△ Importance Sampling

Idea: importance function f $\Longleftrightarrow$ uniform weight

Goal: $\int_a^b g(x)\,dx$ $\qquad Y = \frac{g(x)}{f(x)} \qquad X \sim f$

① Var small $\Leftrightarrow$ $\frac{g}{f}$ close to a const.

② Easy to generate from f

eg. $\int_0^1 \frac{e^{-x}}{1+x^2}\,dx \qquad g(x) = \frac{e^{-x}}{1+x^2}$

Importance function:

$f_0(x) = 1, \quad 0 \le x \le 1 \qquad$ Unif(0, 1)

inefficiency: range of x is too large

$f_1(x) = e^{-x}, \quad 0 < x < \infty \qquad$ Exp(1)

$f_2(x) = \frac{\pi}{1+x^2}, \quad -\infty < x < \infty \qquad$ Cauchy

$f_3(x) = \frac{e^{-x}}{1-e^{-1}}, \quad 0 < x < 1 \qquad$ truncated Exp(1)

$f_4(x) = \frac{4}{\pi} \cdot \frac{1}{1+x^2}, \quad 0 < x < 1 \qquad$ truncated Cauchy

△ Variance reduction with Importance Sampling

$$\theta = \int g(x)\,dx = \int \frac{g(x)}{f(x)} f(x)\,dx$$

$$\hat{\theta} = \frac{1}{m} \sum_{i=1}^{m} \frac{g(x_i)}{f(x_i)} \qquad x_i \sim f$$

$$Var(\hat{\theta}) = \frac{1}{m} Var(Y), \quad Y = \frac{g(x)}{f(x)}$$

$$Var(Y) = Var\left(\frac{g(x)}{f(x)}\right) = E\left\{\frac{g(x)}{f(x)}\right\}^2 - \left[E\left\{\frac{g(x)}{f(x)}\right\}\right]^2$$

$$= \int \frac{g^2(x)}{f^2(x)} f(x)dx - \left[\int \frac{g(x)}{f(x)} f(x)dx\right]^2$$

$$= \int \frac{g^2(x)}{f(x)} dx - \underbrace{\left[\int g(x)dx\right]^2}_{\theta}$$

$$\min_f Var(Y) \iff \min_f \int \frac{g^2(x)}{f(x)} dx$$

$$(HW) \overset{\text{Cauchy-Schwarz}}{\iff} f^*(x) = \frac{|g(x)|}{\int |g(x)| dx} \Rightarrow \begin{cases} \text{① valid density} \\ \text{② shape of } |g(x)| \\ \text{③ useless in practice} \end{cases}$$

△ Stratified Sampling

Goal : $\theta = \int_a^b g(x)dx$

Idea : partition the integral range $[a,b]$ into k disjoint strata

Apply simple MC within each and then aggregate

Remark: ① $a = x_0 < x_1 < \cdots < x_k = b$

② Guaranteed by $\begin{cases} \text{linearity of integral} \\ \text{SLLN} \end{cases}$

Prop: Denote $\hat{\theta}^m$ as the stanard MC

Let $\hat{\theta}^S = \frac{1}{k}\sum_{j=1}^{k}\hat{\theta}_j$ be the stratified sampling

In each stratum, sample size is $m_k = \frac{m}{k}$

Denote (θ_j, σ_j^2) be the mean and variance of $g(u)$ restricted in the j-th stratum

Then $Var(\hat{\theta}^m) \geq Var(\hat{\theta}^S)$

Remark: ① $\hat{\theta}^M$: $u \sim unif(0,1)$ when $a=0, b=1$

$\hat{\theta}^S$: $u | I_j \sim Unif(I_j)$ with density $k \cdot 1_{\{x \in I_j\}}$

② $\theta_j = E\{g(u)|I_j\} = k\int_{I_j} g(u)du$

$\sigma_j^2 = Var\{g(u)|I_j\}$

pf: Independence of $\hat{\theta}_j$'s

$$Var(\hat{\theta}^S) = \frac{1}{k^2}\cdot\sum_{j=1}^{k}\frac{\sigma_j^2}{m_k} = \frac{1}{m\cdot k}\sum_{j=1}^{k}\sigma_j^2$$

r.v. J is a discrete uniform dist. in $\{1,2,\cdots,k\}$

$J=j \Leftrightarrow u = u^{(j)} \Leftrightarrow u \in I_j$

$P(J=j) = \frac{1}{k}$, $j=1,2,\cdots,k$

$$Var(\hat{\theta}^M) = \frac{1}{m}Var(g(u))$$

[Law of total variance (HW)

$$Var(Y) = E\{Var(Y|X)\} + Var\{E(Y|X)\}$$]

$$= \frac{1}{m}\left(Var[E\{g(u)|J\}] + E[Var\{g(u)|J\}]\right)$$

using remarks $= \frac{1}{m}\left\{Var(\theta_J) + E(\sigma_J^2)\right\}$

$\because \sigma_J^2 = \begin{cases} \sigma_1^2 & \frac{1}{k} \\ \sigma_2^2 & \frac{1}{k} \\ \vdots & \\ \sigma_k^2 & \frac{1}{k} \end{cases}$

$\therefore \sigma_J^2 = \sum \sigma_j^2 \mathbb{1}_{\{J=j\}}$

$$= \frac{1}{m} Var(\theta_J) + \frac{1}{m} \cdot \frac{1}{k} \sum_{j=1}^{k} \sigma_j^2$$

$$\geq Var(\hat{\theta}^s)$$

"$=$" holds $\iff Var(\theta_J) = 0$

$\iff \theta_1 = \theta_2 = \cdots = \theta_k$