3. Monte-Carlo-Methoden

In Kapitel 1 haben wir folgende Formel zur risiko-neutralen Bewertung von Optionen eingeführt:

$$V(S_0,0) = e^{-rT}\mathsf{E}_{\mathsf{Q}}\left[\Psi(S_T)|S_0\right] ,$$

wobei $\Psi(S_T)$ den Payoff bezeichnet. Im Black-Scholes-Modell ergab sich daraus der Spezialfall

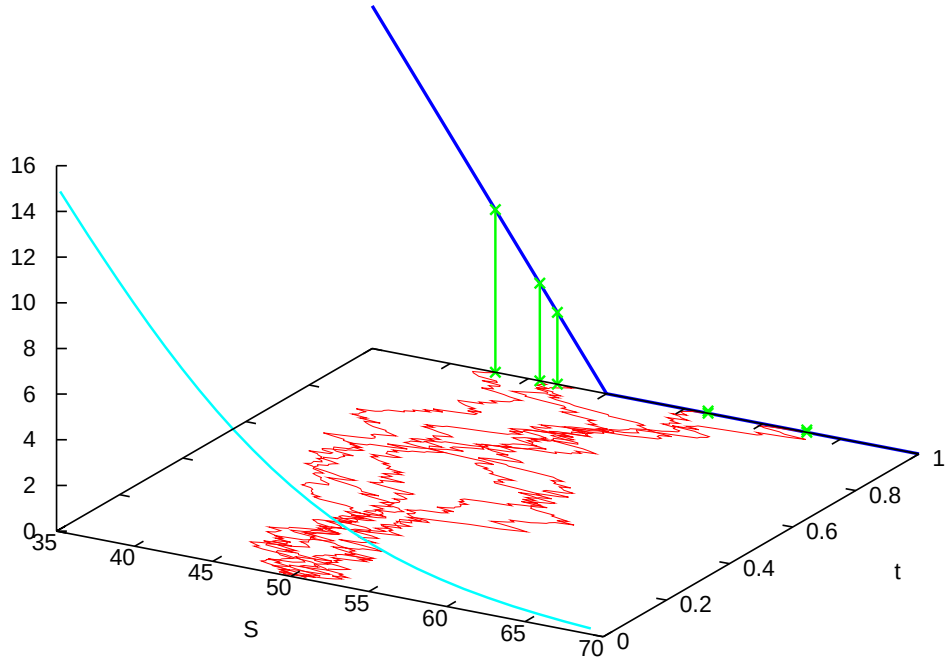
$$V(S_0,0) = e^{-rT}\int_0^\infty \Psi(S_T)\cdot f_{\mathrm{GBM}}(S_T,T;S_0,r,\sigma)\,\mathrm{d}S_T\,. \qquad \text{(Int)}$$

(Übergangsdichte f_{GBM} vgl. Abschnitt 1.5D.) Die daraus resultierende PDGl des Black-Scholes-Modells wird in Kapitel 4 diskutiert und gelöst. Solche PDGlen existieren für allgemeinere Modelle in der Regel nicht. Auf Grund dessen werden wir nun Monte-Carlo-Methoden herleiten, die in jedem Fall benutzt werden können.

Es gibt zwei Varianten, obigen Erwartungswert zu berechnen:

1) Das Integral (Int) wird mit numerischen Quadratur-Methoden berechnet.
2) Man wendet Monte-Carlo-Simulationen an. D.h. man "spielt" mit Hilfe von Zufallszahlen Börsenkurse S_t durch unter den angenommenen (risiko-neutralen) Wahrscheinlichkeitsverteilungen, die zu dem jeweiligen Modell passen. Dies ist der Hauptteil der Simulation; anschließend wird der Mittelwert der Payoff-Werte gebildet und diskontiert.

In diesem Kapitel diskutieren wir ausschließlich die zweite Variante.



Fünf simulierte Börsenkurs-Pfade mit Payoff-Werten.

Bezeichnungen (Wiederholung)

Das einfache skalare Modell mit Antrieb durch einen Wiener-Prozess ist gegeben durch

$$\mathrm{d}X_t = a(X_t,t)\,\mathrm{d}t + b(X_t,t)\,\mathrm{d}W_t\,. \qquad \text{(SDE)}$$

Wir verwenden ein diskretes Gitter

$$\ldots < t_{j-1} < t_j < t_{j+1} < \ldots\ ,$$

mit äquidistanter Gitterweite h bzw. $\Delta t = t_{j+1} - t_j$. Es sei y_j eine Approximation zu X_{t_j} mit $y_0 := X_0$.

Beispiel: Aus Kapitel 1 kennen wir die **Euler-Diskretisierung**

$$\begin{aligned} y_{j+1} &= y_j + a(y_j,t_j)\,\Delta t + b(y_j,t_j)\,\Delta W_j\,, \\ t_j &= j\Delta t\,, \\ \Delta W_j &= W_{t_{j+1}} - W_{t_j} = Z\sqrt{\Delta t} \\ &\text{mit } Z \sim \mathcal{N}(0,1)\,. \end{aligned} \qquad \text{(Euler)}$$

3.1 Approximations-Fehler

Definition

Für einen gegebenen Pfad des Wiener-Prozesses W_t nennen wir eine Lösung von X_t von (SDE) eine *starke* Lösung. Falls der Wiener-Prozess frei ist, so nennen wir X_t eine *schwache* Lösung.

Bei starken Lösungen wird die numerische Diskretisierung mit dem *gleichen* W_t berechnet, welches für (SDE) zu Grunde liegt. Dies hat den Vorteil, dass die pfadweise Differenz $X_t - y_t$ untersucht werden kann (insbesondere auf Konvergenzverhalten, mit $y_0 = X_0$).

Bezeichnung: Schreibe y_t^h für eine numerisch mit Schrittweite h berechnete Näherung y an der Stelle t.

