STA 5106
Computational Methods in Statistics I

Department of Statistics
Florida State University

Class 20
November 12, 2019

Review: Monte Carlo Method

- The main goal in the **Monte Carlo method** is to estimate the quantity Θ, where
$$\Theta = \int g(x) f(x) dx = E[g(X)],$$
for a random variable X distributed with the density $f(x)$.
- $X_1, X_2, \ldots, X_n$ are i.i.d. samples from $f(x)$. Then one can approximate Θ by the quantity: $\hat{\Theta}_n = \frac{1}{n} \sum_{i=1}^{n} g(X_i).$
- The mean
$$E(\hat{\Theta}_n) = \frac{1}{n} \sum_{i=1}^{n} E[g(X_i)] = \frac{1}{n} \sum_{i=1}^{n} \Theta = \Theta$$
- The variance
$$\text{var}(\hat{\Theta}_n) = E[(\hat{\Theta}_n - \Theta)^2] = \frac{1}{n} \text{var}(g(X))$$

Review: Variance Reduction Methods

- **Variance Reduction by Conditioning:** Y and $E(Y|Z)$ have the same means but for the variances
$$var(Y) \geq var(E(Y|Z)).$$
Therefore $E(Y|Z)$ is a better random variable to estimate Θ.

- **Variance Reduction using Control Variates.** Assume $E(f(X)) = \mu$ is known for a given function f. Let
$$W = g(X) + a(f(X) - \mu),$$
where
$$a = \frac{-\operatorname{cov}(g(X), f(X))}{\operatorname{var}(f(X))}.$$
Then
$$\operatorname{var}(W) = \operatorname{var}(g(X)) - \frac{\operatorname{cov}(f(X), g(X))^2}{\operatorname{var}(f(X))}.$$

6.4 Importance Sampling

Importance Sampling

- Another technique commonly used for reducing variance in Monte Carlo methods is **importance sampling**.

- Instead of sampling from $f(x)$, the importance sampling samples from another density $h(x)$, and computes the estimate of Θ using averages of $g(x)f(x)/h(x)$ instead of $g(x)$ evaluated on those samples.

- Mathematically, we can rearrange the definition of Θ as follows:

$$\Theta = \int g(x) f(x) dx = \int \frac{g(x) f(x)}{h(x)} h(x) dx.$$

- $h(x)$ can be any density function as long as the *support* of $h(x)$ contains the *support* of $f(x)$.

Estimate

- Generate samples $X_1, X_2, \ldots, X_n$ from the density $h(x)$ and compute the estimate:

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

- It can be seen that the mean of $\hat{\Theta}_n$ is Θ and its variance is:

$$\mathrm{var}(\hat{\Theta}_n) = \frac{1}{n}\left(Var_h\left(\frac{g(X) f(X)}{h(X)}\right)\right).$$

- It is possible that a suitable choice of h can reduce the estimator variance below that of the classical Monte Carlo estimator.

Example

- **Example 5** Let X be a Cauchy random variable with parameters (0,1), i.e. X has the density function:

$$f(x) = 1/[\pi(1 + x^2)],$$

and $g(x) = 1_{\{x>2\}}$ be an indicator function. We are interested in estimating:

$$\Theta = \int g(x) f(x) dx = P\{X > 2\}.$$

We will do so using the notion of importance sampling although it is not too difficult to compute the exact value of Θ analytically to be $(\mathrm{atan}(\infty) - \mathrm{atan}(2))/\pi = 0.15$.

(note: the primitive function of $1/(1+ x^2)$ is $\mathrm{atan}(x)$).

Example

- Consider $X_1, X_2, \ldots, X_n$ to be i.i.d. samples from the Cauchy density $f(x)$ and with $g(x)$ being the indicator function, $\hat{\Theta}_n$ is just the frequency of sampled values larger than 2:
$$\hat{\Theta}_n = \frac{1}{n}\sum_{i=1}^{n} 1_{\{X_i > 2\}}.$$
Variance of this estimator is simply $\Theta(1-\Theta)/n$ or $0.126/n$.

- Alternative, we can utilize the fact that the density $f(x)$ is symmetric around 0, and Θ is just half of the probability $\Pr\{|X| > 2\}$. Assuming X_i's to be i.i.d. Cauchy, we have
$$\hat{\Theta}_n = \frac{1}{2n}\sum_{i=1}^{n} 1_{\{|X_i| > 2\}},$$
and the variance of this estimator is $\Theta(1-2\Theta)/(2n)$ or $0.052/n$.

Example

- If we rewrite Θ as the following integral:

$$\Theta = \frac{1}{2} - \int_0^2 \frac{1}{\pi(1+x^2)} dx,$$

- We can obtain another Monte Carlo estimator of Θ. Let X_1, $X_2, \ldots, X_n$ be samples from a uniform random variable taking values between 0 and 2, and define an estimator:

$$\hat{\Theta}_n = \frac{1}{2} - \frac{1}{n} \sum_{i=1}^{n} \frac{2}{\pi(1+X_i^2)}$$

- One can show that this estimator is also unbiased and its variance is given by: $0.028/n$.

Example

- If we rewrite Θ as the integral:

$$\Theta = \int_0^{1/2} \frac{x^{-2}}{\pi(1+x^{-2})} dx,$$

Using i.i.d samples from U[0, 1/2] and evaluating average of the function $g(x) = 1/[2\pi(1+x^2)]$ one can further reduce the estimator variance.

Example

- This example shows that a suitable split of the integrand between $f(x)$ and $g(x)$ can lead to a reduction in variance.
- The question is: what should be new density $h(x)$ that leads to an estimator with minimum variance?
- Theoretically, it is easier to answer that question, i.e. it is easy to write down the optimal $h(x)$, but it may not be easy to sample from this optimal density.
- In practice, one can try many different densities and use experiments to select the best one.