DATA 405 Assignment 3

Rin Meng, 51940633

Question 1

(a) Quantile Function: The CDF of V is:

$$F_V(x) = \begin{cases} 1 - e^{-x^2}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

To find the quantile function, solve for x in terms of u:

$$\begin{aligned} F_V(x) = u \Rightarrow 1 - e^{-x^2} &= u \\ e^{-x^2} &= 1 - u \\ -x^2 &= \ln(1-u) \\ x &= \sqrt{-\ln(1-u)} \end{aligned}$$

Therefore, the quantile function $Q(u)$ for V is:

$$Q(u) = \sqrt{-\ln(1-u)}$$

Then, we can write the `rmyV` function using the quantile function:

```
rmyV <- function(n) {
  u <- runif(n)
  x <- sqrt(-log(1 - u))
  return(x)
}
```

(b) The density function can be found by differentiating $F_v(x)$,

$$f_v(x) = \frac{d}{dx} F_V(x) = \frac{d}{dx}(1 - e^{-x^2}) = 2xe^{-x^2}, x > 0$$

Then we can start coding for the simulation and plotting:

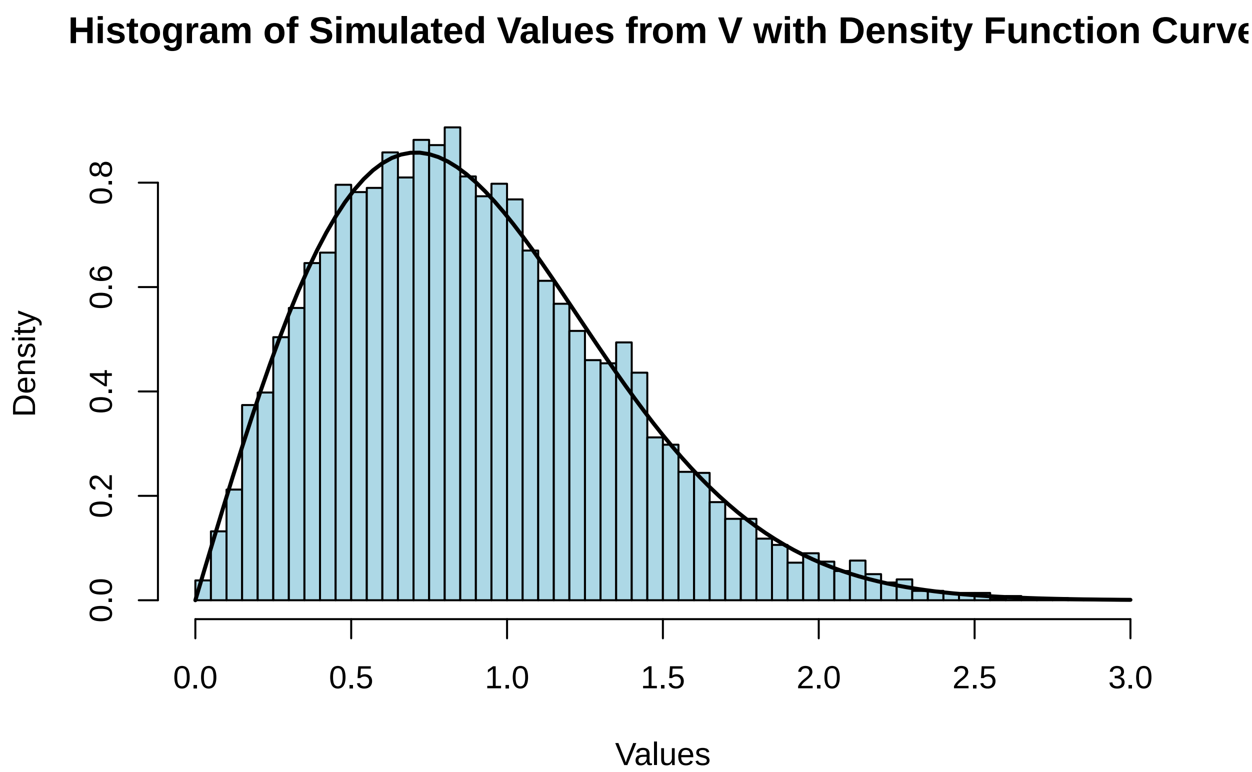
```
dV <- function(x){
  ifelse(x > 0, 2 * x * exp(-x^2), 0)
}

set.seed(51940633)
```

```r
simVals <- rmyV(10000)

hist(simVals, probability = TRUE, breaks = 50,
main = "Histogram of Simulated Values from V with Density Function Curve",
xlab = "Values", col = "lightblue", border = "black")

curve(dV(x), add = TRUE, col = "black", lwd = 2)
```



