Chapter 5 Monte Carlo Integration and Variance Reduction

△ MC Integration

eg. given function $g(x)$.

integral $\int_a^b g(x)\,dx$

$-\infty \le a < b \le +\infty$

this is our target

eg. consider a r.v. $X \sim f$

$E\{g(x)\} = \int_{-\infty}^{+\infty} g(x) f(x)\,dx$

this is our tool

Define an operator $E_n\{g(x)\} = \frac{1}{n}\sum_{i=1}^{n} g(x_i)$

as the empirical average, other notations like $\overline{g(x)}$, $\overline{g_n(x)}$

In theory, $E_n\{g(x)\}$ is a "good" estimator of $E\{g(x)\}$

"good":

① unbiased $E\left(E_n\{g(x)\}\right) = E(g(x))$

② consistent $E_n\{g(x)\} \xrightarrow{a.s.} E\{g(x)\}$ as $n \to +\infty$

(SLLN)

③ asymptotic normality (CLT)

$\sqrt{n}\left(E_n\{g(x)\} - E\{g(x)\}\right) \xrightarrow{d} N(0, \sigma^2)$

eg. Let $\theta = \int_0^1 g(x)\,dx = \int_{-\infty}^{\infty} g(x) f(x)\,dx$

where $f(x) = 1_{[0,1]}(x) = 1 \ (0 \le x \le 1)$

$\hat{\theta} = \frac{1}{m}\sum_{i=1}^{m} g(x_i)$ where $\{x_i\}$'s are generated from $\text{Unif}(0,1)$

eg. $\theta = \int_a^b g(x)\,dx$

① change variables $y = \frac{x-a}{b-a} \in (0,1)$

$\theta = \int_0^1 g(y(b-a)+a)(b-a)\,dy$

② Generate from Unif (a,b)

$$f(x) = 1_{[a,b]}(x) \cdot \frac{1}{b-a}$$

$$\theta = \int_{-\infty}^{+\infty} g(x) f(x)\, dx \cdot (b-a)$$

Steps of simple MC estimator

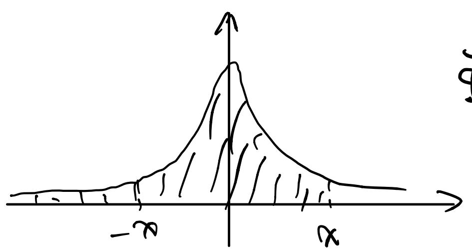
① Generate $X_1, \dots, X_m$ from Unif (a,b)

② Compute $\overline{g(x)}_m = \frac{1}{m} \sum_{i=1}^{m} g(X_i)$

③ Calculate $\hat{\theta} = (b-a)\, \overline{g_m(X)}$

eg. (MC integration over an unbounded interval)

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt \qquad (\text{c.d.f. of } N(0,1))$$



$$\Phi(+\infty) = 1 \quad \& \quad \Phi(x) + \Phi(-x) = 1$$

$$\Phi(x) = \begin{cases} \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}}\, dt & x \geq 0 \\ 1 - \Phi(-x) & x < 0 \end{cases}$$

Let $\theta = \int_0^x e^{-t^2/2}\, dt$, $x \geq 0$

Direct method : Unif $(0, x)$

when x changes, the generating procedures change

Indirect method : change of variables

$t = xy$, then $y = \frac{t}{x} \in (0,1)$

$\theta = \int_0^1 x \cdot \exp\{-\frac{(xy)^2}{2}\}\, dy$, where x fixed

$\theta = E_Y\left(x \cdot \exp\{-\frac{(xY)^2}{2}\}\right)$, where $Y \sim$ Unif $(0,1)$

Steps: ① Generating $u_1, \cdots, u_m$ from $Unif(0,1)$

② $\hat{\theta} = \overline{g_m(u)} = \frac{1}{m}\sum_{i=1}^{m} x \cdot \exp\{-\frac{(x u_i)^2}{2}\}$

