.22 einfach

$$a \equiv 1 \mod 4 \text{ und } b \text{ ungerade}. \tag{3.5}$$

Im Beispiel 3.8 sind diese Bedingungen offenbar verletzt ($a = 2^{16} + 3$ und $b = 0$) und Konsequenz ist ein verkürzter maximaler Zyklus.

Beispiel 3.11 *In der Programmiersprache C++ gibt es einen Generator namens* ***drand48****:*

$$\textit{Modul} = 2^{48}, \, a = 25214903917, \, b = 11.$$

Die Zykluslänge ist maximal, da die Bedingungen (3.5) *erfült sind.* □

Beispiel 3.12 *Von D. Knuth wurde der Generator*

$$\textit{Modul} = 2^{16}, \, a = 137, \, b = 187$$

vorgeschlagen. Die Zykluslänge ist maximal, da die Bedingungen (3.5) *erfüllt sind.* □

Beispiel 3.13 *Ein weiterer Generator:*

$$\textit{Modul} = 2^{16}, \, a = 193, \, b = 73.$$

Die Zykluslänge ist maximal, da die Bedingungen (3.5) *erfüllt sind.* □

Von Knuth ([10]) stammen folgende Empfehlungen für die Wahl von a, b, m:

1. Wähle den Modul m groß.
2. Für a gelte $\sqrt{m} < a < m - \sqrt{m}$.
3. Ist $m = 2^e$, dann sollte gelten $a \mod 8 = 5$, ist $m = 10^e$, dann sollte gelten $a \mod 200 = 21$.
4. Das Ziffernmuster von a sollte unregelmäßig sein.
5. b sollte ungerade sein und nicht durch 5 teilbar.

Man beachte, dass die Forderung „b sollte ungerade sein" den unerwünschten Fall $b = 0$ ausschaltet. In diesem Fall kann es ja zur konstanten Folge $x^i = 0, i \geq k$, kommen.

Im Allgemeinen sollte man niemals Teile einer Pseudozufallszahl als weitere Zufallszahl benützen. Will man aus großen Zufallszahlen kleine gewinnen, d.h. will man von Zufallszahlen x^i in $[0, m)$ auf Zufallszahlen y^i in $[0, r)$ übergehen, so sollte man dies nicht gemäß

$$y^i := x^i \mod r$$

durchführen, sondern etwa mittels

$$y^i := \lfloor x^i / w \rfloor \mod r \,, \quad w \text{ geeignet}.$$

Beispiel 3.14 *Betrachte den Spezialfall* $a = 31415821, b = 1, m = 10^8$. *Der Startwert* $x^0 :=$ 1234567 *liefert Zahlen* $x^1, \dots, x^{20}$, *mit den letzten Bits* $8, 9, 0, 1, 2, 3, 4, \dots, 5, 6, 7$. *Wir beobachten also eine vollkommene Regelmäßigkeit der letzten Bits.*

Hier ist eine nichtlineare Variante eines Kongruenzgenerators, der **Inverse Kongruenz-Generator**:

$$x^i = a\overline{x^{i-1}} + b \mod m \tag{3.6}$$

wobei $\overline{x_{i-1}}$ das multiplikativ „inverse" Element zu x_{i-1} ist, d.h.

$$\overline{x_{i-1}}x_{i-1} \equiv 1 \mod m\,.$$

Falls das inverse Element nicht existiert, setzen wir $\overline{x_{i-1}} := 0$, was dann nur weiterhilft, wenn $b \neq 0$ ist. Wir wissen aus der elementaren Algebra, dass für einen Modul m, der eine Primzahl ist, alle Elemente $x \neq 0$ ein Inverses besitzen.

Von diesen inversen Generatoren ist der Vorteil bekannt, dass die „geometrische Korrelation" (siehe Abschnitt 3.5.2), dass Tupel auf Geraden und Tripel auf Ebenen liegen, nicht auftritt. Der Nachteil dieser Generatoren liegt in dem Aufwand der Multiplikation und ganz besonders im Aufwand der Inversion: die Ermittlung des Inversen ist mit einem Aufwand, der $O(\log_2(m))$ beträgt, verbunden.

3.3 Ein Zufallsbit-Generator

Ist man lediglich an einer Zufallsbitfolge interessiert, kann man den Münzwurf heranziehen: man notiert eine Null, wenn Kopf fällt, eine Eins, wenn Zahl fällt. Sicherlich ist dies kein effizientes Verfahren, denn es dauert, bis eine „lange" Zufallsfolge zustande kommt.

Anfang der 1980er Jahre entstand im Umkreis der Kryptologie eine Vorstellung davon, wie man die Generierung von Zufallszahlen komplexitätstheoretisch angehen könnte. Damit wurde ein neuer Qualitätsstandard gesetzt: eine Pseudozufallsfolge sollte sich durch keinen (effizienten) Algorithmus von einer echten Zufallsfolge unterscheiden lassen. Der erste konkrete Ansatz ist der Blum-Blum-Shub-Generator (BBS-Generator); siehe [1]. Der BBS-Generator setzt bei der Quadratrest-Vermutung an.

Definition 3.15 *Sei* m *ein Modul in* $\mathbb{N}, n \geq 2$. *Eine ganze Zahl* $x \in \mathbb{Z}$ *heißt* **quadratischer Rest** *modulo* m, *wenn es* $a \in \mathbb{N}$ *gibt mit* $x^2 \equiv a \mod m$. *Anderenfalls heißt* a **nichtquadratischer Rest** *modulo* m. □

Eine Quadratzahl ist ein quadratischer Rest modulo jedes Moduls m. Umgekehrt gibt es Zahlen, die selbst keine Quadratzahlen sind, aber ein quadratischer Rest modulo eines Moduls sind. Die Entscheidung, ob eine Zahl ein quadratischer Rest ist, ist nicht einfach zu entscheiden. Eine Hilfe dabei ist der Chinesische Restsatz (siehe Satz 3.21), denn er hilft dabei, aus der Primzahlzerlegung $m = p_1^{r_1} \cdots p_l^{r_l}$ eines Moduls m aus den Quadratresten modulo m auf Quadratreste modulo $p_i, i = 1, \dots, l$, zu schließen und umgekehrt. Damit ist es dann naheliegend, den Sachverhalt für einen Modul m zu klären, der eine Primzahl p ist. Hier der Spezialfall, dass -1 ein quadratischer Rest ist, vollkommen geklärt: -1 ist quadratischer Rest modulo p genau dann, wenn p nicht kongruent 3 modulo p ist; beachte, dass jede Primzahl $\neq 2$ kongruent 1 oder 3 ist.

Vor allem in der Kryptologie stellt sich vielfach die Aufgabe für eine vorgegebene Zahl und einen bekannten Modul zu entscheiden, ob diese Zahl ein quadratischer Rest ist. Diese Fragestellung wird als **Quadratische-Reste-Problem** bezeichnet. Ist der Modul eine Primzahl, so kann dies recht einfach entschieden werden. Andernfalls stellt es sich teilweise recht schwierig dar. Insbesondere besagt die Quadratische-Reste-Annahme, dass es für bestimmte Moduln praktisch nicht möglich ist, diese Frage zu entscheiden. Allerdings, für kleinere Zahlen m können die quadratischen Reste relativ rasch berechnet werden: es genügt, die Zahlen $0 \le x \le m/2$ zu betrachten, denn x^2 und $(x + k \cdot m)^2$ haben denselben Rest.

Algorithm 3.2 Der Zufallsbit-Generator nach Blum, Blum, Shub

EIN Verschiedene Primzahlen p, q mit $p \mod 4 = 3, q \mod 4 = 3$. Iterationsanzahl k_{it}.

Schritt 0 Berechne $m := pq$.

Schritt 1 Wähle $s \in [1, m-1] \cap \mathbb{Z}$ mit s teilerfremd zu m.

Schritt 2 Berechne $x^0 := s^2 \mod m$.

Schritt 3 Für $k = 0, \ldots, k_{it} - 1$ berechne $x^{k+1} := (x^k)^2 \mod m$.

Schritt 4 Für $k = 0, \ldots, k_{it}$ berechne $z^k := x^k \mod 2$.

AUS „Zufallsbits“ $z^0, \ldots, z^{k_{it}}$.

Es ist wichtig, dass die Primzahlen p, q groß und etwa gleich dimensioniert sind. Die letzten $\log_2 m$ Bits gelten als kryptographisch sicher.

3.4 Fibonacci-Generatoren

Eine Abhilfe bei den Problemen mit den affinen Kongruenz–Generatoren schafft der so genannte **lagged Fibonacci Generator**. Der Name *lagged Fibonacci* bezieht sich auf die Rekursion zweier Terme der Form

$$x^i = x^{i-k} \diamond x^{i-l}, \; i = l+1, l+2, \ldots,$$

mit geeigneten *lags* $k, l \in \mathbb{N}$, $k \le l$. Dabei steht $\diamond$ für eine der zwei Operationen:

- Addition modulo 2^e
- Multiplikation modulo 2^e

Der Startwert solcher Generatoren sind Vektoren der Länge l, deren Einträge z.B. mit affinen Kongruenz Generatoren erzeugt werden können.

3.4.1 Additive lagged Fibonacci Generatoren

Die Periode eines additiven lagged Fibonacci Generators

$$x^i = x^{i-k} + x^{i-l} \mod m \tag{3.7}$$

für $m = 2^e$ ist $2^{e-1}(2^l - 1) \approx 2^{e+l-1}$.

Maximale Perioden werden erzeugt mit den Paaren $(k, l) = (5,2), (7,1), (7,3), (17,3), \ldots, (521, 32)$.

Beispiel 3.16 *`Rand17` ist ein additiver lagged Fibonacci Generator:*

$$x^i = x^{i-17} + x^{i-5} \mod 2^{32}$$

Er ist mit 17 Startwerte zu versehen, seine Periodenlänge ist $2.8 \cdot 10^{14}$. *Die Gütetests (siehe unten) weisen ihn als recht guten Generator aus.* □

3.4.2 Multiplikative lagged Fibonacci Generatoren

Der multiplikative lagged Fibonacci Generator

$$x^i = x^{i-k} \cdot x^{i-l} \mod m \tag{3.8}$$

hat für $m = 2^e$ die Periode $2^{e-3}(2^l - 1)$. Zwar ist diese Periode um den Faktor 4 kleiner als die der additiven lagged Fibonacci Generatoren, aber sie haben den Vorteil, dass eine bessere Streuung der Zufallszahlen erreicht wird.

Dieser Generator ist ein weiteres Beispiel eines nichtlinearen Generators, die Affinität der bisher betrachteten Generatoren ist ja nicht wirklich weit entfernt von der Linearität. Die Nichtlinearität ist offensichtlich, sie wird deutlich in der Abhängigkeit von den Startwerten.

3.5 Die Güte von Zufallszahlen

Wie soll man nun gute und weniger gute Generatoren auseinanderhalten? Ein Merkmal haben wir schon kennengelernt: „Ausschöpfung" des zur Verfügung stehenden Zahlraumes. Ferner gibt es **theoretische** und **empirische Tests** für die Güte von Generatoren. Theoretische Tests setzen am Generator selbst an, empirische Tests nehmen sich die erzeugten Zahlenfolgen vor. Kriterien, die untersucht werden, sind Gleichverteiltheit, Unkorreliertheit, Effizienz, d.h. Schnelligkeit. Auf die Schnelligkeit wollen wir aber nicht eingehen.

3.5.1 Geometrische Korrelation

Da die Abfolge der Pseudozufallszahlen durch den Generator festgelegt ist, besteht eine Korrelation zwischen aufeinander folgenden Zufallszahlen. Um Aufschluss über die Korreliertheit/Unkorreliertheit zu erhalten, sollte man daher Paare, Trippel,... von Zufallszahlen betrachten und deren „geometrische Verteilung" untersuchen.[6]

Wir betrachten nun Kongruenz-Generatoren und „skalieren" dazu die Zufallszahlen mit Modul m gemäß

$$u^i := \frac{x^i}{m} \in [0,1]\,,\; i \in \mathbb{N}_0\,.$$

Bei den linearen Kongruenz-Generatoren stellt man nun fest, dass diese Punkte den Raum nicht ausfüllen, sondern sich auf n-dimensionalen Hyperebenen aufreihen. Dieser Effekt kann sich als großes Problem erweisen, wenn man Zufallszahlen zur Erzeugung von Punkten im k-dimensionalen Raum einsetzt (wie etwa bei der Monte-Carlo-Methode).

Vergleichen wir die geometrische Verteilung der Paare (u^{i+1}, u^i) in $[0,1]\times[0,1]$ für die Generatoren aus Beispiel 3.12 und Beispiel 3.13. Man kann Geraden entdecken, worauf alle Zufallszahlen liegen, 21 im ersten Fall, 8 im zweiten Fall; die Streifen dazwischen sind frei von den erzeugten Zufallspaaren. Der maximale Abstand von solchen Streifen ist bei beiden Generatoren dementsprechend ziemlich verschieden: $\frac{1}{\sqrt{274}}$ bei Beispiel 3.12, $\frac{1}{\sqrt{32}}$ bei Beispiel 3.13. Dies bedeutet, dass der Generator 3.12 größeres Vertrauen genießen sollte.

Betrachten wir den Generator RANDU genauer. Wir haben hier

$$\begin{aligned} x^{i+1} &\equiv (65539)^2 x^{i-1} \mod 2^{31} \\ &\equiv (2^{16}+3)^2 x^{i-1} \mod 2^{31} \\ &\equiv (6x^i - 9x^{i-1}) \mod 2^{31} \end{aligned}$$

[6]Eine „Anwendung" der folgenden Pärchenbildung ist die Abtastung eines Schachbretts: will man jedes Feld zufällig besuchen, so braucht man zufällige Koordinaten.

und daher

$$x^{i+1} - 6x^i + 9x^{i-1} = k \cdot 2^{31} \quad \text{mit } k \in \mathbb{Z}\,.$$

Dies bedeutet für „normalisierten“ Zufallszahlen $u_i := x^i 2^{-31}$

$$u^{i+1} - 6u^i + 9u^{i-1} = k \quad \text{mit } k \in \mathbb{Z}\,. \tag{3.9}$$

Da $0 < u^{i+1} < 1$ gilt – beachte, dass daher $u^{i+1} - 6u^i + 9u^{i-1}$ nur die Werte $-5, -4, \ldots, 9$ annehmen kann – müssen die Tripel (u^{i+1}, u^i, u^{i-1}) auf Ebenen in $[0,1]^3$ liegen, und zwar auf nicht mehr als 15 Stück, die jeweils einen gleichen Abstand $1/\sqrt{118}$ haben. Neben der mangelnden Maximalität der Zykluslänge ein weiterer Nachteil dieses Generators. Von vielen wird er daher für den Müllhaufen der Informatik-Geschichte freigegeben.

Bemerkung 3.17 *Der derzeit „beste“ Generator ist wohl der Mersenne-Twister-19937-Generator, liefert er doch Zufallszahlen, die als Tupel selbst im* 623*-dimensionalen Raum noch gleichverteilt sind und der eine Periodenlänge von* $2^{19937} - 1 (\approx 4,3 \cdot 10^{6001})$ *besitzt; siehe [14]. Diese Periodenlänge erklärt auch den Namen des Algorithmus: Sie ist eine Mersenne-Primzahl und einige Eigenschaften des Algorithmus resultieren aus dieser Eigenschaft. Die geometrische Korreliertheit ist sehr gering; siehe oben.*

Der Generator gehört der Klasse der twisted generalized feedback shift register *an. Das Wort „Twisted“ bezieht sich auf eine bestimmte Transformation des Algorithmus, durch die diese hochgradige Gleichverteilung sichergestellt wird. Er wird in allen modernen Pogrammbibliotheken wie z.B. der* GNU Scienitic Library *eingesetzt.* □

3.5.2 Statistische Tests

Die erzeugten Zahlen können durch statistische Tests auf ihre Gleichverteilung untersucht werden. Tests, die in Gebrauch sind:

- Chi-Quadrat-Test
- Kolmogorov-Smirnov-Test
- Poker-Test
- Run-Test

Im Allgemeinen führen erst Kombinationen von Tests zu aussagekräftigen Ergebnissen. Wir besprechen unten den Chi-Quadrat-Test etwas genauer, zunächst machen wir ein paar Anmerkungen zu den übrigen genannten Möglichkeiten.

Eine gewisse Verfeinerung des Chi-Quadrat-Tests stellt der **Kolmogorov-Smirnov-Test** dar. Dies ist ein statistischer Test auf Übereinstimmung zweier Wahrscheinlichkeitsverteilungen. Mit seiner Hilfe kann geprüft werden, ob zwei Zufallsvariablen die gleiche Verteilung besitzen. Es liegt nahe, wie dies hier angewendet werden kann.

Ein weiteres Hilfsmittel der Statistik, Verteilungseigenschaften nachzuprüfen, sind so genannte **Run-Tests**. Hier wird etwa untersucht, ob die Anzahl und Länge gleichbleibender/aufsteigender Zahlenfolgen in Übereinstimmung mit der Gleichverteilung sind.

Der **Poker-Test** betrachtet Gruppen zu je 5 aufeinanderfolgenden Zahlen und beobachtet, welches der folgenden 7 Muster – manchmal betrachtet man nur 5 Fälle – mit dem Quintupel übereinstimmt.

Alle verschieden	*abcde*
Ein Paar	*aabcd*
Zwei Paare	*aabbc*
Drei Gleiche	*aaabc*
Full house	*aaabb*
Vier gleiche	*aaaab*
Fünf gleiche	*aaaaa*

Auf diese Anzahlen wird ein Chi-Quadrat-Test angewendet, um herauszufinden, ob die empirische Verteilungsfunktion in Übereinstimmung mit der Gleichverteilung ist.

3.5.3 Der Chi-Quadrat-Test

Der von Karl Pearson um 1900 herum entwickelte **Chi-Quadrat-Test (χ^2-Test)** ist eines der ältesten und mächtigsten Testverfahren der Statistik. In der einfachsten Form dient es der Prüfung der Verträglichkeit von beobachteten relativen Häufigkeiten – hier in einer Zufallsfolge – mit hypothetischen Wahrscheinlichkeiten. Er testet nach Zuordnung der erzeugten Zufallszahlen zu Kategorien, ob die Kategorie der einer gegebenen Verteilung, hier die Gleichverteilung, entsprechende Anzahl enthält. Wir teilen dazu die vorliegenden Zufallszahlen $x^i, i = 1, \dots, n$, in disjunkte Kategorien $K_j, j = 1, \dots, l$, die den Raum der möglichen Zufallszahlen ausschöpfen, ein, wobei das Eintreten der Kategorie K_j unter der Annahme der Gleichverteilung durch die Wahrscheinlichkeit p_j verknüpft sei; k_j sei die Anzahl der Zufallszahlen in der Kategorie K_j. Damit ist auch klar, dass

$$k_1 + \cdots + k_l = n$$

gilt. Diese Identität begründet auch, dass man (nur) von $l-1$ **Freiheitsgraden** in der „Testanordnung“ spricht.

Das Ziel ist ein Test der einfachen Hypothese

$$H_0 : p_j = k_j/n \text{ für jedes } j = 1, \dots, l$$

gegen die Alternative

$$H_1 : p_j \neq k_j/n \text{ für ein } j \in \{1, \dots, l\}\,.$$

Die auf Pearson zurückgehende Idee besteht nun darin, eine handhabare Testgröße anzugeben, die es gestattet, bei einer kritischen Größe die Hypothese H_0 (mit Recht) abzulehnen. Diese Testgröße ist

$$\chi^2 := \sum_{j=1}^{l} \frac{(k_j - np_j)^2}{np_j}\,;$$

man nennt sie die **χ^2-Statistik** mit $l-1$ Freiheitsgraden. Wie aus der Darstellung ersichtlich, erhalten „seltene Kategorien“ eine hohe Gewichtung und der χ^2-Wert ist umso größer, je stärker die Abweichung zwischen beobachteter und theoretischer Verteilung ist. Wenn er eine bestimmte Schranke c überschreitet, so verwirft man die Hypothese der Übereinstimmung beider Verteilungen. Wie üblich, gibt man sich ein Testniveau $\alpha > 0$ (zugelassene Wahrscheinlichkeit für einen Fehler erster Art) vor und passt dann die kritische Größe c an. Es ist nun der Vorteil des χ^2-Tests, dass c aus einer Tabelle in Abhängigkeit von der Anzahl der Freiheitsgrade (für große n) und α abgelesen werden kann; siehe etwa [9], Seite 259.

Beispiel 3.18 *Betrachte die Bitkette der Länge 50:*

$$10101\,00000\,01111\,01000\,10001\,01011\,00110\,01000\,10001\,00010\,.$$

Wir finden 19 1-Bits und 31 0-Bits. Da bei einer unterstellten Gleichverteilung für das Auftreten eines 1-Bits mit der Wahrscheinlichkeit $\frac{1}{2}$ zu rechnen ist, erhalten wir für den χ^2-Wert:

$$\chi^2 = \frac{(19-25)^2}{25} + \frac{(31-25)^2}{25} = \frac{36}{25} + \frac{36}{25} = 2.88\,.$$

Die Tabelle in [9] weist als kritischen Wert $c = 2.71$ für $\alpha := 0.1$ und $c = 3.84$ für $\alpha = 0.05$ aus (Freiheitsgrad $=1$). Damit lehnen wir die Hypothese zu H_0 im Fall $\alpha := 0.1$ ab und verwerfen H_0 im Fall $\alpha := 0.05$ nicht. □

3.6 Anhang: Beweis des Satzes zur Periodenlänge

Wir folgen bei den Überlegungen [5]. Zunächst Vorbereitungen.

Lemma 3.19 *Für $a \in \mathbb{Z}$ und $k \in \mathbb{N}_0$ setzen wir*

$$S_k(a) := 0\,,\ \text{falls } k = 0\,,\ S_k(a) := 1 + a + \cdots + a^{k-1}\,,\ \text{falls } k \geq 1\,. \tag{3.10}$$

Damit gilt (in $\mathbb{Z}$) :

$$S_{rk}(a) := S_r(a^k)S_k(a)\,,\ r \in \mathbb{N}\,. \tag{3.11}$$

Beweis:
Dies rechnet man mittels vollständiger Induktion so nach:
$r = 1 : S_{rk}(a) = S_k(a) = S_1(a^k)S_k(a)$, da $S_1(a^k) = 1$ ist.
$r + 1 :$

$$\begin{aligned} S_{(r+1)k}(a) &= S_{rk+k}(a) = S_{rk}(a) + a^{rk} + \cdots + a^{rk+k-1} \\ &= S_r(a^k)S_k(a) + a^{rk}S_k(a) = S_k(a)(S_r(a^k) + a^{rk}) = S_k(a)S_{r+1}(a^k)\,. \end{aligned}$$

■

Lemma 3.20 *Sei p eine Primzahl und a eine ganze Zahl mit $a = 1 \mod p$ bzw. $a = 1 \mod 4$, falls $p = 2$. Dann gilt für alle $n, k \in \mathbb{N}$:*

(a) $S_p(a^{p^n}) = 0 \mod p\,,\ S_p(a^{p^n}) \neq 0 \mod p^2\,.$

(b) $S_{p^n}(a) = 0 \mod p^n\,,\ S_{p^n}(a) \neq 0 \mod p^{n+1}\,.$

(c) $S_k(a) = 0 \mod p^n \iff p^n | k\,.$

Beweis:
Zu a). Wir führen den Beweis nur für $p > 2$.
Wir haben auf Grund der Voraussetzung $a = 1 + jp$ mit $j \in \mathbb{Z}$. Mit der Binomialformel erhält man $a^t - 1 = tjp + cp^2$ mit einer Konstante c. daher

$$S_p(a^{p^n}) - p = \sum_{k=0}^{p-1}(a^{kp^n} - 1) = p^n jp \frac{1}{2} p(p-1) + \tilde{c}p^2\,,$$

mit einer weiteren Konstanten $\tilde{c}$. Daraus liest man die Behauptung ab.
Zu b). Wir beweisen die Behauptung durch Induktion nach n.
$n = 1$. Der Fall $p = 2$ ist trivial. Sei also p ungerade. Nach Voraussetzung ist $a = 1 + jp$ mit einer ganzen Zahl j, also $a^k = 1 + kjp \mod p^2$, $k \in \mathbb{N}$. Dann ist

$$S_p(a) - p = \sum_{k=0}^{p-1}(a^k - 1) = \sum_{k=0}^{p-1} kjp = jp\frac{p(p-1)}{2} = 0 \mod p^2\,.$$

Damit ist der Induktionsbeginn schon klar.
$n+1$. Der Induktionsschluss ergibt sich aus der Formel

$$S_{p^{n+1}}(a) = S_p(a^{p^n})S_{p^n}(a)\,,$$

die sich aus (3.11) ergibt, und a).
Zu c).
Sei p^m die höchste Potenz von p mit $p^m|k$, also $k = p^m l$ mit $p \not| l$. Nach (3.11) ist

$$S_k(a) = S_l(a^{p^m})S_{p^m}(a)\,.$$

Es gilt $S_l(a^{p^m}) = l \mod p$, $S_l(a^{p^m}) \neq 0 \mod p$; dies verifiziert man wie oben. Daraus schließt man

$$p^n|S_k(a) \iff p^n|S_{p^m}(a)\,,$$

und mit Lemma 3.20

$$p^n|S_k(a) \iff n \leq m \text{ d.h. } p^n|S_k(a) \iff p^n|k\,;$$

beachte dabei a). ■

Bevor wir das Hauptergebnis über lineare Kongruenz–Pseudozufallsgeneratoren beweisen, noch ein wichtiges Resultat für das Rechnen in Kongruenzen, das man oft zur Vereinfachung von Argumentationen verwenden kann.

Satz 3.21 (Chinesischer Restsatz) *Ist $m = p_1^{k_1}\cdot\dots\cdot p_r^{k_r}$ die Primfaktorzerlegung von m, so ist $\mathbb{Z}_m$ isomorph zu $\mathbb{Z}_{p_1^{k_1}}\times\dots\times\mathbb{Z}_{p_r^{k_r}}$, d.h. es gibt eine bijektive Abbildung $g : \mathbb{Z}_m \longrightarrow \mathbb{Z}_{p_1^{k_1}}\times\dots\times\mathbb{Z}_{p_r^{k_r}}$, für die gilt:*

$$g([u]+[v]) = g([u]) + g([v])\,,\; g([u]\cdot[v]) = g([u])\cdot g([v])\,,\; u,v \in \mathbb{Z}\,. \tag{3.12}$$

Dabei wird sowohl Addition als auch Multiplikation in $\mathbb{Z}_{p_1^{k_1}}\times\dots\times\mathbb{Z}_{p_r^{k_r}}$ komponentenweise erklärt.

Beweis:
Offenbar haben $\mathbb{Z}_m$ und $\mathbb{Z}_{p_1^{k_1}}\times\dots\times\mathbb{Z}_{p_r^{k_r}}$ gleich viele Elemente. Also genügt es zeigen, dass g injektiv ist. Dazu reicht nun zu zeigen, dass aus $g([u]) = ([0],\dots,[0])$ folgt: $[u] = [0]$.
Aus $g([u]) = ([0],\dots,[0])$ folgt $p_i^{k_i}|u, i = 1,\dots,r$. Dann folgt aber $m = p_1^{k_1}\cdots p_r^{k_r}|u$. Also $[u] = [0]$. ■

Satz 3.22 *Mit $m,a,b \in \mathbb{Z}$, $m \geq 2$ betrachte die Abbildung*

$$f : \{0,\dots,m-1\} \ni x \longmapsto ax+b \mod m \in \{0,\dots,m-1\}\,. \tag{3.13}$$

Für beliebiges $x^0 \in \{0,\dots,m-1\}$ sei die Folge $(x^n)_{n\in\mathbb{N}}$ definiert durch

$$x^{n+1} := f(x^n)\,,\; n \in \mathbb{N}_0\,.$$

Genau dann ist diese Folge periodisch mit der maximalen Periodenlänge m, wenn folgende Bedingungen erfüllt sind:

a) *b und m sind teilerfremd.*

b) *$p|(a-1)$ für alle Primteiler p von m;*

c) *$4|(a-1)$ falls $4|m$;*

Beweis:
Ist $m = p_1^{k_1} \cdot \dots \cdot p_r^{k_r}$ die Primfaktorzerlegung von m, so hat man nach dem chinesischen Restsatz, dass $\mathbb{Z}_m$ isomorph zu $\mathbb{Z}_{p_1^{k_1}} \times \dots \times \mathbb{Z}_{p_r^{k_r}}$ vermöge der Abbildung g in Satz 3.21 mit den Eigenschaften aus (3.12) ist. Es ist daher leicht einzusehen, dass es genügt, den Satz für den Fall zu beweisen, dass m eine Primzahlpotenz p^k ist. Vorüberlegung: Es gilt

$$x^{i+1} - x^i = f(x^i) - f(x^{i-1}) = a(x^i - x^{i-1}) \, , i = 1, 2, \dots \, ,$$

und daher

$$x^n - x^0 = \sum_{i=0}^{n-1} a^i (x^1 - x^0) = S_n(a)(x^1 - x^0) \, , \, x^n - x^0 = S_n(a)(x^1 - x^0) \mod m \, , \, n \in \mathbb{N} \, . \quad (3.14)$$

Wir zeigen nun die Notwendigkeit der Bedingungen. Es sei also vorausgesetzt, dass f einen Zyklus der maximalen Länge m erzeugt. Beachte: $m = p^k$.
Falls p nicht $a-1$ teilt, sind $m, a-1$ teilerfremd und die Gleichung $(a-1)x = b$ ist in $\mathbb{Z}_m$ lösbar, d.h. f besitzt einen Fixpunkt. Dann kann aber kein Zyklus der Länge m existieren. b) ist damit gezeigt.
Es teile 4 die Zahl m, also $m = 2^k$ mit $k \geq 2$. Nach b) wissen wir schon, dass $2|(a-1)$ gilt; a ist also ungerade. Es muß also noch gezeigt werden, dass der Fall $a = 3 \mod 4$ nicht auftreten kann. Wäre dies doch der Fall, so würde $S_2(a) = 1 + a = 0 \mod 4$ sein und daher mit $S_{2i}(a) = S_i(a^2)S_2(a)$ (siehe Lemma 3.20) schließlich $S_{2i}(a) = 0 \mod 4$ für alle i gelten. Mit (3.14) folgt daraus $x^{2i} = x^0 \mod 4$ und $x^{2i+1} = x^1 \mod 4$ für alle i; es könnte also keinen Zyklus maximaler Länge m geben. b) ist damit gezeigt.
In einem Zyklus der Länge m muß insbesondere das Element 0 vorkommen; wir dürfen daher o. E. $x^0 = 0$ und daher $x^1 = b$ voraussetzen. Dann sagt (3.14), dass $x^n = S_n(a)b$ gilt. Ist b nicht invertierbar mod m, kann das Element 1 niemals im Zyklus auftreten. Damit ist auch a) klar.
Wir zeigen nun die Hinlänglichkeit der angeführten Bedingungen. Beachte $m = p^k$.
Ist $m = 2$, dann ist $b = 1$ wegen a) und ein Zyklus der Länge 2 existiert. Im Fall $p = 2$ dürfen wir also $4|m$ annehmen und die Voraussetzungen von Lemma 3.20 sind erfüllt. Es genügt zu zeigen, dass für $x^0 = 0$ ein Zyklus der Länge m erzeugt wird. Aus $x^0 = 0$ folgt mit (3.14), dass $x^n = S_n(a)b \mod m$ gilt. Da b invertierbar mod m ist, ist dann $x^n = x^0 = 0$ gleichbedeutend mit $S_n(a) = 0 \mod m$. Aus Lemma 3.20 c) folgt $m|n$ und damit die Behauptung. ■

3.7 Bibliographische Anmerkungen

Die Grundlagen modularer Arithmetik kann man etwa finden bei J. Wolfart und O. Forster (siehe [5, 17]). Den Beginn mit den linearen Kongruenz Generatoren hat Lehmer gemacht; siehe [12]. Einen Überblick über historische Kuriositäten geben [2, 6]; siehe auch [8] und [16]. Eine Monographie zur Erzeugung von Zufallszahlen und zu Monte Carlo Methoden ist [7, 15]. Dort findet man auch etwas zu den so genannten „Feedback shift register generators“. In [3, 4] werden „nichtlineare Generatoren“ behandelt. Ein sehr viel tiefliegender Standpunkt bezüglich nichtlinearer Generatoren wird in [1] eingenommen. Die feinsten Untersuchungen zur Güte findet man in vielen Arbeiten von G. Marsaglia. Wir geben einen Hinweis auf die Arbeit, in der die Korrelation von Zufallszahlen erstmals genauer untersucht wurde ([13]). Statistische Tests sind unter anderem beschrieben in [10, 11].

3.8 Übungen

1.) Überprüfe, ob der affine Generator

$$x^{i+1} := ax^i + b \mod m$$

mit $a = 3432, b = 6789, m = 9973$ eine maximale Periode hat.

2.) Überprüfe, ob der affine Generator

$$x^{i+1} := ax^i + b \mod m$$

mit $a = 1229, b = 1, m = 2048$ eine maximale Periode hat.

3.) Betrachte den affinen Generator

$$x^{i+1} := ax^i + b \mod m$$

mit $a = 7, b = 6, m = 64$. Zeige, dass es Startwerte x^0 gibt, zu denen Orbits mit Periodenlänge $1, 2$ und 8 gehören.

4.) Betrachte den affinen Generator

$$x^{i+1} := ax^i + b \mod m\,.$$

Zeige für $n, k \in \mathbb{N}_0$:

(a) $x^n = (a^n x^0 + \frac{a^n - 1}{a - 1} b) \mod m$

(b) $x^{n+k} = (a^k x^n + \frac{a^k - 1}{a - 1} b) \mod m$

5.) Überprüfe, ob der affine Generator

$$x^{i+1} := ax^i + b \mod m$$

mit $a = 1711, b = 0, m = 30269$ eine maximale Periode hat.

6.) Betrachte einen affinen Generator

$$x^{i+1} := ax^i + b \mod m\,.$$

Sei m_1 ein Teiler von m.

(a) Setze $y^i := x^i \mod m_1$, $i \in \mathbb{N}$. Zeige:

$$y^{i+1} = ay^i + b \mod m_1\,.$$

(b) Was lässt sich mit Hilfe von a) aus Satz 3.22 zur Periode des Generators schließen?

7.) Seien Zufallszahlen x^i mit dem affinen Kongruenz Generator mit $a = 2, b = 0, m = 11$ erzeugt. Setze $u_i := x^i/m$. Auf wie vielen Geraden in $[0,1]^2$ liegen die Tupel (u_i, u_{i+1})?

8.) Seien Zufallszahlen x^i mit dem affinen Kongruenz Generator mit $a = 1229, b = 1, m = 2048$ erzeugt. Setze $u_i := x^i/m$. Auf wie vielen Geraden in $[0,1]^2$ liegen die Tupel (u_i, u_{i+1})?

9.) Betrachte die folgendermaßen erzeugten Zahlen:

$$x^n := \texttt{fractional part}(n\phi)\,,\ n \in \mathbb{N}_0,\ \text{mit } \phi : \frac{1}{2}(1 + \sqrt{5})\,.$$

(a) Zeige: $x^{i+1} = \texttt{reminder}(x^i + 1)$, $i \in \mathbb{N}_0$. Teste den Generator, indem die Anzahl von 10 000 erzeugten Zufallszahlen in $[0.45, 0.50]$ ermittelt werden.

10.) Betrachte die folgende Variante eines Fibonacci-Generators:

$$x^{i+1} := \begin{cases} x^{i-17} - x^{i-5} \text{ falls } x^{i-17} - x^{i-5} \geq 0 & x^{i-17} - x^{i-5} + 1 \text{ , sonst} \end{cases} .$$

Dabei sind $x^0, \ldots, x^{17}$ in $[0,1]$ „zufällig" vorgegeben.
Implementiere diesen Generator und plotte 10 000 Paare (x^i, x^{i-1}) in $[0,1]^2$.

11.) Es werden Pseudo-Zufallszahlen in $[0,1]$ erzeugt, von denen man unterstellt, dass sie gleichverteilt sind. Man hat folgende „Statistik":

Intervall	$[0.0, 0.1]$	$[0.1, 0.2]$	$[0.2, 0.3]$	$[0.3, 0.4]$	$[0.4, 0.5]$
Häufigkeit	1009	1017	996	1006	999

Intervall	$[0.5, 0.6]$	$[0.6, 0.7]$	$[0.7, 0.8]$	$[0.8, 0.9]$	$[0.9, 1.0]$
Häufigkeit	1007	985	971	992	10018

Führe den χ^2-Quadrattest durch. Was spricht für die unterstellte Gleichverteilung?

12.) Berechne mit dem affinen Kongruenz Generator mit $a = 950706376, b = 0, m = 2^{31} - 1$, mit dem Startwert $x^0 = 1$ 20 Zufallszahlen.

13.) Betrachte einen affinen Generator

$$x^{i+1} := 3x^i + 2^{16} \mod m := 2^{31} .$$

Berechne damit die Zufallszahlen $u^i := x^i/m \in [0,1), i = 0, \ldots, N := 2000$ und plotte die Punktpaare $(u^{i+1}, u^i), i = 0, \ldots, N$, und $(u^{i+1}, u^i, u^{i-1}), i = 1, \ldots, N$,

14.) Zur Berechnung der Quadratzahlen kann die Beziehung

$$(n+1)^2 = n^2 + 2 \cdot n + 1 \ \ (n \in \mathbb{N})$$

verwendet werden. Wieso?

Literaturverzeichnis

[1] L. Blum, M. Blum and M. Shub. A simple unpredictable pseudo-random number generator. *SIAM J. on Computing*, 15:364–383.

[2] U. Dieter. Erzeugung von gleichverteilten Zufallszahlen. *Jahrbuch Überblicke Mathematik*, pages 25–44, 1993.

[3] J. Eichenauer-Herrmann. Inversive congruential pseudorandom numbers: A tutorial. *Int. Stat. Rev.*, 60:167–176, 1992.

[4] J. Eichenauer-Herrmann. Pseudorandom number generation by nonlinear methods. *Int. Stat. Rev.*, 63:247–255, 1995.

[5] O. Forster. *Algorithmische Zahlentheorie*. Vieweg, 1996.

[6] B. Gaertner. Ein Reinfall mit Computer-Zufalsszahlen. *DMV-Mitteilungen*, Ausgabe 2:55–60, 1999.

[7] J.E. Gentle. *Random number generation and Monte Carlo methods*. Springer, 2001.

[8] P. Hellekalek. Good random number generators are (not so) easy to find. *Mathematics and Computers in Simulation*, 46:485–505, 1998.

[9] N. Henze. *Stochastik für Einsteiger*. Vieweg, Braunschweig, 1997.

[10] D.E. Knuth. *The Art of Computer Programming, Vol. 2: Seminumerical Algorithms*. Addison-Wesley, Reading, 1998.

[11] H. Leeb and S. Wegenkittl. Inversive and linear congruential pseudorandom number generators in empirical tests. *ACM Transactions on Modelling and Computer Simulation*, 7:272–286, 1997.

[12] D.H. Lehmer. Mathematical methods in large-scale computing units. xx:141–146, 1951.

[13] G. Marsaglia. Random numbers fall mainly in the planes. *Proc. Nat. Acad. Sci.*, 61:1–25, 1968.

[14] M. Matsumoto and T. Nishimura. Mersenne twister: A 623-dimensionally equidistributed uniform pseudorandom number generator. *ACM Trans. on Modeling and Computer Simulations*, 8:3–30, 1998.

[15] H. Niederreiter. *Random Number Generation and Quasi-Monte-Carlo-Methods*. SIAM, Philadelphia, 1992.

[16] S.K. Park and K.W. Miller. Random number generators: good ones are hard to find. *Communications of the ACM*, 31:1192–1201, 1988.

[17] J. Wolfart. *Einführung in die Zahlentheorie und Algebra*. Vieweg, 1996.

Kapitel 4

Simulationsgrundlagen

Simulation geht aus von einer modellhaften Nachbildung eines realen Vorgangs und besteht darin, Experimente entlang dieses Modells durchzuführen und Informationen über mögliche Aktionen und Ereignisse zu beobachten und auszuwerten. Der grundsätzliche Ansatz bei Simulationsverfahren für die Berechnung von Optionspreisen ist, einen diskontierten Erwartungswert der Auszahlungsfunktion zu bestimmen. Bei der Erledigung dieser Aufgabe fallen unterschiedliche Fragestellungen hinsichtlich der Nutzung von Zufallszahlen an: Normalverteilte Zufallszahlen, Generierung von Zufallszahlen mit speziellen Verteilungen, Monte Carlo-Verfahren. Insbesondere die mehrdimensionale Integration ist ein wichtiges Hilfsmittel für Aufgabenstellungen der Finanzmathematik.

4.1 Erzeugung von Verteilungen: Die Inversionsmethode

Hier unterstellen wir, dass ein Mechanismus vorliegt, mit dem man beliebig viele unabhängige, auf dem Intervall $[0,1]$ gleichmäßig verteilte Zufallszahlen erzeugen kann; siehe Kapitel 3. Sie kommen zustande als Ergebnis einer unabhängigen Wiederholung eines Zufallsexperiments, dessen Ausgang durch eine Zufallsgröße beschreiben wird, die auf $[0,1]$ gleichverteilt ist. Unabhängigkeit ist also die der Zufallsgrößen. Beachte, dass dies eine idealisierte Annahme ist, denn die Probleme mit den Pseudozufallszahlen sind uns wohlbekannt.

