RIGA TECHNICAL UNIVERSITY

Monte Carlo Methods in Financial Engineering

9. Topic

Variance reduction methods

Variance reduction methods

Goal – to increase the efficiency of Monte Carlo simulations by reducing the variance of simulation estimates:

- *Control variates;*
- *Antithetic variates;*
- *Stratified sampling;*
- *Latin hypercube sampling;*
- *Moment matching methods;*
- *Importance sampling.*

Control variates

- The control variate method is one of the most effective and widely applicable techniques to enhance the efficiency of Monte Carlo simulations.
- It utilizes information about estimation errors in known quantities to reduce the errors in estimating unknown quantities.

Control variates

- $Y_1, \ldots, Y_n$ are the n repetitions of simulation outputs
  - For instance, Y_i could represent the discounted payoff of a derivative in the i-th simulated trajectory
- $Y_i \sim i.i.d.$
- The goal is to estimate $E[Y_i]$
- The simplest estimation is the sample mean $\bar{Y} = (Y_1 + \cdots + Y_n)/n$
- Let's assume that in each repetition, another outcome X_i is simulated alongside Y_i.

Control variates

- Let's assume that the pairs $(X_i, Y_i), i = 1, \dots, n$, are $i.i.d.$ and that $E[X]$ from X_i is known.
- Then, for each fixed value of b from the i-th repetition, we can compute

$$Y_i(b) = Y_i - b(X_i - E[X])$$

and then calculate the sample mean

$$\bar{Y}(b) = \bar{Y} - b(\bar{X} - E[X]) = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - b(X_i - E[X])\right) \qquad (1)$$

- This is a control variate estimation; the observed error $\bar{X} - E[X]$ is useful as a control in estimating $E[Y]$.

Control variates

- For each $Y_i(b)$, the variance is given by

$$Var[Y_i(b)] = \sigma_Y^2 - 2b\sigma_X\sigma_Y\rho_{XY} + b^2\sigma_X^2 \equiv \sigma^2(b) \qquad (2)$$

where $\sigma_X^2 = Var[X], \sigma_Y^2 = Var[Y], \rho_{XY}$ - is the correlation between X and Y

- The optimal coefficient b^* minimizes the variance

$$b^* = \frac{\sigma_Y}{\sigma_X}\rho_{XY} = \frac{Cov[X,Y]}{Var[X]} \qquad (3)$$

Control variates

- By substituting this value into the expression in (2) and simplifying, it can be derived that the optimal controlled estimation ratio to uncontrolled estimation is

$$\frac{Var[\bar{Y} - b^*(\bar{X} - E[X])]}{Var[\bar{Y}]} = 1 - \rho_{XY}^2 \qquad (4)$$

- Conclusions and equation (4) can be applied if the optimal coefficient b^* is known.
- Practically, if $E[Y]$ is unknown, it is unlikely that σ_Y and ρ_{XY} will be known.

