Course Notes on Computational Finance

3

Monte Carlo Methods

3. Monte Carlo Methods

Chapter 1 introduced the formula of risk-neutral valuation of European options,

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

where $\Psi(S_T)$ denotes the payoff. In the Black–Scholes model, specifically, this is

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

(For the transition density f_{GBM} see Section 1.5D.)

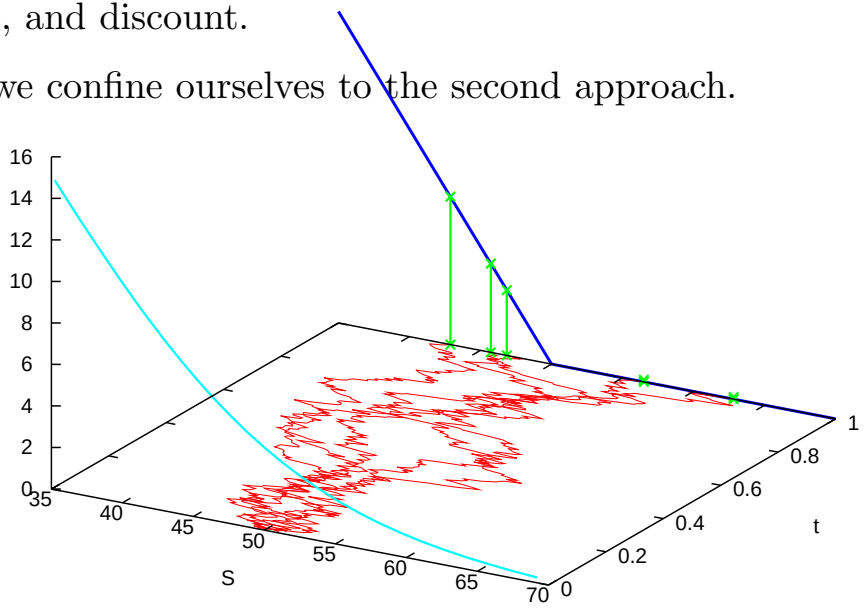
The resulting PDE of the Black–Scholes model will be the topic of Chapter 4. For general models, such PDEs are not always known, or not easy to solve. In such cases we need Monte Carlo methods, which can be applied in all cases.

There are two approaches to calculate the above integral:

1) The integral (Int) is approximated using numerical quadrature.

2) One applies Monte Carlo simulation. That is, one draws random numbers that match the underlying risk-neutral probability, and calculates many paths of asset prices S_t. This is the bulk of the work. To complete, compute the mean of the payoff values, and discount.

In this chapter we confine ourselves to the second approach.



Five simulated asset paths with payoff.

Notations (from Chapter 1)

A scalar SDE driven by a Wiener process is described by

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

We discretize time t with a grid

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

with equidistant step h or $\Delta t = t_{j+1} - t_j$. Let y_j denote an approximation of X_{t_j}, where $y_0 := X_0$.

Example: **Euler discretization**

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

3.1 Approximation Error

Definition

For a given path of the Wiener processes W_t we call a solution X_t of (SDE) a **strong solution**. In case the Wiener process is free, X_t or (X_t, W_T) is called **weak solution**.

For strong solutions the numerical discretization is based on the same W_t as the SDE. This enables to investigate the pathwise difference $X_t - y_t$ for convergence behavior for $h \to 0$.

Notation: We write y_t^h for a numerically (with step length h) calculated approximation y at t, in particular, for $t = T$.

Definition (absolute error)

For a strong solution X_t of (SDE) and an approximation y_t^h the absolute error at $t = T$ is defined as

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

For a GBM, where the analytic solution X_t is known, $\varepsilon(h)$ can be obtained easily:

Set in (SDE) $a(X_t,t) = \alpha X_t$ and $b(X_t,t) = \beta X_t$. Then the solution (see Section 1.5D) for given W_T is

$$X_T = X_0 \exp\left[\left(\alpha - \frac{\beta^2}{2}\right)T + \beta W_T\right] .$$

The expectation $\varepsilon(h)$ can be estimated as mean of a large number of evaluations of $|X_T - y_T^h|$. For Euler's method this empirical investigation reveals the error behavior

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

which is a low accuracy compared to the deterministic case $O(h)$. The result is plausible because ΔW is of the order $O(\sqrt{h})$ (in probability), compare Section 1.4.

Definition (strong convergence)

y_T^h converges strongly to X_T with order $\gamma > 0$, if

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

y_T^h converges strongly if

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

Example

When a and b satisfy global Lipschitz conditions and bounded growth conditions, then the Euler discretization converges strongly with order $\gamma = \frac{1}{2}$.

Note that for several important SDE models (such as CIR, Heston) the global Lipschitz conditions do not hold. Then modifications of the standard Euler may be necessary, also to guarantee $S_t \geq 0$.

How about weak solutions?

In many practical situations the individual paths of X_t are not of interest. Instead, the focus may be on *moments* of X_T. In particular, we would like to know $\mathsf{E}[X_T]$ or $\mathsf{Var}[X_T]$, rather than samples of X_T. For options, the interest is on $\mathsf{E}[\Psi(X_T)]$.

Definition (weak convergence)

y_T^h converges weakly to X_T with respect to a function g with order $\beta > 0$, if

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

and converges weakly with order β, if this holds for all polynomials g.

Example

When a and b are four times continuously differentiable, the Euler method is weakly convergent with order $\beta = 1$.

Importance of g

In case the convergence order β holds for all polynomials, the convergence of all moments follows.

Proof (for the first two moments):

(a) For $g(x) := x$

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

holds, viz, convergence of the mean.

(b) If in addition the convergence order holds for $g(x) := x^2$, then (writing $y := y_T^h$ and $X := X_T$)

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

i.e. convergence of the variance.

