Problem 1

HW1

Hanyuan Chi(chixx105), Zhi Shen(shenx704)

April 09, 2017

```
suppressPackageStartupMessages({
  library(ggplot2)
  library(dplyr)
})
```

Inverse-Transform Sampling

Question 1

Please simulate 2000 random samples from the distribution with the following CDF

$$F(x) = \begin{cases} 1 - e^{-x^2} & \text{if } x \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

- Hints:
  - This question is easier since I provided a CDF to you right away
  - This is a special case of Weibull distribution
    - You will use this (type of a) distribution frequently if you do survival analysis and model "time till death" kind of scenarios.
- Output:
  - Please create a *data.frame* `df1` that contains numeric vector `df1$X` with the generated sample
  - Your file should knit in less than 1 minute

```
N <- 2000L
set.seed(123456) # PLEASE DO NOT CHANGE THE SEED


# Please write your code below
df1 <- data.frame(u = runif(N)) %>%
      mutate(X = sqrt(-log(1-u)))
```

Question 2

Please simulate 2000 random samples from the distribution with the following PDF

$$f(x) = \begin{cases} 3x^2 e^{-x^3} & \text{if } x \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

- Hints:
  - Note that this time I gave your a PDF not a CDF
- Output:
  - Please create a *data.frame* `df2` that contains numeric vector `df2$X` with the generated sample
  - Your file should knit in less than 1 minute

```r
N <- 2000L
set.seed(123456) # PLEASE DO NOT CHANGE THE SEED

# Please write your code below
df2 <- data.frame(u = runif(N)) %>%
      mutate(X = (-log(1-u))^(1/3))
```

Question 3

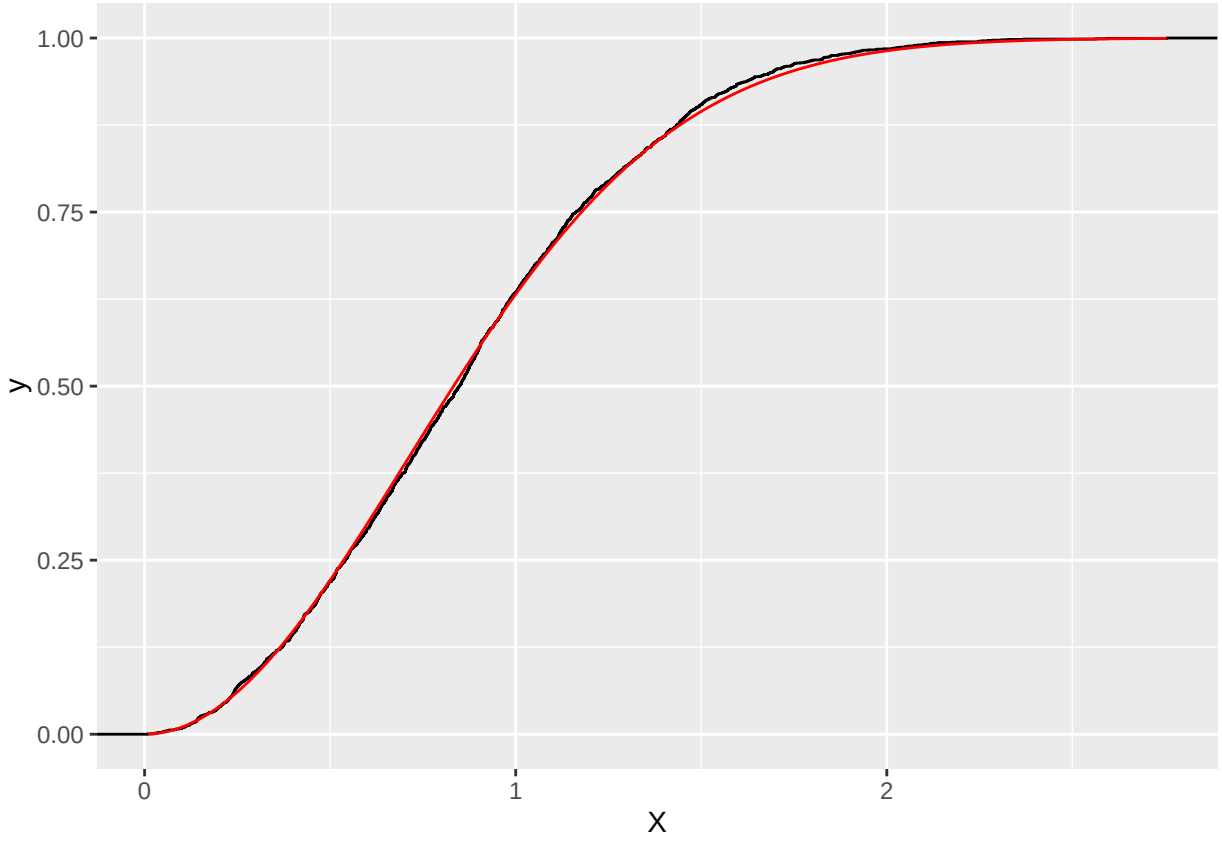
Please verify that your solution for Question 1 is correct by plotting:

- the empirical CDF versus true CDF
- QQPlot of sample quantiles versus true quantiles

Hints:

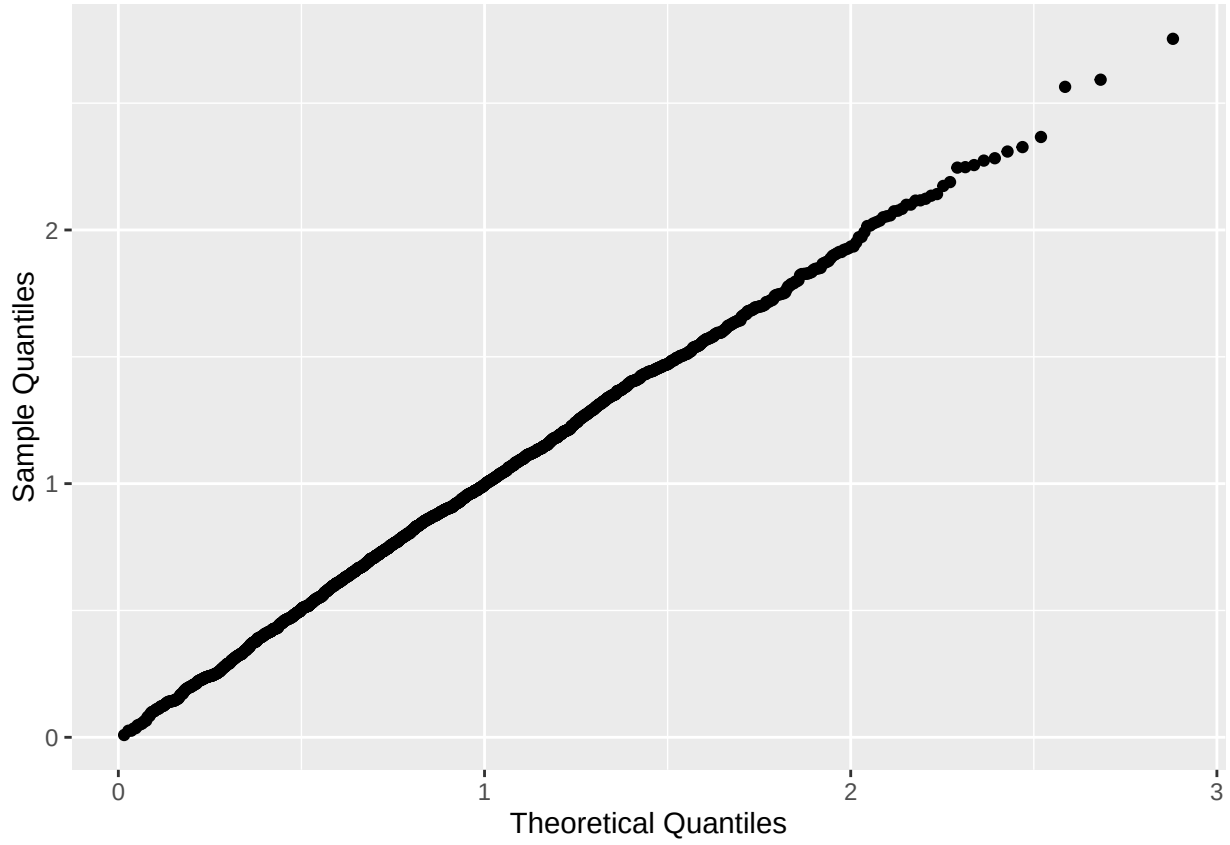
- Take a look at Workday1

```r
p1 <- ggplot(data = df1) +
  stat_ecdf(aes(x = X)) +
  stat_function(aes(x= X),
                fun = function(x) (1-exp(-x^2)),color = 'red')
p1
```




```r
p2 <- ggplot(df1) +
  stat_qq(aes(sample = X),
          distribution = function(p) sqrt(-log(1-p))) +
```

```
  xlab("Theoretical Quantiles") + ylab("Sample Quantiles")
p2
```



