Probability, Computation and Simulation Homework 3

Brian Cervantes Alvarez

October 22, 2024

Problem 1

Suppose we wanted to estimate θ, where

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

Show that generating a random number U and then using the estimator

$$Y = e^{U^2}\left(\frac{1+e^{1-2U}}{2}\right)$$

is better than generating two random numbers U_1 and U_2 and using

$$Z = \frac{e^{U_1^2}+e^{U_2^2}}{2}.$$

We aim to compare the variances of the two estimators Y and Z for estimating

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

Since U_1, U_2 are independent and uniformly distributed on ([0,1]), the variance of Z is

$$\mathrm{Var}(Z) = \mathrm{Var}\left(\frac{e^{U_1^2}+e^{U_2^2}}{2}\right) = \frac{1}{4}\left(\mathrm{Var}(e^{U_1^2})+\mathrm{Var}(e^{U_2^2})\right) = \frac{1}{2}\,\mathrm{Var}(e^{U^2}),$$

where $U \sim \mathrm{Uniform}(0,1)$.

Let $U \sim \mathrm{Uniform}(0,1)$ and set $V = 1-U$. Then U and V are dependent but satisfy $U+V=1$. The estimator Y becomes

$$Y = \frac{e^{U^2}+e^{V^2}}{2}.$$

The variance of Y is

$$\mathrm{Var}(Y) = \mathrm{Var}\left(\frac{e^{U^2}+e^{V^2}}{2}\right) = \frac{1}{4}\left(\mathrm{Var}(e^{U^2})+\mathrm{Var}(e^{V^2})+2\,\mathrm{Cov}(e^{U^2},e^{V^2})\right).$$

Since U and V are symmetrically distributed over ([0,1]), e^{U^2} and e^{V^2} have the same variance:

$$\mathrm{Var}(e^{U^2}) = \mathrm{Var}(e^{V^2}).$$

Thus,

$$\mathrm{Var}(Y) = \frac{1}{2}\,\mathrm{Var}(e^{U^2}) + \frac{1}{2}\,\mathrm{Cov}(e^{U^2}, e^{V^2}).$$

Because U and V are negatively correlated ($V = 1-U$), and e^{x^2} is an increasing function on ([0,1]), e^{U^2} and e^{V^2} are negatively correlated. Therefore,

$$\mathrm{Cov}(e^{U^2}, e^{V^2}) < 0.$$

This implies

$$\mathrm{Var}(Y) < \frac{1}{2}\,\mathrm{Var}(e^{U^2}).$$

From earlier,

$$\mathrm{Var}(Z) = \frac{1}{2}\,\mathrm{Var}(e^{U^2}).$$

Thus,

$$\mathrm{Var}(Y) < \mathrm{Var}(Z).$$

Therefore, the variance of Y is less than that of Z, the estimator Y is better for estimating θ due to its lower variance.

Problem 3

Let X_i, $i = 1, \ldots, 5$, be independent exponential random variables each with mean 1, and consider the quantity

$$\theta = P\left(\sum_{i=1}^{5} iX_i \geq 21.6\right).$$

Part A

Now, we can estimate θ using Monte Carlo simulation by

1. Generating samples of X_i from an exponential distribution with mean 1.
2. Compute the weighted sum $S = \sum_{i=1}^{5} iX_i$.
3. Then, we repeat this process many times to estimate the probability θ by calculating the proportion of times $S \geq 21.6$.

Problem 3

Let X_i, $i = 1, \ldots, 5$, be independent exponential random variables each with mean 1, and consider the quantity

$$\theta = P\left(\sum_{i=1}^{5} iX_i \geq 21.6\right).$$

Part B

To use antithetic variables,

1. Generate uniform random variables $U_i \sim \text{Uniform}(0, 1)$.
2. Compute $X_i = -\ln(U_i)$ and $X_i' = -\ln(1 - U_i)$.
3. Calculate $S = \sum_{i=1}^{5} iX_i$ and $S' = \sum_{i=1}^{5} iX_i'$.
4. Use the average indicator function $\frac{I(S \geq 21.6) + I(S' \geq 21.6)}{2}$ as the estimator.

Problem 3

Let X_i, $i = 1, \ldots, 5$, be independent exponential random variables each with mean 1, and consider the quantity

$$\theta = P\left(\sum_{i=1}^{5} iX_i \geq 21.6\right).$$

Part C

To determine efficiency, we compare the variances of the standard estimator, $\mathrm{Var}(S)$, and the antithetic estimator, $\mathrm{Var}(A)$. If the antithetic estimator has a lower variance, it is more efficient.

