HW1

Haojie Liu

2023-10-05

Problem 1: Suppose X is a geometric random variable with $p = 0.3$, determine the value of k such that $P(X \leq k) \simeq 0.8$

$$X \sim Geometric(0.3)$$
$$P(X \leq k) = 1 - (1-p)^k \simeq 0.8$$
$$1 - 0.7^k \simeq 0.8$$
$$0.7^k \simeq 0.2$$
$$k = \frac{log(0.2)}{log(0.7)}$$

```
paste("k =",log(0.2)/log(0.7))
```

```
## [1] "k = 4.51233802593815"
```

Problem 2: If two random variables, X and Y, are independent, show Var(XY) in terms of the expected values and variances of X and Y.

$$Var(XY) = E((XY)^2) - (E(XY))^2 = E(X^2)E(Y^2) - E(X)^2E(Y)^2$$
$$Var(XY) = (Var(X) + E(X)^2)(Var(Y) + E(Y)^2) - E(X)^2E(Y)^2$$
$$Var(XY) = Var(X)Var(Y) + Var(X)E(Y)^2 + Var(Y)E(X)^2 + E(X)^2E(Y)^2 - E(X)^2E(Y)^2$$
$$Var(XY) = Var(X)Var(Y) + Var(X)E(Y)^2 + Var(Y)E(X)^2$$

Problem 3: At a call center in Los Angeles County, calls come in at an average rate of four calls per minute. Assume that the wait time from one call to the next follows the exponential distribution. Here we concern only with the rate of calling in and ignore the time spent on the phone. Also, assume that the waiting time between calls is independent. That is, a long delay between two calls doesn't make a shorter waiting period for the next call.

$$X \sim exp(4)$$
$$N_X \sim Possion(x4)$$

(a) What is the average time between two successive calls?

$$E(X) = \frac{1}{\lambda} = \frac{1}{4} min$$

(b) What is the probability that exactly five calls occur within a minute?

$$P(N_X = 5) = \frac{e^{-4}4^5}{5!}$$

```
paste("The probability that exactly five calls occur within a minute is", exp(-4)*(4^5)/factorial(5))
```

```
## [1] "The probability that exactly five calls occur within a minute is 0.156293451850532"
```

(c) What is the probability that less than five calls occur within a minute?

$$P(N_X < 5) = \Sigma_{i=0}^{4}\frac{e^{-4}4^i}{i!}$$

```
result <- 0
for(i in 0:4){

  result <- result + exp(-4)*(4^i)/factorial(i)

}

paste("The probability that less than five calls occur within a minute is", result)
```

```
## [1] "The probability that less than five calls occur within a minute is 0.628836935179873"
```

(d) After a call is received, what is the probability that the next call occurs in less than 5 seconds?