Definition (absoluter Fehler)

Für eine starke Lösung X_t von (SDE) und eine Näherung y_t^h ist der absolute Fehler für $t = T$

$$\varepsilon(h) := \mathsf{E}\big[\ |X_T - y_T^h|\ \big]\ .$$

Dieser Wert kann bei GBM, bei der die analytische Lösung X_t bekannt ist, leicht empirisch ermittelt werden:

Setze hierzu in (SDE) $a(X_t,t) = \alpha X_t$ und $b(X_t,t) = \beta X_t$. So erhält man mit der Lösungsformel aus Abschnitt 1.5D für gegebenes W_T den Wert X_T.

Ein Schätzer für den Erwartungswert $\varepsilon(h)$ ist der Mittelwert über eine große Anzahl von Auswertungen von $|X_T - y_T^h|$. Dies liefert empirisch als Ergebnis

$$\varepsilon(h) = \mathcal{O}(h^{\frac{1}{2}})\,,$$

eine im Vergleich zum deterministischen Fall geringe Genauigkeit. (Dieses Resultat ist plausibel wegen $\Delta W = \mathcal{O}\left(\sqrt{h}\right)$, vgl. Abschnitt 1.4.)

Definition (starke Konvergenz)

y_T^h konvergiert *stark* gegen X_T mit Ordnung $\gamma > 0$, wenn

$$\varepsilon(h) = \mathsf{E}\big[\ |X_T - y_T^h|\ \big] = \mathcal{O}(h^\gamma)\,.$$

Man sagt, y_T^h konvergiert stark, wenn

$$\lim_{h\to 0} \mathsf{E}\big[\ |X_T - y_T^h|\ \big] = 0\,.$$

Beispiel

Die Euler-Diskretisierung hat unter geeigneten Voraussetzungen an a und b die Ordnung $\gamma = \frac{1}{2}$.

Beweis: [Kloeden & Platen: Numer. Solution of SDEs] u.a. S.342-344

Was ist mit den schwachen Lösungen?

In vielen Fällen der Praxis sind die jeweiligen Pfade von X_t ohne Interesse. Stattdessen interessiert uns nur ein *Moment* von X_T, d.h. entweder $\mathsf{E}[X_T]$ oder $\mathsf{Var}[X_T]$, ohne Wert auf die *einzelnen* Werte von X_T zu legen.

Definition (schwache Konvergenz)

y_T^h konvergiert *schwach* gegen X_T bzgl. einer Funktion g mit Ordnung $\beta > 0$, wenn

$$\mathsf{E}[g(X_T)] - \mathsf{E}\big[g(y_T^h)\big] = \mathcal{O}(h^\beta)\,,$$

und konvergiert schwach mit Ordnung β, wenn dies für alle Polynome g gilt.

Beispiel

Die Euler-Methode ist unter geeigneten Voraussetzungen an a und b von schwacher Konvergenz mit Ordnung $\beta = 1$.

Bedeutung von g

Wenn die Konvergenzordnung β für alle Polynome gilt, so folgt u.a. Konvergenz aller Momente. Beweis für die ersten beiden Momente:

(a) Für $g(x) := x$ gilt

$$\mathsf{E}[X_T] - \mathsf{E}[y_T^h] = \mathcal{O}(h^\beta)\,,$$

d.h. Konvergenz der Mittel.

(b) Falls die Konvergenzordnung zusätzlich für $g(x) := x^2$ gilt (mit Bezeichnungen $y := y_T^h$ und $X := X_T$)

$$\begin{aligned}\big|\mathsf{Var}[X_T] - \mathsf{Var}[y_T^h]\big| &= \big|\mathsf{E}[X^2] - \mathsf{E}[y^2] - (\mathsf{E}[X])^2 + (\mathsf{E}[y])^2\big| \\ &\le \underbrace{|\mathsf{E}[X^2] - \mathsf{E}[y^2]|}_{=\mathcal{O}(h^\beta)} + \underbrace{|\mathsf{E}[X] + \mathsf{E}[y]|}_{\le \text{const}} \cdot \underbrace{|\mathsf{E}[X] - \mathsf{E}[y]|}_{=\mathcal{O}(h^\beta)}\,,\end{aligned}$$

d.h. Konvergenz der Varianz.

Bemerkung

Starke Konvergenz impliziert schwache Konvergenz bzgl. $g(x) = x$, denn aus den Eigenschaften der Integration

$$| \mathsf{E}[X] - \mathsf{E}[Y] \,| = |\mathsf{E}[X-Y]| \le \mathsf{E}\,[\,|X-Y|\,]$$

folgt

$$\mathsf{E}\,[\,|X-Y|\,] = \mathcal{O}(h^\gamma) \quad \Longrightarrow \quad \mathsf{E}[X] - \mathsf{E}[Y] = \mathcal{O}(h^\gamma).$$

Praktischer Vorteil der schwachen Konvergenz:

Die Inkremente ΔW für die Berechnung von y^h können ersetzt werden durch andere Zufallsvariablen $\Delta\widehat{W}$ mit den gleichen Momenten.

Beispiel:

$\Delta\widehat{W} := \pm\sqrt{\Delta t}$, wobei beide Vorzeichen jeweils mit Wahrscheinlichkeit 1/2 angenommen werden. Dies ist billiger zu approximieren als die Berechnung von $Z \sim \mathcal{U}(0,1)$.

Folgerung: u.a. $\mathsf{E}(\Delta\widehat{W}) = 0$, $\mathsf{E}((\Delta\widehat{W})^2) = \Delta t$ $(\Rightarrow \mathrm{Var}(\Delta\widehat{W}) = \Delta t)$

Mit diesem $\Delta\widehat{W}$ statt ΔW ergibt sich das "vereinfachte Euler-Verfahren". Es ist schwach konvergent mit Ordnung 1.