Definition 4.1 *Zwei Zufallsgrößen $X, Y : \Omega \longrightarrow \mathbb{R}$ auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$ heißen* **unabhängig***, wenn gilt:*

$$P(\{X = x, Y = y\}) = P(\{X = x\})P(\{Y = y\}) \text{ für alle } x, y \in \mathbb{R}\,.$$

□

4.1.1 Die Inversionsmethode

Die Inversionsmethode ist ein Simulationsverfahren, um aus gleichverteilten Zufallszahlen andere Wahrscheinlichkeitsverteilungen zu erzeugen. Sie basiert auf der Tatsache, dass man mit der Inversen einer gegebenen Verteilungsfunktion ganz einfach einen Zusammenhang zwischen der Gleichverteilung und der vorgegebenen Verteilung herstellen kann.

Sei $F : \mathbb{R} \longrightarrow \mathbb{R}$ eine Verteilungsfunktion einer Zufallsgröße X auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$. Auf Grund der Definition der Verteilungsfunktion $(F(x) = P(\{X \leq x\})\,,\, x \in \mathbb{R})$ ist F rechtsstetig, d.h. $\lim_{x \downarrow x_0} F(x) = F(x_0)$ für alle $x_0 \in \mathbb{R}$. Beachte, F muss nicht notwendigerweise stetig und strikt monoton wachsend sein. Daher lässt sich auch nicht sofort eine Inverse zu F angegeben. Wir definieren die **verallgemeinerte Umkehrfunktion** als

$$F^{-1}(y) := \inf\{u \in \mathbb{R} | F(u) \geq y\}\,,\; y \in \mathbb{R}\,;$$

dabei wird $F^{-1}(1)$ gleich ∞ gesetzt, falls $\{u \in \mathbb{R} | F(u) \geq 1\} = \emptyset$ ist. F^{-1} wird oft auch als **Quantilfunktion** bezeichnet.

Satz 4.2 *Sei $F : \mathbb{R} \longrightarrow \mathbb{R}$ eine Verteilungsfunktion einer Zufallsgröße X auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$. Sei U eine auf $[0,1]$ gleichverteilte Zufallsgröße, also $U \sim \mathcal{U}[0,1]$. Dann ist $Y := F^{-1} \circ U$ eine reelle Zufallsgröße, die F als Verteilungsfunktion besitzt.*

Beweis:
Wir haben für $z \in [0,1], x \in \mathbb{R}$ die Äquivalenz

(a) $(F^{-1} \circ U)(z) \leq x$

(b) $U(z) \leq F(x)$

denn:
$(a) \implies (b)$ Sei $\varepsilon > 0$. Dazu gibt es dann $u \in \mathbb{R}$ mit $u \leq x + \varepsilon$ und $F(x+\varepsilon) \geq F(u) \geq U(z)$. Aus der Rechtsstetigkeit von F folgt, da $\varepsilon > 0$ beliebig war, $F(x) \geq U(z)$.
$(b) \implies (a)$ Folgt, da $x \in \{u \in \mathbb{R} | F(u) \geq U(z)\}$.
Damit gilt für $x \in \mathbb{R}$:

$$P(\{Y \leq x\}) = P(\{F^{-1} \circ U \leq x\}) = P(\{U \leq F(x)\}) = F(x)\,.$$

■

Bemerkung 4.3 *Das wiederholte Ziehen von Zufallszahlen $x^1, x^2, \ldots$ kann auf zwei Arten stochastisch interpretiert werden: Zum einen kann man die x^i als verschiedene unabhängige Realisierungen ein und derselben Zufallsgröße X interpretieren, also $x^i = X(\omega_i)$, zum anderen kann man die x^i als Realisierungen unabhängiger und identisch verteilter Zufallsgrößen $X_i \sim X$, also $x^i = X_i(\omega)$ auffassen. Beide Sichtweisen sind äquivalent und wir verwenden beide Interpretation je nach Praktikabilität.* □

Es ist nun klar, wie die Erzeugung einer Zufallsfolge, die nach einer gegebenen Verteilungsfunktion verteilt ist, geschehen kann: Nehme eine Zufallszahl u in $[0,1]$ und setze $x := F^{-1}(u)$, oder anders ausgedrückt: löse die Gleichung $F(x) = u$. Das Problem bei diesem Vorgehen ist, dass in vielen Fällen die Inverse F^{-1} mit vertretbarem Aufwand nicht berechenbar ist.

Algorithm 4.1 Zufallszahlen vorgegebener kontinuierlicher Verteilung

EIN Verteilungsfunktion F. Mechanismus zur Erzeugung von gleichmäßig verteilten Zufallszahlen.

Schritt 1 Erzeuge die verallgemeinerte Inverse F^{-1}.

Schritt 2 Für $k = 1, \ldots, N$ erzeuge eine Zufallszahl u^k und berechne $x^k := F^{-1}(u^k)$.

AUS N Zufallzahlen $x^1, \ldots, x^N$, die nach F verteilt sind.

Aus Satz 4.2 wissen wir, dass die Aussage in AUS korrekt ist.

Beispiel 4.4 *Sei* $F(x) := (1 - e^{-\lambda x})\chi_{[0,\infty)}, x \in \mathbb{R}\,; \lambda > 0$. *Wenn wir den obigen Algorithmus anwenden, erzeugen wir so genannte* **exponentialverteilte Zufallszahlen**. *Wir haben*

$$F^{-1}(y) = -\frac{1}{\lambda}\ln(1-y)\,,\ y \in (0,1]\,.$$

Weil die Zufallszahl $1-y$ *genau wie* y *im Intervall* $[0,1]$ *gleichverteilt ist, folgt, dass die Zufallsfolge* $x^1, \dots, x^N$, *der Wahrscheinlichkeitsdichte* $f(x,\lambda) := \lambda e^{-\lambda x}\chi_{[0,\infty)}(x)$ *folgt.* □

Der Algorithmus 4.1.1 ließe sich grundsätzlich auch zur Erzeugung normal-verteilter Zufallszahlen anwenden. Dazu wäre dann die Inverse von $\mathcal{N}$ zu berechnen. Dies ist aber kein leichtes Unterfangen, insbesondere an den „Rändern“ des Wertebereichs von $\mathcal{N}$. Eine naive numerische Berechnung der Inversen ist im Allgemeinen auch nicht sehr stabil, insbesondere nicht, wenn $u \approx 1$, $u = F(x)$. Dazu gibt es tiefliegende Überlegungen, die mit effizienten Approximationen von $\mathcal{N}^{-1}$ zu tun haben. Wir verfolgen dies nicht weiter (siehe [1, 25]), sondern betrachten in den folgenden Abschnitten Methoden auf anderer Basis.

4.1.2 Diskrete Verteilung

Soll eine Zufallsgröße X simuliert werden, die nur die endlich vielen Werte $x^1, \dots, x^s \in \mathbb{R}$ mit Wahrscheinlichkeit $p_1, \dots, p_s$ annimmt, wenden wir die obige Idee etwas abgewandelt an.

Es ist eine Zufallsgröße Z auf einem geeigneten Wahrscheinlichkeitsraum (Ω, Σ, P) zu finden mit

$$p_i = P(\{Z = x^i\})\,,\ i = 1, \dots, s\,.$$

Wir wählen dazu den Wahrscheinlichkeitsraum als $([0,1], \mathcal{B}([0,1]), \lambda)$. Dabei ist $\mathcal{B}([0,1])$ die Borel-Sigma–Algebra von $[0,1]$ und λ das Lebesguemaß auf $[0,1]$. Nun teilen wir das Intervall $[0,1]$ in s Teilintervalle $I_1, \dots, I_s$ auf mit

$$I_i := [p_1 + \cdots + p_{i-1}, p_1 + \cdots + p_{i-1} + p_i)\,,\ i = 1, \dots, k-1\,,\ I_s := [p_1 + \cdots + p_{s-1}, 1]\,.$$

Damit definieren wir nun die Zufallsgröße Z folgendermaßen:

$$Z(y) := i\,,\ \text{falls } y \in I_i\,.$$

Da für $i = 1, \dots, s$

$$\begin{aligned} P(\{Z = i\}) &= \lambda(\{\omega \in [0,1] | Z(\omega) = i\}) \\ &= \lambda(\{\omega \in [0,1] | \omega \in I_i\}) \\ &= \lambda(I_i) = p_i \end{aligned}$$

ist das Konstruktionsproblem gelöst.

Damit können wir nun auch Zufallszahlen konstruieren, die nach der vorgegebenen Verteilung verteilt sind. Wir verwenden die gerade eingeführten Bezeichnungen.

Algorithm 4.2 Zufallszahlen vorgegebener diskreter Verteilung

EIN Verteilungsparameter $p_1, \dots, p_s$. Mechanismus zur Erzeugung von gleichmäßig verteilten Zufallszahlen.

Schritt 1 Für $k = 1, \dots, N$ erzeuge eine Zufallszahl u^k und setze $z_k := i$ falls $u^k \in I_i$.

AUS N „Zufallszahlen“ $z_1, \dots, z_k$, die diskret nach $p_1, \dots, p_s$ verteilt sind.

Eine diskrete Zufallsgröße, die eine Verteilung besitzt, bei der abzählbare viele Werte mit positiver Wahrscheinlichkeit angenommen werden, kann vollkommen analog simuliert werden.

4.1.3 Anwendung: Normalverteilung

Die Normalverteilung mit Erwartungswert μ und Standardabweichung σ, die wir dann mit $\mathcal{N}(\mu, \sigma^2)$ bezeichnen, hat die Dichte

$$f(x) := \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/(2\sigma^2)}\,,\ x \in \mathbb{R}\,.$$

Ist die Zufallsgröße Z normalverteilt mit Erwartungswert 0 und Standardabweichung 1, so nennen wir sie **Standard-normalverteilt** und schreiben $Z \sim \mathcal{N}(0,1)$. Ist $Z \sim \mathcal{N}(0,1)$, dann ist die Zufallsgröße $X := \sigma Z + \mu$ normalverteilt mit Erwartungswert μ und Standardabweichung σ, also $X \sim \mathcal{N}(\mu, \sigma^2)$. Es ist daher ausreichend, die Standard-Normalverteilung zu realisieren.

Die Normalverteilung nimmt eine Sonderstellung unter den Verteilungen ein. Dies hängt mit den vielfältigen Anwendungen und, damit einhergehend, der Gültigkeit des **Zentralen Grenzwertsatzes** zusammen.

Satz 4.5 (Zentraler Grenzwertsatz) *Sei $(X_i)_{i\in\mathbb{N}}$ eine Folge von identisch verteilten, unabhängigen Zufallsgrößen auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$ mit $\mathbb{E}(X_i) = \mu, \mathbb{V}(X_i) = \sigma^2$. Für die Mittelwertbildung $\overline{X}_n := \frac{1}{n}(X_1 + \cdots + X_n)$ gilt dann:*

$$\lim_n P(\{\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \le x\}) = \mathcal{N}(x) \ \ \textit{für jedes } x \in \mathbb{R}\,. \tag{4.1}$$

Beweis:
Siehe etwa [16].

Bemerkung 4.6 *Existiert das dritte zentrierte Moment $\mathbb{E}((X_1-\mu)^3)$ und ist es endlich, dann ist die Konvergenz in Satz 4.5 sogar gleichmäßig und die Konvergenzgeschwindigkeit ist wenigstens von der Ordnung $1/\sqrt{n}$ (Satz von Berry-Esseen).* □

Es gibt ein einfaches Verfahren, aus gleichverteilten Zufallsgrößen eine Zufallsgröße X zu konstruieren, die eine passable/gute Approximation einer Verteilung $Z \sim \mathcal{N}$ ist.

Seien $X_1, \dots, X_n$ gleichverteilte, unabhängige Zufallsgrößen auf dem Intervall $[0,1]$. Wir wissen, dass der Erwartungswert einer gleichmäßig verteilten Zufallsgröße gleich 1/2 ist. Die Varianz berechnet sich nach

$$\mathbb{V}(X_i) = \mathbb{E}(X_i^2) - \mathbb{E}(X_i)^2 = \int_0^1 X_i^2 dP(u) - \frac{1}{4} = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}\,.$$

Wir setzen

$$Y := \sqrt{\frac{12}{n}}(X_1 + \cdots + X_n - n/2)\,.$$

Es gilt

$$\mathbb{E}(Y) = \sqrt{\frac{12}{n}}(\sum_{i=1}^n \mathbb{E}(X_i) - n/2) = \sqrt{\frac{12}{n}}(\sum_{i=1}^n \frac{1}{2} - n/2) = 0$$

und in analoger Rechnung

$$\mathbb{V}(Y) = \frac{12}{n}\mathbb{V}\left(\sum_{i=1}^n X_i - n/2\right) = 1\,.$$

(Beachte $\mathbb{V}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \mathbb{V}(X_i)$ auf Grund der Tatsache, dass $X_1, \dots, X_n$ als unabhängig angenommen wurden.) Der zentrale Grenzwertsatz besagt, dass Y eine Standard-Normalverteilung approximiert. In der Praxis wählt man $n = 12$.

Die beschriebene Methode hat u.a. den Nachteil, dass für eine „normalverteilte Zufallszahl“ 12 gleichmäßig verteilte Zufallszahlen benötigt werden.

4.1.4 Ein Transformationssatz für Dichten

Hier lassen wir auch vektorwertige Zufallsgrößen zu. Dies ist notwendig, um die Transformation von Verteilungen gewinnbringend und allgemein einsetzen zu können.

Sei $X : \Omega \longrightarrow \mathbb{R}^d$ mit dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$. Betrachte dazu noch einen Zufallsvektor $G : \mathbb{R}^d \longrightarrow \mathbb{R}^d$, wobei auf $\mathbb{R}^d$ die Sigmaalgebra $\mathcal{B}_d$ der Borelschen Mengen mit dem Lebesguemaß λ_d zugrundeliege. Auf $\mathbb{R}^d$ haben wir dann das Wahrscheinlichkeitsmaß P^X, das vermöge X auf $\mathbb{R}^d$ so definiert ist:

$$P^X(A) := P(\{X \in A\}) \, , \, A \in \mathcal{B}_d \, .$$

Ebenso haben wir das Wahrscheinlichkeitsmaß P^G, das vermöge G auf $\mathbb{R}^d$ so definiert ist:

$$P^G(B) := P^X(\{G \in B\}) \, , \, B \in \mathcal{B}_d \, .$$

Damit können wir auch

$$(\mathbb{R}^d, \mathcal{B}_d, P^X) \xrightarrow{G} (\mathbb{R}^d, \mathcal{B}_d, P^G)$$

betrachten. Zusätzlich steht uns noch die Zufallsgröße

$$(\Omega, \Sigma, P) \overset{Y:=G\circ X}{\longrightarrow} (\mathbb{R}^d, \mathcal{B}_d, P^Y)$$

mit der Verteilung P^Y zur Verfügung. Wir halten fest:

Regel 4.7 *Es gilt:* $P^Y = P^G$, *d.h.* $P(\{Y \in B\}) = P^G(B) \, , \, B \in \mathcal{B}_d \, .$

Wenn also G und die Verteilung P^X bekannt ist, kann man die Verteilung von $Y = G \circ X$ dadurch berechnen, dass man P^G bestimmt. Wir gehen dieser Frage nach für den Fall, dass X eine Dichte f besitzt:

$$P^X(A) = \int_{\mathbb{R}^d} \chi_A(x) f(x) dx \, , \, A \in \mathcal{B}_d \, .$$

Damit gilt dann

$$P^G(B) = P^X(\{G \in B\}) = \int_{\mathbb{R}^d} \chi_{\{G\in B\}}(x) f(x) dx \, , \, B \in \mathcal{B}_d \, .$$

Aus der Analysis wissen wir, dass Substitutionssätze den Sachverhalt von Koordinatenwechseln bei der Integration behandeln.

Satz 4.8 *Sei X eine Zufallsgröße auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$ mit Verteilung P^X und Dichte f. Ferner sei $G : \mathbb{R}^d \longrightarrow \mathbb{R}^d$ ein Zufallsvektor, wobei auf $\mathbb{R}^d$ die Sigmaalgebra $\mathcal{B}_d$ der Borelschen Mengen mit dem Lebesguemaß λ_d zugrundeliege. Sei $M \subset \mathbb{R}^d$ eine offene, zusammenhängende Menge, so dass die folgenden Bedingungen gelten:*

(a) *Für $x \notin M$ ist $f(x) = 0$.*

(b) *G ist stetig differenzierbar auf M und es ist $\det DG(x) \neq 0$ für alle $x \in M$.*

(c) *Ist $N = G(M)$ das Bild von M unter der Abbildung G, so ist die Abbildung $G : M \longrightarrow N$ bijekiv mit der Umkehrabblidung H.*

Dann besitzt die Verteilung P^G die Dichte

$$g(y) := \begin{cases} f(H(y)) |\det DH(y)| & , \textit{falls } y \in N \\ 0 & , \textit{sonst} \end{cases} \tag{4.2}$$

Beweis:
Den Beweis führen wir im Anhang 4.9 an. ■

Mit diesem Satz kommt man dem Vorhaben, gleichmäßig verteilte Zufallszahlen in normalverteilte zu transformieren, näher. Mit

$$n = 1, f \equiv 1, g := \mathcal{N}'$$

bleibt ein $h : \mathbb{R} \longrightarrow \mathbb{R}$ – h steht für H^{-1} – zu finden, so dass

$$|\frac{dh}{dy}| = g \tag{4.3}$$

gilt. Nun stellen wir aber fest, dass dies eine Differentialgleichung ist für h, die ohne analytische Lösungsmöglichkeit ist.

4.2 Das Box-Muller-Verfahren

Der Ausweg aus dem Dilemma, dass (4.3) nicht analytisch lösbar ist, ist es, eine Transformation nach $\mathbb{R}^2$ zu verwenden, d.h. die Fragestellung in eine „größere einzubetten".

Betrachte dazu die Transformation

$$y = G(x)\,,\ (y_1, y_2) = (\sqrt{-2\ln(x_1)}\cos(2\pi x_2), \sqrt{-2\ln(x_1)}\sin(2\pi x_2))\,,\ (x_1, x_2) \in M := (0,1)^2\,.$$

Auflösen nach x_1, x_2 ergibt wegen $y_1^2 + y_2^2 = -2\ln(x_1), y_2/y_1 = \tan(2\pi x_2)$

$$x_1 = \exp(-|y|^2/2)\,,\ x_2 = \frac{1}{2\pi}\arctan(y_2/y_1)\,,$$

wobei wir den Hauptzweig der Arcus-Tangensfunktion genommen haben. Also erhalten wir als Umkehrabbildung $H := G^{-1}$

$$H(y) = (\exp(-|y|^2/2), \frac{1}{2\pi}\arctan(y_2/y_1))\,,\ (y_1, y_2) \in \mathbb{R}^2\,.$$

Bemerkung 4.9 *Betrachtet man die Koordinaten x_1, x_2 als gleichverteilte Zufallsgrößen auf $[0,1]$, so wird die Variable R^2 eine exponentialverteilte Zufallsgröße auf $[0,\infty)$ mit Erwartungswert 2, denn*

$$P(\{R^2 \le x\}) = 1 - e^{-x/2}\,,\ x \ge 0\,.$$

Ist dann R gegeben, dann sind die Punkte (Y_1, Y_2) gleichverteilt auf dem Kreis mit Radius R. □

Satz 4.8 zufolge hat $Y := G \circ X$ die Dichte $g = |\det DH|$. Wegen

$$\det(DH(y_1, y_2)) = \det\begin{pmatrix} -y_1\exp(-|y|^2/2) & -y_2\exp(-|y|^2/2) \\ -\frac{1}{2\pi}\frac{y_2}{y_1^2+y_2^2} & \frac{1}{2\pi}\frac{y_1}{y_1^2+y_2^2} \end{pmatrix} = -\frac{1}{2\pi}\exp(-|y|^2/2)$$

erhalten wir

$$g(y) = \frac{1}{2\pi}\exp(-\frac{1}{2}(y_1^2+y_2^2)) = \frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}y_1^2)\frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}y_2^2)\,,\ (y_1, y_2) \in \mathbb{R}^2\,. \tag{4.4}$$

Aus (4.4) lesen wir ab:

- Die Dichte von Y ist das Produkt der Dichten g_1, g_2 seiner Komponenten. Daraus folgt die Unabhängigkeit der Komponenten Y_1, Y_2.
- Die Dichten g_1, g_2 sind standard-normalverteilt.
- Die Zufallsgrößen Y_1, Y_2 sind unabhängig und standard-normalverteilt. Daraus ergibt sich der Box-Muller-Algorithmus.

Algorithm 4.3 Box-Muller-Algorithmus

EIN Zwei gleichverteilte Zufallsgrößen U_1, U_2.

Schritt 1 Setze $\theta := 2\pi U_2$, $\rho := \sqrt{-2\ln(U_1)}$.

Schritt 2 Setze $Z_1 := \rho\cos(\theta)$, $Z_2 := \rho\sin(\theta)$.

AUS Z_1, Z_2 sind unabhängige standard-normalverteilte Zufallsgrößen.

Der Algorithmus ist numerisch recht aufwändig, da Wurzeln, trigonometrische Funktionen und der Logarithmus ausgewertet müssen. Die Polar-Methode von Marsaglia und Bray ([24]) entledigt sich der trigonometrischen Funktionen.

Algorithm 4.4 Marsaglia's Polar-Methode

EIN Zwei gleichverteilte Zufallsgrößen U_1, U_2.

Schritt 1 Setze $V_i := 2U_i - 1$ solange $W := V_1^2 + V_2^2 < 1$.

Schritt 2 Setze $Z_1 := V_1\sqrt{-2\ln(W)/W}$, $Z_2 := V_2\sqrt{-2\ln(W)/W}$.

AUS Z_1, Z_2 sind standard-normalverteilte Zufallsgrößen.

In Schritt 1 sind beide „gezogenen Zufallszahlen" U_1, U_2 zu verwerfen, wenn $W \geq 1$ ist. Die „überlebende Zufallsgröße" $X := (V_1, V_2)$ ist auf dem Einheitskreis $D := \{(V_1, V_2) | V_1^2 + V_2^2 < 1\}$ gleichverteilt mit Dichte $1/\pi$. Wir haben die bijektive Abbildung

$$h : D \ni (V_1, V_2) \longmapsto (V_1^2 + V_2^2, \frac{1}{2\pi}\arctan(V_2/V_1)) \in [0,1]^2 .$$

Nun können wir die Dichte der mittels h transformierten Zufallsgröße $Y := h \circ X$ berechnen. Wie oben erhalten wir, dass Y normalverteilt ist.

Der Vorteil dieser Methode ist die Ersparnis der Auswertung trigonometrischer Funktionen. Allerdings gewinnen wir diesen Vorteil nicht zum Nullpreis, denn es werden alle Punkte (V_1, V_2) verworfen, für die nicht $V_1^2 + V_2^2 < 1$ gilt. Wegen

$$1 - \frac{\mathrm{vol}(\{(v_1, v_2) \in \mathbb{R}^2 | v_1^2 + v_2^2 < 1\})}{\mathrm{vol}([-1,1]^2)} = 1 - \frac{\pi}{4} \approx 0.215$$

werden also ungefähr 21.5 % aller Punkte verworfen. Trotzdem gilt Marsaglia's Polar-Methode im Vergleich zur Box-Muller-Methode als die effizientere.

4.3 Wegwerfmethode/Acceptance-Rejection-Method

Eine sehr allgemein anwendbare Transformationstechnik ist die Erzeugung von Zufallszahlen, die einer schwer zugänglichen Verteilung, deren Dichte g bekannt ist, folgen, aus einer „leicht zugänglichen" Dichte, hier meist die Dichte der Gleichverteilung auf $[0,1]$. Die so genannte **Wegwerfmethode** geht auf J. von Neumann (1951) zurück. Dazu braucht man eine Konstante $c \geq 1$ und einen Bereich (x_1, x_2) – der Träger von f, g sei darin enthalten – so dass gilt:

$$g(x) \leq cf(x) \text{ für alle } x \in (x_1, x_2)\,. \tag{4.5}$$

Wenn f die Dichte der Gleichverteilung auf $(x_1, x_2) := (0,1)$ ist, kann man c etwa als das Maximum von g wählen.

Aus der Wahl von c folgt (wir unterstellen $f(x) > 0$ für $x \in (x_1, x_2)$)

$$a(x) := \frac{g(x)}{cf(x)} \in [0,1] \text{ für alle } x \in (x_1, x_2)\,.$$

Wir können also $a(x)$ als **Akkzeptanzwahrscheinlichkeit** interpretieren. Damit ergibt sich folgendes Vorgehen:

Algorithm 4.5 Wegwerfmethode/Acceptance-Rejection-Method

EIN Wahrscheinlichkeitsdichten f, g mit Träger in (x_1, x_2). Mechanismus zur Generierung von Zufallszahlen, passend zur Dichte f, Mechanismus zur Generierung gleichverteilter Zufallszahlen. Schrankenzahl $c \in [1, \infty)$. N Anzahl der zu erzeugenden Zufallszahlen.

Schritt 0 $k := 1$.

Schritt 1 Erzeuge eine Zufallszahl $x \in [x_1, x_2]$ entsprechend der Verteilung mit Dichte f.

Schritt 2 Erzeuge eine Zufallszahl $u \in [0,1]$ entsprechend der Gleichverteilung.

Schritt 3 Ist

$$u \leq a(x) = \frac{g(x)}{cf(x)},$$

akzeptiere x und setze $x^k := x, k := k+1$, ansonsten verwerfe x.

Schritt 4 Wiederhole die Schritte 1,2,3 solange $k \leq N$.

AUS Zufallszahlen $x^1, \ldots, x^N$, die nach der Dichte g verteilt sind.

Es ist natürlich nun zu klären, weshalb die Aussage in AUS zutrifft. Sei dazu die Zufallsgröße X eine Realisierung der Verteilung mit Dichte f und sei U eine auf $[0,1]$ gleichverteilte Zufallsgröße. Sei Y die Zufallsgröße, die den Rückgabewert x des obigen Algorithmus bedingt auf

$$U \leq a(X) := \frac{g(X)}{cf(X)}$$

beschreibt. Um zu klären, ob der Algorithmus Zufallszahlen mit der Dichte g liefert, haben wir die Verteilung von Y zu analysieren.

Betrachte die Ereignisse

$$A := \{X \leq x\} \; (x \in \mathbb{R})\,, \; B := \{U \leq \frac{g(X)}{cf(X)}\}\,.$$

Wir haben zu bestätigen, dass gilt:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \int_{-\infty}^{x} g(w)dw$$

Wir beginnen mit

$$\begin{aligned}
P(B) &= P\left(\left\{U \le \frac{g(X)}{cf(X)}\right\}\right) \\
&= \int_{-\infty}^{\infty} P\left(\left\{U \le \frac{g(X)}{cf(X)}|X = w\right\}\right) f(w)dw \\
&= \int_{-\infty}^{\infty} P\left(\left\{U \le \frac{g(w)}{cf(w)}\right\}\right) f(w)dw \\
&= \int_{-\infty}^{\infty} \frac{g(w)}{cf(w)} f(w)dw \\
&= \frac{1}{c},
\end{aligned}$$

wobei wir die Gleichverteilung von U und die Tatsache, dass g eine Dichte ist, benutzt haben.

$$\begin{aligned}
P(A \cap B) &= P\left(\left\{U \le \frac{g(X)}{cf(X)}, X \le x\right\}\right) \\
&= \int_{-\infty}^{x} P\left(\left\{U \le \frac{g(X)}{cf(X)}|X = w\right\}\right) f(w)dw \\
&= \int_{-\infty}^{x} P\left(\left\{U \le \frac{g(w)}{cf(w)}\right\}\right) f(w)dw \\
&= \int_{-\infty}^{x} \frac{g(w)}{cf(w)} f(w)dw \\
&= \frac{1}{c} \int_{-\infty}^{x} g(w)dw
\end{aligned}$$

Also hat Y die Dichte g.

Sei

$$p := P\left(\left\{U \le \frac{g(X)}{cf(X)}\right\}\right)$$

die Wahrscheinlichkeit, dass bei einem Durchgang durch den Algorithmus eine Akzeptanz eintritt (Erfolgswahrscheinlichkeit). Wir haben oben ausgerechnet, dass $p = \frac{1}{c}$ gilt. Dann ist die Zufallsgröße, die den ersten Erfolg beim n-maligen Durchlauf notiert, nach

$$(1-p)^{n-1}p$$

verteilt (Geometrische Verteilung). Der Erwartungswert ist bekanntlich $\frac{1}{p} = c$. Wir sehen also, dass $c \approx 1$ wünschenswert ist. Um dies zu erreichen, kann man die Dichte f auf g „passend zuschneiden". Hat etwa die Dichte g eine ausgeprägte Spitze, dann ist die Wahl der Dichte f als Gleichverteilung nicht günstig, denn die Wahl $c = \max\{g(x)|x \in (x_1, x_2)\}$ führt dazu, dass sehr viele Versuche verworfen werden müssen.

Als „Überleitung" zu den Monte Carlo-artigen Integrationsverfahren führen wir das mit der Wegwerf-Methode verwandte **Hit- und Miss-Verfahren** an. Gegeben sei eine stetige Funktion f auf dem Intervall $[0, 1]$. Wir wollen das Integral von f über $[0, 1]$ approximativ berechnen. Wir machen dies unter der Annahme, dass der Graph von f in $[0,1] \times [0,1]$ liegt, was keine wirkliche Einschränkung ist. Man schätzt die Fläche unter dem Graphen, was ja unsere gesuchte Zahl ist, dadurch, dass man N-viele im Quadrat $[0,1] \times [0,1]$ gleichverteilte Punkte erzeugt und den Anteil bestimmt, der unter dem Graphen von f liegt.

Seien zur Analyse dieses Vorgehens X, Y unabhängige auf $[0,1]$ gleichverteilte Zufallsgrößen auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$. Setze

$$Z := \chi_{\{Y \le f(X)\}}\,.$$

Dann ist Z eine Bernoulli-Zufallsgröße (mit Werten 0 und 1). Wir haben

$$\begin{aligned} p &:= P(\{Z=1\}) \;=\; P(\{Y \le f(X)\}) \\ &= \int_0^1 \int_0^1 \chi_{\{Y \le f(X)\}}(x,y)dydx \\ &= \int_0^1 \left(\int_0^{f(x)} 1dy \right) dx \;=\; \int_0^1 f(x)dx \end{aligned}$$

Damit ist der Parameter p der Bernoulli-Zufallsgröße Z gerade gleich der gesuchten Fläche. Andererseits gilt $\mathbb{E}(Z) = p$. Mit dem Gesetz der großen Zahlen haben wir also fast sicher die folgende Konvergenz:

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} z^i = \int_0^1 f(x)dx,$$

wobei $z^1, \dots, z^N$ (unabhängige) Realisierungen von Z sind.

Algorithm 4.6 Hit- und Miss-Verfahren

EIN Stetige Funktion $f : [0,1] \longrightarrow [0,1]$. Mechanismus zur Erzeugung von gleichmäßig verteilten Zufallszahlen auf $[0,1]$.

Schritt 1 Für $k = 1, \dots, N$ erzeuge Zufallszahlen x^k, y^k in $[0,1]$.

Schritt 2 $p := \frac{1}{N}\#\{k | y^k \le f(x^k)\}$.

AUS Schätzwert p für das Integral $\int_0^1 f(x)dx$.

4.4 Numerische Integration

Wir bereiten die Monte Carlo-artigen Integrationsverfahren vor.

4.4.1 Eindimensionale Integration

Wir wollen ein Integral

$$I(f) := \int_0^1 f(t)dt$$

berechnen. Die Wahl des Integrationsintervalls als $[0,1]$ ist offenbar keine Einschränkung, solange wir über beschränkte Intervalle integrieren. Wenn ein Integral über $\mathbb{R}$ berechnet werden soll, kann man sich mit einer Transformation behelfen, die $\mathbb{R}$ auf $[0,1]$ abbildet; siehe aber unten die Fragestellung „offen/abgeschlossen“.

Beispiel 4.10 *Wenn man etwa das Integral* $I := \int_0^\infty \ln(1+x^2)e^{-x}dx$ *berechnen will, so kommt man mit der Transformation* $t := 1 - e^{-x}$ *zu* $I = \int_0^1 \ln(1+\ln(1-t)^2)dt$*. Beachte, dass wir uns eine „Singularität" bei* $t = 1$ *eingehandelt haben.* □

Die numerischen Integrationsverfahren gehen mehr oder minder immer von einer Approximation des Integranden durch eine einfache Funktion aus. Im Allgemeinen sind diese einfachen Funktionen Interpolationspolynome. Als Resultat erhält man dann Quadraturformeln folgender Bauart:

$$Q_n(f) := \sum_{i=1}^{n} w_{n,i} f(x^{n,i}).$$

Dabei sind $w_{n,1}, \ldots, w_{n,n}$ **Gewichte**, $0 \le x^{n,1} < \cdots < x^{n,n} \le 1$ die **Stützstellen** und n der **Grad** der Quadraturformel. Der Fehler $e_n(f)$ der Quadraturformel ist gegeben durch

$$e_n(f) := I(f) - Q_n(f).$$

Es sind dann folgende Fragen zu diskutieren:

Einfachheit Die Gewichte und Stützstellen sollten einfach berechenbar sein.

Exaktheit Ein „kleines" $n = n(r)$ sollte ausreichen, Polynome vom Höchstgrad r exakt zu integrieren.

Offen-Geschlossen Dies meint, ob die Randpunkte $0, 1$ Stützstellen sind: ja (abgeschlossen), nein (offen).

Fehlerdarstellung Der Fehler sollte als „Funktion" des Integranden und der Ingredienzien der Quadraturformel (n, Stützstellen) dargestellt werden können, um Aussagen über die Güte der Quadraturformeln zulassen.

Fehlerschranken und Konvergenzordnung Es lassen sich für bestimmte Glattheitsklassen von Integranden Fehlerschranken angeben, die qualitativ und quantitative Aussagen über die Güte der Quadraturformeln zulassen.

Alle diese Fragen sind in unserer eindimensionalen Situation ausreichend untersucht, insbesondere für einige spezielle Quadraturformel-Klassen wie Newton-Cotes-Formeln, Glenshaw-Curtis-Formeln und Gauss-Formeln. Lehrbücher zur Numerischen Mathematik behandeln die Fragen meist ziemlich erschöpfend.

Eine häufig verwendete Quadraturformel ist die **Sehnen-Trapezregel** zur äquidistanten Schrittweite $h := 1/n$:

$$Q_{n,\mathrm{st}}(f) := {\sum_{i=0}^{n}}' h f(ih), \tag{4.6}$$

wobei $\sum'$ bedeutet, dass der erste und der letzte Summand mit $\frac{1}{2}$ zu zählen ist. Aus der Numerischen Mathematik ist bekannt – der Beweis läuft über die Taylorentwicklung –

$$I(f) - Q_{n,\mathrm{st}}(f) = -\frac{1}{12}h^2 f''(\xi) \text{ mit } \xi \in (0,1), \tag{4.7}$$

$$|I(f) - Q_{n,\mathrm{st}}(f)| \le \frac{1}{12}h^2 \max_{0 \le x \le 1} |f''(x)|, \tag{4.8}$$

falls f zweimal stetig differenzierbar ist. Also sind hier die obigen Fragestellungen positiv beantwortet.

4.4.2 Mehrdimensionale Integration

Betrachte ein Integral

$$I^d(f) := \int_{[0,1]^d} f(x)dx\,.$$

Eine exakte analytische Auswertung ist in der Praxis bei $d >> 1$ nur sehr selten möglich. Quadraturformeln haben in der mehrdimensionalen Situation prinzipiell dasselbe Aussehen:

$$Q_n^d(f) := \sum_{i=1}^{n} w_{n,i} f(x^{n,i})\,.$$

Dabei sind erneut $w_{n,1}, \ldots, w_{n,n}$ **Gewichte**, $x^{n,1}, \ldots, x^{n,n}$ die **Stützstellen** und n der **Grad** der Quadraturformel. Der Fehler $e_n(f)$ der Quadraturformel ist erneut gegeben durch

$$e_n^d(f) := I^d(f) - Q_n^d(f)\,.$$

Die obigen Fragestellungen sind nach wie vor aktuell, aber es kommt eine neue, wesentliche Frage hinzu:

Komplexität Wie sieht die Bilanz des Rechenaufwands in Abhängigkeit von d aus, um eine ausreichende Güte der Quadraturformel zu erreichen?

In der numerischen Finanzmathematik ist dies eine Fragestellung mit Priorität, denn es gibt Aufgabenstellungen, bei denen hochdimensionale Integrale zu berechnen sind.

Beispiel 4.11 *Problemstellungen in der Finanzmathematik, bei denen es auf die Integration hochdimensionaler Integrale hinausläuft, sind häufig anzutreffen.*

*Ein solche Problemstellung ist etwa die Bestimmung des Barwertes PV von Zahlungsflüssen eines Portfolios bestehend aus Hypotheken mit 30-jähriger Laufzeit und monatlichen Zahlungen (***Mortage-Backed-Securities (MBS)***). Der Barwert ist der Erwartungswert über die in die Zinsentwicklung einfließende $\mathcal{N}(0,\sigma^2)$-verteilte Zufallsgröße ξ_k für $k = 1, 2, \ldots, d := 30 \cdot 12 = 360$:*

$$PV := \mathbb{E}\left(\sum_{i=1}^{d} u_k m_k\right)\,.$$

Hierbei sind u_k Diskontierungsfaktoren und m_k Zahlungsflüsse, die stochastisch modelliert werden (siehe [5]). Schließlich kann PV umgeschrieben werden zu

$$PV = \int_{\mathbb{R}^d} v(\xi_1, \ldots, \xi_d) g(\xi_1) \cdots g(\xi_1) d\xi_d \cdots d\xi_1$$

wobei g die Dichte einer eindimensionalen Normalverteilung und v eine geeignete Funktion ist. Mit einer Substitution kann das Integral in ein Integral über $[0,1]^d$ umgeschrieben werden. □

Bei der Auswertung von mehrdimensionalen Integralen, also Integralen auf $\mathbb{R}^d, d > 1$, können vier Herangehensweisen bei der Approximationsmethode ausgemacht werden:

(Pseudo-)Zufallszahlen-Methoden Hier nutzt man Zufallspunkte in $\mathbb{R}^d$ als Stützpunkte, wertet den Integranden an den gewählten Zufallspunkten aus und mittelt. Wir sprechen von **Monte Carlo-Methoden**.

Quasi-Zufallszahlen-Methoden Diese Methoden verwenden statt (Pseudo-)Zufallszahlen so genannte **Quasi-Zufallszahlen**. Sie werden deterministisch berechnet und füllen den Raum $\mathbb{R}^d$ sukzessive gleichmäßig aus. Wir sprechen von **Quasi-Monte Carlo-Methoden**.

Gitter-Methoden Diese Methoden übertragen die Interpolationsmethoden vom Eindimensionalen mehr oder minder Eins zu Eins ins Mehrdimensionale. Man wählt in jeder Dimension k Stützstellen, etwa n_k Stück, und wertet dann den Integranden an $n_1 \cdots n_d$ Stützpunkten aus.

Smolyak-Methoden Diese Methoden dünnen die mehrdimensionalen Gitter aus ohne entscheidend an Effizienz zu verlieren. Sie können als spezielle Quasi-Monte Carlo-Methoden angesehen werden: die Stützpunkte werden in einer sehr geschickten Weise gewählt.

Bei all den vier Methoden oben wählen wir Stützpunkte aus; wir nennen die Gesamtheit dieser Stützpunkte hier **Stichproben**. Der Fehler der Integration hängt von zwei Eigenschaften ab:

- Punkteverteilung der Stichproben
- Änderungsverhalten der zu integrierenden Funktion

Für das Änderungsverhalten einer Funktion ist eine passende Definition der Variation einer Funktion einzuführen. Die Verteilung der Punkte einer Stichprobe kann durch den Begriff der Diskrepanz beschrieben werden.

4.5 Monte Carlo-Methoden

Hier geben wir einen allgemeinen Überblick über Monte Carlo-Verfahren, der zeigen soll, dass es ein allgemeineres Vorgehen ist, als die Schilderungen zur Integration vorgeben.[1]

Monte Carlo-Methoden benutzen zur Simulation von realen Vorgängen Zufallszahlen. Im Allgemeinen ist der Aufwand, zu guten Ergebnissen zu gelangen, groß.[2] Wir werden sehen, dass sie sich gerade in der Berechnung von Optionspreisen die Monte Carlo-Simulation auf Grund ihrer großen Flexibilität bewähren. Zwei Ansätze, die in einem einfachen Zusammenhang stehen, führen zu den Methoden der Monte Carlo-Simulation: (Hochdimensionale) Integration und Schätzung von Erwartungswerten. Beide Aufgaben fallen bei numerischen Fragestellungen in der Optionspreisermittlung miteinander verknüpft an.

4.5.1 Grundidee

Eine allgemeine Definition könnte sein:

> *Monte Carlo ist ein Vorgehen zur Approximation eines Erwartungswertes einer Funktion einer Zufallsgröße durch den Mittelwert einer Stichprobe, gezogen mit der Zufallsgröße.*

[1] Als Entdecker der Monte Carlo–Simulation gilt G.L.L. de Buffon (1707-1788), der als erster die Zahl π durch den Wurf einer Nähnadel auf eine karierte Tischdecke berechnete. Seine Idee war, dass über den zufälligen Winkel zwischen Nadel und parallelem Karomuster der Tischdecke die Zahl π steckt. Mit dem 10 000-maligen Wurf der Nadel konnte er so die Zahl π auf mehrere Stellen genau berechnen.

Die genaue Herkunft der Bezeichnung für dieses Verfahren ist umstritten, fest steht, dass der Begriff „Monte Carlo“ wohl das erste Mal im zweiten Weltkrieg als Deckname für eine geheime Forschung im Bereich des amerikanischen Atomwaffenprogramms (Manhattan-Projekt/Neutronendiffusion), an dem J. v. Neumann und S. Ulam beteiligt waren, verwendet wurde. Vermutlich wurde der Name von einem 1862 in Monaco gegründeten Spielcasino abgeleitet.