Question 2

(a) The cumulative distribution of the function X can be written as:

$$F_X(x) = \int f_X(x)\,dx = \int 3x^2\,dx = x^3$$

So the CDF for $F_X(x)$ is:

$$F_X(x) = \begin{cases} x^3, & x \in [0,1] \\ 0, & x < 0 \\ 1, & x > 1 \end{cases}$$

(b) Now we can write the quantile function $Q(u)$, found by solving $F_X(x) = u$ for x, where u is a uniform random variable on $[0, 1]$.

$$F_X(x) = u \Rightarrow x^3 = u$$

$$x = u^{1/3}$$

So then the quantile function $Q(u)$ is,

$$Q(u) = u^{1/3}$$

Now we can write the R function to generate random variates:

```
rmyX <- function(n){
  u <- runif(n)
  x <- u^(1/3)
  return(x)
}
```

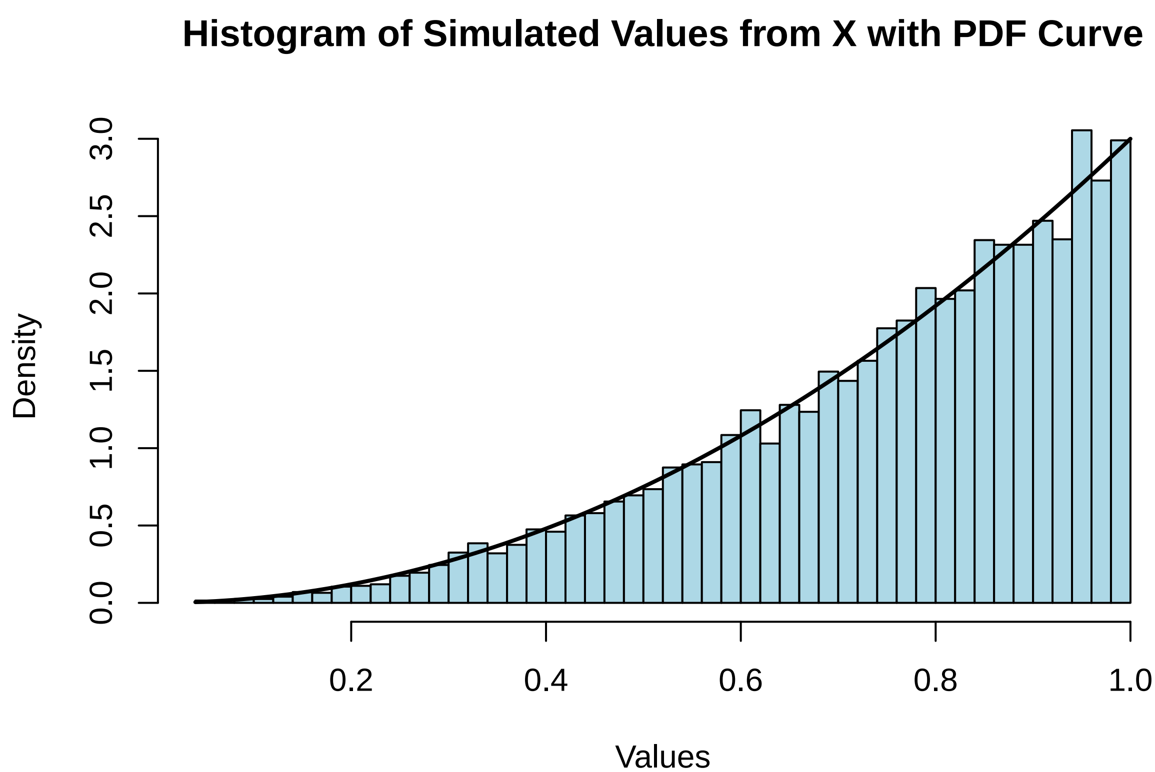
Now we define the PDF function for x, and simulate the distribution:

```
dX <- function(x){
  ifelse(x >= 0 & x <= 1, 3 * x^2, 0)
}

set.seed(51940633)
simVals <- rmyX(10000)

hist(simVals, probability = TRUE, breaks = 50,
main = "Histogram of Simulated Values from X with PDF Curve",
xlab = "Values", col = "lightblue", border = "black")

curve(dX(x), add = TRUE, col = "black", lwd = 2)
```



Question 3

(a) The CDF of Y is given as:

$$G(y) = pF_V(y) + (1-p)F_X(y)$$

Where $F_v(y) = 1 - e^{-y^2}$ is from question 1 and

$$F_X(y) = \begin{cases} y^3, & x \in [0,1] \\ 0, & y < 0 \\ 1, & y > 1 \end{cases}$$

from question 2.

