

# Importance Sampling Techniques for Variance Reduction

AMS 553: Simulation Modeling and Analysis

Kai Li, Wenbo Du, Zhe Zhou, Zeyu Dong

Applied Mathematics and Statistics, Stony Brook University

December 1, 2022

# Introduction

We will use the Monte Carlo method to find the average value of a function  $g(x)$  over an interval  $(a, b)$ . This is given by

$$\frac{1}{b-a} \int_a^b g(x) dx.$$

The simple Monte Carlo method generates a large number of replicates  $X_1, \dots, X_m$  uniformly distributed on  $[a, b]$  and estimates the average by the sample mean

$$\frac{1}{m} \sum_{i=1}^m g(X_i).$$

This converges to  $\frac{1}{b-a} \int_a^b g(x) dx$  with probability 1 by the strong law of large numbers.

# Introduction

There are two limitations for this method:

- ① It cannot apply to unbounded intervals.
- ② It could be inefficient to draw samples uniformly across the interval if the function  $g(x)$  is not very uniform.

In this project, we will introduce three different sampling methods: importance sampling, stratified sampling, and stratified importance sampling to reduce variance in Monte Carlo simulations. We will analyze the variance reduction for each method and compare their performance in simulations.

# Importance Sampling

Suppose  $X$  is a random variable with density function  $f(x)$  supported on  $A$ , then

$$\int_A g(x)dx = \int_A \frac{g(x)}{f(x)} f(x)dx = \mathbb{E}_f \left[ \frac{g(X)}{f(X)} \right].$$

Estimate  $\mathbb{E}_f[g(X)/f(X)]$  by simple Monte Carlo integration. That is, compute the average

$$\frac{1}{m} \sum_{i=1}^m \frac{g(X_i)}{f(X_i)},$$

where the random variables  $X_1, \dots, X_m$  are generated from the distribution with density  $f(x)$ . The density  $f(x)$  is called the importance function.

# Importance Sampling

The distribution of  $X$  can be chosen to reduce the variance of the sample mean estimator. The minimum variance is obtained when

$$f(x) = \frac{|g(x)|}{\int_A |g(x)| dx}.$$

This result shows when choosing the distribution  $f(x)$ , we want to make sure the value of  $\frac{g(x)}{f(x)}$  is close to a constant, or intuitively, we want  $f(x)$  to have a similar “shape” as  $g(x)$ . In the meanwhile, though, the variable with density  $f(\cdot)$  should be reasonably easy to simulate.

# Stratified Sampling

Another approach to variance reduction is stratified sampling. This method aims to reduce the variance of the estimator by dividing the interval into strata and estimating the integral on each stratum with a smaller variance. Linearity of the integral operator and the strong law of large numbers implies that the sum of these estimates converges to  $\int g(x)dx$  with probability 1. In stratified sampling, the number of replicates  $m$  and number of replicates  $m_j$  to be drawn from each of  $k$  strata are fixed so that  $m = m_1 + \dots + m_k$ , with the goal that

$$\text{Var} \left( \hat{\theta}_k(m_1, \dots, m_k) \right) < \text{Var} \left( \hat{\theta} \right),$$

where  $\hat{\theta}_k(m_1, \dots, m_k)$  is the stratified estimator and  $\hat{\theta}$  is the standard Monte Carlo estimator based on  $m = m_1 + \dots + m_k$  replicates.

# Stratified Sampling

## Proposition

Denote the standard Monte Carlo estimator with  $M$  replicates by  $\hat{\theta}^M$ , and let

$$\hat{\theta}^S = \frac{1}{k} \sum_{j=1}^k \hat{\theta}_j$$

denote the stratified estimator with equal size  $m = M/k$  strata. Denote the mean and variance of  $g(U)$  on stratum  $j$  by  $\theta_j$  and  $\sigma_j^2$ , respectively.

Then  $\text{Var}(\hat{\theta}^M) \geq \text{Var}(\hat{\theta}^S)$ .

There can be more reduction in variance using stratification when the means of the strata are widely dispersed than if the means of the strata are approximately equal. For integrands that are monotone functions, stratification should be an effective way to reduce variance.