3.2 Stochastische Taylorentwicklung

(Hilfsmittel zur Analyse von Konvergenzordnungen)

Zur Motivation betrachten wir zunächst den deterministischen autonomen Fall

$$\frac{\mathrm{d}}{\mathrm{d}t} X_t = a(X_t).$$

Die Kettenregel für beliebiges $f \in C^1(\mathbb{R})$ besagt

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t} f(X_t) &= \frac{\mathrm{d}f(X)}{\mathrm{d}X} \cdot \frac{\mathrm{d}X}{\mathrm{d}t} \\ &= a(X_t)\frac{\mathrm{d}}{\mathrm{d}X} f(X_t) =: Lf(X_t)\,. \\ \Longrightarrow f(X_t) &= f(X_{t_0}) + \int_{t_0}^{t} \underbrace{Lf(X_s)}_{=:\tilde{f}}\, \mathrm{d}s\end{aligned}$$

Diese Version wird mit

$$\tilde{f}(X_s) := Lf(X_s)$$

in sich selbst eingesetzt:

$$\begin{aligned}f(X_t) =& f(X_{t_0}) + \int_{t_0}^{t} \left\{ \tilde{f}(X_{t_0}) + \int_{t_0}^{s} L\tilde{f}(X_z)\,\mathrm{d}z \right\} \mathrm{d}s \\ =& f(X_{t_0}) + \tilde{f}(X_{t_0}) \int_{t_0}^{t} \mathrm{d}s + \int_{t_0}^{t}\int_{t_0}^{s} L\tilde{f}(X_z)\,\mathrm{d}z\,\mathrm{d}s \\ =& f(X_{t_0}) + Lf(X_{t_0})(t-t_0) + \int_{t_0}^{t}\int_{t_0}^{s} L^2 f(X_z)\,\mathrm{d}z\,\mathrm{d}s\,.\end{aligned}$$

Dies ist die Taylorentwicklung in Integralform, entwickelt bis einschließlich des linearen Terms mit Restglied in Form eines Doppelintegrals. Wenn man diesen Prozess fortsetzt, erhält man die deterministische Taylorentwicklung in Integralform.

Als nächstes betrachten wir den **stochastischen Fall**, d.h. die Itô-Taylor-Entwicklung. Literatur: [P. Kloeden & E. Platen: Numerical SDE]

Das Itô-Lemma auf $f(X)$ und die autonome SDE

$$\mathrm{d}X_t = a(X_t)\,\mathrm{d}t + b(X_t)\,\mathrm{d}W_t$$

angewendet ergibt

$$\mathrm{d}f(X_t) = \underbrace{\{a(X_t)\frac{\partial}{\partial x}f(X_t) + \frac{1}{2}\left(b(X_t)\right)^2\frac{\partial^2}{\partial x^2}f(X_t)\}}_{=:L^0 f(X_t)}\,\mathrm{d}t + \underbrace{b(X_t)\frac{\partial}{\partial x}f(X_t)}_{=:L^1 f(X_t)}\,\mathrm{d}W_t\,,$$

d.h.

$$\boxed{f(X_t) = f(X_{t_0}) + \int_{t_0}^{t} L^0 f(X_s)\,\mathrm{d}s + \int_{t_0}^{t} L^1 f(X_s)\,\mathrm{d}W_s\,.} \qquad (*)$$

Im Spezialfall $f(x) = x$ liefert dies die Ausgangs-SDE

$$X_t = X_{t_0} + \int_{t_0}^{t} a(X_s)\,\mathrm{d}s + \int_{t_0}^{t} b(X_s)\,\mathrm{d}W_s\,.$$

Wir wenden nun $(*)$ an für geeignete $\tilde{f}$, zunächst mit $\tilde{f} := a$ und $\tilde{f} := b$, und erhalten

$$\begin{aligned} X_t = X_{t_0} &+ \int_{t_0}^{t}\left\{a(X_{t_0}) + \int_{t_0}^{s} L^0 a(X_z)\,\mathrm{d}z + \int_{t_0}^{s} L^1 a(X_z)\,\mathrm{d}W_z\right\}\mathrm{d}s \\ &+ \int_{t_0}^{t}\left\{b(X_{t_0}) + \int_{t_0}^{s} L^0 b(X_z)\,\mathrm{d}z + \int_{t_0}^{s} L^1 b(X_z)\,\mathrm{d}W_z\right\}\mathrm{d}W_s\,, \end{aligned}$$

mit

$$\begin{aligned} L^0 a &= aa' + \frac{1}{2}b^2 a'' & L^0 b &= ab' + \frac{1}{2}b^2 b'' \\ L^1 a &= ba' & L^1 b &= bb'\,. \end{aligned}$$

Dies schreiben wir

$$X_t = X_{t_0} + a(X_{t_0})\int_{t_0}^{t}\mathrm{d}s + b(X_{t_0})\int_{t_0}^{t}\mathrm{d}W_s + R\,,$$

mit

$$\begin{aligned} R = &\int_{t_0}^{t}\int_{t_0}^{s} L^0 a(X_z)\,\mathrm{d}z\,\mathrm{d}s + \int_{t_0}^{t}\int_{t_0}^{s} L^1 a(X_z)\,\mathrm{d}W_z\,\mathrm{d}s \\ &+ \int_{t_0}^{t}\int_{t_0}^{s} L^0 b(X_z)\,\mathrm{d}z\,\mathrm{d}W_s + \int_{t_0}^{t}\int_{t_0}^{s} L^1 b(X_z)\,\mathrm{d}W_z\,\mathrm{d}W_s\,. \end{aligned}$$

