planetmath.org

Math for the people, by the people.

Monte Carlo simulation

Canonical name	MonteCarloSimulation
Date of creation	2013-03-22 17:20:09
Last modified on	2013-03-22 17:20:09
Owner	stevecheng (10074)
Last modified by	stevecheng (10074)
Numerical id	5
Author	stevecheng (10074)
Entry type	Topic
Classification	msc 65C05
Classification	msc 62-00
Classification	msc 60-00
Synonym	Monte Carlo
Synonym	statistical sampling
Synonym	stochastic simulation
Related topic	AcceptanceRejectionMethod

Monte Carlo simulation is the use of randomized numerical experiments to evaluate mathematical expressions.

In the typical problems addressed by Monte Carlo simulation, the search space or sample space is countably or uncountably infinite. In contrast to determistic algorithms that sweep a finite subset of points in the search space in order to derive a solution to the problem, Monte Carlo simulation randomizes the selection of points in the hope that good representatives for the solution of the problem are chosen.

Monte Carlo integration

The term Monte Carlo simulation most often refers to integration by randomized methods. The problem is to evaluate the expectation or integral

$$\mathbb{E}X = \int X \, d\mathbb{P}$$

of a real-valued random variable X on a probability space. The expectation is approximated by taking the sample mean of independent observations $\{X_i\}$ with the same distribution as X:

$$\mathbb{E}X \approx \frac{1}{n} \sum_{i=1}^{n} X_i \, .$$

By the strong law of large numbers, the sample mean converges almost surely, as $n \to \infty$, to the true mean $\mathbb{E}X$.

Typically the distribution of X is not known in closed form; otherwise we would be able to use the standard numerical quadrature techniques to compute the one-dimensional integral representing $\mathbb{E}X$, and one-variable quadrature in general tend to work better than the Monte Carlo method.

Rather, the random variable $X = f(Y)$ is expressed as some function f of another random variable Y, and we know how to compute f and randomized samples of Y. Then

$$\mathbb{E}X = \mathbb{E}[f(Y)] \approx \frac{1}{n} \sum_{i=1}^{n} f(Y_i)$$

for independent samples $\{Y_i\}$ for the random variable Y.

The realizations Y_i may be obtained by generating random (or pseudo-random) samples according to the known distribution for Y. Or, in some

cases, they may be obtained from pre-tabulated observations, e.g. based on past observations in the real world.

Bounds on error

If $X_1, X_2, \ldots$ are identically and independently distributed, with mean m and variance v, then

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{X_i - m}{\sqrt{v}}$$

converges in distribution, as $n \to \infty$, to the standard normal distribution. Then for any $\delta > 0$, and for large n, we have

$$\Pr\left(-\delta < \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{X_i - m}{\sqrt{v}} < \delta\right) \approx 2\Phi(\delta) - 1\,,$$

approximating the true distribution with the Gaussian cumulative distribution function Φ. Setting $(1-\alpha)\times 100\% = 2\Phi(\delta) - 1$ as the required confidence level, we have $\delta = \Phi^{-1}(1 - \alpha/2)$.

Thus, given an observation of the sample mean from a run of the Monte Carlo simulation, with $(1 - \alpha) \times 100\%$ confidence, the true mean $\mathbb{E}X$ is approximately within

$$\pm\delta\sqrt{\frac{v}{n}} = \pm\sqrt{\frac{v}{n}}\,\Phi^{-1}(1 - \alpha/2)$$

of the sample mean.

Since the computed sample mean is random, it is in general not possible to obtain actual error *bounds*, in the usual sense of the word, as with determistic algorithms. Instead, the confidence interval size, or the standard deviation of $(1/n)\sum_{i=1}^{n} X_i$, is used as a measure of the error of Monte Carlo simulation; this is sometimes called a *probabilistic error bound*.

Observe that the probabilistic error bound of the Monte Carlo simulation decreases as $1/\sqrt{n}$.

Comparison with other methods

Simplicity. The chief advantage of Monte Carlo simulation, compared to the other numerical methods that can solve the same problem, is that

it is conceptually very simple. It does not require specific knowledge of the form of the solution or its analytic properties. Monte Carlo is also relatively easy to implement on a computer.

Slowness. The main disadvantage of Monte Carlo integration is that it is slow. Many samples may be required — on the order of thousands or even millions — to obtain acceptable precision in the answer. In particular, since the probabilistic error bound decreases as the reciprocal square root of the number of iterations, to achieve one more decimal digit of precision in the answer requires $10^2 = 100$ times more the number of iterations.

Other advantages of Monte Carlo.

Independence of dimension. The amount of work to obtain the same amount of precision is independent of the dimension d of the underlying random variables.

Thus Monte Carlo integration is practically the only method to numerically compute high-dimensional integrals. Traditional quadrature techniques generally require an amount of work exponential in the number of dimensions d, since they require sampling a grid in d-dimensional space.

On the other hand, Monte Carlo integration is generally not competitive with quadrature for low-dimensional integration (e.g. $d = 1$ or $d = 2$).

Unrestricted choice of functions. The functions to integrate with Monte Carlo can be practically arbitrary. No smoothness conditions or boundedness conditions are needed, for example, providing the integral is finite.

(However, irregularities in the integrand may impact the accuracy of the result.)

Easily parallelizable. Many computer processors can be participating in a Monte Carlo simulation simultaneously. Each simulation is independent of another.

Other disadvantages of Monte Carlo.

Error may depend on distribution. The estimate of the expectation $\mathbb{E}X$ may be impacted severely if the distribution of X is heavily skewed or has heavier-than-normal tails. A non-random numerical method may avoid these deficiencies, or at least not be as severely impacted.

Difficulty in estimating error. There are no hard bounds on the error of the computed result. The probabilistic error bound, which is essentially based on the variance, may not be a good measure of the error for skewed distributions.

Black-box approach. With some types of analytical approximation, one can study the behavior of the solution if the initial parameters are changed. This generally is hard to do with the black-box approach of Monte Carlo.

Variance reduction

There are some variance reduction techniques that can be used to reduce the error in the result of a Monte Carlo simulation. However, they generally cannot overcome the slow $1/\sqrt{n}$ decrease of the error inherent in Monte Carlo.

References

[1] James E. Gentle. *Random Number Generation and Monte Carlo Methods*, second edition. Springer, 2003.

[2] "`http://en.wikipedia.org/wiki/Monte_Carlo_method`Monte Carlo method". Wikipedia, The Free Encyclopedia. Accessed 27 June 2007.