Monte Carlo Methods

Lecture notes for MAP001169

Based on Script by Martin Sköld

adopted by Krzysztof Podgórski

Contents

Part I

Simulation and Monte-Carlo Integration

Chapter 1

Simulation and Monte-Carlo integration

1.1 Issues in simulation

1.2 Raw ingredients

Chapter 2

Simulating from specified distributions

2.1 Transforming uniforms

2.2 Transformation methods

2.3 Rejection sampling

2.4 Conditional methods

Chapter 3

Monte-Carlo integration

Many quantities of interest to statisticians can be formulated as integrals,

$$\tau = E(\phi(X)) = \int \phi(x) f(x)\, dx, \tag{3.1}$$

where $X \in \mathbf{R}^d$, $\phi : \mathbf{R}^d \mapsto \mathbf{R}$ and f is the probability density of X. Note that probabilities correspond to ϕ being an indicator function, i.e.

$$P(X \in A) = \int \mathbf{1}\{x \in A\} f(x)\, dx,$$

where

$$\mathbf{1}\{x \in A\} = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{else} \end{cases} \tag{3.2}$$

When dimension n is large and/or ϕf complicated, the integration in (3.1) can often not be performed analytically. Monte-Carlo integration is a numerical method for integration based on the *Law of Large Numbers* (LLN). The algorithm goes as follows:

Algorithm 3.1 (Basic Monte-Carlo Integration)**.**

1. Draw N values $x_1, \ldots, x_N$ independently from f.
2. Approximate $\tau = E(\phi(X))$ by

$$t_N = t(x_1, \ldots, x_N) = \frac{1}{N} \sum_{i=1}^{N} \phi(x_i).$$

As an example of this, suppose we wish to calculate $P(X < 1, Y < 1)$ where (X, Y) are bivariate normal distribution with correlation 0.5 and having standard normal distribution for marginals. This can be written as

$$\int \mathbf{1}\{x < 1, y < 1\} f(x,y)\,dx\,dy \tag{3.3}$$

where f is the bivariate normal density. Thus, provided we can simulate from the bivariate normal, we can estimate this probability as

$$n^{-1} \sum_{i=1}^{n} \mathbf{1}\{x_i < 1, y_i < 1\} \tag{3.4}$$

which is simply the proportion of simulated points falling in the set defined by $\{(x,y); x < 1, y < 1\}$. Here we use the approach from Example **??** for simulating bivariate normals. R code to achieve this is

```
 bvnsim=function(n,m,s,r){
  x=rnorm(n)*s[1]+m[1]
  y=rnorm(n)*s[2]*sqrt(1-r^2)+m[2]+(r*s[2])/s[1]*(x-m[1])
  bvnsim=matrix(0,ncol=2,nrow=n)
  bvnsim[,1]=x
  bvnsim[,2]=y
  bvnsim
}
```

To obtain an estimate of the required probability on the basis of, say, 1000 simulations, we simply need

```
X=bvnsim(1000,c(0,0),c(1,1),.5);
mean((X[,1]<1)&(X[,2]<1))
```

I got the estimate 0.763 doing this. A scatterplot of the simulated values is given in Figure 3.1.

Example 3.1. For a non-statistical example, say we want to estimate the integral

$$\begin{aligned}\tau &= \int_0^{2\pi} x \sin[1/\cos(\log(x+1))]^2\,dx \\ &= \int (2\pi x \sin[1/\cos(\log(x+1))]^2)(\mathbf{1}\{0 \le x \le 2\pi\}/(2\pi))\,dx,\end{aligned}$$

where, of course, the second term of the integrand is the $U[0, 2\pi]$ density function. The integrand is plotted in Figure 3.2, and looks to be a challenge for many numerical methods.

Monte-Carlo integration in R proceeds as follows:

```
x=runif(10000)*2*pi
tn=mean(2*pi*x*sin(1/cos(log(x+1)))^2)
tn
[1] 8.820808
```