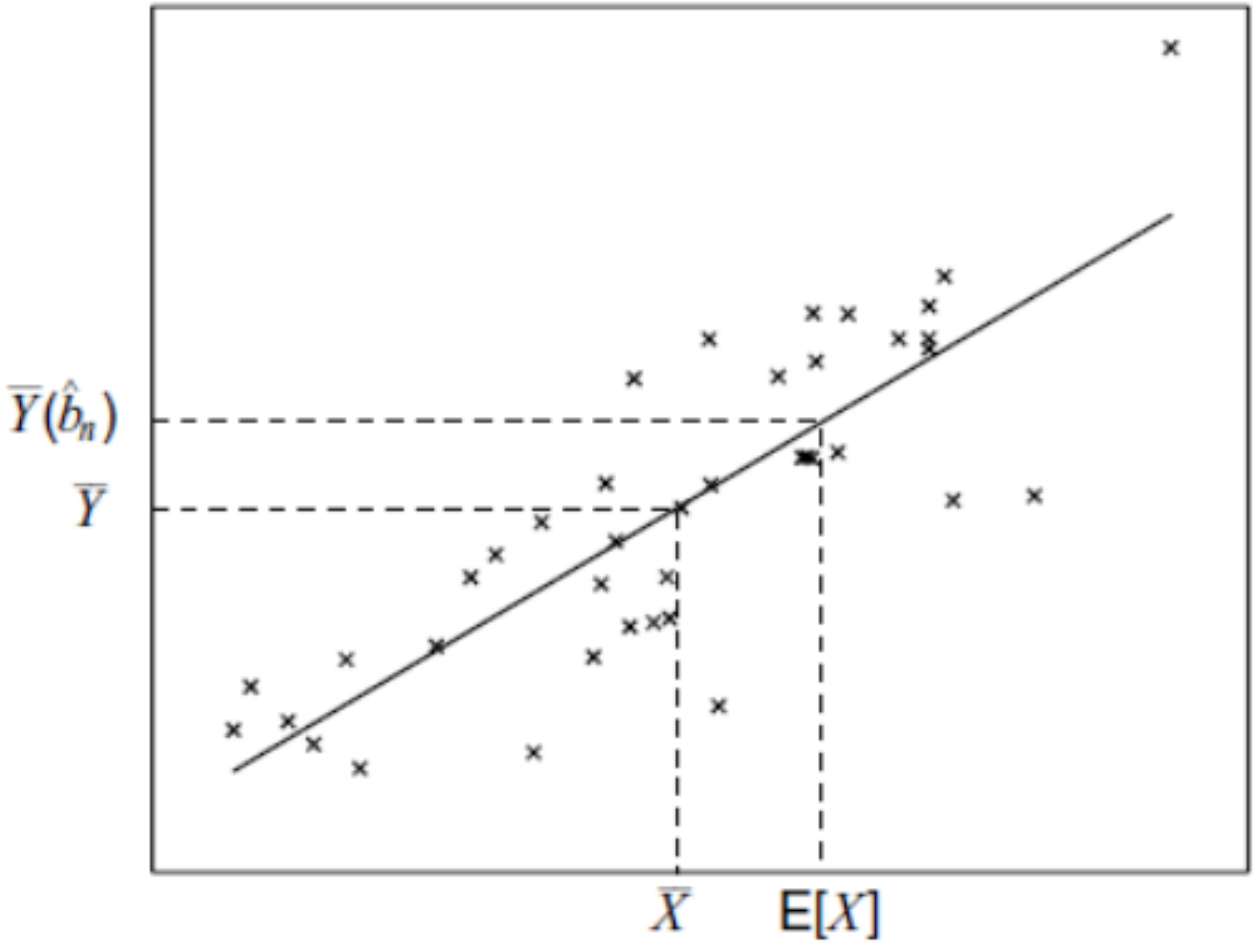
Control variates

- Then the estimation can use

$$\hat{b}_n = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2} \qquad (5)$$

- Equation (5) provides the slope for the least square regression line drawn though points $(X_i, Y_i), i = 1, \ldots, n$.
- The relationship between the controlled variable and regression is useful in the statistical analysis of controlled variable estimations and offers the opportunity for graphical interpretation of this method.

Control variates



Antithetic variates

- Antithetic variates method attempts to reduce dispersion by introducing negative dependency between pairs of repetitions.
- There are various variants of this method, but the most commonly used one is based on the observation that if U is uniformly distributed in the interval $[0,1]$, then $1 - U$ is also uniformly distributed in the interval $[0,1]$.
- Therefore, when generating a trajectory using input data $U_1, \ldots, U_n$, another trajectory can be generated using $1 - U_1, \ldots, 1 - U_n$, without changing the underlying simulation process.

Antithetic variates

- The variables U_i and $1 - U_i$ form an antithetic pair in the sense that a high value of one variable is associated with a low value of the other.
- This means that unusually high or low output values from the first trajectory can be balanced out by values from the opposite trajectory, thereby reducing dispersion.

Antithetic variates

- These observations can be extended to other distributions using the inverse transform method: both $F^{-1}(U)$ and $F^{-1}(1-U)$ follow distribution F, but they are opposite of each other because F^{-1} is monotonic.
- In terms of symmetry concerning the original distribution, $F^{-1}(1-u)$ and $F^{-1}(u)$ represent the same magnitude but with opposite signs.

Antithetic variates

- If simulations are performed using independent standard normal random variables, antithetic variables can be introduced by creating pairs from the sequence $Z_1, Z_2, \ldots i.i.d. N(0,1)$ variables and $-Z_1, -Z_2, \ldots i.i.d. N(0,1)$ variables.
- If Z_i is used to simulate increments in Brownian motion, then $-Z_i$ simulates the increments mirrored (reflected) from the original trajectory.
- Thus, by simulating pairs using the original trajectory and its mirrored reflection, it can result in lower dispersion.