Remark

Strong convergence implies weak convergence with respect to $g(x) = x$.

Because the properties of integration

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

lead to

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

Practical advantage of weak convergence:

The increments ΔW needed to calculate y^h can be replaced by other random variables $\widehat{\Delta W}$ with matching first moments. The weak-convergence order survives.

Example

$\widehat{\Delta W} := \pm\sqrt{\Delta t}$, where both signs occur with probability 1/2.
(cheaper to approximate than $Z \sim \mathcal{N}(0,1)$)

This implies $\mathsf{E}(\widehat{\Delta W}) = 0$ and $\mathsf{E}((\widehat{\Delta W})^2) = \Delta t$ $(\Rightarrow \mathrm{Var}(\widehat{\Delta W}) = \Delta t)$.

When $\widehat{\Delta W}$ replaces ΔW one obtains the "simplified Euler method," which is weakly convergent with order 1.

3.2 Constructing Integrators for SDEs

The derivation of integrators for SDEs can be based on the stochastic Taylor expansion.

A. Stochastic Taylor expansion

(follows [P. Kloeden & E. Platen: Numerical Solution of SDEs])
For motivation we first consider the **deterministic** autonomous case

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

The chain rule for $f \in C^1(\mathbb{R})$ tells

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

Substitute this formula for

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

into itself gives

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

This is the Taylor expansion with remainder term in integral form, here expanded until the linear term; the remainder is a double integral. This process can be continued, and the deterministic Taylor expansion with remainder in integral form results. (All needed derivatives may exist.)

stochastic case (Itô-Taylor expansion)

Applying the Itô lemma on $f(X)$ and the autonomous SDE

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

leads to

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

or

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

Specifically for $f(x) = x$, the equation $(*)$ includes the starting SDE

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

Now apply $(*)$ for suitable $\tilde{f}$, begin with $\tilde{f} := a$ and $\tilde{f} := b$, and obtain

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

This can be written

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

with remainder

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

The integrands are

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

Analogously the integrands in the double integrals in R can be replaced, by applying $(*)$ with proper $\tilde{f}$. Thereby, the double integrals

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

occur as factors. $\mathcal{I}(1,0), \mathcal{I}(0,1), \mathcal{I}(1,1)$ are stochastic variables. By a plausibility argument (replace $\Delta W_s := W_s - W_{t_0}$ by its expectation $\sqrt{s-t_0}$) expect that $I(1,1)$ is the integral of lowest order: $O(\Delta t)$. We begin with this integral, for $\tilde{f} := L^1 b(X)$. From $(*)$ conclude

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

Then R consists of

$$R = \text{ three double integrals } + \underbrace{b(X_{t_0})b'(X_{t_0})}_{=L^1 b(X_{t_0})}\mathcal{I}(1,1) + \text{ two triple integrals}$$

Calculate the double integral $\mathcal{I}(1,1)$:

Let $g(x) := x^2$ and $X_t = W_t$, which solves an SDE with $a = 0$ and $b = 1$. The Itô lemma implies

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

which in turn yields

$$\int_{t_0}^{t} \int_{t_0}^{s} \mathrm{d}W_z \, \mathrm{d}W_s = \int_{t_0}^{t} (W_s - W_{t_0}) \, \mathrm{d}W_s = \int_{t_0}^{t} W_s \, \mathrm{d}W_s - W_{t_0} \int_{t_0}^{t} \mathrm{d}W_s =$$

$$\int_{t_0}^{t} \tfrac{1}{2} \left[\mathrm{d}(W_s^2) - \mathrm{d}s \right] - W_{t_0}(W_t - W_{t_0}) =$$

$$\tfrac{1}{2}(W_t^2 - W_{t_0}^2) - \tfrac{1}{2}(t - t_0) - \tfrac{2}{2} W_{t_0}(W_t - W_{t_0}) = \tfrac{1}{2}(\Delta W_t)^2 - \tfrac{1}{2}\Delta t \, .$$

This confirms the anticipated order $O(\Delta t)$ of $\mathcal{I}(1,1)$.

The above derivation of the stochastic Taylor expansion can be continued. This calls for a systematic definition and notation of the multi-integrals, for example, $\mathcal{I}(0,0,0), \ldots$. In this notation, a "0" stands for a deterministic integration, and a "1" for a stochastic integration.

Application

Attach further leading terms of the stochastic Taylor expansion to obtain integrators of higher order.

Example (Milstein method)

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

The first terms represent the Euler method, and the last term completes the list of $O(\Delta t)$-terms, and improves the low order of strong convergence to 1. The weak order is also 1. (This may be checked empirically.)

Question: What does this result mean in view of SDEs with $b' = 0$?

B. Positivity

As mentioned before, positive solutions are characteristic for many SDEs in finance. This should be preserved by numerical approximations. We discuss this topic for the CIR process, which is part of the Heston model.

Example CIR

$$\mathrm{d}X_t = \kappa(\theta - X_t)\,\mathrm{d}t + \sigma\sqrt{X_t}\,\mathrm{d}W_t$$

with $\kappa, \theta, \sigma > 0$, $X_0 = x_0 > 0$. Positivity of X_t for all t is established by the "Feller condition"

$$\kappa\theta \geq \frac{1}{2}\sigma^2\,,$$

which guarantees a strong enough growth rate. We remark in passing that $b(X) = \sigma\sqrt{X}$ does not satisfy a global Lipschitz condition.

Euler scheme:

$$y_{j+1} = y_j + \kappa(\theta - y_j)\Delta t + \sigma\sqrt{y_j}\,\Delta W_j$$

with $y_0 := x_0$, works as long as $y_j \geq 0$. There is a positive probability that y_{j+1} is negative. When X represents an asset price, an interest rate, or a variance (Heston model), then $y < 0$ must be avoided.

