Introduction to Monte Carlo Simulation With Applications

Professor Karl Sigman
Columbia University
Department of IEOR
New York City
USA

January 20, 2021

Introduction (1)

Suppose our computer could hand us, upon demand, numbers,

$$U_1, U_2, \ldots, U_n,$$

that were independent and identically distributed (iid) samples from the continuous uniform distribution on the unit interval $(0, 1)$:

$$P(U \leq x) = x, \ x \in (0, 1).$$

The density function is given by

$$f(x) = 1, \ x \in (0, 1), \qquad f(x) = 0, \ x \notin (0, 1).$$

QUESTION: What useful/practicle things can we do with these numbers?

Introduction (2)

ANSWER: As we shall see, by using the U_i to construct random variables X with any desired distribution, $F(x) = P(X \leq x)$, we can then

- construct various stochastic processes (random walks, point processes, Markov chains, Brownian motion, etc.) so that we can
- *simulate* complicated models (queueing models in telecommunications, insurance risk models, asset price models, etc.) and numerically estimate quantities of interest (average delay at a call center, probability of ruin, price of an exotic option or derivative).

Outline

- Simple integration example; strong law of large numbers, the central limit theorem, confidence intervals for justifying the method.
- Basics techniques of stochastic simulation: Inverse transform method, acceptance rejection, polar method for generating random numbers X via using uniforms U.
- Simulating a Poisson process, and more generally a renewal process, Markov chains
- The binomial lattice model (BLM) for risky assets; some option pricing
- Simulating Brownian motion, geometric Brownian motion
- Markov Chain Monte Carlo (MCMC) method

Integration (1)

Suppose we wish to compute an integral

$$\alpha = \int_0^1 g(x)dx,$$

for a function g that has no known antiderivative.
Observe that if $U \sim unif(0,1)$, then the expected value of the random variable $g(U)$ is in fact the integral:

$$E(g(U)) = \int_0^1 g(x)dx.$$

This follows from the more general basic fact in probability that if X is a rv with density function $f(x)$, then the expected value of the random variable $g(X)$, for a function g is given by

$$E(g(X)) = \int g(x)f(x)dx.$$

Integration (2)

In the following we shall be applying $X_i = g(U_i)$, where the U_i hence the X_i are iid rvs with mean $E(X) = \alpha = \int_0^1 g(x)dx$.
The famous and fundamental *Strong Law of Large Numbers (SLLN)* in probability theory asserts that:

Theorem (SLLN)

For any iid sequence of random variables X_1, $X_2, \ldots$ with finite mean $E(X)$, it holds that with probability 1:

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} X_j = E(X).$$

Thus

$$E(X) \approx \frac{1}{n} \sum_{j=1}^{n} X_j, \text{ for } n \text{ large}$$

Integration (3)

Thus, using our computer's iid numbers $U_1, U_2, \ldots, U_n$, for large n, and defining $X_1 = g(U_1), \ldots, X_n = g(U_n)$, yields an approximation for our integral:

$$\int_0^1 g(x)dx \approx \frac{1}{n}\sum_{j=1}^{n} g(U_j).$$

Integration (4)

This is an example of *Monte Carlo Simulation*: We can numerically estimate an integral by first expressing it as an expected value, and then applying the SLLN.
This method works for multidimensional integrals as well: For example the $3-d$ integral

$$\int_0^1 \int_0^1 \int_0^1 g(x,y,z)dxdydz = E(g(U_1,U_2,U_3)).$$

We then define $X_1 = g(U_1,U_2,U_3),\ X_2 = g(U_4,U_5,U_6)$ and so on and once again use the approximation

$$\frac{1}{n}\sum_{j=1}^{n} X_j.$$

The precision of this method does not depend on the dimension of the integral!!

Integration (5)

The integration limits $\int_0^1$ can be more generally $\int_a^b$ by using the change of variables $y = (x - a)/(b - a),\ x \in (a, b)$. For example

$$\int_a^b g(x)dx = (b - a)\int_0^1 g(a + (b - a)y)dy.$$

Monte Carlo method

The general theme: We need to compute something which equivalently can be expressed as an expected value, $E(X)$, for some (perhaps very complicated) rv X. We ask the computer to *simulate* iid copies of X, then use the SLLN to estimate $E(X)$ as an empirical average. We will give other examples soon, but first let us recall the notion and usefulness of confidence intervals.

