Monte Carlo Integration

Chapter 4

STATS 102C: Introduction to Monte Carlo Methods

$X \sim N(\mu, \sigma^2)$

$X \sim T(\theta)$

$\hat{\theta}$

$Var(X) = E(X^2) - E(X)^2$

$E(X) = \mu = \int x \cdot f(x)\,dx$

$F_X(x) = \int_{-\infty}^{x} f(t)\,dt = \theta$

$E(X^2) = \int x^2 f(x)\,dx$

$E(X^3)$

$E(e^x) = \int e^x f(x)\,dx$

$E[g(X)] = \int_{-\infty}^{\infty} g(x) f(x)\,dx = \theta$

$\boxed{x_1, \ldots, x_n \sim f(x)}$

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

$x \sim f_x(x)$

$Y = h(x)$

$x = h^{-1}(y) = g(y)$

$\lim_{n \to \infty} P(|\hat{\theta}_n - \theta| < \varepsilon) = 1 \quad \forall \varepsilon$

$f_Y(y) = |g'(y)| \cdot f_x(g(y))$

$g(x) = x$

$f_{x,y}(r, \theta) = |J| \cdot f_{xy}(r, \theta)$

$E(x)$

Introduction

LLN: $\lim_{n\to\infty} P(|\bar{X} - \mu| < \varepsilon) = 1 \quad \forall \varepsilon$

$\bar{X} = \hat{\theta}$, $\mu = \theta$

- Monte Carlo integration is a statistical method based on random sampling.
- Let $g(x)$ be a function and suppose that we want to compute $\int_a^b g(x)dx$ (assuming that this integral exists).
- If X is a random variable with density $f(x)$, then the mathematical expectation of the random variable $g(X)$ is

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx$$

$$\theta = \int_0^1 g(x)\,dx = \int_0^1 g(x) \cdot 1\,dx$$

$X_1, \dots, X_m \sim \text{Unif}(0,1)$

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

$g(x) = e^{-x^2}$

$$\theta = \int_0^1 e^{-x^2} dx$$

$$\sum_{i=1}^{m} \frac{e^{-x_i^2}}{m} = \hat{\theta}$$

Simple Monte Carlo Estimator

- Consider the problem of estimating $\theta = \int_0^1 g(x)dx$. If $X_1, \ldots, X_m$ is a random Uniform (0,1) sample then

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

converges to $E(g(X)) = \theta$ with probability 1, by the Law of Large Numbers.

- Algorithm
  (1) Generate $X_1, \ldots, X_m$, iid from Uniform(a, b).
  (2) Compute $g(\bar{X}) = \frac{1}{m}\sum_{i=1}^{m} g(X_i)$.
  (3) $\hat{\theta} = (b-a)g(\bar{X})$

$\hat{\theta}_{MC}$

- Is $\hat{\theta}$ a consistent estimator?

$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\, dt = \theta$

$$\theta = \int_a^b g(x)\,dx = \int_a^b g(x)\frac{b-a}{b-a}\,dx$$

$$= (b-a)\int_a^b g(x)\underbrace{\frac{1}{b-a}}_{f(x)}\,dx$$

$$x_1, \dots, x_m \sim \mathrm{Unif}(a, b)$$

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

$$MSE(\hat{\theta}) = bias^2 + Var(\hat{\theta})$$

$$E(\hat{\theta}) = E\left(\frac{\sum_1^m g(x_i)}{m}\right)$$

$$= E[g(X)]$$

$$= \theta$$

$$\frac{\sum_1^m E(g(x_i))}{m}$$

$$\frac{\sum_1^m E[g(X)]}{m}$$

$$Var(\hat{\theta}) = Var\left(\frac{\sum_1^m g(x_i)}{m}\right)$$

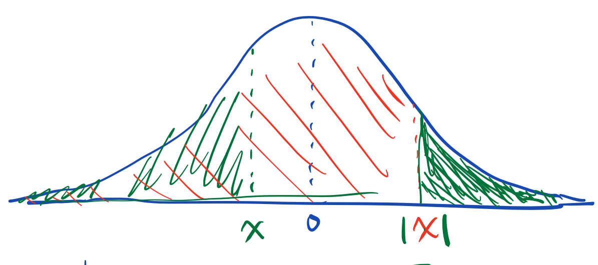
$$= \frac{1}{m} Var[g(X)]$$

$$= \frac{1}{m}\left[E(g^2(X)) - E(g(X))^2\right]$$

$$= \frac{1}{m}\left[\int g^2(x) f(x)\,dx - \theta^2\right]$$

① $m \to \infty$ $\quad var(\hat{\theta}) \approx 0$

② $\int g^2(x) f(x)\,dx \approx \theta^2$



$$F_X(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}\,dt = \theta$$

$$X \sim N(0, 1)$$

Apprach 1

$$x > 0 \qquad\qquad x < 0$$

$$F(0) + \underbrace{\int_0^x \boxed{\frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}} dt}_{\theta} \quad g(x)$$

$$F_X(x) = 1 - F_X(|x|)$$

$$t_1, \ldots, t_m \sim \text{Unif}(0, x)$$

$$= 1 - \int_{-\infty}^{|x|} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

