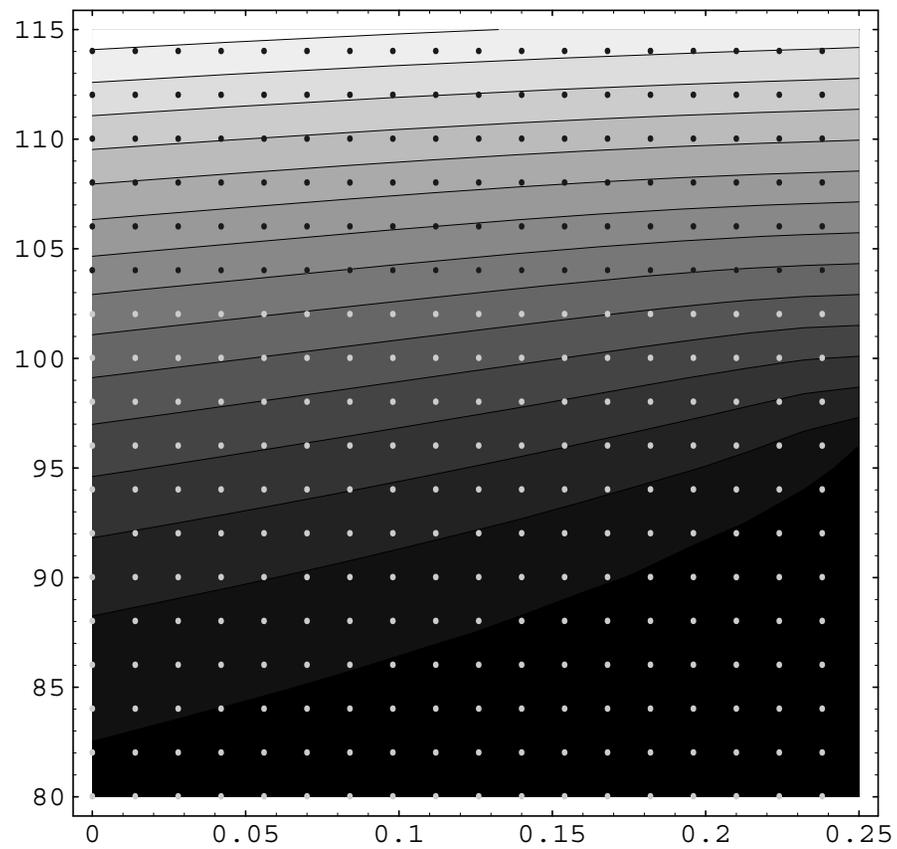


Numerical Methods

All science is dominated
by the idea of approximation.
— Bertrand Russell

Finite-Difference Methods

- Place a grid of points on the space over which the desired function takes value.
- Then approximate the function value at each of these points (p. 836).
- Solve the equation numerically by introducing difference equations in place of derivatives.



Example: Poisson's Equation

- It is $\partial^2\theta/\partial x^2 + \partial^2\theta/\partial y^2 = -\rho(x, y)$, which describes the electrostatic field.
- Replace second derivatives with finite differences through central difference.
- Introduce evenly spaced grid points with distance of Δx along the x axis and Δy along the y axis.
- The finite-difference form is

$$-\rho(x_i, y_j) = \frac{\theta(x_{i+1}, y_j) - 2\theta(x_i, y_j) + \theta(x_{i-1}, y_j)}{(\Delta x)^2} + \frac{\theta(x_i, y_{j+1}) - 2\theta(x_i, y_j) + \theta(x_i, y_{j-1})}{(\Delta y)^2}.$$

Example: Poisson's Equation (concluded)

- In the above, $\Delta x \triangleq x_i - x_{i-1}$ and $\Delta y \triangleq y_j - y_{j-1}$ for $i, j = 1, 2, \dots$
- When the grid points are evenly spaced in both axes so that $\Delta x = \Delta y = h$, the difference equation becomes

$$\begin{aligned} -h^2 \rho(x_i, y_j) &= \theta(x_{i+1}, y_j) + \theta(x_{i-1}, y_j) \\ &+ \theta(x_i, y_{j+1}) + \theta(x_i, y_{j-1}) - 4\theta(x_i, y_j). \end{aligned}$$

- Given boundary values, we can solve for the x_i s and the y_j s within the square $[\pm L, \pm L]$.
- From now on, $\theta_{i,j}$ will denote the finite-difference approximation to the exact $\theta(x_i, y_j)$.

Explicit Methods

- Consider the diffusion equation^a
 $D(\partial^2\theta/\partial x^2) - (\partial\theta/\partial t) = 0, D > 0.$
- Use evenly spaced grid points (x_i, t_j) with distances Δx and Δt , where $\Delta x \triangleq x_{i+1} - x_i$ and $\Delta t \triangleq t_{j+1} - t_j$.
- Employ central difference for the second derivative and forward difference for the time derivative to obtain

$$\left. \frac{\partial\theta(x, t)}{\partial t} \right|_{t=t_j} = \frac{\theta(x, \boxed{t_{j+1}}) - \theta(x, \boxed{t_j})}{\Delta t} + \dots, \quad (119)$$

$$\left. \frac{\partial^2\theta(x, t)}{\partial x^2} \right|_{x=x_i} = \frac{\theta(\boxed{x_{i+1}}, t) - 2\theta(\boxed{x_i}, t) + \theta(\boxed{x_{i-1}}, t)}{(\Delta x)^2} + \dots \quad (120)$$

^aIt is a parabolic partial differential equation.

Explicit Methods (continued)

- Next, assemble Eqs. (119) and (120) into a single equation at (x_i, t_j) .
- But we need to decide how to evaluate x in the first equation and t in the second.
- Since central difference around x_i is used in Eq. (120), we might as well use x_i for x in Eq. (119).
- Two choices are possible for t in Eq. (120).
- The first choice uses $t = t_j$ to yield the following finite-difference equation,

$$\frac{\theta_{i,j+1} - \theta_{i,j}}{\Delta t} = D \frac{\theta_{i+1,j} - 2\theta_{i,j} + \theta_{i-1,j}}{(\Delta x)^2}. \quad (121)$$

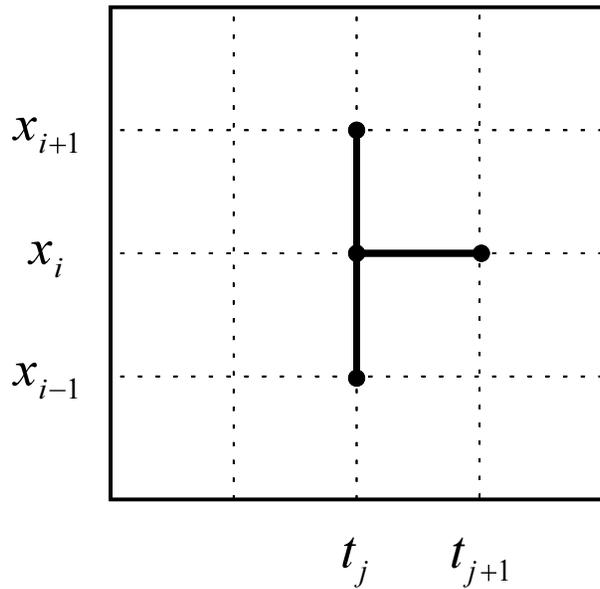
Explicit Methods (continued)

