

Monte Carlo

Monte Carlo

Monte Carlo is useful to estimate the properties of estimates. Assume that we have a population space with a defined process in it and we want to estimate parameters of the process. We can take samples and estimate those parameters. Further, we construct a sampling distribution. The problem with that is that we don't know the actual values of the parameters. What we can simulate the population process and derive samples from it. In this case we know actual values of estimated parameters, by construction. Hence, construction of sample distribution allows us to conclude if parameters are unbiased, consistent and estimate their variance.

Interview Questions

1. Binomial Tree vs Monte-Carlo
 2. Longstaff-Schwartz algorithm for pricing an early exercisable options with Monte-Carlo
 3. Change of Measure. Give an example when a change of measure is useful in MC
 4. Grandma problem with monte carlo
 5. 159 q5.20 Given that a stock price at time T is $N(100,1)$, you want to price a digital call struck at 110 by Monte Carlo simulation. What will happen if you do this? Improve the method
 6. 190 q 6.4 how would you determine pi by Monte Carlo simulation
 7. Asian Option pricing with monte carlo
 8. Central Limit Theorem and types of convergence
 9. $E[X/Y] \neq E[X]/E[Y]$ - bias exists but becomes negligible as the number of replications increases, and the convergence rate of the estimator is unaffected.
- . 3 considerations are particularly important for MC computations: time, bias, variance . Model Discretization error.

Variance Reduction Techniques

1. control variates
 - It exploits information about the errors in estimates of known quantities to reduce the error in an estimate of unknown quantity.
 - The idea is to speedup variance convergence, by doing more computations. This works because MC converges very slowly. To reduce Variance by a factor of 2 we need to perform $4 \times n$ computations.
 - Let Y_1, Y_2, Y_n be outputs from some replications. Now, for each replication we compute X_i . Assume that $E[X_i]$ is known. Then for any b: $Y_i(b) = Y_i - b(X_i - E[X])$. Then the sample mean $\bar{Y}(b) = \bar{Y} - b(\bar{X} - E[X]) = \frac{1}{n} \sum_{i=1}^n (Y_i - b(X_i - E[X]))$. As an estimator of $E[Y]$, the control estimator is unbiased because $E[\bar{Y}(b)] = E[\bar{Y} - b(\bar{X} - E[X])] = E[\bar{Y}] = E[Y]$
 - The variance: $Var[Y_i(b)] = \sigma_Y^2 - 2b\sigma_X\sigma_Y\rho_{XY} + b^2\sigma_X^2 = \sigma^2(b)$. The optimal coefficient $b^* = \frac{\sigma_Y}{\sigma_X}\rho_{XY}$
Then $\frac{\bar{Y} - b^*(\bar{X} - E[X])}{Var[\bar{Y}]} = 1 - \rho_{XY}^2$

- A few observations follow from this expression:
 - With the optimal coefficient b^* , the effectiveness of a control variate, as measured by the variance reduction ratio (4.4), is determined by the strength of the correlation between the quantity of interest Y and the control X . The sign of the correlation is irrelevant because it is absorbed in b^* .
 - If the computational effort per replication is roughly the same with and without a control variate, then (4.4) measures the computational speed-up resulting from the use of a control. More precisely, the number of replications of the Y_i required to achieve the same variance as n replications of the control variate estimator is $n/(1 - \rho_{XY}^2)$.
 - The variance reduction factor $n/(1 - \rho_{XY}^2)$ increases very sharply as $|\rho_{XY}|$ approaches 1 and, accordingly, it drops off quickly as $|\rho_{XY}|$ decreases away from 1. For example, whereas a correlation of 0.95 produces a ten-fold speedup, a correlation of 0.90 yields only a five-fold speed-up; at $|\rho_{XY}| = 0.70$ the speed-up drops to about a factor of two. This suggests that a rather high degree of correlation is needed for a control variate to yield substantial benefits.
- In derivative pricing simulations, underlying assets provide a virtually universal source of control variates. If $S(t)$ is an asset price then $e^{-rt}S(t)$ is a martingale. Then the control variate estimator is $\frac{1}{n} \sum_{i=1}^n (Y_i - b[S_i(T) - e^{rT}S(0)])$. If $Y = e^{-rT}(S(T) - K)^+$
- Any martingale with a known initial value provides a potential control variate precisely because its expectation at any future time is its initial value.
- Options or Bonds could potentially become control variates
- Every instrument that serves as a good hedge also serves as a good control variate, if it can be easily priced. Additionally, dynamic hedge can remove all the variance assuming complete market and continuous trading.