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

$$= 1 - \hat{F}_{MC}(|x|)$$

$$\hat{F}_{MC}(x) = \frac{1}{2} + \hat{\theta}$$

Approach 2

Let x/x and set $y = \frac{t}{x}$ $\quad dt = x\, dy$

$x/x \quad y = \frac{t}{x}$

$dt = x\, dy$

$$\theta = \int_0^1 \underbrace{\boxed{\frac{1}{\sqrt{2\pi}} e^{-\frac{(xy)^2}{2}} x}}_{g(y)} dy$$

$$\int_0^x \lambda e^{-\lambda t} dt$$

$$Y_1, \ldots, Y_m \sim \text{Unif}(0,1)$$

$$\int_0^1$$

$$\hat{\theta} = \frac{\sum g(y_i)}{m}$$

Approach 3 The hit-or-miss approach

Compute $F_X(0.01)$ $\quad F_X(1)$ $\quad F_X(1.1)$

$P_X(X \le 0.01)$ $\quad P(X \le 1)$ $\quad P(X \le 1.1)$

$$x_1, \ldots, x_m \sim N(0, 1)$$

$$\frac{\sum_{i=1}^m I(x_i \le 0.01)}{m} = \hat{F}(0.01) \qquad \frac{\sum_{i=1}^m I(x_i \le 1)}{m} = \hat{F}(1)$$

$$X = \mu + A^T Z$$

Example: Standard Normal CDF

$\Sigma = A^T A$

- Use the Monte Carlo approach to estimate the standard normal cdf
$$F(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$

- R code to generate to estimate standard normal cdf using the simple Monte Carlo method.

```
x <- seq(.1, 2.5, length = 10)
m <- 10000
u <- runif(m)
mc <- numeric(length(x))
for (i in 1:length(x)){
    g <- x[i] * exp(-((u * x[i]) ^ 2) / 2)
    mc[i] <- mean(g) / sqrt(2 * pi) + 0.5
}
```

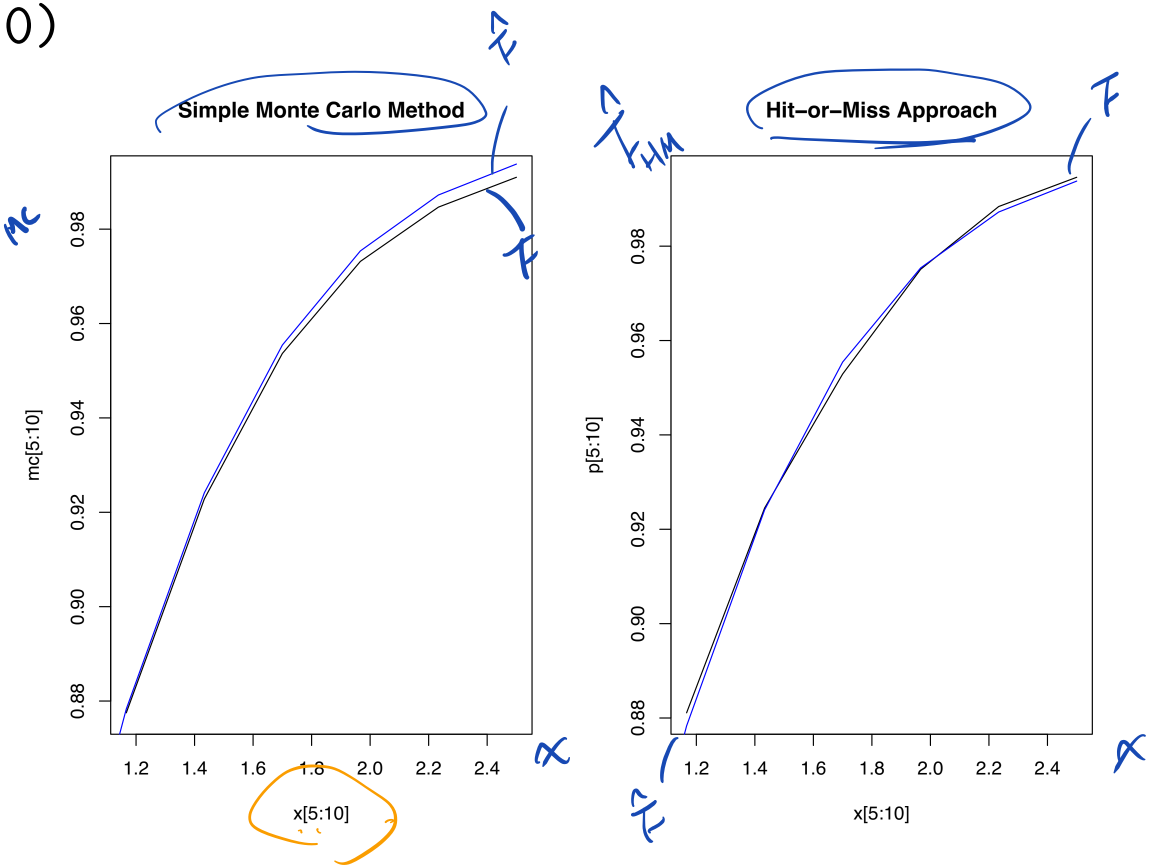
$\mathrm{Var}(\hat{\theta}_{MC})$ $\mathrm{Var}(\hat{\theta}_{HM})$