Review of confidence intervals in statistics (1)

In statistics, we estimate an unknown mean $\mu = E(X)$ of a distribution by collecting n iid samples from the distribution, $X_1, \ldots, X_n$ and using the sample mean

$$\overline{X}(n) = \frac{1}{n} \sum_{j=1}^{n} X_j. \tag{1}$$

This estimate is justified by the SLLN. But how good is our estimate?

Review of confidence intervals in statistics (2)

Letting $\sigma^2 = Var(X)$ denote the variance of the distribution, we have

$$Var(\overline{X}(n)) = \frac{\sigma^2}{n}. \tag{2}$$

The *Central Limit Theorem* asserts that as $n \to \infty$,

$$Z_n \stackrel{\text{def}}{=} \frac{\sqrt{n}}{\sigma}(\overline{X}(n) - \mu),$$

converges *in distribution* to $N(0,1)$, the unit normal distribution. Letting Z denote a $N(0,1)$ rv, we conclude that for n sufficently large, $Z_n \approx Z$ in distribution, meaning that $P(Z_n \leq x) \approx \Theta(x)$ for any $x \in \mathbf{R}$.

Review of confidence intervals in statistics (3)

This allows us to construct *confidence intervals* for our estimate: (letting $z_{\alpha/2}$ be such that $P(Z > z_{\alpha/2}) = \alpha/2$):

> *we say that the interval* $\overline{X}(n) \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ *is a* $100(1-\alpha)\%$ *confidence interval for the mean* μ.

Typically, we would use (say) $\alpha = 0.05$ in which case $z_{\alpha/2} = z_{0.025} = 1.96$, and we thus obtain a 95% confidence interval $\overline{X}(n) \pm (1.96)\frac{\sigma}{\sqrt{n}}$.

Review of confidence intervals in statistics (4)

We would not know σ in practice, so we instead use an estimate for it, the *sample variance* $s^2(n)$ defined by

$$s^2(n) = \frac{1}{n-1}\sum_{j=1}^{n}(X_j - \overline{X}(n))^2.$$

It can be shown that $s^2(n) \to \sigma^2$, with probability 1, as $n \to \infty$ and that $E(s^2(n)) = \sigma^2,\ n \geq 2$.
So, in practice we would use $s(n)$ is place of σ when constructing our confidence intervals. For example, a 95% confidence interval is given by $\overline{X}(n) \pm (1.96)\frac{s(n)}{\sqrt{n}}$.

Review of confidence intervals in statistics (5)

The following recursions can be derived; they are useful when implementing a simulation requiring a confidence interval:

$$\overline{X}(n+1) = \overline{X}_n + \frac{X_{n+1} - \overline{X}(n)}{n+1},$$

$$S(n+1)^2 = \left(1 - \frac{1}{n}\right) S^2(n) + (n+1)(\overline{X}(n+1) - \overline{X}(n))^2.$$

Inverse transform method

Theorem

If $F(x) = P(X \leq x)$ is the cumulative distribution of a desired rv, and if

$$F^{-1}(y) = \min\{x : F(x) \geq y\}, \ y \in [0, 1],$$

(the generalized inverse), then setting $X = F^{-1}(U)$ yields a rv distributed exactly as F.

For example, if X has the exponential distribution at rate λ, $F(x) = 1 - e^{-\lambda x}$, then this leads to the algorithm

$$X = -\lambda^{-1} \ln(U).$$

This method requires having an explicit closed form for F^{-1} which is not always possible (consider the normal distribution for example). Other methods: acceptance-rejection, polar method, and so on.

Inverse-transform method for discrete random variables

Consider a non-negative discrete rv X with probability mass function (pmf) $p(k) = P(X = k),\ k \geq 0$. In this case, the construction $X = F^{-1}(U)$ is explicitly given by: $X = 0$ if $U \leq p(0)$,

$$X = k, \text{ if } \sum_{i=0}^{k-1} p(i) < U \leq \sum_{i=0}^{k} p(i),\ k \geq 1.$$

This is known as the *discrete inverse-transform method.*
For example, if $P(X = 1) = p$ and $P(X = 0) = 1 - p$ is a Bernoulli (p) rv, then we can generate it via Set $X = 0$ if $U \leq 1 - p$, set $X = 1$ if $U > 1 - p$.
We also could do this via: Set $X = 1$ if $U \leq p$, set $X = 0$ if $U > p$.