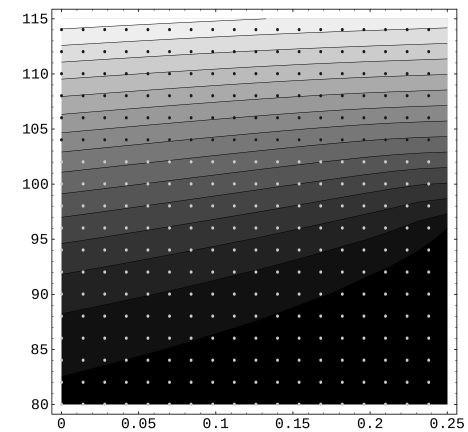
Numerical Methods

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 (see p. 553).
- Solve the equation numerically by introducing difference equations in place of derivatives.



Example: Poisson's Equation

- It is $\partial^2\theta/\partial x^2 + \partial^2\theta/\partial y^2 = -\rho(x, y)$.
- 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, $\Delta x \equiv x_i - x_{i-1}$ and $\Delta y \equiv y_j - y_{j-1}$ for $i, j = 1, 2, \ldots$.
- When the grid points are evenly spaced in both axes so that $\Delta x = \Delta y = h$, the difference equation becomes

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

- Given boundary values, we can solve for the x_is and the y_js 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 $D(\partial^2\theta/\partial x^2) - (\partial\theta/\partial t) = 0$.
- Use evenly spaced grid points (x_i, t_j) with distances Δx and Δt, where $\Delta x \equiv x_{i+1} - x_i$ and $\Delta t \equiv t_{j+1} - t_j$.
- Employ central difference for the second derivative and forward difference for the time derivative to obtain

$$\left.\frac{\partial\theta(x,t)}{\partial t}\right|_{t=t_j} = \frac{\theta(x, t_{j+1}) - \theta(x, t_j)}{\Delta t} + \cdots, \tag{109}$$

$$\left.\frac{\partial^2\theta(x,t)}{\partial x^2}\right|_{x=x_i} = \frac{\theta(x_{i+1}, t) - 2\theta(x_i, t) + \theta(x_{i-1}, t)}{(\Delta x)^2} + \cdots \tag{110}$$

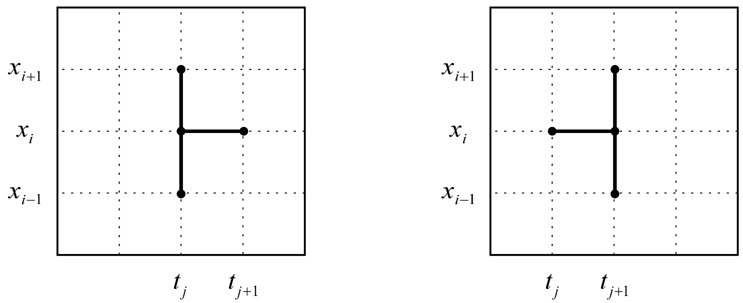
Explicit Methods (continued)

- To assemble Eqs. (109) and (110) into a single equation at (x_i, t_j), 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. (110), we might as well use x_i for x in Eq. (109).
- Two choices are possible for t in Eq. (110).
- The first choice uses $t = t_j$ to yield the following finite-difference equation,

$$\frac{\theta_{i,j+1} - \theta_{i,j}}{\Delta t} = D\,\frac{\theta_{i+1,j} - 2\theta_{i,j} + \theta_{i-1,j}}{(\Delta x)^2}. \tag{111}$$

Explicit Methods (concluded)

- The stencil of grid points involves four values, $\theta_{i,j+1}$, $\theta_{i,j}$, $\theta_{i+1,j}$, and $\theta_{i-1,j}$.
- We can therefore calculate $\theta_{i,j+1}$ from the other three, $\theta_{i,j}, \theta_{i+1,j}, \theta_{i-1,j}$, at the previous time t_j (see figure (a) on next page).
- Starting from the initial conditions at t_0, that is, $\theta_{i,0} = \theta(x_i, t_0)$, $i = 1, 2, \ldots$, we calculate $\theta_{i,1}$, $i = 1, 2, \ldots$, and then $\theta_{i,2}$, $i = 1, 2, \ldots$, and so on.



(a) (b)

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, doubling $(\Delta x)^{-1}$ would imply quadrupling $(\Delta t)^{-1}$, resulting in a running time eight times as much.

