Simulation Methods

Leonid Kogan

MIT, Sloan

15.450, Fall 2010

Outline

Overview

- Simulation methods (Monte Carlo) can be used for option pricing, risk management, econometrics, etc.
- Naive Monte Carlo may be too slow in some practical situations. Many special techniques for variance reduction: antithetic variables, control variates, stratified sampling, importance sampling, etc.
- Recent developments: Quasi-Monte Carlo (low discrepancy sequences).

The Basic Problem

- Consider the basic problem of computing an expectation

$$\theta = \mathrm{E}[f(X)], \quad X \sim pdf(X)$$

- Monte Carlo simulation approach specifies generating N independent draws from the distribution $pdf(X)$, $X_1, X_2, ..., X_N$, and approximating

$$\mathrm{E}[f(X)] \approx \widehat{\theta}_N \equiv \frac{1}{N} \sum_{i=1}^{N} f(X_i)$$

- By Law of Large Numbers, the approximation $\widehat{\theta}_N$ converges to the true value as N increases to infinity.
- Monte Carlo estimate $\widehat{\theta}_N$ is unbiased:

$$\mathrm{E}\left[\widehat{\theta}_N\right] = \theta$$

- By Central Limit Theorem,

$$\sqrt{N}\frac{\widehat{\theta}_N - \theta}{\sigma} \Rightarrow \mathcal{N}(0,1), \quad \sigma^2 = \mathrm{Var}[f(X)]$$

Outline

Generating Random Numbers

- Pseudo random number generators produce deterministic sequences of numbers that appear stochastic, and match closely the desired probability distribution.
- For some standard distributions, e.g., uniform and Normal, MATLAB® provides built-in random number generators.
- Sometimes it is necessary to simulate from other distributions, not covered by the standard software. Then apply one of the basic methods for generating random variables from a specified distribution.

The Inverse Transform Method

- Consider a random variable X with a continuous, strictly increasing CDF function $F(x)$.
- We can simulate X according to

$$X = F^{-1}(U), \quad U \sim Unif[0,1]$$

- This works, because

$$\text{Prob}(X \leqslant x) = \text{Prob}(F^{-1}(U) \leqslant x) = \text{Prob}(U \leqslant F(x)) = F(x)$$

- If $F(x)$ has jumps, or flat sections, generalize the above rule to

$$X = \min\left(x : F(x) \geqslant U\right)$$

The Inverse Transform Method

Example: Exponential Distribution

- Consider an exponentially-distributed random variable, characterized by a CDF
$$F(x) = 1 - e^{-x/\theta}$$
- Exponential distributions often arise in credit models.
- Compute $F^{-1}(u)$

$$u = 1 - e^{-x/\theta} \quad \Rightarrow \quad X = -\theta \ln(1-U) \sim -\theta \ln U$$

The Inverse Transform Method

Example: Discrete Distribution

- Consider a discrete random variable X with values

$$c_1 < c_2 < \cdots < c_n, \quad \text{Prob}(X = c_i) = p_i$$

- Define cumulative probabilities

$$F(c_i) = q_i = \sum_{j=1}^{i} p_j$$

- Can simulate X as follows:
  1. Generate $U \sim Unif[0, 1]$.
  2. Find $K \in \{1, ..., n\}$ such that $q_{K-1} \leqslant U \leqslant q_K$.
  3. Set $X = c_K$.

The Acceptance-Rejection Method

- Generate samples with probability density $f(x)$.
- The acceptance-rejection (A-R) method can be used for multivariate problems as well.
- Suppose we know how to generate samples from the distribution with pdf $g(x)$, s.t.,

$$f(x) \leqslant cg(x), \quad c > 1$$

- Follow the algorithm
  1. Generate X from the distribution $g(x)$;
  2. Generate U from $Unif[0,1]$;
  3. If $U \leqslant f(X)/[cg(X)]$, return X; otherwise go to Step 1.
- Probability of acceptance on each attempt is $1/c$. Want c close to 1.
- See the Appendix for derivations.