In analoger Weise können die Integranden der Doppelintegrale in R durch Anwendung von $(*)$ mit geeignetem $\tilde{f}$ ersetzt werden. Dabei treten die Doppelintegrale

$$\underbrace{\int_{t_0}^{t}\int_{t_0}^{s} \mathrm{d}z\,\mathrm{d}s}_{=:\mathcal{I}(0,0)=\frac{1}{2}(\Delta t)^2}\,, \quad \underbrace{\int_{t_0}^{t}\int_{t_0}^{s} \mathrm{d}W_z\,\mathrm{d}s}_{=:\mathcal{I}(1,0)}, \quad \underbrace{\int_{t_0}^{t}\int_{t_0}^{s} \mathrm{d}z\,\mathrm{d}W_s}_{=:\mathcal{I}(0,1)}\,, \quad \underbrace{\int_{t_0}^{t}\int_{t_0}^{s} \mathrm{d}W_z\,\mathrm{d}W_s}_{=:\mathcal{I}(1,1)}$$

auf. Mit einer Plausibilitätsbetrachtung (ersetze $W_t - W_{t_0}$ durch $\sqrt{t-t_0}$) erwarte, dass $I(1,1)$ das Integral von niedrigster Ordnung ist: $O(\Delta t)$. Wir beginnen mit dem Integral niedrigster Ordnung, also demjenigen mit dem Integranden $\tilde{f} := L^1 b(X)$. Aus $(*)$ folgt dann

$$\int_{t_0}^{t}\int_{t_0}^{s} L^1 b(X_z)\,\mathrm{d}W_z\,\mathrm{d}W_s = L^1 b(X_{t_0}) \int_{t_0}^{t}\int_{t_0}^{s} \mathrm{d}W_z\,\mathrm{d}W_s + \text{zwei Dreifachintegrale}$$

Damit gilt

$$R = \text{ drei Doppelintegrale (s.o.) } + \underbrace{b(X_{t_0})b'(X_{t_0})}_{=L^1 b(X_{t_0})}\mathcal{I}(1,1) \quad + \text{ zwei Dreifachintegrale}$$

Berechnung des Doppelintegrales $\mathcal{I}(1,1)$:

Es sei $g(x) = x^2$ und $X_t = W_t$, d.h. SDE mit $a = 0$ und $b = 1$. Mit dem Itô-Lemma erhält man

$$\mathrm{d}(W_t^2) = \frac{1}{2}2\;\mathrm{d}t + 2W_t\;\mathrm{d}W_t = \mathrm{d}t + 2W_t\;\mathrm{d}W_t,$$

und daraus wiederum

$$\int_0^t\int_0^s \mathrm{d}W_z\mathrm{d}W_s = \int_0^t W_s\mathrm{d}W_s = \frac{1}{2}W_t^2 - \frac{1}{2}t\,.$$

Allgemein gilt

$$\int_{t_0}^{t}\int_{t_0}^{s} \mathrm{d}W_z\,\mathrm{d}W_s = \frac{1}{2}(\Delta W_t)^2 - \frac{1}{2}\Delta t\,.$$

Dies bestätigt die erwartete Ordnung.

Für die allgemeine Herleitung der stochastischen Taylorentwicklung setze die begonnene systematische Definition der nach obigem Prozess entstehenden Mehrfachintegrale fort, wie zum Beispiel $\mathcal{I}(0,0,0),\ldots$. Dabei steht "0" steht für eine deterministische und eine "1" für eine stochastische Integration.

Anwendung auf die Numerik

Durch Hinzufügen weiterer führender Terme der stochastischen Taylorentwicklung erhält man Integrationsverfahren höherer Ordnung.

Beispiel (Milstein-Verfahren)

$$y_{j+1} = y_j + a\,\Delta t + b\,\Delta W_j + \frac{1}{2}bb'\left\{(\Delta W_j)^2 - \Delta t\right\}$$

Die ersten Summanden entsprechen der Euler-Methode, und der letzte ist ein Korrekturterm, durch dessen Hinzufügen ein Verfahren der (starken) Ordnung 1 entsteht. (Nachprüfung z.B. empirisch) Die schwache Ordnung ist ebenfalls 1.

3.3 Monte-Carlo-Methoden bei europäischen Optionen

Wir wollen nun den Wert

$$V(S_0, 0) = e^{-rT} \mathsf{E}_{\mathsf{Q}} \left[\Psi(S_T) \,|\, S(0) = S_0 \right]$$

einer europäischen Option berechnen, wobei Ψ den Payoff bezeichnet.

A. Grundprinzip

Dieses Integral kann mit Monte-Carlo-Methoden approximiert werden. Dazu muss zunächst entschieden werden, welches Modell zugrunde gelegt wird (z.B. Heston- oder Black-Scholes-Modell). Wir betrachten hier insbesondere das klassische Black-Scholes-Modell

$$\mathrm{d}S_t = S_t \left(\mu \, \mathrm{d}t + \sigma \, \mathrm{d}W_t \right) .$$

Zur Realisierung von Q muss $\mu = r$ gesetzt werden. Der Rest erfolgt analog zur **Monte-Carlo-Quadratur** (siehe Vorlesung Numerik II):

Man ersetzt ein zu berechnendes bestimmtes Integral $\int_0^1 f(x) \, \mathrm{d}x = \int_0^1 f(x) \, 1 \, \mathrm{d}x$ durch die Summe

$$\frac{1}{N} \sum_{i=1}^{N} f(x_i) \, ,$$

wobei x_i zufällig gewählte gleichverteilte Punkte im Definitionsbereich D sind. (hier $D = [0, 1]$)

Nach dem Gesetz der großen Zahlen gilt

$$\frac{1}{N} \sum_{i=1}^{N} f(x_i) \longrightarrow \mathsf{E}(f) \quad \text{für } N \to \infty \, .$$

Allgemeiner:

$$\int_D f(x) \mathrm{d}x \approx \frac{\mathrm{Vol}(D)}{N} \sum_{i=1}^{N} f(x_i) \, .$$

Der **Fehler** ist von probabilistischer Natur (vgl. Zentraler Grenzwertsatz), d.h. es gelten Aussagen wie:

> Mit 95% Wahrscheinlichkeit liegt der wahre Wert des Integrals in dem "Konfidenzintervall", welches durch die halbe Breite $a\sigma/\sqrt{N}$ gegeben ist. Dabei ist σ die Standardabweichung, und $a = 1.96$ (bei 95%).