- The stencil of grid points involves four values, $\theta_{i,j+1}$, $\theta_{i,j}$, $\theta_{i+1,j}$, and $\theta_{i-1,j}$.
- Rearrange Eq. (121) on p. 840 as

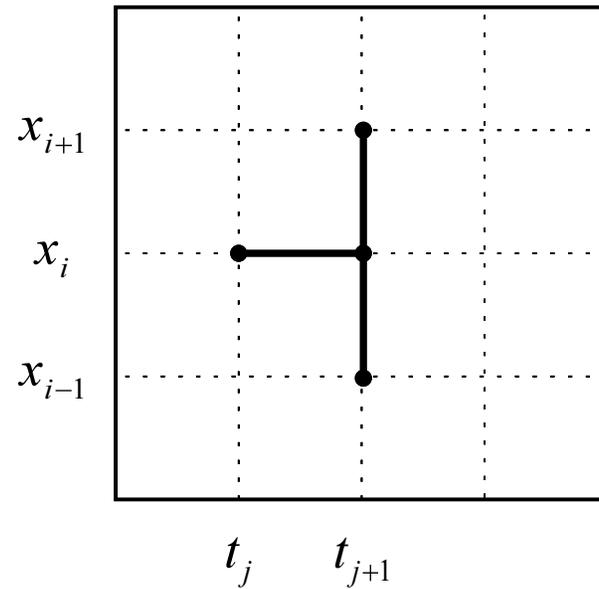
$$\theta_{i,j+1} = \frac{D\Delta t}{(\Delta x)^2} \theta_{i+1,j} + \left(1 - \frac{2D\Delta t}{(\Delta x)^2}\right) \theta_{i,j} + \frac{D\Delta t}{(\Delta x)^2} \theta_{i-1,j}. \quad (122)$$

- We can calculate $\theta_{i,j+1}$ from $\theta_{i,j}$, $\theta_{i+1,j}$, $\theta_{i-1,j}$, at the previous time t_j (see exhibit (a) on next page).

Stencils



(a)



(b)

Explicit Methods (concluded)

- Starting from the initial conditions at t_0 , that is, $\theta_{i,0} = \theta(x_i, t_0)$, $i = 1, 2, \dots$, we calculate

$$\theta_{i,1}, \quad i = 1, 2, \dots .$$

- And then

$$\theta_{i,2}, \quad i = 1, 2, \dots .$$

- And so on.

Stability

- The explicit method is numerically unstable unless

$$\Delta t \leq (\Delta x)^2 / (2D).$$

- A numerical method is unstable if the solution is highly sensitive to changes in initial conditions.
- The stability condition may lead to high running times and memory requirements.
- For instance, halving Δx would imply quadrupling $(\Delta t)^{-1}$, resulting in a running time 8 times as much.

Explicit Method and Trinomial Tree

- Recall Eq. (122) on p. 841:

$$\theta_{i,j+1} = \frac{D\Delta t}{(\Delta x)^2} \theta_{i+1,j} + \left(1 - \frac{2D\Delta t}{(\Delta x)^2}\right) \theta_{i,j} + \frac{D\Delta t}{(\Delta x)^2} \theta_{i-1,j}.$$

- When the stability condition is satisfied, the three coefficients for $\theta_{i+1,j}$, $\theta_{i,j}$, and $\theta_{i-1,j}$ all lie between zero and one and sum to one.
- They can be interpreted as probabilities.
- So the finite-difference equation becomes identical to backward induction on trinomial trees!

Explicit Method and Trinomial Tree (concluded)

- The freedom in choosing Δx corresponds to similar freedom in the construction of trinomial trees.
- The explicit finite-difference equation is also identical to backward induction on a binomial tree.^a
 - Let the binomial tree take 2 steps each of length $\Delta t/2$.
 - It is now a trinomial tree.

^aHilliard (2014).

Implicit Methods

- Suppose we use $t = t_{j+1}$ in Eq. (120) on p. 839 instead.
- The finite-difference equation becomes

$$\frac{\theta_{i,j+1} - \theta_{i,j}}{\Delta t} = D \frac{\theta_{i+1,j+1} - 2\theta_{i,j+1} + \theta_{i-1,j+1}}{(\Delta x)^2}. \quad (123)$$

- The stencil involves $\theta_{i,j}$, $\theta_{i,j+1}$, $\theta_{i+1,j+1}$, and $\theta_{i-1,j+1}$.
- This method is now implicit:
 - The value of any one of the three quantities at t_{j+1} cannot be calculated unless the other two are known.
 - See exhibit (b) on p. 842.

Implicit Methods (continued)

- Equation (123) can be rearranged as

$$\theta_{i-1,j+1} - (2 + \gamma) \theta_{i,j+1} + \theta_{i+1,j+1} = -\gamma \theta_{i,j},$$

where $\gamma \triangleq (\Delta x)^2 / (D \Delta t)$.

- This equation is unconditionally stable.
- Suppose the boundary conditions are given at $x = x_0$ and $x = x_{N+1}$.
- After $\theta_{i,j}$ has been calculated for $i = 1, 2, \dots, N$, the values of $\theta_{i,j+1}$ at time t_{j+1} can be computed as the solution to the following tridiagonal linear system,

Implicit Methods (continued)

$$\begin{bmatrix} a & 1 & 0 & \cdots & \cdots & \cdots & 0 \\ 1 & a & 1 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & a & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & 0 & 1 & a & 1 \\ 0 & \cdots & \cdots & \cdots & 0 & 1 & a \end{bmatrix} \begin{bmatrix} \theta_{1,j+1} \\ \theta_{2,j+1} \\ \theta_{3,j+1} \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \theta_{N,j+1} \end{bmatrix} = \begin{bmatrix} -\gamma\theta_{1,j} - \theta_{0,j+1} \\ -\gamma\theta_{2,j} \\ -\gamma\theta_{3,j} \\ \vdots \\ \vdots \\ -\gamma\theta_{N-1,j} \\ -\gamma\theta_{N,j} - \theta_{N+1,j+1} \end{bmatrix},$$

where $a \triangleq -2 - \gamma$.

Implicit Methods (concluded)

- Tridiagonal systems can be solved in $O(N)$ time and $O(N)$ space.
 - Never invert a matrix to solve a tridiagonal system.
- The matrix above is nonsingular when $\gamma \geq 0$.
 - A square matrix is nonsingular if its inverse exists.

