



Monte Carlo methods

Monte Carlo methods are used for estimating integrals based on pseudo-random numbers. Typically the integrals can not be performed by other means such as numerical methods; this is particularly the case for high-dimensional integrals.

The perhaps earliest application of the simple Monte Carlo method is “*Buffon’s needle*”, see Exercise 2 below (this experiment by Buffon goes back to 1777). Enrico Fermi and Stanislaw Ulam reinvented the method in physics. Fermi worked on neutron diffusion in Rome in the early 1930s. Ulam, and two other pioneers, John von Neumann and Nicholas Metropolis, worked on the Manhattan Project at Los Alamos during the Second World War (the atomic bomb was constructed at Los Alamos). In 1947 Metropolis and von Neumann showed how the Monte Carlo method could solve a number of problems concerned with neutron export in the hydrogen bomb. This work was an impressive success, but they were not able to publish their results, because they were classified as secret. Over the following two years however, they and others applied the method to a variety of more mundane problems in physics, and published a number of papers which drew the world’s attention to this emerging technique. Of particular note to us is the publication in 1953 of the paper by Metropolis and coworkers, in which they describe for the first time the Monte Carlo technique that has come to be known as the *Metropolis algorithm* (which we study later on this course).

As indicated much of the theoretical and practical development of Monte Carlo methods have appeared in physics. In the last twenty years, statisticians have contributed substantially to this development. Due to the exponential growth in computer power and the invention of new algorithms, Monte Carlo methods are widely used today in science and technology.

Exercise 1 (The simple Monte Carlo method)

For specificity, suppose that X is a continuous random variable with density function f (the results and ideas described below apply as well if X is a discrete random variable), and we want to calculate the mean

$$\theta = E(h(X)) = \int h(x)f(x)dx$$

where $h : \mathbb{R}^n \rightarrow \mathbb{R}$ is a function such that the mean exists. Recall the following important result:

Theorem 1 — the strong law of large numbers If X_1, X_2, \dots are iid with density f , then with probability one,

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

converges to θ :

$$P(\hat{\theta}_n \rightarrow \theta \text{ as } n \rightarrow \infty) = 1.$$

Definition: We say then that $\hat{\theta}_n$ is an *consistent estimator* of θ .

1. Show that $\hat{\theta}_n$ is *unbiased*, i.e. $E\hat{\theta}_n = \theta$.
2. Show that a) $\int_0^1 \frac{u}{1+u} du = 1 - \log(2)$ and b) construct a Monte Carlo method for estimating this integral.

Hint: a) Use substitution of u by $x = 1 + u$.

b) Use $Z = U/(1 + U)$ where $U \sim \text{unif}(0,1)$.

3. Use R to estimate the integral for $n = 10, 100, 1000, 10000, 100000, 1000000$. Investigate how the estimates converges.
4. Recall that if the variance $Var(h(X))$ exists, then by the CLT,

$$\sqrt{n}(\hat{\theta}_n - \theta) \overset{\sim}{\rightarrow} N(0, \sigma^2) \quad \text{as } n \rightarrow \infty$$

where $\sigma = \sqrt{Var(h(X))}$. Show that the probability that θ is included in the interval

$$\left[\hat{\theta}_n - \frac{1.96\sigma}{\sqrt{n}}, \hat{\theta}_n + \frac{1.96\sigma}{\sqrt{n}} \right]$$

is approximately 95%; this interval is called a *95% confidence interval* for θ .

5. Consider again the integral in 2. Show that $Var(Z) = 1/2 - (\log(2))^2$. Use this in R to obtain 95% confidence intervals for the estimates in 3.

Hint: Find the derivative of $2(u - \log(u + 1)) - u^2/(1 + u)$ and obtain thereby $E(Z^2)$.

Exercise 2 (Buffon's needle)

The following relates to the webpage <http://www.mste.uiuc.edu/reese/buffon/buffon.html>. See also <http://www.angelfire.com/wa/hurben/buff.html>.

1. Consider the paragraph "The Simplest Case". The problem is not completely formulated because the randomness in the needle drop is not completely specified. a) So how are the random variables D and θ distributed? b) Why is the suggested method for calculating π the same as simple Monte Carlo?

Remark: An actual experiment of this type was carried out by the astronomer R. Wolf in Zurich about 1850 — making this probably the first Monte Carlo procedure. He dropped a needle 5000 times on a ruled grating and got the value 3.1596 for π , about 0.6ths of a percent in error.