The Acceptance-Rejection Method

Example: Beta Distribution

- The beta density is

$$f(x) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1}(1-x)^{\beta-1}, \quad 0 \leqslant x \leqslant 1$$

- Assume $\alpha, \beta \geqslant 1$. Then $f(x)$ has a maximum at $(\alpha - 1)/(\alpha + \beta - 2)$.
- Define

$$c = f\left(\frac{\alpha - 1}{\alpha + \beta - 2}\right)$$

  and choose $g(x) = 1$.
- The A-R method becomes
  1. Generate independent U_1 and U_2 from $Unif[0,1]$ until $cU_2 \leqslant f(U_1)$;
  2. Return U_1.

The Acceptance-Rejection Method

Example: Beta Distribution

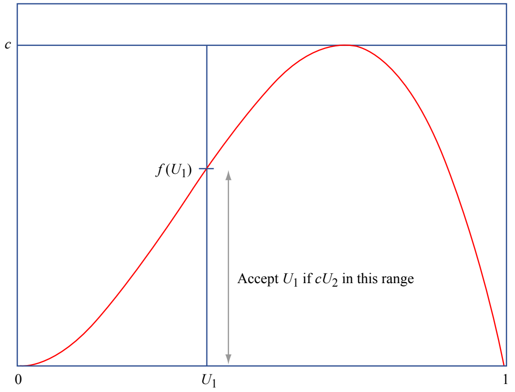


Illustration of the acceptance-rejection method using uniformly distributed candidates.

Image by MIT OpenCourseWare.

Source: Glasserman 2004, Figure 2.8

Outline

Variance reduction

- Suppose we have simulated N independent draws from the distribution $f(x)$. How accurate is our estimate of the expected value $\mathrm{E}[f(X)]$?
- Using the CLT, construct the $100(1-\alpha)\%$ confidence interval

$$\left[\widehat{\theta}_N - \frac{\widehat{\sigma}}{\sqrt{N}} z_{1-\alpha/2}, \widehat{\theta}_N + \frac{\widehat{\sigma}}{\sqrt{N}} z_{1-\alpha/2}\right],$$

$$\widehat{\sigma}^2 = \frac{1}{N} \sum_{i=1}^{N} \left(f(X_i) - \widehat{\theta}_N \right)^2$$

where $z_{1-\alpha/2}$ is the $(1-\alpha/2)$ percentile of the standard Normal distribution.
- For a fixed number of simulations N, the length of the interval is proportional to $\widehat{\sigma}$.
- The number of simulations required to achieve desired accuracy is proportional to the standard deviation of $f(X_i)$, $\widehat{\sigma}$.
- The idea of variance reduction: replace the original problem with another simulation problem, with the same answer but smaller variance!

Antithetic Variates

- Attempt to reduce variance by introducing negative dependence between pairs of replications.
- Suppose want to estimate

$$\theta = \mathrm{E}[f(X)], \quad pdf(X) = pdf(-X)$$

- Note that

$$-X \sim pdf(X) \;\Rightarrow\; \mathrm{E}\left[\frac{f(X)+f(-X)}{2}\right] = \mathrm{E}[f(X)]$$

- Define $Y_i = [f(X_i) + f(-X_i)]/2$ and compute

$$\widehat{\theta}_N^{AV} = \frac{1}{N}\sum_{i=1}^{N} Y_i$$

- Note that Y_i are IID, and by CLT,

$$\sqrt{N}\frac{\widehat{\theta}_N^{AV} - \mathrm{E}[f(X)]}{\sigma_{AV}} \Rightarrow \mathcal{N}(0,1), \quad \sigma_{AV} = \sqrt{\mathrm{Var}[Y_i]}$$

Antithetic Variates

When are they useful?

