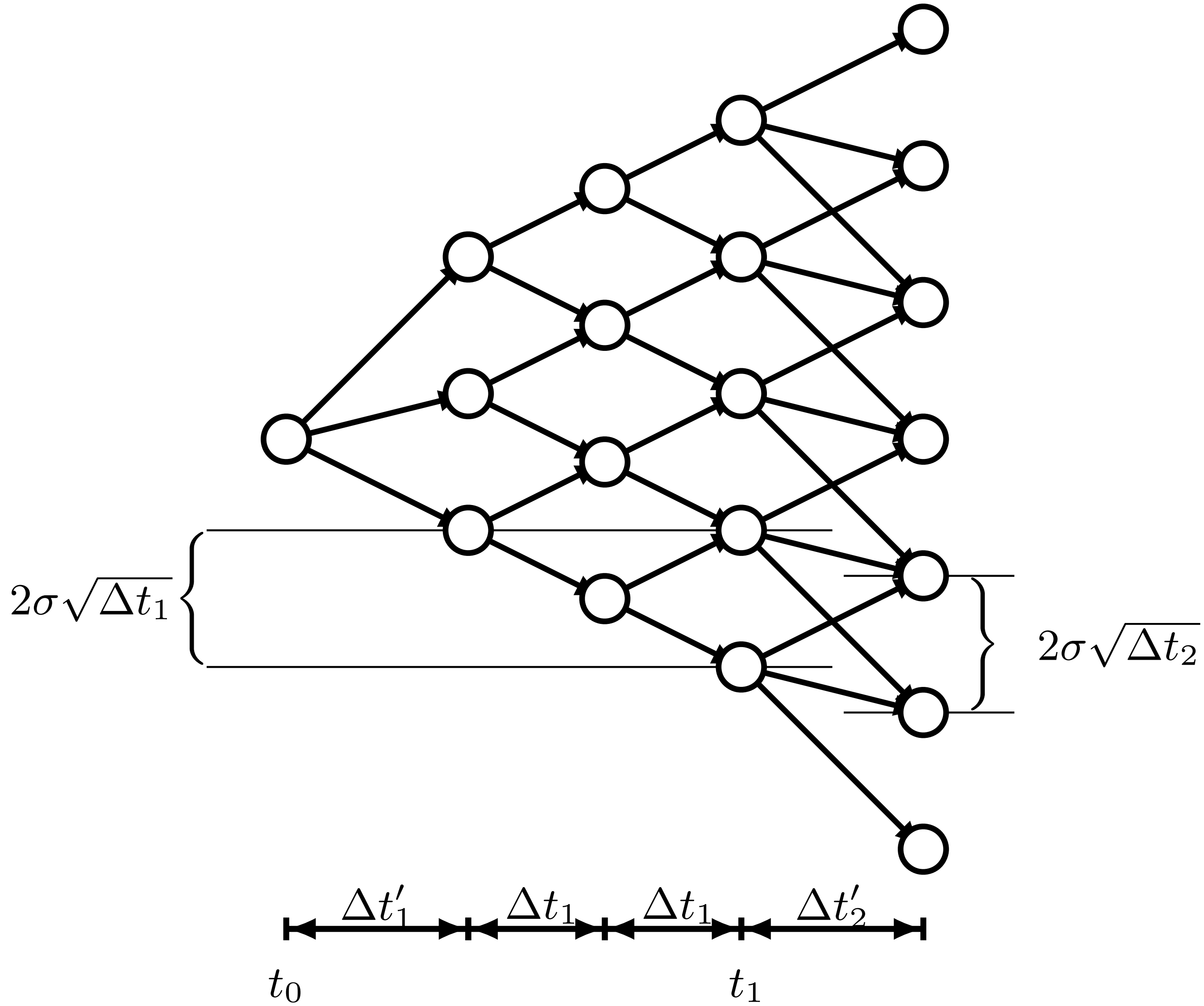
Pricing Discrete Barrier Options

- Barrier options whose barrier is monitored only at discrete times are called discrete barrier options.
- They are more common than the continuously monitored versions.
- The main difficulty with pricing discrete barrier options lies in matching the monitored times.
- Here is why.
- Suppose each period has a duration of Δt and the $\ell > 1$ monitored times are $t_0 = 0, t_1, t_2, \ldots, t_\ell = T$.