Variants

For example, replace $\sqrt{y}$ by $\sqrt{|y|}$ or by $\sqrt{y^+}$. Then the scheme is defined for all $y \in \mathbb{R}$. Another variant calculates

$$y_{j+1} = | \, y_j + \kappa(\theta - y_j)\Delta t + \sigma\sqrt{y_j}\, \Delta W_j \, | \; .$$

Implicit Euler methods can be applied as well, for example, the drift-implicit scheme

$$y_{j+1} = y_j + a(y_{j+1})\Delta t + b(y_j)\, \Delta W_j \; .$$

If this scheme is applied to the SDE of the square root process $\sqrt{X_t}$, then a quadratic equation for y_{j+1} results with a unique positive solution (Exercise !). [A. Alfonsi (2005)]

3.3 Monte Carlo Methods for European Options

The aim is to calculate the value

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

of a European option, where Ψ denotes the payoff and Q a risk-free probability measure.

A. Basic Principle

The integral of this expectation can be approximated by Monte Carlo methods. The first decision is the choice of the market model (as Heston- or Black–Scholes model). Here we focus on the classic Black-Scholes model with GBM,

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

The procedure is analogous to **Monte Carlo quadrature**.

Monte Carlo quadrature

With respect to $\mathcal{U}[0,1]$,

$$\int_0^1 f(x)\,\mathrm{d}x = \int_{-\infty}^{\infty} f(x)\,\mathbf{1}_{[0,1]}\,\mathrm{d}x = \mathsf{E}(f)\,.$$

Applying the law of large numbers, yields

$$\frac{1}{N}\sum_{k=1}^{N} f(x_k) \longrightarrow \mathsf{E}(f) \quad \text{for } N\to\infty\,,$$

where x_k are independent random uniformly distributed numbers in the domain $\mathcal{D} := [0,1]$, because $\mathbf{1}_{[0,1]}$ is the corresponding density. Hence the sum

$$\frac{1}{N}\sum_{k=1}^{N} f(x_k)$$

approximates the integral $\int_0^1 f(x)\,\mathrm{d}x$.

For general domains $\mathcal{D}$ the approximation is

$$\int_{\mathcal{D}} f(x)\,\mathrm{d}x \approx \frac{\mathrm{Vol}(\mathcal{D})}{N}\sum_{k=1}^{N} f(x_k)\,.$$

The **error** is probabilistic. The central limit theorem provides related assertions, for example:

> With 95% probabilty the true value of the integral lies in the confidence interval around the approximate value, which is given by the half width $a\sigma/\sqrt{N}$. For 95% probability the parameter a is $a = 1.96$, and σ is the standard deviation.

So much on MC applied to quadrature. Now the question is, what is the structure of f when options are to be priced under GBM?

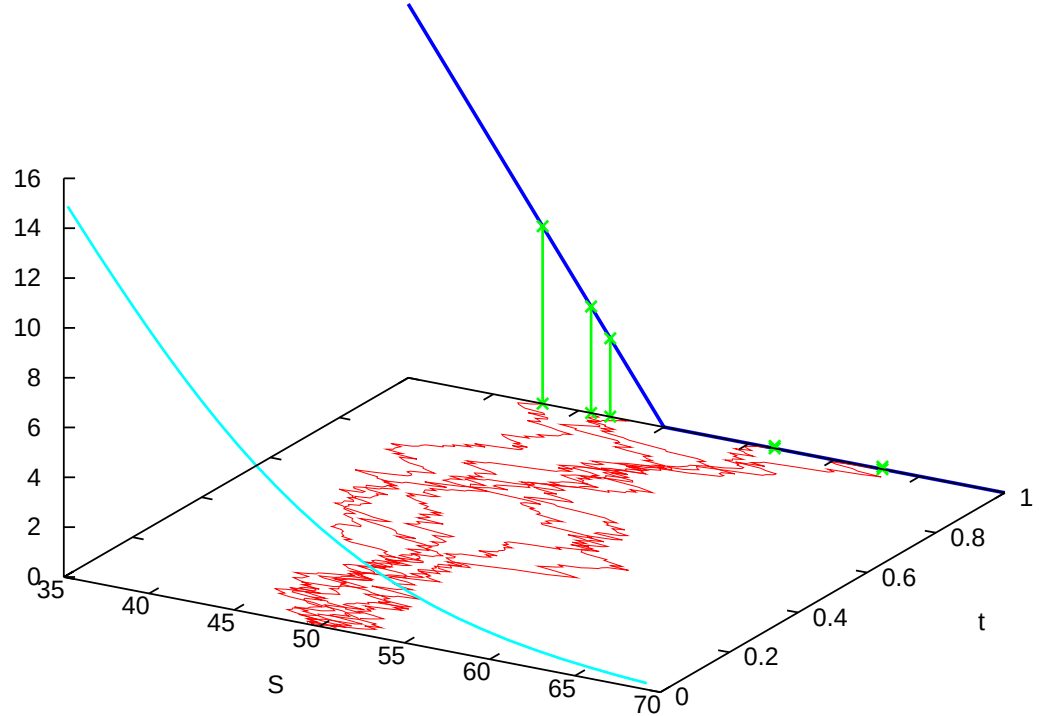
In that case, the density is f_{GBM}, and accordingly the x_k must be distributed lognormally.

Algorithm: Monte Carlo Method for European Options

Simulate N paths of the asset price under the risk-neutral measure Q. Each path starts at S_0, and terminates in $x_k := (S_T)_k$ for $k = 1, ..., N$. Evaluate the payoff $\Psi(S_T)$

$$f(x_k) := \Psi\big((S_T)_k\big)\,,$$

calculate the mean, and discount with factor e^{-rT}. Its expectation yields the true value V as long as f (S_T and Ψ) is unbiased.