- Assume that the computational cost of computing Y_i is roughly twice that of computing $f(X_i)$.
- Antithetic variates are useful if

$$\mathrm{Var}[\widehat{\theta}_N^{AV}] < \mathrm{Var}\left[\frac{1}{2N}\sum_{i=1}^{2N} f(X_i)\right]$$

using the IID property of Y_i, as well as X_i, the above condition is equivalent to

$$\begin{aligned}
\mathrm{Var}[Y_i] &< \frac{1}{2}\mathrm{Var}[f(X_i)] \\
4\mathrm{Var}[Y_i] &= \mathrm{Var}\left[f(X_i) + f(-X_i)\right] = \\
&\mathrm{Var}[f(X_i)] + \mathrm{Var}[f(-X_i)] + 2\mathrm{Cov}[f(X_i), f(-X_i)] = \\
&2\mathrm{Var}[f(X_i)] + 2\mathrm{Cov}[f(X_i), f(-X_i)]
\end{aligned}$$

- Antithetic variates reduce variance if

$$\mathrm{Cov}[f(X_i), f(-X_i)] < 0$$

Antithetic Variates

When do they work best?

- Suppose that f is a monotonically increasing function. Then

$$\mathrm{Cov}[f(X), f(-X)] < 0$$

and the antithetic variates reduce simulation variance. By how much?

- Define

$$f_0(X) = \frac{f(X) + f(-X)}{2}, \quad f_1(X) = \frac{f(X) - f(-X)}{2}$$

- $f_0(X)$ and $f_1(X)$ are uncorrelated:

$$\mathrm{E}[f_0(X) f_1(X)] = \frac{1}{4}\mathrm{E}[f^2(X) - f^2(-X)] = 0 = \mathrm{E}[f_0(X)]\mathrm{E}[f_1(X)]$$

- Conclude that

$$\mathrm{Var}[f(X)] = \mathrm{Var}[f_0(X)] + \mathrm{Var}[f_1(X)]$$

- If $f(X)$ is linear, $\mathrm{Var}[f_0(X)] = 0$, and antithetic variates eliminate all variance!
- Antithetics are more effective when $f(X)$ is close to linear.

Control Variates

- The idea behind the control variates approach is to decompose the unknown expectation $\mathrm{E}[Y]$ into the part known in closed form, and the part that needs to be estimated by simulation.
- There is no need to use simulation for the part known explicitly. The variance of the remainder may be much smaller than the variance of Y.

Control Variates

- Suppose we want to estimate the expected value $\mathrm{E}[Y]$.
- On each replication, generate another variable, X_i. Thus, draw a sequence of pairs (X_i, Y_i).
- Assume that $\mathrm{E}[X]$ is known. How can we use this information to reduce the variance of our estimate of $\mathrm{E}[Y]$?
- Define

$$Y_i(b) = Y_i - b(X_i - \mathrm{E}[X])$$

- Note that $\mathrm{E}[Y_i(b)] = \mathrm{E}[Y_i]$, so $\frac{1}{N}\sum_{i=1}^{N} Y_i(b)$ is an unbiased estimator of $\mathrm{E}[Y]$.
- Can choose b to minimize variance of $Y_i(b)$:

$$\mathrm{Var}[Y_i(b)] = \mathrm{Var}[Y] - 2b\,\mathrm{Cov}[X, Y] + b^2\,\mathrm{Var}[X]$$

- Optimal choice $b^\star$ is the OLS coefficient in regression of Y on X:

$$b^\star = \frac{\mathrm{Cov}[X, Y]}{\mathrm{Var}[X]}$$

Control Variates

- The higher the R^2 in the regression of Y on X, the larger the variance reduction.
- Denoting the correlation between X and Y by ρ_{XY}, find

$$\frac{\mathrm{Var}\left[\frac{1}{N}\sum_{i=1}^{N} Y_i(b^\star)\right]}{\mathrm{Var}\left[\frac{1}{N}\sum_{i=1}^{N} Y_i\right]} = 1 - \rho_{XY}^2$$