Pricing Discrete Barrier Options (continued)

- It is extremely unlikely that *all* monitored times coincide with the end of a period on the tree, meaning Δt divides t_i for all i.
- The binomial-trinomial tree can handle discrete options with ease, however.
- We simply build a binomial-trinomial tree from time 0 to time t_1, followed by one from time t_1 to time t_2, and so on until time t_ℓ.
- See p. 615.



$2\sigma\sqrt{\Delta t_1}$
$2\sigma\sqrt{\Delta t_2}$
$\Delta t_1'$
Δt_1
Δt_1
$\Delta t_2'$
t_0
t_1

Pricing Discrete Barrier Options (concluded)

- This procedure works even if each t_i is associated with a distinct barrier or if each window $[\,t_i, t_{i+1})$ has its own continuously monitored barrier or double barriers.
- If the ith binomial-trinomial tree has n_i periods, the size of the whole tree is

$$O\left(\left(\sum_{i=1}^{\ell} n_i\right)^2\right).$$

Options on a Stock That Pays Known Dividends

- Many ad hoc assumptions have been postulated for option pricing with known dividends.[a]
  1. The one we saw earlier models the stock price minus the present value of the anticipated dividends as following geometric Brownian motion.
  2. One can also model the stock price plus the forward values of the dividends as following geometric Brownian motion.

[a]Frishling (2002).

Options on a Stock That Pays Known Dividends (continued)

- The most realistic model assumes the stock price decreases by the amount of the dividend paid at the ex-dividend date.
- The stock price follows geometric Brownian motion between adjacent ex-dividend dates.
- But this model results in binomial trees that grow exponentially.
- The binomial-trinomial tree can often avoid the exponential explosion for the known-dividends case.

Options on a Stock That Pays Known Dividends (continued)

- Suppose that the known dividend is D dollars and the ex-dividend date is at time t.
- So there are $m \equiv t/\Delta t$ periods between time 0 and the ex-dividend date.
- To avoid negative stock prices, we need to make sure the lowest stock price at time t is at least D, i.e., $Se^{-(t/\Delta t)\sigma\sqrt{\Delta t}} \geq D$.
  - Equivalently,

$$\Delta t \geq \left[\frac{t\sigma}{\ln(S/D)} \right]^2.$$

Options on a Stock That Pays Known Dividends (continued)

- Build a binomial tree from time 0 to time t as before.
- Subtract D from all the stock prices on the tree at time t to represent the price drop on the ex-dividend date.
- Assume the top node's price equals S'.
  - As usual, its two successor nodes will have prices $S'u$ and $S'u^{-1}$.
- The remaining nodes' successor nodes will have prices

$$S'u^{-3}, S'u^{-5}, S'u^{-7}, \ldots ,$$

same as the binomial tree.

Options on a Stock That Pays Known Dividends (concluded)

- For each node at time t below the top node, we build the trinomial connection.
- Note that the binomial-trinomial structure remains valid in the special case when $\Delta t' = \Delta t$ on p. 600.
- Hence the construction can be completed.
- From time $t + \Delta t$ onward, the standard binomial tree will be used until the maturity date or the next ex-dividend date when the procedure can be repeated.
- The resulting tree is called the stair tree.[a]

[a] Dai (R86526008, D8852600) and Lyuu (2004).

Multivariate Contingent Claims

- They depend on two or more underlying assets.
- The basket call on m assets has the terminal payoff
$$\max\left(\sum_{i=1}^{m}\alpha_i S_i(\tau)-X,0\right),$$
where α_i is the percentage of asset i.
- Basket options are essentially options on a portfolio of stocks or index options.
- Option on the best of two risky assets and cash has a terminal payoff of $\max(S_1(\tau), S_2(\tau), X)$.

