# hw 6 solution

Statistical Computing, Jieun Shin

Autumn 2021

**문제 1.**

1. $h(x) = x/(2^x - 1)$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 $U_1, \ldots, U_n \sim U(0,1)$ 을 생성한다.

(2) 역변환법을 통해 $N(0,1)$을 따르는 정규난수 $X$를 생성한다. 정규난수는 $X = (F_X^{-1}(U_1), \ldots, F_X^{-1}(U_n))$와 $X^{'} = (F_X^{-1}(1-U_1), \ldots, F_X^{-1}(1-U_n))$을 생성한다. 단, 이 문제의 경우 $X^{'} = (1 - X_1, \ldots, 1 - X_n)$와 동일하다.

(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) x / (2^x - 1)

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

mean1; var1
```

```
## [1] 1.500168
## [1] 2.552766e-05
```

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

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] 1.499257
```

```
## [1] 1.196732e-06
```

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

```
## [1] 0.9531202
```

**문제 2.**

1. $I = \int_0^{\pi/2} \cos x \, dx$을 적중법과 표본평균 몬테칼로 적분법으로 추정하고자 한다. 먼저 표본평균 몬테칼로 적분법은 다음의 과정을 따른다.

(1) 균일분포의 범위를 $[a, b]$라고 할 때, $a = 0, b = \frac{1}{\pi}$로 지정한 후 $U(0, \frac{\pi}{2})$에서 $x_1, \ldots, x_N$을 생성한다.

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

```r
set.seed(123)
N = 10000
a = 0; b = pi/2; c = 1

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

```
## [1] 1.005431
```

적중법은 다음의 과정을 따른다.

(1) $X$좌표에 해당하는 $N$개의 값 $x_1, \ldots, x_N$을 $U(0, \frac{\pi}{2})$에서 생성하고, $Y$좌표에 해당하는 $N$개의 값 $y_1, \ldots, y_N$를 $U(0, 1)$에서 생성한다.

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

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

c * (b - a) * p  # 추정값
```

```
## [1] 0.9998119
```

결과를 보면 표본평균 몬테칼로 적분법과 적중법 모두 1 근처의 값으로 $I$를 추정하였다.

표본 크기는

$$
\begin{aligned}
P(|\hat{I} - I| < \epsilon) > 1 - \alpha \\
\geq 1 - \frac{Var(\hat{I})}{\epsilon^2} \geq 1 - \alpha \quad \because \text{chebyshev inequality}
\end{aligned}
$$

을 만족하도록 하며, 표본평균 몬테칼로 적분법에서는

$$
\begin{aligned}
1 - \frac{Var(\hat{I})}{\epsilon} &\geq 1 - \alpha \\
\frac{Var(\hat{I})}{\epsilon^2} &\leq \alpha \\
\frac{(b-a)^2}{N}\left(\int_a^b g(x)^2\ dx - I^2\right) &\leq \alpha\epsilon^2 \\
\frac{(b-a)^2}{\alpha\epsilon^2}\left(\int_a^b g(x)^2\ dx - I^2\right) &\leq N
\end{aligned}
$$

을 만족하는 $N$을 구하고, 적중법에서는

$$
\begin{aligned}
1 - \frac{Var(\hat{I})}{\epsilon} &\geq 1 - \alpha \\
\frac{Var(\hat{I})}{\epsilon} &\leq \alpha \\
c^2(b-a)^2\frac{p(1-p)}{N} &\leq \alpha\epsilon^2 \\
\frac{c^2(b-a)^2}{4\alpha\epsilon^2} &\leq N
\end{aligned}
$$

을 만족하는 $N$을 구한다.