- R code to generate to estimate standard normal cdf using the hit-or-miss approach.

```r
z <- rnorm(m)
dim(x) <- length(x)
p <- apply(x, MARGIN = 1,
           FUN = function(x, z){mean(z < x)}, z = z)
```

```
## integer(0)
```

$$\frac{\sqrt{Var(\hat{F}_{MC})}}{\sqrt{Var(\hat{F}_{HM})}}$$




```
## integer(0)
```

$x_1, \ldots, x_m \sim f(x)$

Standard Error of $\hat{\theta}$

$\int_A f(x)\,dx = 1$ $\quad$ $\hat{\theta} = \frac{\sum^m g(x_i)}{m} = \overline{g(x)}$

- If $f(x)$ is a pdf supported on a set A, $E[g(X)] = \int_A g(x) f(x) dx = \theta$

$Var(\hat{\theta}) = Var\left(\frac{\sum g(x_i)}{m}\right)$

$= \frac{1}{m} Var(g(x))$

- To estimate θ, generate $x_1, \ldots, x_m \sim f(x)$, and compute $\hat{\theta} = \frac{\sum_{i=1}^m g(x_i)}{m}$

$= \frac{1}{m}\sigma^2$

- When the distribution of X is unknown, we substitute for F_X the empirical distribution $\hat{F}_m$ of the sample $x_1, \ldots, x_m$. The variance of $\hat{\theta}$ can be estimated by

$$SE(\hat{\theta}) = \sqrt{Var(\hat{\theta}_{MC})} = \sqrt{\frac{1}{m^2}\sum_{i=1}^m [g(x_i) - \hat{\theta}]^2}$$

$\hat{\sigma}^2_{MLE} = \frac{\sum (g(x) - \overline{g(x)})^2}{m}$

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

- The Central Limit Theorem implies that

$$\frac{\hat{\theta} - E(\hat{\theta})}{\sqrt{Var(\hat{\theta})}} = \frac{\hat{\theta} - \theta}{SE(\hat{\theta})}$$

coverges in distribution to $N(0, 1)$ as $m \to \infty$.

To estimate $\int_a^b g(x)\,dx = \theta$

$$\theta = (b-a)\int_a^b g(x)\frac{1}{b-a}\,dx = (b-a)\,\underline{E_X(g(x))},\quad X \sim Unif(a,b)$$

$$E_X(g(x)) = \frac{\theta}{b-a}$$

$$\hat{\theta} = \underline{(b-a)\cdot\overline{g(x)}},\qquad \overline{g(x)} = \frac{\sum_{i=1}^{m} g(x_i)}{m}$$

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

$$Var\left(\frac{\sum_1^m g(X_i)}{m}\right)$$

$$\underline{Var(\hat{\theta}_{MC})} = Var\left((b-a)\,\overline{g(x)}\right)$$
$$= (b-a)^2\frac{\sigma^2}{m}$$
$$= (b-a)^2\frac{\hat{\sigma}^2}{m}$$