Crank-Nicolson Method

- Take the average of explicit method (121) on p. 840 and implicit method (123) on p. 847:

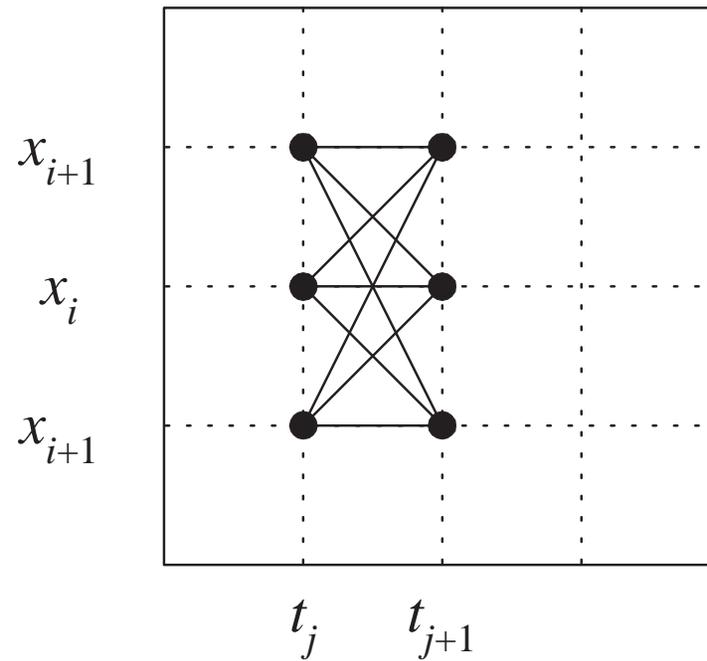
$$\begin{aligned} & \frac{\theta_{i,j+1} - \theta_{i,j}}{\Delta t} \\ = & \frac{1}{2} \left(D \frac{\theta_{i+1,j} - 2\theta_{i,j} + \theta_{i-1,j}}{(\Delta x)^2} + D \frac{\theta_{i+1,j+1} - 2\theta_{i,j+1} + \theta_{i-1,j+1}}{(\Delta x)^2} \right). \end{aligned}$$

- After rearrangement,

$$\gamma\theta_{i,j+1} - \frac{\theta_{i+1,j+1} - 2\theta_{i,j+1} + \theta_{i-1,j+1}}{2} = \gamma\theta_{i,j} + \frac{\theta_{i+1,j} - 2\theta_{i,j} + \theta_{i-1,j}}{2}.$$

- This is an unconditionally stable implicit method with excellent rates of convergence.

Stencil



Numerically Solving the Black-Scholes PDE (94) on p. 687

- See text.
- Brennan and Schwartz (1978) analyze the stability of the implicit method.

Monte Carlo Simulation^a

- Monte Carlo simulation is a sampling scheme.
- In many important applications within finance and without, Monte Carlo is one of the few feasible tools.
- When the time evolution of a stochastic process is not easy to describe analytically, Monte Carlo may very well be the only strategy that succeeds consistently.

^aA top 10 algorithm (Dongarra & Sullivan, 2000).

The Big Idea

- Assume X_1, X_2, \dots, X_n have a joint distribution.
- $\theta \triangleq E[g(X_1, X_2, \dots, X_n)]$ for some function g is desired.
- We generate

$$\left(x_1^{(i)}, x_2^{(i)}, \dots, x_n^{(i)}\right), \quad 1 \leq i \leq N$$

independently with the same joint distribution as (X_1, X_2, \dots, X_n) .

- Output $\bar{Y} \triangleq (1/N) \sum_{i=1}^N Y_i$, where

$$Y_i \triangleq g\left(x_1^{(i)}, x_2^{(i)}, \dots, x_n^{(i)}\right).$$

The Big Idea (concluded)

- Y_1, Y_2, \dots, Y_N are independent and identically distributed random variables.
- Each Y_i has the same distribution as

$$Y \triangleq g(X_1, X_2, \dots, X_n).$$

- Since the average of these N random variables, \bar{Y} , satisfies $E[\bar{Y}] = \theta$, it can be used to estimate θ .
- The strong law of large numbers says that this procedure converges almost surely.
- The number of replications (or independent trials), N , is called the sample size.

Accuracy

- The Monte Carlo estimate and true value may differ owing to two reasons:
 1. Sampling variation.
 2. The discreteness of the sample paths.^a
- The first can be controlled by the number of replications.
- The second can be controlled by the number of observations along the sample path.

^aThis may not be an issue if the financial derivative only requires discrete sampling along time, such as the *discrete* barrier option.

Accuracy and Number of Replications

- The statistical error of the sample mean \bar{Y} of the random variable Y grows as $1/\sqrt{N}$.
 - Because $\text{Var}[\bar{Y}] = \text{Var}[Y]/N$.
- In fact, this convergence rate is asymptotically optimal.^a
- So the variance of the estimator \bar{Y} can be reduced by a factor of $1/N$ by doing N times as much work.
- This is amazing because the same order of convergence holds independently of the dimension n .

^aThe Berry-Esseen theorem.

Accuracy and Number of Replications (concluded)

- In contrast, classic numerical integration schemes have an error bound of $O(N^{-c/n})$ for some constant $c > 0$.
- The required number of evaluations thus grows exponentially in n to achieve a given level of accuracy.
 - The curse of dimensionality.
- The Monte Carlo method is more efficient than alternative procedures for multivariate derivatives for n large.

Monte Carlo Option Pricing

- For the pricing of European options on a dividend-paying stock, we may proceed as follows.
- Assume

$$\frac{dS}{S} = \mu dt + \sigma dW.$$

- Stock prices S_1, S_2, S_3, \dots at times $\Delta t, 2\Delta t, 3\Delta t, \dots$ can be generated via

$$\begin{aligned} S_{i+1} \\ = S_i e^{(\mu - \sigma^2/2) \Delta t + \sigma \sqrt{\Delta t} \xi}, \quad \xi \sim N(0, 1), \end{aligned} \quad (124)$$

by Eq. (87) on p. 621.

Monte Carlo Option Pricing (continued)

- If we discretize $dS/S = \mu dt + \sigma dW$ directly, we will obtain

$$S_{i+1} = S_i + S_i \mu \Delta t + S_i \sigma \sqrt{\Delta t} \xi.$$

- But this is locally normally distributed, not lognormally, hence biased.^a
- Negative stock prices are also possible.^b
- In practice, this is not expected to be a major problem as long as Δt is sufficiently small.

^aContributed by Mr. Tai, Hui-Chin (R97723028) on April 22, 2009.

^bContributed by Mr. Chen, Yu-Hsing (B06901048, R11922045) on May 5, 2023.

Monte Carlo Option Pricing (continued)

Non-dividend-paying stock prices in a risk-neutral economy can be generated by setting $\mu = r$ and $\Delta t = T$.^a

```
1:  $C := 0$ ; {Accumulated terminal option value.}
2: for  $i = 1, 2, 3, \dots, N$  do
3:    $P := S \times e^{(r - \sigma^2/2)T + \sigma\sqrt{T}\xi}$ ,  $\xi \sim N(0, 1)$ ;
4:    $C := C + \max(P - X, 0)$ ;
5: end for
6: return  $Ce^{-rT}/N$ ;
```

^aIt is sometimes called a one-shot simulation (Brigo & Mercurio, 2006).