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 $1/n$:

$$f(n) = f + \frac{c}{n} + o\left(\frac{1}{n}\right). \tag{70}$$

- 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,

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

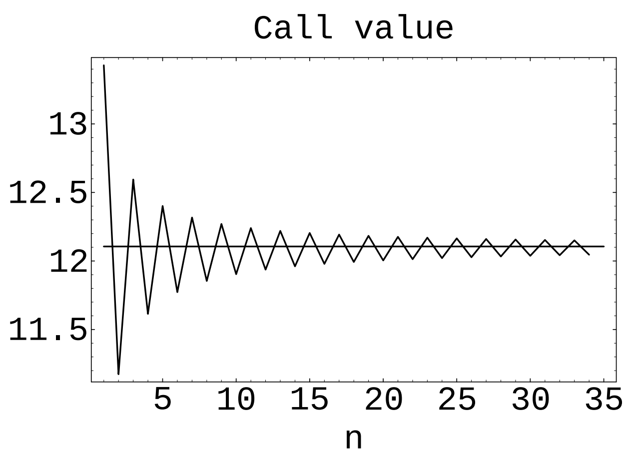
- A better approximation to the desired f is

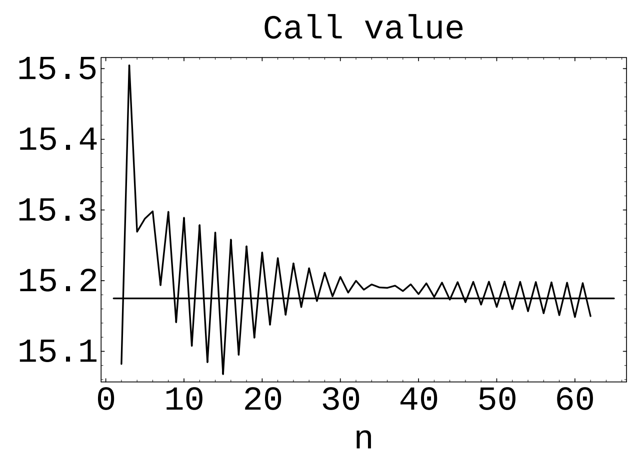
$$f = \frac{n_1 f(n_1) - n_2 f(n_2)}{n_1 - n_2}. \tag{71}$$

- 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. (70) on p. 629.
- But the plots on p. 253 (redrawn on next page) demonstrate that convergence to the true option value oscillates with n.
- Extrapolation is inapplicable at this stage.

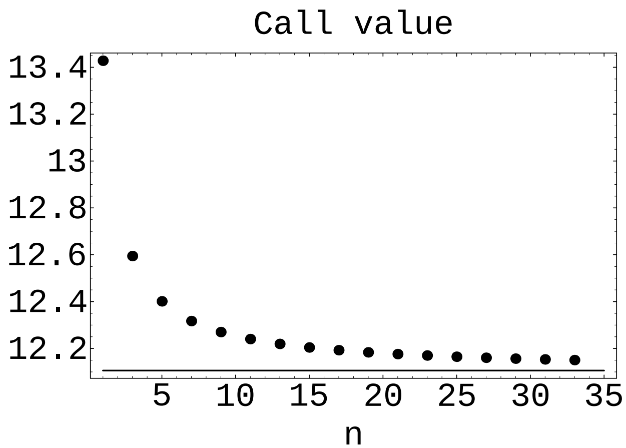

Call value
13
12.5
12
11.5
5
10
15
20
25
30
35
n

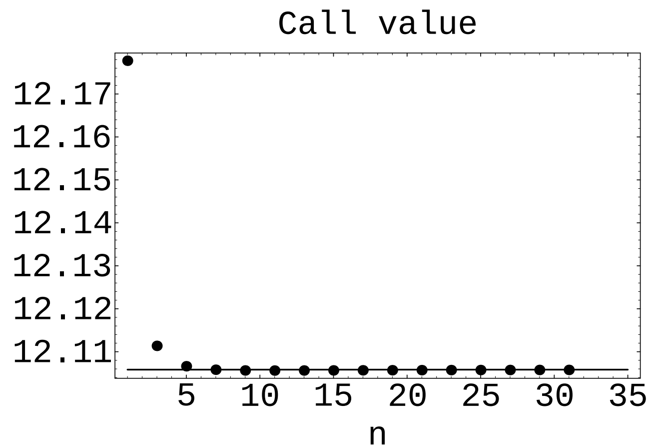

Call value
15.5
15.4
15.3
15.2
15.1
0
10
20
30
40
50
60
n

Improving BOPM with Extrapolation (concluded)

- Take the at-the-money option in the left plot on p. 632.
- The sequence with odd n turns out to be monotonic and smooth (see the left plot on p. 634).[a]
- Apply extrapolation (71) on p. 630 with $n_2 = n_1 + 2$, where n_1 is odd.
- Result is shown in the right plot on p. 634.
- The convergence rate is amazing.
- See Exercise 9.3.8 of the text (p. 111) for ideas in the general case.