16
14
12
10
8
6
4
2
0
35
40
45
50
55
60
65
70
S
0
0.2
0.4
0.6
0.8
1
t

Examples of a payoff Ψ

1.) *Binary* or *digital* option, e.g., binary call:

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

2.) *Barrier* option with barrier B, e.g. a *down-and-out* call option with

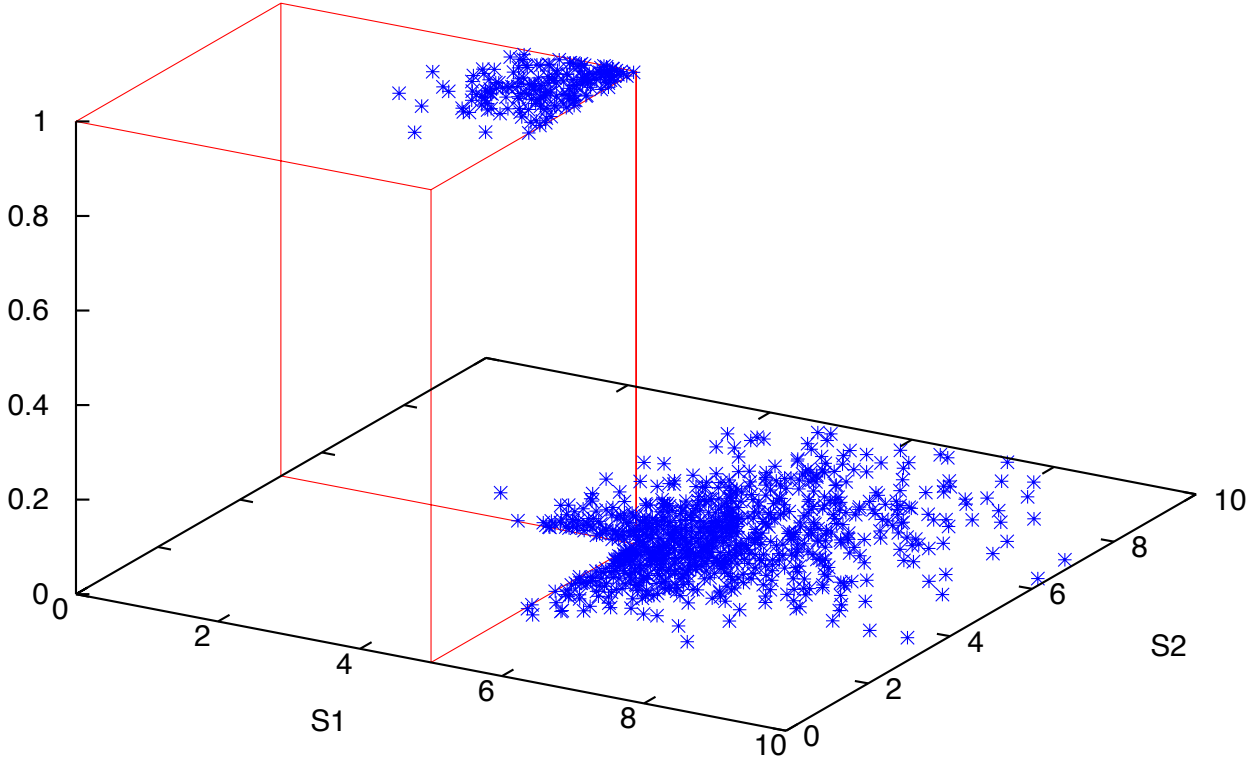
$$\Psi(S) = \begin{cases} 0 & \text{in case } S_t \le B \text{ for a } t \text{ in } 0 \le t \le T \\ (S_T - K)^+ & \text{elsewhere} \end{cases}$$

For this path-dependent exotic option the entire path S_t on $0 \le t \le T$ is of interest. (meaningful for $S_0 > B > K$; illustration in (S,t,V)-space under consideration of boundary conditions along $S = B$)

3.) *two-asset cash-or-nothing* put: The payoff is 1 in case the inequalities $S_1(T) < K_1$ and $S_2(T) < K$ hold, where $S_1(t), S_2(t)$ denote the prices of the two assets.

Hint: Many analytic solution formulas can be found in [E.G. Haug: Option Pricing Formulas].

two-asset cash-or-nothing put; payoff and 1000 MC simulations



(further illustrations in www.compfin.de)

Implementation of the Monte Carlo Method

For the GBM model the true solution

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

can be applied. For options that are not path-dependent this requires only one random number for each path, for generating W_T and S_T. For more general models without analytic solution formula, one must resort to numerical integration (say, with Euler's method). Then Monte Carlo consists of two loops: the outer loop of sampling ($k = 1, ..., N$), and the inner loop of the integration ($j = 1, ..., M$; $\Delta t = \frac{T}{M}$; $t_j = j\Delta t$). For path-dependent GBM models, the analytic formula can be applied in a piecewise fashion,

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

for all j, with $\Delta W = \sqrt{\Delta t}\,Z$, $Z \sim \mathcal{N}(0,1)$.

Dimension. Monte Carlo methods work in the same way for high-dimensional problems. The costs are essentially independent on the dimension. This is an important advantage of Monte Carlo.