# Stratified Importance Sampling

Stratified importance sampling is a modification to the importance sampling method of estimation  $\theta = \int g(x)dx$ .

Choose a suitable importance function  $f$ . Suppose that  $X$  is generated with density  $f$  and cdf  $F$  using the probability integral transformation. If  $M$  replicates are generated, the importance sampling estimate  $\theta$  has variance  $\sigma^2/M$ , where  $\sigma^2 = \text{Var}(g(X)/f(X))$ .



# Stratified Importance Sampling

For the stratified importance sampling estimate, divide the real line into  $k$  intervals  $I_j = \{x : a_{j-1} \leq x < a_j\}$  with endpoints  $a_0 = -\infty$ ,  $a_j = F^{-1}(j/k)$ , where  $j = 1, \dots, k-1$ , and  $a_k = \infty$ . On each subinterval define  $g_j(x) = g(x)$  if  $x \in I_j$  and  $g_j(x) = 0$  otherwise. Now, we have  $k$  parameter to estimate,

$$\theta_j = \int_{a_{j-1}}^{a_j} g_j(x) dx, \quad j = 1, \dots, k$$

and  $\theta = \theta_1 + \dots + \theta_k$ . The conditional densities provide the importance functions on each subinterval. That is, on each subinterval  $I_j$ , the conditional density  $f_j$  of  $X$  is defined by

$$\begin{aligned} f_j(x) &= f_{X|I_j}(x|I_j) = \frac{f(x, a_{j-1} \leq x < a_j)}{\mathbb{P}(a_{j-1} \leq x < a_j)} \\ &= \frac{f(x)}{1/k} = kf(x), \quad a_{j-1} \leq x < a_j. \end{aligned}$$

# Stratified Importance Sampling

Let  $\sigma_j^2 = \text{Var}(g_j(X)/f_j(X))$ . For each  $j = 1, \dots, k$  we simulate an importance sample size  $m$ , compute the importance sampling estimator  $\hat{\theta}_j$  of  $\theta_j$  on the  $j$ th subinterval, and compute  $\hat{\theta}^{SI} = \frac{1}{k} \sum_{j=1}^k \hat{\theta}_j$ . Then by independence of  $\hat{\theta}_1, \dots, \hat{\theta}_k$ ,

$$\text{Var}(\hat{\theta}^{SI}) = \text{Var}\left(\sum_{j=1}^k \hat{\theta}_j\right) = \sum_{j=1}^k \frac{\sigma_j^2}{m} = \frac{1}{m} \sum_{j=1}^k \sigma_j^2.$$

# Stratified Importance Sampling

Denote the importance sampling estimator by  $\hat{\theta}^I$ . In order to determine whether  $\hat{\theta}^{SI}$  is better estimator of  $\theta$  than  $\hat{\theta}^I$ , we need to check that  $\text{Var}(\hat{\theta}^{SI})$  is smaller than the variance without stratification. The variance is reduced by stratification if

$$\frac{\sigma^2}{M} > \frac{1}{m} \sum_{j=1}^k \sigma_j^2 = \frac{k}{M} \sum_{j=1}^k \sigma_j^2 \implies \sigma^2 - k \sum_{j=1}^k \sigma_j^2 > 0.$$

# Stratified Importance Sampling

## Proposition

*Suppose  $M = mk$  is the number of replicates for an importance sampling estimator  $\hat{\theta}^I$ , and  $\hat{\theta}^{SI}$  is a stratified importance sampling estimator, with estimates  $\hat{\theta}_j$  for  $\theta_j$  on the individual strata, each with  $m$  replicates. If  $\text{Var}(\hat{\theta}^I) = \sigma^2/M$  and  $\text{Var}(\hat{\theta}_j) = \sigma^2/m, j = 1, \dots, k$ , then*

$$\sigma^2 - k \sum_{j=1}^k \sigma_j^2 \geq 0$$

*with equality if and only if  $\theta_1 = \dots = \theta_k$ . Hence, stratification never increases the variance, and there exists a stratification that reduces the variance except when  $g(x)$  is constant.*