Chapter 6: Monte Carlo Integration

Monte Carlo integration is a statistical method based on <u>random sampling</u> in order to approximate integrals. This section could alternatively be titled,

"Integrals are hard, how can we avoid doing them?"



https://xked.com/2117/

1 A Tale of Two Approaches

Consider a one-dimensional integral.

 $\int_{a}^{b} f(z) dz$

The value of the integral can be derived analytically only for a few functions, f. For the rest, numerical approximations are often useful.

Why is integration important to statistics?

Many quantities of interest in informatical statistics can be expressed as the expectation Many quantities of interval. of a function of a random variable. $E[g(X)] = \int g(X) f(X) dx$ $\frac{1}{X}$ integrand.

1.1 Numerical Integration

Idea: Approximate $\int_a^b f(x) dx$ via the sum of many polygons under the curve f(x).

To do this, we could partition the interval [a, b] into m subintervals $[x_i, x_{i+1}]$ for $i=0,\ldots,m-1$ with $x_0=a$ and $x_m=b$.

Within each interval, insert k+1 nodes, so for $[x_i, x_{i+1}]$ let x_{ij}^* for $j=0,\ldots,k,$ then

$$\int\limits_a^b f(x)dx = \sum\limits_{i=0}^{m-1}\int\limits_{x_i}^{x_{i+1}}f(x)dx pprox \sum\limits_{i=0}^{m-1}\sum\limits_{j=0}^k A_{ij}f(x_{ij}^*)$$
theights.

for some set of constants, A_{ij} . f(7*;) 0.4 -0.3 -× 0.2-0.1 -0.0 --2 2 x, × 2, x,=a X_m=b. X4

1.2 Monte Carlo Integration

How do we compute the mean of a distribution?

Example 1.1 Let $X \sim Unif(0, 1)$ and $Y \sim Unif(10, 20)$.

```
x <- seq(0, 1, length.out = 1000)
f <- function(x, a, b) 1/(b - a)
ggplot() +
  geom_line(aes(x, f(x, 0, 1))) +
  ylim(c(0, 1.5)) +
  ggtitle("Uniform(0, 1)")

y <- seq(10, 20, length.out = 1000)
ggplot() +
  geom_line(aes(y, f(y, 10, 20))) +
  ylim(c(0, 1.5)) +
  ggtitle("Uniform(10, 20)")</pre>
```



$$E(x) = \int_{0}^{\infty} x \cdot f(x) dx \qquad E[Y] = \int_{10}^{\infty} y f(y) dy \\ = \int_{0}^{\infty} x \cdot 1 dx \qquad = \int_{10}^{\infty} \frac{y}{10} dy \\ = \left[\frac{x^{2}}{2}\right]_{0}^{1} = \frac{1}{d}. \qquad = \frac{1}{10} \left[\frac{y^{2}}{2}\right]_{10}^{20} = 15.$$

1.2.1 Notation

 $\theta = parameter of interest (unknown).$ $\hat{\theta} = estimator of <math>\theta$, statistic (sometimes we write \bar{X} , s^2 , etc. instead $\neq \hat{\Theta}$).

Distribution of $\hat{\theta} = \text{Sampling distribution}$.

$$E[\hat{\theta}] = \text{theoretical mean of the scorphized days of } \hat{\theta}$$

"on average, what is the value of $\hat{\theta}$?"
 $Var(\hat{\theta}) = Variance of scorphized days of $\hat{\theta}$$

$$\begin{array}{l} & \hat{E}[\hat{\theta}] = \text{ estimated trees of sampling din q } \hat{\theta} \\ \\ & \hat{V}\hat{ar}(\hat{\theta}) = \text{ estimated variance of sampling din of } \hat{\theta}. \\ & \hat{se}(\hat{\theta}) = \sqrt{Var(\hat{\theta})} \\ & \hat{se}(\hat{\theta}) = \sqrt{Var(\hat{\theta})}. \end{array}$$

1.2.2 Monte Carlo Simulation

What is Monte Carlo simulation?

Computer simulation that generatic a large # of scopple from a dsn. The dsn characterizes the population from which a sample is drawn. (Sound's like Ch. 3).

1.2.3 Monte Carlo Integration

To approximate $\theta = E[X] = \int x f(x) dx$, we can obtain an iid random sample X_1, \ldots, X_n from f and then approximate θ via the sample average

Example 1.2 Again, let $X \sim Unif(0, 1)$ and $Y \sim Unif(10, 20)$. To estimate E[X] and E[Y] using a Monte Carlo approach,

Now consider E[g(X)].

$$heta=E[g(X)]=\int\limits_{-\infty}^{\infty}g(x)f(x)dx.$$

The Monte Carlo approximation of θ could then be obtained by

1.

2.

Definition 1.1 *Monte Carlo integration* is the statistical estimation of the value of an integral using evaluations of an integrand at a set of points drawn randomly from a distirbution with support over the range of integration.

Example 1.3

Why the mean?

Let $E[g(X)] = \theta$, then

and, by the strong law of large numbers,

Example 1.4 Let $v(x) = (g(x) - \theta)^2$, where $\theta = E[g(X)]$, and assume $g(X)^2$ has finite expectation under f. Then

$$Var(g(X)) = E[(g(X) - \theta)^2] = E[v(X)].$$

We can estimate this using a Monte Carlo approach.

1.2 Monte Carlo Integration

Monte Carlo integration provides slow convergence, i.e. even though by the SLLN we know we have convergence, it may take us a while to get there.

But, Monte Carlo integration is a **very** powerful tool. While numerical integration methods are difficult to extend to multiple dimensions and work best with a smooth integrand, Monte Carlo does not suffer these weaknesses.

•

1.2.4 Algorithm

The approach to finding a Monte Carlo estimator for $\int g(x)f(x)dx$ is as follows.

1.

2.

3.

4.

Example 1.5 Estimate $\theta = \int_0^1 h(x) dx$.

Example 1.6 Estimate $\theta = \int_a^b h(x) dx$.

Another approach:

Example 1.7 Monte Carlo integration for the standard Normal cdf. Let $X \sim N(0, 1)$, then the pdf of X is

$$\phi(x) = f(x) = rac{1}{\sqrt{2\pi}} \mathrm{exp}igg(-rac{x^2}{2}igg), \qquad -\infty < x < \infty$$

and the cdf of X is

$$\Phi(x)=F(x)=\int\limits_{\infty}^{x}rac{1}{\sqrt{2\pi}} ext{exp}igg(-rac{t^{2}}{2}igg)dt.$$

We will look at 3 methods to estimate $\Phi(x)$ for x > 0.

1.2 Monte Carlo Integration

1.2.5 Inference for MC Estimators

The Central Limit Theorem implies

So, we can construct confidence intervals for our estimator

1.

2.

But we need to estimate $Var(\hat{\theta})$.

So, if $m \uparrow \text{then } Var(\hat{\theta}) \downarrow$. How much does changing *m* matter?

Example 1.8 If the current $se(\hat{\theta}) = 0.01$ based on *m* samples, how many more samples do we need to get $se(\hat{\theta}) = 0.0001$?

Is there a better way to decrease the variance? Yes!