[a]This can be proved; see Chang and Palmer (2007).


Call value
13.4
13.2
13
12.8
12.6
12.4
12.2
5 10 15 20 25 30 35
n

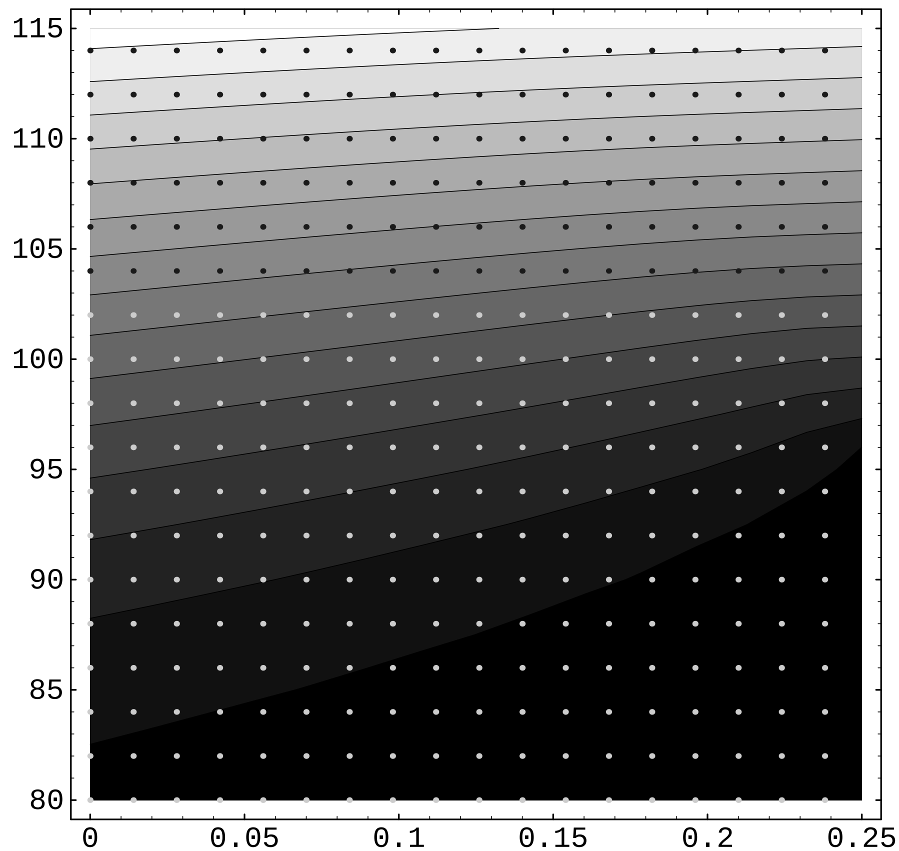

Call value
12.17
12.16
12.15
12.14
12.13
12.12
12.11
5 10 15 20 25 30 35
n

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. 638).
- Solve the equation numerically by introducing difference equations in place of derivatives.


115
110
105
100
95
90
85
80
0
0.05
0.1
0.15
0.2
0.25

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

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

- 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

$$\left.\frac{\partial\theta(x,t)}{\partial t}\right|_{t=t_j} = \frac{\theta(x, t_{j+1}) - \theta(x, t_j)}{\Delta t} + \cdots, \tag{72}$$

$$\left.\frac{\partial^2\theta(x,t)}{\partial x^2}\right|_{x=x_i} = \frac{\theta(x_{i+1}, t) - 2\theta(x_i, t) + \theta(x_{i-1}, t)}{(\Delta x)^2} + \cdots. \tag{73}$$

Explicit Methods (continued)

- Next, assemble Eqs. (72) and (73) 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. (73), we might as well use x_i for x in Eq. (72).
- Two choices are possible for t in Eq. (73).
- The first choice uses $t = t_j$ to yield the following finite-difference equation,

$$\frac{\theta_{i,j+1} - \theta_{i,j}}{\Delta t} = D \, \frac{\theta_{i+1,j} - 2\theta_{i,j} + \theta_{i-1,j}}{(\Delta x)^2}. \tag{74}$$