[2] *Monte Carlo Methoden sind extrem schlecht; sie sollten nur dann verwendet werden, wenn sämtliche Alternativen noch schlechter sind (Alan Sokal, 1997).*

Gemeinsam ist in der Analyse des Vorgehens die Anwendung des Gesetzes der großen Zahl. In mathematische Terme übersetzt bedeutet dies:

> **Gegeben** sei eine Zufallsgröße X (mit Werten in $\mathbb{R}^d$) auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$.
> **Gegeben** sei weiterhin eine Funktion f.
> **Wähle** eine Stichprobe $x^1, \dots, x^n$, gezogen mit der Zufallsgröße X.
> **Berechne** den Monte Carlo-Schätzer $\hat{\theta}_n$ für $\theta := \mathbb{E}(f(X)) := \mathbb{E}(f \circ X)$ durch
>
> $$\hat{\theta}_n := \frac{1}{n} \sum_{k=1}^{n} f(x^k)\,.$$

Alternativ könnten wir bei $\hat{\theta}_n$ auch von der Zufallsgröße

$$f_n(X) := \frac{1}{n} \sum_{k=1}^{n} f(X_k)$$

sprechen, wobei $X_1, \dots, X_n$ nach X verteilt sind. Wir beobachten, dass $f_n(X)$ ein **erwartungstreuer (unbiased) Schätzer** von $\mathbb{E}(f(X))$ ist:

$$\mathbb{E}(f_n(X)) = \mathbb{E}\left(\frac{1}{n} \sum_{k=1}^{n} f(X_k)\right) = \frac{1}{n} \sum_{k=1}^{n} \mathbb{E}(f(X_k)) = \mathbb{E}(f(X))\,. \tag{4.9}$$

Bemerkung 4.12 *In der Literatur wird die Monte Carlo-Simulation im Allgemeinen für den „einfacheren Fall“ $f = id$ dargestellt.* □

Falls $\theta = \mathbb{E}(f(X))$ existiert, dann besagt das schwache Gesetz der großen Zahl, dass für jedes (kleine) $\varepsilon > 0$ gilt:

$$\lim_n P(\{|f_n(X) - \mathbb{E}(f(X))| \geq \varepsilon\}) = 0\,. \tag{4.10}$$

Für stärkere Aussagen benötigt man auch noch die Existenz der Varianz. Wir gehen von einer kontinuierlichen Zufallsgröße aus und nehmen an, dass $f(X_1), \dots, f(X_n)$ unabhängige Realisierungen von $f(X)$ sind. Dann

$$\mathbb{V}(f_n(X)) = \mathbb{V}\left(\frac{1}{n} \sum_{k=1}^{n} f(X_i)\right) = \frac{1}{n}\mathbb{V}(f(X)) = \frac{1}{n} \int_M (f(x) - \theta)^2 g(x)dx\,.$$

Der zentrale Grenzwertsatz besagt: Existieren $\theta = \mathbb{E}(f(X))$ und $\sigma^2 := \mathbb{V}(f(X))$, so ist $S_n := f(X_1) + \cdots + f(X_n)$ für große n approximativ $\mathcal{N}(n\theta, n\sigma^2)$ verteilt, oder anders aufgeschrieben:

$$\lim_n P\left(\left\{\frac{S_n/n - \theta}{\sigma/\sqrt{n}} \leq x\right\}\right) = \mathcal{N}(x)\,,\ x \in \mathbb{R}\,; \tag{4.11}$$

siehe Satz 4.5 und Bemerkung 4.6.

Aus (4.10) und (4.11) folgt, dass $\hat{\theta}_n$ für große n ein guter Schätzer für θ ist. Damit ist die Grundidee der Monte Carlo-Simulation beschrieben.

Bemerkung 4.13 *In der obigen Formulierung haben wir nicht unterstellt, dass X eine Gleichverteilung besitzt. Dies ist auch angebracht, um eine ausreichende Flexibilität vorzuhalten, welche z.B. gefragt ist, wenn wir später etwa die Varianzreduktion betrachten.* □

Wir wollen mit zwei Anwendungen zeigen, dass der obige Rahmen allgemein genug abgesteckt ist. Unser Hauptinteresse wird die Integration sein, die wir später einordnen werden.

Beispiel 4.14 *Sei Y eine Zufallsgröße auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$. Wir wollen die Wahrscheinlichkeit des Ereignisses $P(\{Y \in A\})$ schätzen. Wir können dies tun, indem das Problem als die Schätzung des Erwartungswertes der Indikatorfunktion $\chi_{\{Y\in A\}}$ aufgeschrieben wird:*

$$P(\{Y \in A\}) = \mathbb{E}(\chi_{\{Y\in A\}})$$

□

Beispiel 4.15 *Wir betrachten die Summe der Werte einer Funktion $Q : M \longrightarrow \mathbb{R}$. Sei Y die Zufallsgröße, die jeden Wert ω in M mit gleicher Wahrscheinlichkeit p annimmt ($\sum_{\omega\in M} p = 1$). Dann kann diese Summe über die Werte der Funktion Q wieder als Erwartungswert geschrieben werden:*

$$\sum_{\omega\in M} Q(\omega) = \frac{1}{p}\sum_{\omega\in M} Q(\omega)p = \frac{1}{p}\mathbb{E}(Q(Y)) .$$

□

4.5.2 Konvergenzgeschwindigkeit

Numerische Tests zeigen, dass die Monte Carlo-Schätzer $\hat{\theta}_n$ für wachsende Größe der Stichproben relativ langsam gegen θ konvergieren. Wir wollen die Konvergenz noch etwas anders betrachten.

Definition 4.16 *Sei $p \in [0,1]$ und $n \in \mathbb{N}$. Ein Intervall I der Form $[\theta - \varepsilon, \theta + \varepsilon]$ heißt* **Konfidenzintervall** *des Monte Carlo-Schätzers $\hat{\theta}_n$ für θ, wenn gilt:*

$$P\left(\frac{1}{n}\sum_{k=1}^{n} f(X_k) \in I\right) = p .$$

□

Halten wir nun p aus der Definition 4.16 fest, so können wir untersuchen, wie groß diese Konfidenzintervalle in Abhängigkeit von n sind.

Satz 4.17 *Sei $p \in (0,1)$. Dann existiert $k > 0$ und eine Folge $(p_n)_{n\in\mathbb{N}}$ mit $\lim_n p_n = p$, so dass die p_n-Konfidenzintervalle I_n von der Form*

$$I_n = \left[\theta - \frac{k\sigma}{\sqrt{n}}, \theta + \frac{k\sigma}{\sqrt{n}}\right]$$

sind.

Beweis:
Wähle $x \in \mathbb{R}$, so dass mit der Verteilungsfunktion F der Standard-Normalverteilung $F(x) - F(-x) = p$ gilt. Auf Grund der Konvergenz (4.11) existiert eine Folge $\delta_n^{\pm}$ mit

$$\left|P\left(\left\{\frac{S_n - n\theta}{\sigma\sqrt{n}} \le \pm x\right\}\right) - F(\pm x)\right| = \delta_n^{\pm} , \lim_n \delta_n^{\pm} = 0 .$$

Dann ergibt eine einfache Rechnung

$$P\left(\left\{\left|\frac{S_n - n\theta}{\sigma\sqrt{n}}\right| \le x\right\}\right) = F(x) - F(-x) + \delta_n^{+} - \delta_n^{-} = p + \delta_n^{+} - \delta_n^{-} =: p_n .$$

Es gilt $\lim_n p_n = 0$ und

$$|\frac{S_n - n\theta}{\sigma\sqrt{n}}| \leq x \text{ genau dann, wenn } \frac{1}{n}S_n \in \left[\theta - \frac{k\sigma}{\sqrt{n}}, \theta + \frac{k\sigma}{\sqrt{n}}\right]$$

für $k := x$. ■

Die Kernaussage des Satzes 4.17 ist, wie schnell die Konfidenzintervalle in Abhängigkeit von n schrumpfen: die Breite der Intervalle schrumpft wie $\frac{1}{\sqrt{n}}$. Damit erklärt sich die langsamme Konvergenz einer Monte Carlo-Simulation. An dieser Konvergenzgüte lässt sich nichts ändern. Um die Konvergenzgüte zu verbessern, bleibt also nur, die Konstante σ zu verkleinern. Wir werden uns nun also im nächsten Abschnitt mit Maßnahmen beschäftigen, die hierbei behilflich sein können. Zunächst skizzieren wir noch die Monte Carlo-Integration.

4.5.3 Monte-Carlo-Integration

Wir wollen ein Integral

$$I^d(f) := \int_{[0,1]^d} f(x)dx$$

berechnen. Wir identifizieren dazu eine Zufallsgröße X mit Träger in $(0,1)^d$ und Dichte g und eine Funktion G, so dass der Erwartungswert θ von $G \circ X$ den Wert des Integrals ergibt:

$$\theta = \mathbb{E}(G \circ X) = \int_{[0,1]^d} G(x)g(x)dx = I^d(f).$$

Dies ist eine Aufgabenstellung, wie wir sie oben diskutiert haben.

In den meisten Fällen wird X als Zufallsgröße mit einer Gleichverteilung auf $[0,1]^d$ und G als Integrand f gewählt:

$$\theta = \mathbb{E}(f \circ X) = \int_{\mathbb{R}^d} f(x)g(x)dx = \int_{[0,1]^d} 1f(x)dx = I^d(f).$$

Diesen Fall verfolgen wir weiter.

Der Erwartungswert $\theta := \mathbb{E}(f(X))$ kann näherungsweise durch eine Summe berechnet werden mittels einer Stichprobe $x^{n,1}, \dots, x^{n,n}$ von Punkten, die in $[0,1]^d$ gleichverteilt sind:

$$\theta = \mathbb{E}(f \circ X) \approx \hat{\theta}_n := Q_n^d(f) := \frac{1}{n}\sum_{i=1}^{n} f(x^{n,i}).$$

$\hat{\theta}_n$ ist ein „erwartungstreuer Schätzer (unbiased)“ von θ, denn

$$\mathbb{E}(\frac{1}{n}\sum_{i=1}^{n} f \circ X_i) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}(f \circ X_i) = \mathbb{E}(f \circ X) = \theta,$$

wobei $X_1, \dots, X_n$ unabhängige auf $[0,1]^d$ gleichverteilte Zufallsgrößen sind, die die Stützpunkte $x^{n,i}$ realisieren. Die Varianz dieser Schätzmethode ist

$$\mathbb{V}(\hat{\theta}_n) = \frac{1}{n^2}\sum_{i=1}^{n} \mathbb{V}(f \circ X_i) = \frac{1}{n}\mathbb{V}(f \circ X_i) = \frac{1}{n}\int_{[0,1]^d} \left(f(x) - \int_{[0,1]^d} f(u)d(u)\right)^2 dx.$$

Das Integral

$$\sigma_f^2 := \int_{[0,1]^d} \left(f(x) - \int_{[0,1]^d} f(u)du\right)^2 dx$$

ist ein Maß der Rauhheit/Glattheit des Integranden f.

Die Qualität des Schätzers $\hat{\theta}_n$ haben wir in den beiden vorhergehenden Abschnitten diskutiert. Übersetzt bedeuten diese Ergebnisse u.a., dass für fast alle Realisierungen der Folge X_i für die Mittelwertbildung der klassische Konvergenzbegriff aus der reellen Analysis gilt. Somit ist $\hat{\theta}_n$ stark konsistenter Schätzer für $\int_{[0,1]^d} f(x)\,dx$, d.h. für große n ist θ_n mit hoher Wahrscheinlichkeit eine gute Näherung des Integrals $\int_{[0,1]^d} f(x)\,dx$. Im folgenden Satz halten wir dies fest und ergänzen es um ein quantitatives Resultat.

Satz 4.18 *Sei $f : [0,1]^d \longrightarrow \mathbb{R}$ Lebesgue-integrierbar und es gelte*

$$\sigma_f^2 := \int_{[0,1]^d} \left(f(x) - \int_{[0,1]^d} f(u)du \right)^2 < \infty\,. \tag{4.12}$$

Dann gilt:

(a) $\lim_n Q_n^d(f) = \int_{[0,1]^d} f(x)dx$ *λ_d-fast sicher*

(b) $\lim_n \lambda_d \left(\frac{\sigma_f}{\sqrt{n}}\, a < e_n^d(f) < \frac{\sigma_f}{\sqrt{n}}\, b \right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{1}{2}t^2} dt\,, \ -\infty < a < b < \infty\,.$

(c) $|I(f) - Q_n^d(f)| \le \sigma_f/\sqrt{n}\,,\ n \in \mathbb{N}\,.$

Hierbei ist λ_d das Lebesgue-Maß in $\mathbb{R}^d$.

Beweis:
(a) folgt aus dem Gesetz der großen Zahlen, (b) aus dem zentralen Grenzwertsatz (von Lindeberg-Lévy); siehe etwa [21]. Zu (c) führen wir für $d = 1$ einen elementaren Beweis an.

$$\begin{aligned}
\mathbb{E}(I(f) - \hat{\theta}_n)^2 &= \mathbb{E}\left(I(f) - \frac{1}{n}\sum_{k=1}^n f(x^k) \right)^2 \\
&= \mathbb{E}\left(\frac{1}{n}\sum_{k=1}^n (I(f) - f(x^k)) \right)^2 \\
&= \frac{1}{n^2}\int_0^1 \cdots \int_0^1 \left(\sum_{k=1}^n (I(f) - f(x^k)) \right)^2 dx_1 \cdots dx_n \\
&= \frac{1}{n^2}\int_{[0,1]^d} \left(\sum_{k=1}^n (I(f) - f(x^k))^2 + 2\sum_{k<j} (I(f) - f(x^k))(I(f) - f(x^j)) \right) dx_1 \cdots dx_n \\
&= \frac{1}{n^2}\sum_{k=1}^n \int_0^1 (I(f) - f(x^k))^2 dx_k \\
&\quad + \frac{2}{n^2}\sum_{k<j} \int_0^1 (I(f) - f(x^k))dx_k \int_0^1 (I(f) - f(x^j))dx_j \\
&= \frac{1}{n}\int_0^1 (I(f) - f(x))^2 dx \\
&= \frac{\sigma_f^2}{n}
\end{aligned}$$

■

Die **Nachteile der Monte Carlo-Integration** sind:

- Die Konvergenzaussagen in Satz 4.18 sind „probabilistischer Natur“. Sie schließen nicht aus, dass das Ergebnis stark abweicht.

- Es ist bekannt, dass das Ergebnis der Monte Carlo-Integration stark vom verwendeten Pseudo-Zufallsgenerator, der bei der Auswahl der Stützpunkte eingesetzt wird, abhängen kann.

- Der Fehler ist von der Größenordnung $O(n^{-\frac{1}{2}})$ für jede Dimension d. Dies bedeutet, dass 100-mal mehr Funktionsauswertungen benötigt werden, um eine Stelle mehr an Genauigkeit zu erreichen.

Fassen wir die **Vorteile der Monte Carlo-Integration** zusammen:

- Wir wissen, dass die Sehnentrapezregel einen Aufwand von $O(\varepsilon^{-d/2})$ erfordert, wenn man einen Fehler $\varepsilon > 0$ erreichen will: $O(n^d)$ Funktionsauswertungen, Genauigkeit $O(n^{-2})$. Dagegen benötigt die Monte Carlo-Integration einen Aufwand von $O(\varepsilon^{-2})$, wenn man einen Fehler $\varepsilon > 0$ erreichen will: $O(n)$ Funktionsauswertungen, Genauigkeit $O(n^{-1/2})$. Wir stellen also fest, dass die Monte Carlo-Integration für $d >> 1$ eine bessere Bilanz vorweist.

- Die Behandlung der Integrationsgrenzen ist unwesentlich.

- Die Genauigkeit kann kontinuierlich gesteigert werden (durch Hinzunahme weiterer Punkte).

- Der Fehler ist leicht abschätzbar, denn die Varianz σ_f lässt sich ohne großen Aufwand durch die diskrete „Varianz“ $\sigma_{n,f} := \sqrt{\frac{1}{n-1}\sum_{i=1}^n (f(x^{n,i}) - Q_n^d(f))^2}$ approximieren.

4.6 Varianzreduktion

Die Idee der Varianzreduktion ist: Simuliere statt der Zufallsgrößen X_i andere Zufallsgrößen Y_i mit gleichem Erwartungswert, aber kleinerer Varianz. Varianzreduktion hat auch damit zu tun, wie Ereignisse zu simulieren sind, die eine kleine Wahrscheinlichkeit besitzen.

4.6.1 Abtrennung des Hauptteils

Wir betrachten wieder das Integral

$$I^d(f) := \int_{[0,1]^d} f(x)dx$$

und nehmen an, dass das Integral

$$I^d(\psi) := \int_{[0,1]^d} \psi(x)dx$$

analytisch berechenbar ist. Nun kann man hinschreiben:

$$I^d(f) = \int_{[0,1]^d} (f(x) - \psi(x))dx + \int_{[0,1]^d} \psi(x)dx$$

und daraus den Schätzer

$$\theta_n(f;\psi) := \frac{1}{n}\sum_{k=1}^n (f(x^k) - \psi(x^k)) + \int_{[0,1]^d} \psi(x)dx = \theta_n(f) - \theta_n(\psi) + I^d(\psi)\,,$$

ableiten, wobei

$$\theta_n(f) = \frac{1}{n}\sum_{k=1}^{n} f(x^k)\,,\ \theta_n(\psi) = \frac{1}{n}\sum_{k=1}^{n} \psi(x^k)$$

ist; $x^1, \ldots, x^n$ sind gleichverteilte Zufallszahlen in $[0,1]$. Der Integrand ψ – wir nennen ihn in unserem Zusammenhang **Hauptteil** – sollte dem Integranden f möglichst ähnlich sein, damit die Varianz von $\theta_n(f;\psi)$ kleiner als die Varianz von $\theta_n(f)$ wird. Aber warum sollte dieses Vorgehen die Varianz reduzieren?

Bekanntlich ist die Kovarianz von zwei Zufallsgrößen X, Y definiert als

$$\mathrm{Cov}(X,Y) := \mathbb{E}((X-\mathbb{E}(X))(Y-\mathbb{E}(Y))) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)\,.$$

Eine einfache Rechnung zeigt

Regel 4.19 *Cov*$(X,Y) \le \frac{1}{2}(\mathbb{V}(X) + \mathbb{V}(Y))$.

Man kann erwarten, dass aus der Ähnlichkeit von ψ und f sowohl die Ähnlichkeit von $I^d(f)$ und $I^d(\psi)$ als auch die von $\theta_n(f)$ und $\theta_n(\psi)$ folgt. Entsprechend sollte die Kovarianz davon nahe der oberen Schranke sein, also

$$\mathrm{Cov}(\theta_n(f), \theta_n(\psi)) \approx \frac{1}{2}(\mathbb{V}(\theta_n(f)) + \mathbb{V}(\theta_n(\psi)))\,,$$

was dann die Ungleichung

$$\mathrm{Cov}(\theta_n(f), \theta_n(\psi)) \ge \frac{1}{2}\mathbb{V}(\theta_n(\psi))$$

impliziert. Für die Varianz des neuen Schätzers $\theta_n(f;\psi)$ gilt dann

$$\mathbb{V}(\theta_n(f;\psi)) = \mathbb{V}(\theta_n(f)) + \mathbb{V}(\theta_n(\psi)) - 2\mathrm{Cov}(\theta_n(f), \theta_n(\psi)) \lessapprox \mathbb{V}(\theta_n(f))\,.$$

Folglich ist die Varianz unter unseren Annahmen an den Hauptteil zumindest nicht größer geworden.

Beispiel 4.20 *Die Wahl des Hauptteils kann z.B. so erfolgen:*

$$\psi(x) = \begin{cases} f(x) & \textit{falls } x \le c \\ 0 & \textit{falls } x > c \end{cases}$$

wobei $c \in (0,1)$. *Hierbei unterstellen wir, dass das Integral von* f *über* $[0,c]$ *bekannt sei.* □

4.6.2 Antithetische Variablen

Es soll der Erwartungswert $\mathbb{E}(f(X))$ mittels Monte Carlo-Simulation approximiert werden. Die Idee ist nun, die Abbildung f durch eine Funktion $\tilde{f}$ zu ersetzen mit der Eigenschaft

$$\mathbb{E}(\tilde{f}(X)) = \mathbb{E}(f(X))\,,\ \mathbb{V}(\tilde{f}(X)) < \mathbb{V}(f(X))\,. \tag{4.13}$$

Dies kann man mit **antithetischen (entgegengesetzten) Variablen** versuchen. Die Bezeichnung rührt vom Spezialfall

$$\tilde{f}(X) := \frac{1}{2}(f(X) + f(-X)) \tag{4.14}$$

her. In der Monte Carlo-Simulation führt dies zum Ersetzen von

$$\frac{1}{n}\sum_{k=1}^{n} f(X_k)$$

durch
$$\frac{1}{n}\sum_{k=1}^{n}\frac{1}{2}(f(X_k)+f(-X_k))\,.$$
Jede mit X gezogene Zufallszahl x^k geht also zweimal in den Schätzer
$$\hat\theta_n := \frac{1}{n}\sum_{k=1}^{n}\frac{1}{2}(f(x^k)+f(-x^k))$$
ein. Wie sieht es mit der Varianz $\mathbb{V}(\tilde f(X))$ aus? Wir beweisen ein Lemma, das weiterhelfen kann.

Lemma 4.21 *Sei X eine Zufallsgröße und sei $a,b:\mathbb{R}\longrightarrow\mathbb{R}$ ein ungleiches Paar, d.h. a monoton wachsend und b monoton fallend oder a monoton fallend und b monoton wachsend. Dann gilt $Cov(a(X),b(X))\le 0$.*

Beweis:
Wähle eine weitere Zufallsgröße X, so dass X,Y unabhängig und identisch verteilt sind. Dann haben wir auf Grund der Monotonieannahme
$$\begin{aligned}
0 &\ge \mathbb{E}((a(X)-a(Y))(b(X)-b(Y))) \quad \text{(fast sicher)}\\
&= \mathbb{E}(a(X)b(X))-\mathbb{E}(a(X)b(Y))-\mathbb{E}(a(Y)b(X))+\mathbb{E}(a(Y)b(Y))\\
&= 2\mathbb{E}(a(X)b(X))-2\mathbb{E}(a(X))\mathbb{E}(b(X))\\
&= 2\mathrm{Cov}(a(X),b(X))
\end{aligned}$$

■

Folgerung 4.22 *Sei X standard-normalverteilt, sei $f:\mathbb{R}\longrightarrow\mathbb{R}$ monoton wachsend und sei $\tilde f(x) := \frac{1}{2}(f(x)+f(-x)), x\in\mathbb{R}$. Dann gilt:*

(a) $\mathbb{E}(\tilde f(X)) = \mathbb{E}(f(X))$.

(b) $\mathbb{V}(\tilde f(X)) \le \frac{1}{2}\mathbb{V}(f(X))$.

Beweis:
Ist X standard-normalverteilt, dann ist es auch $-X$. Also gilt
$$\mathbb{E}(\tilde f(X)) = \frac{1}{2}(\mathbb{E}(f(X))+\mathbb{E}(f(-X))) = \frac{1}{2}(\mathbb{E}(f(X))+\mathbb{E}(f(X))) = \mathbb{E}(f(X))$$
und
$$\mathbb{V}(f)(X)) = \mathbb{V}(f)(-X))\,.$$
Zur Untersuchung der Varianz von $\tilde f(X)$ verwenden wir
$$\mathbb{V}(\tilde f(X)) = \frac{1}{2}(\mathbb{V}(f(X))+\mathrm{Cov}(f(X),f(-X)))\,.$$
Nun können wir Lemma 4.21 anwenden, da a,b mit $a(x):=f(x), b(x):=f(-x)$ ein ungleiches Paar ist, und erhalten
$$\mathrm{Cov}(f(X),f(-X))\le 0\,.$$

■

Beispiel 4.23 *Die Anwendung des obigen Vorgehens ist im Zusammenhang mit der Erzeugung eines Pfades der geometrischen Brownschen Bewegung, wie wir sie im nächsten Kapitel betrachten werden, zu sehen. Wir diskutieren dies schon mal an.*

Der faire Preis einer Call-Option kann (unter Verwendung der geometrischen Brownschen Bewegung) dargestellt werden als

$$f(Z) := \left(S_0 \exp((r - \frac{1}{2}\sigma^2)T + \sigma\sqrt{T}Z) - K\right)^+ .$$

Dabei ist Z eine standard-normalverteilte Zufallsgröße. Da f monoton wachsend in Z ist, ist das vorhergehende Lemma anwendbar mit $a(Z) := f(Z), b(Z) := f(-Z)$. ■

Das folgende Beispiel zeigt, wie die Vorgehensweise bei gleichverteilten Verteilungen genutzt werden kann.

Beispiel 4.24 *Wir wollen*

$$\theta := \mathbb{E}(e^U) = \int_0^1 e^x dx \quad (U \sim \mathcal{U}[0,1])$$

mit auf $[0,1]$ gleichverteilten Zufallsgröße U durch Zufallszahlen simulieren.

Klar, $\theta = e - 1$. Wir berechnen Kennzahlen, die klären, dass der Einsatz von antithetischen Variablen sehr erfolgreich sein wird.

$$Cov(e^U, e^{1-U}) = \mathbb{E}(e^U e^{1-U}) - \mathbb{E}(e^U)\mathbb{E}(e^{1-U}) = e - (e-1)^2 = -0.2342 .$$

$$\mathbb{V}(e^U) = \mathbb{E}(e^{2U}) - \mathbb{E}(e^U)^2 = \int_0^1 e^{2x} dx - (e-1)^2 = 0.2420 .$$

Nutzen wir die antithetischen Zufallsgrößen $U, 1-U \sim \mathcal{U}[0,1]$, so erhalten wir

$$\mathbb{V}\left(\frac{1}{2}(e^U + e^{1-U})\right) = \frac{1}{2}(\mathbb{V}(e^U) + Cov(e^U, e^{1-U})) = 0.0039 .$$

Also haben wir unter Nutzung antithetischer Variablen eine beträchtliche Varianzreduktion erzielt.

Für zwei unabhängige Zufallsgrößen $U_1, U_2 \sim \mathcal{U}[0,1]$ erhalten wir

$$\mathbb{V}\left(\frac{1}{2}(e^{U_1} + e^{U_2})\right) = \frac{1}{2}\mathbb{V}(e^{U_1}) = 0.1210 .$$

Die Reduktion ist nicht sehr beeindruckend. □

4.6.3 Kontroll-Zufallsgrößen

Wir wollen den Erwartungswert $\theta := \mathbb{E}(X)$ simulieren. Wir nehmen an, dass eine weitere Zufallsgröße Y zur Verfügung steht, deren Erwartungswert $\mu := \mathbb{E}(Y)$ wir kennen. Dann ist

$$Z := X + r(Y - \mu)$$

ein erwartungstreuer Schätzer, denn es ist $\mathbb{E}(X) = \theta$. Um den besten Skalar r^* zu bestimmen, berechnen wir die Varianz von X.

$$\mathbb{V}(Z) = \mathbb{V}(X + r(Y - \mu)) = \mathbb{V}(X) + r^2\mathbb{V}(Y) + 2r\text{Cov}(X, Y)$$

Die rechte Seite wird minimal für

$$r^* := -\frac{\mathrm{Cov}(X,Y)}{\mathbb{V}(Y)} . \tag{4.15}$$

Die Varianz von $Z := X + r^*(Y-\mu)$ berechnet sich zu

$$\mathbb{V}(X + r^*(Y-\mu)) = \mathbb{V}(X) - \frac{\mathrm{Cov}(X,Y)^2}{\mathbb{V}(Y)} . \tag{4.16}$$

Die Zufallsgröße Y heißt in unserem Zusammenhang **Kontroll-Zufallsgröße (control variate)**. Die Effektivität des Vorgehens ist abzulesen an

$$\mathbb{V}(X + r^*(Y-\mu)) = \mathbb{V}(X)(1 - \mathrm{Corr}(X,Y)^2) , \tag{4.17}$$

wobei

$$\mathrm{Corr}(X,Y) := \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathbb{V}(X)\mathbb{V}(Y)}}$$

die **Korrelation** von X,Y ist. Um dies nutzen zu können, benötigen wir aber zur Berechnung von r^* $\mathrm{Cov}(X,Y)$ und $\mathbb{V}(Y)$. Dazu können wir die Stichproben $x^1,\dots,x^n$ und $y^1,\dots,y^n$ nutzen, die wir zur Schätzung von $\mathbb{E}(X)$ benötigen, denn

$$\mathrm{Cov}(X,Y) \approx \frac{1}{n-1}\sum_{k=1}^{n}(x^k - \hat\theta_n)(y^k - \hat\mu_n) \, , \; \mathbb{V}(Y) \approx \frac{1}{n-1}\sum_{k=1}^{n}(y^k - \hat\mu_n)^2 \, ,$$

wobei

$$\hat\theta_n := \frac{1}{n}\sum_{k=1}^{n} x^k \approx \mathbb{E}(X) \, , \; \hat\mu_n := \frac{1}{n}\sum_{k=1}^{n} y^k \approx \mathbb{E}(Y) \, .$$

4.6.4 Selektives Simulieren (Importance sampling)

Mit **Importance Sampling** soll der Versuch gemacht werden, die Verteilung der gezogenen Zufallszahlen so an das Problem anzupassen, dass die „wichtigen Ereignisse“ auch wirklich eintreten.

Beispiel 4.25 *Wir können das Integral $\int_0^1 f(x)dx$ mit einer Gleichverteilung in $[0,1]$ simulieren. Wir könnten aber auch mit einer auf $[0,5]$ gleichverteilten Zufallsgröße W simulieren, wenn wir f mit Null auf $\mathbb{R}$ fortsetzen. Dies bedeutet $\int_0^1 f(x)dx = 5\mathbb{E}(f(W))$ und ergibt den Schätzer*

$$\frac{5}{n}\sum_{k=1}^{n} f(x^k) \;\text{ mit } x^1,\dots,x^n \text{ gleichverteilt in } [0,5] \, .$$

Aber dieses Vorgehen ist nicht sehr effektiv, denn 80% der Funktionsauswertungen tragen nichts zum Ergebnis bei. □

Was ist nun im Blick bei „Importance Sampling“? Es ist die Wahl einer Verteilung, nach der die Stichprobe gezogen wird, mit dem Ziel eine kleinere Varianz zu erhalten. Sei g eine (bekannte) Dichte einer Zufallsgröße Z auf dem Wahrscheinlichkeitsraum $(\Omega,\mathcal{F},P)$. Sei $G : \mathbb{R} \longrightarrow \mathbb{R}$. Wir wollen

$$\theta := \mathbb{E}_g(G(Z)) = \int_{\mathbb{R}} G(z)g(z)dz \tag{4.18}$$

berechnen. Dazu wollen wir die Dichte g gegen eine andere Dichte auswechseln.

Sei h eine weitere Dichte mit

$$h(z) = 0 \implies G(z)g(z) = 0, z \in \mathbb{R}\,. \tag{4.19}$$

Damit definieren wir den **Likelihood-Quotient** L von g und h durch

$$L(z; g, h) := \begin{cases} \frac{g(z)}{h(z)} & \text{falls } h(z) \neq 0 \\ 0 & \text{sonst} \end{cases} \tag{4.20}$$

Dann haben wir

$$\theta = \mathbb{E}_g(G(Z)) = \int_D G(z)g(z)dz = \int_D G(z)L(z; g, h)h(z)dz = \mathbb{E}_h\left(G(Z)L(Z; g, h)\right)\,, \tag{4.21}$$

wobei $\mathbb{E}_h$ bedeutet, dass der Erwartungswert bezüglich der Verteilungsdichte h genommen wird. Der Likelihood-Quotient $L(z; g, h)$ wird genutzt, um den Wechsel in der Wahrscheinlichkeit zu kompensieren. Zum Verständnis von L folgende Überlegung. Da g, h Dichten sind, können wir für ein Ereignis $A \in \mathcal{F}$ definieren:

$$p(A) := \int_A g(z)dz\,,\; q(A) := \int_A h(z)dz\,.$$

Die Bedingung der **absoluten Stetigkeit** von p bezüglich q ist:

$$\text{Ist } A \in \mathcal{F} \text{ mit } p(A) > 0, \text{ so ist } q(A) > 0\,. \tag{4.22}$$

Beachte, dass unter der Bedingung (4.22) das Maß q einer Menge $A \in \mathcal{F}$ einen positiven Wert zuweisen kann, selbst, wenn $p(A) = 0$ ist. Die Bedingung (4.22) stellt sicher, dass nach dem Satz von **Radon-Nikodym** eine Ableitung existiert:

$$\int_A g(z)dz = p(A) = \int_A \frac{\partial p}{\partial q}(z)dq(z) = \int_A \frac{\partial p}{\partial q}(z)h(z)dz\,,\; A \in \mathcal{F}\,.$$

Wenn wir den Sachverhalt mit den Erwartungswerten in (4.21), der Definition in (4.20) und der Voraussetzung (4.19) zusammenbringen, dann wird klar, warum ein Likelihood-Quotient L auch Radon-Nikodym-Quotient genannt wird.

Wir haben also gesehen, dass

$$\hat{\theta} = G(Z)\frac{g(Z)}{h(Z)}$$

ein erwartungstreuer Schätzer von θ ist, wenn wir Z mit h simulieren. Dies ergibt den Schätzer

$$\theta_n(Z; h) = \frac{1}{n}\sum_{k=1}^{n} G(Z_k)\frac{g(Z_k)}{h(Z_k)} \text{ wobei jedes } Z_k \text{ nach } h \text{ verteilt ist.}$$

Die Varianz von $\hat{\theta}$ ist

$$\mathbb{V}_h\left(G(Z)\frac{g(Z)}{h(Z)}\right) := \int_{\mathbb{R}} \left(G(z)\frac{g(z)}{h(z)} - \theta\right)^2 h(z)dz\,.$$

Wenn wir annehmen, dass G nichtnegativ ist (h sollte ja eine Dichte sein, also nichtnegativ), dann ist klar, dass h proportional zu θ sein sollte, d.h.

$$h(z) \propto G(z)g(z)\,,\; z \in \mathbb{R}\,,$$

mit $1/\theta$ als Proportionalitätskonstante. Denn diese Wahl führt zu der idealen Situation

$$\mathbb{E}_g(G(Z)) = \mathbb{E}_h\left(G(Z)\frac{g(Z)}{h(Z)}\right),\ \mathbb{V}_h\left(G(Z)\frac{g(Z)}{h(Z)}\right) = 0\,.$$

Was bedeutet dies? Wir hätten eine einfache Dichte h zu wählen, um damit die Stichprobe zu simulieren. Dazu müssten wir aber $1/\theta$ kennen, eine sinnlose Situation, denn θ wollen wir ja gerade über den Schätzer $\theta_n(Z;h)$ berechnen. Was wir aber lernen, ist, dass die Reduktion der Varianz gelingen sollte, wenn es möglich ist, h „nahezu“ proportional zu Gg zu wählen.

Fassen wir zusammen: Eine gute Wahl von h, um die Varianz zu reduzieren, sollte berücksichtigen:

- $h(z) = 0 \implies g(z) = 0, z \in \mathbb{R}$.
- h sollte nahezu proportional zu Gg sein.
- Die Zufallsgröße mit Dichte h sollte so einfach sein, dass damit leicht zu simulieren ist.
- Die Dichte h sollte einfach auszuwerten sein.

Beispiel 4.26 *Wir wollen die Wahrscheinlichkeit $\gamma := P(\{Z > c\}), c > 0$, schätzen, wobei die Zufallsgröße Z exponentiell verteilt sei mit Parameter $\lambda > 0$. Die Dichte von Z ist $z \longmapsto \chi_{[0,\infty)}\lambda e^{-\lambda z}$, der Erwartungswert $1/\lambda$.*

Algorithm 4.7 Simulation einer „kleinen“ Wahrscheinlichkeit

EIN Parameter λ, c. Iterationsanzahl $N \in \mathbb{N}$.

Schritt 0 $p := 0$.

Schritt 1 Für $k = 0, \ldots, N$

- erzeuge eine mit λ exponentiell verteilte Zufallszahl z;
- falls $z > c$ setze $p := p + 1$.

AUS $\gamma_N := p/N$ Näherung für $\gamma := P(\{Z > c\})$.

Ist der Wert c wesentlich größer als $\mathbb{E}(Z)$, so ist $\{Z > 0\}$ ein seltenes Ereignis. Importance Sampling sollte also genutzt werden, um das seltene Ereignis häufiger zu erzeugen. Wähle dazu die Dichte $h(x) := \chi_{[0,\infty)}(x)\mu e^{-\mu x}$ mit $0 < \mu < \lambda$. Der Likelihood-Quotient ist $L(z; g, h) := \lambda\mu^{-1}e^{(\mu-\lambda)z}$. Damit führen wir den Algorithmus 4.26, wobei wir λ in Schritt 1 durch μ ersetzen.

Die gewählte Dichte h ist wieder eine Exponentialverteilung. Sie ist nicht gewählt im Sinne der oben dargestellten Überlegungen. □

Bemerkung 4.27 *„Stratified Sampling“ ist eine Art lokales „Importance Sampling“: die Dichte der Wahl wird „gezwungen“, in bestimmten Regionen ähnlich zur Ausgangsdichte zu sein.* □

4.6.5 Varianzreduktion durch Bedingung

Wir benötigen bedingte Erwartungswerte. Sei X eine Zufallsgröße auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$ und sei Σ eine Teilsigma-Algebra von $\mathcal{F}$. Dann ist $Z := \mathbb{E}(X|\Sigma)$ eine Zufallsgröße auf dem Wahrscheinlichkeitsraum (Ω, Σ, P) mit

- Z ist messbar bezüglich Σ;
- $\int_\Omega Z\chi_C dP = \int_\Omega X\chi_C dP$ für alle $C \in \Sigma$.

Eine solche Zufallsgröße existiert (unter schwachen Voraussetzungen).

Seien nun X, Y Zufallsgrößen auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$. Dann setzen wir

$$\mathbb{E}(X|Y) := \mathbb{E}(X|\Sigma_Y)$$

wobei Σ_Y die kleinste Sigma-Algebra ist, die alle Urbilder $\overset{-1}{Y}(B)$ enthält, wobei B die meßbaren Mengen im Meßraum des Wertebereichs von Y sind. Damit gilt

$$\begin{aligned} \mathbb{E}(X) &= \mathbb{E}(\mathbb{E}(X|Y)) && (4.23) \\ \mathbb{V}(X) &= \mathbb{E}(\mathbb{V}(X|Y)) + \mathbb{V}(\mathbb{E}(X|Y)) && (4.24) \end{aligned}$$

Als Konsequenz haben wir

$$\begin{aligned} \mathbb{V}(X) &\geq \mathbb{E}(\mathbb{V}(X|Y)) && (4.25) \\ \mathbb{V}(X) &\geq \mathbb{V}(\mathbb{E}(X|Y))) && (4.26) \end{aligned}$$

Aus (4.23) wissen wir, dass $\mathbb{E}(X|Y)$ ein treuer Schätzer für $\mathbb{E}(X)$ ist. Wir können (4.26) zur Varianzreduktion nutzen. In der Anwendung simulieren wir mit Y statt mit X.

Beispiel 4.28 *Eine Variante von Marsaglias Methode kann genutzt werden, die Zahl π zu approximieren. Wir wählen zwei auf $[0,1]$ gleichverteilte Zufallsgrößen U_1, U_2. Dazu setzen wir $V_i := 2U_i - 1, i = 1, 2$, und definieren die Zählgröße*

$$Z = \begin{cases} 1 & \textit{falls } V_1^2 + V_2^2 \leq 1 \\ 0 & \textit{sonst} \end{cases} .$$

Klar, $\mathbb{E}(Z) = \pi/4$. Wir berechnen nun den Erwartungswert der Zufallsgröße Z bedingt auf V_1.

$$\begin{aligned} P(\{V_1^2 + V_2^2 \leq 1 | V_1 = v\}) &= P(\{v^2 + V_2^2 \leq 1 | V_1 = v\}) \\ &= P(\{V_2^2 \leq 1 - v^2\}) \ \ (\textit{da } V_1, V_2 \textit{ unabhängig sind}) \\ &= P(\{-\sqrt{1-v^2} \leq V_2 \leq \sqrt{1-v^2}\}) \\ &= \frac{1}{2}\int_{-\sqrt{1-v^2}}^{\sqrt{1-v^2}} dx \ \ (\textit{da } V_2 \textit{ gleichverteilt in } [-1,1] \textit{ ist}) \\ &= \sqrt{1-v^2} \end{aligned}$$

Also gilt

$$\mathbb{E}(Z|V_1) = \int_1^1 \sqrt{1-v^2}\frac{1}{2}dv = \int_0^1 \sqrt{1-v^2}dv = \pi/4 = \mathbb{E}\left(\sqrt{1-U_2^2}\right) .$$

Dies bedeutet, dass die Simulation mit $U := \sqrt{1-U_2^2}$ ein erwartungstreuer Schätzer ist. Die Varianz von U berechnet sich nach

$$\mathbb{V}(U) = \mathbb{E}(1-U_2^2) - (\pi/4)^2 = \frac{2}{3} - (\pi/4)^2 \approx 0.0498 .$$

Andererseits hat die Bernoulli-Zufallsgröße Z den Erwartungswert $\pi/4$ und Varianz

$$\mathbb{V}(Z) = \left(\frac{\pi}{4}\right)\left(1 - \frac{\pi}{4}\right) \approx 0.1686 .$$

Also haben wir beim Übergang von Z zu U um 70.44% die Varianz reduziert. □

4.7 Hochdimensionale Integration

Hier skizzieren wir die Analyse der Verfahren, die wir im obigen Überblick angegeben haben. Die Monte Carlo-Integration ist schon skizziert, also haben wir uns nun mit mehr oder minder deterministischen Vorgehensweisen zu beschäftigen.

