Monte-Carlo method II

variance reduction technique

\- importance sampling-

Woojoo Lee

Monte-Carlo integration can have large variation.

This is about the efficiency problem.

How can we imporve efficiency, i.e. reducing the variance of Monte-Carlo estimate ?

Importance sampling

Suppose that $\theta(= \int h(x)f(x)dx)$ is the parameter of interest.
The key idea is using the following expression:

$$\theta = \int (h(x)f(x)/g(x))g(x)dx.$$

Then, we can estimate θ by using

$$\hat{\theta}_g = \frac{1}{n}\sum_{i=1}^{n} \psi(Y_i)$$

where $\psi(x) = h(x)f(x)/g(x)$ and $Y_i \sim g(x)$.

Q) Compute the variance of $\hat{\theta}_g$.

$\rightarrow$ Our aim is to select g which minimizes $\mathrm{Var}(\hat{\theta}_g)$

When does $\frac{1}{n}\sum_{i=1}^{n}(h(Y_i)\frac{f(Y_i)}{g(Y_i)})$ have smaller variance than $\frac{1}{n}\sum_{i=1}^{n} h(X_i)$?

If we note that these two random quantities have the same mean, this problem is:

Find $g(x)$ satisfying

$$\int (h(x)\frac{f(x)}{g(x)})^2 g(x)dx \leq \int (h(x))^2 f(x)dx$$

Ripley (1987).
Suppose that X follows a Cauchy distribution:

$$f(x) = \frac{1}{\pi(1+x^2)}.$$

We want to find $P(X > 1.9)$. Note that

$$P(X > 1.9) = \int_{1.9}^{\infty} f(x)dx = \int_{-\infty}^{\infty} 1_{(x>1.9)} f(x)dx$$

Q) Compute $P(X > 1.9)$ directly.

Q) Find a Monte-Carlo approach for computing $P(X > 1.9)$.

Q) Compute the variance of the suggested estimator for $P(X > 1.9)$.

Now consider the importance sampling.

$$\begin{aligned} P(X > 1.9) &= \int_{-\infty}^{\infty} 1_{(x>1.9)} f(x) dx \\ &= \int_{-\infty}^{\infty} 1_{(x>1.9)} \frac{f(x)}{g(x)} g(x) dx \end{aligned}$$

We note that for large x,

$$f(x) \propto \frac{1}{1+x^2} \approx \frac{1}{x^2}.$$

Q) Suggest a good canidadate for $g(x)$.

Q) Suggest how to generate random variables from $g(x)$.

From the importance sampling, out estimator is

$$\hat{\theta}_g = \frac{1}{n}\sum_{i=1}^{n}\frac{1_{(Y_i>1.9)}f(Y_i)}{g(Y_i)}$$

where $Y_i \sim g(x)$.

Q) Compute the variance of $\hat{\theta}_g$ analytically.

Implementation in R

The usual Monte-Carlo Approach :

```
#### P(X>1.9) where X follows Cauchy(0,1)
#### true value = 0.15-atan(1.9)/pi = 0.1542

cau19<-function(n){
        x<-rcauchy(n,0,1)
        prob<-mean((x>1.9))
        return(prob)
}
```

The importance sampling approach :

```
imp19<-function(n){
        x<-1.9/runif(n,0,1)
        prob<-mean((x>1.9)*(1/(pi*(1+x^2)))/(1.9/x^2))
        return(prob)
}
```

Pros and Cons of importance sampling

Pros

1. Generally applicable
2. Not always, but often available in multi-dimensional problems

Cons

1. Still difficult to find a proper $g(x)$.

When the normalizing constant of $f(x)$ is unknown, can we still use the importance sampling technique ?