Putting it into Practice

```r
set.seed(202425)
N <- 100000
# Standard estimator
standardResults <- replicate(N, {
  X <- rexp(5); S <- sum((1:5) * X); S >= 21.6
})
thetaS <- mean(standardResults)
varS <- var(standardResults)
# Antithetic estimator
antitheticResults <- replicate(N / 2, {
  U <- runif(5); X <- -log(U); X_prime <- -log(1 - U)
  S <- sum((1:5) * X); S_prime <- sum((1:5) * X_prime)
  c(S >= 21.6, S_prime >= 21.6)
})
# Compute the mean of each pair (average of S and S')
antitheticMeans <- rowMeans(matrix(antitheticResults, ncol = 2))
thetaA <- mean(antitheticMeans)
varA <- var(antitheticMeans)
print(paste0("Variance of standard estimator:", varS))
```

```
[1] "Variance of standard estimator:0.142069459794319"
```

```r
print(paste0("Variance of antithetic estimator:", varA))
```

```
[1] "Variance of antithetic estimator:0.0700363111261794"
```

If $\mathrm{Var}(S) < \mathrm{Var}(A)$, then the antithetic variables method is more efficient. Based on the simulation, the antithetic estimator shows reduced variance, indicating increased efficiency. To emphasis, ‘varA

Problem 10

In certain situations, a random variable X, whose mean is known, is simulated to obtain an estimate of $P\{X \leq a\}$ for a given constant a. The raw simulation estimator from a single run is I, where

$$I = \begin{cases} 1 & \text{if } X \leq a, \\ 0 & \text{if } X > a. \end{cases}$$

Because I and X are negatively correlated, a natural attempt to reduce the variance is to use X as a control variable and use an estimator of the form

$$I + c(X - \mathbb{E}[X]).$$

Part A

For $X \sim \text{Uniform}(0, 1)$, we know the following,

- $\mathbb{E}[X] = 0.5$
- $\text{Var}(X) = \frac{1}{12}$
- $P(X \leq a) = a$
- $\text{Var}(I) = a(1 - a)$

Using this, we can compute the covariance between I and X by,

$$\text{Cov}(I, X) = \mathbb{E}[IX] - \mathbb{E}[I]\mathbb{E}[X] = \frac{a^2}{2} - a \times 0.5 = \frac{a(a-1)}{2}$$

The optimal c^* is calculated by,

$$c^* = -\frac{\text{Cov}(I, X)}{\text{Var}(X)} = -\frac{\frac{a(a-1)}{2}}{\frac{1}{12}} = -6a(a-1)$$

Now, the variance reduction,

$$\text{Var}(I) - \frac{\text{Cov}(I, X)^2}{\text{Var}(X)}$$

$$\text{Percentage Reduction} = \frac{\text{Cov}(I, X)^2}{\text{Var}(I)\,\text{Var}(X)} \times 100\% = 3a(1-a) \times 100\%$$

Problem 10

In certain situations, a random variable X, whose mean is known, is simulated to obtain an estimate of $P\{X \leq a\}$ for a given constant a. The raw simulation estimator from a single run is I, where

$$I = \begin{cases} 1 & \text{if } X \leq a, \\ 0 & \text{if } X > a. \end{cases}$$

Because I and X are negatively correlated, a natural attempt to reduce the variance is to use X as a control variable and use an estimator of the form

$$I + c(X - \mathbb{E}[X]).$$

Part B

For $X \sim \text{Exponential}(1)$, we know the following attributes,

- $\mathbb{E}[X] = 1$
- $\text{Var}(X) = 1$
- $P(X \leq a) = 1 - e^{-a}$
- $\text{Var}(I) = (1 - e^{-a})e^{-a}$

Next, compute $\text{Cov}(I, X)$,

$$\text{Cov}(I, X) = \mathbb{E}[IX] - \mathbb{E}[I]\mathbb{E}[X] = (-ae^{-a} + 1 - e^{-a}) - (1 - e^{-a})(1) = -ae^{-a}$$