(b) We can now generate random variates from Y,

```
rmyY <- function(n, p) {
  # Step 1: Determine which distribution to sample from
  from_V <- rbinom(n, 1, p)  # Vector of 0s and 1s, where 1 means sample from V

  # Step 2: Generate samples from the appropriate distribution
  samples <- numeric(n)
  samples[from_V == 1] <- rmyV(sum(from_V == 1))  # Generate from V where from_V == 1
  samples[from_V == 0] <- rmyX(sum(from_V == 0))  # Generate from X where from_V == 0

  return(samples)
}
```

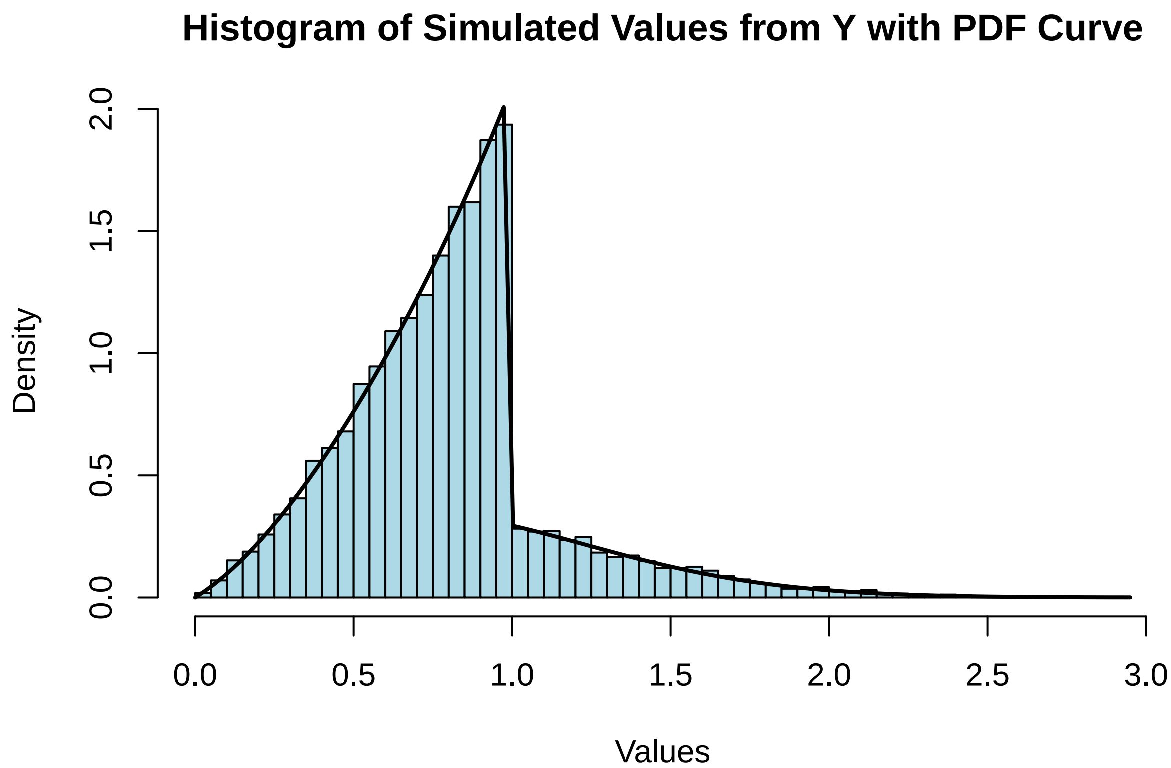
(c) Simulate the distribution of Y, for cases where p = 0.4:

```
gY <- function(y, p) {
  p * (2 * y * exp(-y^2)) + (1 - p) * (ifelse(y >= 0 & y <= 1, 3 * y^2, 0))
}

set.seed(51940633)  # for reproducibility
simulated_values_Y <- rmyY(10000, p = 0.4)

hist(simulated_values_Y, probability = TRUE, breaks = 50,
     main = "Histogram of Simulated Values from Y with PDF Curve",
     xlab = "Values", col = "lightblue", border = "black")

curve(gY(x, p = 0.4), add = TRUE, col = "black", lwd = 2)
```


Histogram of Simulated Values from Y with PDF Curve
Density
0.0
0.5
1.0
1.5
2.0
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Values