4.7.1 Gitter-Verfahren/Produktformeln

Unter Produktformel werden Quadraturformeln verstanden, die dadurch entstehen, dass eine eindimensionale Quadraturformel „tensoriert“ wird. Wir beschreiben nur den einfachen Fall der äquidistanten Zerlegung in jeder Dimension. Seien also $0 < h < 2h < \cdots < nh = 1$ Stützstellen in $[0,1]$. Dann lautet die Sehnentrapezregel

$$Q_{n,st}(f) := \sum_{i=0}^{n}{}' \, hf(ih) .$$

Als Quadraturformel für das Integral in $\mathbb{R}^d$ ergibt sich

$$Q^d_{n,st}(f) = \sum_{i_1=0}^{n}{}' \cdots \sum_{i_d=0}^{n}{}' h^d f(i_1 h, \ldots, i_d h) \tag{4.27}$$

Als Fehlerabschätzung erhält man (mit der Fehlerdarstellung der eindimensionalen Quadraturformel)

$$|I^d(f) - Q^d_{n,st}(f)| \le cn^{-2} , \tag{4.28}$$

falls f zweimal stetig differenzierbar ist.

Zur Auswertung der Quadraturformel ist ein Aufwand von $O(n^d)$ zu betreiben, während die Fehlerschranke dimensionsunabhängig $O(n^{-2})$ ist. Das Verhältnis von Ertrag zu Aufwand wird also für $d \to \infty$ immer schlechter. Man fasst diese Beobachtung unter dem Stichwort **Fluch der Dimension/curse of dimensionality** zusammen.

Ein weiterer Nachteil ist, dass die Gitter bei Verkleinerung der Gittermaschenweite im Allgemeinen neu berechnet werden müssen.

4.7.2 Quasi-Monte Carlo-Integration

Hier wollen wir die Integration durch eine Quadraturformel annähern, die statt (Pseudo)-Zufallszahlen so genannte **Quasizufallszahlen** verwenden. Das Kriterium bei Erzeugung dieser Zahlen ist das „maximale Ausweichen“ dieser Zahlen voneinander. Es existieren verschiedene Verfahren zur Erzeugung solcher Zahlenfolgen, die nach den Entwicklern benannt sind, etwa: Hammersly, Halton, van der Corput, Faure, Sobol, Niederreiter, Weyl. Nach dieser Aufzählung ist nun auch klar, dass es Quasi-Zufallszahlen gibt; siehe unten.

Die Diskrepanz definiert ein Maß für die gleichmäßige Verteilung einer endlichen Menge von Punkten $x^1, \ldots, x^N$. Es sei $Q \subset [0,1]^d$ ein Quader und seien $x^1, \ldots, x^N \in [0,1]^d$. Die Idee hinter der Definition der Diskrepanz ist, dass bei einer gleichmäßig verteilten Punktmenge, die Anzahl der Punkte in Q gerade dem Volumen von Q entspricht.

Definition 4.29

(a) Die **Diskrepanz** *einer Menge* $X := \{x^1, \ldots, x^N\} \subset [0,1]^d$ *ist*

$$D(X) := \sup_{Q \text{ Quader in } [0,1]^d} \Big| \frac{\#\{i | x_i \in Q\}}{N} - vol(Q) \Big|$$

(b) *Bezieht man das Supremum auf Quader der Form* $[0,y_1)\times\cdots\times[0,y_d)$, *so bezeichnet man die resultierende Größe* $D^*(X)$ *als* **Stern-Diskrepanz**.

(c) *Für eine unendliche Menge* $X=\{x^1,x^2,\dots\}$ *nutzen wir die Bezeichnung*

$$D_N := D(\{x^1,x^2,\dots,x^N\})\,,\; N\in\mathbb{N}$$

bzw.

$$D_N^* := D^*(\{x^1,x^2,\dots,x^N\})\,,\; N\in\mathbb{N}\,.$$

□

Hat man eine Folge von Punkten $X:=\{x^1,x^2,\dots\}\subset[0,1]^d$, so sind diese Punkte x^i je gleichmäßiger in $[0,1]^d$ verteilt, desto näher die Diskrepanz D_N bei 0 liegt. Das Kriterium

$$\lim_N D_N = 0$$

wird als Definition für eine **gleichmäßig verteilte Menge** von Punkten in $[0,1]^d$ angesehen. Offensichtlich ist die Stern-Diskrepanz etwas leichter zu untersuchen.

Regel 4.30 *Sei* $X:=\{x^1,x^2,\dots\}\subset[0,1]^d$. *Ohne Beweis geben wir an:*

$$D_N, D_N^* \;\geq\; 0 \tag{4.29}$$

$$D_N^* \;\leq\; D_N \leq 2^d D_N^* \tag{4.30}$$

$$D_N^* \;\geq\; 1/2N \text{ für } d=1 \tag{4.31}$$

Definition 4.31 *Eine Menge* $X:=\{x^1,x^2,\dots\}\subset[0,1]^d$ *heißt von* **niederer Diskrepanz**, *wenn*

$$D_N \leq C_d\frac{(\ln N)^d}{N}$$

gilt mit einer von N *unabhängigen Konstante* $C_d\in[0,\infty)$. □

Der Faktor $1/N$ in der obigen Schranke zeigt schnellere Konvergenz als $1/\sqrt{N}$, wie wir ihn im Zusammenhang mit der Monte Carlo-Simulation kennen. Allerdings haben wir noch den Faktor $(\ln N)^d$ zu berücksichtigen. Er wächst aber nur bescheiden und insgesamt haben wir

$$\left(\frac{(\ln N)^d}{N}\right)\cdot\left(\frac{1}{\sqrt{N}}\right)^{-1} = o(1) \text{ für } N\to\infty\,.$$

Beispiel 4.32 *Sei* $N\in\mathbb{N}, N\geq 2$. *Wir definieren die Menge*

$$M_N := \{1/N, 2/N, \dots, N/N\}\,.$$

M_N *hat die Diskrepanz* $D(M_N)=1/N$. *Der Nachteil der Zahlen aus* M_N *ist, dass beim Übergang von* N *zu* $N+1$ *die Menge* M_{N+1} *völlig neu berechnet werden muss.* □

Beispiel 4.33 *Sei* $m\in\mathbb{N}, m\geq 2$. *Seien die Zahlen* $M_m:=\{u_0,u_1,\dots\}\in[0,1]$ *das Ergebnis eines affin-linearen Zufallsgenerators mit Modul* m. *Dann gilt* $M_m\subset\{0,1/m,\dots,(m-1)/m\}$. *Für die Diskrepanz erhalten wir*

$$D_N \geq \Big|\frac{\#\{i|x_i\in Q\}}{N} - vol(Q)\Big| = vol(Q) = \frac{1}{2m+2}\,,$$

wobei $Q:=[1/(2m+2), 1/(m+1)]$. *Also trifft* $\lim_N D_N = 0$ *nicht zu.* □

Für die Fehlerschranke der Quasi-Monte Carlo-Integration benötigen wir das passende Maß für die Variation einer Funktion f. Diese wollen wir nun skizzieren.

Betrachte in $\mathbb{R}^d$ Quader

$$Q = [u_1^-, u_1^+] \times \cdots \times [u_d^-, u_d^+]$$

mit $0 \leq u_i^- \leq u_i^+ \leq 1, i = 1, \ldots, d$. Jeder dieser Quader hat Ecken, die die Koordinaten $u_i^\pm$ haben. Wir setzen

$$\mathcal{E}(Q) := \{\text{Ecken von } Q | \text{Koordinaten haben gerade viele } + \text{–Bezeichnungen}\},$$

$$\mathcal{O}(Q) := \{\text{Ecken von } Q | \text{Koordinaten haben ungerade viele } - \text{–Bezeichnungen}\}.$$

Definiere damit

$$\Delta(f; Q) := \sum_{u \in \mathcal{E}(Q)} f(u) - \sum_{u \in \mathcal{O}(Q)} f(u).$$

Sei $\mathcal{P}$ die Menge aller disjunkter Zerlegungen des Einheitskubus $[0,1]^d$ in Quader der Form Q. Wir setzen

$$V^{(d)}(f) := \sup_{P \in \mathcal{P}} \sum_{Q \in P} |\Delta(f; Q)|$$

Für jedes $1 \leq k \leq d$ und alle $1 \leq i_1 < i_2 < \cdots < i_k \leq d$ betrachte die Restriktion von f auf $[0,1]^k$ in folgendem Sinne:

$$f(u_1, u_2, \ldots, u_j, \ldots, u_d) \text{ mit } u_j = 1 \text{ falls } j \notin \{i_1, \ldots, u_k\}.$$

Damit wird durch $V^{(k)}(f, i_1, \ldots, i_k)$ die Anwendung von $V^{(k)}$ auf diese Restriktion erklärt. Schließlich definieren wir:

$$V(f) := \sum_{k=1}^{d} \sum_{1 \leq i_1 < \cdots i_k \leq d} V^{(k)}(f, i_1, \ldots, i_k). \tag{4.32}$$

$V(f)$ ist die **Variation von Hardy-Krause**.

Satz 4.34 (Koksma-Hlawka-Ungleichung) *Sei $f : [0,1]^d \longrightarrow \mathbb{R}$ und $x^1, \ldots, x^N \in [0,1]^d$. Dann gilt:*

$$\Big| \frac{1}{N} \sum_{i=1}^{N} f(x^i) - \int_{[0,1]^d} f(x)\, dx \Big| \leq V(f) D_N^*(x^1, \ldots, x^N). \tag{4.33}$$

Beweis:
Für den Beweis siehe [26]. ∎

Vergleichen wir das Resultat „Koksma-Hlawka-Ungleichung“ mit der Monte Carlo-Integration.

- Die Abschätzung (4.33) ist eine deterministische Abschätzung, während in der Monte Carlo-Integration nur probabilistische Schranken vorliegen.
- Will man das Resultat von Koksma-Hlawka als Integrationsverfahren umdeuten, so sollten die Punkte der Quadraturformel so gewählt werden, dass sie kleine Stern-Diskrepanz besitzen.
- Die Größen $V(f), D_N^*(x^1, \ldots, x^N)$ sind schwer zu berechnen, während, wie wir schon festgestellt haben, die Konstante σ_f gut zu approximieren ist.

Die negative Einschätzung der Berechenbarkeit von $V(f)$ ist abzuschwächen, denn $V^{(d)}(f)$ hat die Darstellung

$$V^{(d)}(f) = \int_{[0,1]} \cdots \int_{[0,1]} \Big| \frac{\partial^d f}{\partial u_1 \cdots \partial u_d} \Big| du_1 \cdots du_d \,, \tag{4.34}$$

wenn die partielle Ableitung von f, wie sie in (4.34) vorkommt, existiert und stetig ist.

Bemerkung 4.35 *Die Abschätzung* (4.33) *ist scharf in folgendem Sinne: Für alle $x^1, \ldots, x^N$ und für alle $\varepsilon > 0$ gibt es $f \in C^\infty([0,1]^d)$ mit $V(f) = 1$ und*

$$\Big| \frac{1}{N} \sum_{i=1}^{N} f(x^i) - \int_{[0,1]^d} f(x)\, dx \Big| > V(f) D_N^*(x^1, \ldots, x^N) \,.$$

□

Kommen wir nun zu Beispielen für Quasizufallszahlen.

Van-der-Corput Folgen, werden mit der Dualentwicklung natürlicher Zahlen erzeugt, und zwar durch Bit-Umkehr. Sei also

$$i = (d_j \ldots d_0)_2 = \sum_{k=0}^{j} d_k 2^k$$

die Dualdarstellung von $i \in \mathbb{N}$. Dann heißt

$$\Phi_2(i) := x^i := (.d_0 \ldots d_j)_2 = \sum_{k=0}^{j} d_k 2^{-k-1}$$

die **i-te van der Corput-Zahl**. Beispielsweise sind

$$\frac{1}{2}, \frac{1}{4}, \frac{3}{4}, \frac{1}{8}, \frac{5}{8}, \frac{3}{8}$$

die ersten 6 van der Corput-Zahlen. Der Vorteil gegenüber den Zahlen, die wir in den Beispielen 4.32, 4.33 betrachtet haben, ist, dass bereits berechnete Zahlen immer mitverwendet werden können. Klar, die Basis $b = 2$ lässt sich gegen jede beliebige Basiszahl $b \in \mathbb{N}, b \geq 2$, austauschen.[3] Alle diese van der Corput-Zahlen lassen sich algorithmisch einfach durch Division mit Rest bestimmen. Sie entsprechen also einer Liste von Zahlen, die total den Anspruch der Zufälligkeit verloren haben. Was sie aber auszeichnet, ist die Tatsache der niedrigen Diskrepanz. Es gilt nämlich, wenn wir die Basis b zu Grunde legen:

$$\limsup_N \frac{N D_N}{\log N} = \begin{cases} \dfrac{b^2}{4(b+1)\log b} & \text{, falls } b \text{ gerade} \\[2ex] \dfrac{b-1}{4 \log b} & \text{, falls } b \text{ ungerade} \end{cases} \tag{4.35}$$

Den Beweis dieser Tatsache findet man in [26].

Die Konstruktion der van der Corput-Zahlen kann man nun nutzen, um Folgen in $[0,1]^d$ zu erzeugen. Dazu wähle man für jede Dimension j eine Basis b_j, erzeuge damit die van der Corput-Folge $(x_i^j)_{i\in\mathbb{N}}$. Damit bilde man dann die Vektoren

$$x^i := (x_1^i, \ldots, x_d^i) \in [0,1]^d \,.$$

Im Allgemeinen nimmt man als Basen die ersten d Primzahlen.

Diese so konstruierte Folge von Punkten nennt man eine Folge von **Halton-Punkten**. Die Eigenschaft der niedriger Diskrepanz überträgt sich von den van der Corput-Zahlen auf die Halton-Punkte.

[3] van der Corput (1935) hat sie für die Basis 2 als erster betrachtet.

4.7.3 Dünne Gitter

Siehe [10] für einen Überblick.

4.8 Anhang: Korreliert verteilte Zufallszahlen

Bei der Simulation einer mehrdimensionalen Brownschen Bewegung benötigt man im Allgemeinen Zufallsgrößen, die einer korrelierten mehrdimensionalen Verteilung folgen.

Reellwertige korrelierte Zufallsgrößen und damit abgeleitete Pseudozufallszahlen lassen sich recht einfach erzeugen. Seien X_1, X_2 unabhängige Zufallsgrößen und sei ρ die Korrelation. Dann erhält man durch

$$Z_1 := X_1\,,\ Z_2 := \rho X_1 + \sqrt{1-\rho^2}X_2$$

zwei entsprechend korrelierte Zufallsgrößen. Im vektorwertigen Fall betrachten wir nur den Fall der Normalverteilung.

Definition 4.36 *Ein Zufallsvektor $X = (X_1, \ldots, X_d)$ heißt $\mathcal{N}(\mu, \Sigma)$–verteilt, wenn X die Dichte f mit*

$$f(x) = \frac{1}{(2\pi)^n \det(\Sigma)} \exp(-\frac{1}{2}\langle x-\mu, \Sigma^{-1}(x-\mu)\rangle\,,\ x \in \mathbb{R}^d,$$

besitzt. Dabei ist $\mu \in \mathbb{R}, \Sigma$ eine symmetrische positiv definite Matrix in $\mathbb{R}^{d,d}$. □

In Definition 4.36 heißt $\Sigma = (\Sigma_{ij})_{i,j=1,\ldots,d}$ **Kovarianz-Matrix**. Wir haben

$$\Sigma_{ij} = \mathbb{E}(\langle X_i - \mu_i, X_j - \mu_j\rangle)\,,$$

wobei $\mu = (\mu_1, \ldots, \mu_d) = (\mathbb{E}(X_1), \ldots, \mathbb{E}(X_d))$ der Erwartungsvektor von X ist. Die Matrix $S = (S_{ij})_{i,j=1,\ldots,d}$ mit den Einträgen

$$S_{ij} := \frac{\Sigma_{ij}}{\sqrt{\Sigma_{ii}\Sigma_{jj}}}$$

heißt die **Korrelation**.

Seien $Z_1, \ldots, Z_d$ unabhängige auf $[0,1]$ standard-normalverteilte Zufallsgrößen. Wir setzen $Z := (Z_1, \ldots, Z_d)$. Sei f die Dichte von Z. Sie ist offenbar

$$f(x) := \frac{1}{\sqrt{(2\pi)^n}} \exp(-\frac{1}{2}\langle x, x\rangle)\,,\ x \in \mathbb{R}^d\,. \tag{4.36}$$

Das Ziel ist nun, eine Zufallsgröße zu konstruieren, die $\mathcal{N}(\mu, \Sigma)$–verteilt ist. Seien also $\mu \in \mathbb{R}^d$ und $\Sigma \in \mathbb{R}^{d,d}$ vorgegeben. Wir setzen voraus, dass Σ eine symmetrische positiv definite Matrix in $\mathbb{R}^{d,d}$ ist. Wir zerlegen Σ nach Cholesky

$$\Sigma = LL^T$$

mit einer Matrix $L = (L_{ij})_{i,j=1,\ldots,d}$, die eine untere Dreiecksmatrix ist. Damit definieren wir den Zufallsvektor G und die Zufallsgröße Y durch

$$Y := G \circ Z \text{ mit } G(x) := \mu + Lx\,,\ x \in \mathbb{R}^d\,.$$

Satz 4.8 ist anwendbar mit $M := \mathbb{R}^d$. Offenbar ist $DG(z) = L$ und $\det DG(z) = \sqrt{\det(\Sigma)}, z \in \mathbb{R}^d$. Da Σ positiv definit ist, ist L regulär und G ist invertierbar auf ganz $\mathbb{R}^d$. Wir haben $G^{-1}(y) = L^{-1}(y-\mu)$.

Folgerung 4.37 *Der Zufallsvektor* $Y := \mu + LZ$ *ist* $\mathcal{N}(\mu, \Sigma)$*-verteilt.*

Beweis:
Aus Satz 4.8 wissen wir über die Dichte g von Y:

$$g(y) = \frac{1}{\sqrt{\det(\Sigma)}} f(L^{-1}(y-\mu)) \, , \, y \in \mathbb{R}^d \, .$$

Daraus liest man die Behauptung mit der Darstellung von f aus (4.36) ab. ■

Beispiel 4.38 *Wir wollen einen normalverteilten Zufallsvektor mit dem Nullvektor als Erwartungsvektor und mit Kovarianz-Matrix*

$$\Sigma := \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

mit $|\rho| \leq 1, \sigma_1 > 0, \sigma_2 > 0$ *konstruieren.*

Mit dem Ansatz

$$L = \begin{pmatrix} a & 0 \\ b & c \end{pmatrix}$$

liefert die Zerlegung $\Sigma = LL^T$ *durch Koeffizientenvergleich die Zahlen* a, b, c *und wir erhalten*

$$L = \begin{pmatrix} \sigma_1 & 0 \\ \rho\sigma_2 & \sigma_2\sqrt{1-\rho^2} \end{pmatrix} \, .$$

Sind also Z_1, Z_2 *unabhängige und standard-normalverteilte Zufallsgrößen, dann stellt*

$$X := (X_1, X_2) := (\sigma_1 Z_1, \sigma_2(\rho Z_1 + \sqrt{1-\rho^2} Z_2))$$

einen normalverteilten Zufallsvektor dar mit Erwartungsvektor $(0,0)$ *und Kovarianz-Matrix* Σ *.* □

4.9 Anhang: Beweis des Dichtetransformationssatzes

Satz 4.39 *Sei* X *eine Zufallsgröße auf dem Wahrscheinlichkeitsraum* (Ω, Σ, P) *mit Verteilung* P^X *und Dichte* f *. Ferner sei* $G : \mathbb{R}^d \longrightarrow \mathbb{R}^d$ *ein Zufallsvektor, wobei auf* $\mathbb{R}^d$ *die Sigmaalgebra* $\mathcal{B}_d$ *der Borelschen Mengen mit dem Lebesguemaß* λ_d *zugrundeliege. Sei* $M \subset \mathbb{R}^d$ *eine offene, zusammenhängende Menge, so dass die folgenden Bedingungen gelten:*

(a) *Für* $x \notin M$ *ist* $f(x) = 0$ *.*

(b) G *ist stetig differenzierbar auf* M *und es ist* $\det DG(x) \neq 0$ *für alle* $x \in M$ *.*

(c) *Ist* $N = G(M)$ *das Bild von* M *unter der Abbildung* G*, so ist die Abbildung* $G : M \longrightarrow N$ *bijekiv mit der Umkehrabblidung* H *.*

Dann besitzt die Verteilung P^G *die Dichte*

$$g(y) := \begin{cases} f(H(y)) |\det DH(y)| & , \textit{falls } y \in N \\ 0 & , \textit{sonst} \end{cases} \tag{4.37}$$

Beweis:
Es ist zu zeigen, dass mit der Funkton g aus (4.37) gilt:

$$P^G(B) = \int_{\mathbb{R}^d} \chi_B(y)g(y)dy\,,\; B \in \mathcal{B}_d\,.$$

Beachte, dass

$$|\det DG(H(y))|^{-1} = |\det DH(y)|$$

für $y \in N$ gilt.

Sei also $B \in \mathcal{B}_d$ gegeben. Zunächst zerlegen wir die Menge B folgendermaßen:

$$B = B \cap N \cup B \cap \mathbb{R}^d \backslash N =: B_1 \cup B_2\,.$$

Da B_2 disjunkt zu N ist, muss das Urbild $\{G \in B_2\}$ ganz im Komplement $\mathbb{R}^d \backslash M$ von M liegen, denn aus $x \in M$ würde $G(x) \in N$ folgen. Da $f(x) = 0$ ist für alle $x \in N$ und die Funktion g aus (4.37) auf $\mathbb{R}^d \backslash N$ verschwindet, ist (siehe Regel (4.7))

$$P^G(B_2) = P^X(\{G \in B_2\}) = \int_{\mathbb{R}^d} \chi_{\{G \in B_2\}}(x)f(x)dx = 0 = \int_{\mathbb{R}^d} \chi_{B_2}(y)g(y)dy\,.$$

Da die Abbildung G nur auf M und nicht notwendigerweise auf dem ganzen $\mathbb{R}^d$ umkehrbar ist, gibt es eventuell Punkte $x \notin M$, die durch G in die Menge B_1 abgebildet werden:

$$\{G \in B_1\} = \{x \in M | G(x) \in B_1\} \cup \{x \notin M | G(x) \in B_1\} =: U_1 \cup U_2$$

mit

$$P^G(B_1) = P^X(\{G \in B_1\}) = P^X(U_1) + P^X(U_2)\,.$$

Wegen $U_2 \subset \mathbb{R}^d \backslash M$ ist

$$P^X(U_2) = \int_{\mathbb{R}^d} \chi_{U_2}(x)f(x)dx = 0\,.$$

Die Funktion G eingeschränkt auf die Menge U_1 als Definitionsbereich ist eine Parametrisierung der Menge B_1, denn U_1 ist das Urbild von B_1, G ist auf M umkehrbar und erfüllt alle Voraussetzungen für den Substitutionssatz für Integrale. Daher haben wir

$$P^G(B_1) = P^X(U_1) = \int_{\mathbb{R}^d} \chi_{U_1}(x)f(x)dx = \int_{\mathbb{R}^d} \chi_{B_1}(y)g(y)dy\,.$$

Zusammen ergibt dies

$$P^G(B) = P^G(B_1) + P^G(B_2) = \int_{\mathbb{R}^d} (\chi_{B_1}(y) + \chi_{B_2}(y))g(y)dy = \int_{\mathbb{R}^d} \chi_B(y)g(y)dy\,.$$

$\blacksquare$

4.10 Bibliographische Anmerkungen

Die Erzeugung von Zufallszahlen beliebiger Verteilung wird zum Beispiel in [8, 19] betrachtet; eine Übersicht findet man in [7]. Der Spezialfall von normalverteilten Zufallszahlen hat großes Interesse gefunden, was seinen Grund in der Approximation der Optionspreise, beschrieben durch Erwartungswerte der Kursentwicklung eines Basisobjekts entlang einer stochastischen Differentialgleichung. Zur Realisierung unterschiedlicher Verteilungen findet man Informationen etwa in [4, 11, 18, 20, 22].

Numerische Überlegungen zu der Approximation der kummulativen Normalverteilung werden z.B. in [1, 25] diskutiert. Zu Fragen der Wegwerf–Methoden, insbesondere zu Vor- und Nachteilen, verweisen wir auf [9, 11].

Klassische Quellen zu Monte Carlo-Methoden sind [14, 11, 29]. Etwas spezieller Untersuchungen findet man in [12, 23, 15, 30]. Zu Gitter-Methoden in der numerischen Integration siehe etwa [3, 6, 13, 27, 28, 31, 32, 33, 34]. Zur Diskrepanz bestimmter Folgen findet man etwas in [2] und [26].

Die Fragen der Varianzreduktion werden diskutiert u.a. in [11, 17, 30].

4.11 Übungen

1.) Sei die Zufallsgröße U gleichverteilt auf $[0,1]$. Gib eine Zufallsgröße X mit Wahrscheinlichkeitsverteilung F an, wobei X die Dichte

$$f(x) := \begin{cases} \frac{1}{x^2} & x \in [1,\infty) \\ 0 & \text{sonst} \end{cases}$$

habe.

2.) Die Weibull-Verteilung hat die Dichte $f(x) := ax^{a-1}e^{-x^a}, x > 0$, mit $a > 0$. Berechne die Verteilungsfunktion und die Quantilfunktion.

3.) Die Zufallsgröße habe die Dichte $f(x) := \frac{1}{2}x\chi_{[0,2]}(x)$, $x \in \mathbb{R}$. Welche Dichte hat die Zufallsgröße $Y := X(2-X)$?

4.) Die Zufallsgröße habe die Dichte $f(x) := e^{-x}, x \in \mathbb{R}$. Welche Dichte hat $Y := X^{\frac{1}{2}}$?

5.) Sei die Zufallsgröße U gleichverteilt auf $[0,1]$. Gib eine Zufallsgröße X mit Wahrscheinlichkeitsverteilung F an, wobei X die Dichte

$$f(x) := \frac{\gamma}{\pi}\frac{1}{(x-\mu)^2+\gamma^2}\,,\ x \in \mathbb{R}\,,$$

habe.

6.) Seien die Zufallsgrößen unabhängige und mit dem Parameter λ exponentiell verteilt. Welche Verteilung besitzt die Zufallsgröße $Y := (Y_1, Y_2)$ mit $Y_1 := X_1 + X - 2$ und $Y_2 := X_1/X_2$? Sind Y_1, Y_2 unabhängige Zufallsgrößen?
Hinweis: Nutze den Dichte-Transformationssatz.

7.) Seien X, Y identisch verteilte Zufallsgrößen auf $\Omega := \{1,\dots,n\}, n \geq 2$, also

$$P(\{X = i\}) = 1/n, 1 \leq i \leq n\,.$$

Ist $Z := X + Y$ auf $\{2,\dots,2n\}$ gleichmäßig verteilt? Begründung.

8.) Um die Verteilung der Summe $Y := X_1 + X_2$ zweier Zufallsgrößen zu berechnen, ergänzt man die Funktion $G_1(x_1,x_2) := x_1 + x_2$ durch die Komponente $G_2(x_1,x_2) := x_2$ und wendet den Dichtetransformationssatz an. Führe dieses „Programm“ für die unabhängigen Zufallsvariablen X_1, X_2 durch, wenn sie exponentiell verteilt sind mit Parameter λ in der Dichte

$$f_1(x) := f_2(x) := \begin{cases} \lambda e^{-\lambda x} & \text{, falls } x > 0 \\ 0 & \text{, falls } x \leq 0 \end{cases}$$

9.) Betrachte die Verteilungsfunktion

$$F(x) := \frac{2}{\pi}\arcsin(\sqrt{x})\,,\ 0 \le x \le 1\,, = 0 \text{ für } x < 0, = 1 \text{ für } x > 1\,.$$

Konstruiere eine Zufallsgröße X mit Verteilung F unter Verwendung einer auf $[0,1]$ gleichverteilten Zufallsgröße U.

10.) Betrachte die Verteilungsfunktion

$$F(x) := (1 - e^{-2x(x-b)})\,,\ x \ge b\,, = 0 \text{ sonst}\,.$$

Konstruiere eine Zufallsgröße X mit Verteilung F unter Verwendung einer auf $[0,1]$ gleichverteilten Zufallsgröße U.

11.) Bei N-maliger Wiederholung eines Bernoulli-Experiments (Erfolgswahrscheinlichkeit $q \in [0,1]$) ist die Wahrscheinlichkeit, k-mal Erfolg zu beobachten,

$$b(N,k) = \binom{N}{k} q^k (1-q)^{N-k}\,.$$

Finde eine Rekursionsformel bezüglich k zur Berechnung von $b(N,k)$.

12.) Betrachte die Cauchy-Verteilung. Ihre Dichte f ist

$$f(x) := \frac{1}{\pi}\frac{\gamma}{(x-\mu)^2 + \gamma^2}\,,\ x \in \mathbb{R}\,.$$

Was lässt sich über den Erwartungswert und die Varianz sagen?

13.) Betrachte eine Zufalllsgröße X mit Dichte $f(x) := \max(0, 1 - |x|)\,,\ x \in \mathbb{R}$.

(a) Berechne die Verteilungsfunktion F von X.

(b) Berechne die Quantilfunktion F^{-1}.

(c) Zeige: $F \circ X$ ist gleichverteilt auf $[0,1]$.

14.) Eine radioaktive Probe wirft Partikel in zufällige Richtungen aus. Dieser Probe wird im Abstand d ein horizontaler photographischer Schirm (x-Gerade) gegenübergestellt; die Probe hat in der $x-y$–Ebene die Koordinaten $(0,d)$. Die Partikel-Konzentration, die im Winkel $\Phi(\omega)$ auf den Schirm fällt, werde mit $X(\omega)$ bezeichnet; also $\tan(\Phi(\omega)) = X(\omega)/d$. Der Winkel Φ kann als gleichverteilte Zufallsgröße auf $[-\pi/2, \pi/2]$ angesehen werden.

(a) Berechne die Verteilungsfunktion von X.

(b) Berechne die Dichte von X.

15.) X und Y seien zwei Zufallsgrößen mit $\mu = \mathbb{E}(X), \nu = \mathbb{E}(Y), \sigma^2 := \mathbb{V}(X) > 0, \tau := \mathbb{V}(Y) > 0, \sigma_{X,Y} := \mathrm{Cov}(X,Y)$. Die beiden (abgeleiteten) Zufallsgrößen U und V seien definiert durch

$$U := 7 - 3X + 5Y\,,\ V := X - 2Y - 4\,.$$

(a) Berechne die Erwartungswerte und Varianzen von U und V sowie die Kovarianz von U, V.

(b) Sind U, V unabhängig?

16.) Es gilt

$$\pi = \frac{8}{\ln(2)}\int_0^{\ln(2)} \frac{x}{e^x + 2e^{-x} - 2}dx$$

Berechne damit π mit einer Monte Carlo-Simulation.

17.) Seien Z_1, Z_2 zwei unabhängige, standard-normalverteilte Zufallsgrößen. Berechne die Dichte von $\max\{Z_1, Z_2\}$.

18.) Sei $F : \mathbb{R} \longrightarrow \mathbb{R}$ die Verteilungsfunktion der Zufalllsgröße X .
Zeige: Ist F stetig, so ist $F \circ X$ eine gleichverteilte Zufalllsgröße.

19.) Berechne das Integral

$$I := \int_{[0,1]^5} \exp(-x_1x_2x_3x_4x_5)dx_1 \ldots dx_5$$

mit der Monte Carlo-Simulation, wobei 1000 gleichverteilte Punkte verwendet werden sollen.

20.) Wir sagen, dass eine Zufallsgröße X nach $\mathrm{logist}(a, b)$ verteilt ist, wenn X die Verteilungsfunktion

$$F(x) := \frac{1}{1 + e^{\frac{x-a}{b}}} \, , \; x \in \mathbb{R} \, ,$$

hat.
Zeige: Ist U gleichverteilt auf $[0, 1]$, dann ist $X := \ln(U) - \ln(1 - U)$ nach $\mathrm{logist}(0, 1)$ verteilt.

21.) Betrachte das Monte Carlo-Verfahren zur Berechnung des Integrals $\int_0^1 f(x)dx$ mittels $\hat{I} := \frac{1}{N}\sum_{i=1}^N f(x^i)$ und Invarianz $\sigma_I^2 = \frac{1}{N-1}\sum_{i=1}^N (f(x^i)^2 - N\hat{I}^2)$.

(a) Zeige für die Iteration

$$\alpha_i := \alpha_{i-1} + \frac{1}{i}(f(x^i) - \alpha_{i-1}) \, , \; \beta_i := \beta_{i-1} + \frac{i-1}{i}(f(x^i) - \alpha_{i-1})^2 \, ,$$

dass gilt: $\hat{I} = \alpha_N, \sigma_T^2 = \frac{1}{N-1}\beta_N$.

(b) Warum ist diese iterative Vorgehen stabiler als eine direkte Berechnung von $\hat{I}, \sigma_I^2$?

22.) Seien $X_1, \ldots, X_n$ reellwertige, identisch verteilte Zufallsgrößen auf dem Wahrscheinlichkeitsraum (Ω, Σ, P) ; F sei die Verteilungsfunktion dieser Zufallsgrößen. Beweise:

(a) $P(\{\max(X_1, \ldots, X_n)\} \le x) = F(x)^n \, , \; n \in \mathbb{N}$.

(b) $P(\{\min(X_1, \ldots, X_n)\} \le x) = 1 - (1 - F(x))^n \, , \; n \in \mathbb{N}$.

Was ist die Dichte von $X := \max(X_1, \ldots, X_n)$, falls $X_1, \ldots, X_n$ gleichverteilt auf $[0, 1]$ sind.

23.) Erzeuge mit Hilfe der Acceptance-Rejection-Methode Zufallszahlen, die nach der Verteilungsfunktion

$$F(x) := \begin{cases} 0 & x \le 0 \\ \frac{1}{2}x^5 + \frac{1}{2}x^4 & \\ 1 & x \ge 1 \end{cases}$$

verteilt sind unter Nutzung der Gleichverteilung.

24.) Erzeuge mit Hilfe der Acceptance-Rejection-Methode Zufallszahlen, die nach der Verteilungsdichte

$$f(x) := \begin{cases} 0 & x \le 0 \\ 30(x^2 - 2x^3 + x^4) & 0 < x < 1 \\ 1 & x \ge 1 \end{cases}$$

verteilt sind unter Nutzung der Gleichverteilung.

25.) Sei U gleichverteilt auf $[0,1]$. Betrachte das Paar $(U,\tilde{U})$ mit $\tilde{U} := (\frac{1}{3} - U) \mod 1$ bzw. $\tilde{U} := (\frac{2}{3} - U) \mod 1$. Sind die Variablen $U,\tilde{U}$ identisch verteilt und sind sie negativ korreliert?

26.) Berechne den Wert des Integrals $\int_0^1 e^x dx$ mit der Monte Carlo-Methode, wobei 50 Paare der auf $[0,1]$ gleichverteilten antithetischen Variablen $U_1, U_2 := 1 - U_1$ genutzt werden sollen.

27.) Erzeuge mit Marsaglia's Polar-Methode

$$V_i := 2U_i - 1 \text{ solange } W := V_1^2 + V_2^2 < 1\,,\; Z_1 := V_1\sqrt{-2\ln(W)/W}\,,\; Z_2 := V_2\sqrt{-2\ln(W)/W}$$

ausgehend von zwei gleichmäßig auf $[0,1]$ verteilten Zufalllsgrößen U_1, U_2 unter Nutzung des Kongruenzgenerators (Langland) mit

$$a = 142412240584757, b = 11, m = 2^{48}\,,$$

10 000 normal verteilte Pseudozufallszahlen. Plotte dazu ein Histogramm, das die erreichte „Normalverteilung“ illustriert.

28.) Realisiere unter Verwendung der Gleichverteilung f auf $[0,1]$ die Wegwerfmethode für die Beta-Verteilung

$$g(x) := B(\alpha_1,\alpha_2)^{-1} x^{\alpha_1 - 1}(1-x)^{\alpha_2 - 1}\,,\; 0 \le x \le 1\,,$$

wobei

$$B(\alpha_1,\alpha_2) := \int_0^1 x^{\alpha_1-1}(1-x)^{\alpha_2-1}dx = \frac{\Gamma(\alpha_1)\Gamma(\alpha_2)}{\Gamma(\alpha_1+\alpha_2)}\,,\; \alpha_1 \ge 1, \alpha_2 \ge 1$$

ist.

29.) Betrachte mit $N \in \mathbb{N}$

$$I_f := \int_0^1 f(t)dt\,,\; Q_f^N := 1/N \sum_{i=1}^N f(x^i)$$

für $f(t) := 5t^4$. Dabei sind die Stützstellen $x^1, \ldots, x^N$ gleichverteilte Zufallszahlen in $[0,1]$. Berechne unter Nutzung eines Zufallsgenerators eine Näherung Q_f^N für $N = 2^k, k = 2, \ldots, 11$, plotte den Fehler $e_f^N := |I_f - Q_f^N|$ und vergleiche mit dem theoretischen Fehler $c/\sqrt{N}$ (welches c?).

30.) Zeige für die Diskrepanzen einer Folge von Punkten $X := \{x^1, \ldots, x^N, \ldots\} \subset [0,1]^d$:

(a) $D_N \ge 0$

(b) $D_N^* \le D_N \le 2^m D_N^*$

(c) $D_N^* \ge 1/2N$

Literaturverzeichnis

[1] J.D. Bearsley and S.G. Springer. The percentage points of the normal distribution. *Applied Statistics*, 26:118–121, 1977.

[2] P. Boyle, M. Broadie and P. Glassermann. Monte carlo methods for security pricing. *J. Econ. Dyn. Control*, 21:1267–1321, 1997.

[3] H.-J. Bungartz and S. Dirstorfer. Multivariate quadrature on adaptive sparse grids. *Computing*, 71:89–114, 2003.

[4] R.E. Caflisch. Monte Carlo and quasi-Monte Carlo methods. *Acta Numerica*, pages 1–49, 1998.

[5] R.E. Caflisch, W. Morokoff and A. Owen. Valuation of mortage backed securities using Brownian bridges to reduce effective dimension. *J. Comput. Finance*, 1, 1997.

[6] R. Cools, E. Novak and K. Ritter. Smolyak's construction of curbature formulas of arbitrary trigonometric degree. *Computing*, 62:147–162, 1999.

[7] L. Devroye. *Non-uniform random variate generation*. Springer, New York, 1986.

[8] L. Devroye and R. Neininger. Density approximation and exact simulation of random variables which are solutions of fixed-point equations. *Advances of Applied Probability*, 34:441–468, 2002.

[9] G.S. Fishman. *Monte Carlo: Concepts, Algorithms and Apllications*. Springer, New York, 1996.

[10] T. Gerstner and M. Griebel. Numerical integration using sparse grids. *Numerical Algorithms*, 18:209–232, 1998.

[11] P. Glasserman. *Monte Carlo Methods in Financial Engineering*. Springer, Baltimore, 2003.

[12] P. Glasserman, P. Heidelberger and P. Shahabuddin. Asymptotically optimal importance sampling and stratification for pricing path-dependent options. *Mathematical Finance*, 9:117–152, 1999.

[13] M. Griebel and P. Oswald. Tensor-product-type subspace splittings and multilevel iterative methods for anisotropic problems. *Advances of Computational Mathematics*, 4:171–206, 1995.

[14] J.M. Hammersley and D.C. Handscomb. *Monte Carlo Methods*. Methuen & Co. Ltd., London, 1964.

[15] J. Hartinger, R. Kainhofer and R. Tichy. Quasi-monte carlo algorithms for unbounded, weighted integration problems. *Jornal of Complexity*, 20:558–654, 2004.

[16] N. Henze. *Stochastik für Einsteiger*. Vieweg, Braunschweig, 1997.

[17] C. Hickernell, C. Lemieux and A.B. Owen. Control variates for quasi-monte carlo. *Statist. Sci.*, 20:1–31, 2005.

[18] D.J. Higham. *An introduction to financial option valuation*. Cambridge University Press, 2004.

[19] W. Hörmann, J. Leydold and G. Derflinger. *Automatic Nonuniform Random Variate Generation*. Springer, Berlin, 2004.

[20] P. Jäckel. *Monte Carlo Methods in Finance*. Wiley, New York, 2001.

[21] U. Krengel. *Einführung in die Wahrscheinlichkeitstheorie und Statistik*. Vieweg, Braunschweig, 1991.

[22] A.M. Law and W.D. Kelton. *Simulation Modeling and Applications*. McGraw-Hill, Singapore, 2000.

[23] P. L'Ecuyer and C. Lemieux. Variance reduction via lattice rules. *Management Sciences*, 46:1214–1235, 2000.

[24] G. Marsaglia and T.A. Bray. A convenient method for generating normal variables. *SIAM Review*, 6:260–264, 1964.

[25] B. Moro. The full monte. *Risk*, 8:57–58, 1995.

[26] H. Niederreiter. *Random Number Generation and Quasi-Monte-Carlo-Methods*. SIAM, Philadelphia, 1992.

[27] E. Nowak and K. Ritter. High dimensional integration of smooth functions over cubes. *Numerische Mathematik*, 75:79–97, 1996.

[28] C. Reisinger and G. Wittum. Efficient hierarchicak approximation of high-dimensional option pricing. *SIAM J. SCi. Comput*, 29:440–458, 2007.

[29] R.Y. Rubinstein. *Simulation and the Monte Carlo-Method*. Wiley & Sons, New York, 1981.