Antithetic variates

- The objective of the method is to estimate the mathematical expectation $E[Y]$, and then, by implementing one of the processes of creating antithetic variables, to create a sequence of pairs of observations $(Y_1, \bar{Y}_1), (Y_2, \bar{Y}_2), \dots, (Y_n, \bar{Y}_n)$
- $(Y_1, \bar{Y}_1), (Y_2, \bar{Y}_2), \dots, (Y_n, \bar{Y}_n)\ are\ i.i.d.$
- $\forall i$: $Y_i\ and\ \bar{Y}_i$ share the same distribution, although they are typically not independent.

Antithetic variates

- The estimation of the antithetic variables is the average value from all $2n$ observations

$$\hat{Y}_{AN} = \frac{1}{2n}\left(\sum_{i=1}^{n} Y_i + \sum_{i=1}^{n} \tilde{Y}_i\right) = \frac{1}{n}\sum_{i=1}^{n} \frac{Y_i + \tilde{Y}_i}{2}$$

- $\hat{Y}_{AN}$ is the sample mean from $n\ independent$ observations

$$\left(\frac{Y_1 + \tilde{Y}_1}{2}\right), \ldots, \left(\frac{Y_n + \tilde{Y}_n}{2}\right)$$

Antithetic variates

- By using the fact that Y_i and $\tilde{Y}_i$ have the same variance, if they have the same distribution, then the variance $Var[Y_i + \tilde{Y}_i] = 2Var[Y_i] + 2Cov[Y_i, \tilde{Y}_i]$.
- Thus, the condition for forming antithetic variables to reduce dispersion is

$$Cov[Y_i, \tilde{Y}_i] < 0$$

- Antithetic variables are highly effective when the relationship between input and output is close to linear.

Stratified sampling

- Stratified sampling generally refers to any sampling mechanism that involves selecting observations from specific subgroups of strata.
- The objective is to estimate $E[Y]$, where Y is a real value.
- If $A_1, \ldots, A_K$ are non-overlapping subsets of the real number line where $P(Y \in U_i A_i) = 1$, then

$$E[Y] = \sum_{i=1}^{K} p_i E[Y|Y \in A_i], \qquad p_i = P(Y \in A_i)$$

Stratified sampling

- The strata is defined as the second variable X.
- This stratification variable can take values from an arbitrary set.
- $P(X \in U_i A_i) = 1$
- $E[Y] = \sum_{i=1}^{K} p_i E[Y|X \in A_i], \quad p_i = P(X \in A_i) \qquad (6)$
- In some applications, Y is a function of X (for instance, X could represent discrete trajectories of asset prices, and Y could be the discounted payoff of a derivative). However, in general, they can be dependent on each other without fully specifying one another.

Stratified sampling

- To use equation (6) for stratified sampling, it's necessary to generate pairs $(X_{ij}, Y_{ij}), j = 1, \dots, n_i$, considering the conditional distribution (X, Y) given $X \in A_i$.
- If $q_i = n_i/n$ is the fraction of observations from the i-th stratum, $i = 1, \dots, K$, then the estimation of $E[Y]$ can be found as

$$\hat{Y} = \frac{1}{n} \sum_{i=1}^{K} \frac{p_i}{q_i} \sum_{j=1}^{n_i} Y_{ij}$$

- Minimizing the dispersion of this observation over q_i, one can find a distribution rule that is at least as effective as a proportional distribution.

Latins Hypercube Sampling

- Latins Hypercube Sampling is an extension of stratified sampling for multidimensional cases.
- Its complexity becomes apparent in the straightforward case of sampling a d-dimensional hypercube $[0,1)^d$.
- Dividing each coordinate into K strata creates K^d strata for the hypercube. Hence, the sample size needs to be at least K^d to ensure inclusion of each stratum in the sample.

Latins Hypercube Sampling

- The Latin hypercube Sampling method processes all coordinates equally and avoids the possibility of the sample size growing exponentially, which could occur in full stratification when stratifying only the multi-dimensional marginal distributions of the multidimensional joint distribution.
- The simplest description of sampling involves creating a sample from a uniform distribution within a until hypercube.
- Let’s fix the dimension d and the sample size K.