- In practice, $b^\star$ is not known, but is easy to estimate using OLS.
- Two-stage approach:
  - Simulate N_0 pairs of (X_i, Y_i) and use them to estimate $\widehat{b}^\star$.
  - Simulate N more pairs and estimate $\mathrm{E}[Y]$ as

$$\frac{1}{N}\sum_{i=1}^{N} Y_i - \widehat{b}^\star(X_i - \mathrm{E}[X])$$

Control Variates

Example: Pricing a European Call Option

- Suppose we want to price a European call option using simulation.
- Assume constant interest rate r. Under the risk-neutral probability $\mathbf{Q}$, we need to evaluate

$$\mathrm{E}_0^{\mathbf{Q}}\left[e^{-rT}\max(0, S_T - K)\right]$$

- The stock price itself is a natural control variate. Assuming no dividends,

$$\mathrm{E}_0^{\mathbf{Q}}\left[e^{-rT}S_T\right] = S_0$$

- Consider the Black-Scholes setting with

$$r = 0.05,\ \sigma = 0.3,\ S_0 = 50,\ T = 0.25$$

- Evaluate correlation between the option payoff and the stock price for different values of K. $\widehat{\rho}^2$ is the percentage of variance eliminated by the control variate.

Control Variates

Example: Pricing a European Call Option

K	40	45	50	55	60	65	70
$\widehat{\rho}^2$	0.99	0.94	0.80	0.59	0.36	0.19	0.08

Source: Glasserman 2004, Table 4.1

- For in-the-money call options, option payoff is highly correlated with the stock price, and significant variance reduction is possible.
- For out-of-the-money call options, correlation of option payoff with the stock price is low, and variance reduction is very modest.

Control Variates

Example: Pricing an Asian Call Option

- Suppose we want to price an Asian call option with the payoff

$$\max(0, \overline{S}_T - K), \quad \overline{S}_T \equiv \frac{1}{J}\sum_{j=1}^{J} S(t_j), \quad t_1 < t_2 < \cdots < t_J \leqslant T$$

- A natural control variate is the discounted payoff of the European call option:

$$X = e^{-rT}\max(0, S_T - K)$$

- Expectation of the control variate under **Q** is given by the Black-Scholes formula.
- Note that we may use multiple controls, e.g., option payoffs at multiple dates.
- When pricing look-back options, barrier options by simulation can use similar ideas.

Control Variates

Example: Stochastic Volatility

- Suppose we want to price a European call option in a model with stochastic volatility.
- Consider a discrete-time setting, with the stock price following

$$S(t_{i+1}) = S(t_i) \exp\left((r - \sigma(t_i)^2/2)(t_{i+1} - t_i) + \sigma(t_i)\sqrt{t_{i+1} - t_i}\,\varepsilon^{\mathbf{Q}}_{i+1} \right)$$

$$\varepsilon^{\mathbf{Q}}_i \overset{\text{IID}}{\sim} \mathcal{N}(0,1)$$

- $\sigma(t_i)$ follows its own stochastic process.
- Along with $S(t_i)$, simulate another stock price process

$$S(t_{i+1}) = S(t_i) \exp\left(\left(r - \frac{\widetilde{\sigma}^2}{2} \right)(t_{i+1} - t_i) + \widetilde{\sigma}\sqrt{t_{i+1} - t_i}\,\varepsilon^{\mathbf{Q}}_{i+1} \right)$$

- Pick $\widetilde{\sigma}$ close to a typical value of $\sigma(t_i)$.
- Use the same sequence of Normal variables $\varepsilon^{\mathbf{Q}}_i$ for $\widetilde{S}(t_i)$ as for $S(t_i)$.
- Can use the discounted payoff of the European call option on $\widetilde{S}$ as a control variate: expectation given by the Black-Scholes formula.