Monte Carlo Option Pricing (concluded)

Pricing Asian options is also easy.

```
1:  $C := 0$ ;  
2: for  $i = 1, 2, 3, \dots, N$  do  
3:    $P := S$ ;  $M := S$ ;  
4:   for  $j = 1, 2, 3, \dots, n$  do  
5:      $P := P \times e^{(r - \sigma^2/2)(T/n) + \sigma\sqrt{T/n} \xi}$ ;  
6:      $M := M + P$ ;  
7:   end for  
8:    $C := C + \max(M/(n + 1) - X, 0)$ ;  
9: end for  
10: return  $Ce^{-rT}/N$ ;
```

How about American Options?

- Standard Monte Carlo simulation is inappropriate for American options because of early exercise.
 - Given a sample path S_0, S_1, \dots, S_n , how to decide which S_i is an early-exercise point?
 - What is the option price at each S_i if the option is not exercised?
- It is difficult to determine the early-exercise point based on one single path.^a
- But Monte Carlo simulation can be modified to price American options with small biases.^b

^aUnless, of course, the exercise boundary is given (recall pp. 403ff).
Contributed by Mr. Chen, Tung-Li (D09922014) on May 5, 2023.

^bLongstaff & Schwartz (2001). See pp. 927ff.

Obtaining Profit and Loss of Delta Hedge^a

- Profit and loss of delta hedge should be calculated under the real-world probability measure.^b
- So stock prices should be sampled from

$$\frac{dS}{S} = \mu dt + \sigma dW.$$

- Suppose backward induction on a tree under the risk-neutral measure is performed for the delta.^c

^aContributed by Mr. Lu, Zheng-Liang (D00922011) on August 12, 2021.

^bRecall p. 713.

^cBecause, say, no closed-form formulas are available for the delta.

Obtaining Profit and Loss of Delta Hedge (concluded)

- Note that one needs a delta per stock price.
- So Nn trees are needed for the distribution of the profit and loss from N paths with $n + 1$ stock prices per path.
- These are a lot of trees!
- How to do it efficiently to generate plots like that on p. 656?

Delta and Common Random Numbers

- In estimating delta, it is natural to start with the finite-difference estimate

$$e^{-r\tau} \frac{E[P(S + \epsilon)] - E[P(S - \epsilon)]}{2\epsilon}.$$

- $P(x)$ is the terminal payoff of the derivative security when the underlying asset's initial price equals x .
- Use simulation to estimate $E[P(S + \epsilon)]$ first.
- Use another simulation to estimate $E[P(S - \epsilon)]$.
- Finally, apply the formula to approximate the delta.
- This is also called the bump-and-revalue method.

Delta and Common Random Numbers (concluded)

- This method is not recommended because of its high variance.
- A much better approach is to use common random numbers to lower the variance:

$$e^{-r\tau} E \left[\frac{P(S + \epsilon) - P(S - \epsilon)}{2\epsilon} \right].$$

- Here, the *same* random numbers are used for $P(S + \epsilon)$ and $P(S - \epsilon)$.
- This holds for gamma and cross gamma.^a

^aFor multivariate derivatives.

Problems with the Bump-and-Revalue Method

- Consider the binary option with payoff

$$\begin{cases} 1, & \text{if } S(T) > X, \\ 0, & \text{otherwise.} \end{cases}$$

- Then

$$P(S+\epsilon) - P(S-\epsilon) = \begin{cases} 1, & \text{if } S + \epsilon > X \text{ and } S - \epsilon < X, \\ 0, & \text{otherwise.} \end{cases}$$

- So the finite-difference estimate per run for the (undiscounted) delta is 0 or $O(1/\epsilon)$.
- This means high variance.

Problems with the Bump-and-Revalue Method (concluded)

- The price of the binary option equals

$$e^{-r\tau} N(x - \sigma\sqrt{\tau}).$$

- It equals *minus* the derivative of the European call with respect to X .
 - It also equals $X\tau$ times the rho of a European call (p. 364).
- Its delta is

$$\frac{N'(x - \sigma\sqrt{\tau})}{S\sigma\sqrt{\tau}}.$$

Gamma

- The finite-difference formula for gamma is

$$e^{-r\tau} E \left[\frac{P(S + \epsilon) - 2 \times P(S) + P(S - \epsilon)}{\epsilon^2} \right].$$

- For a correlation option with multiple underlying assets, the finite-difference formula for the cross gamma $\partial^2 P(S_1, S_2, \dots) / (\partial S_1 \partial S_2)$ is:

$$e^{-r\tau} E \left[\frac{P(S_1 + \epsilon_1, S_2 + \epsilon_2) - P(S_1 - \epsilon_1, S_2 + \epsilon_2)}{4\epsilon_1 \epsilon_2} - \frac{P(S_1 + \epsilon_1, S_2 - \epsilon_2) + P(S_1 - \epsilon_1, S_2 - \epsilon_2)}{4\epsilon_1 \epsilon_2} \right].$$

Gamma (continued)

- Choosing an ϵ of the right magnitude can be challenging.
 - If ϵ is too large, inaccurate Greeks result.
 - If ϵ is too small, unstable Greeks result.
- This phenomenon is sometimes called the curse of differentiation.^a

^aAit-Sahalia & Lo (1998); Bondarenko (2003).

Gamma (continued)

- In general, suppose (in some sense)

$$\frac{\partial^i}{\partial \theta^i} e^{-r\tau} E[P(S)] = e^{-r\tau} E \left[\frac{\partial^i P(S)}{\partial \theta^i} \right]$$

holds for all $i > 0$, where θ is a parameter of interest.^a

– A common requirement is Lipschitz continuity.^b

- Then Greeks become integrals.
- As a result, we avoid ϵ , finite differences, and resimulation.

^aThe $\partial^i P(S)/\partial \theta^i$ within $E[\cdot]$ may not be partial differentiation in the classic sense.

^bBroadie & Glasserman (1996).

Gamma (continued)

- This is indeed possible for a broad class of payoff functions.^a
 - Roughly speaking, any payoff function that is equal to a sum of products of differentiable functions and indicator functions with the right kind of support.
 - For example, the payoff of a call is

$$\max(S(T) - X, 0) = (S(T) - X)I_{\{S(T) - X \geq 0\}}.$$

- The results are too technical to cover here (see next page).

^aTeng (R91723054) (2004); Lyuu & Teng (R91723054) (2011).