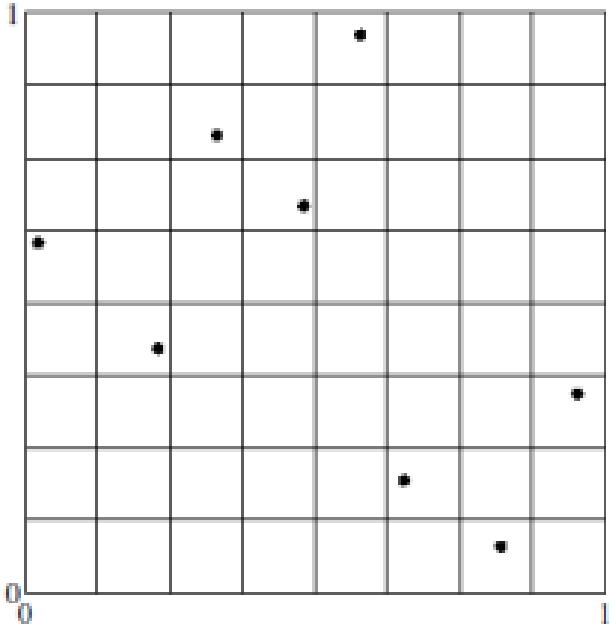
Latins Hypercube Sampling

- For each coordinate $i = 1, \ldots, d$, independently generate a stratified sample $V_i^{(1)}, \ldots, V_i^{(K)}$ from the unit interval, using K equally probable strata; each $V_i^{(j)}$ is uniformly distributed in the interval $[(j-1)/K\,, j/K)$.
- If the d stratified samples are arranged in columns, each row forms the coordinates of a point from $[0,1)^d$

$$\begin{matrix} V_1^{(1)} & V_2^{(1)} & \cdots & V_d^{(1)} \\ V_1^{(2)} & V_2^{(2)} & \cdots & V_d^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ V_1^{(K)} & V_2^{(K)} & \cdots & V_d^{(K)} \end{matrix}$$

Latins Hypercube Sampling

- Latins Hypercube Sampling's "marginal" stratification property is demonstrated in the figure showing a sample of size $K = 8$ with dimension $d = 2$. When projecting points with two coordinates, it illustrates that exactly one point falls into each of the 8 sub-divisions (or bins).



Moment matching methods

- The objective is to ensure that the average value obtained from a specific sample via simulation matches precisely with the population value (i.e., values that would be obtained by taking an infinite number of repetitions).
- In financial engineering, reconciling samples and populations often means ensuring a finite in the formation of base asset prices.

Moment matching methods

- *Moment matching* is based on transforming simulated trajectories.
- The idea behind trajectory transformation for moment matching, in its simplest form, can be demonstrated by modeling with an interest rate r where one underlying asset $S(t)$ is placed, simulated using a risk-free measure.
- If no dividends are paid, then $E[S(t)] = e^{rt}S(0)$.

Moment matching methods

- Let's assume that n independent repetitions of the process are simulated as $S_1, \dots, S_n$, and define the process of sample means

$$\bar{S}(t) = \frac{1}{n}\sum_{i=1}^{n} S_i(t)$$

Moment matching methods

- In general, the final sample mean of n simulations may not match $E[S(t)]$; simulation could create an incorrect base asset price, meaning that

$$e^{-rt}\bar{S}(t) \neq S(0)$$

  On the right-hand side, there-s the current price of the asset, and on the left-hand side, there's its simulated estimation.

Moment matching methods

- The possibility to rectify this involves transforming the simulated trajectory using either:

$$\tilde{S}_i(t) = S_i(t) \cdot \frac{E[S(t)]}{\bar{S}(t)}, \qquad i = 1, \dots, n$$

or

$$\tilde{S}_i(t) = S_i(t) + E[S(t)] - \bar{S}(t), \qquad i = 1, \dots, n$$

then, using $\tilde{S}_i(t)$ instead of S_i to determine the derivative security price.

- The application of these corrections ensures that the sample mean from $\tilde{S}_1(t), \dots, \tilde{S}_n(t)$ will precisely match $E[S(t)]$.

Importance sampling

- Importance sampling aims to reduce variance by changing the probability measure used for generating trajectories.
- When transitioning from the target probability measure to the risk-neutral measure, our goal is typically to obtain a more convenient representation of the expected value.
- In this method, measures are adjusted to assign greater weight to significant output values, thereby increasing the effectiveness of sample generation.