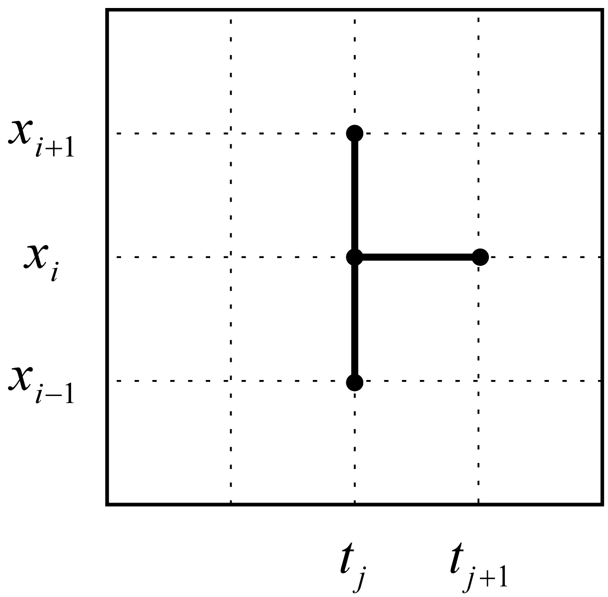
Explicit Methods (continued)

- 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. (74) on p. 642 as

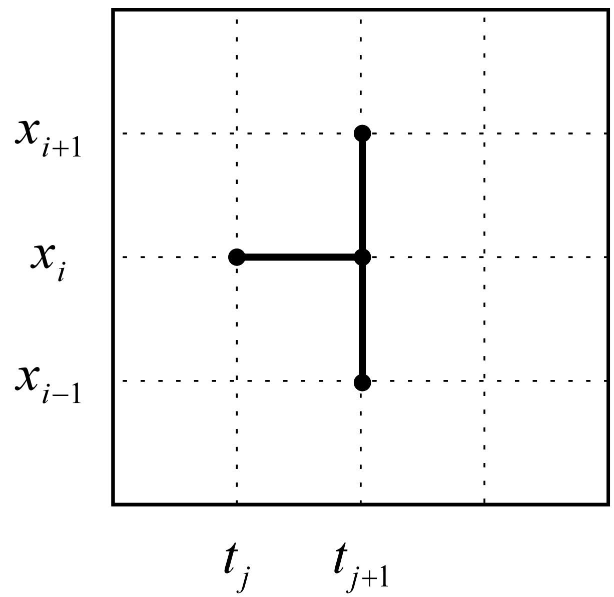
$$\theta_{i,j+1} = \frac{D\Delta t}{(\Delta x)^2}\,\theta_{i+1,j} + \left(1 - \frac{2D\Delta t}{(\Delta x)^2}\right)\theta_{i,j} + \frac{D\Delta t}{(\Delta x)^2}\,\theta_{i-1,j}.$$

- 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, $\theta_{i,0} = \theta(x_i, t_0)$, $i = 1, 2, \ldots$, we calculate

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

- And then

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

- 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 eight times as much.

Explicit Method and Trinomial Tree

- Rearrange Eq. (74) on p. 642 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 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 trinomial trees.

Implicit Methods

- Suppose we use $t = t_{j+1}$ in Eq. (73) on p. 641 instead.
- The finite-difference equation becomes

$$\frac{\theta_{i,j+1} - \theta_{i,j}}{\Delta t} = D \, \frac{\theta_{i+1,j+1} - 2\theta_{i,j+1} + \theta_{i-1,j+1}}{(\Delta x)^2}. \tag{75}$$

- 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:
  - 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. 644.

Implicit Methods (continued)

- Equation (75) 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 (74) on p. 642 and implicit method (75) on p. 648:

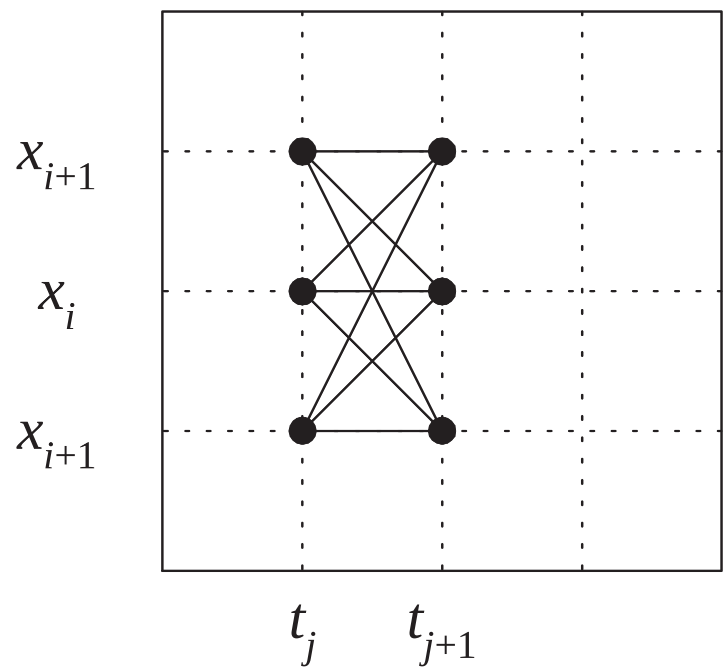
$$\begin{aligned}&\frac{\theta_{i,j+1}-\theta_{i,j}}{\Delta t}\\ =\quad&\frac{1}{2}\left(D\,\frac{\theta_{i+1,j}-2\theta_{i,j}+\theta_{i-1,j}}{(\Delta x)^2}+D\,\frac{\theta_{i+1,j+1}-2\theta_{i,j+1}+\theta_{i-1,j+1}}{(\Delta x)^2}\right).\end{aligned}$$

- 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

- See 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 $(X_1, X_2, \ldots, X_n)$.
- 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 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.

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

$$S_{i+1} = S_i e^{(\mu - \sigma^2/2)\,\Delta t + \sigma \sqrt{\Delta t}\,\xi}, \quad \xi \sim N(0,1) \tag{76}$$

when $dS/S = \mu\, dt + \sigma\, dW$.

Monte Carlo Option Pricing (continued)

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

$$S_{i+1} = S_i + (\mu - \sigma^2/2) \, \Delta t + \sigma \sqrt{\Delta t} \; \xi.$$

- But this is locally normally distributed, not lognormally, hence biased.[a]
- In practice, this is not expected to be a major problem as long as Δt is sufficiently small.

---

[a] Contributed by Mr. Tai, Hui-Chin (`R97723028`) on April 22, 2009.

Monte Carlo Option Pricing (concluded)

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

$C := 0$;
for $i = 1, 2, 3, \ldots, m$ **do**
  $P := S \times e^{(r-\sigma^2/2)\,T + \sigma\sqrt{T}\,\xi}$;
  $C := C + \max(P - X, 0)$;
end for
`return` Ce^{-rT}/m;

- Pricing Asian options is easy (see text).

How about American Options?

- Standard Monte Carlo simulation is inappropriate for American options because of early exercise (why?).
- It is difficult to determine the early-exercise point based on one single path.
- But Monte Carlo simulation can be modified to price American options with small biases (p. 709ff).[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).

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 gammas $\partial^2 P(S_1, S_2, \dots)/(\partial S_1 \partial S_2)$ is:
$$e^{-r\tau}\, E\left[\, \frac{P(S_1+\epsilon_1, S_2+\epsilon_2) - P(S_1-\epsilon_1, S_2+\epsilon_2)}{4\epsilon_1\epsilon_2} \right.$$
$$\left. \frac{-P(S_1+\epsilon_1, S_2-\epsilon_2) + P(S_1-\epsilon_1, S_2-\epsilon_2)}{} \,\right].$$

Gamma (concluded)

- 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.
- Need formulas for Greeks which are integrals (thus avoiding ϵ, finite differences, and resimulation).[a]

[a] Lyuu and Teng (`R91723054`) (2008).

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 $t_1, t_2, \ldots, t_n$.
- A sample path $S(t_0), S(t_1), \ldots, S(t_n)$ is produced.
  - Here, $t_0 = 0$ is the current time, and $t_n = T$ is the expiration time of the option.