Question 4

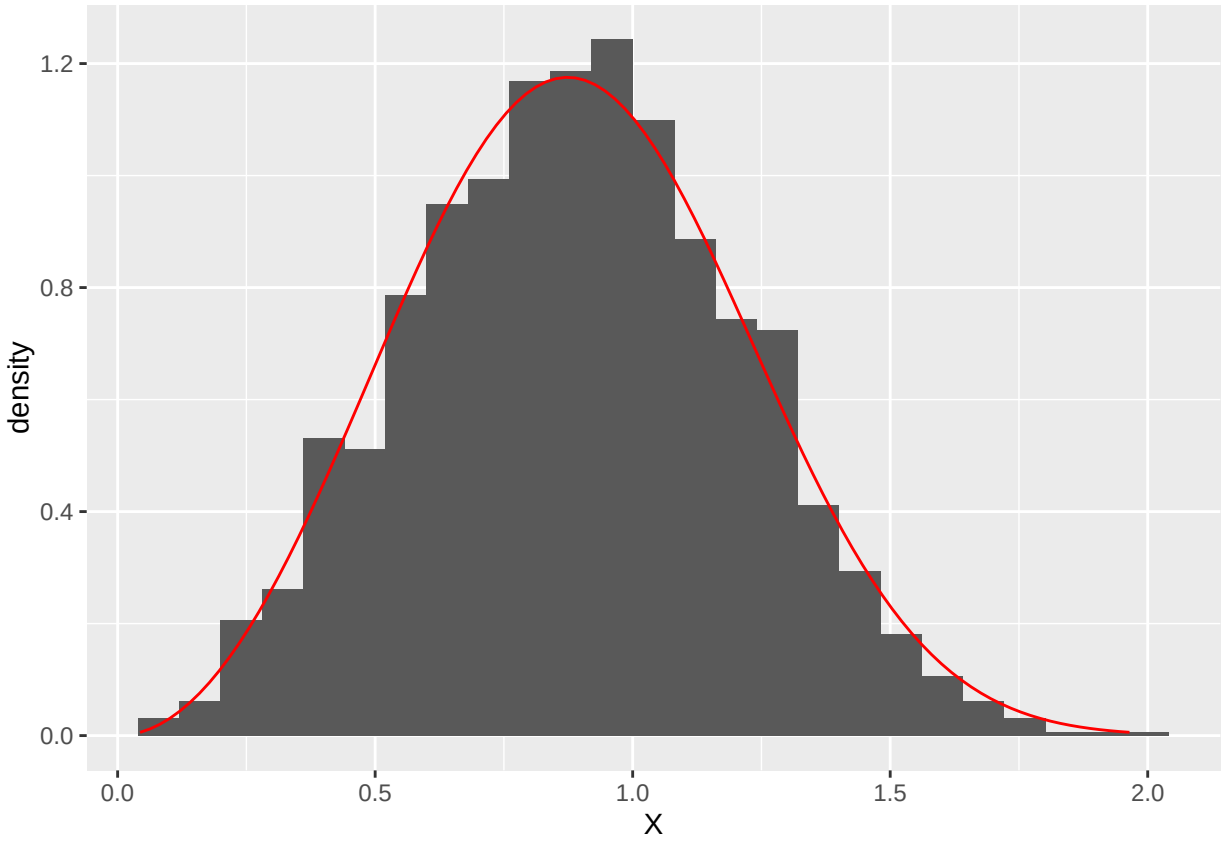
Please verify that your solution for Question 2 is correct by plotting the histogram of the random outcomes versus true PDF

Hints:

- Take a look at Workday1

```
p3 <- ggplot(data = df2) +
  geom_histogram(aes(x=X, y=..density..),bins = 25) +
  stat_function(aes(X), fun = function(x) (3*x^2*exp(-x^3)), color = 'red')

p3
```


1.2
0.8
0.4
0.0
density
0.0
0.5
1.0
1.5
2.0
X