Gamma (continued)

- Suppose $h(\theta, x) \in \mathcal{H}$ with pdf $f(x)$ for x and $g_j(\theta, x) \in \mathcal{G}$ for $j \in \mathcal{B}$, a finite set of natural numbers.
- Then

$$\begin{aligned}
 & \frac{\partial}{\partial \theta} \int_{\mathfrak{R}} h(\theta, x) \prod_{j \in \mathcal{B}} \mathbf{1}_{\{g_j(\theta, x) > 0\}}(x) f(x) dx \\
 = & \int_{\mathfrak{R}} h_{\theta}(\theta, x) \prod_{j \in \mathcal{B}} \mathbf{1}_{\{g_j(\theta, x) > 0\}}(x) f(x) dx \\
 & + \sum_{l \in \mathcal{B}} \left[h(\theta, x) J_l(\theta, x) \prod_{j \in \mathcal{B} \setminus l} \mathbf{1}_{\{g_j(\theta, x) > 0\}}(x) f(x) \right]_{x=\chi_l(\theta)},
 \end{aligned}$$

where

$$J_l(\theta, x) = \text{sign} \left(\frac{\partial g_l(\theta, x)}{\partial x_k} \right) \frac{\partial g_l(\theta, x) / \partial \theta}{\partial g_l(\theta, x) / \partial x} \text{ for } l \in \mathcal{B}.$$

Gamma (concluded)

- Similar results have been derived for Levy processes.^a
- Formulas are also available for credit derivatives.^b
- In queueing networks, this is called infinitesimal perturbation analysis (IPA).^c

^aLyu, Teng (R91723054), & S. Wang (2013).

^bLyu, Teng (R91723054), Tseng, & S. Wang (2014, 2019).

^cCao (1985); Y. C. Ho & Cao (1985).

Biases in Pricing Continuously Monitored Options with Monte Carlo

- We are asked to price a continuously monitored up-and-out call with barrier H .
- The Monte Carlo method samples the stock price at n discrete time points t_1, t_2, \dots, t_n .
- A sample path

$$S(t_0), S(t_1), \dots, S(t_n)$$

is produced.

- Here, $t_0 = 0$ is the current time, and $t_n = T$ is the expiration time of the option.

Biases in Pricing Continuously Monitored Options with Monte Carlo (continued)

- If all of the sampled prices are below the barrier, this sample path pays $\max(S(t_n) - X, 0)$.
- Repeat these steps and average the payoffs for a Monte Carlo estimate.

```

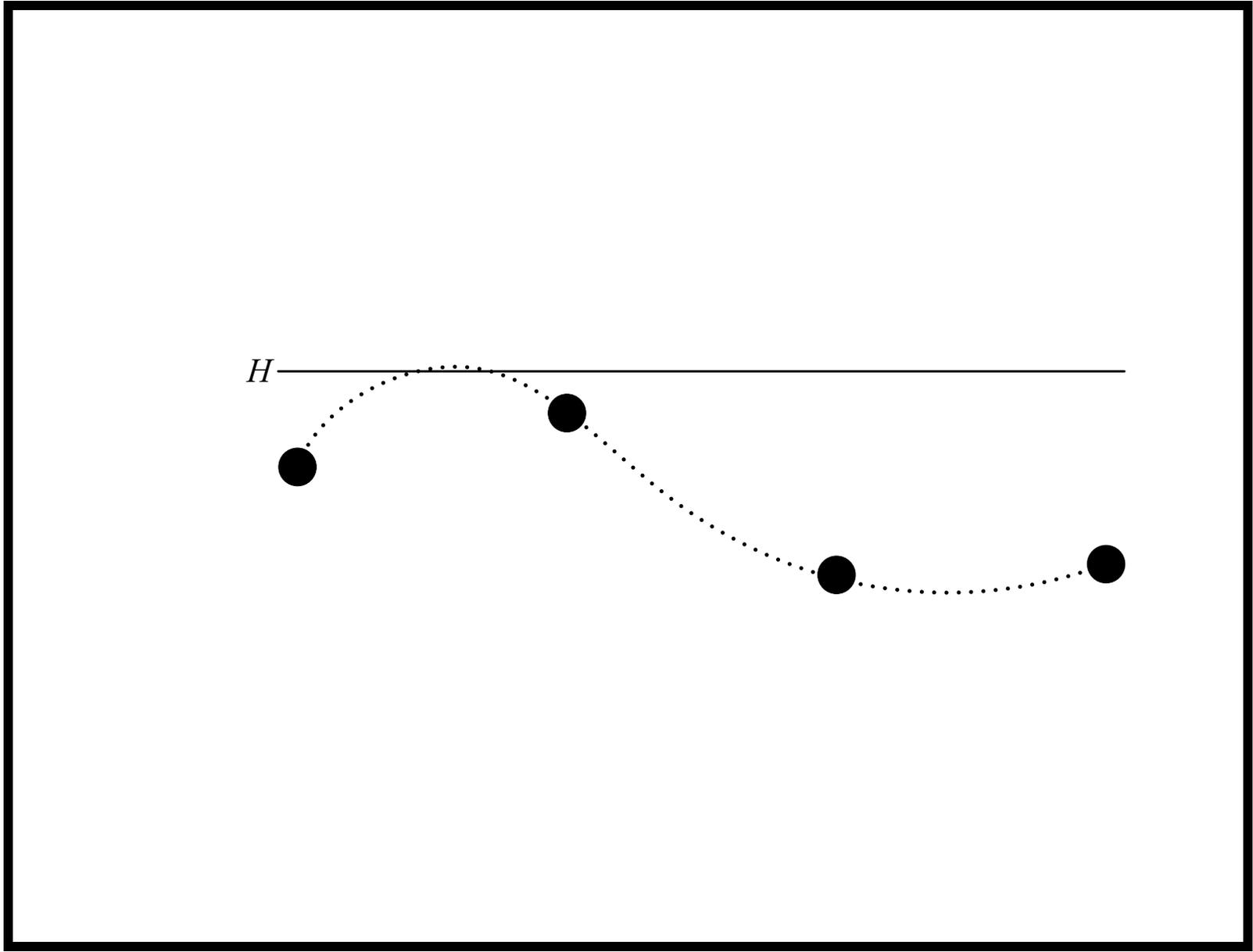
1:  $C := 0$ ;
2: for  $i = 1, 2, 3, \dots, N$  do
3:    $P := S$ ;  $\text{hit} := 0$ ;
4:   for  $j = 1, 2, 3, \dots, n$  do
5:      $P := P \times e^{(r-\sigma^2/2)(T/n)+\sigma\sqrt{(T/n)}\xi}$ ; {By Eq. (124) on p.
      860.}
6:     if  $P \geq H$  then
7:        $\text{hit} := 1$ ;
8:       break;
9:     end if
10:  end for
11:  if  $\text{hit} = 0$  then
12:     $C := C + \max(P - X, 0)$ ;
13:  end if
14: end for
15: return  $Ce^{-rT}/N$ ;

```

Biases in Pricing Continuously Monitored Options with Monte Carlo (continued)

- This estimate is biased.^a
 - Suppose none of the sampled prices on a sample path equals or exceeds the barrier H .
 - It remains possible for the continuous sample path that passes through them to hit the barrier *between* sampled time points (see plot on next page).
 - Hence the knock-out probability is underestimated.

