hw 6 solution

Statistical Computing, Jieun Shin

Autumn 2022

문제 1.

1. $I = \int_1^2 e^{-x}dx$을 적중법과 표본평균 몬테칼로 적분법으로 추정하고자 한다. 먼저 표본평균 몬테칼로 적분법은 다음의 과정을 따른다.

(1) 균일분포의 범위를 $[a, b]$라고 할 때, $a = 1, b = 2$로 지정한 후 $U(1, 2)$에서 $x_1, \ldots, x_N$을 생성한다.

(2) 추정값 $\hat{I}_M = (b-a)\frac{1}{N}\sum_{i=1}^{N} g(x_i) = \frac{1}{N}\sum_{i=1}^{N} e^{-x_i}$을 계산한다.

```
set.seed(123)
N = 10000
a = 1; b = 2; c = 1

# 표본평균 몬테칼로 적분법
x = runif(N, a, b)
(b - a) * mean( exp(-x) )  # 추정값
```

```
## [1] 0.2329801
```

적중법은 다음의 과정을 따른다.

(1) X좌표에 해당하는 N개의 값 $x_1, \ldots, x_N$을 $U(0, 2)$에서 생성하고, Y좌표에 해당하는 N개의 값 $y_1, \ldots, y_N$를 $U(0, 1)$에서 생성한다.

(2) 추정값 $\hat{I}_H = c(b-a)\hat{p} = \frac{1}{N}\sum_{i=1}^{N} I(y_i \le g(x_i))$을 계산한다. 여기서 $\hat{p}$는 전체 난수발생 갯수 중 $g(x_i)$ 아래에 들어간 난수의 갯수이다.

```
# 적중법
x = runif(N, a, b)
y = runif(N, 0, c)
p = mean( y <= exp(-x))

c * (b - a) * p                # 추정값
```

```
## [1] 0.2339
```

```
c^2 * (b-a)^2 * p*(1-p) / N  # 추정치의 분산    #######구해야 함
```

```
## [1] 1.791908e-05
```

결과를 보면 표본평균 몬테칼로 적분법과 적중법 모두 0.233 근처의 값으로 I를 추정하였다.

```
# 참값
g = function(x) exp(-x)

integrate(g, 1, 2)
```

```
## 0.2325442 with absolute error < 2.6e-15
```

문제 2.

1. $h(x) = \sqrt{1-x^2}$라고 할 때, 참 값 $I = \mathbb{E}[h(X)]$을 추정하는 방법 중 대조변수법을 이용한 추정량 $\hat{I}_2$을 다음의 과정에 따라 정의할 수 있다.

(1) 먼저 $U_1, \dots, U_n \sim U(0,1)$ 을 생성한다.

(2) 역변환법을 통해 $N(0,1)$을 따르는 정규난수 X를 생성한다. 정규난수는 $X = (F_X^{-1}(U_1), \dots, F_X^{-1}(U_n))$와 $X^{'} = (F_X^{-1}(1-U_1), \dots, F_X^{-1}(1-U_n))$을 생성한다.

(3) $N = 2n$이라 하면 대조변수법을 사용한 추정량은 다음과 같이 표현된다.

$$\hat{I}_2 = \frac{1}{N}\sum_{i=1}^{N/2}[h(X_i) + h(X_i^{'})]$$

2. $\hat{I}_1$와 $\hat{I}_2$의 분산은 다음과 같이 구할 수 있다.

$$\mathrm{Var}(\hat{I}_1) = \mathrm{Var}\left(\frac{1}{N}\sum_{i=1}^{N} h(X_i)\right) = \frac{1}{N}\mathrm{Var}(h(X_i)),$$

$$\mathrm{Var}(\hat{I}_2) = \mathrm{Var}\left(\frac{1}{N}\sum_{i=1}^{N/2}[h(X_i) + h(X_i^{'})]\right) = \frac{1}{2N}\left[\mathrm{Var}(h(X_i)) + \mathrm{Var}(h(X_i^{'})) + 2\mathrm{Cov}(h(X_i), h(X_i^{'}))\right]$$

결과를 보면 두 추정량은 1.5로 비슷한 값을 가지고 분산은 대조변수법의 분산이 더 작으며, 대조변수법에서 분산 감소를 비율이 약 95%임을 확인하였다.

```r
set.seed(123)
N = 10000
n = N / 2

h = function(x) sqrt(1-x^2)
integrate(h, 0, 1)
```

```
## 0.7853983 with absolute error < 0.00011
```

```r
rh = function(n){
  u = sort(runif(N))
  csum = cumsum(h(u))
}

# 표본평균 몬테칼로 추정량
X = runif(N)
h1 = h(X)
mean1 = mean(h1)     # 추정치
var1 = var(h1) / N # 추정치의 분산

mean1; var1
```

```
## [1] 0.788531
```

```
## [1] 4.848323e-06
```

```r
# 대조변수법을 이용한 추정량
X1 = runif(n)
X2 = -X1
X = c(X1, X2)
```

```r
mean2 = mean(h(X))    # 추정치
var2 = ( var(h(X1)) + var(h(X2)) + 2 * cov(h(X1), h(X2)) ) / (2 * N)  # 추정치의 분산

mean2; var2
```

```
## [1] 0.7871315
```

```
## [1] 9.745845e-06
```

```r
# 분산 감소 비율 비교
( var1 - var2) / var1
```

```
## [1] -1.010148
```