Correlated Trinomial Model[a]

- Two risky assets S_1 and S_2 follow $dS_i/S_i = r\,dt + \sigma_i\,dW_i$ in a risk-neutral economy, $i = 1, 2$.
- Let

$$\begin{aligned} M_i &\equiv e^{r\Delta t}, \\ V_i &\equiv M_i^2(e^{\sigma_i^2 \Delta t} - 1). \end{aligned}$$

  - $S_i M_i$ is the mean of S_i at time Δt.
  - $S_i^2 V_i$ the variance of S_i at time Δt.

[a] Boyle, Evnine, and Gibbs (1989).

Correlated Trinomial Model (continued)

- The value of $S_1 S_2$ at time Δt has a joint lognormal distribution with mean $S_1 S_2 M_1 M_2 e^{\rho\sigma_1\sigma_2\Delta t}$, where ρ is the correlation between dW_1 and dW_2.
- Next match the 1st and 2nd moments of the approximating discrete distribution to those of the continuous counterpart.
- At time Δt from now, there are five distinct outcomes.

Correlated Trinomial Model (continued)

- The five-point probability distribution of the asset prices is (as usual, we impose $u_i d_i = 1$)

Probability	Asset 1	Asset 2
p_1	$S_1 u_1$	$S_2 u_2$
p_2	$S_1 u_1$	$S_2 d_2$
p_3	$S_1 d_1$	$S_2 d_2$
p_4	$S_1 d_1$	$S_2 u_2$
p_5	S_1	S_2

Correlated Trinomial Model (continued)

- The probabilities must sum to one, and the means must be matched:

$$\begin{aligned} 1 &= p_1 + p_2 + p_3 + p_4 + p_5, \\ S_1 M_1 &= (p_1 + p_2)\, S_1 u_1 + p_5 S_1 + (p_3 + p_4)\, S_1 d_1, \\ S_2 M_2 &= (p_1 + p_4)\, S_2 u_2 + p_5 S_2 + (p_2 + p_3)\, S_2 d_2. \end{aligned}$$

Correlated Trinomial Model (concluded)

- Let $R \equiv M_1 M_2 e^{\rho\sigma_1\sigma_2\Delta t}$.
- Match the variances and covariance:

$$\begin{aligned} S_1^2 V_1 &= (p_1+p_2)((S_1u_1)^2-(S_1M_1)^2)+p_5(S_1^2-(S_1M_1)^2) \\ &\quad +(p_3+p_4)((S_1d_1)^2-(S_1M_1)^2), \\ S_2^2 V_2 &= (p_1+p_4)((S_2u_2)^2-(S_2M_2)^2)+p_5(S_2^2-(S_2M_2)^2) \\ &\quad +(p_2+p_3)((S_2d_2)^2-(S_2M_2)^2), \\ S_1S_2R &= (p_1u_1u_2+p_2u_1d_2+p_3d_1d_2+p_4d_1u_2+p_5)\,S_1S_2. \end{aligned}$$

- The solutions are complex (see text).

Correlated Trinomial Model Simplified[a]

- Let $\mu_i' \equiv r - \sigma_i^2/2$ and $u_i \equiv e^{\lambda\sigma_i\sqrt{\Delta t}}$ for $i = 1, 2$.
- The following simpler scheme is good enough:

$$\begin{aligned} p_1 &= \frac{1}{4}\left[\frac{1}{\lambda^2}+\frac{\sqrt{\Delta t}}{\lambda}\left(\frac{\mu_1'}{\sigma_1}+\frac{\mu_2'}{\sigma_2}\right)+\frac{\rho}{\lambda^2}\right], \\ p_2 &= \frac{1}{4}\left[\frac{1}{\lambda^2}+\frac{\sqrt{\Delta t}}{\lambda}\left(\frac{\mu_1'}{\sigma_1}-\frac{\mu_2'}{\sigma_2}\right)-\frac{\rho}{\lambda^2}\right], \\ p_3 &= \frac{1}{4}\left[\frac{1}{\lambda^2}+\frac{\sqrt{\Delta t}}{\lambda}\left(-\frac{\mu_1'}{\sigma_1}-\frac{\mu_2'}{\sigma_2}\right)+\frac{\rho}{\lambda^2}\right], \\ p_4 &= \frac{1}{4}\left[\frac{1}{\lambda^2}+\frac{\sqrt{\Delta t}}{\lambda}\left(-\frac{\mu_1'}{\sigma_1}+\frac{\mu_2'}{\sigma_2}\right)-\frac{\rho}{\lambda^2}\right], \\ p_5 &= 1-\frac{1}{\lambda^2}. \end{aligned}$$

- It cannot price 2-asset 2-barrier options accurately.[b]

[a] Madan, Milne, and Shefrin (1989).
[b] See Chang, Hsu, and Lyuu (2006) for a solution.