[30] W. Sandmann. *Simulation seltener Ereignisse mittels importance sampling unter besonderer Berücksichtigung Markovscher Modelle*. PhD thesis, Universität Bonn, 2004.

[31] S.A. Smolyak. Quadrature and interpolation formulas for tensor products of certain classes of functions. *Sov. Mathematics Doklacy*, 4:240–243, 1963.

[32] X. Wang and K.-T. Fang. Effective dimension and quasi-Monte Carlo algorithm. *J. Complexity*, 19:101–124, 2003.

[33] X. Wang and I.H. Sloan. Why are high-dimensional finance problems often of low effective dimension. *SIAM J. Sci. Comput.*, 27:159–183, 2005.

[34] C. Zenger. Sparse grids. Proc. 6th GAMM Seminar, pages 241–251, 1991.

Kapitel 5

Bewertung von Optionen mittels Monte Carlo Methoden

Der Wert einer plain vanilla-Option kann sehr einfach über die Black-Scholes Formel bestimmt werden. Diese Einfachheit ist für komplexere Optionstypen nicht gegeben, schon der Preis der Bermuda-Optionen kann nur mit Hilfe numerischer Verfahren ermittelt werden. Wir besprechen hier die Bewertung allgemeiner Optionen durch stochastische Bewertungsverfahren und nutzen dabei Monte Carlo-Simulation der Kursentwicklung der zugrundeliegenden Basisobjekte.

Wir werden nur wenige Details der Rechtfertigung der stochastischen Differentialgleichung und der daraus resultierenden Kursentwicklung des Basisobjekts darlegen können; wir halten uns im Wesentlichen an die numerische Umsetzung der Optionspreise. Im Anhang skizzieren wir die Vielfalt der Optionen, die einer Bewertung zugänglich gemacht werden sollten.

5.1 Brownsche Bewegung und stochastische Differentialgleichungen

Wir bereiten hier die Simulation des Erwartungswertes einer Auszahlungsfunktion in ihren Bausteinen vor.

5.1.1 Brownsche Bewegung

Definition 5.1 *Ein (kontinuierlicher, zeitlich unbegrenzter) stochastischer Prozess ist eine Familie $(X_t)_{t\geq 0}$ von Zufallsgrößen auf einem Maßraum $(\Omega, \mathcal{F}, P)$ mit Werten im Maßraum*[1] $(\mathbb{R}^d, \mathcal{B}_d)$, *versehen mit dem Lebesgueschen Maß λ_d.*

Er heißt **stetig**, *wenn die Pfade $X(\cdot,\omega) : [0,\infty) \ni t \longmapsto X(t,\omega) := X_t(\omega) \in Z$ stetig für alle $\omega \in \Omega$ sind.* □

Liegt ein stochastischer Prozess $(X_t)_{t\geq 0}$ von Zufallsgrößen auf einem Maßraum $(\Omega, \mathcal{F}, P)$ vor, so kann man dazu die folgende Familie $\hat{\mathcal{F}} := (\mathcal{F}_t)_{t\geq 0}$ dazu betrachten:

$$\mathcal{F}_t := \text{kleinste } \sigma\text{-Algebra, bezüglich der alle } X_u, u \leq t, \text{ meßbar sind}\,.$$

Man sagt, dass $\hat{\mathcal{F}}$ eine **Filtration** ist und dass $(X_t)_{t\geq 0}$ zu $\hat{\mathcal{F}}$ **adaptiert** ist. Umgekehrt heißt ein stochastischer Prozess $(X_t)_{t\geq 0}$ **adaptiert bezüglich einer Filtration** $\hat{\mathcal{F}} = (\mathcal{F}_t)_{t\geq 0}$, wenn jedes X_t meßbar bezüglich $\mathcal{F}_t$ ist.

Eine (übliche) Filtration $\hat{\mathcal{F}} = (\mathcal{F}_t)_{t\geq 0}$ hat folgende Eigenschaften:

[1]Dies ist nicht die allgemeinste Situation, man könnte den Maßraum bis zum topologischen Raum verallgemeinern.

- $\mathcal{F}_t$ σ-Algebra für alle $t \geq 0$;
- $\mathcal{F}_s \subset \mathcal{F}_t \subset \mathcal{F}$;
- $\mathcal{F}_0$ enthält alle Nullmengen (Vollständigkeit);
- $\mathcal{F}_t = \cap_{s>t}\mathcal{F}_t$ (Rechtsstetigkeit).

Eine große Bedeutung hat eine Filtration im Zusammenhang mit Martingalen.

Definition 5.2 *Ein stochastischer Prozess $(X_t)_{t\geq 0}$, der zur Filtration $\hat{\mathcal{F}} := (\mathcal{F}_t)_{t\geq 0}$ adaptiert ist, heißt ein* **Martingal**, *falls gilt:*

$$\mathbb{E}(|X_t|) < \infty, 0 \leq t\,,\; \mathbb{E}(X_t|\mathcal{F}_s) = X_s\,,\; 0 \leq s < t\,.$$

□

Definition 5.3 *Ein* **Standard-Wiener Prozess** $(W_t)_{t\geq 0}$ *ist ein stetiger stochastischer Prozess mit folgenden Eigenschaften:*

(a) $W_0 = 0$ *(mit Wahrscheinlichkeit 1).*

(b) $W_t - W_s$ *ist normalverteilt mit Erwartungswert 0 und Varianz* $t - s$ *für alle* $0 \leq s < t$.

(c) Die Inkremente $W_{t_2} - W_{t_1}, W_{t_4} - W_{t_3}$ *sind unabhängige Zufallsgrößen für alle* $0 \leq t_1 < t_2 \leq t_3 < t_4$.

□

In der Definition (5.3) stecken redundante Informationen.[2] Ferner haben wir den Prozess auf dem Halbstrahl $[0,\infty)$ erklärt, in Anwendungen zieht man sich oft auf ein Intervall $[0,T]$ zurück.

Regel 5.4 *Sei* $(W_t)_{t\geq 0}$ *ein Wiener Prozess. Dann gilt:*

$$\begin{aligned}
\mathbb{E}(W_t - W_s) &= 0 \textit{ für } 0 \leq s < t\,. && (5.1)\\
\mathbb{E}(W_t) &= \mathbb{E}(W_t - W_0) = 0 \textit{ für } t \geq 0\,. && (5.2)\\
\mathbb{V}(W_t - W_s) &= \mathbb{E}((W_t - W_s)^2) = t - s \textit{ für } 0 \leq s < t\,. && (5.3)\\
W_s, W_t - W_s &\textit{ sind unabhängig für alle } 0 \leq s < t && (5.4)\\
W_{t+\Delta t} - W_t &\textit{ ist normalverteilt mit Erwartungswert } 0,\textit{ Varianz } \Delta t\,;\; t \geq 0, \Delta t > 0 && (5.5)
\end{aligned}$$

Diese Regeln führen hin zu einer numerischen Simulation eines Wiener Prozesses. Wir beschreiben dies in einer etwas allgemeineren Situation.

Ist $(W_t)_{t\geq 0}$ ein Standard-Wiener-Prozess, so nennt man den stochastischen Prozess

$$X_t := \mu t + \sigma W_t \tag{5.6}$$

eine **Brownsche Bewegung** mit Drift μ und Volatilität σ. Damit lassen sich auch stochastische Prozesse darstellen, die tendenziell eher fallen ($\mu < 0$) oder tendenziell eher steigen ($\mu > 0$). Dabei gilt

$$X_t - X_s \sim \mathcal{N}\left(\mu(t-s), \sigma^2(t-s)\right)\,,\; 0 \leq s < t\,. \tag{5.7}$$

[2] Das, was wir heute physikalisch als Brownsche Bewegung bezeichnen, geht auf J. Ingenhousz (1785/Holzstaubteilchen in Alkohol) und R. Brown (1827/Teilchen in Flüssigkeiten) zurück. Albert Einstein (und unabhängig von ihm Marian Smoluchowski) definierte 1905 offenbar ohne Kenntnis von Bacheliers Arbeiten den Wiener-Prozess/die Brownsche Bewegung in seiner heutigen Gestalt. Einsteins Motivation war es, die Bewegung der brownschen Partikel durch die molekulare Struktur des Wassers zu erklären. Einen Beweis für die wahrscheinlichkeitstheoretische Existenz des Prozesses blieb Einstein allerdings schuldig. Dieser gelang erst 1923 Norbert Wiener, der dabei neue Hilfsmittel von Lebesgue und Borel auf dem Gebiet der Maßtheorie ausnutzen konnte.

Bei der Diskretisierung einer solchen Brownschen Bewegung kommt uns folgende Beobachtung zu Pass: Sei Δt eine Schrittweite und seien dazu $t_0 := 0, t_1 := \Delta t, t_2 := 2\Delta t, \ldots, t_j := j\Delta t$ Zeitschritte. Dann haben wir

$$W_{t_j} = \sum_{k=1}^{j} (W_{t_k} - W_{t_{k-1}}) .$$

Da die Inkremente $W_{t_1} - W_{t_0}, \ldots, W_{t_k} - W_{t_{k-1}}$ unabhängig sind (siehe (c) in Definition (5.3)) und da jedes Inkrement normalverteilt ist mit Erwartungswert 0 und Varianz Δt (beachte Regeln (5.3), (5.2)), ist jedes skalierte Inkrement $W_{t_k} - W_{t_{k-1}}, k = 1, \ldots, j$, standard-normalverteilt. Verwendet man nun die obige Diskretisierung des Zeitintervalls $[0,T]$, so kommt man zu einem Berechnungsverfahren für den Prozess aus (5.6), das wir nun so zusammenfassen:

Algorithm 5.1 Simulation eines Wiener Prozesses

EIN Parameter μ, σ, Endzeit T. Anzahl der Zerlegungspunkte N. Mechanismus zur Erzeugung von standard-normal-verteilten Zufallszahlen.

Schritt 0 $\Delta t := T/N$, $t_0 := 0$.

Schritt 1 Für $k = 1, \ldots, N$ erzeuge eine Zufallszahl u und setze $X_k := X_{k-1} + \mu\Delta t + \sigma u\sqrt{\Delta t}$.

AUS Diskrete Approximation eines Pfades des Wiener Prozesses im Zeitraum $[0,T]$.

Das Ergebnis aus obigem Algorithmus (5.1.1) kann durch lineare Interpolation zu einem stetigen Pfad vervollständigt werden.

Jeder Pfad des Wiener Prozesses ist überall nichtdifferenzierbar fast sicher. Dies kann man im wesentlichen aus einem Differenzenquotienten

$$\frac{W_{t+\Delta t} - W_t}{\Delta t}$$

für den Standard-Wiener-Prozess ablesen, denn der Zähler hat Varianz Δt . Es ist, wenn wir schon die Integration eines Wiener-Prozesses im Blick haben, noch schlimmer, denn die Variation der Wiener-Pfade ist (fast sicher für alle ω) unendlich:

$$\lim_n \sum_{k=1}^{2^n} |W_{k2^{-n}}(\omega) - W_{(k-1)2^{-n}}(\omega)| = \infty .$$

5.1.2 Itô-Prozesse

Wie wir schon in Abschnitt (1.2) angedeutet haben, wird die Kursentwicklung eines Basisobjekts mit einer stochastischen Differentialgleichung folgendermaßen modelliert (siehe (1.7)):

$$dS = \mu S\, dt + \sigma S\, dW_t \tag{5.8}$$

Hierbei ist μ ein so genannter **Driftterm**, σ die **Volatilität**; es werden keine Dividenden modelliert, der Zinssatz r kommt später ins Spiel. Die zufällige Kursentwicklung des Basisobjekts wird durch den Wiener Prozess (die Brownsche Bewegung) W „gesteuert“. Wie kann man daraus den stochastischen Prozess $(S_t)_{t\geq 0}$ ableiten, der dieser Gleichung (5.8) genügt? Wie soll man die Differentialgleichung (5.8) verstehen? Dies wollen wir nun vorbereiten.

Aus (5.8) folgt für die Variable $Y := \ln(S/S_0)$ – wir werden später auf diese Transformation zurückkommen – in formaler Weise

$$dY = \mu\, dt + \sigma\, dW_t, \tag{5.9}$$

eine „Differentialgleichung", die ohne den Term mit dem Wiener-Prozess trivial ist, mit dem Wiener-Prozess das Mindeste sein sollte, was eine klare Bedeutung haben sollte. Als Integralgleichung geschrieben haben wir für den stochastischen Prozess $(Y_t)_{t\geq 0}$, der (5.9) lösen soll,

$$Y_t = Y_0 + \mu \int_0^t dt + \sigma \int_0^t d\,W_s = \mu t + \sigma\, W_t\,,\ t \geq 0\,.$$

Definition 5.5 *Sei $(W_t)_{t\geq 0}$ ein Wiener-Prozess und seien $(a_t)_{t\geq 0}, (b_t)_{t\geq 0}$ stochastische Prozesse (mit gewissen Eigenschaften, die wir hier nicht anführen). Ein stochastischer Prozess $(X_t)_{t\geq 0}$ mit $X_0 = x_0$, der der Gleichung*

$$X_t = x_0 + \int_0^t a_s\, ds + \int_0^t b_s\, dW_s\,,\ t \geq 0\,, \tag{5.10}$$

genügt, wird als **Itô-Prozess** *bezeichnet.* □

Dies ist noch eine sehr formale Definition: es bleibt z.B. offen, wie $\int_0^t b_s\, dW_s$ zu verstehen ist und wann das Integral einen Sinn macht.[3] Die Bezeichnung „Drift" und „Diffusionsrate"[4] für $(a_t)_{t\geq 0}$ bzw. $(b_t)_{t\geq 0}$ ist üblich und ist nahezu in Übereinstimmung mit den Redensarten in der Finanzmathematik.
Meist schreibt man Itô-Prozesse in differentieller Form. Für (5.10) bedeutet dies

$$dX = a_t\, dt + b_t\, dW_t \tag{5.11}$$

Wir holen nun weiter aus, um die wesentlichen Aspekte etwas deutlicher zu sehen. Betrachte

$$dX = a(X,t)\, dt + b(X,t)\, dW_t \tag{5.12}$$

Hier modelliert $b(X,t)$ ein multiplikatives „Rauschen" (es sei denn, b ist unabhängig von x, dann haben wir ein additives Rauschen). Um über Existenz- und Eindeutigkeitsansätze reden zu können, schreibt man (5.12), wie schon bei gewöhnlichen Differentialgleichungen, um in eine Integralgleichung:

$$X_t = X_{t_0} + \int_{t_0}^t a(X_s,s)\, ds + \int_{t_0}^t b(X_s,s)\, dW_s\,,\ t \geq t_0\,. \tag{5.13}$$

Hier ist X_{t_0} der Anfangsprozess. In vielen Fällen ist X_{t_0} deterministisch, also $X_{t_0} = x_0 \in \mathbb{R}$, was wir hier auch nun so zunächst annehmen wollen. Das erste Integral ist mehr oder minder klar, denn wenn wir ein Elementarereignis im zugrundeliegenden Wahrscheinlichkeitsraum einsetzen, entsteht ein „Riemanninteral". Das zweite Integral bedarf aber einer Erläuterung, denn Integration bezüglich des stochastischen Prozesses $(W_t)_{t\geq 0}$ ist ja nicht in der klassischen Integrationstheorie verstehbar. Übergehen wir dies zunächst und betrachten den Fall, der für (5.8) ausreicht und etwas einfacher ist, nämlich den Fall

$$a(x,t) = ax, b(x,t) = bx, a, b \text{ Konstanten}\,,\ t_0 = 0\,.$$

[3] Auf Kiyoshi Itô (* 1915) geht die stochastische Integration zurück. Für seine Verdienste erhielt er 2006 den Gauß-Preis; siehe `http://www.mat.univie.ac.at/ schachermayer/pubs/preprnts/prpr0133.pdf`.

[4] Die Bezeichnung kommt aus der Molekularphysik. Dort modelliert man Diffusionen die Ortsveränderung eines Moleküls infolge einer deterministischen Komponente (Drift) und einer erratischen Komponente.

Wir interpretieren dann den Term $b\,dW_s$ als Inhomogenität f in der gewöhnlichen Differentialgleichung

$$y' = a + f(t) \tag{5.14}$$

mit Anfangswert $y(0) = y_0$, wobei y für $\ln(x)$ steht. Die Lösungsformel lautet bekanntermaßen

$$y(t) = y(0) + at + \int_0^t f(s)ds,$$

was nach Ersetzen von f zu

$$y(t) = y(0) + at + \int_0^t b\,d\,W_s = y(0) + at + b\,W_t \tag{5.15}$$

führt, wobei wir $W_0 = 0$ verwendet haben. Nun schreiben wir dies wieder für X_t auf:

$$X_t = x_0 \exp(at + bW_t)\,,\ t \geq 0\,. \tag{5.16}$$

Zuallererst, es ist nicht klar, wie der Term $\exp(bW_t)$ zu verstehen ist. Ferner und noch bedeutender, die Formel in (5.16) ist nicht korrekt, denn die Integration, die wir auf dem Weg zu (5.15) ausgeführt haben, ist nicht korrekt, es fehlt ein Zusatzterm (wenn wir schon eine gültige Integrationstheorie im Auge haben). Hier ist die korrekte Formel, schon als Prozess aufgeschrieben:

$$X_t = x_0 \exp((a - \frac{1}{2}b^2)t + bW_t)\,,\ t \geq 0\,. \tag{5.17}$$

Nun wollen wir skizzieren, was es mit dem obigen „Fehler“ bei der Integration auf sich hat. Das Problem kann nur in der Integration in (5.10) liegen. Wir werden keine Details der Entwicklung eines Integrals präsentieren, dazu verweisen wir etwa auf [9].

Natürlich wird man es bei der Integration wieder mit der Approximationstechnik „Riemannsummen“ versuchen, wie bei Riemann-Stieltjes-Integralen. Es ist aber keine einfache Übertragung möglich, da ein Wiener-Prozess in jedem Zeitintervall eine unbeschränkte Variation besitzt. Um ein Integral

$$\int_0^\tau f_s\,dW_s$$

mit einem stochastischen Prozess $f = (f_t)_{t\geq 0}$ zu definieren, approximiert man f durch einfachere Prozessvariablen:

$$\tilde{f}_s := \sum_{j\in J} f_{t_j^*}\chi_{[t_j,t_{j+1}]}\,,\ \text{ d.h. } \tilde{f}_s(\omega,t) := \sum_{j\in J} f_{t_j^*}(\omega)\chi_{[t_j,t_{j+1}]}(t)\,,$$

wobei t_j^* gewählte Punkte in $[t_j, t_{j+1}]$ sind. Damit betrachtet man die „Riemannsummen“

$$\sum_{j\in J} f_{t_j^*}(W_{j+1} - W_j)\,.$$

Das Integral $\int_0^\tau f_s\,dW_s$ soll nun der Grenzwert dieser „Riemannsummen“ sein, in welchem Sinne soll hier nicht diskutiert werden, jedenfalls kann es so eingerichtet werden, dass die erwarteten Eigenschaften (Linearität, Monotonie, ...) nachgewiesen werden können; der Erwartungswert spielt dabei eine wesentliche Rolle. Wichtig ist aber folgende Anmerkung: es ist nicht gleichgültig – im Unterschied zur Riemann-Integrierbarkeit – wie die Punkte t_j^* in $[t_j, t_{j+1}]$ gewählt werden. Zwei ausgezeichnete Möglichkeiten bieten sich an:

- $\mathbf{t_j^* = t_j}$ Diese Wahl führt zum **Itô–Integral**, mit der Bezeichnung
$$\int_0^\tau f_s \, dW_s \,,$$
die wir schon vorweggenommen haben.

- $\mathbf{t_j^* = \frac{1}{2}(t_{j+1} + t_j)}$ Diese Wahl führt zum **Stratonovich–Integral** mit der Bezeichnung
$$\int_0^\tau f_s \circ dW_s \,.$$

So groß sollte die Überraschung bezüglich der Unterschiede nicht sein, denn nimmt man etwa den Spezialfall $f_t := W_t$, so kann man beim Produkt in den „Riemannsummen“ im Fall der Itô-Wahl mit Unabhängigkeit der Inkremente argumentieren, im anderen Falle nicht. Jeder der beiden Integralbegriffe hat seine Vor- und Nachteile. In der Finanzmathematik hat sich der Itô–Integralbegriff durchgesetzt, vor allem dank der Tatsache, dass sich ein enger Zusammenhang mit der Martingaltheorie herstellen lässt. Wir können darauf nicht eingehen, ein kleines Resultat soll den Zusammenhang herstellen.

Regel 5.6

(a) Der Wiener-Prozess ist ein Martingal.

(b) Der Prozess $(W_t^2 - t)_{t \geq 0}$ ist ein Martingal.

Zu (a): Dies folgt mit der Unabhängigkeit der Inkremente so:
$$\mathbb{E}(W_t|\mathcal{F}_s) = \mathbb{E}(W_t - W_s|\mathcal{F}_s) + W_s = \mathbb{E}(W_t - W_s) + W_s = W_s \,,\; 0 \leq s < t \,.$$
Zu (b): Dies überlassen wir dem Leser.

Die uns vom Riemann-Integral vertrauten Integrationsregeln lassen sich nicht Eins zu Eins übertragen. Einen Hinweis soll uns das folgende Beispiel geben.

Beispiel 5.7 *Sei $f : [0,\infty) \longrightarrow \mathbb{R}$ differenzierbar mit $f(0) = 0$. Dann gilt auf Grund von Ergebnissen aus der Riemann-Integration*
$$\int_0^t f(s) df(s) = \int_0^t f(s) f'(s) ds = \frac{1}{2} f(t)^2 \,.$$
Dies vergleichen wir nun mit Integration „entlang“ des Wiener-Prozesses.

Sei $(W_t)_{t\geq 0}$ ein Wiener-Prozess. Differenzierbarkeit von $t \longmapsto W_t$ unterstellend würde
$$\int_0^t W_s \, dW_s = \int_0^{W_t} w \, dw = \frac{1}{2} W_t^2$$
erwarten lassen. Dies kann nicht zutreffen, denn mit einer Zerlegung des Intervalls $[0,t]$ durch die Zerlegungspunkte $0 = t_0 < t_1 < \cdots < t_n = t$ erhalten wir
$$\begin{aligned}
\frac{1}{2}\sum_{k=0}^{n-1}(W_{t_{k+1}} - W_{t_k})^2 &= \frac{1}{2}\sum_{k=0}^{n-1} W_{t_{k+1}}^2 - \sum_{k=0}^{n-1} W_{t_k} W_{t_{k+1}} + \frac{1}{2}\sum_{k=0}^{n-1} W_{t_k}^2 \\
&= \frac{1}{2}\sum_{k=0}^{n-1} W_{t_{k+1}}^2 - \sum_{k=0}^{n-1} W_{t_k} W_{t_{k+1}} + \sum_{k=0}^{n-1} W_{t_k}^2 - \frac{1}{2}\sum_{k=0}^{n-1} W_{t_k}^2 \\
&= \frac{1}{2} W_t^2 - \sum_{k=0}^{n-1} W_{t_k}(W_{t_{k+1}} - W_{t_k})
\end{aligned}$$

also
$$\sum_{k=0}^{n-1} W_{t_k}(W_{t_{k+1}} - W_{t_k}) \; = \; \frac{1}{2}W_t^2 - \frac{1}{2}\sum_{k=0}^{n-1}(W_{t_{k+1}} - W_{t_k})^2 \, .$$
Wegen
$$\mathbb{E}((W_{t_{k+1}} - W_{t_k})^2) \; = \; t_{k+1} - t_k$$
kann sich im Grenzwert $n \to \infty$ nur
$$\int_0^t W_s \, dW_s \; = \; \frac{1}{2}W_t^2 - \frac{1}{2}t$$
ergeben. □

Wir halten fest, dass nun für geeignete stochastische Prozesse $(a_t)_{t\geq 0}$, $(b_t)_{t\geq 0}$ die Integrale
$$\int_0^t a_s \, ds \, , \int_0^t b_s \, dW_s$$
existieren. Wesentlich dabei sind die Adaptiertheit und die Existenz der Erwartungswerte der Prozesse.

5.1.3 Das Itô-Lemma

Wir wollen nun das Itô-Lemma präsentieren. Dieses Lemma sagt etwas aus über die Transformation von Itô-Prozessen.

Als Vorbereitung auf das Itô-Lemma bringen wir einige Aussagen für einen Wiener-Prozess in der Schreibweise von infinitesimalen Größen.

Regel 5.8 *Sei $(W_t)_{t\geq 0}$ ein Wiener Prozess. Dann gilt für alle $t \geq 0$:*
$$\begin{aligned} \mathbb{E}(dW_t) &= 0 & (5.18)\\ \mathbb{E}(dW_t \, dt) &= \mathbb{E}(dW_t) \, dt \; = \; 0 & (5.19)\\ \mathbb{E}((dW_t)^2) &= dt & (5.20) \end{aligned}$$

Die Regel (5.18) ist die Umschreibung der Tatsache, dass Inkremente des Wiener-Prozesses Erwartungswert Null besitzen. Die Regel (5.18) folgt aus der Tatsache, dass der Erwartungswert eines Produkts einer Zufallsgröße und eines Skalars die Konstante mal den Erwartungswert der Zufallsgröße ist. Die Regel (5.20) ist eine Umformulierung der Aussage über die Varianz der Inkremente des Wiener-Prozesses.

Regel 5.9 *Sei $(W_t)_{t\geq 0}$ ein Wiener Prozess. Dann gilt für alle $t \geq 0$:*
$$\begin{aligned} \mathbb{V}((dW_t)^2) &= \mathbb{E}((dW_t)^4) - (\mathbb{E}((dW_t)^2))^2 \; = \; 3(dt)^2 - (dt)^2 \; = \; 0 & (5.21)\\ \mathbb{E}((dW_t \, dt)^2) &= \mathbb{E}((dW_t)^2) \, (dt)^2 \; = \; 0 & (5.22)\\ \mathbb{V}(dW_t \, dt) &= \mathbb{E}((dW_t \, dt)^2) - (\mathbb{E}(dW_t \, dt))^2 \; = \; 0 & (5.23) \end{aligned}$$

(5.21) folgt aus der Tatsache, dass das 4.-te Moment einer Standard-Normalverteilung den Wert 3 hat und der Interpretation des Quadrats des Infinitesimals dt als Null. (5.22) folgt in ähnlicher Weise, (5.23) folgt aus (5.19) und (5.22).

Regel 5.10 *Sei $(W_t)_{t\geq 0}$ ein Wiener Prozess. Wir benutzen für alle $t \geq 0$:*

$$(dW_t)^2 = dt \tag{5.24}$$

$$dW_t\, dt = 0 \tag{5.25}$$

$$(dt)^2 = 0 \tag{5.26}$$

Wir wissen aus (5.20), dass dt der Erwartungswert von $(dW_t)^2$ ist, und da wegen (5.21) die Varianz von $(dW_t)^2$ Null ist, können wir $(dW_t)^2$ als Konstante betrachten, was (5.24) bedeutet. (5.25) liegt eine ähnliche Interpretation zugrunde, (5.26) haben wir schon „akzeptiert".

Satz 5.11 (Itô-Lemma) *Sei $(X_t)_{t\geq 0}$ ein Itô-Prozess, der (5.10) genügt. Ist dann $h : \mathbb{R}_+ \times \mathbb{R} \longrightarrow \mathbb{R}$ eine im ersten Argument einmal und im zweiten Argument zweimal stetig differenzierbare Funktion, so ist auch $Y_t := h(t, X_t)$ ein Itô-Prozess; er folgt*

$$dY_t = \frac{\partial h}{\partial t}(t, X_t) + \frac{\partial h}{\partial x}(t, X_t) dX_t + \frac{1}{2}\frac{\partial^2 h}{\partial x^2}(t, X_t)\,(dX_t)^2 \tag{5.27}$$

wobei beim Einsetzen von dX_t aus (5.10) in (5.27) die Regeln (5.24),(5.25),(5.26) zu benutzen sind.

Beweis:
Eine Taylorentwicklung für h ergibt in Kurzschreibweise:

$$dh(t,x) = \frac{\partial h}{\partial t} dt + \frac{\partial h}{\partial x} dx + \frac{1}{2}\left(\frac{\partial^2 h}{\partial t^2}(dt)^2 + \frac{\partial^2 h}{\partial x \partial t} dx dt + \frac{\partial^2 h}{\partial x^2}(dx)^2\right).$$

Hier kann $(dt)^2$ und $dx\, dt$ Null gesetzt werden und wir erhalten

$$dh(t,x) = \frac{\partial h}{\partial t} dt + \frac{\partial h}{\partial x} dx + \frac{1}{2}\frac{\partial^2 h}{\partial x^2}(dx)^2\,.$$

■

Beispiel 5.12 *Wir wollen das Integral*

$$I := \int_0^t W_s\, dW_s,$$

was ja eine Zufallsgröße ist, betrachten; siehe Beispiel 5.7. Dazu setzen wir $X_t := W_t, h(t,x) := \frac{1}{2}x^2, Y_t = h(t, W_t)$. Mit dem Itô-Lemma folgt

$$dY_t = W_t\, dW_t + \frac{1}{2}(d\,W_t)^2 = W_t\, d\,W_t + \frac{1}{2}dt$$

und daher

$$d(\frac{1}{2}W_t^2) = W_t\, d\,W_t + \frac{1}{2}dt\,.$$

Also erhalten wir

$$I = \frac{1}{2}W_t^2 - \frac{1}{2}t$$

mit $\mathbb{E}(I) = 0$ (was zu beweisen ist). Berechnen wir die Varianz des Integrals.

$$\begin{aligned}
\mathbb{V}\left(\int_0^t W_s\, dW_s\right) &= \mathbb{V}\left(\frac{1}{2}(W_t)^2 - \frac{1}{2}t\right) = \mathbb{E}\left(\left(\frac{1}{2}(W_t)^2 - \frac{1}{2}t\right)^2\right) \\
&= \frac{1}{4}\mathbb{E}((W_t)^4 - 2t(W_t)^2 + t^2) = \frac{1}{4}(3t^2 - 2t^2 + t^2) = \frac{1}{2}t^2
\end{aligned}$$

□

5.1.4 Geometrische Brownsche Bewegung und Finanzmathematik

Werfen wir einen grundsätzlichen Blick auf das Black-Scholes-Modell in Form einer stochastischen Differentialgleichung, insbesondere auf die Annahmen über den Prozess $(S_t)_{t\geq 0}$, die explizit oder implizit dahinterstehen.

- Es gelten die „allgemeinen Geschäftsbedingungen“ aus 2.1 und der Markt sei arbitragefrei (siehe 2.2).
- Es ist kein Gedächtnis im Modell abgebildet, was damit beschrieben werden kann, dass ein Markov-Prozess modelliert wird. Als Interpretation für das Marktgeschehen bedeutet dies, dass die Vergangenheit vollständig im Kurs des Basisobjekts abgebildet ist.
- Die Renditen
$$R_{t,\Delta t} := \frac{\Delta S_t}{S_t} := \frac{S_{t+\Delta t} - S_t}{S_t}$$
sind die relevanten Größen im Modell.
- Das Basisobjekt hat eine konstante Volatilität σ, d.h. die Varianz der Renditen wächst wie $\sigma^2 O(\Delta t)$:
$$\mathbb{V}\left(\frac{\Delta S_t}{S_t}\right) = \sigma^2 O(\Delta t)\,.$$
- Die Renditen sind normalverteilt, also
$$\frac{\Delta S_t}{S_t} \sim \mathcal{N}(\mu\Delta t, \sigma\sqrt{\Delta})\,.$$
Dies können wir auch so hinschreiben:
$$\ln\left(\frac{S_{t+\Delta t}}{S_t}\right) \sim \mathcal{N}((\mu - \frac{1}{2}\sigma^2)\Delta t, \sigma\sqrt{\Delta})\,.$$
- Die Renditen sind (stochastisch) unabhängig voneinander.
- Die Verteilungen der Renditen hängen nicht vom Startzeitpunkt t ab, sondern nur von der Zeitdifferenz.

Wir werden sehen, dass diese Annahmen für die geometrische Brownsche Bewegung zutreffen.

Definition 5.13

(a) *Ein stochastischer Prozess* $(X_t)_{t\geq 0}$ *ist eine* **arithmetische Brownsche Bewegung**, *wenn er der stochastischen Differentialgleichung*
$$dX = adt + bdW_t \quad (\textit{ mit } X_0 = x_0 > 0)$$
und Konstanten a, b *genügt.*

(b) *Ein stochastischer Prozess* $(X_t)_{t\geq 0}$ *ist eine* **geometrische Brownsche Bewegung**, *wenn er der stochastischen Differentialgleichung*
$$dX = aXdt + bXdW_t \quad (\textit{ mit } X_0 = x_0 > 0)$$
und Konstanten a, b *genügt.*

□

Arithmetische und geometrische Brownsche Bewegungen sind Itô-Prozesse.

Bemerkung 5.14 *Ist $(S_t)_{t\geq 0}$ der lösende Prozess einer arithmetischen Brownschen Bewegung*

$$dS = a\,dt + b\,dW_t\,,\; S_0 = x_0,\;\; (a, b \text{ konstant}),$$

so gilt

$$\mathbb{E}(S_t) = x_0 + at\,,\; \mathbb{V}(S_t) = b^2 t\,,\; S_t \sim \mathcal{N}(x_0 + at, b^2 t)\,.$$

Diesen Prozess hat Bachelier zur Modellierung von Aktienkursen vorgeschlagen. Er ist aber zur Modellierung von Aktienkursen nicht wirklich geeignet, denn S_t kann negative Werte annehmen und der Erwartungswert der Zuwächse ist unabhängig vom aktuellen Wert der Aktie. □

Nehmen wir eine geometrische Brownsche Bewegung $S = (S_t)_{t\geq 0}$ für die Kursentwicklung eines Basisobjekts (etwa einer Aktie) an, so haben wir dann[5]

$$dS = \mu S\,dt + \sigma S\,dW_t\,, \;\text{ mit } S_0 \text{ als deterministischen Anfangswert} \tag{5.28}$$

für eine Driftrate $\mu \in \mathbb{R}$ und eine Volatilität $\sigma > 0$. Wäre $\sigma = 0$, wäre sie einfach zu integrieren. Der Term „$\sigma\,dW_t$“ beschreibt die zufälligen Veränderungen des Kurses. Damit ist nun auch die vorläufige Beschreibung in Abschnitt 1.2 eingeordnet. Beachte, dass der logarithmierte Prozess einer arithmetischen Bewegung folgt:

$$d\ln(S_t) = \mu\,dt + \sigma\,dW_t\,. \tag{5.29}$$

Führt man den Prozess $X_t := \mu t + \sigma W_t$ ein, dann geht die Gleichung (5.28) in die Form

$$dS = SdX_t$$

über. Diese Gleichung kann als stochastisches Analogon zur einfachsten deterministischen Differentialgleichung $dy = ydx$ oder $y' = y$ angesehen werden. Es ist also nicht verwunderlich, dass sich als Lösung von (5.28) ein Exponential ergeben wird.

Zur Integration von (5.28) verwenden wir im Itô-Lemma

$$h(t, x) := \ln(\frac{x}{S_0})\,,\; x \geq 0\,,$$

und sehen, dass für den Prozess $Y_t := h(t, S_t)$ der Gleichung

$$dY_t = (\mu - \frac{1}{2}\sigma^2)dt + \sigma dW_t$$

folgt. Durch Integration erhalten wir

$$Y_t = Y_0 + (\mu - \frac{1}{2}\sigma^2)t + \sigma W_t) \text{ d.h. } S_t = S_0 \exp\left((\mu - \frac{1}{2}\sigma^2)t + \sigma W_t\right)\,, \tag{5.30}$$

was wir in (5.16) schon festgestellt hatten. Allerdings müssen wir feststellen, dass $h(t, x) = \ln(\frac{x}{S_0})$ nicht für $x \leq 0$ und/oder $S_0 =$ definiert ist. Aber wir können den Spieß nun umdrehen und das Itô-Lemma mit

$$f(t, x) := x_0 \exp\left((\mu - \frac{1}{2}\sigma^2)t + \sigma x\right)$$

[5] Dieses Modell für die Bewertung von Optionen geht auf P. Samuelson (1915-2009) zurück; siehe [15]. Er erhielt 1970 den Nobelpreis für Wirtschaftswissenschaften.

anwenden und als Ergebnis erhalten wir, dass $S_t := f(t, W_t)$ der stochastischen Gleichung (5.28) folgt. Also bleibt nur noch die Frage nach der Eindeutigkeit der Lösung von (5.28). Es herrscht Eindeutigkeit, wir verzichten auf den Beweis dieser Tatsache.

Aus der obigen Ableitung lesen wir ab, dass die Inkremente dY_t normalverteilt sind: $dY_t \sim \mathcal{N}((\mu - \frac{1}{2}\sigma^2)dt, \sigma dt)$. Man spricht daher davon, dass S_t **log–normalverteilt** ist.

Folgerung 5.15 *Für die geometrische Brownsche Bewegung $(S_t)_{t\geq 0}$ aus (5.30) gilt:*

(a) $\mathbb{E}(S_t) = \exp(\mu t)S_0\,,\ t \geq 0\,.$

(b) $\mathbb{V}(S_t) = S_0^2 \exp(2\mu t)(\exp(\sigma^2 t) - 1)\,,\ t \geq 0\,.$

(c) $\mathbb{E}(S_t|S_u, 0 \leq u \leq s) = \mathbb{E}(S_t|S_s) = e^{\mu(t-s)}S_s\,,\ s < t\,.$

Beweis:
Wir verwenden die Darstellung in (5.30).
Zu (a) Es ist im Wesentlichen der Erwartungswert des Itô-Prozesses $Y_t := \exp(\sigma W_t)$ zu berechnen. Anwendung der Itô-Formel liefert

$$Y_t = Y_0 + \int_0^t \sigma \exp(\sigma W_s)\, dW_s + \int_0^t \frac{1}{2}\sigma^2 \exp(\sigma W_s)\, ds\,.$$

Mit Eigenschaften des Itô-Integrals, insbesondere wegen

$$\mathbb{E}\left(\int_0^t \exp(\sigma W_s)\, dW_s\right) = 0,$$

folgt

$$\mathbb{E}(Y_t) = \mathbb{E}(Y_0) + \frac{1}{2}\sigma^2 \int_0^t \mathbb{E}(Y_s)\, ds\,.$$

Wegen $Y_0 \equiv 1$ ist die Lösung dieser deterministischen Integralgleichung

$$\mathbb{E}(\exp(\sigma W_t)) = \exp(\frac{1}{2}\sigma^2 t)\,.$$

Zu (b) Folgt aus

$$\mathbb{V}(S_t) = \mathbb{E}((S_t)^2) - \mathbb{E}(S_t)^2 = S_0^2 \exp((2\mu + \sigma^2)t) - S_0^2 \exp(2\mu t)\,, \tag{5.31}$$

wobei die Rechnung zu $\mathbb{E}((S_t)^2)$ analog zu (a) auszuführen ist.
(c) überlassen wir dem Leser. ■

Beachte, dass in Folgerung (5.15) σ stets quadratisch auftritt, das Vorzeichen von σ spielt also keine Rolle. Ferner stellt man fest, dass der Parameter σ den Erwartungswert $\mathbb{E}(S_t)$ nicht beeinflusst.

Bemerkung 5.16 *Mit* (5.31) *erklärt sich die Forderung* (2.11) *zur Berechnung der Zahlen u, d, p zum Binomialmodell in Abschnitt 2.2.3.* □

Beispiel 5.17 *Betrachte die geometrische Brownsche Bewegung $(S_t)_{t\geq 0}$. Dann haben wir, dass $\ln(S_T/S_0)$ normalverteilt mit Erwartungswert $\mu' := (\mu - \frac{1}{2}\sigma^2)T$ und Varianz $\sigma' := \sigma\sqrt{T}$ ist.*

Betrachten wir den Spezialfall

$$T = 3, \mu := 0.1, \sigma := 0.3, S_0 := 60\,.$$

Wir setzen $\mu' := 0.165, \sigma' := 0.520$. *Wir wollen berechnen, welches Kursintervall am Ende von 3 Jahren mit einer Wahrscheinlichkeit* $p := 99.73\%$ *nicht verlassen wird; der Erwartungswert von* $\ln(S_3)$ *ist* $\mu'' := \mu' + \ln(S_0) = 4.314$. *Dazu verwenden wir den für eine Normalverteilung bekannten Zusammenhang*

$$P(\mu'' - 3\sigma' \leq \ln(S_3) \leq \mu'' + 3\sigma') = 0.9973\,.$$

Dies ergibt für das $3\sigma'$*-Intervall des logarithmisierten Kurses das Intervall* $[2.755, 5.873]$ *und für den Kurs selbst das Intervall* $[15.73, 355.78]$. □

In der Folge ist es wichtig, dass der Drift-Parameter in einer „risiko-neutralen Welt“ auf Grund der Annahme der Arbitragefreiheit des Marktes gleich dem Zinssatz r gesetzt werden muss und daher μ, ein Parameter, der ja wieder aus dem Markt geschätzt werden müsste, eliminiert werden kann. Wir gehen auf die genaue Begründung unten ein. Als Resultat dieser Vorgehensweise, die man unter dem Stichwort **Martingal-Ansatz** zusammenfasst, ist, dass Optionspreise als diskontierte Erwartungswerte beschrieben werden können, wobei das Maß unter dem der Erwartungswert zu bilden ist, das so genannte **äquivalente Martingalmaß** ist.