B. Accuracy

a) Denote

$$\hat{\mu} := \frac{1}{N}\sum_{k=1}^{N} f(x_k)\,, \quad \hat{s}^2 := \frac{1}{N-1}\sum_{k=1}^{N}\left(f(x_k)-\hat{\mu}\right)^2\,,$$

and $\mu = \mathsf{E}(\hat{\mu})$. According to the **central limit theorem**, the approximation $\hat{\mu}$ obeys $\mathcal{N}(\mu, \sigma^2)$,

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

with distribution function F. In practice σ^2 is replaced by its approximation $\hat{s}^2$. The error behaves as $\frac{\hat{s}}{\sqrt{N}}$. To reduce this statistical error, either reduce the numerator *(variance reduction)*, or enlarge the denominator. The latter means to increase the number of simulations, and is very costly. For example, to gain one additional correct decimal, the error must be reduced by a factor $\frac{1}{10}$, which amounts to raise the costs by a factor of $100 = (\frac{1}{10})^{-2}$.

b) In several cases, the computation of $f(x_i)$ gives rise to another error, namely, the **bias**.

Let $\hat{x}$ be an estimator of the true x that is to be estimated, then the bias is defined as

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

Examples

1.) For a *lookback* option the payoff involves the variable

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

An approximation is

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

Clearly $\hat{x} \le x$. Almost surely $\hat{x}$ underestimates x, i.e. $\mathsf{E}[\hat{x}] < x$. Hence $\text{bias}(\hat{x}) \neq 0$.

2.) Compared to the analytic solution of GBM, the Euler method provides biased results. For GBM,

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

is unbiased, whereas the Euler step

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

is biased.

(Both examples are asymptotically unbiased for $M \to \infty$.)

To reduce the errors, there are several possibilities, which must be compared for costs and tradeoffs. Either

- apply variance reduction,
- increase N, or
- reduce the bias (M larger, Δt smaller),

or apply all these measures. Increasing M and N should be balanced.

The overall error is measured by the mean square error:

Definition (mean square error)

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

As is easily verified,

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

C. Variance Reduction

There are several methods of variance reduction. The simplest is the method of **antithetic variates**. As for the crude MC, paths are simulated with random numbers $Z_1, Z_2, \ldots$ Let us denote the MC approximation $\hat{V}$. The idea of antithetic variates is to use in parallel the numbers $-Z_1, -Z_2, \ldots$, which are also $\sim \mathcal{N}(0,1)$, to calculate "mirror paths" S_t^- from which the payoff values $\Psi(S_T^-)$ are calculated. This leads to a second Monte Carlo value V^-. By construction, $\mathsf{Var}(\hat{V}) = \mathsf{Var}(V^-)$. The effort to calculate V^- is slightly lower than that for $\mathsf{Var}(\hat{V})$ because the Z's are recycled. The mean

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

satisfies

$$\begin{aligned}
\mathsf{Var}(V_{\mathrm{AV}}) &= \frac{1}{4}\mathsf{Var}(\hat{V} + V^-) \\
&= \frac{1}{4}(\mathsf{Var}\hat{V} + \mathsf{Var}V^- + 2\mathsf{Cov}(\hat{V}, V^-)) \\
&= \frac{1}{2}\mathsf{Var}\hat{V} + \frac{1}{2}\mathsf{Cov}(\hat{V}, V^-)
\end{aligned}$$

The anti-symmetric construction of the mirror paths inspires some confidence that the results are negatively correlated, $\mathsf{Cov}(\hat{V}, V^-) < 0$. This holds in case the dependence of the output V on the input Z is monotonic. For $\mathsf{Cov}(\hat{V}, V^-) < 0$ the effect is

$$\mathsf{Var}(V_{\mathrm{AV}}) < \frac{1}{2}\mathsf{Var}(\hat{V})\,.$$

This approach at most doubles the costs. In comparison, an error reduction (factor $< \frac{1}{2}$) by merely increasing N requires at least fourfold costs.

Example GBM: Let the index k in V_k label the MC simulation, $k = 1, \ldots, N$, of GBM. For a payoff Ψ draw $Z_k \sim \mathcal{N}(0,1)$ and calulate the pairs $\hat{V}_k, V_k^-$ and the antithetic variate $V_{\mathrm{AV},k}$ as follows (for $t_0 = 0$):

$$\hat{V}_k = \Psi\left(S_0 \exp\left\{(r - \frac{\sigma^2}{2})T + \sigma\sqrt{T}\,Z_k\right\}\right)$$
$$V_k^- = \Psi\left(S_0 \exp\left\{(r - \frac{\sigma^2}{2})T - \sigma\sqrt{T}\,Z_k\right\}\right)$$
$$V_{\mathrm{AV},k} = \frac{1}{2}(\hat{V}_k + V_k^-)$$

For each k, $\hat{V}_k$ and V_k^- are dependent, but the independence of $Z_k \sim \mathcal{N}(0,1)$ makes the $V_{\mathrm{AV},k}$ for $k = 1, \ldots, N$ independent, and MC is applied: The mean, discounted with factor e^{-rT}, approximates V.

Notice that MC has not been developed for the simple vanilla options. The potential of MC is needed for exotic options, in particular, in high-dimensional situations.

3.4 Monte Carlo Methods for American Options

A. Stopping Time

Examples of decisions in the financial market include selling an asset, or exercising an American option. Let us call the decision "stopping," and the time instant of the decision "stopping time" τ. Such decisions can only be based on the information available so far. Accordingly, a stopping time must be **non-anticipating**: That is, for any time t one must know whether the decision is made, i.e. whether $\tau \le t$ or $\tau > t$.

This characterization of a stopping time can be defined formally with the means of stochastics, building on the underlying process S_t:

Recall the filtration $\mathcal{F}_t$: A stochastic process S_t is called $\mathcal{F}_t$-adapted, if S_t is $\mathcal{F}_t$-measurable for all t. The natural filtration $\mathcal{F}_t^S$ is the smallest sigma-algebra over $\{S_s \mid 0 \le s \le t\}$, augmented by the P-null sets. S_t is $\mathcal{F}_t^S$-adapted. Filtrations represent the amount of information available at time t. Hence we require for the set $\{\tau \le t\}$ of all decisions until t

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

which is the $\mathcal{F}_t$-measurability of τ.