^aShevchenko (2003).



Biases in Pricing Continuously Monitored Options with Monte Carlo (concluded)

- The bias can be lowered by increasing the number of observations along the sample path.
 - For trees, the knock-out probability may *decrease* as the number of time steps is increased.
- However, even daily sampling may not suffice.
- The computational cost also rises as a result.

Brownian Bridge Approach to Pricing Barrier Options

- We desire an unbiased estimate which can be calculated efficiently.
- The above-mentioned payoff should be multiplied by the probability p that a *continuous* sample path does *not* hit the barrier conditional on the sampled prices.
 - Formally,

$$p \triangleq \text{Prob}[S(t) < H, 0 \leq t \leq T \mid S(t_0), S(t_1), \dots, S(t_n)].$$

- This methodology is called the Brownian bridge approach.

Brownian Bridge Approach to Pricing Barrier Options (continued)

- As a barrier is not hit over a time interval if and only if the maximum stock price over that period is at most H ,

$$p = \text{Prob} \left[\max_{0 \leq t \leq T} S(t) < H \mid S(t_0), S(t_1), \dots, S(t_n) \right].$$

- Luckily, the conditional distribution of the maximum over a time interval given the beginning and ending stock prices is known.

Brownian Bridge Approach to Pricing Barrier Options (continued)

Lemma 22 Assume S follows $dS/S = \mu dt + \sigma dW$ and define^a

$$\zeta(x) \triangleq \exp \left[-\frac{2 \ln(x/S(t)) \ln(x/S(t + \Delta t))}{\sigma^2 \Delta t} \right].$$

(1) If $H > \max(S(t), S(t + \Delta t))$, then

$$\text{Prob} \left[\max_{t \leq u \leq t + \Delta t} S(u) < H \mid S(t), S(t + \Delta t) \right] = 1 - \zeta(H).$$

(2) If $h < \min(S(t), S(t + \Delta t))$, then

$$\text{Prob} \left[\min_{t \leq u \leq t + \Delta t} S(u) > h \mid S(t), S(t + \Delta t) \right] = 1 - \zeta(h).$$

^aHere, Δt is an arbitrary positive real number.

Brownian Bridge Approach to Pricing Barrier Options (continued)

- Lemma 22 gives the probability that the barrier is not hit in a time interval, given the starting and ending stock prices.
- For our up-and-out^a call, choose $n = 1$.
- As a result,

$$p = \begin{cases} 1 - \exp \left[-\frac{2 \ln(H/S(0)) \ln(H/S(T))}{\sigma^2 T} \right], & \text{if } H > \max(S(0), S(T)), \\ 0, & \text{otherwise.} \end{cases}$$

^aSo $S(0) < H$ by definition.

Brownian Bridge Approach to Pricing Barrier Options (continued)

The following algorithm works for up-and-out *and* down-and-out calls.

```
1:  $C := 0$ ;  
2: for  $i = 1, 2, 3, \dots, N$  do  
3:    $P := S \times e^{(r-q-\sigma^2/2)T + \sigma\sqrt{T} \xi(i)}$ ;  
4:   if ( $S < H$  and  $P < H$ ) or ( $S > H$  and  $P > H$ ) then  
5:      $C := C + \max(P - X, 0) \times \left\{ 1 - \exp \left[ -\frac{2 \ln(H/S) \times \ln(H/P)}{\sigma^2 T} \right] \right\}$ ;  
6:   end if  
7: end for  
8: return  $C e^{-rT} / N$ ;
```

Brownian Bridge Approach to Pricing Barrier Options (concluded)

- The idea can be generalized.
- For example, we can handle more complex barrier options.
- Consider an up-and-out call with barrier H_i for the time interval $(t_i, t_{i+1}]$, $0 \leq i < m$.
- This option contains m barriers.
- Multiply the probabilities for the m time intervals to obtain the desired probability adjustment term.

Pricing Barrier Options without Brownian Bridge

- Let T_h denote the amount of time for a process X_t to hit h for the *first* time.
- It is called the first passage time or the first hitting time.
- Suppose X_t is a (μ, σ) Brownian motion:

$$dX_t = \mu dt + \sigma dW_t, \quad t \geq 0.$$

Pricing Barrier Options without Brownian Bridge (continued)

- The first passage time T_h follows the inverse Gaussian (IG) distribution with probability density function:^a

$$\frac{|h - X(0)|}{\sigma t^{3/2} \sqrt{2\pi}} e^{-(h - X(0) - \mu x)^2 / (2\sigma^2 x)}.$$

- For pricing a barrier option with barrier H by simulation, the density function becomes

$$\frac{|\ln(H/S(0))|}{\sigma t^{3/2} \sqrt{2\pi}} e^{-[\ln(H/S(0)) - (r - \sigma^2/2)x]^2 / (2\sigma^2 x)}.$$

^aA. N. Borodin & Salminen (1996), with Laplace transform $E[e^{-\lambda T_h}] = e^{-|h - X(0)|\sqrt{2\lambda}}$, $\lambda > 0$.

Pricing Barrier Options without Brownian Bridge (concluded)

- Draw an x from this distribution.^a
- If $x > T$, a knock-in option fails to knock in, whereas a knock-out option does not knock out.
- If $x \leq T$, the opposite is true.
- If the barrier option survives at maturity T , then draw an $S(T)$ to calculate its payoff.
- Repeat the above process and average the discounted payoff.

^aThe IG distribution can be very efficiently sampled (Michael, Schucany, & Haas, 1976).

Brownian Bridge Approach to Pricing Lookback Options^a

- By Lemma 22(1) (p. 885),

$$\begin{aligned} F_{\max}(y) &\triangleq \text{Prob} \left[\max_{0 \leq t \leq T} S(t) < y \mid S(0), S(T) \right] \\ &= 1 - \exp \left[-\frac{2 \ln(y/S(0)) \ln(y/S(T))}{\sigma^2 T} \right]. \end{aligned}$$

- So F_{\max} is the conditional distribution function of the maximum stock price.

^aEl Babsiri & Noel (1998).

Brownian Bridge Approach to Pricing Lookback Options (continued)

- A random variable with that distribution can be generated by $F_{\max}^{-1}(x)$, where x is uniformly distributed over $(0, 1)$.^a
- Note that

$$x = 1 - \exp \left[-\frac{2 \ln(y/S(0)) \ln(y/S(T))}{\sigma^2 T} \right].$$

^aThis is called the inverse-transform technique (see p. 259 of the text-book).