Betrachte den Preis $C = C(S,t)$ einer europäischen Call-Option in Abhängigkeit von $S \in [0,\infty)$ und $t \in [0,\infty)$. Der Kurs des Basisobjekts S werden beschrieben durch eine allgemeine stochastische Differentialgleichung:

$$dS = f(S,t)\,dt + g(S,t)\,dW_t\,, \quad \text{mit } S_0 \text{ als Anfangswert.} \tag{5.32}$$

Wir nehmen an, dass C stetig differenzierbar ist und zweimal stetig differenzierbar bezüglich der ersten Komponente ist. Damit können wir das Itô-Lemma anwenden auf die Abbildung C. Es gilt in Kurzschreibweise:

$$dC = \left(\frac{\partial C}{\partial t} + \frac{\partial C}{\partial S} f + \frac{1}{2}\frac{\partial^2 C}{\partial S^2} g^2\right) dt + \frac{\partial C}{\partial S} g dW_t$$

Betrachtet man den Fall der geometrischen Brownschen Bewegung mit

$$f(S,t) := \mu S\,,\; g(S,t) := \sigma S\,,$$

dann erhält man

$$dC = \left(\mu S\frac{\partial C}{\partial S} + \frac{\partial C}{\partial t} + \frac{1}{2}\frac{\partial^2 C}{\partial S^2}\sigma^2 S^2\right) dt + \sigma S\frac{\partial C}{\partial S} dW_t \tag{5.33}$$

Nun kommen wir zur Frage, weshalb in der Black-Scholes statt μ der Zinssatz r auftaucht. Dazu betrachten wir ein Portfolio Π mit dem Wert $\Pi = C - \alpha S$. dann gilt $d\Pi = dC - \alpha dS$. Setzt man nun dC und dS in diese Gleichung ein, so erhält man

$$d\Pi = \sigma S\left(\frac{\partial C}{\partial S} - \alpha\right) dW_t + \left(\mu S\frac{\partial C}{\partial S} + \frac{1}{2}\sigma^2 S^2\frac{\partial^2 C}{\partial S^2} + \frac{\partial C}{\partial t} - \alpha\mu S\right) dt$$

Damit Π risikolos ist, also nicht mehr von den Schwankungen des Basisobjekts abhängt, wählt man $\alpha = \frac{\partial C}{\partial S}$. Dies liefert:

$$d\Pi = \left(\frac{\partial C}{\partial t} + \frac{1}{2}\sigma^2 S^2\frac{\partial^2 C}{\partial S^2}\right) dt \tag{5.34}$$

Das Portfolio wächst wie eine Anleihe. Auf Grund der Arbitragefreiheit muss im Zeitraum $[t, t+dt]$ gelten:

$$d\Pi = r\Pi dt\,,\; r\Pi dt = \left(\frac{\partial C}{\partial t} + \frac{1}{2}\sigma^2 S^2\frac{\partial^2 C}{\partial S^2}\right) dt\,,$$

wobei nun der Zinssatz r ins Spiel gekommen ist. Mit $\Pi = C - \alpha S = C - \frac{\partial C}{\partial S} S$ folgt

$$\frac{\partial C}{\partial t} + rS\frac{\partial C}{\partial S} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 C}{\partial S^2} = rC \tag{5.35}$$

Dies ist die **Black-Scholes-Gleichung**!

Es ist an der Zeit, etwas über **Dividendenzahlungen** zu berichten. Die Dividende ist der Teil des Gewinns, den eine Aktiengesellschaft an ihre Aktionäre (oder eine Genossenschaft an ihre Mitglieder) ausschüttet. Die Höhe der Dividende orientiert sich am Gewinn des Unternehmens, wobei in der Regel ein Teil des Gewinns einbehalten und den Rücklagen zugeführt wird. Die Höhe der Dividende wird vom Vorstand vorgeschlagen und von der Hauptversammlung mit einfacher Mehrheit beschlossen. Die Dividendenzahlung erfolgt meist am Tag nach der Hauptversammlung.

In der Finanzmathematik unterscheidet man zwischen diskreten und kontinuierlichen Dividendenzahlungen. Nehmen wir zunächst einen Zeitpunkt $t_D \in (0, T)$ für eine Dividendenzahlung an. Ausgezahlt wird qS_{t_D-}, wobei $0 < q < 1$ und S_{t_D-} der Kurs des Basisobjekts unmittelbar vor dem Zeitpunkt der Dividendenzahlung ist. Aus Arbitragegründen muss der Kurs des Basisobjekts um die Dividendenzahlung gemindert werden, da sonst ein sofortiger risikoloser Gewinn durch den Kauf des Basisobjekts zur Zeit t_D- und sofortigem Verkauf nach der Dividendenauszahlung erzielt werden könnte. Also haben wir

$$S_{t_D+} = S_{t_D-} - qS_{t_D-}$$

und der Kurs des Basisobjekts ist über t_D hinweg nicht stetig.

Der Inhaber einer Option auf das in Frage stehende Basisobjekt erhält keine Dividendenzahlung, da er das Basisobjekt ja nicht besitzt. Dies hat zur Folge, dass sich der Optionspreis zu den Zeitpunkten t_D- und t_D+ nicht unterscheiden darf. Also muss für den Preis der Option gelten:

$$C(S, t_D-) = C(S, t_D+)$$

Das eben Gesagte kann nun auf endlich viele Zeitpunkte für Dividendenzahlungen angewendet werden. Die durch die Unstetigkeiten sehr viel schwieriger gewordene Bewertung der Option kann man dann so vornehmen, dass man die unstetige Entwicklung von S_t in ihre stetige Anteile zerlegt und über jedes entstehende Teilintervall die Bewertung vornimmt.

Geht man zum „Grenzwert kontinuierlicher Dividendenausschüttung“ über, wird in einem „Zeitintervall dt die Dividende $q\,S\,dt$ ausgezahlt. Als Konsequenz haben wir – wie im diskreten Fall – den stochastischen Prozess anzupassen:

$$dS = (\mu - q)S\,dt + \sigma\,S\,dW_t$$

Auf demselben Weg wie im Fall ohne Dividendenzahlungen kommt man dann zur modifizierten Black-Scholes-Gleichung:

$$\frac{\partial C}{\partial t} + (r-q)S\frac{\partial C}{\partial S} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 C}{\partial S^2} = rC \tag{5.36}$$

Hierbei ist stets $0 \leq q \leq r$.

Die Auszahlungsfunktion ist nicht abzuändern, die Randbedingung bei $S = 0$ bleibt erhalten, eine Veränderung ist aber in der asymptotischen Randbedingung $S \to \infty$ vorzunehmen. Sie ist zu ersetzen durch

$$\lim_{S\to\infty} \left(C(S,t) - Se^{-q(T-t)}\right) = 0 \tag{5.37}$$

Die Reduktion dieser Aufgabe auf ein klassisches Anfangs- Randwertproblem für die Wärmeleitungsgleichung kann wieder durch eine angepasste Transformation erfolgen.

Die Put-Call-Parität ist bei kontinuierlicher Dividendenzahlung neu zu fassen. Sie lautet

$$Se^{-q(T-t)} + P(S,t) - C(S,t) = Ke^{-r(T-t)}\,,\; S > 0, t \in [0,T]\,, \tag{5.38}$$

wobei C und P der Preis des europäischen Calls bzw. Puts ist; siehe Satz 2.6.

Bemerkung 5.18 *Der Preis einer europäischen Call-Option (mit konstanter Volatilität) kann wiederum formelmäßig angegeben werden (siehe Abschnitt 1.2):*

$$C(S,t) = e^{-q(t-T)}S\mathcal{N}(d_+(\sigma)) - Ke^{-r(T-t)}\mathcal{N}(d_-(\sigma))\,, \tag{5.39}$$

Hierbei ist

$$d_\pm(\sigma) = \frac{\ln(\frac{S}{K}) + ((r-q) \pm \frac{\sigma^2}{2})(T-t)}{\sigma\sqrt{T-t}}\,,\; \sigma \geq 0\,,$$

und $\mathcal{N}$ die Verteilungsfunktion der Standardnormalverteilung, also

$$\mathcal{N}(a) := \int_{-\infty}^{a} \frac{1}{\sqrt{2\pi}} \exp(-\frac{s^2}{2})ds\,,\; a \in \mathbb{R}\,.$$

*Bewiesen wurde diese Formel erstmals in [**?**].* □

Das Black-Scholes-Modell, das wir in der stochastischen Fassung als geometrische Brownsche Bewegung und in der deterministischen Form als parabolische Anfangs- Randwertaufgabe kennengelernt haben, kann heutezutage als **Standard-Referenz-Modell** angesehen werden, und dies, obwohl viele der zur Ableitung herangezogenen Annahmen in der Realität verletzt sind. Hierzu einige Anmerkungen, ohne den Anspruch auf Vollständigkeit erheben zu wollen.

- Die Modellannahmen lassen keine **Sprünge in der Kursentwicklung** des Basisobjekts zu, geschweige denn eine vorzeitige Terminierung von Laufzeiten auf Grund von Übernahmen. Eine Entwicklung weg von dieser Annahme ist die Betrachtung von **Jump-Modellen**.

- Die Identität von Soll- und Habenzinsen ist unrealistisch.

- Die Annahme der Konstanz der Volatilität ist vor allem deswegen problematisch, weil damit zwar eine quantitative Ungewissheit vorhanden ist – die Kalibrierung des Modells kann hier Abhilfe schaffen – wesentlich schwerwiegender ist hingegen die qualitative Ungewissheit bezüglich der Frage, ob stochastisch oder deterministisch modelliert werden soll. Auf die stochastische Modellierung der Volatilität kommen wir noch zurück.

- Im Zusammenhang mit der Beobachtung des „Smile“ haben wir schon das widersprüchliche Verhalten der Marktpreise hinsichtlich Restlaufzeiten angesprochen: in der Nähe des Verfallsdatums stimmt das Verhalten der Optionspreise nicht mit dem Modell überein. Eine Beobachtung hierzu ist: Call-Optionen scheinen fast gleichmäßig über die Zeit an Wert zu verlieren, leicht zunehmend jedoch mit kürzer werdender Restlaufzeit. Die zu beobachtende Sensitivität des Optionspreises in der Nähe des Verfallsdatums ist allerdings erklärbar: eine kleine Veränderung des Kurses des Basisobjekts entscheidet bei Optionen *at the money* über Ausübung und Verfall.

- Die Normalverteilung der logarithmierten Renditen ist zentral in der Modellierung. Dass diese Annahme sehr im Gegensatz zu den Realitäten ist, kann man deutlich sehen, wenn man etwa ein Black-Scholes-Modell für DAX-Optionen analysiert: die gewichtete Summe von Aktienkursen, die zwar einzeln log-normal-verteilt sein mögen, ist im Allgemeinen nicht log-normalverteilt.

Es gibt Überlegungen zu den obigen Problemen, die im Einzelfall nahelegen, dass kein wesentliches Fehlerpotential begründet wird.

5.2 Monte Carlo–Simulation

Wir betrachten eine Option, die sich auf ein Basisobjekt bezieht. Die Einschränkung, dass wir nur Optionen, bezogen auf **ein** Basisobjekt betrachten, kommentieren wir später. Der Zufall in der Kursentwicklung des Basisobjekts soll mit Hilfe einer stochastischen Differentialgleichung modelliert werden.

5.2.1 Problemstellung

Basisobjekt

Wir nehmen an, dass ein Modell für die Dynamik des Basisobjekts in Form eines stochastischen Prozesses $S := (S_t)_{t\in[0,\infty)}$, der der geometrischen Brownschen Bewegung mit Drift μ und Volatilität σ nachgebildet ist, genügt:

$$dS = \mu(S,t)\,dt + \sigma(S,t)\,dW_t\,,\ t \in [0,T]\,,\ x = S_0\,. \tag{5.40}$$

x steht für den Anfangswert der Dynamik und kann als **aktueller Kurs des Basisobjekts** angesehen werden. $W = (W_t)_{t\in[0,T]}$ ist ein Wiener-Prozess auf $[0,T]$, wobei T für das Verfallsdatum einer auf S bezogenen Option steht.
Beachte, dass der Prozess von drei „Parametern“ beeinflusst wird: dem Anfangswert x, der Drift μ und der Volatilität σ.

Auszahlungsfunktion

Betrachte eine Auszahlungsfunktion Φ, die von den Kursen des Basisobjekts zu den Zeitpunkten $0 = t_0 < t_1 < \cdots < t_m = T$ abhängt. Der Erwartungswert dieser Auszahlungsfunktion

$$v := \mathbb{E}\left(\Phi(S_{t_1},\dots,S_{t_m})\right)\,, \tag{5.41}$$

soll bestimmt werden. Der Erwartungswert ist zu bilden bezüglich des Wahrscheinlichkeitsmaßes, das der Brownschen Bewegung zu Grunde liegt.

Aufgabe

Die Aufgabe besteht darin, den Wert v bei gegebenen Parametern x, μ, σ, β zu berechnen. Die Abhängigkeit von v bezüglich x, μ, σ ist implizit gegeben. Ein weiterer Parameter, der implizit noch in v versteckt ist, ist der Ausübungspreis.

5.2.2 Simulation einer Auszahlungsfunktion

Die Monte Carlo-Simulation der Berechnung entlang der obigen Bausteine sieht nun so aus:

Algorithm 5.2 Monte Carlo-Simulation einer Auszahlungsfunktion

EIN Auszahlungsfunktion Φ, Anfangswert $x \in \mathbb{R}$, Drift μ, Volatilität σ .

Schritt 1 Für $k = 1, \ldots, N$ integriere die stochastische Differentialgleichung (5.40) mit den Pfaden $S^k = (S_t^k)_{t \in [0,T]}$.

Schritt 2 Werte die Auszahlungsfunktion aus: Ergebnisse $\Phi(S_{t_1}^k, \ldots, S_{t_m}^k)$, $k = 1, \ldots, N$.

Schritt 3 Berechne den Mittelwert $\hat{E} := \frac{1}{N} \sum_{k=1}^N \Phi(S_{t_1}^k, \ldots, S_{t_m}^k)$

AUS „Schätzwert“ $\hat{E}$ als Approximation für den Auszahlungspreis v .

Der obige Algorithmus ist nun zu ergänzen um die Konkretisierung des Schrittes 1. Dazu ein Algorithmus zur Diskretisierung einer stochastischen Differentialgleichung und ein Algorithmus zur Simulation der Brownswchen Bewegung heranzuziehen. Dabei ist zu beachten, dass nach Möglichkeit die Zeitpunkte $t_1, \ldots, t_m$ Zerlegungspunkte der Diskretisierung der stochastischen Differentialgleichung sind.

5.3 Sensitivitäten

Wir betrachten wieder den Erwartungswert v aus (5.41). Allerdings vereinfachen wir die Auszahlungsfunktion, indem wir $m = 1$ setzen. Der allgemeinere Fall bringt nur technische Schwierigkeiten, aber keine grundsätzlichen.

5.3.1 Parameterabhängiger Erwartungswert

Wir wollen also

$$v := \mathbb{E}\left(\Phi(S_T)\right) \tag{5.42}$$

betrachten. $\Phi(S_T)$ ist also eine Zufallsgröße auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$, die implizit von Parametern abhängt; diese Abhängigkeit wollen wir studieren. Stellvertretend für diese Parameter nehmen wir ein $\alpha \in A \subset \mathbb{P}$ und wollen $\Phi(S_T)$ als $\Phi(S_T) = \Phi(S_T)(\alpha) := Y(\alpha)$ betrachten; hierbei ist $\mathbb{P}$ ein Vektorraum. Damit stellen wir dann die folgende vereinfachte Aufgabe:

> Studiere die Sensitivität von
> $$u(\alpha) := \mathbb{E}(Y(\alpha))$$
> bezüglich α .

Wir unterstellen für die allgemeinen Überlegungen stets eine hinreichende Glattheit der Abbildung u .

Woran sind wir nun interessiert? Es ist die Variation der Größe u für Variationen des Parameters α . Wenn wir an die ins Auge gefassten Konkretisierungen denken, gibt es zwei wesentliche Unterschiede zu bedenken: Wird etwa α als der Anfangswertwert x der stochastischen Differentialgleichung aufgefasst, so ist α eine reelle Zahl; steht α für die Volatilität σ, so ist α ein Objekt in einem Funktionenraum und Variationen sind wesentlich schwieriger zu handbaben.

5.3.2 Differenzen-Sensitivitäten

Hier sei α stets ein Parameter in $\mathbb{R}$, Erweiterungen auf den Fall $\alpha \in \mathbb{R}^d, d > 1$, sind naheliegend. Sensitivität bezüglich α meint, Schätzer für die Ableitung u' anzugeben. Es liegt nahe dies mit Differenzenquotienten und Monte Carlo-Simulation anzugehen.

Algorithm 5.3 Monte Carlo-Simulation zur Schätzung der Sensitivität

---

EIN Familie von Zufallsgrößen $(Y(\alpha))_{\alpha \in A}$, Parameter $\alpha \in A$. Schrittweite $h > 0$. Stichprobengröße n.

Schritt 1 Simuliere unabhängige Realisierungen $Y_1(\alpha), \dots, Y_n(\alpha)$ und unabhängige Realisierungen $Y_1(\alpha+h), \dots, Y_n(\alpha+h)$.

Schritt 2 Berechne die Mittelwerte

$$\hat{Y}(\alpha) := \frac{1}{n}\sum_{i=1}^{n} Y_i(\alpha)\,,\ \hat{Y}(\alpha+h) := \frac{1}{n}\sum_{i=1}^{n} Y_i(\alpha+h)\,.$$

Schritt 3 Berechne den Schätzwert

$$\theta_n(\alpha) := D_n(\alpha) := \frac{\hat{Y}(\alpha+h) - \hat{Y}(\alpha)}{h}\,.$$

AUS „Schätzwert" $\theta_n(\alpha)$ für $u'(\alpha)$.

Die obige Formulierung macht deutlich, dass wir keine Unabhängigkeit bei der Realisierung der Stichproben $Y_1(\alpha), \dots, Y_n(\alpha)$ und $Y_1(\alpha+h), \dots, Y_n(\alpha+h)$ unterstellen.

Wenn u zweimal differenzierbar in α ist, dann erhalten wir mit

$$u(\alpha+h) = u(\alpha) + u'(\alpha)h + \frac{1}{2}u''(\alpha)h^2 + o(h^2)$$

für den Bias des Schätzers:

$$\mathrm{bias}(D_n(\alpha)) := \mathbb{E}(D_n(\alpha) - u'(\alpha)) = \frac{1}{2}u''(\alpha)h + o(h)\,. \tag{5.43}$$

Für die zentralen Differenzen

$$C_n(\alpha) := \frac{1}{2h}(\hat{Y}(\alpha+h) - \hat{Y}(\alpha-h))$$

(der obige Algorithmus 5.3.2 kann dahingehend abgeändert werden) erhält man

$$\mathrm{bias}(C_n(\alpha)) := \mathbb{E}(C_n(\alpha) - u'(\alpha)) = \frac{1}{6}u'''(\alpha)h^2 + o(h^2)\,. \tag{5.44}$$

Die höhere Genauigkeit wird erkauft durch größeren Aufwand.

Wir berechnen die Varianz für die Schätzer $D_n(\alpha)$ und $C_n(\alpha)$:

$$\mathbb{V}(D_n(\alpha)) = h^{-2}\mathbb{V}(\hat{Y}(\alpha+h) - \hat{Y}(\alpha)) \tag{5.45}$$
$$\mathbb{V}(C_n(\alpha)) = (2h)^{-2}\mathbb{V}(\hat{Y}(\alpha+h) - \hat{Y}(\alpha-h))\,. \tag{5.46}$$

Diese Identitäten zeigen uns die Konsequenzen auf, die sich aus dem Wunsch, die Schrittweite h klein zu wählen, um den bias klein zu machen, ergeben: es resultiert möglicherweise eine große Varianz für die Schätzer. Dass dies eine realistische „Bedrohung" ist, erkennt man für den Fall, dass die Paare $Y(\alpha), Y(\alpha+h)$ und $Y_i(\alpha), Y_i(\alpha+h)$, $i = 1, \dots, n$, unabhängig und identisch verteilt sind. Dann erhalten wir

$$\mathbb{V}(\hat{Y}(\alpha+h) - \hat{Y}(\alpha)) = \frac{1}{n}\mathbb{V}(Y(\alpha+h) - Y(\alpha))$$

Also sollte man nun die Fälle $q = 0, q = 1, q = 2$ in

$$\mathbb{V}(Y(\alpha+h) - Y(\alpha)) = O(h^q) \tag{5.47}$$

unterscheiden.
Fall $q = 0$: Diese Situation kann man erhalten, wenn $Y(\alpha), Y(\alpha+h)$ unabhängig sind, denn dann ist

$$\mathbb{V}(Y(\alpha+h) - Y(\alpha)) = \mathbb{V}(Y(\alpha+h)) + \mathbb{V}(Y(\alpha)) \approx 2\mathbb{V}(Y(\alpha)),$$

falls $\alpha \longmapsto \mathbb{V}(Y(\alpha))$ stetig ist.
Fälle $q = 1, 2$: Hier ist eine Abhängigkeit der Zufallsgrößen $Y(\alpha+h), Y(\alpha)$ vonnöten.

Betrachten wir den Fall $q = 0$ etwas genauer, wobei wir auch annehmen, dass $Y(\alpha), Y(\alpha+h)$ unabhängig sind. Zur Analyse der Situation führen wir den mittleren quadratischen Fehler eines Schätzers ein:

Definition 5.19 *Sei X eine reellwertige Zufallsgröße auf einem Wahrscheinlichkeitsraum und sei f eine (messbare) reellwertige Funktion. Wenn $\theta \in \mathbb{R}$ geschätzt werden soll, dann ist die* **mittlere quadratische Abweichung** *eines Schätzers $f(X)$ für θ wie folgt definiert:*

$$\mathit{MSE}(f, \theta) := \mathbb{E}\left(|f(X) - \theta|^2\right) .$$

□

Aus der Definition folgt unmittelbar:

$$\mathrm{MSE}(f, \theta) = \mathrm{Bias}^2(f(X)) + \mathbb{V}(f(X)) , \tag{5.48}$$

wobei die Verzerrung (Bias) als $\mathrm{Bias}(f(X)) := \mathbb{E}(f(X) - \theta)$ definiert ist. Ist also der Schätzer erwartungstreu, dann entspricht die mittlere quadratische Abweichung der Varianz.

Kommen wir zurück zur ursprünglichen Problematik. Wir haben

$$\mathrm{MSE}\left(\frac{\hat{Y}(\alpha+h) - \hat{Y}(\alpha))}{h}, u'(\alpha)\right) = O(\frac{1}{2}h^2 + \frac{1}{nh^2}) \tag{5.49}$$

und folgern daraus, dass die Fehlerschranke auszubalanzieren ist, d.h. dass die Wahl von n, h gemäß

$$\frac{1}{2}h^2 = \frac{1}{nh^2} \tag{5.50}$$

erfolgen sollte. Wir erhalten in diesem Falle

$$\mathrm{MSE}\left(\frac{\hat{Y}(\alpha+h) - \hat{Y}(\alpha))}{h}, u'(\alpha)\right) = h(n)^2 \text{ mit } h(n)^4 = \frac{2}{n} . \tag{5.51}$$

Also

$$\lim_n \mathrm{MSE}\left(\frac{\hat{Y}(\alpha+h(n)) - \hat{Y}(\alpha))}{h(n)}, u'(\alpha)\right) = 0 . \tag{5.52}$$

Hier finden wir die typische Situation eines schlechtgestellten Problems vor, wie wir es im Anhang **??** illustrieren mit dem wohl einfachsten Problem dieser Art, der numerischen Differentiation.

Bemerkung 5.20 *Die obigen Überlegungen lassen sich ganz analog auf die zweite Ableitung von u übertragen. Man braucht lediglich die Taylorentwicklung um einen Term zu erweitern und wieder auszubalanzieren.* □

Die oben besprochenen Differenzenmethoden lassen sich nun zurückübersetzen in die Aufgabe, $\mathbb{E}(\Phi(S_T))$ zu schätzen.

5.3.3 Likelihood-Quotienten-Methode

Wir betrachten wieder die Aufgabe, $\mathbb{E}(\Phi(S_T))$ zu schätzen und nehmen an, dass die Dichte der Verteilung des Endwerts S_T bekannt sei. Also nehmen wir

$$v = \int \Phi(S)p(S)dS \tag{5.53}$$

an. Hierbei ist p die Dichte der Verteilung von S_T. Diese Dichte enthalte die Abhängigkeit von einem Parameter α, und wir haben als Konsequenz $v = u(\alpha)$. Dann ist, ohne die Voraussetzungen zu formulieren,

$$u'(\alpha) = \int \Phi(S)\frac{\partial p}{\partial \alpha}(S)dS = \int \Phi(S)\frac{\partial \ln(p)}{\partial \alpha}(S)p(S)dS = \mathbb{E}\left(\Phi\frac{\partial \ln(p)}{\partial \alpha}\right) \tag{5.54}$$

Damit können wir $u'(\alpha)$ als Erwartungswert mittels Monte Carlo-Simulation berechnen.

In derselben Weise kann man auch höhere Ableitungen von u berechnen:

$$u''(\alpha) = \mathbb{E}\left(\Phi\left(\frac{\partial^2 \ln(p)}{\partial \alpha^2} + \left(\frac{\partial \ln(p)}{\partial \alpha}\right)^2\right)\right) \tag{5.55}$$

Wenn der Endwert S_T als Endwert $\hat{S}_N$ der diskreten Sequenz $\hat{S}_0, \hat{S}_1, \dots, \hat{S}_N = \hat{S}$ zustande kommt, dann können wir die Dichte p mit Hilfe von Übergangswahrscheinlichkeiten zerlegen gemäß

$$p(\hat{S}) = \prod_{k=1}^{N} p(\hat{S}_k|\hat{S}_{k-1})\,. \tag{5.56}$$

Dann ergibt sich

$$u'(\alpha) = \int \Phi(\hat{S}_N)\sum_{k=1}^{N}\frac{\partial \ln(p(\hat{S}_k|\hat{S}_{k-1}))}{\partial \alpha}d\hat{S}_1\cdots d\hat{S}_N \tag{5.57}$$

Der größte Vorteil dieser Vorgehensweise ist, dass keine Voraussetzungen an die Funktion Φ zu stellen sind. Dies ist wichtig in Situationen, in denen Φ eine nichtdifferenzierbare Auszahlungsfunktion darstellt, wie dies etwa bei einer Vanilla-Option der Fall ist. Der Nachteil ist, dass wir die Verteilungsdichte p kennen müssen.

5.3.4 Pfadabhängige Überlegungen

Im Abschnitt 5.3.2 sind wir von einem Differenzenquotienten von u ausgegangen. Die folgenden Überlegungen stützen sich auf die Gültigkeit von

$$\frac{d}{d\alpha}\mathbb{E}\left(\Phi(S_T)\right) = \mathbb{E}\left(\frac{d}{d\alpha}\Phi(S_T)\right) = \mathbb{E}\left(\frac{d\Phi}{dS}\frac{dS_T}{d\alpha}\right)\,. \tag{5.58}$$

Mit der Monte Carlo-Methode kann der Erwartungswert simuliert werden.

Delta Hier ist α als Anfangswert S_0 der Dynamik zu interpretieren.

Vega Hier ist α als Volatilität σ zu interpretieren. Da σ als Parameter in der stochastischen Differentialgleichung vorkommt, ist eine Differentiation des Endwertes S_T heikel. Hierzu sind umfangreiche Vorarbeiten zu erledigen, auf die wir hier nicht eingehen können.

Rho Für die Sensitivität bezüglich der Drift ist α als μ zu interpretieren. Wiederum sind umfangreiche Vorarbeiten zu erledigen, auf die wir hier nicht eingehen können.

5.4 Anhang: Der Zoo der Optionen

Zu den „einfachen“ Optionen gehören neben den europäischen auch die amerikanischen Call- und Put-Optionen; man spricht auch von **naked positions**. Daneben gibt es zusammengesetzte Positionen:

hedge position Neben der Vanilla-Option werden noch Basisobjekte gehalten.

spread Es werden zwei Vanillas eines Types einmal long und einmal short gehalten.

combinations Es werden zwei Vanillas verschiedenen Typs nur long oder nur short gehalten.

Diese und weitere Varianten wollen wir nun skizzieren und damit die Vielfalt der Optionen aufzeigen, die auf den Märkten gehandelt werden; ohne Gewähr für Vollständigkeit!

5.4.1 Optionen mit speziellen Ausübungsrechten

Bei den folgenden Optionen ist die Auszahlung wie bei den einfachen (Vanilla-)Optionen, aber die Ausübungsrechte sind unterschiedlich. Stets: T Laufzeit, K Ausübungpreis.

Bermuda-Option Wir haben diese Option schon kurz angesprochen; siehe Abschnitt 2.4. Ihr Preis liegt, bei gleicher Laufzeit und gleichem Ausübungspreis, zwischen dem der europäischen und amerikanischen Option.

Cap-Option Cap-Optionen werden automatisch ausgeübt, sobald der Wert des Basisobjekts eine vorgegebene Schranke H über (Call-Option) bzw. unterschreitet (Put-Option). Cap-Optionen besitzen daher eine obere Schranke für die Auszahlung.

Shout-Option Shout-Optionen geben dem Halter praktisch zwei Ausübungszeitpunkte. Zu einem beliebigen Zeitpunkt $\bar{t} < T$ kann der Halter den aktuellen Preis $S_{\bar{t}}$ festhalten (shouten) und erhält zum Ausübungszeitpunkt T bei einer Call-Option folgende Auszahlung:

$$\max\{(S_{\bar{t}} - K)^+, (S_T - K)^+\}\,.$$

5.4.2 Kombinierte Handelsstrategien

Unterschiedliche Optionstypen können kombiniert werden, um komplexe Absicherungsziele anzustreben.

Spread (Spreize) Ein Spread kombiniert Calls (oder Puts), die sich auf das dasselbe Basisobjekt beziehen, aber unterschiedliche Verfallsdaten und/oder unterschiedliche Ausübungspreise haben.

Straddle (Grätsche) Ein Straddle kombiniert einen Put und einen Call mit identischen Basispreisen und Laufzeiten. Etwa kaufe einen Put und einen Call, jeweils mit Verfallstag T und Ausübungspreis K. Die Auszahlung zur Zeit T ist

$$V(S_T, T) = (S_T - K)^+ + (K - S_T)^+ .$$

Der Kauf eines Straddles lohnt sich also, wenn der Kurs des Basiswertes sich signifikant von K unterscheidet, etwa wenn der Kurs deutlich steigt oder fällt.

Strangle (Würger?) Ein Strangle besteht aus dem gleichzeitigen Kauf (oder Verkauf) von Calls und Puts gleicher Anzahl mit unterschiedlichen Ausübungspreisen. Etwa kaufe einen Call mit Verfallstag T und Ausübungspreis K_2 und kaufe einen Put mit Verfallstag T und Ausübungspreis K_1. Die Auszahlung zur Zeit T ist

$$V(S_T, T) = (S_T - K_2)^+ + (K_1 - S_T)^+ .$$

Die Käufer eines Strangles erwarten sehr große Kursschwankungen, ausgehend vom Kurs K . Ein Strangle ist preiswerter als ein Straddle, da die Gewinnchancen kleiner sind.

Butterfly spread Kaufe einen Call C_1 mit Ausübungspreis K_1 und einen Put P_1 mit Ausübungspreis K_2 sowie verkaufe einen Call C_2 und einen Put P_2 jeweils mit Ausübungspreis K. Alle vier Optionen haben den Verfallstag T. Die Auszahlung zur Zeit T ist

$$V(S_T, T) = (S_T - K_1)^+ + (K_2 - S_T)^+ - (S_T - K)^+ - (K - S_T)^+ .$$

Der Käufer eines Butterfly spread erwartet stagnierende Kurse um den Wert K bzw. geringe Kursschwankungen.

Ein **Optionsschein (Warrant)** auf Aktien berechtigt zum Bezug von Aktien einer Gesellschaft unter genau festgelegten Konditionen und ist im allgemeinen mit der Emission einer Optionsanleihe verbunden; zur Gestaltung von Preisen siehe [12, 17].

5.4.3 Pfadabhängige Optionen

Pfadabhängige Optionen sind spezielle (exotische) Optionen, bei denen zur Bestimmung der Auszahlung nicht nur der Preis des Basisobjekts zum aktuellen Zeitpunkt, sondern die gesamte Preishistorie zugrunde gelegt wird.

Lookback-Option Bei Lookback-Optionen wird zur Ermittlung der Auszahlung das Maximum oder Minimum der Kurse im Ausübungszeitraum betrachtet. Man unterscheidet Lookback-Optionen mit festem und variablem Ausübungspreis. Die Auszahlung zur Zeit T ist:

$$\begin{cases} (\max_{0\le t\le T} S_t - K)^+ & \text{, falls fester Ausübungspreis} \\ (K - \min_{0\le t\le T} S_t)^+ & \text{, falls variabler Ausübungspreis} \end{cases}$$

Lookback-Optionen sind daher relativ teuer.

So genannte **russische Optionen** sind Lookback-Optionen, die ewig, also mit $T = \infty$ laufen.

Barrier-Optionen Dies ist eine Option, die wertlos oder wertvoll wird, wenn der Kurs des Basiswerts eine vorher festgelegte Schranke vor dem Verfallstag über- bzw. unterschreitet. Ein Beispiel ist der Down-and-out Call; diese Option wird wertlos, wenn der Basiswert innerhalb der Laufzeit der Option eine vorgegebene Schranke B unterschreitet. Die Auszahlungsfunktion lautet:

$$V(S_T,T) = (S_T - K)^+ \chi_{\{\min S_t \geq B\}},$$

wobei $\chi_{\{\min S_T \geq B\}} = 1$, falls $\min_{0\leq t\leq T} S_t \geq B$ und $\chi_{\{\min S_T \geq B\}} = 0$, falls $\min_{0\leq t\leq T} S_t < B$.

Asiatische Option Bei asiatischen Optionen wird nicht der Wert des Basisobjekts am Ende des Ausübungszeitraums, sondern über sein zeitliches Mittel über einen gegebenen Zeitraum mit dem Ausübungspreis K verglichen. Man unterscheidet zwischen arithmetischen, geometrischen, diskreten und kontinuierlichen Mitteln:

$$\begin{cases} \left(\frac{1}{m}\sum_{i=1}^m S_{t_i} - K\right)^+ & \text{(diskret arithmetisch)} \\ \left((\prod_{i=1}^m S_{t_i})^{\frac{1}{m}} - K\right)^+ & \text{(diskret geometrisch)} \\ \left(\frac{1}{T}\int_0^T S_t dt - K\right)^+ & \text{(kontinuierlich arithmetisch)} \\ \left(\exp(\frac{1}{T}\int_0^T S_t\, dt - K\right)^+ & \text{(kontinuierlich geometrisch)} \end{cases}$$

Asiatische Optionen schützen den Verkäufer der Option vor Manipulationen des Kurses des Basisobjekts kurz vor dem Verfallstag, da diese durch die Durchschnittsbildung herausgemittelt werden.

5.4.4 Noch exotischere Optionen

Binäre Optionen Diese Option wird wertlos, wenn der Kurs des Basiswerts zum Verfallstag eine festgelegte Schranke K über- oder unterschreitet. Im Gegensatz zu europäischen Optionen ist die Höhe des Auszahlungsbetrags B jedoch unabhängig vom Kurs des Basiswerts. Die Auszahlungsfunktion eines Calls lautet:

$$V(S_T,T) = \begin{cases} B & \text{, falls } S_T > K \\ 0 & \text{, falls } S_T \leq K \end{cases}.$$

Compound-Optionen Mit dem Kauf einer Compound-Option erwirbt man das Recht, zum Verfallstag eine andere Option mit Verfallstag $T' > T$ zum Ausübungspreis K zu kaufen bzw. zu verkaufen. Ein Beispiel ist eine „Put auf Call"-Option mit der Auszahlungsfunktion

$$V(S_T,T) = (K - C_T)^+$$

Chooser-Optionen Bei diesen Optionen kann man zum Verfallstag wählen, ob man einen europäischen Call C oder einen europäischen Put P mit Verfallstag $T' > T$ erhalten möchte. Diese Optionen haben die Auszahlungsfunktion

$$V(S_T,T) = \max\{C_T, P_T\}.$$

5.5 Bibliographische Anmerkungen

Eine gute Quelle für stochastische Differentialgleichungen ist [9, 13, 14, 16], die Brownsche Bewegung wird ausführlich behandelt in [8].

Die Monte Carlo-Methode für die Bewertung von Optionen wird beschrieben u.a. in [10, 6, 7, 11], insbesondere zu Greeks konsultiere man [1, 2]. Eine etwas neuere Entwicklung stellen die Mehrgitter-Monte Carlo-Methoden dar; siehe dazu etwa [3, 4, 5].

5.6 Übungen

1.) Betrachte die Anfangswertaufgabe

$$y' = \xi(\omega) y \, , \; y(0,\omega) = 1 \, , \; t \in [0,T] \, .$$

Dabei ist $\xi : \Omega \longrightarrow \mathbb{R}$ eine Zufallsgröße.

(a) Leite einen Lösungskandidaten her.

(b) Berechne $\mathbb{E}(y(T))$, wenn ξ auf $(-1,1)$ gleichverteilt ist.

(c) Berechne $\mathbb{E}(y(T))$, wenn ξ standard-normalverteilt ist.

2.) Zeige für den Wiener-Prozess $(W_t)_{t\geq 0}$:

$$\mathbb{E}(W_t^{2k}) = \frac{(2k!)}{2^k \, k!} t^k \, , \; k \in \mathbb{N}_0 \, .$$

3.) Sei $f : \mathbb{R} \longrightarrow \mathbb{R}$ stetig. Zeige

$$\mathbb{E}(f \circ W_t) = \frac{1}{\sqrt{2\pi t}} \int_{\mathbb{R}} f(x) e^{-\frac{x^2}{2t}} dx \, ,$$

falls das Integral auf der rechten Seite existiert.

4.) Zeige, dass der Prozess $(W_{t+h} - W_t)_{h\geq 0}$ der Inkremente eines Wiener-Prozesses für alle $t \geq 0$ identisch verteilt ist.

5.) Zeige für den Wiener-Prozess $(W_t)_{t\geq 0}$ und $t_0 \geq 0$, dass $(\tilde{W}_t := W_t - W_{t_0})_{t\geq 0}$ ein Wiener-Prozess ist.

6.) Zeige für den Wiener-Prozess $(W_t)_{t\geq 0}$ und $c > 0$, dass $(\tilde{W}_t := \frac{1}{c} W_{c^2 t})_{t\geq 0}$ ein Wiener-Prozess ist.

7.) Berechne das Integral

$$I := \int_0^t s \, d W_s \, .$$

Hinweis: Wir erwarten $I = tW_t$.

8.) Berechne das Doppelintegral

$$\int_{t_0}^{t} \int_{t_0}^{s} dW_r \, dW_s$$

9.) Löse die stochastische Differentialgleichung

$$dX_t = -\alpha X_t dt + \sigma dW_t \, .$$

10.) Was besagt das Itô-Lemma für $Y_t := W_t / \exp(W_t)$?

11.) Löse die stochastische Differentialgleichung

$$dX_t = dt + 2\sqrt{X_t} dW_t \, .$$

12.) Betrachte einen Wiener-Prozess $(W_t)_{t\geq 0}$. Zeige: Der Zufallsvektor $(W_u, W_s, W_t) \in \mathbb{R}^3$ mit $u < s < t$ ist normal-verteilt mit Erwartungswert $(0,0,0)$ und Kovarianzmatrix

$$\Sigma = \begin{pmatrix} u & u & u \\ u & s & s \\ u & s & t \end{pmatrix}$$

13.) Welcher stochastischen Differentialgleichung genügt das Exponential $(\exp(W_t - t/2))_{t\geq 0}$, wobei $(W_t)_{t\geq 0}$ ein Wiener-Prozess ist?

14.) S_1 und S_2 seien die Preise zweier Aktien zum Zeitpunkt $T = 1$:

$$S_i = \exp(\sigma Z_i - \sigma^2/2), i = 1, 2\,.$$

Dabei sind Z_1, Z_2 zwei unabhängige, standard-normalverteilte Zufallsgrößen und $\sigma = 0.5$. Man schätze den Erwartungswert von $M := \max\{S_1, S_2\}$ mittels Monte Carlo-Simulation.

15.) Betrachte den Prozess

$$X_t = x_0 e^{-\lambda t} + \int \sigma e^{-\lambda(t-s)} dW_s\,,\ t \geq 0\,.\ \ (\lambda, \sigma > 0)\,.$$

Welchen Itô-Prozess erhält man, wenn man das Itô-Lemma mit $h(t, x) := e^{-\lambda t}(x_0 + x)$ benutzt.

16.) Betrachte das Bachelier-Modell eines Finanzmarktes, das heißt, der Aktienkurs S_t ist durch

$$S_t := S_0 + S_0 \sigma W_t$$

gegeben. Berechne den Preis einer europäischen Call-Option mit Verfallsdatum T und Ausübungspreis K und vergleiche den Wert mit dem entsprechenden Preis im Black-Scholes-Modell.

17.) Eine „standardisierte Black-Scholes-Gleichung“ hat die Form

$$\frac{\partial v}{\partial t} = \frac{\partial^2 v}{\partial x^2} + (\gamma - 1)\frac{\partial v}{\partial x} - \gamma v\,,\ (x, t) \in \mathbb{R}^2\,.$$

Zeige: Falls $u : \mathbb{R}^2 \longrightarrow \mathbb{R}$ eine Lösung der klassischen Wärmeleitungsgleichung

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2}$$

ist, dann existiert eine Funktion $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$, so dass

$$v : \mathbb{R}^2 \ni (x, t) \longmapsto v(x, t) := e^{f(x,t)} u(x, t) \in \mathbb{R}$$

die Black-Scholes-Gleichung löst.
Hinweis: Versuche einen linearen Ansatz für f.