Hit-or-miss: Assume we can generate from $N(0,1)$

$$\Phi(x) = P(Z \le x) \qquad Z \sim N(0,1)$$
$$= E\{1_{(Z \le x)}\}$$
$$\hat{\Phi}(x) = E_m\{1_{(Z \le x)}\} = \frac{1}{m}\sum_{i=1}^{m} 1(Z_i \le x)$$

Steps: ① Generating $Z_1, \cdots, Z_m \sim N(0,1)$

② Evaluating each indicator $1(Z_i \le x)$, $i=1,\cdots,m$

③ $\hat{\Phi}(x) = \frac{1}{m}\sum_{i=1}^{m} 1(Z_i \le x)$

△ Uncertainty quantification of $\hat{\theta}$

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

$E\hat{\theta} = \theta$

Assume unbiasedness first, then compare the variance

$$Var\,\hat{\theta} = \frac{1}{m^2}\sum_{i=1}^{m} Var(g(x_i)) = \frac{1}{m} Var(g(x))$$
$$= \frac{1}{m}\sigma^2, \text{ where } \sigma^2 = Var(g(x))$$

In practice, σ^2 is unknown and is needed to be estimated

"plug-in" estimator. $\sigma^2 = Var(g(x)) = E[g(x) - E\{g(x)\}]^2$

replace E by E_m: $\hat{\sigma}^2 = E_m[g(x) - E_m\{g(x)\}]^2$

$E_m\{g(x)\} = \frac{1}{m}\sum_{i=1}^{m} g(x_i) = \overline{g_m(x)}$

$$\hat{\sigma}^2 = \frac{1}{m}\sum_{i=1}^{m}\{g(x_i) - \overline{g_m(x)}\}^2$$

Then, $\widehat{Var}(\hat{\theta}) = \frac{1}{m}\hat{\sigma}^2 = \frac{1}{m^2}\sum_{i=1}^{m}\{g(x_i) - \overline{g_m(x)}\}^2$

standard error $\rightarrow$ $s.e.(\hat{\theta}) = \sqrt{\widehat{Var}(\hat{\theta})} = \frac{1}{m}\left[\sum_{i=1}^{m}\{g(x_i) - \overline{g_m(x)}\}^2\right]^{\frac{1}{2}}$

because, in statistics, m is always very large

so we don't care about the biasedness.

CLT: $\frac{\hat{\theta}-\theta}{s.e.(\hat{\theta})} \xrightarrow{d} N(0,1)$, $m \to +\infty$

△ Variance and Efficiency

Efficiency: two estimators $\hat{\theta}_1$ and $\hat{\theta}_2$ are unbiased,

if $Var(\hat{\theta}_1) < Var(\hat{\theta}_2)$, then we say $\hat{\theta}_1$ is

more efficient than $\hat{\theta}_2$.

Quantity $\frac{Var(\hat{\theta}_2) - Var(\hat{\theta}_1)}{Var(\hat{\theta}_2)} \times 100\%$

is the percentage of improvement of $\hat{\theta}_1$ over $\hat{\theta}_2$

① $\theta = \int_a^b g(x)\,dx = E_X\, g(x)$, $X \sim Unif(a,b)$

$\hat{\theta} = (b-a)\frac{1}{m}\sum_{i=1}^{m} g(X_i)$

$Var(\hat{\theta}) = \frac{(b-a)^2}{m} Var(g(x))$

② $F(x) = \int_{-\infty}^{x} f(t)\,dt = E(1\{X \le x\})$, $X \sim f$

$\hat{F}(x) = \frac{1}{m}\sum_{i=1}^{m} 1\{X_i \le x\}$

$Var(\hat{F}(x)) = \frac{1}{m}F(x)(1-F(x))$

△ Variance Reduction

- Antithetic variable
- Control Variate
- Importance Sampling
- Stratified Sampling