2. Antithetic variates

- The method of antithetic variates attempts to reduce variance by introducing negative dependence between pairs of replications. The most applicable form is based on the observation that if U is uniformly distributed on $[0, 1]$ then $1 - U$ is too. Hence, if we generate a path using as inputs U_1, U_2, \dots, U_n , we can generate a second path using $1 - U_1, 1 - U_2, \dots, 1 - U_n$.
- The variables U_i and $1 - U_i$ form an antithetic pair in the sense that a large value computed from the first path could be balanced out by the value computed by the antithetic path, thus reducing the variance.
- This observation extends to other distributions through the inverse transform method: $F^{-1}(U)$ and $F^{-1}(1 - U)$ both have distribution F but are antithetic to each other because F^{-1} is monotone. In particular, if $Z_1, Z_2, \dots, Z_n \text{ i.i.d. } N(0, 1)$ antithetic variables could be formed by $-Z_1, -Z_2, \dots, -Z_n$.

3. stratified sampling

- Stratified sampling refers broadly to any sampling mechanism that constrains the fraction of observation drawn from specific subsets (or strata) of the sample space. Suppose, more specifically, that our goal is to estimate $E[Y]$ with Y real-valued, and let A_1, \dots, A_K be disjoint subsets of the real line for which $P(Y \in \cup_i A_i) = 1$. Then $E[Y] = \sum_{i=1}^K P(Y \in A_i)E[Y|Y \in A_i] = \sum_{i=1}^K p_i E[Y|Y \in A_i]$. In random sampling, we generate independent Y_1, \dots, Y_n having the same distribution as Y . The fraction of these samples falling in A_i will not in general equal p_i , though it would approach p_i as the sample size n increased. In stratified sampling, we decide in advance what fraction of the samples should be drawn from each stratum A_i ; each observation drawn from A_i is constrained to have the distribution of Y conditional on $Y \in A_i$.
- Stratifying nonuniform distributions. Let F be a CDF on the real line and let

$$F^{-1}(u) = \inf x : F(x) \leq u$$

be its generalised inverse. Given probabilities p_1, \dots, p_K summing to 1, define

$$a_0 = -\infty, a_1 = F^{-1}(p_1), a_2 = F^{-1}(p_1 + p_2), \dots, a_K = F^{-1}(p_1 + \dots + p_K) = F^{-1}(1)$$

Define strata

$$A_i = (a_{i-1}, a_i] \text{ for } i = 1, \dots, K$$

To use the sets A_1, \dots, A_K for stratified sampling, we need to be able to generate samples of Y conditional on $Y \in A_i$. If $U \sim \text{Unif}[0, 1]$ then $V = a_{i-1} + U(a_i - a_{i-1})$ is uniformly distributed between a_{i-1} and a_i and then $F^{-1}(V)$ has a distribution of Y conditional on $Y \in A_i$

4. Latin hypercube sampling

5. Moment matching methods

6. importance sampling

- Importance sampling attempts to reduce variance by changing the probability measure from which paths are generated. We change measures to try to give more weight to “important” outcomes thereby increasing sampling efficiency.

Consider the problem of estimating

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

, where $X \in \mathbb{R}^d$ is a r.v. with probability density f , and $h : \mathbb{R}^d \rightarrow \mathbb{R}$. Let g be any other probability density on \mathbb{R}^d s.t. $f(x) > 0 \Rightarrow g(x) > 0 \forall x$. Then

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

. This integral can be interpreted as an expectation with respect to the density g :

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

. If X_i are now independent draws from g , the importance sampling estimate associated with g is

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

. The weight $\frac{f(X_i)}{g(X_i)}$ is the likelihood ratio or Radon-Nikodym derivative. It follows that $\tilde{\alpha}_g$ is an unbiased estimator of α

- To compare variances with and without importance sampling it therefore suffices to compare second moments:

$$\tilde{E}\left[\left(h(X) \frac{f(X)}{g(X)}\right)^2\right] = E\left[\left(h(X)^2 \frac{f(X)}{g(X)}\right)\right]$$

This could be larger or smaller than the second moment $E[h(X)^2]$ without importance sampling. Successful importance sampling lies in the art of selecting an effective importance sampling density g .

- Nevertheless, this optimal choice of g does provide some useful guidance: in designing an effective importance sampling strategy, we should try to sample in proportion to the product of h and f . In option pricing applications, h is typically a discounted payoff and f is the risk-neutral density of a discrete path of underlying assets. In this case, the “importance” of a path is measured by the product of its discounted payoff and its probability density.

References:

1. Monte Carlo Methods in Financial Engineering. Paul Glasserman
2. Quant Job Interview Questions and Answers. Mark Joshi, Nick Denson