Extrapolation

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

\[
  f(n) = f + \frac{c}{n} + o \left( \frac{1}{n} \right). \tag{66}
\]
- Assume that \( c \) is an unknown constant independent of \( n \).
  - Convergence is basically monotonic and smooth.

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 converges at the rate of \( 1/n \), consistent with Eq. (66) on p. 571.
- But the plots on p. 256 (redrawn on next page) demonstrate that convergence to the true option value oscillates with \( n \).
- Extrapolation is inapplicable at this stage.

Extrapolation (concluded)

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

\[
  f(n_1) = f + \frac{c}{n_1},
\]

\[
  f(n_2) = f + \frac{c}{n_2}.
\]
- A better approximation to the desired \( f \) is

\[
  f = \frac{n_1 f(n_1) - n_2 f(n_2)}{n_1 - n_2}. \tag{67}
\]
- This estimate should converge faster than \( 1/n \).
- The Richardson extrapolation uses \( n_2 = 2n_1 \).
Improving BOPM with Extrapolation (concluded)

- Take the at-the-money option in the left plot on p. 574.
- The sequence with odd $n$ turns out to be monotonic and smooth (see the left plot on p. 576).
- Apply extrapolation (67) on p. 572 with $n_2 = n_1 + 2$, where $n_1$ is odd.
- Result is shown in the right plot on p. 576.
- The convergence rate is amazing.
- See Exercise 9.3.8 of the textbook (p. 111) for ideas in the general case.

**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. 579).
- 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 $\Delta x$ along the $x$ axis and $\Delta y$ 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 $x_i$s and the $y_j$s within the square $[-L, L]$.\[825\times \]

- 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 $\Delta x$ and $\Delta t$, 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

$$\frac{\partial \theta(x, t)}{\partial t} \bigg|_{t=t_j} = \theta(x, t_{j+1}) - \theta(x, t_j) \quad \frac{\Delta}{\Delta t},$$

$$\frac{\partial^2 \theta(x, t)}{\partial x^2} \bigg|_{x=x_i} = \theta(x_{i+1}, t) - 2\theta(x_i, t) + \theta(x_{i-1}, t) \quad \frac{(\Delta x)^2}{(\Delta x)^2}.$$
Explicit Methods (continued)

- To assemble Eqs. (68) and (69) 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. (69), we might as well use \(x_i\) for \(x\) in Eq. (68).
- Two choices are possible for \(t\) in Eq. (69).
- The first choice uses \(t = t_j\) to yield the following finite-difference equation,

\[
\frac{\theta_{i,j+1} - \theta_{i,j}}{\Delta t} = D \frac{\theta_{i+1,j} - 2\theta_{i,j} + \theta_{i-1,j}}{(\Delta x)^2}, \quad (70)
\]

Stencils

(a) \[X_j, X_{j+1}, \quad t_j, t_{j+1}\]

(b) \[X_i, X_{i+1}, \quad t_j, t_{j+1}\]

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.

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 \((\Delta x)^2\) would imply quadrupling \((\Delta t)^2\), resulting in a running time eight times as much.
Implicit Methods

- If we use \( t = t_{j+1} \) in Eq. (69) instead, the finite-difference equation becomes

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

(71)

- 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. 585).
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 (70) on p. 583 and implicit method (71) on p. 588:
  \[
  \frac{\theta_i,j+1 - \theta_i,j}{\Delta t} = \frac{1}{2} \left( D \frac{\theta_{i+1,j} - 2\theta_i,j + \theta_{i-1,j}}{(dx)^2} + D \frac{\theta_{i,j+1} - 2\theta_i,j + \theta_{i,j-1}}{(dx)^2} \right).
  \]
- After rearrangement,
  \[
  \gamma \theta_i,j+1 + \frac{\theta_{i+1,j} - 2\theta_i,j + \theta_{i-1,j}}{2} = \gamma \theta_i,j + \frac{\theta_{i,j+1} - 2\theta_i,j + \theta_{i,j-1}}{2}.
  \]
- This is an unconditionally stable implicit method with excellent rates of convergence.

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.
- 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.