Frage: Welche Struktur hat $f(x)$ bei der Bewertung von Optionen?

Verteilung: Die Dichte ist hier f_{GBM}, entsprechend müssen die x_i lognormal-verteilt sein.

Algorithmus: Monte-Carlo-Verfahren bei europäischen Optionen

Führe N Simulationen des Aktienkurses entsprechend dem risikoneutralen Maß Q aus, startend in S_0, und erhalte so $x_i := (S_T)_i$ für $i = 1, ..., N$. Werte jeweils den Payoff $\Psi(S_T)$ aus:

$$f(x_i) = \Psi\big((S_T)_i\big) \, ,$$

mit anschließender Mittelwertbildung und dann Diskontierung mit Faktor e^{-rT}.

Beispiele für den Payoff Ψ

1.) *Binary-* oder *Digital*-Option, hier Binär-Call:

$$\Psi(S_T) = \mathbf{1}_{S_T>K} = \begin{cases} 1 & \text{falls } S_T > K \\ 0 & \text{sonst} \end{cases}$$

2.) *Barrier*-Option mit Barriere B, z.B. die *down-and-out* Call Option mit

$$\Psi(S) = \begin{cases} 0 & \text{falls } S_t \le B \text{ für ein } 0 \le t \le T \\ (S_T - K)^+ & \text{sonst} \end{cases}$$

Bei dieser pfadabhängigen exotischen Option ist der gesamte Kursverlauf auf S_t $0 \le t \le T$ von Interesse. (sinnvoll nur für $S_0 > B$; Illustration dreidimensional im (S,t,V)-Raum unter Berücksichtigung der Randbedingung entlang $S = B$)

3.) *two-asset cash-or-nothing* Put: Die Auszahlung ist 1 wenn die Ungleichungen $S_1(T) < K_1$ und $S_2(T) < K$ gelten, wobei $S_1(t), S_2(t)$ die Preise der beiden Assets sind.

Hinweis: Viele analytische Lösungsformeln finden sich in [E.G. Haug: Option Pricing Formulas].

Ausführungen der Monte-Carlo-Methode

Beim Black-Scholes-Modell kann die analytische Formel für S_t

$$S_t = S_0 \exp\left\{(r - \tfrac{1}{2}\sigma^2)t + \sigma W_t\right\}$$

verwendet werden. Bei nicht-pfadabhängigen Optionen muss dann nur für $t = T$ für jeden generierten Pfad $(S_t)_i$ der Zufallsgenerator einmal benutzt werden, um W_T und damit S_T zu ermitteln. Alternativ, wichtig bei allgemeineren Modellen, bei denen keine analytische Formel existiert, muss (z.B. mit dem Euler-Verfahren) numerisch integriert werden. Das Monte-Carlo-Verfahren besitzt dann außer der äußeren Schleife ($i = 1, ..., N$) noch eine innere Schleife, z.B. $1 \le j \le m$, wobei $\Delta t = \frac{T}{m}$ die Schrittweite des Integrators ist. Wenn GBM angenommen wird, kann bei pfadabhängigen Optionen auch lokal für jedes t_j die analytische Formel verwendet werden:

$$S_{t_{j+1}} = S_{t_j} \exp\left\{(r - \tfrac{1}{2}\sigma^2)\,\Delta t + \sigma\,\Delta W\right\}$$

mit $\Delta W = \sqrt{\Delta t}\,Z$, $Z \sim \mathcal{N}(0,1)$.

B. Genauigkeit

a) Es sei

$$\hat{\mu} := \frac{1}{N}\sum_{i=1}^{N} f(x_i)$$

$$\hat{s}^2 := \frac{1}{N-1}\sum_{i=1}^{N} \left(f(x_i) - \hat{\mu}\right)^2 .$$

Die Näherung $\hat{\mu}$ verhält sich nach dem ZGWS wie $\mathcal{N}(\mu, \sigma^2)$,

$$\mathsf{P}(\hat{\mu} - \mu \le a\frac{\sigma}{\sqrt{N}}) = F(a)$$

mit Verteilungsfunktion F. In der Praxis wird σ^2 durch die Näherung $\hat{s}^2$ ersetzt. Der Fehler verhält sich also wie $\frac{\hat{s}}{\sqrt{N}}$. Um diesen Fehler verringern zu können, muss entweder

der Zähler kleiner werden (*Varianzreduktion*) oder der Nenner größer, d.h. mehr Simulationen durchgeführt werden. Die zweite Möglichkeit ist sehr aufwändig, denn um z.B. eine weitere korrekte Dezimalstelle zu gewinnen, müsste der Fehler um den Faktor $\frac{1}{10}$ reduziert werden, woraus eine *Aufwandssteigerung um den Faktor 100* folgt! Ein wichtiger Vorteil von Monte-Carlo-Methoden ist, dass der Fehler im Wesentlichen dimensionsunabhängig ist.

b) Wenn die Berechnung von $f(x_i)$ erwartungstreu ist, ist der obige Fehler der einzige. Ansonsten kommt ein weiterer Fehler hinzu, nämlich die *Verzerrung* (*bias*).

Es sei $\hat{x}$ ein Schätzer für x, so ist die Verzerrung definiert als

$$\text{bias}(\hat{x}) := \mathsf{E}[\hat{x}] - x.$$

Beispiele

1.) Bei einer *Lookback*-Option gilt

$$x := \mathsf{E}\left[\max_{0\le t\le T} S_t\right] .$$

Eine Approximation hierzu ist

$$\hat{x} := \max_{0\le j\le m} S_{t_j} .$$

$\hat{x}$ wird x fast sicher unterschätzen, d.h. $\mathsf{E}[\hat{x}] < x$ fast sicher. $\hat{x}$ liefert also nur verzerrte Ergebnisse, hier gilt $\text{bias}(\hat{x}) \neq 0$.