Explicit Method and Trinomial Tree

- Rearrange Eq. (111) as
$$\theta_{i,j+1} = \frac{D\Delta t}{(\Delta x)^2}\,\theta_{i+1,j} + \left(1 - \frac{2D\Delta t}{(\Delta x)^2}\right)\theta_{i,j} + \frac{D\Delta t}{(\Delta x)^2}\,\theta_{i-1,j}.$$
- 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 therefore be interpreted as probabilities.
- So the finite-difference equation becomes identical to backward induction on trinomial trees.
- The freedom in choosing Δx corresponds to similar freedom in the construction of the trinomial trees.

Implicit Methods

- If we use $t = t_{j+1}$ in Eq. (110) instead, the finite-difference equation becomes
$$\frac{\theta_{i,j+1} - \theta_{i,j}}{\Delta t} = D\,\frac{\theta_{i+1,j+1} - 2\theta_{i,j+1} + \theta_{i-1,j+1}}{(\Delta x)^2}. \qquad (112)$$
- The stencil involves $\theta_{i,j}$, $\theta_{i,j+1}$, $\theta_{i+1,j+1}$, and $\theta_{i-1,j+1}$.
- This method is implicit because the value of any one of the three quantities at t_{j+1} cannot be calculated unless the other two are known (see figure (b) on p. 559).

Implicit Methods (continued)

- Equation (112) 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 $\gamma \equiv (\Delta x)^2/(D\Delta t)$.
- 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 $i = 1, 2, \ldots, N$, 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)

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

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

Implicit Methods (concluded)

- Tridiagonal systems can be solved in $O(N)$ time and $O(N)$ space.
- 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 (111) and implicit method (112):

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

- 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.

Numerically Solving the Black-Scholes PDE

- We focus on American puts.
- The technique can be applied to any derivative satisfying the Black-Scholes PDE as only the initial and the boundary conditions need to be changed.
- The Black-Scholes PDE for American puts is

$$\frac{1}{2}\sigma^2 S^2 \frac{\partial^2 P}{\partial S^2} + (r-q)\, S\, \frac{\partial P}{\partial S} - rP + \frac{\partial P}{\partial t} = 0$$

  with $P(S,T) = \max(X-S, 0)$ and $P(S,t) = \max(\overline{P}(S,t), X-S)$ for $t < T$.
- $\overline{P}$ denotes the option value at time t if it is not exercised for the next instant of time.

Numerically Solving the Black-Scholes PDE (continued)

- After the change of variable $V \equiv \ln S$, the option value becomes $U(V,t) \equiv P(e^V, t)$ and

$$\frac{\partial P}{\partial t} = \frac{\partial U}{\partial t}, \quad \frac{\partial P}{\partial S} = \frac{1}{S}\frac{\partial U}{\partial V}, \frac{\partial^2 P}{\partial^2 S} = \frac{1}{S^2}\frac{\partial^2 U}{\partial V^2} - \frac{1}{S^2}\frac{\partial U}{\partial V}.$$

- The Black-Scholes PDE is now transformed into

$$\frac{1}{2}\sigma^2 \frac{\partial^2 U}{\partial V^2} + \left(r - q - \frac{\sigma^2}{2}\right)\frac{\partial U}{\partial V} - rU + \frac{\partial U}{\partial t} = 0$$

  subject to $U(V,T) = \max(X - e^V, 0)$ and $U(V,t) = \max(\overline{U}(V,t), X - e^V)$, $t < T$.

Numerically Solving the Black-Scholes PDE (concluded)

- Along the V axis, the grid will span from $V_{\min}$ to $V_{\min} + N \times \Delta V$ at ΔV apart for some suitably small $V_{\min}$.
- So boundary conditions at the lower ($V = V_{\min}$) and upper ($V = V_{\min} + N \times \Delta V$) boundaries will have to be specified.
- S_0 as usual denotes the current stock price.

Explicit Method

- The explicit scheme for the Black-Scholes differential equation is

$$\begin{aligned} 0 &= \frac{1}{2}\sigma^2 \frac{U_{i+1,j} - 2U_{i,j} + U_{i-1,j}}{(\Delta V)^2} \\ &+ \left(r - q - \frac{\sigma^2}{2}\right)\frac{U_{i+1,j} - U_{i-1,j}}{2\Delta V} - rU_{i,j} + \frac{U_{i,j} - U_{i,j-1}}{\Delta t} \end{aligned}$$

  for $1 \leq i \leq N-1$.
- The computation moves backward in time.
- There are $N-1$ difference equations.

