hw 5 solution

Statistical Computing, Jieun Shin

Autumn 2021

문제 1.

1. 표준 라플라스 분포$f(x) = \frac{1}{2}e^{-|x|}, \quad -\infty < x < \infty$를 따르는 확률난수를 역변환법에 의해 발생시키고자 한다.

먼저 표준 라플라스 분포의 cdf를 구하자.

(1) $x < 0$일 때,

$$F_X(x) = \mathbb{P}(X \le x) = \int_{-\infty}^{x} \frac{1}{2}e^t dt = \frac{1}{2}[e^t]_{-\infty}^{x} = \frac{1}{2}(e^x - 0) = \frac{1}{2}e^x$$

이고,

(2) $x \ge 0$일 때,

$$F_X(x) = \mathbb{P}(X \le x) = \frac{1}{2} + \int_{0}^{x} \frac{1}{2}e^{-t} dt = \frac{1}{2} - \frac{1}{2}[e^{-t}]_0^x = \frac{1}{2} - \frac{1}{2}(e^{-x} - 1) = 1 - \frac{1}{2}e^{-x}$$

이다.

그러면 표준라플라스 분포를 따르는 확률난수는 $u \sim \text{unif}(0,1)$에 대해 다음의 cdf의 역함수를 이용해 구할 수 있다. 이 때 u의 범위에 따라 달라짐을 유의해야 한다.

$$F^{-1}(u) = \begin{cases} \log(2u), & 0 < u < 0.5 \\ -\log(2(1-u)), & 0.5 < u < 1 \end{cases}$$

코드

```
LaplaceDist = function(N){  # 생성할 난수의 크기를 지정하게 하였음
  n = N / 2
  u1 = runif(n, 0, 1)
  u2 = runif(n, 0, 1)

  x1 = log(2 * u1)
  x2 = -log(2 * (1 - u2))

  return(c(x1, x2))
}
```

2. 먼저 난수의 크기를 10, 100, 500, 1000일 때 생성한 분포의 평균과 이론적인 평균 0과의 차이를 계산하였다. 대체로 난수를 많이 발생할수록 표본평균이 이론적인 평균과 가까워짐을 확인하였다.

```
set.seed(100)
for(N in c(10, 100, 500, 1000)){
  x = LaplaceDist(N)
  cat("N = ", N, ", mean = ", mean(x), ", error = ", abs(0 - mean(x)), '\n')
}
```

```
## N =  10 , mean =  -0.2987063 , error =  0.2987063
## N =  100 , mean =  0.1254538 , error =  0.1254538
## N =  500 , mean =  0.06887125 , error =  0.06887125
## N =  1000 , mean =  0.02601886 , error =  0.02601886
```

문제 2.

첫 번째 방법은 유한한 값을 출력하는 반면, 두 번째 방법은 무한대로 발산하므로 첫 번째 방법의 결과가 더 정확하다.

```r
N = 1000
x = runif(N)

sum(log(x))
```

```
## [1] -1006.048
```

```r
log(prod(x))
```

```
## [1] -Inf
```

문제 3.

1. 로지스틱 분포 $f(x) = \frac{e^x}{(1+e^x)^2}, \quad -\infty < x < \infty$를 따르는 확률난수를 역변환법에 의해 발생시키고자 한다.

먼저 로지스틱 분포의 cdf를 구한다:

$$F_X(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^{x} \frac{e^x}{(1+e^x)^2} dt = \int_{0}^{e^x} \frac{1}{(1+y)^2} dy = 1 - \frac{1}{1+e^x}.$$