$$\frac{1}{m^2}\sum_1^m \underbrace{\boxed{Var(g(x_i))}}_{\sigma^2}$$

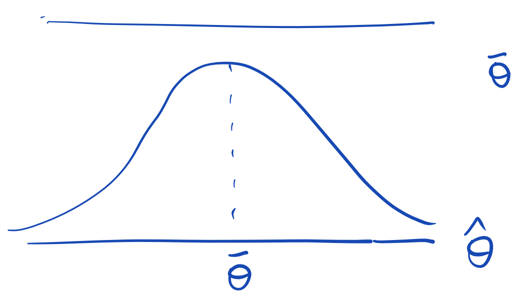
$$\hat{\sigma}^2 = \frac{\sum\left(g(x_i) - \overline{g(x)}\right)^2}{m}$$

$$Var(\hat{\theta}_{HM}) = \frac{\hat{F}(x)\,(1-\hat{F}(x))}{m}$$

Ch 7

Bootstrap Estimation of $SE(\hat{\theta})$

$\hat{\theta}^{(1)}\ \hat{\theta}^{(2)}\ \hat{\theta}^{(3)} \cdots\ \hat{\theta}^{(B)} \rightarrow$ estimate $SE(\hat{\theta})$



$$\bar{\theta} = \frac{\sum\hat{\theta}}{B}$$

$$SE(\hat{\theta}) = \sqrt{\frac{\sum_{j=1}^{B}\left(\hat{\theta}^{(j)} - \bar{\theta}\right)^2}{B-1}}$$

$$MSE(\hat{\theta}) = Var(\hat{\theta}) + bias^2$$

bias $\hat{\theta} - \bar{\theta}$ $\quad\because$ By LLN $E(\bar{\theta}) \simeq \theta$

$$\underline{MSE(\hat{\theta})} = E\left((\hat{\theta}-\theta)^2\right)$$

$$\widehat{MSE}(\hat{\theta}) = \frac{\sum\left(\hat{\theta}^{(j)} - \bar{\theta}\right)^2}{B}$$

Estimation of $\underline{MSE}(\hat{\theta})$ MC method

$$MSE(\hat\theta) = E[\underline{(\hat\theta - \theta)^2}] = \int (\hat\theta - \theta)^2 \cdot f(x)\, dx$$

$x^{(1)}, x^{(2)}, \ldots, x^{(B)} \sim f(x)$

$\downarrow \quad \downarrow \quad\quad \downarrow$

$\hat\theta^{(1)}_{MC} \quad \hat\theta^{(2)}_{MC} \quad \ldots \quad \hat\theta^{(B)}_{MC}$

$$\widehat{MSE}(\hat\theta) = \underline{\frac{1}{B}\sum_{j=1}^{B} (\hat\theta^{(j)}_{MC} - \bar\theta)^2}$$

$X^{(j)}_{1:m}$

E.g. Estimate the MSE of a trimmed mean for $N(0,1)$

$x^1, \ldots, x^B \sim f(x)$ and $\underline{x^{(1)}}\ x^{(2)}, \ldots\ x^{(B)}$ ordered saple

$x^1_{1:m}$

$$\theta = \mu \leftarrow \underline{\bar{X}_{[-k]}} = \frac{1}{m-2k}\sum_{i=k+1}^{m-k} X_{(i)}$$

$k=1$

for ($j = 1$ to B) {

MC
- generate $x_1^{(j)}, \ldots, x_m^{(j)} \sim N(0,1)$
- obtain $x^{(j)}_{(1)}, \ldots, x^{(j)}_{(m)}$
- compute $\hat\theta^{(j)} = \frac{1}{m-2} \cdot \sum_{i=2}^{m-1} x^{(j)}_{(i)}$

}

$B = 100$

$\underline{50}$

$$\widehat{MSE}(\hat\theta) = \frac{1}{B}\sum_{j=1}^{B} (\hat\theta^{(j)} - \bar\theta)^2$$

E.g. To estimate $\underline{E|X_1 - X_2|} = \theta$ $\quad X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right)$

① $\begin{pmatrix} x_1^{(j)} \\ x_2^{(j)} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right)$

$$\sum_{i=1}^{m} \frac{|x_1^{(j)} - x_2^{(j)}|}{m} = \hat\theta^{(j)}$$

$\begin{pmatrix} x_1^{(1)} \\ x_2^{(1)} \end{pmatrix} \ldots\ldots \begin{pmatrix} x_1^{(B)} \\ x_2^{(B)} \end{pmatrix}$

$SE(\hat\theta)$

Variance and Efficiency

- Supposed $\hat{\theta}_1$ and $\hat{\theta}_2$ are two estimators for θ, then $\hat{\theta}_1$ is more efficient than $\hat{\theta}_2$ if
$$\boxed{\frac{Var(\hat{\theta}_1)}{Var(\hat{\theta}_2)}} < 1.$$

- Supposed the variances of estimators $\hat{\theta}_i$ are unknown, we can estimate efficiency by substituting a sample estimate of the variance for each estimator.