Explicit Method (continued)

- Regroup the terms to obtain

$$U_{i,j-1} = aU_{i-1,j} + bU_{i,j} + cU_{i+1,j},$$

where

$$\begin{aligned} a &\equiv \left(\left(\frac{\sigma}{\Delta V} \right)^2 - \frac{r - q - \sigma^2/2}{\Delta V} \right) \frac{\Delta t}{2}, \\ b &\equiv 1 - r\Delta t - \left(\frac{\sigma}{\Delta V} \right)^2 \Delta t, \\ c &\equiv \left(\left(\frac{\sigma}{\Delta V} \right)^2 + \frac{r - q - \sigma^2/2}{\Delta V} \right) \frac{\Delta t}{2}. \end{aligned}$$

Explicit Method (continued)

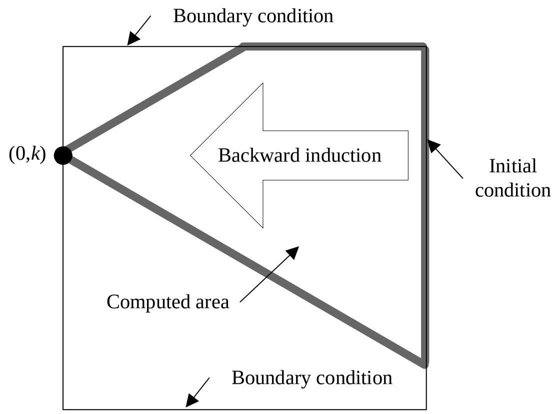
- These $N - 1$ equations express option values at time step $j - 1$ in terms of those at time step j.
- For American puts, we assume for U's lower boundary that the first derivative at grid point $(0, j)$ for every time step j equals $-e^{V_{\min}}$.
- This essentially makes the put value $X - S = X - e^V$.
- So $U_{0,j-1} = U_{1,j-1} + (e^{V_{\min}+\Delta V} - e^{V_{\min}})$.
- For the upper boundary, we set $U_{N,j-1} = 0$.
- The put's value at any grid point at time step $j - 1$ is therefore an explicit function of its values at time step j.

Explicit Method (concluded)

- $U_{i,j}$ is set to the greater of the value derived above and $X - e^{V_{\min}+i\times\Delta V}$ for early-exercise considerations.
- Repeating this process as we move backward in time, we will eventually arrive at the solution at time zero, $U_{k,0}$.
  - k is the integer so that $V_{\min} + k \times \Delta V$ is closest to $\ln S_0$.
- By the stability condition, given ΔV, the value of Δt must be small enough for the method to converge.
  - The conditions to satisfy are $a > 0$, $b > 0$, and $c > 0$.

Region of Influence

- The explicit method evaluates all the grid points in a rectangle.
- But we are only interested in the single grid point at time zero, $(0, k)$, that corresponds to the current stock price.
- The grid points that may influence the desired value form a triangular subset of the rectangle.
- This triangle could be truncated further by the two boundary conditions (see figure on next page).
- Only those points within the truncated triangle need be evaluated.



Implicit Method

The partial differential equation now becomes the following $N-1$ difference equations,

$$\begin{aligned} 0 &= \frac{1}{2}\sigma^2 \frac{U_{i+1,j} - 2U_{i,j} + U_{i-1,j}}{(\Delta V)^2} \\ &+ \left(r - q - \frac{\sigma^2}{2}\right) \frac{U_{i+1,j} - U_{i-1,j}}{2\Delta V} - rU_{i,j} + \frac{U_{i,j+1} - U_{i,j}}{\Delta t} \end{aligned}$$

for $1 \leq i \leq N-1$.

Implicit Method (continued)

Regroup the terms to obtain

$$aU_{i-1,j} + bU_{i,j} + cU_{i+1,j} = U_{i,j+1},$$

where

$$\begin{aligned} a &\equiv \left(-\left(\frac{\sigma}{\Delta V}\right)^2 + \frac{r-q-\sigma^2/2}{\Delta V}\right)\frac{\Delta t}{2}, \\ b &\equiv 1 + r\Delta t + \left(\frac{\sigma}{\Delta V}\right)^2 \Delta t, \\ c &\equiv -\left(\left(\frac{\sigma}{\Delta V}\right)^2 + \frac{r-q-\sigma^2/2}{\Delta V}\right)\frac{\Delta t}{2}. \end{aligned}$$

Implicit Method (continued)