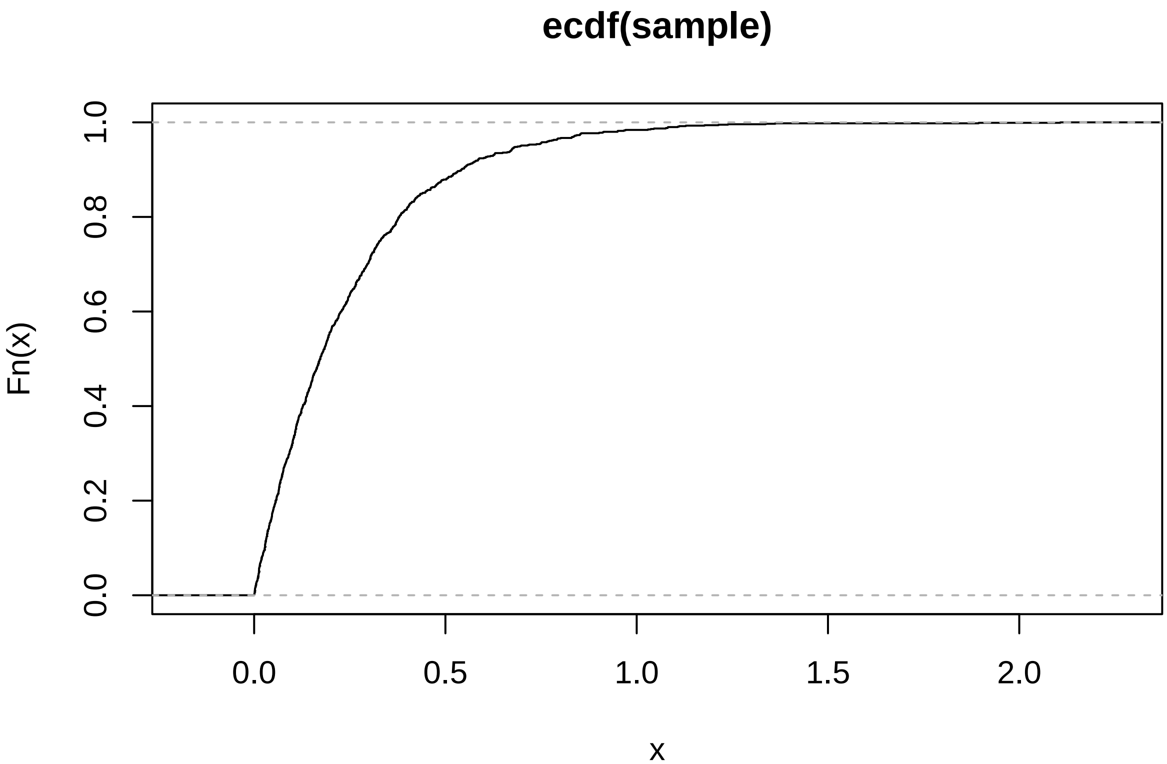
$$x = \frac{5}{60} = \frac{1}{12}$$
$$P(X < \frac{1}{12}) = 1 - e^{-4\frac{1}{12}}$$

```
paste("The probability that the next call occurs in less than 5 seconds is", 1-exp(-4/12))
```

```
## [1] "The probability that the next call occurs in less than 5 seconds is 0.283468689426211"
```

(e) Use rexp function in R to generate 1000 samples with the λ you derived from (d), and then draw a EDF plot with ecdf function. Finally, add confidence interval to your EDF plot.

```
sample <- rexp(1000, 4)
plot(ecdf(sample))
```



Problem 4: Let X be a continuous random variable with cumulative density function F(x). Let $u \sim Unif[0,1]$.

(a) Let $Y = F^{-1}$. Show that $P(Y \leq x) = F(x)$.

$$P(F^{-1}(U) \leq x) = P(U \leq F(x)) = \int_0^{F_X(x)} 1dt = F(x)$$

Therefore $P(Y \leq x) = F(x)$

(b) Show that $F(x) \sim Unif[0,1]$.

Problem 5: Let $X \sim$ two-parameter $Exponential(\lambda, \eta)$ with probability density function

$$f(x) = \lambda e - \lambda(x - \eta), x \geq \eta$$

where λ and η are positive constant.

(a) Find the cumulative density function F(x)

$$F(x) = 1 - e^{-\lambda(x-\eta)}$$

(b) Write an algorithm using the CDF inverse method to generate random numbers from Exponential(λ, η)

$$u = 1 - e^{-\lambda(x-\eta)}$$
$$ln(1-u) = -\lambda(x - \eta)$$

$$\frac{ln(1-u)}{\lambda} = \eta - x$$
$$\eta - \frac{ln(1-u)}{\lambda} = x$$

```
sample_exp <- function(lambda, eta, n){

  u <- runif(n)
  return(eta - ((log(1-u))/lambda))
}
```

(c) Write R code to generate 10,000 random numbers and plot the histogram. You may set λ =the last digit of your UID and η = the second last digit. (If the digit is 0, please have it plus 1)