- Variance Reduction: There are several approaches to reducing the variance in the sample mean estimator of $E[g(X)]$.

Antithetic Variables

to reduce $var(\hat{\theta}_{MC})$

Y,

Y

Underlying principle: The antithetic variates technique consists, for every sample path obtained, in taking its antithetic path - that is given a path $\{\varepsilon_1, \ldots, \varepsilon_M\}$ to also take $\{-\varepsilon_1, \ldots, -\varepsilon_M\}$. The advantage of this technique is twofold: it reduces the number of normal samples to be taken to generate N paths, and it reduces the variance of the sample paths, improving the precision. Wikiepedia

Proposition 6.1[1]**:**
If $Y_1, Y_2, \ldots, Y_d$ are *independent*, and f and h are *increasing functions*, then

$$E[f(Y)h(Y)] \geq E[f(Y)]E[h(Y)]$$

[1] Statistical Computing with R, 2nd edition

Two identically r.v. Y and Y'

$\frac{Y+Y'}{2}$ $\quad E[\frac{Y+Y'}{2}] = E[Y] = \mu$

$Var(\frac{Y+Y'}{2}) = \frac{1}{4}[Var(Y) + Var(Y') + 2\underline{Cov(Y, Y')}]$

to make $cov(Y, Y') < 0$

suppose $X_j = \underline{F_X^{-1}(u)}$, $j=1, \dots, m$ $\quad u \sim Unif(0,1)$

$Y_j = g[\underline{F_X^{-1}(u)}]$, $j=1, \dots, m$

u and $1-u \sim Unif(0,1)$

$Corr(\underline{u, 1-u}) = -1$

$Y_j' = g[\underline{F_X^{-1}(1-u)}]$, $j=1, \dots, m$

Q: under what conditions are Y_j and Y_j' negatively correlated?

Ans: If $g()$ is monotone, $\underline{Y_j \text{ and } Y_j' \text{ are negatively correlated.}}$

$Y_j = \underline{g[F_X^{-1}(u)]} = f(u)$ $\quad$ assume $g()$ is increasing func.

$-Y_j' = -g[F_X^{-1}(1-u)] = h(u)$ $\quad \underline{cov(Y_j, Y_j')} < 0$

$E[f(u)h(u)] \geq E(f(u))\,E(h(u))$ $= E(Y_j Y_j') - E(Y_j)E(Y_j')$

$E[-Y_j Y_j'] \geq E(Y_j)\,E(-Y_j')$

$E(Y_j Y_j') \leq E(Y_j)\,E(Y_j')$

Algorithm

$u_1, u_2, \cdots u_{m/2} \sim Unif(0,1)$

$Y_j = f(u_j)$

$Y_j' = h(u_j) = f(1-u_j)$

$\hat{\theta} = \sum_{i=1}^{m/2} \frac{Y_j + Y_j'}{m}$

E.g. $x \geq 0$, $F_X(x) + 0.5$ $g(u)$

$$\theta = \int_0^1 \frac{x}{\sqrt{2\pi}} e^{-\frac{1}{2}(xu)^2} du \qquad u \sim Unif(0,1)$$

$Y_j = g(u_j) \Rightarrow f(u)$

$Y_j' = g(1-u_j) \Rightarrow h(u)$

$\hat{\theta} = \sum_{i=1}^{m/2} \frac{Y_j + Y_j'}{m}$

Example: Standard Normal CDF

Repeat the estimation of the standard normal cdf using antithetic variables, and find the approximate reduction in variance.

R function to generate to estimate standard normal cdf with or without antithetic sampling.

```r
MC_phi <- function(x, m = 10000, anti = TRUE){
     u <- runif(m / 2)
     if(!anti)
      v <- runif(m / 2)
     else
      v <- 1 - u

     u <- c(u, v)
     mc <- numeric(length(x))
     for(i in 1:length(x)){
        g <- x[i] * exp(-(u * x[i]) ^ 2 / 2)
        mc[i] <- mean(g) / sqrt(2 * pi) + 0.5
     }
     mc
}
```

$F(0)$

The R code to compute the approximate reduction in variance at a $x = 1.3$.