2.) Im Vergleich zur analytischen Lösung liefert auch das Euler-Verfahren verzerrte Ergebnisse. Bei GBM ist die Verwendung von

$$S_{t_{j+1}} = S_{t_j} \exp\left\{(r - \tfrac{1}{2}\sigma^2)\,\Delta t + \sigma\,\Delta W\right\}$$

unverzerrt, während der Eulerschritt

$$S_{t_{j+1}} = S_{t_j}(1 + r\,\Delta t + \sigma\,\Delta W)$$

verzerrt (*biased*) ist. (Übung)

Es stellt sich also die Frage, wo man mehr Aufwand betreiben sollte: entweder mit

- Varianzreduktion, oder
- eine größere Anzahl N an Pfaden simulieren, oder
- bias verkleinern, d.h. m vergrößern,

oder mit mehreren dieser Maßnahmen.

Bemerkung (mean square error)

Das folgende Fehlermaß beschreibt den Gesamtfehler

$$\text{MSE}(\hat{x}) := \text{mean square error} := \mathsf{E}\left[(x - \hat{x})^2\right] .$$

Eine einfache Rechnung zeigt:

$$\begin{aligned}\text{MSE}(\hat{x}) &= \left(\mathsf{E}[\hat{x}] - x\right)^2 + \mathsf{E}\left[(\hat{x} - \mathsf{E}(\hat{x}))^2\right] \\ &= \left(\text{bias}(\hat{x})\right)^2 + \mathsf{Var}(\hat{x}) .\end{aligned}$$

C. Methoden der Varianzreduktion

Die einfachste (aber vielleicht am wenigsten wirksame) der vielen möglichen Methoden der Varianzreduktion ist die Methode der Antithetischen Variablen (*antithetic variates*). Diese basiert auf folgender Idee: Einerseits werden "normal" mit Zufallszahlen $Z_1, Z_2, \ldots$ Pfade S_t berechnet mit einer MC-Näherung, die wir hier mit $\hat{V}$ bezeichnen. Mit maximal dem doppelten Aufwand verwenden wir jeweils parallel die Zahlen $-Z_1, -Z_2, \ldots$, welche auch $\sim \mathcal{N}(0,1)$ sind, um jeweils gespiegelte "Zwillings-Pfade" S_t^- zu berechnen, aus welchen der Payoff $\Psi(S_T^-)$ ermittelt wird. Wir erhalten so einen *zweiten* Monte-Carlo-Wert V^-. Für den Mittelwert

$$V_{\rm AV} := \frac{1}{2}(\hat{V} + V^-)$$

gilt stets $\mathsf{Var}(\hat{V}) = \mathsf{Var}(V^-)$, und meist $\mathsf{Cov}(\hat{V}, V^-) < 0$. Für den Fall $\mathsf{Cov}(\hat{V}, V^-) < 0$ folgt mit einer elementaren Rechnung

$$\mathsf{Var}(V_{\rm AV}) = \frac{1}{4}(\mathsf{Var}\hat{V} + \mathsf{Var}V^- + 2\mathsf{Cov}(\hat{V}, V^-)) < \frac{1}{2}\mathsf{Var}(\hat{V})\,.$$

(Die nächst-komplexere Methode der Varianzreduktion ist die Methode *control variates*, vgl. Literatur.)

Hinweis: Für den eindimensionalen Fall von Vanilla-Optionen wurde MC nicht entwickelt. Die Möglichkeiten von MC entfalten sich bei exotischen Optionen, vor allem im höherdimensionalen Fall.

3.4 Monte-Carlo-Methoden bei amerikanischen Optionen

Vorbemerkung zur Filtration $\mathcal{F}_t$: Ein stochastischer Prozess X_t heißt $\mathcal{F}_t$-adaptiert, wenn X_t $\mathcal{F}_t$-messbar ist für alle t. Die natürliche Filtration $\mathcal{F}_t^X$ ist die kleinste Sigma-Algebra über $\{X_s \,|\, 0 \le s \le t\}$, vereinigt mit den P-Nullmengen. X_t ist bezüglich $\mathcal{F}_t^X$ adaptiert. Filtrationen repräsentieren die Informationsmenge zum Zeitpunkt t.

A. Stoppzeiten

Das Ausüben einer amerikanischen Option geschieht an einer "Stoppzeit" τ, welche vom zugrundeliegenden Prozess abhängt. Das Ausüben ist eine Entscheidung, die nur auf Informationen beruht, die zum Entscheidungszeitpunkt vorliegen. Deswegen darf eine zu definierende "Stoppzeit" nicht vorgreifend sein *(non-anticipating)*: Für jedes t brauchen wir Gewissheit, ob die Entscheidung gefallen ist, d.h. ob $\tau \le t$ oder $\tau > t$. Fordere also für die Menge $\{\tau \le t\}$ aller Entscheidungen bis t

$$\{\tau \le t\} \in \mathcal{F}_t\ .$$

Das ist die $\mathcal{F}_t$-Messbarkeit von τ.

Definition (Stoppzeit)

Eine Stoppzeit τ bezüglich einer Filtration $\mathcal{F}_t$ ist eine Zufallsvariable, welche für alle t $\mathcal{F}_t$-messbar ist.