결과를 보면 몬테칼로 적분법에서 필요한 표본 크기는 약 107,632,281개, 적중법에서 필요한 표본 크기는 약 61,685,028개로 적중법에서 필요한 표본크기가 더 적다.

```r
# 표본 크기 구하기
alpha = 0.01
e = 10^(-3)

g = function(x) cos(x)
g1 = function(x) x * cos(x)          # g1(X) = x cos(x)
g2 = function(x) x * cos(x)^2        # g2(X) = x cos^2(x)

I = integrate(g, a, b)$value    ; I # I의 이론적인 값 = 1
```

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

```r
I1 = integrate(g1, a, b)$value  ; I1 # E[g(X)]의 이론적인 값 = 0.5708
```

```
## [1] 0.5707963
```

```r
I2 = integrate(g2, a, b)$value  ; I2 # E[g(X)^2]의 이론적인 값 = 0.3669
```

```
## [1] 0.3668503
```

```r
# 표본평균 몬테칼로 적분법의 표본크기
(b - a)^2 * ( I1 - I2^2 ) / ( alpha * e^2 )
```

```
## [1] 107632281
```

```r
# 적중법의 표본크기
( (c * (b - a)^2) ) / (4 * alpha * e^2)
```

```
## [1] 61685028
```

2. $f_1(x)$을 주함수로 한 주표본기법 추정치를 $\hat{I}_{f1}$, $f_2(x)$을 주함수로 한 주표본기법 추정치를 $\hat{I}_{f2}$라 하자. 다음 과정에 따라 $\hat{I}_{f1}$와 $\hat{I}_{f2}$을 추정할 수 있다.

(1) $U(0, \frac{\pi}{2})$로부터 $x_1, \ldots, x_N$을 생성한다.

(2) 각 추정치 $\hat{I}_{f1} = \frac{1}{N}\sum_{i=1}^{N}\frac{g(x_i)}{f_1(x_i)} = \frac{\pi}{2}\frac{1}{N}\sum_{i=1}^{N}\cos x_i$와 $\hat{I}_{f2} = \frac{1}{N}\sum_{i=1}^{N}\frac{g(x_i)}{f_2(x_i)} = \frac{1}{N}\sum_{i=1}^{N}\frac{\cos x_i}{\frac{4}{\pi}(1-\frac{2}{\pi}x_i)}$를 계산하여 구한다.

주표본 추정치에 대한 분산은 $\mathrm{Var}(\hat{I}) = \frac{1}{N}\left(\int_a^b \frac{g^2(x_i)}{f_1(x_i)}dx - I^2\right)$ 이므로 $\hat{I}_{f1}$의 분산은 $\mathrm{Var}(\hat{I}_{f1}) = \frac{1}{N}\left(\int_0^{\frac{\pi}{2}} \frac{\cos^2 x_i}{\frac{2}{\pi}}dx - I^2\right)$ 으로, $\hat{I}_{f2}$의 분산은 $\mathrm{Var}(\hat{I}_{f1}) = \frac{1}{N}\left(\int_0^{\frac{\pi}{2}} \frac{\cos^2 x_i}{\frac{4}{\pi}(1-\frac{2}{\pi}x_i)}dx - I^2\right)$으로 계산할 수 있다.

결과를 보면 주함수가 $f_1$일 때와 주함수가 $f_2$일 때 모두 주표본 추정치는 약 1이며, 분산을 비교하면 $f_1$의 분산이 $f_2$의 분산보다 더 작게 나타났다.

```r
g = function(x) cos(x)
f1 = function(x) (2 / pi)
f2 = function(x) (1 - (2 / pi) * x) * (4 / pi)

gf1 = function(x) cos(x)^2 * (pi / 2)
gf2 = function(x) cos(x)^2 / ( (1 - (2 / pi) * x) * (pi / 4) )

set.seed(123)
u = runif(N, a, b)

# 주함수를 f1으로 했을 때의 주표본 추정치와 분산
mean( g(u) / f1(u) ) ; (integrate(gf1, a, b)$value - I^2) / N
```

```
## [1] 1.005431
## [1] 2.337006e-05
```

```r
# 주함수를 f2으로 했을 때의 주표본 추정치와 분산
mean( g(u) / f2(u) ) ; (integrate(gf2, a, b)$value - I^2) / N
```

```
## [1] 1.075948
## [1] 6.482776e-05
```