Monte Carlo Integration

Suppose we want to evaluate $\theta = \int_x \phi(x)f(x)dx$, and that it is analytically impossible to get a closed form representation for it. Suppose also that it is not possible, or very difficult to evaluate numerically θ. One way is to use simulation in what is called Monte Carlo integration. This is a straightforward consequence of the Strong Law of Large Numbers:

Theorem 0.1 *A sequence of random variables $X_1, X_2, \dots$ with finite expectations in a probability space is said to satisfiy the strong law of large numbers if $\frac{1}{n}\sum_{k=1}^{n}(X_k - \mathrm{E}\, X_k) \xrightarrow{a.s.} 0,$, where a.s. stands for convergence almost surely.*

Let $X_1, X_2, \dots$ be a sequence of independent random variables, with finite expectations. The strong law of large numbers holds if one of the following conditions is satisfied:

1. *The random variables are identically distributed;*
2. *For each n, the variance of X_n is finite, and $\sum_{n=1}^{\infty} \frac{\mathrm{Var}(X_n)}{n^2} < \infty$.*

The above form Kolmogorov's Strong Law of Large Numbers theorems.

Operationally, we take a large enough sample of independent[1] identically distributed observations $X_1, X_2 \dots, X_n$ from a density $f(x)$. Then using the above, and of course assuming that θ exists, yields the estimator $\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} \phi(X_i)$ as the Monte Carlo approximation of θ.

Note that we do have several choices of ϕ and f with our preferred choice being one that provides low standard error.

Example: Suppose we want to evaluate $\theta = \Pr(C > 2)$, where $C \sim Cauchy(0, 1)$. Analytically, we have the θ given by $1 - F(2) = 1/2 - \arctan 2 \approx 0.1476$.

Suppose we want to use Monte Carlo integration to calculate the above. There are quite a few options, provided by different choices of ϕ and f.

1. Let $X_1, X_2 \dots, X_n$ be a simple random sample from $C(0, 1)$. Then

$$\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} \mathrm{I}(X_i > 2);$$

   and for this estimate, $\mathrm{Var}(\hat{\theta}) \approx \frac{0.126}{n}$. In this case, $\phi(x) = \mathrm{I}(x > 2)$.

[1] This is not really necessary, and can be relaxed, and indeed is, in the case of Markov Chain Monte Carlo.

2. An alternative to the above would to define $\phi(x) = \frac{1}{2}\,\mathrm{I}(|x| > 2)$, and proceed with the same sampling scheme as above.

$$\hat{\theta} = \frac{1}{2n}\sum_{i=1}^{n}\mathrm{I}(|X_i| > 2);$$

   In this case, we note that $\mathrm{Var}(\hat{\theta}) \approx \frac{0.052}{n}$. Clearly this, this estimate has more efficiency over the previous one, and results in a variance reduction of about 2.4 times. Hence, the Monte Carlo integral as here has greater accuracy.

3. Note that

$$1 - 2\theta = \int_{-2}^{2} f(x)dx = 2\int_{0}^{2} f(x)dx.$$

   Hence we can perform Monte Carlo integration by drawing an independent, identically distributed sample $X_1, X_2, \ldots, X_n$ from $U(0, 2)$, and taking$\phi(x) = 2f(x)$. In this case, we can show that $\mathrm{Var}(\hat{\theta}) \approx \frac{0.028}{n}$.

4. Also,

$$\theta = \int_{2}^{\infty}\frac{dx}{\pi(1+x^2)} = \int_{0}^{\frac{1}{2}}\frac{y^{-2}dy}{\pi(1+y^{-2})} \equiv \int_{0}^{\frac{1}{2}} g(y)dy.$$

   So, we obtain independently identically distributed samples $X_1, X_2, \ldots, X_n$ from $U(0, \frac{1}{2})$, and take $\phi(x) = \frac{f(x)}{2}$. Then, $\mathrm{Var}(\hat{\theta}) \approx \frac{9.3\times 10^{-5}}{n}$, with a variance reduction of about 1,350 times!

This example, and its different implementations provided by Ripley (1987) indicates that the choice of f and ϕ can greatly influence the efficiency of the Monte Carlo integration.

Hit-or-Miss Monte Carlo

Let $\phi(x)$ be such that it is bounded on $[a, b]$ where $0 \leq \phi \leq c$. Then, in order to evalute $\theta = \int_a^b \phi(x)dx$, we can write

$$\theta = c(b-a) \times \text{proportion of } [a, b] \times [0, c] \text{ under}\phi.$$

This means that

$$\theta = c(b-a)\Pr(V \leq \phi(U))$$

where $U \sim U(a, b)$ and $V \sim U(0, c)$. We can this get n independent samples $(U_i, V_i); i = 1, 2, \ldots, n$ and our Monte Carlo integral estimate to be:

$$\tilde{\theta} = c(b-a) \times \text{proportion of cases for which } V_i < \phi(U_i).$$

Then, $\mathrm{E}\,\hat{\theta} = \theta$ and $\mathrm{Var}(\hat{\theta}) = \theta\frac{c(b-a)-\theta}{n}$. However, we have the

Theorem 0.2 *Suppose* $\theta = \int_a^b \phi(x)dx$*; where* $0 \leq \phi(x) \leq c$*. Then,*

$$\mathrm{Var}(\tilde{\theta}) \leq \mathrm{Var}(\hat{\theta}),$$

with equality if and only if $\phi(x) \equiv c$*,*