Extrapolation

- It is a method to speed up numerical convergence.
- Say $f(n)$ converges to an unknown limit f at rate of $1/n$:

$$f(n) = f + \frac{c}{n} + o\left(\frac{1}{n}\right). \tag{57}$$

- Assume c is an unknown constant independent of n.
  - Convergence is basically monotonic and smooth.

Extrapolation (concluded)

- From two approximations $f(n_1)$ and $f(n_2)$ and by ignoring the smaller terms,

$$\begin{aligned} f(n_1) &= f + \frac{c}{n_1}, \\ f(n_2) &= f + \frac{c}{n_2}. \end{aligned}$$

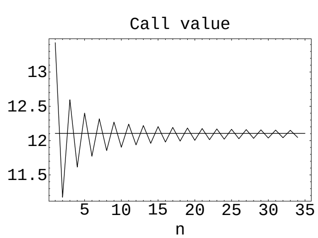
- A better approximation to the desired f is

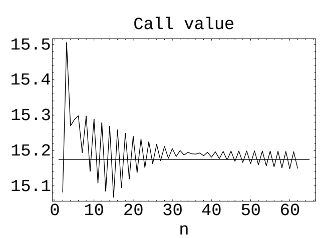
$$f = \frac{n_1 f(n_1) - n_2 f(n_2)}{n_1 - n_2}. \tag{58}$$

- This estimate should converge faster than $1/n$.
- The Richardson extrapolation uses $n_2 = 2n_1$.

Improving BOPM with Extrapolation

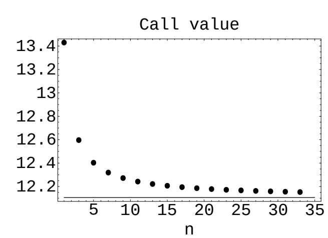
- Consider standard European options.
- Denote the option value under BOPM using n time periods by $f(n)$.
- It is known that BOPM convergences at the rate of $1/n$, consistent with Eq. (57) on p. 535.
- But the plots on p. 242 (redrawn on next page) demonstrate that convergence to the true option value oscillates with n.
- Extrapolation is inapplicable at this stage.

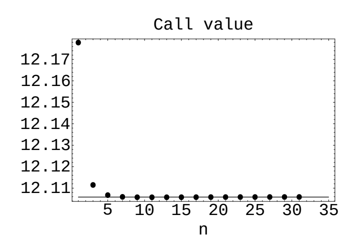




Improving BOPM with Extrapolation (concluded)

- Take the at-the-money option in the left plot on p. 538.
- The sequence with odd n turns out to be monotonic and smooth (see the left plot on p. 540).
- Apply extrapolation (58) on p. 536 with $n_2 = n_1 + 2$, where n_1 is odd.
- Result is shown in the right plot on p. 540.
- The convergence rate is amazing.
- See Exercise 9.3.8 of the textbook (p. 111) for ideas in the general case.



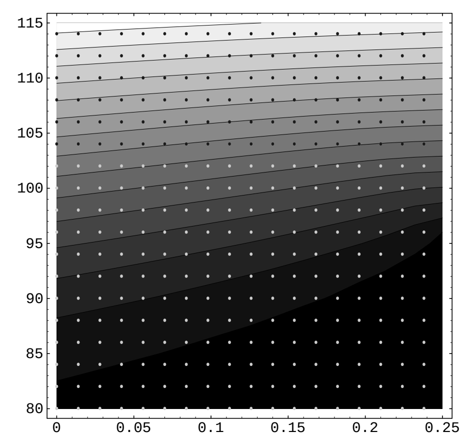


Numerical Methods

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. 544).
- 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

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

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

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

- Given boundary values, we can solve for the x_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

$$\left.\frac{\partial\theta(x,t)}{\partial t}\right|_{t=t_j} = \frac{\theta(x,t_{j+1}) - \theta(x,t_j)}{\Delta t} + \cdots, \tag{59}$$

$$\left.\frac{\partial^2\theta(x,t)}{\partial x^2}\right|_{x=x_i} = \frac{\theta(x_{i+1},t) - 2\theta(x_i,t) + \theta(x_{i-1},t)}{(\Delta x)^2} + \cdots. \tag{60}$$

Explicit Methods (continued)

- To assemble Eqs. (59) and (60) 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. (60), we might as well use x_i for x in Eq. (59).
- Two choices are possible for t in Eq. (60).
- The first choice uses $t = t_j$ to yield the following finite-difference equation,

$$\frac{\theta_{i,j+1} - \theta_{i,j}}{\Delta t} = D\,\frac{\theta_{i+1,j} - 2\theta_{i,j} + \theta_{i-1,j}}{(\Delta x)^2}. \tag{61}$$