(B)
```
n <- 50
MC1 <- MC2 <- numeric(n)
x0 <- 1.3
for(i in 1:n){
  MC1[i] <- MC_phi(x0, m = 1000, anti = FALSE)
  MC2[i] <- MC_phi(x0, m = 1000)
}
(var(MC1) - var(MC2)) / var(MC1)
```

(B)

```
## [1] 0.9569375
```

~96% reduction in variance

Control Variates

- To reduce the variance in a Monte Carlo estimator of $\theta = E[g(X)]$ is the use of control variates. Suppose that there is a function f, such that $u = E[f(X)]$ is known, and $f(x)$ is correlated with $g(x)$. Then for any constant c, we can prove that $\hat{\theta}_c = g(x) + c[f(x) - \mu]$ is an unbiased estimator of θ.
- $Var(\hat{\theta}_c) = Var[g(X)] + 2cCov[g(X), f(X)] + c^2 Var[f(X)]$
- $Var(\hat{\theta}_c)$ is minimized at $c = c^*$, where

$$c^* = -\frac{Cov[g(X), f(X)]}{Var[f(X)]}$$

- and minimum variance is

$$Var(\hat{\theta}_{c^*}) = Var[g(X)] - \frac{Cov[g(X), f(X)]^2}{Var[f(X)]}$$

$$\hat{\theta}_c = g(x) + c[f(x) - \mu]$$

$E(f(x)) = \mu$ is known

$E[g(x)] = \theta$

$$E[\theta_c] = \theta$$

$$Var(\hat{\theta}_c) \le Var(\hat{\theta})$$

$$Var(\hat{\theta}_c) = Var(g(x) + c[f(x) - \mu])$$

$$= Var(g(x)) + 2c\, Cov[g(x), f(x)] + c^2 Var[f(x)]$$

$$\min_c Var(\hat{\theta}_c)$$

$\frac{\partial}{\partial c} Var(\hat{\theta}_c) = 0$ and solve it for c^*

$$c^* = -\frac{Cov[g(x), f(x)]}{Var(f(x))}$$

$$Var(\hat{\theta}_c) = Var[g(x)] - 2\frac{Cov[g(x), f(x)]^2}{Var(f(x))} + \frac{Cov[g(x), f(x)]^2}{Var[f(x)]}$$

Control variate

$$= Var[g(x)] - \frac{Cov[g(x), f(x)]^2}{Var(f(x))} \le Var[g(x)]$$

The percent reduction $= 100 \cdot \dfrac{\frac{Cov[g(x), f(x)]^2}{Var(f(x))}}{Var(g(x))}$

$$= 100 \frac{Cov[g(x), f(x)]^2}{Var(g(x)) \cdot Var(f(x))}$$

$$= 100 \cdot Corr(g(x), f(x))^2$$

$$g(u) = \frac{e^{-u}}{1+u^2} \qquad u \sim \text{Unif}(0,1)$$

$$\text{corr}(g(u), f(u)) \uparrow$$

$$f(u) = e^{-u},$$

$$E(f(u)) = \int_0^1 e^{-u} \, du = 1 - e^{-1}$$

Algorithm

① find $f(x)$ and $E[f(x)] = \mu$

② $u \sim \text{Unif}(0,1)$ $\quad u_1, \dots, u_m$

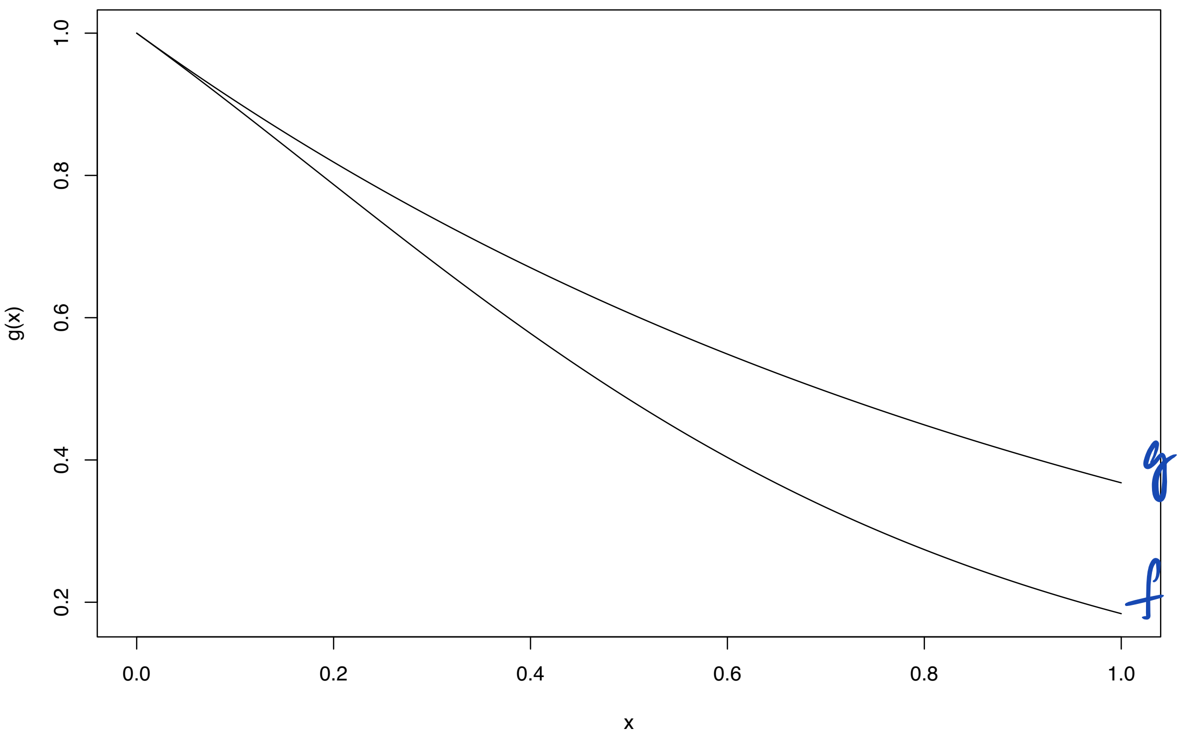
③ Compute $c^* = - \dfrac{\text{Cov}[g(u), f(u)]}{\text{Var}[f(u)]}$