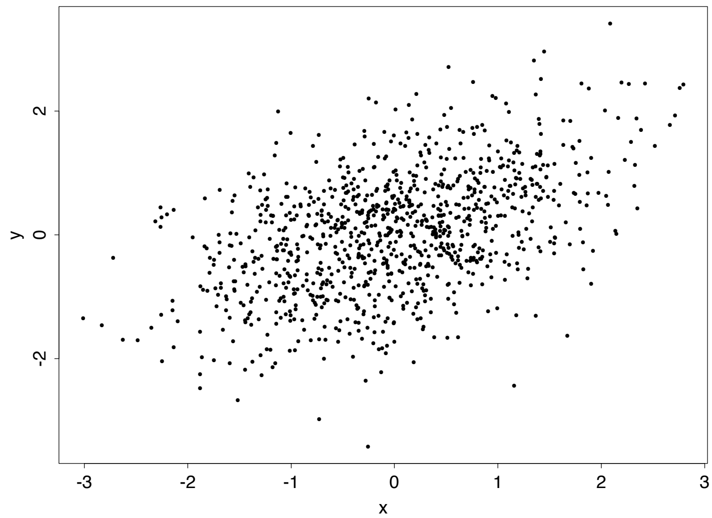

Figure 3.1: Simulated bivariate normals

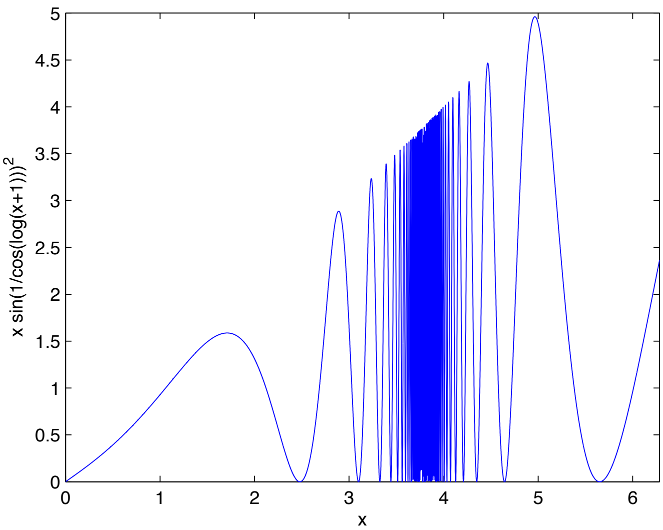

Figure 3.2: An attempt at plotting $x \sin(1/\cos(\log(x+1)))^2$.

Maple, using `evalf` on the integral, gave 8.776170832. A larger run of the Monte-Carlo algorithm shows that this might be an overestimate and that the true value is close to 8.756.

We suggested the motivation comes from the LLN. There are many versions of this celebrated theorem, we will provide a simple mean-square version. First note that if $X_1, \ldots, X_n$ is a sequence of random variables and $T_n = t(X_1, \ldots, X_n)$ for a function t, we say that T_n *converges in the mean square sense* to a fixed value τ if

$$E(T_n - \tau)^2 \to 0 \text{ as } n \to \infty.$$

Theorem 3.1 (A Law of Large Numbers)**.** *Assume $Z_1, \ldots, Z_n$ is a sequence of independent random variables with common means $E(Z_i) = \tau$ and variances $Var(Z_i) = \sigma^2$. If $T_n = n^{-1} \sum_{i=1}^{n} Z_i$, we have*

$$E(T_n - \tau)^2 = \frac{\sigma^2}{n} \to 0 \text{ as } n \to \infty. \tag{3.5}$$

Proof. Simple and straightforward; exercise. □

The above theorem tells us that with $Z_i = \phi(X_i)$ where X_i are independent with density f, the arithmetic mean of $Z_1, \ldots, Z_n$ converges in mean square error to $\tau = E(g(X))$. Moreover, it gives the precise rate of the error: $(E(T_n - \tau)^2)^{1/2} = O(n^{-1/2})$ and this rate is *independent of dimension d*. This is in contrast to deterministic methods for numerical integration, like the trapezoidal rule and Simpson's rule, that have errors of $O(n^{-2/d})$ and $O(n^{-4/d})$ respectively. Monte-Carlo integration is to be preferred in high dimensions (greater than 4 and 8 respectively). Another advantage is that we can reuse the drawn values $x_1, \ldots, x_N$ to estimate other expectations with respect to f without much extra effort.

More precise information on the Monte-Carlo error $(T_n - \tau)$ is given by celebrated result no. 2: the *Central Limit Theorem* (CLT).

Theorem 3.2 (Central Limit Theorem)**.** *Assume $Z_1, \ldots, Z_n$ is a sequence of i.i.d. random variables with common means $E(Z_i) = \tau$ and variances $Var(Z_i) = \sigma^2$. If $T_n = n^{-1} \sum_{i=1}^{n} Z_i$, we have*

$$P\Big(\frac{\sqrt{n}(T_n - \tau)}{\sigma} \le x\Big) \to \Phi(x) \text{ as } n \to \infty, \tag{3.6}$$

where Φ is the distribution function of the $N(0,1)$ distribution.

Proof. Almost as simple, but somewhat less straightforward than LLN. Look it up in a book. □

Slightly less formally, the CLT tells us that the difference $T_n - \tau$ has, at least for large n, approximately an $N(0, \sigma^2/n)$ distribution. With this information we can approximate probabilities like $P(|T_n - \tau| > \epsilon)$, and perhaps more importantly find ϵ such that $P(|T_n - \tau| > \epsilon) = 1 - \alpha$ for some specified confidence level α. To cut this discussion short, the random interval

$$[T_n - 1.96\hat{\sigma}/\sqrt{n}, T_n + 1.96\hat{\sigma}/\sqrt{n}] \tag{3.7}$$

will cover the true value τ with approximately 95% probability. Here $\hat{\sigma}$ is your favourite estimate of standard deviation, e.g. based on

$$\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(z_i - \bar{z})^2, \tag{3.8}$$

and 1.96 is roughly $\Phi^{-1}(0.95)$, the standard Normal 95% quantile.