18.) Berechne die Griechen $\Delta, \mathcal{V}$ und Rho im Black-Scholes-Modell mit konstanter Volatilität und kontinuierlicher Dividendenzahlung.

19.) Sei $(W_t)_{t\geq 0}$ eine Brownsche Bewegung auf $(\Omega, \mathcal{F}, P)$ adaptiert zu $\mathcal{F} = (\mathcal{F}_t)_{t\geq 0}$. Zeige, dass die folgenden Prozesse $(M_t)_{t\geq 0}$ jeweils ein Martingal sind:

(a) $M_t := W_t^2 - t$.

(b) $M_t := W_t^3 - 3tW_t$.

20.) Betrachte das Black-Scholes-Modell mit genau einer Aktie, konstantem Zins und konstanter Volatilität σ. Die Aktie schütte zeitstetig den Anteil $a \in \mathbb{R}$ des aktuellen Kurswertes als Dividende aus. In diesem Fall gilt die stochastische Differentialgleichung

$$dS = S(r - a)dt + \sigma S dW_t$$

für den Kurswert der Aktie. Bestimme den lösenden Prozess $(S_t)_{t\geq 0}$ und den Preis der europäischen Call-Option mit Ausübungspreis K zum Verfallsdatum T.

21.) Betrachte das Vasicek-Zinsstrukturmodell:

$$dr = (\alpha - \beta r)dt + \gamma dW_t \,,\; r_0 = c \in \mathbb{R}\,.$$

Dabei sind α, β, γ positive reelle Zahlen und $W = (W_t)_{t\geq 0}$ ist ein Standard-Wiener-Prozess. Zeige, dass der lösende Prozess $r = (r_t)_{t\geq 0}$ zu jeder Zeit $t \geq 0$ die Verteilung

$$\mathcal{N}\left(\frac{\alpha}{\beta} + \exp(-\beta t)(r_0 - \frac{\alpha}{\beta}), \frac{\gamma^2}{2\beta}(1 - \exp(-2\beta t))\right)$$

besitzt.

22.) Ein Aktienkurs sei gegeben durch

$$S := S_0 \exp(\sigma Z - \sigma^2/2)\,.$$

Dabei sei $S_0 := 100, \sigma^2 = 0.2$ und Z standard-normalverteilt. Man schätze $\mathbb{E}(S)$ mit der Methode der antithetischen Variablen.

Literaturverzeichnis

[1] N. Chen and P. Glasserman. Mallivian greeks without malliavin calculus. *xxx*, 2006.

[2] E. Fournié, J.M. Lasry, J. Lebuchoux, P.-L. Lions and N. Touzi. An applications of Malliavin calculus to Monte Carlo methods in finance. 3:391–412, 1999.

[3] M. Giles. Multi-level Monte Carlo path simulation. *Operations Research*, 56:981–986, 2008.

[4] M. Giles and B.J. Waterhouse. Multi-level quasi-Monte Carlo path simulation. *Radon Series Comp. Appl. Math.*, 8:1–18, 2009.

[5] M. Giles, D.J. Higham and X. Mao. Analyzing multi-level Monte Carlo for options with non-globally Lipschittz payoff. *Finance and Stochastics*, 13:403–413, 2009.

[6] P. Glasserman. *Monte Carlo Methods in Financial Engineering.* Springer, Baltimore, 2003.

[7] P. Glasserman and K.-K. Kim. Moment explosions and stationary distributions in affine diffusion models. *Math. Finance*, 19:189–214, 2010.

[8] I. Karatzas and S. Shreve. *Brownian motion and stochastic calculus.* Springer, New York, 1991.

[9] B. Øksendahl. *Stochastic Differential Equations.* Springer, New York, 1998.

[10] Y. Lai and J. Spanier. Applications of monte carlo/quasi-monte carlo methods in finance: Option pricing. In *Proceedings of a conference held at the Claremont Graduate Univ.* Springer, 1998.

[11] B. Lapeyre and J. Lelong. A framework for adaptive Monte Carlo procedures. *arXiv:1001.3551v1*, pages 1–16, 2010.

[12] B. Lauterbach and P. Schultz. Pricing warrants: an empirical study of the black-scholes-model and its alternatives. *J. of Finance*, 45:1181–1209, 1990.

[13] K.-S. Moon, A. Szepessi, R. Tempone and G. Zouaris. Stochastic differential equations: models and numerics. Technical report, August 2008.

[14] P.E. Protter. *Stochastic integration and differential equations.* Springer, New York, 2004.

[15] P. Samuelson. Rational theory of warrant pricing. *Industrial Management Review*, 6:13–32, 1965.

[16] J.M. Steele. *Stochastic calculus and financial applications.* Springer, New York, 2001.

[17] R.J. Vanderbei and M.C. Pinar. Pricing american warrants by linear programming. *SIAM Review*, 51:767–782, 2009.

Kapitel 6

Diskretisierung bei der Bewertung von Optionen

(Finite-)Differenzen-Verfahren haben den Vorteil, dass sie einen sehr einfachen Zugang zulassen, Finite-Elemente-Verfahren, erfordern etwas mehr Vorbereitung, sie sind aber in den Resultaten besser und insbesondere für allgemeinere Fragestellungen (amerikanische Optionen, mehrdimensionale Probleme,…) geeigneter. Diese werden wir in Kapitel **??** beschreiben.

Abschließend skizzieren wir Diskretisierungen, die ihren Ursprung in der Fouriertransformation haben.

6.1 Differenzenverfahren: Black-Scholes-Gleichung

Wir lösen die Black-Scholes-Gleichung numerisch, indem wir Ableitungen bezüglich eines Gitters mittels finiter Differenzen approximieren.

6.1.1 Explizites Differenzenverfahren/konstante Volatilität

Durch die Approximation mittels finiter Differenzen entsteht ein Problem der numerischen linearen Algebra, allerdings nur dann, wenn wir das nichtbeschränkte Intervall $(0,\infty)$ für die Variable S „künstlich“ zu einem beschränkten Intervall machen. Dazu haben wir $\overline{S}$ zu wählen und dazu die passende Randbedingung. Da wir wissen, dass $\lim_{S\to\infty}(C(S,t)-S)=0$ gilt (siehe (1.4)), können wir die Randbedingung

$$C(\overline{S},t)=\overline{S} \tag{6.1}$$

festlegen. Dabei muss $\overline{S}$ so gewählt werden, das es deutlich größer als der Ausübungspreis K ist. Damit ist nun folgende Aufgabe numerisch zu lösen:

$$\begin{aligned}
\frac{\partial C}{\partial t} &= -\frac{1}{2}\sigma^2(S,t)S^2\frac{\partial^2 C}{\partial S^2} - rS\frac{\partial C}{\partial S} + rC = 0\,,\ S\in(0,\overline{S}), t\in(0,T), && (6.2)\\
C(0,t) &= 0\,,\ C(\overline{S},t) \;=\; \overline{S}\,,\ t\in(0,T), && (6.3)\\
C(S,T) &= (S-K)^+\,,\ S\in(0,\overline{S})\,. && (6.4)
\end{aligned}$$

Dabei ist σ eine Funktion von S und t . Ist σ hinreichend glatt, dann besitzt diese Aufgabe eine eindeutige Lösung. Auf die Regularität kommen wir noch zurück. In diesem Anschnitt betrachten wir den Fall konstanter Volatilität.

Wir schreiben die diskrete Gleichung auf und erklären das zugrundeliegende Vorgehen anschließend:

$$\frac{V_i^{j+1}-V_i^j}{\Delta t}+\frac{1}{2}\sigma^2 S_i^2\frac{V_{i+1}^j-2V_i^j-V_{i-1}^j}{(\Delta S)^2}+rS_i\frac{V_i^j-V_{i-1}^j}{\Delta S}-rV_i^j=0 \tag{6.5}$$

für $i = 1, \ldots, N-1$, $j = M-1, \ldots, 0$. Dazu kommen die Endbedingung

$$V_i^0 = \max(S_i - K)^+ \, , \; i = 1, \ldots, N-1 \tag{6.6}$$

und die Randbedingungen

$$V_0^j = 0 \, , \; j = M, \ldots, 0 \, , \; V_N^j = \overline{S} \, , \; j = M, \ldots, 0 \, . \tag{6.7}$$

V_i^j steht für die Näherung für $C(i\Delta S, j\Delta t)$, $i = 0, \ldots, N, j = 0, \ldots, M$. Daraus ergibt sich, dass wir

$$\Delta x := \frac{\overline{S}}{N} \, , \; \Delta t := \frac{T}{M}$$

gesetzt haben. Da wir die Gleichung von „hinten nach vorne“ rechnen, sind $V_i^0, i = 0, \ldots, N$, die Größen, die berechnet werden sollen.

Die Approximation der Ableitungen durch Differenzenquotienten ist naheliegend, drei Anmerkungen sind wichtig:

- Die Zeitdiskretisierung ist implizit, d.h. auf jeder Zeitschicht $j\Delta t$ ist ein lineares Gleichungssystem zu lösen.
- Die Umsetzung des Terms rS_i kann auch durch $\max(rS_i, 0)$ erfolgen, um sicherzustellen, dass durch die Diskretisierung die Nichtnegativität von S nicht verletzt wird.
- Die Diskretisierung von $\frac{\partial C}{\partial S}$ durch $\frac{V_i^j - V_{i-1}^j}{\Delta S}$ entspricht dem so genannten **upwind-scheme**.

Algorithm 6.1 Explizites Differenzenverfahren zur Black-Scholes-Gleichung

---

EIN Basiskurs S_0, Ausübungszeitpunkt T, Ausübungspreis K, Zinssatz r, Volatilität σ . Maximaler Basispreis $\overline{S}$, Diskretiserungsparameter $M, N \in \mathbb{N}$.

Schritt 0 $\Delta x := \frac{\overline{S}}{N}$, $\Delta t := T/n$.

Schritt 1 Für $j = M-1, \ldots, 0$ löse (6.5),(6.7) unter Verwendung von (6.6) bei $j = M-1$.

AUS Näherungen $V_i^0, i = 0, \ldots, N$, für die Optionspreise $C(i\Delta S, 0)$ zu den Basispreisen $S_i = i\Delta x, i = 0, \ldots, M$.

Hier sind einige Anmerkungen zur Umsetzung.

- Der Aufgabe „Löse (6.5),(6.7)“ kann eine nützliche Form gegeben werden. Dazu formuliert man sie als lineares Gleichungssystem, zu dessen Lösung dann aus der günstigen Struktur (Bandstruktur) des Gleichungssystems Nutzen gezogen werden kann (LU-Zerlegung, Rückwärtssubstitution).
- Die Forderung nach Stabilität der Rechenvorschrift (keine „dramatische“ Anfachung von Daten- und Verfahrensfehlern) bringt keine Schrittweitenbeschränkung ins Spiel, wie dies etwa bei einem expliziten Vorgehen der Fall wäre; siehe unten. Trotz allem eintretende numerische Instabilitäten lassen sich in der Regel aus den Ergebnissen ablesen.
- Die im Risikomanagement wichtigen Sensitivitäten (Greeks) Δ, Γ, Θ lassen sich parallel zur Bestimmung der Preise V_i^j leicht berechnen.

- Durch Interpolation können die Optionspreise V_i^j zu einer Preisfunktion $C : [0,\overline{S}] \longrightarrow \mathbb{R}$ zusammengesetzt werden.

Bemerkung 6.1 *Die Diskretisierung, die dem Algorithmus 6.1.1 zu Grunde liegt, kann auch explizit erfolgen:*

$$\frac{V_i^{j+1}-V_i^j}{\Delta t}+\frac{1}{2}\sigma^2 S_i^2\frac{V_{i+1}^{j+1}-2V_i^{j+1}-V_{i-1}^{j+1}}{(\Delta S)^2}+\max(rS_i,0)\frac{V_i^{j+1}-V_{i-1}^{j+1}}{\Delta S}-rV_i^{j+1}=0 \qquad (6.8)$$

für $i=1,\ldots,N-1$, $j=M-1,\ldots,0$. *Dann wird die Berechnung einfacher, es kommt aber die Schrittweitenbeschränkung* $\Delta t = O((\Delta x)^2)$ *ins Spiel. Ein guter Kompromiss zwischen beiden Vorgehensweisen ist das Crank-Nicolson-Verfahren, das eine bessere Konvergenzordnung zeigt. Dieses formulieren wir im nächsten Abschnitt.* □

6.1.2 Crank-Nicolson–Diskretisierung/Konstante Volatilität

Wir betrachten erneut die Aufgabe (6.2), (6.2), (6.2) mit konstanter Volatilität. Um die Begrifflichkeit etwas von der speziellen Situation der Finanzmathematik zu lösen, machen wir eine Zeitumkehr mittels $s := T-t$ und betrachten logarithmisierte Kurse $x := \ln(S)$. Damit wird aus der Aufgabe (6.2), (6.2), (6.2) folgende Anfangs- Randwertaufgabe für den transformierten Optionspreis $v(x,s) := C(e^x, T-s)$:

$$\begin{aligned}
\frac{\partial c}{\partial s} &= \frac{1}{2}\sigma^2\frac{\partial^2 v}{\partial x^2}+(r-\frac{1}{2}\sigma^2)\frac{\partial v}{\partial x}-rv=0\,,\ x\in(0,\overline{x}:=\ln(\overline{S})), s\in(0,T), &(6.9)\\
v(0,s) &= 0\,,\ v(\overline{x},s)\ =\ \overline{x}\,,\ s\in(0,T), &(6.10)\\
v(x,s) &= (e^x-K)^+\,,\ x\in(0,\overline{x})\,. &(6.11)
\end{aligned}$$

Die Zeitschrittweiten der Differenzenapproximation seien Δs und Δx. V_i^j seien die Approximationen für $v(j\Delta s, i\Delta x)$. Wir definieren formal Operatoren:

$$\begin{aligned}
\delta_i^2 V_i^j &:= V_{i+1}^j-2V_i^j+V_{i-1}^j\\
\delta_{2i} V_i^j &:= V_{i+1}^j-V_{i-1}^j\\
D &:= -\frac{1}{2}\Delta s(\Delta x)^{-2}\sigma^2\delta_\cdot^2-\frac{1}{2}\Delta s(\Delta x)^{-1}(r-\frac{1}{2}\sigma^2)\delta_{2\cdot}-r\Delta s
\end{aligned}$$

Eine Mittelung des expliziten Verfahrens (siehe 6.1) und des impliziten Verfahrens (siehe 6.5) führt zum so genannten **Crank-Nicolson-Verfahren**, das wir mit den oben eingeführten Operatoren so aufschreiben können:

$$(I+\frac{1}{2}D)V_i^{j+1}=(I-\frac{1}{2}D)V_i^j\,,\ i=1,\ldots,N-1\,,\ j=0,\ldots,M-1\,. \qquad (6.12)$$

Die Rand- und Anfangsbedingungen sind wie im impliziten Fall zu wählen:

$$\begin{aligned}
V_0^j &= 0\,,\ V_N^j=\overline{S}\,,\ j=0,\ldots,M\,, &(6.13)\\
V_i^0 &= \max(e^{i\Delta x}-K)^+\,,\ i=1,\ldots,N-1\,. &(6.14)
\end{aligned}$$

Dieses Verfahren können wir nun ohne Mühe in der Form eines Algorithmus wie in 6.1.1 aufschreiben.

Was zeichnet dieses Verfahren aus? Es ist die Stabilität des Verfahrens und die Konvergenzordnung. Die Stabilität sichert die Robustheit gegenüber Datenfehlern, die bewiesene Konvergenzordnung 2 bedeutet, dass die aus der diskreten Approximation durch lineare Interpolation zusammengesetzte Funktion in der L_2-Norm quadratisch konvergiert gegen die Lösung, wenn die Schrittweiten gegen Null konvergieren. Allerdings setzen diese Aussagen die Glattheit der Anfangsdaten voraus. Dies ist aber hier nicht der Fall, denn die vorgegebene Anfangsfunktion in

$$v(x,0) = (e^x - K)^+ \,,\; x \in (0, \overline{x})\,,$$

ist nicht glatt; sie hat in der ersten Ableitung eine Unstetigkeit; aus der Theorie weiß man, dass die Konvergenzordnung vermindert wird. Dies macht sich dann drastisch bemerkbar bei der numerischen Realisierung, wenn man die Griechen aus den Approximationen berechnet, insbesondere dann, wenn man sich die L_∞-Konvergenz anschaut; sie sind ja verknüpft mit Ableitungen der Lösung.

Wie kann diese nichtglatte Anfangsvorgabe nun berücksichtigt werden? Dazu gibt es den Vorschlag, die ersten Zeitschritte implizit mit kleinerer Schrittweite durchzuführen; siehe [31, 9]. Zwei implizite Halb-Schritte sehen so aus:

$$(I + \frac{1}{4}D)V_i^{j+\frac{1}{2}} = V_i^j \,,\; i = 1, \ldots, N-1\,. \tag{6.15}$$

$$(I + \frac{1}{4}D)V_i^{j+1} = V_i^{j+\frac{1}{2}} \,,\; i = 1, \ldots, N-1\,. \tag{6.16}$$

Der Vorschlag, diese impliziten Zeitschritte am Anfang zu machen, hat eine einleuchtende Begründung. Da parabolische partielle Differentialgleichungen klassische Lösungen haben, die die Anfangsdaten (extrem) glätten, kann man von einem impliziten Schritt erwarten, dass er die Nichtdifferenzierbarkeit des Anfangsdatums in der Black-Scholes-Gleichung „beseitigt“.

Bemerkung 6.2 *Die Beschreibung des Crank-Nicolson-Differenzenschrittes 6.12 beinhaltet die expliziten und impliziten Verfahren, wenn man $(I+\frac{1}{2}D)^{-1}$ durch $(I-\frac{1}{2}D)$ bzw. $(I-\frac{1}{2}D)^{-1}$ durch $(I+\frac{1}{2}D)$ approximiert.* □

6.1.3 Crank-Nicolson-Verfahren/lokaler Volatilität

Wir wollen uns wieder mit der numerischen Lösung der Aufgabe

$$\frac{\partial C}{\partial t} + \frac{1}{2}\sigma(S,t)^2 S^2 \frac{\partial^2 C}{\partial S^2} + (r-q)S\frac{\partial C}{\partial S} - rC \;=\; 0\,,\; S \in (0,\infty), t \in (0,T)\,, \tag{6.17}$$

$$C(0,t) \;=\; 0\,,\; C(\overline{S},t) - e^{-q(T-t)}\overline{S} \;=\; 0\,,\; t \in (0,T)\,, \tag{6.18}$$

$$C(S,T) \;=\; (K-S)^+\,,\; S > 0\,. \tag{6.19}$$

mittels eines Differenzenverfahren beschäftigen, und zwar in der Situation, dass eine lokale Volatilität angenommen wird. Dabei ist zunächst $\sigma : [0,\infty) \times [0,T] \longrightarrow \mathbb{R}$ eine stetige Funktion.

Wir wählen eine Ortschrittweite Δx und eine Zeitschrittweite Δt :

$$\Delta x := \overline{S}/N\,,\; \Delta t := T/M\,.$$

Damit ergeben sich die Gitterpunkte

$$(x_i, t_j) := (i\Delta x, j\Delta t)\,,\; i = 0, \ldots, N, j = 0, \ldots, M\,.$$

Schließlich benutzen wir die Kurzschreibweise

$$v_i^j := V(x_i, t_j) \text{ bzw. } \eta_i^j := \frac{1}{2}\sigma(x_i, t_j)^2\,,\; i = 0, \ldots, N, j = 0, \ldots, M\,,$$

wobei wir die Parameter K, T unterdrückt haben. Die Diskretisierung der Ableitungen in der Aufgabe erfolgt mit Differenzenquotienten gemäß

$$\frac{\partial V}{\partial S}(x_i, t_j) \approx \frac{v_{i+1}^j - v_{i-1}^j}{2\Delta x}$$

und

$$\frac{\partial^2 V}{\partial S^2}(x_i, t_j) \approx \frac{v_{i+1}^j - 2v_i^j + v_{i-1}^j}{(\Delta x)^2}\,.$$

Damit lassen sich nun die Gleichungen

$$\frac{v_i^{j+1} - v_i^j}{\Delta t} = -\eta_i^j x_i^2 \frac{v_{i+1}^j - 2v_i^j + v_{i-1}^j}{(\Delta x)^2} - (r-q)x_i \frac{v_{i+1}^j - v_{i-1}^j}{2\Delta x} + rv_i^j \tag{6.20}$$

$$\frac{v_i^{j+1} - v_i^j}{\Delta t} = -\eta_i^{j+1} x_i^2 \frac{v_{i+1}^{j+1} - 2v_i^{j+1} + v_{i-1}^{j+1}}{(\Delta x)^2} - (r-q)x_i \frac{v_{i+1}^{j+1} - v_{i-1}^{j+1}}{2\Delta x} + rv_i^{j+1} \tag{6.21}$$

hinschreiben. Das **Crank-Nicolson-Verfahren** entsteht durch das arithmetische Mittel der beiden Gleichungen:

$$\frac{v_i^{j+1} - v_i^j}{\Delta t} = -\frac{x_i^2}{2(\Delta x)^2}\Big(\eta_i^{j+1}(v_{i+1}^{j+1} - 2v_i^{j+1} + v_{i-1}^{j+1}) + \eta_i^j(v_{i+1}^j - 2v_i^j + v_{i-1}^j)\Big) \tag{6.22}$$

$$-\frac{(r-q)x_i}{4(\Delta x)^2}(v_{i+1}^{j+1} - v_{i-1}^{j+1} + v_{i+1}^j - v_{i-1}^j) + \frac{1}{2}r(v_i^{j+1} + v_i^j) \tag{6.23}$$

Dies ist eine Gleichung auf jeder der $M-1$ Zeitschichten. Aus den Rand- und Anfangsbedingungen erhalten wir

$$v_0^j = 0\,,\ v_{N+1}^j = x_N^j + \Delta x\,,\ v_i^M = (i\Delta x - K)^+\,,\ i = 0, \ldots, N, j = 0, \ldots, M\,. \tag{6.24}$$

Diesen Gleichungen für die $M \times N$ Unbekannten kann man nun ausgehend von der Zeitschicht $j = M$, hier sind alle Unbekannten bekannt, für $j = M-1, \ldots, 0$ nach vorne rechnen. Auf jeder Zeitschicht ist ein lineares Gleichungssystem zu lösen. Dabei kann man die Struktur der Tridiagonalität ausnutzen. Ist die Volatilität nicht zeitabhängig, muss man das Gleichungssystem für alle Zeitschichten nur einmal generieren und das Vorgehen entspricht im Wesentlichen dem Algorithmus 6.1.1.

Es sind nun die Anmerkungen zum Crank-Nicolson-Verfahren aus dem letzten Abschnitt zu bedenken, insbesondere, was das Start-up-Vorgehen betrifft, um die Nichtglattheit der Anfangsdaten zu berücksichtigen.

6.2 Die Diskretisierung der Dupire-Gleichung

Wir werden später die Gleichung nach Dupire etwas genauer betrachten. Hier siue sie schon einmal angeführt. Sie ist eine Gleichung, die mit Rand- und Anfangsbedingungen zusammen, den Preis C einer europäischen Option in Abhängigkeit vom Ausübungspreis K und dem Ausübungsdatum T beschreibt. Sie lautet:

$$\frac{\partial C}{\partial T} = \frac{1}{2}K^2\sigma(K,T)^2\frac{\partial^2 C}{\partial x^2} - (q-r)K\frac{\partial C}{\partial K} - qC \tag{6.25}$$

Dazu kommen die Rand- und Endbedingungen

$$C(K, T; S, t) = 0, \lim_{K\to\infty} C(K, T; S, t) = 0 \tag{6.26}$$

bzw.

$$C(K,t;S,t) = (S-K)^+ \, , \, K \in \mathbb{R}_+ \, . \tag{6.27}$$

Dabei haben σ, r, q wieder die üblichen Bedeutungen.

Zur Diskretisierung kann man nun die Ideen aus den letzten Abschnitten heranziehen.

6.3 Diskretisierung einer stochastischen Differentialgleichung

Wir skizzieren die wesentlichen Fragen bei der Diskretisierung von stochastischen Differentialgleichungen.

6.3.1 Das Euler-Verfahren

Sei $(W_t)_{t\in[0,T]}$ ein Wiener-Prozess mit Filtration $\hat{\mathcal{F}}$. Betrachte die (Itô-)stochastische Differentialgleichung

$$dX = a(X,t)dt + b(X,t)dW_t \, , \, X_0 =: x_0 \in \mathbb{R} \text{ gegeben} \, . \tag{6.28}$$

Eine **Lösung** in $[0,T]$ ist ein stochastischer Prozess $(X_t)_{t\in[0,T]}$, adaptiert bezüglich $\hat{\mathcal{F}}$, so dass die Gleichheit

$$X_t = X_0 + \int_0^t a(X_s,s)ds + \int_0^t b(X_s,s)dW_s \, , \, t \in [0,T] \, , \tag{6.29}$$

gilt mit Wahrscheinlichkeit 1. Die derartige Lösung – wir nennen sie eine **starke Lösung** – hängt also offensichtlich von der vorgegebenen Filtration $\hat{\mathcal{F}}$ (und dem damit zusammenhängenden Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$) ab. Eine Alternative zum obigen Lösungsbegriff sind schwache Lösungen; siehe [18].

Um eine Integralgleichung (6.28) zu lösen, sind nun Annahmen zu den Prozessen a, b zu machen, damit bei einem Itô–Prozess X_t, eingesetzt in die rechte Seite der Gleichung, wieder ein Itô–Prozess entsteht. Dies führt Wachstumsannahmen für a, b. Was die Existenzfrage betrifft, an einem Fixpunktargument wird man nicht vorbeikommen. Dazu sind dann Lipschitzannahmen nötig. Wir verweisen auf die Literatur. Jedenfalls können wir annehmen, dass es hinreichend allgemeine Voraussetzungen gibt, die die Lösbarkeit von (6.28) sichern und als Konsequenz sicherstellen, dass die Lösung X_t selbst wieder ein Itô–Prozess ist.

Wir diskretisieren nach der **expliziten Euler-Methode**, welche im Zusammenhang mit stochastischen Differentialgleichungen das Verfahren von **Euler-Maruyama** heißt. Die auftretenden infinitesimalen Inkremente werden dabei durch finite Zuwächse ersetzt. Dazu wählen wir ein Zeitgitter $0 = t_0 < t_1 < \cdots < t_N = T$ und führen die Bezeichnungen

$$\Delta t_j := t_{j+1} - t_j, X_j := X_{t_j}, W_j := W_{t_j}, \Delta W_j := W_{j+1} - W_j$$

ein. Damit stellt sich das Diskretisierungsverfahren so dar:

$$X_{j+1} = X_j + a(t_j, X_j)\Delta t_j + b(t_j, X_j)\Delta W_j \, , \, j = 0, \ldots, N-1 \, . \tag{6.30}$$

Durch rekursives Einsetzen entsteht

$$X_N = X_0 + \sum_{j=0}^{N-1} a(t_j, X_j)\Delta t_j + \sum_{j=0}^{N-1} b(t_j, X_j)\Delta W_j \, . \tag{6.31}$$

Damit ist klar, was ein Konvergenzsatz für das Vorgehen leisten soll: die Integralgleichung in (6.29) sollte als Grenzwert aus (6.31) für $h := \max \Delta t_j \to 0$ entstehen. Dieses Verfahren ist wegen seiner Einfachheit eine Methode, die direkt auch auf Systeme übertragen werden kann.

Algorithm 6.2 Simulation mit dem Euler–Maruyama-Verfahren

EIN Stochastische Differentialgleichung

$$X_t = X_0 + \int_0^t a(X_s, s)ds + \int_0^t b(X_s, s)dW_s\,.$$

Parameter: Endzeit T, Anzahl der Zerlegungspunkte N. Mechanismus zur Erzeugung von standard-normal-verteilten Zufallszahlen.

Schritt 0 $h := T/N$. $t_0 := 0$.

Schritt 1 Für $k = 0, \ldots, N-1$

- $t_{k+1} := t_k + h$
- erzeuge eine Zufallszahl u^k und setze $\Delta W_k := u^k\sqrt{h}$
- setze $X_{k+1} := X_k + a(X_k, t_k)h + b(X_k, t_k)\Delta W_k$

AUS Diskrete Approximation $X_0, \ldots, X_N$ der gegebenen stochastischen Differentialgleichung in $[0, T]$.

Das Ergebnis aus obigem Algorithmus kann durch lineare Interpolation zu einem stetigen Pfad vervollständigt werden.

Wir haben schon die Idee der Varianzreduktion kennengelernt. Sie kann man mit dem Euler-Maruyama-Verfahren in Verbindung bringen. Dazu verwende man in Schritt 1 die Zufallszahl u^k bzw. $\Delta W_k := u^k\sqrt{h}$ zweimal und erzeugt die „diskreten Pfade“ gemäß

$$X_{k+1}^+ := X_k + a(X_k^+, t_k)h + b(X_k^+, t_k)\Delta W_k\,,\ X_{k+1}^- := X_k^- + a(X_k^-, t_k)h - b(X_k^-, t_k)\Delta W_k\,. \tag{6.32}$$

Bemerkung 6.3 *Wenn wir den Algorithmus 6.3.1 zur Berechnung des Optionspreises entlang der geometrischen Bewegung*

$$dS = rS\,dt + \sigma S\,dW_t$$

verwenden, ergeben sich unter Verwendung antithetischer Variablen gemäß (6.32) *zwei Schätzer für die Auszahlungsfunktion:*

$$\theta_n^+ := (S_n^+ - K)^+\,,\ \theta_n^- := (S_n^- - K)^+\,,$$

Diese fügen wir zu einem Schätzer zusammen:

$$\hat{\theta_n} := \frac{1}{2}(\theta_n^+ + \theta_n^-)\,.$$

Daraus ergibt sich dann eine beträchtliche Genauigkeitssteigerung für die Diskretisierungsmethode. □

Man kann das Euler-Maruyama-Verfahren zu einem halbimpliziten Verfahren abändern, indem man den deterministischen Teil implizit gestaltet, d.h. (6.30) wird ersetzt durch

$$X_{j+1} = X_j + a(t_j, X_{j+1})\Delta t_j + b(t_j, X_j)\Delta W_j\,,\ j = 0, \ldots, N-1\,. \tag{6.33}$$

Hier ist ein (im Allgemeinen) nichtlineares Gleichungssystem zu lösen. Der stochastische Anteil kann jedoch nicht implizit behandelt werden, die Definition des Itô-Integrals erzwingt, zumindest wenn man auf größere „Vorsichtsmaßnahmen“verzichtet, die Auswertung der Funktion b am jeweils linken Rand der Diskretisierungsintervalle.

Um die Güte eines numerischen Verfahrens für Differentialgleichungen zu beurteilen, ist die Konvergenzordnung von Bedeutung. Im Gegensatz zu gewöhnlichen Differentialgleichungen ergeben sich bei stochastischen Differentialgleichungen zwei verschiedene Konzepte.

Die Lösung einer stochastischen Differentialgleichung ist eine Funktion der Zeit $t \geq 0$ und in den Elementarereignissen $\omega \in \Omega$. Zu einem festen Zeitpunkt $T > 0$ kann man den gemittelten absoluten Fehler über alle Elementarereignisse der numerischen Approximation betrachten. Daraus resultiert ein erstes Konvergenzkonzept.

Definition 6.4 *Sei $(X_t)_{t\in[0,T]}$ die Lösung einer stochastischen Differentialgleichung und sei $(X_t^h)_{t\in[0,T]}$ die Näherung aus einem numerischen Verfahren mit Schrittweite $h > 0$. Man sagt, dass die Approximation X_T^h* **stark mit Ordnung (mindestens)** *γ gegen X_T konvergiert, falls für alle hinreichend kleinen Schrittweiten h die Abschätzung*

$$\mathbb{E}(|X_T^h - X_T|) \leq Ch^{\gamma} \tag{6.34}$$

mit einer von h unabhängigen Konstanten $C > 0$ gilt. □

Ein zweites Konzept ergibt sich daraus, dass man häufig nicht an der Lösung einer stochastischen Differentialgleichung selbst interessiert ist, sondern nur an Maßzahlen wie etwa den Momenten der Lösung. Auf den Erwartungswert bezogen, resultiert das folgende Konvergenzkonzept.

Definition 6.5 *Sei $(X_t)_{t\in[0,T]}$ die Lösung einer stochastischen Differentialgleichung und sei $(X_t^h)_{t\in[0,T]}$ die Näherung aus einem numerischen Verfahren mit Schrittweite $h > 0$. Man sagt, dass die Approximation X_T^h* **schwach mit Ordnung (mindestens)** *β gegen X_T konvergiert, falls für alle hinreichend kleinen Schrittweiten h die Abschätzung*

$$|\mathbb{E}(X_T^h) - \mathbb{E}(X_T)| \leq Ch^{\beta} \tag{6.35}$$

mit einer von h unabhängigen Konstanten $C > 0$ gilt. □

Aus der starken Konvergenzordnung γ folgt die schwache Konvergenzordnung $\beta := \gamma$, denn

$$\begin{aligned} |\mathbb{E}(X_T^h) - \mathbb{E}(X_T)| &= |\mathbb{E}(X_T^h - X_T)| = \left| \int_\Omega (X_T^h - X_T) dP \right| \\ &\leq \int_\Omega |X_T^h - X_T| dP = \mathbb{E}(|X_T^h - X_T|) \end{aligned}$$

Das Euler-Maruyama-Verfahren besitzt die schwache Konvergenzordnung $\gamma = 1$ und die starke Konvergenzordnung $\beta = \frac{1}{2}$.

6.3.2 Das Milstein-Verfahren

Um numerische Verfahren höherer Konvergenzordnung zu konstruieren, nutzt man stochastische Taylorentwicklungen, also Vorgehensweisen, die bei gewöhnlichen Differentialgleichungen nicht sehr interessant sind. Damit entstehen dann Verfahren, die den stochastischen Anteil besser approximieren. Es sollte nicht sehr überraschen, dass Verfahren, die in die Zukunft greifen (implizite Verfahren), wie sie bei gewöhnlichen Differentialgleichungen üblich sind, in der Situation von stochastischen Differentialgleichungen nicht weiterhelfen: es ist wie beim Integralbegriff, Unabhängigkeitsargumente fehlen dann.

Das **Milstein**–Verfahren, das wir nun aufschreiben wollen, ist ein Verfahren, das auf einer Taylorentwicklung beruht. Es besitzt die starke Konvergenzordnung $\gamma = 1$ und die schwache Konvergenzordnung $\beta = 1$.

Algorithm 6.3 Das Verfahren von Milstein

EIN Stochastische Differentialgleichung

$$X_t = X_0 + \int_0^t a(X_s, s)ds + \int_0^t b(X_s, s)dW_s$$

mit $b' := \frac{\partial b}{\partial x}$.
Parameter: Endzeit T, Anzahl der Zerlegungspunkte N. Mechanismus zur Erzeugung von standard-normal-verteilten Zufallszahlen.

Schritt 0 $h := T/N$. $t_0 := 0$.

Schritt 1 Für $k = 0, \dots, N-1$

- $t_{k+1} := t_k + h$
- erzeuge eine Zufallszahl u^k und setze $\Delta W_k := u^k\sqrt{h}$
- setze $X_{k+1} := X_k + a(X_k, t_k)h + b(X_k, t_k)\Delta W_k + \frac{1}{2}b(X_k, t_k)b'(X_k, t_k)((\Delta W_k)^2 - h)$.

AUS Diskrete Approximation $X_0, \dots, X_N$ der gegebenen stochastischen Differentialgleichung in $[0, T]$.

Das Ergebnis aus obigem Algorithmus kann durch lineare Interpolation zu einem stetigen Pfad vervollständigt werden. Die Ableitung des Diffusionsterms b kann man unter Verwendung von Differenzenquotienten vermeiden.

Für das Milstein-Verfahren gilt das über die Varianzreduktion im Zusammenhang mit dem Euler-Maruyama Gesagte natürlich entsprechend.

6.4 Fourier-Transformation und Bewertung von Optionen

Wir wollen die Verwendung der Fouriertransformation zur Bewertung von Optionen skizzieren. Analytisch bedeutet dies, dass die Verteilungsdichte der logarithmierten Preise bekannt ist. Numerisch liegt der Vorteil darin, dass die schnelle Fouriertransformation zur Verfügung steht.

Wir nehmen einen Markt an, der den „Geschäftsbedingungen“ genügt und in dem Arbitragefreiheit herrscht. Der Preis einer europäischen Call-Option kann dann als der diskontierte Erwartungswert bezüglich der risiko-neutralen Wahrscheinlichkeit aufgeschrieben werden:

$$C(S, 0; T, K) \quad = \quad e^{-rT}\mathbb{E}((S_T - K)^+) \tag{6.36}$$

Dabei ist S_T der Endwert eines stochastischen Prozesses, der zur Zeit $t = 0$ bei S startet und die Entwicklung des Basiskurses beschreibt.

6.4.1 Eine Preisformel und seine Fouriertransformierte

Eine Umformulierung der Preisformel, die aus der Black-Scholes-Gleichung resultiert (siehe (1.2)), führt zu folgender Darstellung des Preises aus (6.36):

$$C(S, t; T, K) \quad = \quad e^{-r(t-T)}\left(F(t, T)\mathcal{S}(S_T > K) - K\mathcal{P}(S_T - K)\right) , \; t \in [0, T] . \tag{6.37}$$

Im Vergleich mit der Black-Scholes-Preisformel stellen wir fest:

$$F(t,T) = e^{r(T-t)}S \tag{6.38}$$
$$\mathcal{S}(S_T > K) = e^{r(T-t)}\mathcal{N}(d_+(\sigma)) \tag{6.39}$$
$$\mathcal{P}(S_T > K) = \mathcal{N}(d_-(\sigma)) \tag{6.40}$$

Hierbei ist folgende Interpretation üblich (siehe [12]):

- $F(t,T)$: Vorwärtspreis des Basisobjekts.
- $\mathcal{S}(S_T > K)$: „Wahrscheinlichkeit“, dass der Basiskurs zur Zeit T *in the money* ist.
- $\mathcal{P}(S_T > K)$: Risiko-neutrale Wahrscheinlichkeit, dass der Basiskurs zur Zeit T *in the money* ist.

Wir nehmen an:

Annahmen 6.6

1. *k ist der logarithmierte Ausübungspreis: $k = \ln(K)$.*
2. *s_T ist der logarithmierte Basiskurs zur Zeit T: $s_T = \ln(S_T)$.*
3. *Die risiko-neutrale Verteilung von s_T hat eine Dichte q_T.*

Unter diesen Voraussetzungen können wir die **charakteristische Funktion** Φ_T der Dichte q_T. definieren:

$$\Phi_T(u) := \mathbb{E}(e^{is_T}) = \int_{-\infty}^{\infty} e^{ius} q_T(s)ds\,,\ u \in \mathbb{R}\,. \tag{6.41}$$

Die Dichte q_T lässt sich mit Hilfe der inversen Fouriertransformation aus Φ_T (unter geeigneten Voraussetzungen) zurückgewinnen:

$$q_T(s) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-ius}\Phi_T(u)du\,,\ s \in \mathbb{R}\,. \tag{6.42}$$

Mit der charakteristischen Funktion der Dichte q_T können wir die Terme in der Preisformel so darstellen (siehe [8, 11]):

$$\mathcal{S}(S_T > K) = \frac{1}{2} + \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{e^{-iuk}\Phi_T(u-i)}{iu\Phi_T(-i)}du \tag{6.43}$$
$$\mathcal{P}(S_T > K) = \frac{1}{2} + \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{e^{-iuk}\Phi_T(u-i)}{iu}du \tag{6.44}$$
$$\tag{6.45}$$

Ist die Dichte q_T bekannt, dann kann mit der Fouriertransformation Φ_T berechnet werden. Mit der inversen Fouriertransformation kann man dann $\mathcal{S}(S_T > K), \mathcal{P}(S_T > K)$ berechnen. Beide Male kann man also die Fouriertransformation verwenden. Numerisch bedeutet dies, dass nach Diskretisierung die schnelle Fouriertransformation (siehe Anhang 6.5) zum Einsatz kommen kann.

Bemerkung 6.7 *Ein Preisprozess, der zur Diskontierung benutzt wird, wird als* **Numeraire(Bezugsgröße)** *bezeichnet. Wir arbeiteten bisher immer mit dem Numeraire der risikofreien Anlage, was ein deterministischer Prozess ist. Bei einer Änderung des Diskontierungsprozesses ändert sich ebenfalls das risiko-neutrale Wahrscheinlichkeitsmaß. Das Hilsmittel Numeraire wird besonders attraktiv, wenn wir es beim Modell mit mehreren beteiligten Prozessen zu tun haben, etwa mit zwei beteiligten Basisobjekten.* □

6.4.2 Eine gewichtete Preisformel

Ein anderer Weg, den Optionspreis unter Einsatz der Fouriertransformation zu berechnen, benutzt ebenfalls die charakteristischen Funktion der logarithmierten Basiskurse. Damit stellt sich der Preis der Option folgendermaßen dar:

$$C_T(k) \;:=\; C(S,0;T,K) \;=\; \int_k^\infty e^{-rT}(e^s - e^k)q_T(s)ds \tag{6.46}$$

Da $\lim_{k\to-\infty} C(S,0;T,e^k) = S$ ist, ist der Call-Preis $k \longmapsto C_T(k)$ nicht integrierbar. Eine Anwendung der Fouriertransformation ist also nicht unmittelbar möglich. Wir modifizieren daher den Call-Preis. Der einfachste Weg ist, den Callpreis mit einem Gewicht zu versehen, das dieses Defizit behebt. Wir nehmen ein exponentielles Gewicht an und definieren

$$c_T(k;\alpha) \;:=\; e^{\alpha k} C_T(k)\,,\; k \in [0,\infty)\,, \tag{6.47}$$

mit $\alpha \in \mathbb{R}$. Unter der Annahme, dass $k \longmapsto c_T(k;\alpha)$ integrierbar ist, definieren wir

$$\psi_T(\omega;\alpha) \;:=\; \int_{-\infty}^\infty e^{i\omega k} c_T(k;\alpha)dk\,,\; \omega\in\mathbb{R}\,. \tag{6.48}$$

Klar, $k \longmapsto c_T(k;\alpha)$ ist integrierbar genau dann, wenn $\psi_T(0;\alpha)$ endlich ist.