Control Variates

Example: Hedges as Control Variates

- Suppose, again, that we want to price the European call option on a stock with stochastic volatility.
- Let $C(t, S_t)$ denote the price of a European call option with some constant volatility $\widetilde{\sigma}$, given by the Black-Scholes formula.
- Construct the process for discounted gains from a discretely-rebalanced delta-hedge.
    - The delta is based on the Black-Scholes model with constant volatility $\widetilde{\sigma}$.
    - The stock price follows the true stochastic-volatility dynamics.

$$V(T) = V(0) + \sum_{i=1}^{I-1} \frac{\partial C(t_i, S(t_i))}{\partial S(t_i)} \left[e^{-rt_{i+1}} S(t_{i+1}) - e^{-rt_i} S(t_i) \right], \quad t_I = T$$

- Under the risk-neutral probability $\mathbf{Q}$,

$$\mathrm{E}_0^{\mathbf{Q}}[V(T)] = V(0) \quad \text{(Check using iterated expectations)}$$

- Can use $V(T)$ as a control variate. The better the discrete-time delta-hedge, the better the control variate that results.

Hedges as Control Variates

Example

- Consider a model of stock returns with stochastic volatility under the risk-neutral probability measure

$$
\begin{aligned}
S((i+1)\Delta) &= S(i\Delta)\exp\left((r - v(i\Delta)/2)\Delta + \sqrt{v(i\Delta)}\sqrt{\Delta}\varepsilon_{i+1}^{\mathbf{Q}}\right) \\
v((i+1)\Delta) &= v(i\Delta) - \kappa(v(i\Delta) - \overline{v})\Delta + \gamma\sqrt{v(i\Delta)\Delta}u_{i+1}^{\mathbf{Q}} \\
&\varepsilon_i^{\mathbf{Q}} \overset{\text{IID}}{\sim} \mathcal{N}(0,1), \quad u_i^{\mathbf{Q}} \overset{\text{IID}}{\sim} \mathcal{N}(0,1), \quad \operatorname{corr}(\varepsilon_i^{\mathbf{Q}}, u_i^{\mathbf{Q}}) = \rho
\end{aligned}
$$

- Price a European call option under the parameters

$$r = 0.05,\ \ T = 0.5,\ \ S_0 = 50,\ \ K = 55, \Delta = 0.01$$

$$v_0 = 0.09,\ \ \overline{v} = 0.09, \kappa = 2,\ \ \rho = -0.5,\ \ \gamma = 0.1, 0.2, 0.3, 0.4, 0.5$$

- Perform 10,000 simulations to estimate the option price. Report the fraction of variance eliminated by the control variate.

Hedges as Control Variates

Example

γ	0.1	0.2	0.3	0.4	0.5
$\widehat{\rho}^2$	0.9944	0.9896	0.9799	0.9618	0.9512
	Naive Monte Carlo				
$\widehat{C}$	2.7102	2.6836	2.5027	2.5505	2.4834
$S.E.(\widehat{C})$	0.0559	0.0537	0.0506	0.0504	0.0483
	Control variates				
$\widehat{C}$	2.7508	2.6908	2.6278	2.5544	2.4794
$S.E.(\widehat{C})$	0.0042	0.0056	0.0071	0.0088	0.0107

Outline

1. Generating Random Numbers
2. Variance Reduction
3. Quasi-Monte Carlo

Quasi-Monte Carlo

Overview

- Quasi-Monte Carlo, or *low-discrepancy methods* present an alternative to Monte Carlo simulation.
- Instead of probability theory, QMC is based on number theory and algebra.
- Consider a problem of integrating a function, $\int_0^1 f(x)\,dx$.
- Monte Carlo approach prescribes simulating N draws of a random variable $X \sim Unif[0,1]$ and approximating

$$\int_0^1 f(x)\,dx \approx \frac{1}{N}\sum_{i=1}^{N} f(X_i)$$

- QMC generates a deterministic sequence X_i, and approximates

$$\int_0^1 f(x)\,dx \approx \frac{1}{N}\sum_{i=1}^{N} f(X_i)$$

- Monte Carlo error declines with sample size as $O(1/\sqrt{N})$. QMC error declines almost as fast as $O(1/N)$.