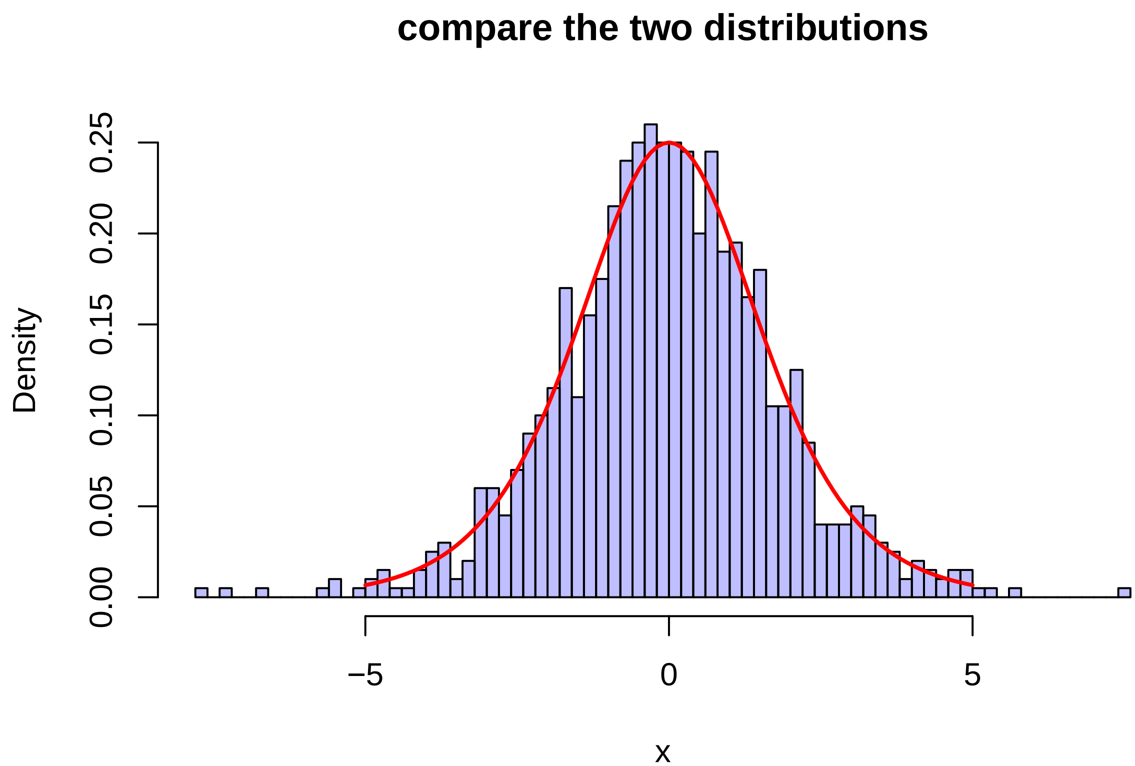
그 다음 $u \sim \text{unif}(0,1)$를 생성하여 cdf의 역함수

$$F^{-1}(u) = \log\left(\frac{1}{1-u} - 1\right)$$

에 의해 로지스틱 분포를 따르는 확률난수를 생성할 수 있다. 두 히스토그램을 비교해 보았을 때 역법환법으로 생성한 분포 (파란색)가 이론적 분포 (붉은색)와 비슷하게 생성됨을 확인할 수 있다.

```r
set.seed(123)
N = 1000
u = runif(N)
x = log(1/(1 - u) - 1)

# 그래프 비교
hist(x, breaks = 100, col = rgb(0, 0, 1, 1/4), freq = FALSE,
     main = "compare the two distributions", xlab = "x")      #  히스토그램: 역변환법으로 생성한 분포
curve(dlogis, from = -5, to = 5, add = TRUE, col = 'red', lwd = 2) # curve: 이론적 분포
```



2. 표준 정규분포를 따르는 난수를 발생시키기 위한 기각법 알고리즘을 유도해보자.

목표가 되는 분포 $f(x)$를 표준정규분포라 하자. 기각법 알고리즘을 위해 $f(x) = C \cdot h(x) \cdot g(x), \quad C \geq 1$로 나누어야 한다. 여기서 $h(x)$를 로지스틱 분포의 pdf라 하면 각 $C, g(x)$를 다음과 같이 설정할 수 있다:

$$C = \frac{4}{\sqrt{2\pi}}, \quad h(x) = \frac{e^x}{(1+e^x)^2}, \quad g(x) = \frac{(1+e^x)^2}{4e^x} e^{-\frac{x^2}{2}}.$$

기각법 알고리즘

다음의 과정을 10,000번 반복한다.

(1) $u \sim \text{unif}(0,1)$

(2) $x = \log\left(\frac{1}{1-u} - 1\right)$

(3) 만약 $u \leq \frac{h(x)}{Cf(x)}$이면 x를 반환하고, 그렇지 않으면 (1)로 돌아간다.

두 히스토그램을 비교해 보았을 때 기각법으로 생성한 분포 (파란색)가 이론적 분포 (붉은색)와 비슷하게 생성됨을 확인할 수 있다.

코드

```
set.seed(123)
N = 10000

# algorithm
u1 =  runif(N)          # step 1
u2 =  runif(N)
```

```
x = log(1/(1 - u1) - 1) # step 2

g = exp(-0.5 * x^2) * exp(-x) * (1 + exp(x))^2 / 4
y = ifelse(u2 <= g, x, NA)  # stpe 3
y = y[!is.na(y)]

# 그래프 비교
hist(y, breaks = 100, col = rgb(0, 0, 1, 1/4), freq = FALSE,
          main = "compare the two distributions", xlab = "x")   #  히스토그램: 기각법으로 생성한 분포
curve(dnorm, from = -5, to = 5, add = TRUE, col = 'red', lwd = 2) # curve: 이론적 분포
```