A similar result to the central limit theorem also holds for the median and general sample quantiles:

Theorem 3.3. *Assume $Z_1, \ldots, Z_n$ is a sequence of i.i.d. random variables with distribution function $F(z - \tau)$ such that $F(0) = \alpha$ and that at zero F has density $f(0) > 0$. Then*

$$P(\sqrt{C_\alpha n}(Z_{(\lceil n\alpha \rceil)} - \tau) \leq x) \to \Phi(x) \text{ as } n \to \infty, \tag{3.9}$$

where $C_\alpha = \alpha(1-\alpha)f^2(0)$ and Φ is the distribution function of the $N(0,1)$ distribution.

Exercise 3.1. *Let (X_1, X_2, X_3) have the trivariate exponential distribution with density proportional to*

$$\exp(-x_1 - 2x_2 - 3x_3 - \max(x_1, x_2, x_3)), \quad x_i > 0, \quad i = 1, \ldots, 3.$$

Construct an algorithm that draws from (X_1, X_2, X_3) using the rejection method, proposing a suitable vector of independent exponentials.

Use basic Monte-Carlo integration to produce an approximate 95% accuracy interval for the probability $P(X_1^2 + X_2^2 \leq 2)$.

Exercise 3.2. *Let $\pi(k)$ be the number of primes less than k. How can you approximate $\pi(10^9)$ without having to check all integers less than 10^9? You* could *use the famous* prime-number theorem*, which says that $\pi(k) \approx k/\log(k)$ for large k. See the following Wikipedia link for more details on historical and mathematical aspects of this result: Prime Number Theorem. We will not this "deterministic result". Instead, let X be uniformly distributed on the odd numbers $\{1, 3, \ldots, 10^9 - 1\}$ (but remember that 2 is also a prime). Let ψ be an indicator of a prime number, i.e. it is a function that takes value one if its argument is prime and zero otherwise.*

Find the (simple) relation between the expected value $E(\psi(X))$ and $\pi(10^9)$. Then use Monte-Carlo method to approximate $\pi(10^9)$ by sampling $X_1, \ldots, X_n$ from X averaging $\psi(X_i)$, $i = 1, \ldots, n$. By what a result in probability theory averaging approximates the expected value of $E(\psi(X))$. You might find R package 'primes' with its `is_prime` *function useful here. Provide with the error assessment. Compare your result with the prime-number theorem.*

Bias and the Delta method

It is not always that we can find a function t_n such that $E(T_n) = \tau$. For example we might be interested in $\tau = h(E(X))$ for some specified smooth function h. If $\bar{X}$ again is the arithmetic mean, then a natural choice is $T_n = h(\bar{X})$. However, unless h is linear, $E(T_n)$ is not guaranteed to equal τ. This calls for a definition: the *bias* of t (when viewed as an estimator of τ), $T_n = t(X_1, \ldots, X_n)$ is

$$\text{Bias}(t) = E(T_n) - \tau. \tag{3.10}$$

The concept of bias allows us to more fully appreciate the concept of mean square error, since

$$E(T_n - \tau)^2 = \text{Var}(T_n) + \text{Bias}^2(t), \tag{3.11}$$

(show this as an exercise). The mean square error equals variance plus squared bias. In the above mentioned example, a Taylor expansion gives an impression of the size of the bias. Roughly we have with $\mu = E(X)$

$$\begin{aligned} E(T_n - \tau) &= E[h(\bar{X}) - h(\mu)] \\ &\approx E(\bar{X} - \mu)h'(\mu) + \frac{E(\bar{X} - \mu)^2}{2} h''(\mu) \\ &= \frac{\text{Var}(X)}{2n} h''(\mu). \end{aligned} \tag{3.12}$$

And it is reassuring that (3.12) suggests a small bias when sample size n is large. Moreover, since variance of T_n generally is of order $O(n^{-1})$ it will dominate the $O(n^{-2})$ squared bias in (3.11) suggesting that bias is a small problem here (though it can be a serious problem if the above Taylor expansions are not valid).

We now turn to the variance of T_n. First note that while $\text{Var}(\bar{X})$ is easily estimated by e.g. (3.8), estimating $\text{Var}(h(\bar{X}))$ is not so straightforward. An useful result along this line is the *Delta Method*

Theorem 3.4 (The Delta method)**.** *Let r_n be an increasing sequence and S_n a sequence of random variables. If there is μ such that h is differentiable at μ and*

$$P(r_n(S_n - \mu) \le x) \to F(x), \text{ as } n \to \infty$$

for a distribution function F, then

$$P(r_n(h(S_n) - h(\mu)) \le x) \to F(x/|h'(\mu)|).$$

Proof. Similar to the Taylor expansion argument in (3.12). □

This theorem suggests that if $S_n = \bar{X}$ has variance σ^2/r_n, then the variance of $T_n = h(S_n)$ will be approximately $\sigma^2 h'(\mu)^2/r_n$ for large n. Moreover, if S_n is asymptotically normal, so is T_n.

Exercise 3.3. *Implement Monte Carlo evaluation of integral*

$$I = \int_0^{2\pi} x^{2|sinx|} e^{x\cos^{3/2}x} dx.$$

Analyze the error of your evaluation. Suppose that one is interested in the accuracy of I^{-2} *from the obtained approximation of* I*. Apply the delta method to assess this accuracy.*