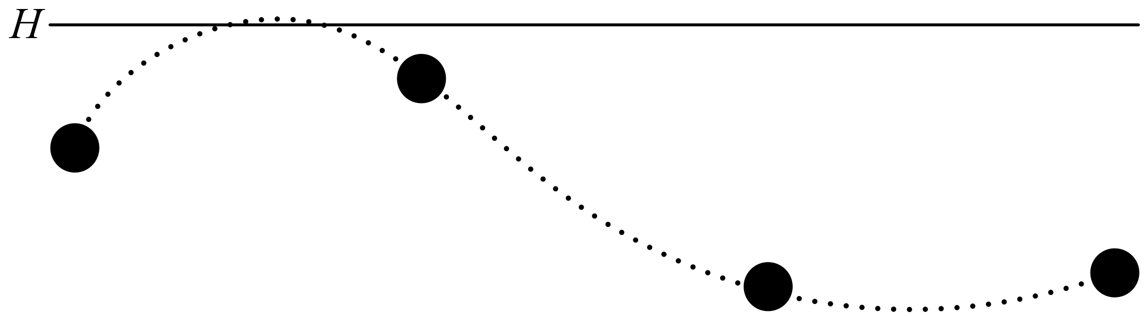
Biases in Pricing Continuously Monitored Options with Monte Carlo (continued)

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

$C := 0;$
for $i = 1, 2, 3, \ldots, m$ **do**
    $P := S;$ hit $:= 0;$
    for $j = 1, 2, 3, \ldots, n$ **do**
        $P := P \times e^{(r-\sigma^2/2)\,(T/n)+\sigma\sqrt{(T/n)}\;\xi};$
        if $P \geq H$ **then**
            hit $:= 1;$
            `break`;
        end if
    end for
    if hit $= 0$ **then**
        $C := C + \max(P - X, 0);$
    end if
end for
`return` $Ce^{-rT}/m;$

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

- This estimate is biased.
  - 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).



H

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

- The bias can certainly be lowered by increasing the number of observations along the sample path.
- 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 efficiently.
- So 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.
- This methodology is called the Brownian bridge approach.
- Formally, we have

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

Brownian Bridge Approach to Pricing Barrier Options (continued)

- As a barrier is hit over a time interval if and only if the maximum stock price over that period is at least H,

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

- 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 19 *Assume S follows $dS/S = \mu\, dt + \sigma\, dW$ and define*

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

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

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

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

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

Brownian Bridge Approach to Pricing Barrier Options (continued)

- Lemma 19 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 call, choose $n = 1$.
- As a result,

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

Brownian Bridge Approach to Pricing Barrier Options (continued)

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

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 $(t_i, t_{i+1}]$, $0 \leq i < n$.
- This option thus contains n barriers.
- It is a simple matter of multiplying the probabilities for the n time intervals properly to obtain the desired probability adjustment term.

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.

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)

- The antithetic-variates method does not sample n more numbers from ξ for the second estimate $g(\boldsymbol{x}')$.
- Instead, generate the sample path $\boldsymbol{x}' \equiv (x'_1, x'_2 \ldots, x'_n)$ from $-\xi_1, -\xi_2, \ldots, -\xi_n$.
- Compute $g(\boldsymbol{x}')$.
- 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.

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.
  - 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

$$\mathrm{Var}[\,W\,] = \mathrm{Var}[\,X\,] + \beta^2\,\mathrm{Var}[\,Y\,] + 2\beta\,\mathrm{Cov}[\,X, Y\,], \tag{77}$$

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

$$\beta^2\,\mathrm{Var}[\,Y\,] + 2\beta\,\mathrm{Cov}[\,X, Y\,] < 0. \tag{78}$$

Control Variates (concluded)

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

Choice of Y

- In general, the choice of Y is ad hoc, and experiments must be performed to confirm the wisdom of the choice.
- Try to match calls with calls and puts with puts.[a]
- On many occasions, Y is a discretized version of the derivative that gives μ.
  - Discretely monitored geometric average-rate option vs. the continuously monitored geometric average-rate option given by formulas (30) on p. 344.
- For some choices, the discrepancy can be significant, such as the lookback option.[b]

[a]Contributed by Ms. Teng, Huei-Wen (`R91723054`) on May 25, 2004.
[b]Contributed by Mr. Tsai, Hwai (`R92723049`) on May 12, 2004.

Optimal Choice of β

- Equation (77) on p. 689 is minimized when

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

  which was called beta in the book.

- For this specific β,

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

  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.
- Typically, neither $\mathrm{Var}[\,Y\,]$ nor $\mathrm{Cov}[\,X, Y\,]$ 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.
  - 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 $\beta > 0$.

A Pitfall

- A potential pitfall is to sample X and Y independently.
- In this case, $\mathrm{Cov}[\,X, Y\,] = 0$.
- Equation (77) on p. 689 becomes

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

- So whatever Y is, the variance is *increased*!

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.