Definition (stopping time)

A stopping time τ with respect to a filtration $\mathcal{F}_t$ is a random variable with vaalues in $[0,T] \cup \{\infty\}$, which is $\mathcal{F}_t$-measurable for all t.

The importance of stopping times for American-style options is highlighted by the following result of Bensoussan (1984):

Let $\Psi(S_t)$ be a payoff, e.g. $\Psi(S_t) = (K - S_t)^+$.
The value of an American option is

$$V(S,0) = \sup_{\substack{0 \le \tau \le T \\ \tau \text{ stopping time}}} \mathsf{E}_{\mathsf{Q}} \left[\mathrm{e}^{-r\tau} \, \Psi(S_\tau) \mid S_0 = S \right] \qquad (*)$$

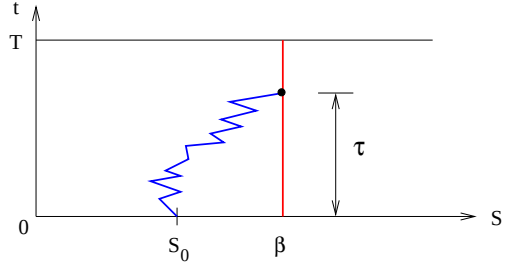
The stopping time τ is with respect to a natural filtration $\mathcal{F}_t$ of S_t.

Examples of stopping times

1) Define the *hitting time*

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

for given $\beta > S_0$. If no such t exists, set $\tau := \infty$.
Clearly, a hitting time is non-anticipating: Setting τ in case of hitting amounts to set a flag in case of hitting, and for any time t check whether the flag is set.*



2) Define t^* as the time instant at which $\max_{0 \leq t \leq T} S_t$ is reached. This is no stopping time! Because for arbitrary t one can not decide whether $t^* \leq t$ or $t^* > t$.

3) A stopping time is given by

$$\tau := \min \{ t \leq T \mid (t, S_t) \in \textit{stopping region} \}$$

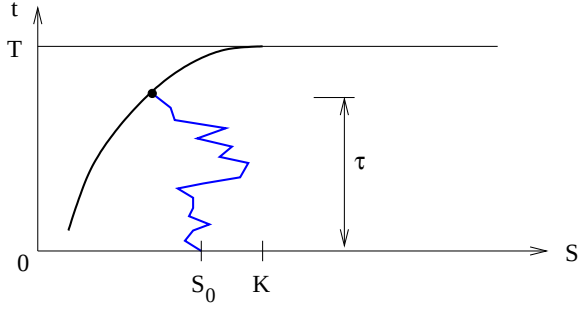
(This is a hitting time for S_t hitting the early-exercise curve; cf. Section 4.5.)

---

* See [Hunt & Kennedy (2000)] for a formal proof that this τ is stopping time.

B. Parametric Methods

In $(*)$ the supremum over *all* stopping times is taken. Now we construct a special stopping time. Similar as in Example 1 or 3 we define a curve in the (S,t)-half strip, which is supposed to approximate the early-exercise curve.



This defines a special stopping strategy $\tilde{\tau}$ by the event of hitting the curve. Assume that β is a vector of parameters defining the curve. Then the stopping rule and $\tilde{\tau}$ depend on β. This special β-depending stopping strategy $\tilde{\tau}$ leads to a **lower bound**

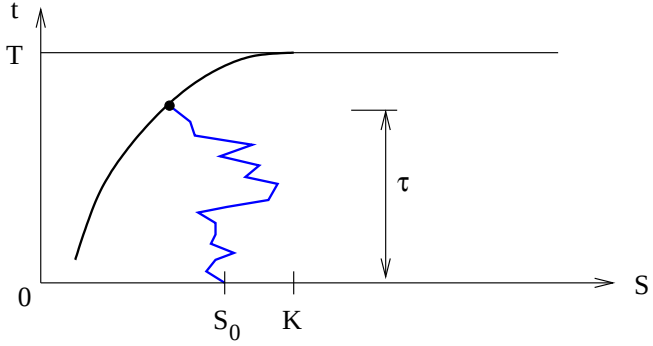
$$V^{\mathrm{low}(\beta)}(S,0) := \mathsf{E}_{\mathsf{Q}}\left[\,\mathrm{e}^{-r\tilde{\tau}}\Psi(S_{\tilde{\tau}}) \mid S_0 = S\,\right] \;\leq V(S,0)\,.$$

Application: Obviously, $V(S,0)$ can be approximated via suitable β-defined stopping curves as

$$\sup_{\beta} V^{\mathrm{low}(\beta)}\,.$$

This idea of an optimal stopping strategy leads to the **procedure**:

Construct a curve depending on a parameter vector β such that the curve approximates the early-exercise curve. The stopping strategy is to stop when the path S_t crosses the curve defined by β.



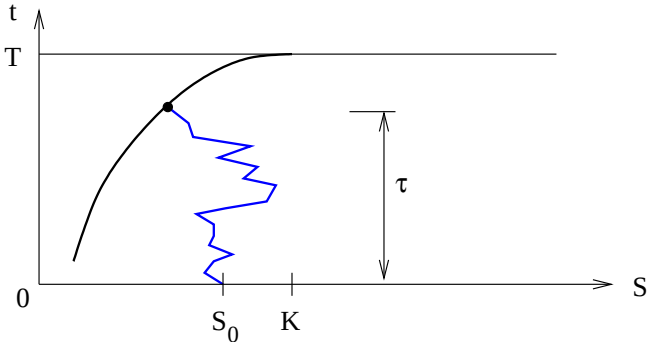
(illustration for a put)

For N such paths evaluate the payoff, and evaluate (approximate) the value $V^{\text{low}(\beta)}$ as crude MC does.