Brownian Bridge Approach to Pricing Lookback Options (continued)

- Equivalently,

$$\begin{aligned} & \ln(1 - x) \\ = & -\frac{2 \ln(y/S(0)) \ln(y/S(T))}{\sigma^2 T} \\ = & -\frac{2}{\sigma^2 T} \{ [\ln(y) - \ln S(0)] [\ln(y) - \ln S(T)] \}. \end{aligned}$$

Brownian Bridge Approach to Pricing Lookback Options (continued)

- There are two solutions for $\ln y$.
- But only one is consistent with $y \geq \max(S(0), S(T))$:

$$\ln y = \frac{\ln(S(0)S(T)) + \sqrt{\left(\ln \frac{S(T)}{S(0)}\right)^2 - 2\sigma^2 T \ln(1-x)}}{2}.$$

Brownian Bridge Approach to Pricing Lookback Options (concluded)

The following algorithm works for the lookback put on the maximum.

- 1: $C := 0$;
- 2: **for** $i = 1, 2, 3, \dots, N$ **do**
- 3: $P := S \times e^{(r-q-\sigma^2/2)T + \sigma\sqrt{T} \xi^{(i)}}$; {By Eq. (124) on p. 860.}
- 4: $Y := \exp \left[\frac{\ln(SP) + \sqrt{(\ln \frac{P}{S})^2 - 2\sigma^2 T \ln[1 - U(0,1)]}}{2} \right]$;
- 5: $C := C + (Y - P)$;
- 6: **end for**
- 7: **return** Ce^{-rT} / N ;

Variance Reduction

- The *statistical* efficiency of Monte Carlo simulation can be measured by the variance of its output.
- If this variance can be lowered without changing the expected value, fewer replications are needed.
- Methods that work in this manner are called variance-reduction techniques.
- Such techniques become practical when the added costs are outweighed by the reduction in sampling.

Variance Reduction: Antithetic Variates

- We want to estimate $E[g(X_1, X_2, \dots, X_n)]$.
- Let Y_1 and Y_2 be random variables with the same distribution as $g(X_1, X_2, \dots, X_n)$.

- Then

$$\text{Var} \left[\frac{Y_1 + Y_2}{2} \right] = \frac{\text{Var}[Y_1]}{2} + \frac{\text{Cov}[Y_1, Y_2]}{2}.$$

- $\text{Var}[Y_1]/2$ is the variance of the Monte Carlo method with two *independent* replications.
- The variance $\text{Var}[(Y_1 + Y_2)/2]$ is smaller than $\text{Var}[Y_1]/2$ when Y_1 and Y_2 are *negatively* correlated.

Variance Reduction: Antithetic Variates (continued)

- For each simulated sample path X , a second one is obtained by *reusing* the first path's random numbers.
- This yields a second sample path Y .
- Two estimates are then obtained: One based on X and the other on Y .
- If N independent sample paths are generated, the antithetic-variates estimator averages over $2N$ estimates.

Variance Reduction: Antithetic Variates (continued)

- Consider process $dX = a_t dt + b_t \sqrt{dt} \xi$.
- Let g be a function of n samples X_1, X_2, \dots, X_n on the sample path.
- Suppose one simulation run has realizations $\xi_1, \xi_2, \dots, \xi_n$ for the normally distributed fluctuation term ξ .
- This generates samples x_1, x_2, \dots, x_n .
- The first estimate is then $g(\mathbf{x})$, where $\mathbf{x} \triangleq (x_1, x_2, \dots, x_n)$.

Variance Reduction: Antithetic Variates (concluded)

- The antithetic-variates method does not sample n more numbers from ξ for the second estimate $g(\mathbf{x}')$.
- Instead, generate the sample path $\mathbf{x}' \triangleq (x'_1, x'_2, \dots, x'_n)$ from $-\xi_1, -\xi_2, \dots, -\xi_n$.
- Compute $g(\mathbf{x}')$.
- Output $(g(\mathbf{x}) + g(\mathbf{x}'))/2$.
- Repeat the above steps.

Variance Reduction: Conditioning

- We are interested in estimating $E[X]$.
- Suppose here is a random variable Z such that $E[X | Z = z]$ can be efficiently and precisely computed.
- $E[X] = E[E[X | Z]]$ by the law of iterated conditional expectations.
- Hence the random variable $E[X | Z]$ is also an unbiased estimator of $E[X]$.

Variance Reduction: Conditioning (concluded)

- As

$$\text{Var}[E[X | Z]] \leq \text{Var}[X],$$

$E[X | Z]$ has a smaller variance than observing X directly.

- First, obtain a random observation z on Z .
- Then calculate $E[X | Z = z]$ as our estimate.
 - There is no need to resort to simulation in computing $E[X | Z = z]$.
- The procedure is repeated to reduce the variance.

Control Variates

- Use the analytic solution of a “similar” yet “simpler” problem to improve the solution.
- Suppose we want to estimate $E[X]$ and there exists a random variable Y with a known mean $\mu \triangleq E[Y]$.
- Then $W \triangleq X + \beta(Y - \mu)$ can serve as a “controlled” estimator of $E[X]$ for any constant β .
 - However β is chosen, W remains an unbiased estimator of $E[X]$ as

$$E[W] = E[X] + \beta E[Y - \mu] = E[X].$$

Control Variates (continued)

- Note that

$$\text{Var}[W] = \text{Var}[X] + \beta^2 \text{Var}[Y] + 2\beta \text{Cov}[X, Y], \quad (125)$$

- Hence W is less variable than X if and only if

$$\beta^2 \text{Var}[Y] + 2\beta \text{Cov}[X, Y] < 0. \quad (126)$$

Control Variates (concluded)

- The success of the scheme clearly depends on both β and the choice of Y .
 - For pricing American options, choose Y to be the otherwise identical European option and μ the Black-Scholes formula.^a
 - For pricing Arithmetic Asian options, choose Y to be the otherwise identical geometric Asian option, μ the formula (58) on p. 447, and $\beta = -1$.
- This approach is often much more effective than the antithetic-variates method.^b

^aHull & White (1988).

^bBoyle, Broadie, & Glasserman (1997).

Choice of Y

- In general, the choice of Y is ad hoc,^a and experiments must be performed to assess the choice.
- Try to match calls with calls and puts with puts.^b
- On many occasions, Y is a discretized version of the derivative that gives μ .
 - Discretely monitored geometric Asian option vs. the continuously monitored version.^c
- The discrepancy can be large (e.g., lookback options).^d

^aBut see Dai (B82506025, R86526008, D8852600), C. Chiu (B90201037, R94922072), & Lyuu (2015, 2018).

^bContributed by Ms. Teng, Huei-Wen (R91723054) on May 25, 2004.

^cPriced by formulas (58) on p. 447.

^dContributed by Mr. Tsai, Hwai (R92723049) on May 12, 2004.

Optimal Choice of β

- Equation (125) on p. 905 is minimized when

$$\beta = -\text{Cov}[X, Y] / \text{Var}[Y].$$

– It is called beta.

- For this specific β ,

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

where $\rho_{X,Y}$ is the correlation between X and Y .

Optimal Choice of β (continued)

- The variance can never increase with the optimal choice.
- The stronger X and Y are correlated, the greater the reduction in variance.
- For example, if this correlation is nearly perfect (± 1), we could control X almost exactly.