Es sei $\Psi(S_t)$ ein Payoff, z.B. $\Psi(S_t) = (K - S_t)^+$, so gilt folgendes Resultat von Bensoussan (1984):

$$V(S,t) = \sup_{t \le \tau \le T} \mathsf{E}_{\mathsf{Q}} \left[e^{-(\tau - t)r} \, \Psi(S_\tau) \mid \mathcal{F}_t \right] \qquad (*)$$

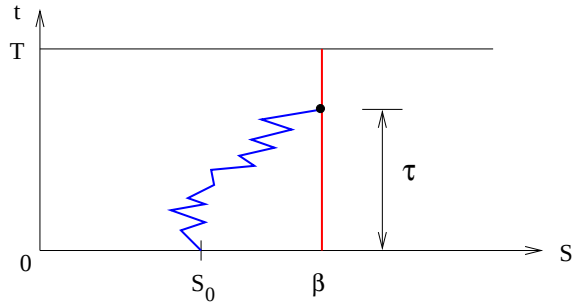
wobei τ eine Stoppzeit ist bezüglich einer natürlichen Filtration $\mathcal{F}_t$ zu S_t.

Beispiele für Stoppzeiten

1) Es sei

$$\tau := \inf\{t > 0 \mid S_t \ge \beta\}$$

für gegebenes $\beta > 0$. (engl. "*hitting time*") Falls kein solches t existiert, setze $\tau := \infty$. Für den formalen Beweis, dass dieses τ eine Stoppzeit ist, siehe z.B. [Hunt & Kennedy (2000)].



2) Definiere t^* als denjenigen Zeitpunkt, an dem $\max_{0 \le t \le T} S_t$ erreicht wird. Dies ist keine Stoppzeit! Denn für ein beliebiges t kann nicht entschieden werden, ob $t^* \le t$ oder $t^* > t$.

3) Es sei

$$\tau := \min\{t \le T \mid (t, S_t) \in \textit{stopping region}\} \quad \text{(vgl. Abschnitt 4.5 im folgenden Kapitel).}$$

B. Parametrische Methoden

In $(*)$ wird das Supremum über *alle* Stoppzeiten gebildet. Wir konstruieren nun eine *spezielle* Stoppzeit. Ähnlich wie in Beispiel 1 oder 3 definieren wir eine Kurve im (S,t)-Halbstreifen, welche eine Näherung zur early-exercise Kurve darstellen soll. Damit ist eine spezielle Stopp-Strategie $\tilde{\tau}$ definiert durch das Überschreiten dieser Kurve. Wenn β einen Vektor von Parametern repräsentiert, der die Kurve definiert, dann hängt $\tilde{\tau}$ von β ab. Die spezielle β-abhängige Stopp-Strategie $\tilde{\tau}$ führt zu einer unteren Schranke

$$V^{(\beta)}(S,t) := \mathsf{E}_{\mathsf{Q}} \left[e^{-(\tilde{\tau} - t)r} \Psi(S_{\tilde{\tau}}) \,|\, \mathcal{F}_t \right] \le V(S,t) \,.$$

Anwendung

Offensichtlich kann $V(S,t)$ über geeignete Kurven näherungsweise als

$$\sup_\beta V^{(\beta)}$$

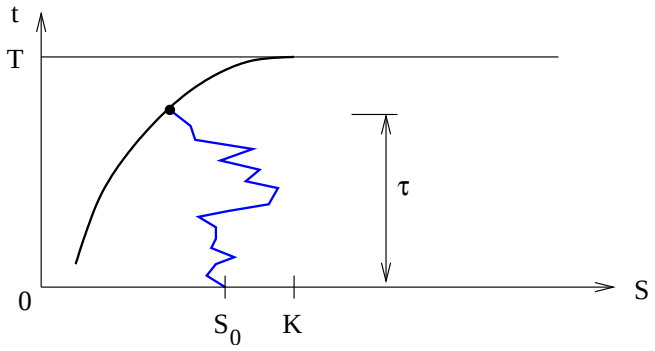
charakterisiert werden, über eine optimale Stopp-Strategie. Diese Idee führt auf einen brauchbaren Ansatz:

> Man konstruiere eine Kurve in Abhängigkeit von einem Parametervektor β derart, dass sie die *early-exercise*-Kurve approximiert. Stopp-Strategie: Stoppe wenn der Pfad S_t die durch β definierte Kurve überschreitet. Dann rechnet man für gegebenes β den Wert $V^{(\beta)}$ mit "gewöhnlichen" Monte-Carlo-Methoden aus. Anschließend maximiert

man den Wert $V^{(\beta)}$ durch Wiederholungen des Verfahrens mit "besseren" Parameter-Werten β.

Beispiel mit $\beta \in \mathbb{R}^1$:

Betrachte eine Parabel mit Scheitel in $(S,t) = (K,T)$ mit nur einem Parameter β als Näherung der early-exercise-Kurve eines Puts (vgl. Figur). Man berechne Pfade (z.B. 10000 mal) bis die Parabel erreicht wird ($\rightarrow \tilde{\tau}$) wie in Beispiel 1, woraus man einen Wert $V^{(\beta)}$ erhält. Jeder einzelne Wert $V^{(\beta)}$ wird also mit einem Aufwand berechnet, der etwa dem MC-Verfahren bei einer europäischen Option entspricht. Anschließend Wiederholung für "bessere" β.



Literatur: [Glasserman: Monte Carlo Methods in Financial Engineering (2004)]

C. Regressionsmethoden

Definition (Bermuda-Option)

Eine Bermuda-Option ist eine Option, die nur an einer endlichen Anzahl M von diskreten Zeitpunkten t_i ausgeübt werden kann.

$$\Delta t := \frac{T}{M}, \quad t_i := i\,\Delta t \quad (i = 0, \ldots, M)$$

Wir bezeichnen den Wert einer solchen Bermuda-Option mit $V^{\mathrm{Be(M)}}$. Es gilt auf Grund der zusätzlichen Anzahl an Ausübungsmöglichkeiten

$$V^{\mathrm{eu}} \le V^{\mathrm{Be(M)}} \le V^{\mathrm{am}}\,.$$

Außerdem kann man zeigen, dass

$$\lim_{M\to\infty} V^{\mathrm{Be(M)}} = V^{\mathrm{am}}\,.$$

Für geeignetes M wird $V^{\mathrm{Be(M)}}$ als Näherung von V^{am} verwendet. Wegen des grundsätzlich großen Aufwandes von Monte-Carlo-Verfahren bei amerikanischen Optionen beschränkt man sich typischerweise auf Bermuda-Optionen.