△ General Set-up

n-variate fun. $g(x) = g(x_1, \cdots, x_n)$

Given generated samples

$X^{(j)} = (x_1^{(j)}, \cdots, x_n^{(j)})$, $j = 1, \cdots, m$

Goal: $\theta = E\{g(x)\}$

Tool: $\hat{\theta} = E_m\{g(x)\} = \frac{1}{m}\sum_{j=1}^{m} g(x^{(j)})$

△ Antithetic variable

eg. Two r.v. u_1 and u_2

$$Var\left(\frac{u_1+u_2}{2}\right) = \frac{1}{4}\left(Var(u_1) + Var(u_2) + 2Cov(u_1, u_2)\right)$$

If u_1 and u_2 ind.

$$Var\left(\frac{u_1+u_2}{2}\right) = \frac{1}{4}\left(Var(u_1) + Var(u_2)\right)$$

To reduce the variance, $Cov(u_1, u_2)$ should be negative

Facts. ① u and $1-u$ are nega. corr.

② If $u \sim Unif(0,1)$ then $1-u \sim Unif(0,1)$

Define
$$Y_j = g(X^{(j)}) = g(x_1^{(j)}, \cdots, x_n^{(j)})$$
$$= g(F_X^{-1}(u_1^{(j)}), \cdots, F_X^{-1}(u_n^{(j)}))$$
$$Y_j' = g(F_X^{-1}(1-u_1^{(j)}), \cdots, F_X^{-1}(1-u_n^{(j)}))$$

Question: Y_j and Y_j' have the same distribution, when are Y_j and Y_j' negatively correlated?

Def: $(x_1, \cdots, x_n) \le (y_1, \cdots, y_n) \Leftrightarrow x_j \le y_j, \forall j$

$g(x)$ increasing $\Leftrightarrow g(x_1, \cdots, x_n) \le g(y_1, \cdots, y_n)$ for $x \le y$

Prop. 5.1. $X = (X_1, \cdots, X_n)$ are ind. f, g increasing func.

then $E\{f(X)g(X)\} \geq \{Ef(X)\} \cdot \{Eg(X)\}$

Pf. Induction

$n=1$ suppose X and Y are iid.

$\{f(X)-f(Y)\}\{g(X)-g(Y)\} \geq 0$

$\Rightarrow E[\{f(X)-f(Y)\}\{g(X)-g(Y)\}] \geq 0$

$\Rightarrow E\{f(X)g(X)\} + E\{f(Y)g(Y)\} \geq$

$E\{f(Y)g(X)\} + E\{f(X)g(Y)\}$

$\Rightarrow 2E\{f(X)g(X)\} \geq E\{f(Y)\}E\{g(X)\} + E\{f(X)\}E\{g(Y)\}$

$\Rightarrow 2E\{f(X)g(X)\} \geq E\{f(X)\}E\{g(X)\} + E\{f(X)\}E\{g(X)\}$

$= 2E\{f(X)\}E\{g(X)\}$

If $n-1$ case is true, we need to show n case is true

Conditioning on $X_n = x_n$

$E\{f(X)g(X) \mid X_n = x_n\} = E\{f(X_1, \cdots, X_{n-1}, x_n)g(X_1, \cdots, X_{n-1}, x_n)\}$

$\geq E\{f(X_1, \cdots, x_n)\} E\{g(X_1, \cdots, x_n)\}$

$= E\{f(X) \mid X_n = x_n\} \cdot E\{g(X) \mid X_n = x_n\}$

Taking expectation w.r.t. X_n

we have $E\{f(X)g(X)\} \geq \{Ef(X)\}\{Eg(X)\}$

Corollary 5.1. If $g(x)$ is a monotone fun.

then $Y = g(F_X^{-1}(u_1) \cdots, F_X^{-1}(u_n))$

$$Y' = g\left(F_X^{-1}(1-u_1), \dots, F_X^{-1}(1-u_n)\right)$$

are negn. corr.

<u>Pf</u>: Assume g increasing

then Y and $-Y'$ are both increasing in U

$$-E\{Y \cdot Y'\} \geq -(EY)\cdot(EY')$$

$$\mathrm{Cov}(Y, Y') = E(Y \cdot Y') - (EY)\cdot(EY') \leq 0$$