Optimal Choice of β (continued)

- Typically, neither $\text{Var}[Y]$ nor $\text{Cov}[X, Y]$ is known.
- So we cannot hope to obtain the maximum reduction in variance.
- We can guess a β and hope that the resulting W does indeed have a smaller variance than X .
- A second possibility is to use the simulated data to estimate $\text{Var}[Y]$ and $\text{Cov}[X, Y]$.
 - How to do it efficiently in terms of time and space?

Optimal Choice of β (concluded)

- Observe that $-\beta$ has the same sign as the correlation between X and Y .
- Hence, if X and Y are positively correlated, $\beta < 0$, then X is adjusted downward whenever $Y > \mu$ and upward otherwise.
- The opposite is true when X and Y are negatively correlated, in which case $\beta > 0$.
- Suppose a suboptimal $\beta + \epsilon$ is used instead.
- The variance increases by only $\epsilon^2 \text{Var}[Y]$.^a

^aHan & Y. Lai (2010).

A Pitfall

- A potential pitfall is to sample X and Y *independently*.
- In this case, $\text{Cov}[X, Y] = 0$.
- Equation (125) on p. 905 becomes

$$\text{Var}[W] = \text{Var}[X] + \beta^2 \text{Var}[Y].$$

- So whatever Y is, the variance is *increased!*
- Lesson: X and Y must be correlated.

Problems with the Monte Carlo Method

- The error bound is only probabilistic.
- The probabilistic error bound of $O(1/\sqrt{N})$ does not benefit from regularity of the integrand function.
- The requirement that the points be independent random samples are wasteful because of clustering.
- In reality, pseudorandom numbers generated by completely deterministic means are used.
- Monte Carlo simulation exhibits a great sensitivity on the seed of the pseudorandom-number generator.

Matrix Computation

To set up a philosophy against physics is rash;
philosophers who have done so
have always ended in disaster.
— Bertrand Russell

Definitions and Basic Results

- Let $A \triangleq [a_{ij}]_{1 \leq i \leq m, 1 \leq j \leq n}$, or simply $A \in \mathbf{R}^{m \times n}$, denote an $m \times n$ matrix.
- It can also be represented as $[a_1, a_2, \dots, a_n]$ where $a_i \in \mathbf{R}^m$ are vectors.
 - Vectors are column vectors unless stated otherwise.
- A is a square matrix when $m = n$.
- The rank of a matrix is the largest number of linearly independent columns.

Definitions and Basic Results (continued)

- A square matrix A is said to be symmetric if $A^T = A$.
- A real $n \times n$ matrix

$$A \triangleq [a_{ij}]_{i,j}$$

is diagonally dominant if $|a_{ii}| > \sum_{j \neq i} |a_{ij}|$ for $1 \leq i \leq n$.

– Such matrices are nonsingular.

- The identity matrix is the square matrix

$$I \triangleq \text{diag}[1, 1, \dots, 1].$$

Definitions and Basic Results (concluded)

- A matrix has full column rank if its columns are linearly independent.
- A real symmetric matrix A is positive definite if

$$x^T Ax = \sum_{i,j} a_{ij} x_i x_j > 0$$

for any nonzero vector x .

- A matrix A is positive definite if and only if there exists a matrix W such that $A = W^T W$ and W has full column rank.

Cholesky Decomposition

- Positive definite matrices can be factored as

$$A = LL^T,$$

called the Cholesky decomposition or Cholesky factorization.

- Above, L is a lower triangular matrix.
- It can be computed by Crout's algorithm in quadratic time.^a

^aGolub and Van Loan (1989).

Generation of Multivariate Distribution

- Let $\mathbf{x} \triangleq [x_1, x_2, \dots, x_n]^T$ be a vector random variable with a positive-definite covariance matrix C .
- As usual, assume $E[\mathbf{x}] = \mathbf{0}$.
- This covariance structure can be matched by $P\mathbf{y}$.
 - $\mathbf{y} \triangleq [y_1, y_2, \dots, y_n]^T$ is a vector random variable with a covariance matrix equal to the identity matrix.
 - $C = PP^T$ is the Cholesky decomposition of C .^a

^aWhat if C is not positive definite? See Y. Y. Lai (R93942114) & Lyuu (2007).

Generation of Multivariate Distribution (concluded)

- For example, suppose

$$C = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}.$$

- Then $PP^T = C$, where^a

$$P = \begin{bmatrix} 1 & 0 \\ \rho & \sqrt{1 - \rho^2} \end{bmatrix}.$$

^aRecall Eq. (28) on p. 180.

Generation of Multivariate Normal Distribution

- Suppose we want to generate the multivariate normal distribution with a covariance matrix $C = PP^T$.
 - First, generate independent standard normal distributions y_1, y_2, \dots, y_n .
 - Then

$$P[y_1, y_2, \dots, y_n]^T$$

has the desired distribution.

Multivariate Derivatives Pricing

- Generating the multivariate normal distribution is essential for the Monte Carlo pricing of multivariate derivatives.^a
- For example, the rainbow option on k assets has payoff

$$\max(\max(S_1, S_2, \dots, S_k) - X, 0)$$

at maturity.

- The closed-form formula is a multi-dimensional integral.^b

^aRecall pp. 822ff.

^bJohnson (1987); C. Y. Chen (D95723006) & Lyuu (2009).

Multivariate Derivatives Pricing (concluded)

- Suppose $dS_j/S_j = r dt + \sigma_j dW_j$, $1 \leq j \leq k$, where C is the correlation matrix for dW_1, dW_2, \dots, dW_k .
- Let $C = PP^T$.
- Let ξ consist of k independent random variables from $N(0, 1)$.
- Let $\xi' = P\xi$.
- Similar to Eq. (124) on p. 860, for each asset $1 \leq j \leq k$,

$$S_{i+1} = S_i e^{(r - \sigma_j^2/2) \Delta t + \sigma_j \sqrt{\Delta t} \xi'_j}$$

by Eq. (124) on p. 860.

Least-Squares Problems

- The least-squares (LS) problem is concerned with

$$\min_{x \in \mathbf{R}^n} \| Ax - b \|,$$

where $A \in \mathbf{R}^{m \times n}$, $b \in \mathbf{R}^m$, and $m \geq n$.

- The LS problem is called regression analysis in statistics and is equivalent to minimizing the mean-square error.
- Often abbreviated as

$$Ax = b.$$

Polynomial Regression

- In polynomial regression, $x_0 + x_1x + \cdots + x_nx^n$ is used to fit the data $\{(a_1, b_1), (a_2, b_2), \dots, (a_m, b_m)\}$.
- This leads to the LS problem,

$$\begin{bmatrix} 1 & a_1 & a_1^2 & \cdots & a_1^n \\ 1 & a_2 & a_2^2 & \cdots & a_2^n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & a_m & a_m^2 & \cdots & a_m^n \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix} .$$

- Consult p. 273 of the textbook for solutions.