Wiederholung (Binomial-Methode, vgl. Abschnitt 1.3)

Der Wert einer amerikanischen Option wird dort rückwärts rekursiv berechnet: Für die europäische Option ergibt sich

$$\begin{aligned} V_{j,i} = V(S_{j,i}, t_i) &= e^{-r\Delta t}\,(pV_{j+1,i+1} + (1-p)V_{j,i+1}) \\ &= e^{-r\Delta t}\,\mathsf{E}_{\mathsf{Q}}\left[V(S_{t_{i+1}}, t_{i+1}) \mid S_{t_i} = S_{j,i}\right] \\ &=: V_{j,i}^{\mathrm{cont}} = \text{ Haltewert} \end{aligned}$$

und

$$V_{j,i}^{\rm am} = \max\left\{\Psi(S_{j,i}), V_{j,i}^{\rm cont}\right\} .$$

Denn an t_i entscheidet der Halter der Option, welche der beiden Möglichkeiten {Ausüben, Halten} die bessere ist.

Bei einer Bermuda-Option definieren wir analog den **Haltewert** an t_i:

$$C_i(x) := e^{-r\Delta t}\, \mathsf{E}_{\mathsf{Q}}\left[V(S_{t_{i+1}}, t_{i+1}) \mid S_{t_i} = x\right] .$$

Diese Funktion mit Zuordnung $(x, C_i(x))$ muss approximiert werden.

Allgemeine Rekursion

Setze $V_M(x) = \Psi(x)$.
Für $i = M-1, \ldots, 1$ berechne $C_i(x)$ für $x > 0$ und

$$V_i(x) := V(x, t_i) = \max\left\{\Psi(x),\, C_i(x)\right\} ,$$

$V_0 := V(S_{t_0}, t_0) = C_0(S_0)$.

Dieses Prinzip heißt "Prinzip der Dynamischen Programmierung".

Um die $C_i(x)$ zu berechnen, zieht man Informationen aus durch Simulation erzeugten Pfaden, und approximiert $C_i(x)$ durch eine Regressionskurve $\hat{C}_i(x)$.

Regression (Grundprinzip)

(a) Simuliere N Pfade $S_1(t), ..., S_N(t)$. Berechne und speichere die Werte

$$S_{i,k} := S_k(t_i) , \quad i = 1, ..., M,\ k = 1, ..., N .$$

(b) Setze für $i = M$: $V_{M,k} = \Psi(S_{M,k})$ für alle k.

(c) Für $i = M-1, ..., 1$:
[Die N Paare $(S_{i,k}\,,\, e^{-r\Delta t} V_{i+1,k})$ bilden den Datensatz zur Approximation von $(x, C_i(x))$.]
Approximiere $C_i(x)$ mit geeigneten Basisfunktionen $\phi_0, ..., \phi_L$ (z.B. Monome)

$$C_i(x) \approx \sum_{l=0}^{L} a_l \phi_l(x) =: \hat{C}_i(x)$$

durch die Methode kleinster Quadrate (Numerik I: "least squares"). Bestimme $a_0, \ldots, a_L$ und damit $\hat{C}_i$ über die N Punkte

$$(S_{i,k}\,,\, e^{-r\Delta t} V_{i+1,k}) , \quad k = 1, ..., N$$

und setze

$$V_{i,k} := \max\left\{\Psi(S_{i,k}),\, \hat{C}_i(S_{i,k})\right\} .$$

(d) Setze

$$V_0 := e^{-r\Delta t}\, \frac{1}{N}(V_{1,1} + ... + V_{1,N}) .$$

Aufwendig sind die Schritte (a) und (c). Zu Schritt (d): Der eigentliche Algorithmus von (c) lässt sich für $i = 0$ nicht durchführen, weil $S_{0,k} = S_0$ für alle k. Deswegen die Mittelwertbildung in (d). Konvergenz des Algorithmus wurde bewiesen.

Auf dem Grundprinzip des obigen Regressions-Algorithmus baut der effizientere Algorithmus von Longstaff & Schwartz auf.

Der **Aufwand** bei Amerikanischen Optionen hängt von der Dimension ab, also von der Anzahl der zugrundeliegenden Assets. (Warum?) Wenn die Gesamtrechenzeit für die Bewertung einer amerikanischen Option beschränkt ist, dann beinflusst diese Dimensionsabhängigkeit des Aufwands auch den erreichbaren Fehler! Insofern ist der Fehler bei Monte-Carlo-Methoden *doch* dimensionsabhängig.

Bei Longstaff & Schwartz wird der Algorithmus wie folgt modifiziert:

Ein dynamisches Programmierprinzip für die optimalen Stoppzeiten wird eingearbeitet. Für jeden Pfad wird eine eigene optimale Stoppzeit τ_k vorgesehen für $k = 1, ..., N$ (es genügt wegen $\tau_k = k\Delta t$ die Speicherung des Index k).

Algorithmus

Initialisierung: $\tau_k := M$ für alle k.

Für jedes $i = M-1, ..., 1$:

Schleife über alle Pfade $k = 1, ..., N$:

Falls $\Psi(S_{ik}) \geq \hat{C}_i(S_{ik})$ dann setze $\tau_k := i$.

Andernfalls lasse τ_k unverändert.

Dieser Algorithmus hat den Vorteil, dass über mehrere Zeitschichten hinweg gearbeitet werden kann. Dank einer Modifikation von Christian Jonen [Intern.J.Computer Math. **86** (2009)] ist dies ein effizientes Verfahren.