Quasi-Monte Carlo

- We focus on generating a d-dimensional sequence of low-discrepancy points filling a d-dimensional hypercube, $[0, 1)^d$.
- QMC is a substitute for draws from d-dimensional uniform distribution.
- As discussed, all distributions can be obtained from $Unif[0, 1]$ using the inverse transform method.
- There are many algorithms for producing low-discrepancy sequences. In financial applications, Sobol sequences have shown good performance.
- In practice, due to the nature of Sobol sequences, it is recommended to use $N = 2^k$ (integer k) points in the sequence.

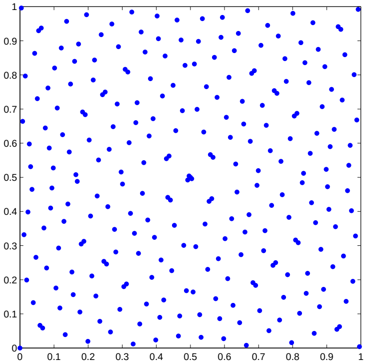
Quasi-Monte Carlo

Illustration

256 points from a 2-dimensional Sobol sequence

MATLAB ® Code

```
P = sobolset(2); X = net(P,256);
```



Quasi-Monte Carlo

Randomization

- Low-discrepancy sequence can be randomized to produce independent draws.
- Each independent draw of N points yields an unbiased estimate of $\int_0^1 f(x)\,dx$.
- By using K independent draws, each containing N points, we can construct confidence intervals.
- Since randomizations are independent, standard Normal approximation can be used for confidence intervals.

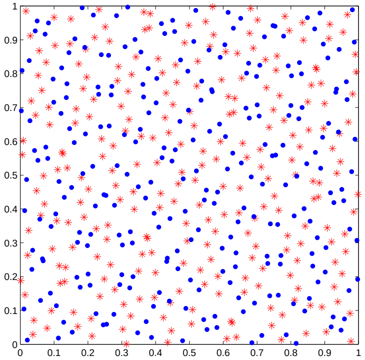
Quasi-Monte Carlo

Randomization

Two independent randomizations using 256 points from a 2-dimensional Sobol sequence

MATLAB® Code

```
q = qrandstream('sobol',2); X = qrand(q,256);
```



Readings

- Campbell, Lo, MacKinlay, 1997, Section 9.4.
- Boyle, P., M. Broadie, P. Glasserman, 1997, "Monte Carlo methods for security pricing," *Journal of Economic Dynamics and Control*, 21, 1267-1321.
- Glasserman, P., 2004, *Monte Carlo Methods in Financial Engineering*, Springer, New York. Sections 2.2, 4.1, 4.2, 7.1, 7.2.

Appendix

Derivation of the Acceptance-Rejection Method

- Suppose the A-R algorithm generates Y. Y has the same distribution as X, conditional on

$$U \leqslant \frac{f(X)}{cg(X)}$$

- Derive the distribution of Y. For any event A,

$$\text{Prob}(Y \in A) = \text{Prob}\left(X \in A | U \leqslant \frac{f(X)}{cg(X)}\right) = \frac{\text{Prob}\left(X \in A, U \leqslant \frac{f(X)}{cg(X)}\right)}{\text{Prob}\left(U \leqslant \frac{f(X)}{cg(X)}\right)}$$

- Note that

$$\text{Prob}\left(U \leqslant \frac{f(X)}{cg(X)} | X\right) = \frac{f(X)}{cg(X)} \quad \text{and therefore}$$

$$\text{Prob}\left(U \leqslant \frac{f(X)}{cg(X)}\right) = \int \frac{f(x)}{cg(x)} g(x)\, dx = \frac{1}{c}$$

- Conclude that

$$\text{Prob}(Y \in A) = c\,\text{Prob}\left(X \in A, U \leqslant \frac{f(X)}{cg(X)}\right) = c \int_A \frac{f(x)}{cg(x)} g(x)\, dx = \int_A f(x)\, dx$$

- Since A is an arbitrary event, this verifies that Y has density f.

15.450 Analytics of Finance
Fall 2010