Wir haben

$$C_T(k) \;=\; \frac{e^{-\alpha k}}{2\pi}\int_{-\infty}^\infty e^{-i\omega k}\psi_T(\omega;\alpha)d\omega \tag{6.49}$$

$$\;=\; \frac{e^{-\alpha k}}{\pi}\int_0^\infty \Re(e^{-i\omega k}\psi_T(\omega;\alpha))d\omega \tag{6.50}$$

da $k \longmapsto C_T(k)$ reellwertig ist, was die Symmetrie von ψ_T impliziert. $\psi_T(\omega;\alpha)$ berechnet sich folgendermaßen:

$$\psi_T(\omega;\alpha) \;=\; \int_{-\infty}^\infty e^{i\omega k}\int_k^\infty e^{\alpha k - rT}(e^s - e^k)q_T(s)ds \tag{6.51}$$

$$= \int_{-\infty}^\infty e^{-rT}q_T(s)\int_{-\infty}^s (e^{s+\alpha k} - e^{(1+\alpha)k})e^{i\omega k}dkds$$

$$= \int_{-\infty}^\infty e^{-rT}q_T(s)\left(\frac{e^{(\alpha+1+i\omega)s}}{\alpha+i\omega} - \frac{e^{(\alpha+1+i\omega)s}}{\alpha+1+i\omega}\right)ds$$

$$= \frac{e^{-rT}\Phi_T(\omega-(\alpha+1)i)}{-(\omega-i\alpha)(\omega-i(\alpha+1))} \tag{6.52}$$

Zur Berechnung von $C_T(K)$ in (6.49) ist im Wesentlichen eine (inverse) Fouriertransformation von $\psi_T(\cdot;\alpha)$ durchzuführen.

- Die Singularität von $\psi_T(\cdot;\alpha)$ in $\alpha = 0$ zeigt, dass $\alpha = 0$ nicht möglich ist.
- Die Wahl $\alpha > 0$ verstärkt Chancen auf Integrabilität des Call-Preises auf der Achse $(-\infty,0]$, führt aber möglicherweise zu Integrabilitätsproblemen auf der Achse $[0,\infty)$. Hinreichend für die Integrabilität von $k \longmapsto c_T(k;\alpha)$ ist, wie wir schon wissen, die Endlichkeit von $\psi_T(0)$. Nach (6.52) ist damit äquivalent die Endlichkeit von $\Phi_T(-(\alpha+1)i)$. Dies bedeutet

$$\mathbb{E}(S_T^{\alpha+1}) \;<\; \infty \tag{6.53}$$

Dies bedeutet, dass das $(\alpha+1)$-te Moment existiert.

- Die Wahl $\alpha < 0$ ist nicht möglich, da schon für $\alpha = 0$ keine Integrabilität vorliegt.

Wir benötigen also zwei Fouriertransformation, um den Optionspreis $C_T(k)$ auszurechnen. Im nächsten Abschnitt skizzieren wir das nötige Vorgehen.

Die Berechnung einer kontinuierlichen Fouriertransformation führt zu zwei Fehlern:**Abschneidefehler** und **Diskretisierungsfehler**.

Der Abschneidefehler kommt ins Spiel, wenn das unendliche $[0,\infty)$, über das zu integrieren ist, abgeschnitten wird. Der Fehler wird beherrschbar, wenn der Integrand hinreichend schnell abklingt. Unter der Voraussetzung (6.53) haben wir

$$|\psi_T(\cdot;\alpha)|^2 \quad \leq \quad \frac{\mathbb{E}(S_T^{\alpha+1})}{(\alpha^2+\alpha-\omega^2)^2+(2\alpha+1)^2\omega^2} \;\leq\; \frac{A}{\omega^4} \tag{6.54}$$

mit einer Konstanten A. Dies impliziert

$$\int_a^\infty \psi_T(\cdot;\alpha)d\omega \quad < \quad \frac{\sqrt{A}}{a}\,. \tag{6.55}$$

Also können wir das Integral $\int_0^\infty |\psi_T(\cdot;\alpha)|d\omega$ abschneiden und erhalten als Schranke für den Abschneidefehler

$$\frac{e^{-\alpha k}}{\pi}\frac{\sqrt{A}}{a} \tag{6.56}$$

Damit ist die Wahl von a gemäß

$$a \quad > \quad \frac{e^{-\alpha k}}{\pi}\frac{\sqrt{A}}{\varepsilon} \tag{6.57}$$

nötig, um den Abschneidefehler kleiner als ε zu machen.

Der Diskretisierungsfehler entspricht im Wesentlichen dem Fehler, der entsteht, wenn man eine Integralformel auf das Integral anwendet. Dieser Fehler wrd beherrschbar, wenn der Integrand hinreichend glatt ist.

6.4.3 Die diskrete Fouriertransformation

Die diskrete Fouriertransformation, die wir nun besprechen wollen, ist eine eigenständige Transformation, die allerdings in Anlehnung an die kontinuierliche Fouriertransformation definiert ist. **Beachte:** Die inverse Fouriertransformation kann umgeschrieben werden in eine Fouriertransformation.

Betrachte eine Funktion $S : [0, 2\pi] \longrightarrow \mathbb{C}$, die 2π-periodisch ist. Wir approximieren das Integral für den k–ten Fourierkoeffizient

$$\hat{S}_k := \int_0^{2\pi} S(t)e^{-ikt}dt\,,$$

durch die Rechteckregel bezüglich einer äquidistanten Zerlegung des Intervalls $[0, 2\pi]$ mit n Teilintervallen. Das Ergebnis ist – $\hat{z}_k$ steht für eine Approximation von $\hat{S}_k$ –

$$\hat{z}_k := \sum_{j=0}^{n-1} S(2j\pi/n)e^{-2\pi ijk/n}\,.$$

Wegen der Identität $e^{-2\pi ijk/n} = e^{-2\pi ij(k+n)/n}$ ist die Zuordnung

$$\mathbb{Z} \ni k \longmapsto \hat{z}_k \in \mathbb{C}$$

n–periodisch. Also ist die ganze Information der Folge der Fourierkoeffizienten von S im Vektor

$$\hat{z} := (\hat{z}_0, \ldots, \hat{z}_{n-1})$$

enthalten. Wir werden sehen, dass dieser Vektor im wesentlichen das Bild von

$$z := (S(0), S(2\pi/n) \ldots, S(2(n-1)\pi/n)) \in \mathbb{C}^{n-1}$$

unter der diskreten Fouriertransformation ist.

Sei $n \in \mathbb{N}, n \geq 2$. Wir setzen

$$\omega_n := \exp(-\frac{2\pi i}{n})$$

und haben $\omega_n^n = 1$. Beachte, dass diese **n-te Einheitswurzel** ω_n etwas verrät über die Konstruktion von regulären Polyedern, denn sie ist eine Lösung der Gleichung $x^n = 1$, und dass ω_n eine zyklische Gruppe der Ordnung n erzeugt: $\omega_n^{ij} = \omega_n^{ij \bmod n}$ für alle $i, j \in \mathbb{Z}$ gilt.

Lemma 6.8 *Wir haben*

$$\frac{1}{n}\sum_{j=0}^{n-1} \omega_n^{(l-k)j} = \begin{cases} 1 & \text{falls } l = k \\ 0 & \text{falls } l \neq k \end{cases}, \; k, l = 0, \ldots, n-1\,. \tag{6.58}$$

Beweis:
Ist $k = l$, dann haben wir nichts zu zeigen. Sei $k \neq l$. O.E. $k > l$.
Dann ist

$$\sum_{j=0}^{n-1} \omega_n^{(l-k)j} = \frac{1 - \omega_n^{(l-k)n}}{1 - \omega_n^{(l-k)}} = 0,$$

da $\omega_n^{(l-k)} \neq 1$ dank der Tatsache, dass $0 < k - l < n$ und $\omega_n^{(l-k)n} = e^{2\pi i(k-l)} = 1$. ■

Betrachte das **trigonometrische Interpolationsproblem**:

> Gegeben Stützstellen $0 = t_0 < \cdots < t_{n-1} < 2\pi$,
> gegeben Stützwerte $y_0, \ldots, y_{n-1} \in \mathbb{C}$.
> Gesucht ist ein trigonometrisches interpolierendes Polynom p vom Grade $n-1$ mit $p(t_k) = y_k, k = 0, \ldots, n-1$.

Dabei heißt ein Polynom p trigonometrisch vom Grade $n-1$, wenn p eine Linearkombination der Funktionen $t \longmapsto 1 = e^{i0t}, \ldots, t \longmapsto e^{i(n-1)t}$ ist.

Satz 6.9 *Betrachte die obige Interpolationsaufgabe für äquidistante Knoten:* $t_k = \frac{2\pi}{n}k, k = 0, \ldots, n-1$. *Dann ist das zugehörige interpolierende trigonometrische Polynom p gegeben durch*

$$p(t) = \sum_{j=0}^{n-1} c_j e^{ijt}, \; t \in \mathbb{R}, \; \textit{wobei } c_j = \frac{1}{n}\sum_{l=0}^{n-1} \omega_n^{jl} y_l \tag{6.59}$$

Beweis:
Da $\omega_n^{jk} = e^{-ijt_k}$ ist, erhalten wir

$$p(t_k) = \sum_{j=0}^{n-1}\left(\frac{1}{n}\sum_{l=0}^{n-1}\omega_n^{jl}y_l\right)\omega_n^{-jk} = \sum_{l=0}^{n-1} y_l\left(\frac{1}{n}\sum_{j=0}^{n-1}\omega_n^{(l-k)j}\right) = y_k\,,\ k = 0,\dots,n-1\,;$$

siehe Lemma 6.8. ■

Definition 6.10 *Die Abbildung* $\hat{\ } : \mathbb{C}^n \longrightarrow \mathbb{C}^n, z \longmapsto \hat{z}$, *mit*

$$\hat{z}_j := \sum_{k=0}^{n-1}\omega_n^{jk} z_k\,,\ j = 0,\dots,n-1\,, \tag{6.60}$$

wird **(eindimensionale) diskrete Fourier Transformation (DFT)** *genannt. Mitunter schreiben wir* DFT_n*, wenn wir die Abhängigkeit von der Größe* n *betonen wollen, oder auch* $DFT_{1,n}$*, wenn wir auch die Dimensionalität vermerken wollen.* □

Beachte, dass $\hat{z}_0$ bis auf den Skalierungsvektor $1/n$ den Mittelwert des Vektors z darstellt.[1]

Bemerkung 6.11 *Satz 6.9 besagt, dass die diskrete Fouriertransformation genutzt werden kann, um im Falle äquidistanter Knoten die Koeffizienten* $c := (c_0,\dots,c_{n-1})$ *des interpolierenden Polynoms zu berechnen:*

$$c = \hat{y} \text{ wobei } y := (y_0,\dots,y_{n-1})\,.$$

□

Lemma 6.12 *Sei* $z = (z_0,\dots,z_{n-1}) \in \mathbb{C}^n$*. Dann gilt:*

$$z_l = \frac{1}{n}\sum_{j=0}^{n-1}\omega_n^{-jl}\hat{z}_j\,,\ l = 0,\dots,n-1\,. \tag{6.61}$$

Beweis:
Wir haben mit Lemma 6.8 für $l = 0,\dots,n-1$

$$\frac{1}{n}\sum_{j=0}^{n-1}\omega_n^{-jl}\hat{z}_j = \frac{1}{n}\sum_{j=0}^{n-1}\sum_{k=0}^{n-1}\omega_n^{-jl}\omega_n^{kj} = \sum_{k=0}^{n-1} z_k\frac{1}{n}\sum_{j=0}^{n-1}\omega_n^{(k-l)j} = z_l\,.$$

■

Definition 6.13 *Die Abbildung* $\check{\ } : \mathbb{C}^n \longrightarrow \mathbb{C}^n, z \longmapsto \check{z}$, *mit*

$$\check{z}_j := \frac{1}{n}\sum_{k=0}^{n-1}\omega_n^{-jk} z_k\,,\ j = 0,\dots,n-1\,, \tag{6.62}$$

heißt **inverse diskrete Fourier Transformation (IDFT)**. □

Folgerung 6.14

(a) $\hat{\ }, \check{\ } : \mathbb{C}^n \longrightarrow \mathbb{C}^n$ *sind lineare und bijektive Abbildungen.*

[1] Hier folgen wir der MATLAB-Konvention:der Vorfaktor $\frac{1}{n}$ wird bei der inversen diskreten Fouriertransformation angebracht; siehe unten.

(b)

$$\sum_{k=0}^{n-1} z_k\overline{w_k} = \frac{1}{n}\sum_{j=0}^{n-1} \hat{z}_j\overline{\hat{w}_j}\,,\ z=(z_0,\dots,z_{n-1}), w=(w_0,\dots,w_{n-1})\,.$$

(c)

$$\sum_{k=0}^{n-1} |z_k|^2 = \frac{1}{n}\sum_{j=0}^{n-1} |\hat{z}_j|^2\,,\ z=(z_0,\dots,z_{n-1})\,.$$

Beweis:
Die Linearität ist klar. Die Bijektivität folgt aus Lemma 6.12. Den Beweis von (b) übergehen wir. Zum Beweis von (c).

$$\begin{aligned}
\sum_{k=0}^{n-1} |z_k|^2 &= \sum_{k=0}^{n-1} \frac{1}{n^2}\left(\sum_{j=0}^{n-1}\omega_n^{-jk}\hat{z}_j\right)\left(\overline{\sum_{l=0}^{n-1}\omega_n^{-lk}\hat{z}_l}\right)\\
&= \sum_{k=0}^{n-1} \frac{1}{n^2}\sum_{j=0}^{n-1}\sum_{l=0}^{n-1}\omega_n^{(l-j)k}\hat{z}_j\overline{\hat{z}_l}\\
&= \sum_{j=0}^{n-1}\hat{z}_j\sum_{l=0}^{n-1}\overline{\hat{z}_l}\frac{1}{n^2}\sum_{k=0}^{n-1}\omega_n^{(l-j)k}\\
&= \sum_{j=0}^{n-1}\hat{z}_j\overline{\hat{z}_j}\frac{1}{n}
\end{aligned}$$

■

Bemerkung 6.15 *Diskrete und inverse diskrete Fouriertransformation unterscheiden sich formelmäßig und aufwandsmäßig nur marginal: hat man eine Implementierung für die eine Transformation, hat man eine für die andere.*

Man kann das Ergebnis $\hat{z}_j$ der diskreten Fouriertransformation von z als Auswertungsergebnis für das Polynom

$$p_z(X) := \frac{1}{n}\sum_{i=0}^{n-1} z_i X^i$$

für $X := \omega_n^j$ ansehen. Die Überraschung ist, dass man aus dem Auswertungsergebnis mit nahezu derselben Transformation die Koeffizienten des Polynoms zurückgewinnen kann. Auf diesen offensichtlichen Hinweis kommen wir noch zurück, wenn wir die Fouriertransformation nutzen für die Faltung. □

Bemerkung 6.16 *Die diskrete Fouriertransformation eines Datenvektors kann als Matrixoperation auf diesem Datenvektor angesehen werden. Sei*

$$\hat{W} \;:=\; \frac{1}{\sqrt{n}}(\omega_n^{kj})_{1\le j,k\le n} \in \mathbb{C}^{n,n}\,,\ \check{W} \;:=\; \frac{1}{\sqrt{n}}(\omega_n^{-kj})_{1\le j,k\le n} \in \mathbb{C}^{n,n}\,.$$

Mit diesen Matrizen haben wir für[2] $z=(z_0,\dots,z_{n-1})\in\mathbb{C}^n$

$$\hat{z}_j \;=\; \frac{1}{\sqrt{n}}(\hat{W}z)_r\,,\ \check{z}_j \;=\; \sqrt{n}(\check{W}z)_r\,,\ j=0,\dots,n-1\,.$$

[2]Wir gehen mit Vektoren $z\in\mathbb{C}^n$ etwas locker um: mal als Zeilenvektor, mal als Spaltenvektor.

Offenbar gilt

$$\hat{W}\check{W}^t = \check{W}\hat{W}^t = E\,;$$

d.h. $\hat{W}$ ist unitär. Hierin liegt auch der Grund, dass die eingangs betrachtete Interpolationsaufgabe so einfach zu lösen war: die Vandermondsche Matrix $\hat{W}$ hat die einfach zu findende Inverse $\check{W}^t$.

$\square$

Es ist offensichtlich, dass die Auswertung der diskreten Fouriertransformation und der inversen diskreten Fouriertransformation in der direkten Umsetzung der definierenden Formeln einen Aufwand von $O(n^2)$ `flops` bedeuten.

6.5 Anhang: Schnelle Fouriertransformation

Hier beschreiben wir die Realisierung der **Schnellen Fouriertransformation (FFT)**.[3] Dies ist eine rekursive Methode, die die Auswertung der diskreten Fouriertransformation mit einem Aufwand von $O(n\log_2(n))$ `flops` realisiert. Dabei unterstellen wir, dass n eine Potenz von 2 ist, was für praktische Anwendungen keine sehr einschränkende Voraussetzung ist.

Sei $z := (z_0, \dots, z_{n-1})$. Wir wollen $Z := \hat{z}$ berechnen. Klar, wenn wir eine schnelle Methode zur Berechnung des Vektors $Z = (Z_0, \dots, Z_{n-1})$ haben, dann haben wir eine schnelle Methode zur Berechnung der inversen Fouriertransformation $\check{z}$; siehe Bemerkung 6.15, Dies leitet sich aus

$$\check{z}_{n-j} = \frac{1}{n}\sum_{k=0}^{n-1}\omega_n^{-k(n-j)}z_k = \sum_{k=0}^{n-1}\frac{1}{n}z_k\omega_n^{-kn+kj} = \sum_{l=0}^{n-1}\frac{1}{n}z_l\omega_n^{kj} = \hat{w}_j\,,\ j = 0, \dots, n-1\,,$$

ab, wobei $w = (\frac{1}{n}z_0, \dots, \frac{1}{n}z_{n-1})$ ist.

Die Einsparung von Aufwand in der Berechnung von Z ergibt sich – wie wir sehen werden – aus folgenden Eigenschaften:

Symmetrie $\omega_n^{(n-k)j} = \overline{\omega_n^{kj}}\,,\ k, j = 0, \dots, n-1\,.$

Periodizität $\omega_n^{kj} = \omega_n^{k(j\pm n)} = \omega_n^{(k\pm n)j}\,,\ k, j = 0, \dots, n-1\,.$

Hierin sind Eigenschaften der multiplikativen zyklischen Gruppe

$$\mathcal{R}_n := \{1, \omega_n, \dots, \omega_n^{n-1}\}$$

subsummiert.

Sei $n = 2^s, s \geq 2$, und $m := n/2 = 2^{s-1}$. Wir wollen die folgende Grundaufgabe der diskreten Fouriertransformation lösen:

Gegeben $z = (z_0, \dots, z_{n-1}) \in \mathbb{C}^n$,
berechne $Z := n\hat{z} = n(\hat{z}_0, \dots, \hat{z}_{n-1}) \in \mathbb{C}^n$.

[3]This transform was introduced in 1965 by Cooley and Tuckey. Actually, the FFT was discovered by Gauss in 1805 - two years before Fourier completed his first big article – but Gauss never published his paper on this subject. It was rediscovered several times.

Wir starten von der folgenden Beobachtung:

$$Z_j = \sum_{l=0}^{n-1} \omega_n^{kj} z_k = \sum_{l=0}^{m-1} \omega_n^{2lj} z_{2l} + \sum_{l=0}^{m-1} \omega_n^{(2l+1)j} z_{2l+1} = \sum_{l=0}^{m-1} (\omega_n^2)^{lj} z_{2l} + \omega_n^j \sum_{l=0}^{m-1} (\omega_n^2)^{lj} z_{2l+1} \,.$$

Wegen

$$\omega_n^2 = \omega_m \,,\, \omega_n^m = -1,$$

erhalten wir

$$Z_j \;=\; U_j + \omega_n^j V_j \,,\, Z_{j+m} \;=\; U_j - \omega_n^j V_j \,,\, j = 0, \ldots, m-1 \,,\text{ wobei} \tag{6.63}$$

$$U_j \;=\; \sum_{l=0}^{m-1} \omega_m^{lj} z_{2l} \,,\, V_j \;=\; \sum_{l=0}^{m-1} \omega_m^{lj} z_{2l+1} \,,\, j = 0, \ldots, m-1 \,. \tag{6.64}$$

Aus diesen Identitäten lesen wir ab, dass die Berechnung der diskreten Fouriertransformation der Größe n ausgeführt werden kann durch zweimalige Berechnung einer diskreten Fourier Transformation der Größe $m = n/2$. Was wir dazu zu tun haben, ist die Zerlegung des Datenvektors z der Größe n in die zwei Datenvektoren z^e, z^o der Größe $m = n/2$ zu organisieren, wobei z^e die Komponenten mit geraden und z^o die Komponenten mit den ungeraden Indizes aufnimmt. Nun können die zwei Aufgaben der Größe m in derselben Weise behandelt werden. Aus $n = 2^s$ mit $s = \log_2(n)$ lesen wir ab, dass das Resultat in $\log_2(n)$ Stufen berechnet werden kann. Man nennt so ein Vorgehen die Methode

n	2	16	256	1024
FT	4	256	65536	1048576
FFT	2	64	2048	10240
Einsparung	50 %	75 %	97 %	99 %

Abbildung 6.1: Schnelligkeitsvergleich

„spalte und herrsche“ **„divide and conquer“** **„divide et impera“**

MATLAB-illustration 6.1

The size n of the problem has to be a power of 2.

$a. * b$ denotes the „multiplication“ of the vectors $a = (a_0, \ldots, a_{n-1})$, $b = (b_0, \ldots, b_{n-1})$ by building the vector $(a_0 b_0, \ldots, a_{n-1} b_{n-1})$.

```
function Z = FFT(z_0, z_1, ..., z_{n-1})
    n = length(z); m = n/2;
    if n = 1
        Z = z return
    end
    U = FFT(z_0, z_2, ..., z_{n-2}),
    V = FFT(z_1, z_3, ..., z_{n-1})
    for k = 0 : m-1
        w_k = exp(-2 * Pi * i * k/n)
    end
    Z = (U + w. * V, U - w. * V)
```

Das obige Vorgehen, das sich aus der rekursiven Realisierung von (6.63), (6.63) ergibt, nennt man **schnelle Fouriertransformation/FFT**. Sie wird beschrieben durch folgendes Fragment eines Algorithmus (Codes für die schnelle Fouriertransformation liegen bei nahezu allen mathematischen Softwarepaketen vor). Das Feld, das das Signal beschreibt, wird als Parameter übergeben und solange in zwei Felder mit geradem und ungeradem Index aufgeteilt, bis es nur noch aus einem Element besteht.

Für die Implementierung ist es nützlich, sich die Verknüpfung der Komponenten des zu transformierenden Vektors in einem so genannten **Schmetterlingsgraphen** anzuschauen; siehe 6.2.

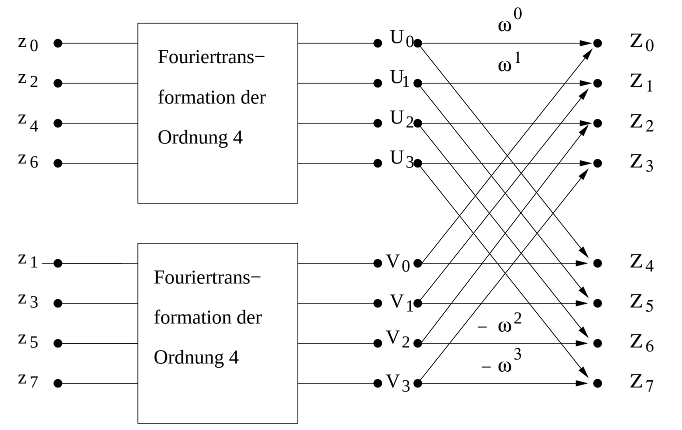

Abbildung 6.2: Schmetterlingsgraph

Seien $C(n)$ die Kosten für die Berechnung der diskreten Fouriertransformation für ein Problem der Größe n. Dann können wir also das Problem mit obigem Vorgehen mit einem Aufwand von

$$T(n) = 2T(n/2) + f(n) \qquad (6.65)$$

berechnen. Hierbei steht $f(n) := 3(n/2)$ für die Kosten in (6.63), die Kosten für die Berechnung von ω_n^j haben wir außer Acht gelassen.

Satz 6.17 *Sei n eine Zweierpotenz. Die Kosten für die Berechnung der diskreten Fouriertransformation von $z \in \mathbb{C}^n$ mit der schnellen Fouriertransformation sind*

$$O(n \log_2 n)\,.$$

Beweis:
Wir setzen $C(s) := T(2^s)$. offenbar $C(1) = 2$. Wir wollen

$$C(s) = 1.5s2^s\,,\ s \in \mathbb{N}\,,$$

zeigen. Wir tun dies induktiv. Offenbar haben wir $C(1) = 2$.

$$C(s+1) = 2C(s) + 3\,2^s = 3\,s2^s + 3\,2^s = 1.5(s+1)2^{s+1}\,.$$

■

Bemerkung 6.18 *Die Rekursionsformel* (6.66) *ist ein Spezialfall der allgemeinen Rekursion*

$$T(n) = aC(n/b) + f(n)\ \ (a \geq 1, b > 1) \qquad (6.66)$$

Dabei steht a für die Anzahl der Unterprobleme in der Rekursion, $1/b$ Teil des Originalproblems, welches durch alle Unterprobleme repräsentiert wird, $f(n)$ für den Aufwand, der durch die Aufteilung der Probleme entsteht.

Die Rekursion wird gelöst durch den **Hauptsatz der Laufzeitfunktionen**, *der für „alle“ Kombinationen von a, b, f das Verhalten von T in der Form*

$$T(n) = O(g(n)) \text{ mit einer aus den Daten } a, b, f \text{ ableitbaren Funktion } g$$

angibt. Im obigen Fall ist $a = 2, b = 2, f(n) = O(n)$ und der Hauptsatz gibt $T(n) = O(n^{\log_b a} \ln n) = O(n \ln n)$ an. □

Beispiel 6.19 *Betrachte die schnelle Fouriertransformation* Z *von* $z = (3, -2, 0, 1)$; *siehe das obige Fragment. Die Rekursion führt auf folgende Schritte:*
$U := 3, V := 0, w := 1, Z := (3 + 1 \cdot 0, 3 - 1 \cdot 0) = (3, 3)$.
$U := -2, V := 1, w := 1, Z := (-2 + 1 \cdot 1, -2 - 1 \cdot 1) = (-1, -3)$.
$U := (3, 3), V := (-1, 3), w := (1, i)$; *daher*

$$Z := ((3,3) + (-1, -3i), (3,3) - (-1, -3i)) = (2, 3-3i, 4, 3+3i)\,.$$

□

Neben der oben dargestellten Vorgehensweise – man bezeichnet sie auch als Radix-2-Methode – existieren noch eine Reihe weiterer Algorithmen, die sich als schnelle Fouriertransformation bezeichnen lassen.

Radix–4–Algorithmus

Der Unterschied besteht darin, dass man verlangt und ausnutzt, dass die Datenpunkte als Vektor der Länge $n := 4^s$ gegeben sind.

Cooley–Tukey–Algorithmus

Dieser Algorithmus berechnet die diskrete Fouriertransformierte DFT_n im Falle $n = pq$ mit $p, q > 1$. Dies gelingt durch Zusammensetzung des Ergebnisses aus einem Ergebnis von DFT_p und DFT_q.

Winograd–Algorithmus

Bei diesem Algorithmus ist nur eine bestimmte Datenlänge möglich:

$$n = \prod_{j=1}^{m} N_j \text{ mit } N_j \in \{2, 3, 4, 5, 7, 8, 9, 16\}\,,$$

wobei noch Einschränkungen zu berücksichtigen sind.

Goertzel–Algorithmus

Dieser Algorithmus ist eine schnelle Methode zur Berechnung einiger weniger Komponenten der Fouriertransformierten.

Im Abschnitt 6.5 haben wir schon darauf hingewiesen, dass die Berechnung der diskreten Fouriertransformation mit einer Multiplikation der Matrix

$$W := (w_n^{jk})_{0 \le j,k \le n-1}$$

mit dem Datenvektor erfolgen kann. Wie läßt sich in diesem Kontext die schnelle Fouriertransformation verstehen? Die Antwort darauf ist: W läßt sich geschickt faktorisieren, und zwar unter Ausnutzung der Periodizität in s „schwach besetzte“ Faktoren $W^1, \ldots, W^s \in \mathbb{C}^{n,n}$, d.h.

$$W = W^s \cdots W^1\,.$$

Die Berechnung der diskreten Fouriertransformierten $\hat{z}$ erfolgt dann so:

$$Z^0 := z\,,\ Z^{k+1} := W^k Z^k\,,\ k = 0, \ldots, s-1\,; \hat{z} := Z^s\,.$$

Machen wir uns die Zerlegung am Beispiel $n = 4$ klar; $\omega := \omega_4$. Hier sieht W so aus:

$$W = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & \omega^1 & \omega^2 & \omega^3 \\ 1 & \omega^2 & 1 & \omega^2 \\ 1 & \omega^3 & \omega^2 & \omega^1 \end{pmatrix}\,.$$

Wir wollen $Z := Wz$ berechnen und vertauschen dazu in Z die Komponenten Z_1, Z_2 und erhalten:

$$\begin{pmatrix} Z_0 \\ Z_2 \\ \hline Z_1 \\ Z_3 \end{pmatrix} = \left(\begin{array}{cc|cc} 1 & 1 & 1 & 1 \\ 1 & \omega^2 & 1 & \omega^2 \\ \hline 1 & \omega^1 & \omega^2 & \omega^3 \\ 1 & \omega^3 & \omega^2 & \omega^1 \end{array}\right) \begin{pmatrix} z_0 \\ z_1 \\ \hline z_2 \\ z_3 \end{pmatrix}, \; \tilde{Z} = \tilde{W} z \,.$$

Offensichtlich erfüllen die zweireihigen Untermatrizen $W_{11}, W_{12}, W_{21}, W_{22}$ von $\tilde{W}$ die Beziehungen

$$W_{12} = W_{11} \,,\; W_{22} = \omega^2 W_{21} \,.$$

Diese Tatsache legt eine Zerlegung von $\tilde{W}$ in das Produkt zweier Blockmatrizen nahe:

$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & \omega^2 & 1 & \omega^2 \\ 1 & \omega^1 & \omega^2 & \omega^3 \\ 1 & \omega^3 & \omega^2 & \omega^1 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & \omega^2 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & \omega^2 \end{pmatrix} \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & \omega^2 & 0 \\ 0 & \omega^1 & 0 & \omega^3 \end{pmatrix} .$$

Die Berechnung kann nun in zwei Teilschritten erfolgen, wobei die spezielle Strukur der Blockmatrizen tatsächlich eine Reduktion der erforderlichen komplexen Multiplikationen mit sich bringt.
1. Teilschritt (Beachte $\omega^2 = -1$)

$$\begin{pmatrix} Z_0^1 \\ Z_1^1 \\ Z_2^1 \\ Z_3^1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & -\omega^0 & 0 \\ 0 & \omega^1 & 0 & -\omega^1 \end{pmatrix} \begin{pmatrix} z_0 \\ z_1 \\ z_2 \\ z_3 \end{pmatrix} .$$

Bei diesem Rechenschritt benötigen wir 4 komplexe Multiplikationen und 2 komplexe Additionen.
2. Teilschritt (Im Hinblick auf den allgemeinen Fall wird $\omega^2 = -1$ nicht verwendet)

$$\begin{pmatrix} Z_0^2 \\ Z_1^2 \\ Z_2^2 \\ Z_3^2 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & \omega^2 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & \omega^2 \end{pmatrix} \begin{pmatrix} Z_0^1 \\ Z_1^1 \\ Z_2^1 \\ Z_3^1 \end{pmatrix} .$$

Dies entspricht zwei diskreten Fouriertransformationen der Größe 2, denn $\omega^2 = \omega_2$. Wir benötigen dazu 2 komplexe Multiplikationen und 4 komplexe Additionen. Insgesamt benötigen wir also 8 komplexe Additionen und 4 komplexe Multiplikationen. Dieser Aufwand entspricht dem Aufwand, der in Satz 6.17 angesagt wird.

Bemerkung 6.20 *Die Umordnungen von Vektoren nach geraden und ungeraden Komponenten kann sehr „billig“ mit Bitumkehr in der Binärdarstellung der Indizes vorgenommen werden. Hier ist beispielshaft der Fall* $n = 4$:

Nummerierung	*0*	*1*	*2*	*3*	*Umordnung*	*0*	*2*	*1*	*3*
	00	*01*	*10*	*11*		*00*	*10*	*01*	*11*

□

6.6 Bibliographische Anmerkungen

Standardergebnisse zu partiellen Differentialgleichungen findet man etwa in [34]. Die Diskretisierung von partiellen Differentialgleichungen, wie sie hier verwendet wird, ist Teil allgemeiner Numerik-Ausbildung. Wir verweisen auf [6, 30].

Die Simulation stochastischer Differentialgleichungen wird etwa betrachtet in [10, 13, 14, 15, 17, 19, 24, 25, 26, 27, 28].

Heston war wohl der Erste, der die Fouriertransformation zur Bewertung von Optionen angewendet hat; siehe [12]. Intensivere Studien findet man in [20, 22, 23, 33]. Zur Theorie der „Numeraire“ siehe [1, 16]. Zur Darstellung der kontinuierlichen Fouriertransformation siehe etwa [4, 21, 32], zur diskreten Fouriertransformation siehe [2] und zur schnellen Fouriertransformation siehe [3, 5, 7].

Eine alternative Methode für die Bewertung von Optionen analytisch und numerisch kann mit Hilfe Mellin-Transformation vorgenommen werden; siehe etwa [29, 35].

6.7 Übungen

1.) Betrachte eine asiatische Option mit Verfallszeit $T = 1$. Der Aktienkurs $S = (S_t)_{t\in[0,T]}$ werde mit
$$dS = r(t)Sdt + \sigma S dW_t\,,\ S_0 = 100\,,$$
modelliert, der Zinssatz sei gegeben durch
$$r(t) := \frac{1}{100}(\sin(2\pi t) + t + 3)$$
und die Volatilität σ sei 0.25. Die Auszahlung zur Zeit T ist gegeben durch
$$\Phi(S) := \left(S_T - \frac{1}{T}\int_0^T S_t dt\right)^+ .$$
Approximiere mit Hilfe von Monte Carlo-Simulationen den Wert der Option durch
$$V_{N,M} := \exp\left(-\int_0^T r(t)dt\right)\frac{1}{N}\sum_{i=1}^N \left(S_{T,i} - \frac{1}{M}\sum_{j=1}^M S_{t_j,i}\right)^+ .$$
Zur Simulation benutze das Euler-Verfahren.
   (a) Kläre das Zustandekommen der obigen Optionspreisformel.
   (b) Plotte den Auszahlungspreis gegen die Anzahl N für $N = 4, 8, 16, 32$ und $M = 100$.
   (c) Ermittle den Optionspreis für $N = 4, 8, 16$ und $M = 100$.

2.) Simuliere mit dem Euler-Verfahren das so genannte **Ho-Lee**-Zinsmodell:
$$dr = \theta(t)dt + \sigma dW_t\,,\ r_0 = c \in \mathbb{R}$$
auf $[0, T]$ für ein c und eine Funktion $\theta : [0, T] \longrightarrow \mathbb{R}$.

3.) Betrachte die stochastische Differentialgleichung
$$dS = (r - q)Sdt + \sigma S dW_t$$
mit dem kanstanten Zinssatz, der Dividendenrate und der konstanten Volatilität.
   (a) Berechne den Prozess $(S_t)_{t\geq 0}$ mit Hilfe des Itô-Lemma.
   (b) Berechne den Erwartungswert $\mathbb{E}(S_T)$ für $T > 0$.

4.) Sei $x = (x_0, \ldots, x_{n-1}) \in \mathbb{R}^n$ und sei $z := \mathrm{DFT}_n(x)$. Betrachte nun $Z := \mathrm{DFT}_{2n}((x_0, 0, x_1, 0, x_3, \ldots, x_{n-1}, 0))$. Läßt sich Z aus z auslesen?

Literaturverzeichnis

[1] S. Benninga, T. Tomas Björk, T. Björk and Z. Wiener. On the use of numeraires in option pricing. *The Journal of Derivatives*, 10:43–58, 2002.

[2] E.O. Brigham. *The fast Fourier transform and its applications.* Englewood Cliffs, Prentice Hall, 1988.

[3] P. Carr and D.B. Madan. Option valuation using the fast Fourier transform. *J. Comput. Finance*, 2:61–73, 1998.

[4] K. Chandrasekharan. *Classical Fourier Transform.* Springer, Berlin, 1989.

[5] J.W. Cooley and J.W. Tuckey. An algorithm for the machine calculation of complex Fourier series. *Math. Comp.*, 19:297–301, 1965.

[6] P. Deuflhard and F. Bornemann. *Numerical Methods II.* de Gruyter, Berlin.

[7] P. Duhamel and M. Vetterli. Fast Fourier transforms: a tutorial review and a state of the art. *Signal Processing*, 19:259–299, 1990.

[8] J. Gil-Pelaez. Note on the inverse theorem. *Biometrika*, 37:481–482, 1951.

[9] M. Giles and R. Carter. Covergence analysis of Crank-Nicolson and Rannacher time-marching. *J. of Computational Finance*, 9:89–112, 2006.

[10] M. Günther and A. Jüngel. *Finanzderivate mit MATLAB.* Vieweg, 2003.

[11] J. Gurland. Inversion formulae for the distribution of ratios. *Ann. of Math. Statistics*, 19:228–237, 1948.

[12] S.L. Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies*, 6:327–343, 1993.

[13] D.J. Higham. An algorithmic introduction to numerical simulation of stochastic differential equation. *SIAM Review*, 43:525–546, 2001.

[14] D.J. Higham, X. Mao and A.M. Stuart. Strong convergence of Euler-type methods for non-linear stochastic differential equations. *SIAM J. Numer. Anal.*, 40:1041–1063, 2002.

[15] M. Hutzenthaler and A. Jentzen. Convergence of the stochastic Euler scheme for locally Lipschitz coefficients. *arXiv:0912.2609vl*, xx:1–56, 2009.

[16] M. Jeanblanc, M. Yor and M. Chesnay. *Mathematical Methods for Financial Markets.* Springer, New York, 2009.

[17] A. Jentzen, P. Kloeden and A. Neuenkirch. Pathwise approximation of stochastic differential equations on domains: higher order convergence rates without global Lipschitz coefficients. *Num. Math.*, 112:41–64, 2008.

[18] I. Karatzas and S. Shreve. *Brownian motion and stochastic calculus.* Springer, New York, 1991.

[19] P. Kloeden and E. Platen. *Numerical Solution of SDE through Computer Experiments.* Springer, Berlin, 1994.

[20] R.W. Lee. Option pricing by transform methods: extensions, unification and error control. *J. Comput. Finance*, 7:xx–xx, 2004.

[21] B. Lenze. *Einführung in die Fourier-Analysis.* Logos, Berlin, 2000.

[22] R. Lord and C. Kahl. Optimal Fourier inversion in semi-analytical option pricing. *Journal of Computational Finance*, 10:1–30, 2007.

[23] R. Lord, F. Fang, F. Bervoets and C.W. Oesterlee. A fast and accurate FFT-based method for pricing early-exercise options under Lévy processes. *SIAM Journal of Scientific Computing*, 30:1678–1705, 2008.

[24] G. Maruyama. Continuous Markov processes and stochastic equations. *Rend. Cic. Math. Palermo*, 4:48–90, 1955.

[25] G.N. Milstein. *Numerical integration of stochastic differential equations.* Kluwer, 1995.

[26] G.N. Milstein and M.V. Tretjakov. Numerical integration of stochastic differential equations with nonglobally Lipschitz coefficients. *SIAM J. Numer. Anal.*, 43:1139–1154, 2005.

[27] K.-S. Moon, A. Szepessi, R. Tempone and G. Zouaris. Convergence rates for adaptive weak approximation of stochastic differential equations. *Stochastic Analysis and Applications*, 23:511–558, 2005.

[28] N.J. Newton. Variance reduction for simulated diffusions. *SIAM J. Appl. Math.*, 54:1780–1805, 1994.

[29] R. Panini and R.P. Srivastav. Option pricing with Mellin transform. *Mathematical and Computer Modelling*, 40:43–56, 2004.

[30] A. Quarterioni, R. Sacco and F. Saleri. *Numerical Mathematics.* Springer, New York, 2000.

[31] R. Rannacher. Finite element solution of diffusion problems with irregular data. *Numerische Mathematik*, 43:309–327, 1984.

[32] R. Steinbrecher. *Bildverarbeitung in der Praxis.* Oldenburg, 1993.

[33] P. Tankov. Fourier transfotrm methods for european option pricing. *xxx*, xx:xx–xx, 2005.

[34] J. Wloka. *Partielle Differentialgleichungen.* Teubner, Stuttgart, 1982.

[35] A. Würfel. Analytische und numerische Lösung der Black-Scholes-Gleichung für europäische und amerikanische Basket-Optionen. Master's thesis, Technische Universität, Institut für Mathematik, 2007.

Teil II

Teil 2: Weiterführende Themen