2. Answer the questions in the paragraph "Questions".

Exercise 3 (The weighted Monte Carlo method: importance sampling)

Suppose that $\theta = P(X > x)$ is unknown, where X is a random variable and x is a real number. When x is large, θ becomes small and the event $\{X > x\}$ is called a *rare event*. Let $1_A(y)$ denote the *indicator function* for $A \subseteq \mathbb{R}$: $1_A(y) = 1$ if $y \in A$ and $1_A(y) = 0$ if $y \notin A$. Since $\theta = P(X > x) = E1_{(x, \infty)}(X)$, we may suggest to estimate θ by $\hat{\theta}_n$ obtained by simple Monte Carlo. However, as shown below, the weighted Monte Carlo method (also called importance sampling) provides a much better method for estimating rare events.

1. Show that $Var \hat{\theta}_n = \theta(1 - \theta)/n$.
2. If e.g. $\theta = 0.001$, then how large need n to be if we want the standard deviation of $\hat{\theta}_n$ to be at most $\theta/10$? (This indicates that a better method than simple Monte Carlo is needed).

3. Verify the following useful result: If Y_1, \dots, Y_n are iid with density q such that

$$q(x) > 0 \text{ whenever } f(x) > 0$$

where f denotes the density of X , then

$$\tilde{\theta}_n = \frac{1}{n} \sum_{i=1}^n 1_{(x, \infty)}(Y_i) \frac{f(Y_i)}{q(Y_i)}$$

is an unbiased consistent estimator of θ (see Theorem 1).

The weighted Monte Carlo method (or importance sampling) is an evaluation of $\tilde{\theta}_n$; more details are given below.

4. Show that if $q(y) = 1_{(x, \infty)}(y)f(y)/\theta$, then $\tilde{\theta}_n = \theta$. (Of course, since θ is assumed to be unknown, we cannot choose q in this way; however, it shows that ideally q should be chosen such that $q(y)$ is large whenever $y > x$ and $f(y)$ is large).
5. a) Suppose that $f(y) = \exp(-y)$ for $y > 0$ (the standard exponential density) and express $\text{Var } \hat{\theta}_n$ in terms of $x > 0$.
 b) Suppose that $q(y) = (1/x) \exp(-y/x)$ for $y > 0$ (the exponential density with parameter $x > 0$). It can be shown that

$$\text{Var } \tilde{X}_n = \frac{1}{n} \left(\frac{x^2}{2x-1} e^{1-2x} - e^{-2x} \right) \quad \text{if } x > 2$$

while $\text{Var } \tilde{X}_n = \infty$ if $0 < x \leq 2$. Plot $\log(\text{Var } \tilde{\theta}_n / \text{Var } \hat{\theta}_n)$ as a function of $x > 2$.

Importance sampling

Exercise 3 shows a simple example of importance sampling when estimating probabilities. In general importance sampling is based on

Theorem 2 Consider a general setting as at the beginning of Exercise 1, and let Y_1, \dots, Y_n be iid with density q such that

$$q(x) > 0 \text{ whenever } f(x) > 0$$

where f denotes the density of X . Then

$$\tilde{\theta}_n = \frac{1}{n} \sum_{i=1}^n h(Y_i) \frac{f(Y_i)}{q(Y_i)} \tag{1}$$

is an unbiased consistent estimator of $\theta = E(h(X))$.

The method of *importance sampling* is an evaluation of (1). Typical in applications, we use (1) for estimating several mean values corresponding to different h functions. We call f the *target density* and q the *instrumental density* or the *importance sampling density*. Moreover,

$$w(Y_i) = \frac{f(Y_i)}{q(Y_i)}, \quad i = 1, \dots, n,$$

are called the *importance weights*. As demonstrated in Exercise 3, the variation of the importance weights should not be too large: if one or a few of the importance weights can be large compared to the others, importance sampling will in general not work well as the variance of $\tilde{\theta}_n$ can be huge. Particularly, problems may be encountered when $E(f(Y)/q(Y)) = \infty$. Apart from this, there is normally no restriction on the choice of the instrumental density q . Furthermore, the same sample (generated from q) can be used repeatedly for estimating different mean values.

A good discussion on importance sampling (and many other Monte Carlo methods, including inversion and rejection) can be found in

Robert, C. P. and Casella, G. (1999). *Monte Carlo Statistical Methods*. Springer.