④ $$\hat{\theta}_c = \sum_{i=1}^{m} \frac{g(u_i) + c^* \cdot [f(u_i) - \mu]}{m}$$

Example: $\theta = \int_0^1 \frac{e^{-x}}{1+x^2} dx$

```
set.seed(99999)
f <- function(u){
      exp(-u)
}
g <- function(u){
      exp(-u)/(1 + u^2)
}
curve(g)
curve(f,add = T)
```

这是随便选了个跟g像的



```r
u <- runif(10000)
B <- f(u)
A <- g(u)
cor(A, B)
```

```
## [1] 0.9995069
```

```r
a <- -cov(A, B) / var(B) #est of c*
m <- 100000
u <- runif(m)
T1 <- g(u) #w/o control variates
T2 <- T1 + a * (f(u) -  (1 - exp(-1)) )#w/ control variates

cbind(mean(T1), mean(T2))
```

$I(\hat{\theta}_f)$ $I(\hat{\theta}_c)$

```
##             [,1]      [,2]
## [1,] 0.525736 0.5247585
```

```r
(var(T1) - var(T2))/var(T1)
```

```
## [1] 0.9989983
```

$$\hat{\Theta}_c = \underline{c\,\hat{\Theta}_1 + (1-c)\,\hat{\Theta}_2}$$

$\hat{\theta}_1$ and $\hat{\theta}_2$ are unbiased estimators of Θ

$$E(\hat{\theta}_c) = \Theta$$

$$\begin{aligned}
Var(\hat{\theta}_c) &= Cov(c\hat{\theta}_1 + (1-c)\hat{\theta}_2,\ c\hat{\theta}_1 + (1-c)\hat{\theta}_2) \\
&= Cov(c(\hat{\theta}_1 - \hat{\theta}_2) + \hat{\theta}_2,\ c(\hat{\theta}_1 - \hat{\theta}_2) + \hat{\theta}_2) \\
&= Var(\hat{\theta}_2) + 2c\,Cov(\hat{\theta}_1 - \hat{\theta}_2,\ \hat{\theta}_2) + c^2\,Var(\hat{\theta}_1 - \hat{\theta}_2)
\end{aligned}$$

$\underline{Corr(\hat{\theta}_1, \hat{\theta}_2) = -1}$ $\Rightarrow$ $Cov(\hat{\theta}_1, \hat{\theta}_2) = -Var(\hat{\theta}_1)$

$\hat{\theta}_1$ and $\hat{\theta}_2$ are identically distributed $Var(\hat{\theta}_1) = Var(\hat{\theta}_2)$

$$\min_c Var(\hat{\theta}_c) \Rightarrow c^* = \frac{1}{2}$$

$$\hat{\theta}_c = \frac{1}{2}\hat{\theta}_1 + \frac{1}{2}\hat{\theta}_2$$

Several control variates

$$\hat{\theta}_c = g(X) + \sum_{i=1}^{K} c_i^* (f_i(X) - \mu_i)$$

$$E(\hat{\theta}_c) = \Theta$$

Simple linear regression

$$Var(\hat{\theta}_c) \le Var(\hat{\theta})$$

Stratified Sampling

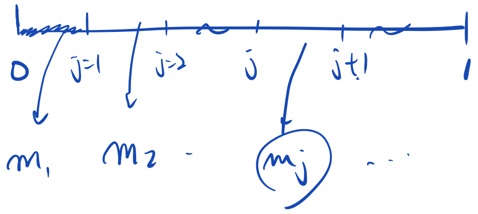
The stratified sampling method reduces the variance of the estimator by dividing the interval into strata and estimating the integral on each of the strata with smaller variance.