Next attempt to maximize the lower bound $V^{\text{low}(\beta)}$ by repeating the procedure for "better" parameters β.

Example with $\beta \in \mathbb{R}^1$:

Consider a parabola with peak in $(S,t) = (K,T)$, which is defined by one parameter β only. As illustrated in the figure, this parabola can be seen as an approximation of the early-exercise curve of a put. Calculate N paths (e.g. $N = 10000$) until the left branch of the parabola ($\rightarrow \tilde{\tau}$) or $t = T$ is reached. Similar as in Example 1 this hitting time gives rise to an approximation $V^{\mathrm{low}(\beta)}$. Each evaluation of $V^{\mathrm{low}(\beta)}$ costs as much as MC for a European option. Then repeat the procedure with a better β. — This $V^{\mathrm{low}(\beta)}$ will not converge to $V(S,0)$. A systematic error will remain because the eary-exercise curve can not be approximated so well with a simple parabola.



To complete the procedure, one should also construct an upper bound V^{up}.
[P. Glasserman: Monte Carlo Methods in Financial Engineering (2004)]

C. Regression Methods

Definition (Bermudan option)

A Bermudan option is an option that can be exercised only at a finite number M of discrete time instances t_j.

Specifically for

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

we denote the value of a Bermudan option $V^{\mathrm{Be(M)}}$. Because of the additional exercise possibilities,

$$V^{\mathrm{Eu}} \le V^{\mathrm{Be(M)}} \le V^{\mathrm{Am}}$$

holds. One can show

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

For suitable M the value $V^{\mathrm{Be(M)}}$ is used as approximation of V^{Am}. The linear convergence suggests working with a few moderate values of M and apply Richardson extrapolation. In this way, the high costs of Monte Carlo for American options can be kept at a tolerable level.

Recall from the binomial method (Section 1.3): The value of an American option is calculated recursively in a backward fashion, where the continuation values $V^{\rm cont}$ are defined as European options on a strip, and for each t_j

$$V^{\rm Am} = \max\left\{ \Psi(S),\, V^{\rm cont} \right\} .$$

Because at each t_j the holder of the option decides which of the two possibilities {exercise, hold} is optimal.*

For a Bermudan option we define the continuation value at t_j analogously:

$$C_j(x) := {\rm e}^{-r\Delta t}\, \mathsf{E}_{\mathsf{Q}}\big[V(S_{t_{j+1}}, t_{j+1}) \mid S_{t_j} = x \big] .$$

These functions $C_j(x)$ must be approximated.

* principle of dynamic programming

General Recursion

Set $V_M(x) \equiv \Psi(x)$.
For $j = M-1, \ldots, 1$
construct $C_j(x)$ for $x > 0$;
$V_j(x) := V(x, t_j) = \max\{\Psi(x),\, C_j(x)\}$ for grid points x.
$V_0 := V(S_{t_0}, t_0) = \max\{\, \Psi(S_0),\, C_0(S_0)\,\}$.

Below we define special x by a **stochastic grid**.

To calculate the functions $C_j(x)$ with Monte Carlo, one draws information out of paths established by simulation, and approximates $C_j(x)$ by a regression curve $\hat{C}_j(x)$.

Regression (basic version)

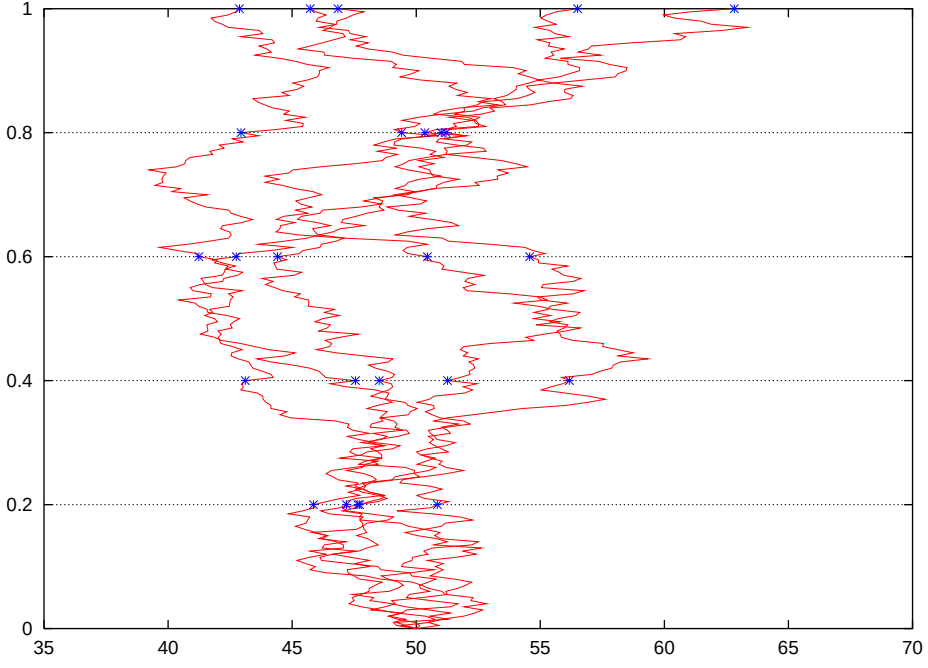
(a) Simulate N paths $S_1(t), \ldots, S_N(t)$

Regression (basic version)

