STA 5106 Computational Methods in Statistics I

Department of Statistics
Florida State University

Class 19
November 7, 2019

Chapter 6

Monte Carlo Methods

6.1 Introduction

Monte Carlo Method

- Integration plays a very important role in statistical inference. Many inference problems can be written as integrals under some given probability measure.

- In many situations, this probability measure is too difficult to analytically integrate out, and hence, numerical approximations are used.

- One technique for numerical approximation is via sampling, that is, to approximate the integral using samples generated from the given probability measure.

- This technique is called **Monte Carlo method** and is gaining wide acceptance among the statisticians.

6.2 Monte Carlo Method

Goal

- The main goal in this technique is to estimate the quantity Θ, where

$$\Theta = \int g(x) f(x) dx = E[g(X)],$$

for a random variable X distributed according to the density function $f(x)$.

- $g(x)$ is any function on $\mathbf{R}$ such that and $E[g(X)^2]$ are bounded.

- Suppose that we have tools to simulate independent and identically distributed samples from $f(x)$, call them $X_1, X_2, \ldots, X_n$, then one can approximate Θ by the quantity:

$$\hat{\Theta}_n = \frac{1}{n} \sum_{i=1}^{n} g(X_i).$$

Unbiased Estimator

- $\hat{\Theta}_n$ is an unbiased estimator of Θ because:

$$E(\hat{\Theta}_n) = \frac{1}{n}\sum_{i=1}^{n} E[g(X_i)] = \frac{1}{n}\sum_{i=1}^{n} \Theta = \Theta$$

- The variance of the estimator is:

$$\begin{aligned}
\mathrm{var}(\hat{\Theta}_n) &= E[(\hat{\Theta}_n - \Theta)^2] \\
&= E[(\frac{1}{n}\sum_{i=1}^{n}(g(X_i) - \Theta))^2] \\
&= \frac{1}{n} E[(g(X) - \Theta)^2] \\
&= \frac{1}{n}\mathrm{var}(g(X))
\end{aligned}$$

Convergence

- Therefore, as n gets larger the variance of $\hat{\Theta}_n$ goes down to zero and it converges to the mean value Θ.
- This setup is called the **classical Monte Carlo approach** where the samples from $f(x)$ are generated in an i.i.d. fashion.
- It is possible to obtain better estimators, than the classical estimator, by reducing the variance of the estimator.
- In the next section we describe two techniques to reduce the variance.

6.3 Variance Reduction Techniques

Variance Reduction by Conditioning

- Let Y and Z be two random variables:

$$var(Y) = E((Y - E(Y))^2) = E(Y^2) - E(Y)^2$$

$$var(Y|Z) = E((Y - E(Y|Z))^2|Z)$$

$$= E(Y^2|Z) - (E(Y|Z))^2$$

- Reorganizing these equations and using law of nested expectations, we obtain

$$E(var(Y|Z)) = E(Y^2) - E((E(Y|Z))^2)$$

$$var(E(Y|Z)) = E((E(Y|Z))^2) - E(Y)^2$$

- Adding the two equations

$$var(Y) = E(var(Y|Z)) + var(E(Y|Z)).$$

Variance Reduction by Conditioning

- Now compare the two random variables Y and $E(Y|Z)$, both have the same means but comparing the variances

$$var(Y) \geq var(E(Y|Z)).$$

- Therefore $E(Y|Z)$ is a better random variable to simulate and average to estimate Θ.

- Of course, an important issue is how to find an appropriate Z such that $E(Y|Z)$ has significantly lower variance than Y.

Variance Reduction using Control Variates

- Our goal is to estimate Θ, the expected value of a function g of random variables X.
- Assume that we know the expected value of another function f of these random variables, call it μ.
- For any constant a, define a random variable W according to
$$W = g(X) + a(f(X) - \mu) .$$
- We can utilize the sample averages of W to estimate Θ since $E(W) = \Theta$.
- Studying the variance of W
$$var(W) = var(g(X)) + a^2 var(f(X)) + 2a\operatorname{cov}(g(X), f(X)) .$$

Variance Reduction using Control Variates

- Considering *var*(*W*) as a function of *a*, and looking for the minimizer we get

$$a = \frac{-\operatorname{cov}(g(X), f(X))}{\operatorname{var}(f(X))}.$$

- The resulting variance of *W* is given by

$$\operatorname{var}(W) = \operatorname{var}(g(X)) - \frac{\operatorname{cov}(f(X), g(X))^2}{\operatorname{var}(f(X))}.$$