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 calculate $\theta_{i,j+1}$ from $\theta_{i,j}, \theta_{i+1,j}, \theta_{i-1,j}$, 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

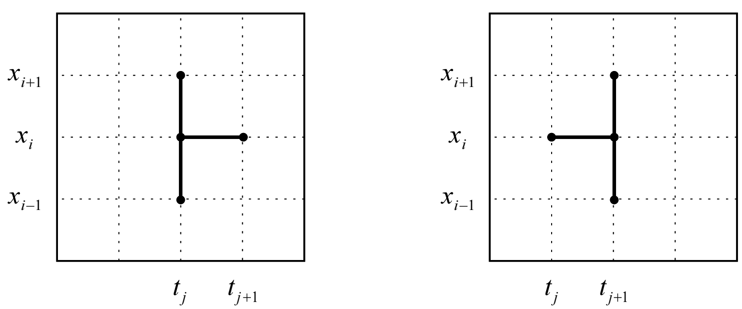
$$\theta_{i,1}, \quad i = 1, 2, \ldots,$$

and then

$$\theta_{i,2}, \quad i = 1, 2, \ldots,$$

and so on.

Stencils



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

Explicit Method and Trinomial Tree

- Rearrange Eq. (61) on p. 548 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. (60) on p. 547 instead, the finite-difference equation becomes

$$\frac{\theta_{i,j+1} - \theta_{i,j}}{\Delta t} = D\,\frac{\theta_{i+1,j+1} - 2\theta_{i,j+1} + \theta_{i-1,j+1}}{(\Delta x)^2}. \tag{62}$$

- 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. 550).

Implicit Methods (continued)

- Equation (62) 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 (61) on p. 548 and implicit method (62) on p. 553:

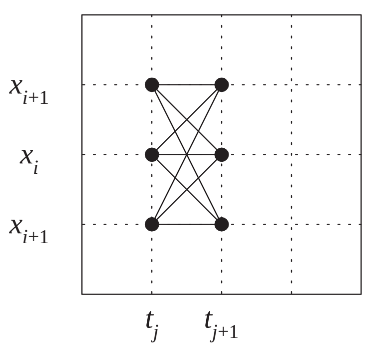
$$\begin{aligned} & \frac{\theta_{i,j+1} - \theta_{i,j}}{\Delta t} \\ = \; & \frac{1}{2} \left(D \, \frac{\theta_{i+1,j} - 2\theta_{i,j} + \theta_{i-1,j}}{(\Delta x)^2} + D \, \frac{\theta_{i+1,j+1} - 2\theta_{i,j+1} + \theta_{i-1,j+1}}{(\Delta x)^2} \right). \end{aligned}$$

- 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

- 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.
- The details of the linear systems are in the text.

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.

[a] A top 10 algorithm according to Dongarra and Sullivan (2000).

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.

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.

[a]This may not be an issue if the derivative only requires discrete sampling along the time dimension.

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

$$S_{i+1} = S_i e^{(\mu - \sigma^2/2)\,\Delta t + \sigma\sqrt{\Delta t}\,\xi}, \quad \xi \sim N(0,1) \tag{63}$$

  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 (see text).

Pricing American 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.
- Monte Carlo simulation can be modified to price American options with small biases (see p. 604).[a]

[a]Longstaff and Schwartz (2001).

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.

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 gammas (for multivariate derivatives).

Variance Reduction: Antithetic Variates

- We are interested in estimating $E[\, g(X_1, X_2, \dots, X_n)\,]$, where $X_1, X_2, \dots, X_n$ are independent.
- Let Y_1 and Y_2 be random variables with the same distribution as $g(X_1, X_2, \dots, X_n)$.
- Then
$$\mathrm{Var}\left[\frac{Y_1 + Y_2}{2}\right] = \frac{\mathrm{Var}[\,Y_1\,]}{2} + \frac{\mathrm{Cov}[\,Y_1, Y_2\,]}{2}.$$
  - $\mathrm{Var}[\,Y_1\,]/2$ is the variance of the Monte Carlo method with two (independent) replications.
- The variance $\mathrm{Var}[\,(Y_1 + Y_2)/2\,]$ is smaller than $\mathrm{Var}[\,Y_1\,]/2$ 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 $X_1, X_2, \dots, X_n$ on the sample path.
- We are interested in $E[\, g(X_1, X_2, \dots, X_n)\,]$.
- 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 estimate is then $g(\boldsymbol{x})$, where $\boldsymbol{x} \equiv (x_1, x_2 \dots, x_n)$.

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 $\boldsymbol{x}' \equiv (x'_1, x'_2 \dots, x'_n)$ generated by $-\xi_1, -\xi_2, \dots, -\xi_n$.
- We then output $(g(\boldsymbol{x}) + g(\boldsymbol{x}'))/2$.
- Repeat the above steps for as many times as required by accuracy.

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.