The system of equations can be written in matrix form,

$$\begin{bmatrix} b^* & c & 0 & \cdots & \cdots & \cdots & 0 \\ a & b & c & 0 & \cdots & \cdots & 0 \\ 0 & a & b & c & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & 0 & a & b & c \\ 0 & \cdots & \cdots & \cdots & 0 & a & b \end{bmatrix} \begin{bmatrix} U_{1,j} \\ U_{2,j} \\ U_{3,j} \\ \vdots \\ \vdots \\ \vdots \\ U_{N-1,j} \end{bmatrix} = \begin{bmatrix} U_{1,j+1} - K \\ U_{2,j+1} \\ U_{3,j+1} \\ \vdots \\ \vdots \\ U_{N-2,j+1} \\ U_{N-1,j+1} \end{bmatrix},$$

where $b^* \equiv a + b$ and $K \equiv a(e^{V_{\min}+\Delta V} - e^{V_{\min}})$.

Implicit Method (concluded)

- The values of

$$U_{1,j}, U_{2,j}, \ldots, U_{N-1,j}$$

can be obtained by inverting the tridiagonal matrix.

- As before, at every time step and before going to the next, we should set the option value to the intrinsic value of the option if the latter is larger.

Monte Carlo Simulation

- Monte Carlo simulation is a sampling scheme.
- In many important applications within finance and without, Monte Carlo is one of the few feasible tools.
- It is also one of the most important elements of studying econometrics.
- 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.

The Big Idea

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

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

independently with the same joint distribution as $(X_1, X_2, \ldots, X_n)$ and set

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

The Big Idea (concluded)

- $Y_1, Y_2, \ldots, Y_N$ are independent and identically distributed random variables.
- Each Y_i has the same distribution as $Y \equiv g(X_1, X_2, \ldots, X_n)$.
- Since the average of these N random variables, $\overline{Y}$, satisfies $E[\,\overline{Y}\,] = \theta$, 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.

Example

- Suppose we want to evaluate the definite integral $\int_a^b g(x)\,dx$ numerically.
- Consider the random variable $Y \equiv (b-a)\,g(X)$.
  - X is uniformly distributed over $[\,a,b\,]$.
  - Note that $\text{Prob}[\,X \le x\,] = (x-a)/(b-a)$ for $a \le x \le b$.

Example (concluded)

- Note that

$$\begin{aligned} E[\,Y\,] &= (b-a)\,E[\,g(X)\,] \\ &= (b-a)\int_a^b \frac{g(x)}{b-a}\,dx \\ &= \int_a^b g(x)\,dx. \end{aligned}$$

- So any unbiased estimator of $E[\,Y\,]$ can be used to evaluate the integral.

Accuracy

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

Accuracy and Number of Replications

- The statistical error of the sample mean $\overline{Y}$ of the random variable Y grows as $1/\sqrt{N}$.
  - Because $\text{Var}[\,\overline{Y}\,] = \text{Var}[\,Y\,]/N$.
- In fact, this convergence rate is asymptotically optimal by the Berry-Esseen theorem.
- So the variance of the estimator $\overline{Y}$ 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.

Accuracy and Number of Replications (concluded)

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

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 improve efficiency in this manner are called variance-reduction techniques.
- Such techniques become practical when the added costs are outweighed by the reduction in sampling.

Monte Carlo Option Pricing

- For the pricing of European options on a dividend-paying stock, we may proceed as follows.
- Stock prices $S_1, S_2, S_3, \ldots$ at times $\Delta t, 2\Delta t, 3\Delta t, \ldots$ can be generated via

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

  when $dS/S = \mu\, dt + \sigma\, dW$.
- Non-dividend-paying stock prices in a risk-neutral economy can be generated by setting $\mu = r$.
- Pricing Asian options is easy.

Pricing American-Style Options

- Standard Monte Carlo simulation is inappropriate for American options because of early exercise.
- It is difficult to determine the early-exercise point based on one single path.
- Recent work: Monte Carlo simulation can be modified to price American options with bias.

Variance Reduction: Antithetic Variates

- We are interested in estimating $E[\, g(X_1, X_2, \ldots, X_n)\,]$, where $X_1, X_2, \ldots, X_n$ are independent.
- Let Y_1 and Y_2 be random variables with the same distribution as $g(X_1, X_2, \ldots, X_n)$.
- 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 random numbers on which the first path is based.
- 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 $X_1, X_2, \ldots, X_n$ on the sample path.
- We are interested in $E[\, g(X_1, X_2, \ldots, X_n)\,]$.
- Suppose one simulation run has realizations $\xi_1, \xi_2, \ldots, \xi_n$ for the normally distributed fluctuation term ξ.
- This generates samples $x_1, x_2, \ldots, x_n$.
- The estimate is then $g(\boldsymbol{x})$, where $\boldsymbol{x} \equiv (x_1, x_2 \ldots, x_n)$.

