

Problem Sheet 4

1. Let x be a random variable following uniform distribution $U[0, 1]$. Let function $f(x)$ be

$$f(x) = \begin{cases} \frac{1}{2}x & \text{if } 0 \leq x \leq \frac{2}{3}, \\ 2x - 1 & \frac{2}{3} \leq x \leq 1. \end{cases}$$

Suppose that we want to compute $E[f(x)]$ using the stratified sampling methods with the two strata: $\Omega_1 = [0, 2/3]$ and $\Omega_2 = [2/3, 1]$. Also assume that the total sample size is $N = 100$. What is the optimal allocation of the samples. What is the estimator variance associated to the optimal allocation?

Solution:

We want to estimate $\mathbb{E}[f(X)]$ using the stratified sampling method with the domain divided into two strata:

$$\Omega_1 = \left[0, \frac{2}{3}\right], \quad \Omega_2 = \left[\frac{2}{3}, 1\right].$$

The probabilities of the strata are proportional to the lengths of the intervals:

$$P(\Omega_1) = \frac{2}{3}, \quad P(\Omega_2) = \frac{1}{3}.$$

The variance of $f(X)$ in each stratum is given by:

$$\text{Var}[f(X)|\Omega_i] = \mathbb{E}[f(X)^2|\Omega_i] - (\mathbb{E}[f(X)|\Omega_i])^2,$$

where $i \in \{1, 2\}$.

For Stratum $\Omega_1 = [0, \frac{2}{3}]$:

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

Expected value:

$$E[f(X) \mid \Omega_1] = \int_0^{2/3} \frac{1}{2}x \cdot \frac{1}{2/3} dx = \frac{1}{6}.$$

Second moment:

$$E[f(X)^2 \mid \Omega_1] = \int_0^{2/3} \left(\frac{1}{2}x\right)^2 \cdot \frac{1}{2/3} dx = \frac{1}{36}.$$

Variance:

$$\text{Var}[f(X) \mid \Omega_1] = \frac{1}{36} - \left(\frac{1}{6}\right)^2 = \frac{1}{108}.$$

For Stratum $\Omega_2 = [\frac{2}{3}, 1]$:

$$f(x) = 2x - 1.$$

Expected value:

$$E[f(X) \mid \Omega_2] = \int_{2/3}^1 (2x - 1) \cdot \frac{1}{1/3} dx = \frac{2}{3}.$$

Second moment:

$$E[f(X)^2 \mid \Omega_2] = \int_{2/3}^1 (2x - 1)^2 \cdot \frac{1}{1/3} dx = \frac{10}{9}.$$

Variance:

$$\text{Var}[f(X) \mid \Omega_2] = \frac{10}{9} - \left(\frac{2}{3}\right)^2 = \frac{2}{9}.$$

Optimal Allocation of Samples

Note that here one can use the Lagrangian method to derive the optimal allocation. It is also OK to use the formula directly. Here we use the formula directly.

The optimal allocation is proportional to:

$$N_i \propto P(\Omega_i) \sqrt{\text{Var}[f(X) \mid \Omega_i]}.$$

For Ω_1 :

$$N_1 \propto \frac{2}{3} \cdot \sqrt{\frac{1}{108}}.$$

For Ω_2 :

$$N_2 \propto \frac{1}{3} \cdot \sqrt{\frac{2}{9}}.$$

Normalize N_1 and N_2 so that $N_1 + N_2 = 100$:

$$N_1 = 67, \quad N_2 = 33.$$

Variance of the Stratified Estimator

The variance is given by:

$$\text{Var}[\hat{\mu}_{\text{strat}}] = \sum_{i=1}^2 \frac{P(\Omega_i)^2 \cdot \text{Var}[f(X) \mid \Omega_i]}{N_i}.$$

Substituting values:

$$\text{Var}[\hat{\mu}_{\text{strat}}] = \frac{\left(\frac{2}{3}\right)^2 \cdot \frac{1}{108}}{67} + \frac{\left(\frac{1}{3}\right)^2 \cdot \frac{2}{9}}{33}.$$

Simplify:

$$\text{Var}[\hat{\mu}_{\text{strat}}] \approx 0.00029.$$