To find the optimal c^*, do the following,

$$c^* = -\frac{\text{Cov}(I, X)}{\text{Var}(X)} = ae^{-a}$$

Here is the variance reduction,

$$\text{Percentage Reduction} = \frac{(\text{Cov}(I, X))^2}{\text{Var}(I)\,\text{Var}(X)} \times 100\% = \frac{a^2 e^{-2a}}{(1 - e^{-a})e^{-a}} \times 100\% = \frac{a^2 e^{-a}}{1 - e^{-a}} \times 100\%$$

Problem 10

In certain situations, a random variable X, whose mean is known, is simulated to obtain an estimate of $P\{X \leq a\}$ for a given constant a. The raw simulation estimator from a single run is I, where

$$I = \begin{cases} 1 & \text{if } X \leq a, \\ 0 & \text{if } X > a. \end{cases}$$

Because I and X are negatively correlated, a natural attempt to reduce the variance is to use X as a control variable and use an estimator of the form

$$I + c(X - \mathbb{E}[X]).$$

Part C

The indicator I is 1 when $X \leq a$ and 0 otherwise. Larger values of X (greater than a) correspond to $I = 0$. Therefore, as X increases, I tends to decrease, indicating a negative correlation between I and X.

Problem 12

Part A

In Exercise 1, $\theta = \int_0^1 e^{x^2}\,dx$. We can use e^x as a control variable since its expected value $\mathbb{E}[e^X]$ is known for $X \sim \text{Uniform}(0,1)$. The control variate estimator is:

$$Y = e^{U^2} + c(e^U - \mathbb{E}[e^U])$$

where c is chosen to minimize the variance (calculated in part b).

Problem 12

Part B

Putting it into Practice

```
set.seed(202425)
N <- 100
U <- runif(N)
eU2 <- exp(U^2)
eU <- exp(U)
E_eU <- (exp(1) - 1)
cov_eU2_eU <- cov(eU2, eU)
var_eU <- var(eU)
cStar <- -cov_eU2_eU / var_eU
# Control variate estimator
YControl <- eU2 + cStar * (eU - E_eU)
varControl <- var(YControl)
print(paste0("Optimal c*: ", cStar))
```

```
[1] "Optimal c*: -0.988045042115672"
```

```
print(paste0("Variance of control variate estimator:", varControl))
```

```
[1] "Variance of control variate estimator:0.0134246377049571"
```

Problem 12

Part C

```r
e_U2_antithetic <- exp(U^2) + exp((1 - U)^2)
YAntithetic <- e_U2_antithetic / 2
varAntithetic <- var(YAntithetic)
print(paste0("Variance of antithetic estimator:", varAntithetic))
```

```
[1] "Variance of antithetic estimator:0.0303702833131082"
```

Problem 12

Part D

By comparing $\mathrm{Var}(C)$ and $\mathrm{Var}(A)$, we determine which method provided greater variance reduction. In our simulation, it appears that $\mathrm{Var}(C) < \mathrm{Var}(A)$, so the control variate estimator is more efficient.

Problem 15

Show that in estimating

$$\theta = \mathbb{E}\left[\sqrt{1-U^2}\right]$$

it is better to use U^2 rather than U as the control variate. Use simulation to approximate the necessary covariances.

We compare the effectiveness of using U and U^2 as control variates by doing a mini-simulation example,

Putting it into Practice

```r
set.seed(202425)
N <- 10000
U <- runif(N)
Y <- sqrt(1 - U^2)
# Using U as control variate
C1 <- U
EC1 <- 0.5
covYC1 <- cov(Y, C1)
varC1 <- var(C1)
c1_star <- -covYC1 / varC1
Y1 <- Y + c1_star * (C1 - EC1)
var_Y1 <- var(Y1)
# Using U^2 as control variate
C2 <- U^2
EC2 <- 1/3
cov_Y_C2 <- cov(Y, C2)
varC2 <- var(C2)
c2_star <- -cov_Y_C2 / varC2
Y2 <- Y + c2_star * (C2 - EC2)
varY2 <- var(Y2)
print(paste0("Variance using U as control variate:", var_Y1))
```

```
[1] "Variance using U as control variate:0.00759764689972037"
```

```r
print(paste0("Variance using U^2 as control variate:", varY2))
```

```
[1] "Variance using U^2 as control variate:0.00163425781626914"
```

The variance when using U^2 as the control variate, $\mathrm{Var}(Y_2)$, is smaller than when using U, $\mathrm{Var}(Y_1)$. Thus, U^2 is a better control variate for estimating θ in this scenario.