Variance Reduction: Antithetic Variates (concluded)

- We do not sample n more numbers from ξ for the second estimate.
- The antithetic-variates method computes $g(\boldsymbol{x}')$ from the sample path $\boldsymbol{x}' \equiv (x_1', x_2' \ldots, x_n')$ generated by $-\xi_1, -\xi_2, \ldots, -\xi_n$.
- We then output $(g(\boldsymbol{x}) + g(\boldsymbol{x}'))/2$.

Variance Reduction: Conditioning

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

Variance Reduction: Conditioning (concluded)

- As $\text{Var}[\,E[\,X \mid Z\,]\,] \leq \text{Var}[\,X\,]$, $E[\,X \mid Z\,]$ has a smaller variance than observing X directly.
- First obtain a random observation z on Z.
- Then calculate $E[\,X \mid Z = z\,]$ as our estimate.
  - There is no need to resort to simulation in computing $E[\,X \mid Z = z\,]$.
- The procedure can be repeated a few times 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 $\mu \equiv E[\,Y\,]$.
- Then $W \equiv X + \beta(Y - \mu)$ can serve as a "controlled" estimator of $E[\,X\,]$ for any constant β.
  - β can scale the deviation $Y - \mu$ to arrive at an adjustment for X.
  - However β is chosen, W remains an unbiased estimator of $E[\,X\,]$.

Control Variates (continued)

- Note that

$$\text{Var}[\,W\,] = \text{Var}[\,X\,] + \beta^2\,\text{Var}[\,Y\,] + 2\beta\,\text{Cov}[\,X, Y\,], \tag{113}$$

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

$$\beta^2\,\text{Var}[\,Y\,] + 2\beta\,\text{Cov}[\,X, Y\,] < 0. \tag{114}$$

- The success of the scheme clearly depends on both β and the choice of Y.

Control Variates (concluded)

- For example, arithmetic average-rate options can be priced by choosing Y to be the otherwise identical geometric average-rate option's price and $\beta = -1$.
- This approach is much more effective than the antithetic-variates method.
- In general, the choice of Y is ad hoc.

Optimal Choice of β

- Equation (113) is minimized when

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

which was called beta earlier in the book.

- For this specific β,

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

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

- The stronger X and Y are correlated, the greater the reduction in variance.

Optimal Choice of β (continued)

- For example, if this correlation is nearly perfect (± 1), we could control X almost exactly, eliminating practically all of its variance.
- Typically, neither $\text{Var}[\,Y\,]$ nor $\text{Cov}[\,X, Y\,]$ is known.
- Therefore, we cannot obtain the maximum reduction in variance.
- We can guess these values 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 these quantities.

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 $\beta > 0$.

Problems with the Monte Carlo Method

- The error bound is only probabilistic.
- The probabilistic error bound of $\sqrt{N}$ 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.

Quasi-Monte Carlo Methods

- The low-discrepancy sequences (or quasi-random sequences) address the above-mentioned problems.
- It is a deterministic version of the Monte Carlo method in that random samples are replaced by deterministic quasi-random points.
- If a smaller number of samples suffices as a result, efficiency has been gained.
- Aim is to select deterministic points for which the deterministic error bound is smaller than Monte Carlo's probabilistic error bound.

Problems with Quasi-Monte Carlo Methods

- Their theories are valid for integration problems, but may not be directly applicable to simulations because of the correlations between points in a quasi-random sequence.
- This problem may be overcome by writing the desired result as an integral.
- But the integral often has a very high dimension.

Problems with Quasi-Monte Carlo Methods (concluded)

- The improved accuracy is generally lost for problems of high dimension or problems in which the integrand is not smooth.
- No theoretical basis for empirical estimates of their accuracy, a role played by the central limit theorem in the Monte Carlo method.

Assessment

- The results are somewhat mixed.
- The application of such methods in finance seems promising.
- A speed-up as high as 1,000 over the Monte Carlo method, for example, is reported.
- The success of the quasi-Monte Carlo method when compared with traditional variance-reduction techniques is problem dependent.
- For example, the antithetic-variates method outperforms the quasi-Monte Carlo method in bond pricing.