Importance sampling

- For example, one need to evaluate

$$a = E[h(X)] = \int h(x)f(x)dx$$

where X is a random element from $\mathbb{R}^d$ with density function f, and h is a function from $\mathbb{R}^d$ to $\mathbb{R}$.

Importance sampling

- The standard Monte Carlo estimation is

$$\hat{a} = \hat{a}(n) = \frac{1}{n}\sum_{i=1}^{n} h(X_i)$$

  where $X_1, \ldots, X_n$ are independent random variables obtained from the density function f.
- Let g be another density function in the space $\mathbb{R}^d$ such that it holds that $f(x) > 0 \Rightarrow g(x) > 0$ for all $x \in \mathbb{R}^d$.

Importance sampling

- Then alternatively, we can represent a as

$$a = \int h(x) \frac{f(x)}{g(x)} g(x) dx$$

- This integral can be interpreted as the explanation with respect to the density function g; therefore

$$a = \tilde{E}\left[h(X) \frac{f(X)}{g(X)}\right]$$

- Here, the notation $\tilde{E}$ denotes that the mathematical expectation is taken using X, which is distributed according to g.

Importance sampling

- If $X_1, \ldots, X_n$ are independently obtained using g, the *importance sampling* estimate related to g is

$$\hat{a}_g = \hat{a}_g(n) = \frac{1}{n}\sum_{i=1}^{n} h(X_i)\frac{f(X_i)}{g(X_i)}$$

- The weights $f(X_i)/g(X_i)$ represents the *likelihood ratio* or the *Radon-Nikodym derivative* obtained with X_i.
- $\tilde{E}[\hat{a}_g] = a$, hence $\hat{a}_g$ is an unbiased estimator for a.

Importance sampling

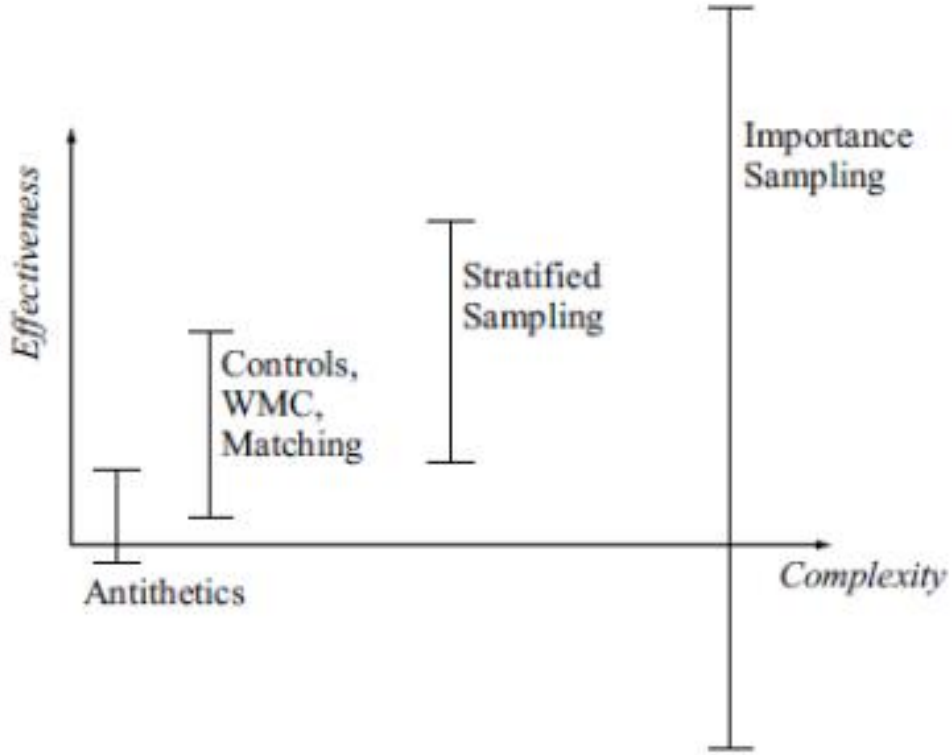
- To compare variances with and without *importance sampling*, it suffices to compare their second moments.
- Successful application of this method relies on the skill to choose the *importance sampling* density function g effectively.

Notes

- The choice of method depends on the available information and the allocated time for adapting general methods to a specific problem.
- Understanding the strengths and weaknesses of alternative methods and having a good knowledge of their effective application examples are crucial aspects in method selection.

Notes

- In the chart, methods are arranged according to their complexity and depict the typical effectiveness of each



Thank You!