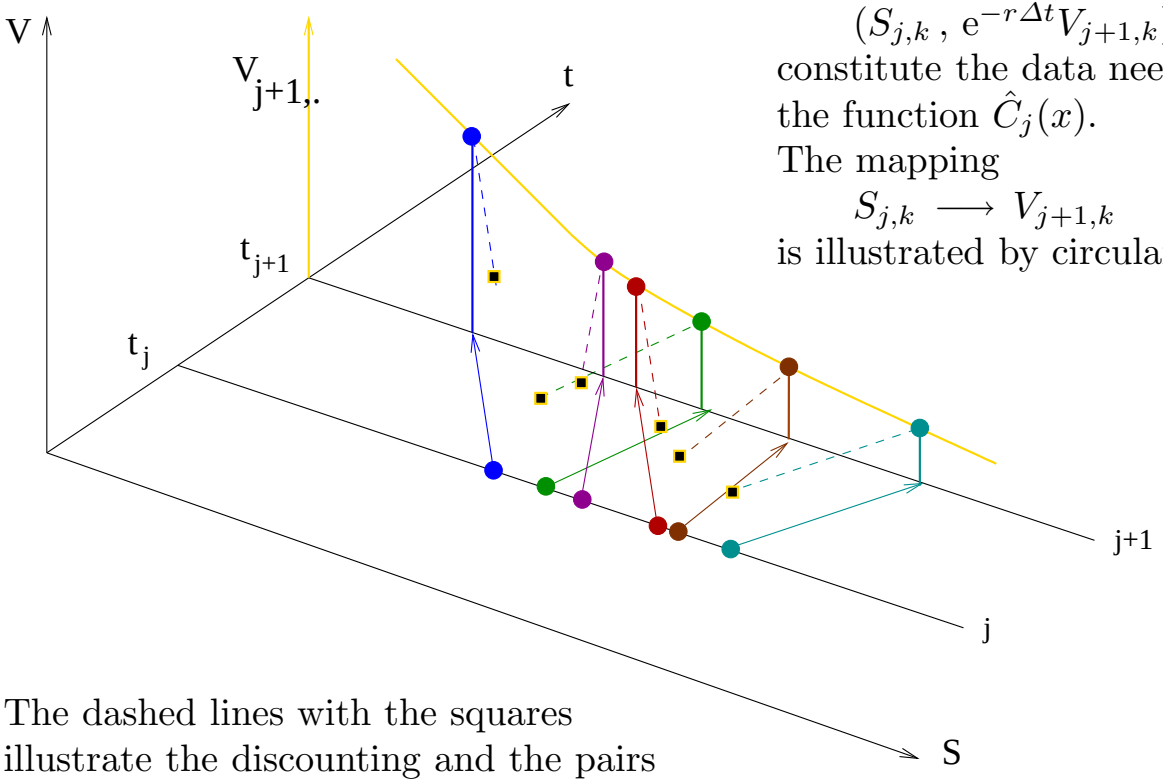
(a) Simulate N paths $S_1(t), \ldots, S_N(t)$: Calculate and store the values

$$S_{j,k} := S_k(t_j)\,, \quad j = 1, \ldots, M,\ k = 1, \ldots, N\,.$$

Illustration: five trajectories and points $(S_{j,k}, t_j)$ for $j = 1, \ldots, 5$, $k = 1, \ldots, 5$



Mapping C, illustration for $N = 6$



The N pairs $(k = 1, \ldots, N)$

$$(S_{j,k}\ ,\ \mathrm{e}^{-r\Delta t} V_{j+1,k})$$

constitute the data needed to calculate the function $\hat{C}_j(x)$.
The mapping

$$S_{j,k} \longrightarrow V_{j+1,k}$$

is illustrated by circular dots.

The dashed lines with the squares illustrate the discounting and the pairs $(S_{j,k}\ ,\ \mathrm{e}^{-r\Delta t} V_{j+1,k})$, which enter a least squares procedure to generate $\hat{C}_j(x)$.

Regression (continue)

(b) For $j = M$ set $V_{M,k} = \Psi(S_{M,k})$ for all k.

(c) For $j = M-1, ..., 1$:
Approximate $C_j(x)$ with suitable base functions $\phi_0, ..., \phi_L$ (e.g. monomials)

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

To this end, apply least-squares minimization on the N points

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

to get the coefficients $a_0, \ldots, a_L$ and thus $\hat{C}_j$.
Evaluation:

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

(d) Set

$$V_0 := \max\left\{\,\Psi(S_0),\, \mathrm{e}^{-r\Delta t}\frac{1}{N}(V_{1,1} + ... + V_{1,N})\,\right\}.$$

Costs

The expensive steps are (a) and (c). Step (d) is needed because the algorithm of (c) can not be applied for $j = 0$, since $S_{0,k} = S_0$ for all k. Instead, the mean in (d) is taken. Convergence of the algorithm was proved.

Based on this regression framework, the algorithm of Longstaff & Schwartz (2001) is built, as well as the even more efficient algorithm by C. Jonen (2009). In particular, step (c) offers potential for improvements.

The costs of step (c) of the above algorithm for American options do depend on the dimension. (Why?) When the entire computing time for pricing an option is limited, then the dependence of the dimension restricts the achievable accuracy. In this sense, the error of MC depends on the dimension.

Supplement to Section 3.4

Longstaff & Schwartz modify the algorithm as follows:

A dynamical-programming principle is incorporated for the optimal stopping times. Each path has its own stopping time τ_k for $k = 1, ..., N$. (It suffices to store the index k since $\tau_k = k\Delta t$.) This algorithm takes advantage of the possibility to work across several time levels. Due to a modification of C. Jonen [Intern. J. Computer Math. **86** (2009); PhD 2011] this is an efficient method.

Algorithm

Initialization: $\tau_k := M$ for all k.

For each $j = M-1, ..., 1$:

loop over all paths $k = 1, ..., N$:

In case $\Psi(S_{j,k}) \geq \hat{C}_j(S_{j,k})$ set $\tau_k := j$.

Otherwise leave τ_k unchanged.

For further hints on regression, and on the computation of sensitivities (Greeks), consult [R. Seydel: Tools for Computational Finance]. See also Topic 6 in the *Topics for CF*.