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 $m = m_1 + \cdots + m_k$, with the goal that

$$Var(\hat{\theta}_k(m_1, \ldots, m_k)) < Var(\hat{\theta}),$$

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

stratified sampling $\quad \underline{\theta = E[g(x)]} = \underline{\int_0^1 g(x)\,dx}$



$M = m_1 + m_2 + \cdots + m_k$

① $\underline{\hat{\theta}_j} = \dfrac{\sum_{i=1}^{m_j} g(x_i)}{m_j}$

$\underline{E[g(x|J=j)]} = \theta_j$

$\underline{\hat{\theta}_s} = \dfrac{\sum_{j=1}^{k} \hat{\theta}_j}{k}$

$E(\hat{\theta}_s) = \theta$

$\underline{E[\underline{E[g(x|J)]}]} \overset{?}{=} E(g(x))$

$E[E[g(x|J)]] = E[g(x|J=1)] \cdot P(J=1) +$

$E[g(x|J=2)] \cdot P(J=2) +$

$\vdots$

$\underline{E[g(x|J=k)]} \cdot P(J=k)$

$P(A,B) = P(A|B) \cdot P(B)$

$= \sum_{j=1}^{k} \sum_{i=1}^{m_j} g(x_i)\, P(X|J=j)\, P(J=j)$

$= \sum_{j=1}^{k} \underline{\sum_x g(x)} \cdot P(\underline{x, J=j})$

$= \sum_x g(x) \underline{\sum_{j=1}^{k} P(x, J=j)}$

$= \sum g(x) \cdot P(x) = E(g(x)) = \theta$

$Var(\hat{\theta}_s) = Var\left(\dfrac{\sum_1^k \hat{\theta}_j}{(k)}\right)$

$= \dfrac{1}{k^2} \sum_{j=1}^{k} \underline{Var(\hat{\theta}_j)}$

$= \dfrac{1}{k^2} \sum_{j=1}^{k} \dfrac{\sigma_j^2}{m_j}$

$= \dfrac{1}{k^2} \sum_{j=1}^{k} \dfrac{\sigma_j^2}{m}$

$\underline{m_1 = m_2 = \cdots m_j = \cdots m_k = m}$

$= \dfrac{1}{k^2 m} \sum_{j=1}^{k} \sigma_j^2$

$= \dfrac{1}{kM} \sum_{j=1}^{k} \sigma_j^2$

$\underline{m = M/k}$

$Var(\hat{\theta}_{MC}) = \dfrac{\sigma^2}{M}$

$\overline{\left| E(g(x))^2 \right|}$

$$\mathrm{Var}[E(g(x)|J)] = \underbrace{E[E(g(x)|J)^2]}_{①} - E[E(g(x|J))]^2$$

$$\mathrm{Var}[g(x|J)] = E[g(x|J)^2] - E[g(x)|J]^2$$

$$E[\mathrm{var}[g(x)|J]] = E[g(x)^2] - \boxed{E[E[g(x)|J]^2]}$$

$$\mathrm{Var}[E[g(x)|J]] = \mathrm{var}(g(x)) - E[\mathrm{var}(g(x)|J)]$$

$$\mathrm{var}(g(x)) = \underbrace{\mathrm{Var}[E[g(x)|J]]}_{\text{between-group variance}} + \underbrace{E[\mathrm{var}(g(x)|J)]}_{\text{within-group variance}}$$

$$\mathrm{var}(\hat{\theta}_{MC}) = \mathrm{var}(\overline{g(x)}) = \frac{1}{M}\mathrm{var}(g(x))$$

$$\mathrm{var}(\hat{\theta}_{MC}) = \frac{1}{M}\left[\mathrm{var}[\underbrace{E[g(x)|J]}_{\theta_J}] + E[\mathrm{var}(g(x)|J)]\right]$$

$$= \frac{1}{M}\left[\mathrm{var}(\theta_J) + E(\sigma_J^2)\right] \qquad p(J=j) = \frac{1}{K}$$

$$= \frac{1}{M}\left[\mathrm{var}(\theta_J) + \sum_{j=1}^{K}\frac{\sigma_j^2}{K}\right]$$

$$= \frac{1}{M}\mathrm{Var}(\theta_J) + \frac{1}{MK}\sum_{j=1}^{K}\sigma_j^2$$

$$\mathrm{var}(\hat{\theta}_{MC}) \geq \mathrm{var}(\hat{\theta}_S)$$