\(^a\)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
  \[
  (x_1^{(i)}, x_2^{(i)}, \ldots, x_n^{(i)}), \quad 1 \leq i \leq N
  \]
  independently with the same joint distribution as
  $(X_1, X_2, \ldots, X_n)$ and set
  \[
  Y_i \equiv g(x_1^{(i)}, x_2^{(i)}, \ldots, x_n^{(i)}).
  \]

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 $\text{Prob}[X \leq x] = (x - a)/(b - a)$ for $a \leq x \leq b$.

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, $\bar{Y}$,
  satisfies $E[\bar{Y}] = \theta$, it can be used to estimate $\theta$.
- 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 (concluded)

- Note that
  \[
  E[Y] = (b-a) \cdot E[g(X)]
  = (b-a) \int_a^b \frac{g(x)}{b-a} \, dx
  = \int_a^b g(x) \, dx.
  \]
- 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 (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.

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 by the Berry-Esseen theorem,
- 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 $n$.

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
  \[ S_{t+1} = S_t e^{(\mu - \sigma^2/2) \Delta t + \sigma \sqrt{\Delta t} \xi}, \quad \xi \sim N(0, 1) \]
  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 (see text).

Delta and Common Random Numbers

- In estimating delta, it is natural to start with the finite-difference estimate
  \[ e^{r\tau} \frac{E[P(S + \epsilon)] - E[P(S - \epsilon)]}{2\epsilon}, \]
  where $P(x)$ 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.

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.

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 $P(S - \epsilon)$.
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 $dX = a_1 dt + b \sqrt{dt} \xi$.
- 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 $\xi$.
- This generates samples $x_1, x_2, \ldots, x_n$.
- The estimate is then $g(x)$, where $x \equiv (x_1, x_2, \ldots, x_n)$.

Variance Reduction: Antithetic Variates (concluded)

- We do not sample $n$ more numbers from $\xi$ for the second estimate.
- The antithetic-variates method computes $g(x')$ from the sample path $x' \equiv (x'_1, x'_2, \ldots, x'_n)$ generated by $-\xi_1, -\xi_2, \ldots, -\xi_n$.
- We then output $(g(x) + g(x'))/2$.
- 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 $E[X | Z = z]$ 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 $E[X]$.

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 $W = X + \beta(Y - \mu)$ can serve as a “controlled” estimator of $E[X]$ for any constant $\beta$.
  - $\beta$ can scale the deviation $Y - \mu$ to arrive at an adjustment for $X$.
  - However $\beta$ is chosen, $W$ remains an unbiased estimator of $E[X]$ (why?).

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 can be repeated a few times to reduce the variance.

Control Variates (continued)

- Note that
  \[
  \text{Var}[W] = \text{Var}[X] + \beta^2 \text{Var}[Y] + 2\beta \text{Cov}[X, Y],
  \] (72)
- Hence $W$ is less variable than $X$ if and only if
  \[
  \beta^2 \text{Var}[Y] + 2\beta \text{Cov}[X, Y] < 0,
  \] (73)
- The success of the scheme clearly depends on both $\beta$ 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.

Optimal Choice of $\beta$

- Equation (72) on p. 614 is minimized when
  \[ \beta = -\frac{\text{Cov}[X,Y]}{\text{Var}[Y]}, \]
  which was called beta earlier in the book.
- For this specific $\beta$,
  \[ \text{Var}[W] = \text{Var}[X] - \frac{\text{Cov}[X,Y]^2}{\text{Var}[Y]} = (1 - \rho^2_{X,Y}) \text{Var}[X], \]
  where $\rho_{X,Y}$ is the correlation between $X$ and $Y$.
- The stronger $X$ and $Y$ are correlated, the greater the reduction in variance.

Choice of $Y$

- In general, the choice of $Y$ is ad hoc, and experiments must be performed to confirm the wisdom of the choice.
- On many occasions, $Y$ is a discretized version of the derivative that gives $\mu$.
  - Discretely monitored geometric average-rate option vs. the continuously monitored geometric average-rate option given by formulas (31) on p. 337.
- For some choices, the discrepancy can be significant, such as the lookback option.\(^a\)

\(^a\)Contributed by Mr. Tsai, Hwai (r92723049) on May 12, 2004.

Optimal Choice of $\beta$ (continued)

- For example, if this correlation is nearly perfect ($\pm 1$), 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 $\beta$ (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 $\beta > 0$.

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 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.

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.