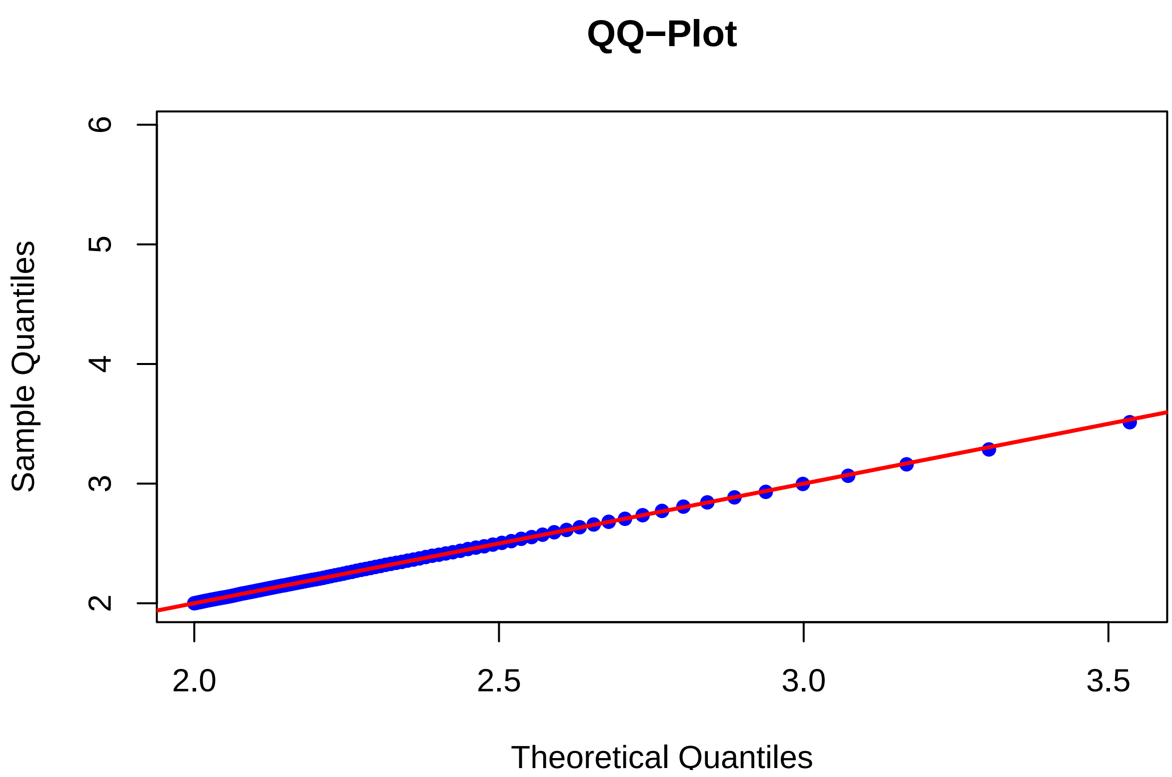
```
result <- sample_exp(3,2,10000)
```

(d) Compare the sample quantiles with the theoretical quantiles.

```
# Theoretical quantiles
p <- seq(0, 1, by = 0.01)
theoretical_quantiles <- 2 - log(1 - p) / 3

# Sample quantiles
sample_quantiles <- quantile(result, probs = p)

# QQ-plot
plot(theoretical_quantiles, sample_quantiles, main = "QQ-Plot",
     xlab = "Theoretical Quantiles", ylab = "Sample Quantiles", pch = 16, col = "blue")
abline(a = 0, b = 1, col = "red", lwd = 2)
```



Problem 6: The Pareto distribution function has a probability density as shown below with the shape parameter α and the scale parameter β.

$$\frac{\alpha\beta^\alpha}{x^{\alpha+1}}$$

$$x > \beta$$

(a) Find the cumulative density function F(x)

$$F(x) = \int_\beta^x \frac{\alpha\beta^\alpha}{t^{\alpha+1}} dt = 1 - (\frac{\beta}{x})^\alpha$$

(b) Write an algorithm using the CDF inverse method to generate random numbers from $Pareto(\alpha = 5, \beta = 1)$

$$u = 1 - (\frac{\beta}{x})^\alpha$$

$$1 - u = (\frac{\beta}{x})^\alpha$$

$$\frac{log(1-u)}{\alpha} = log(\beta) - \log(x)$$

$$log(\beta) - \frac{log(1-u)}{\alpha} = log(x)$$

$$x = e^{log(\beta) - \frac{log(1-u)}{\alpha}}$$

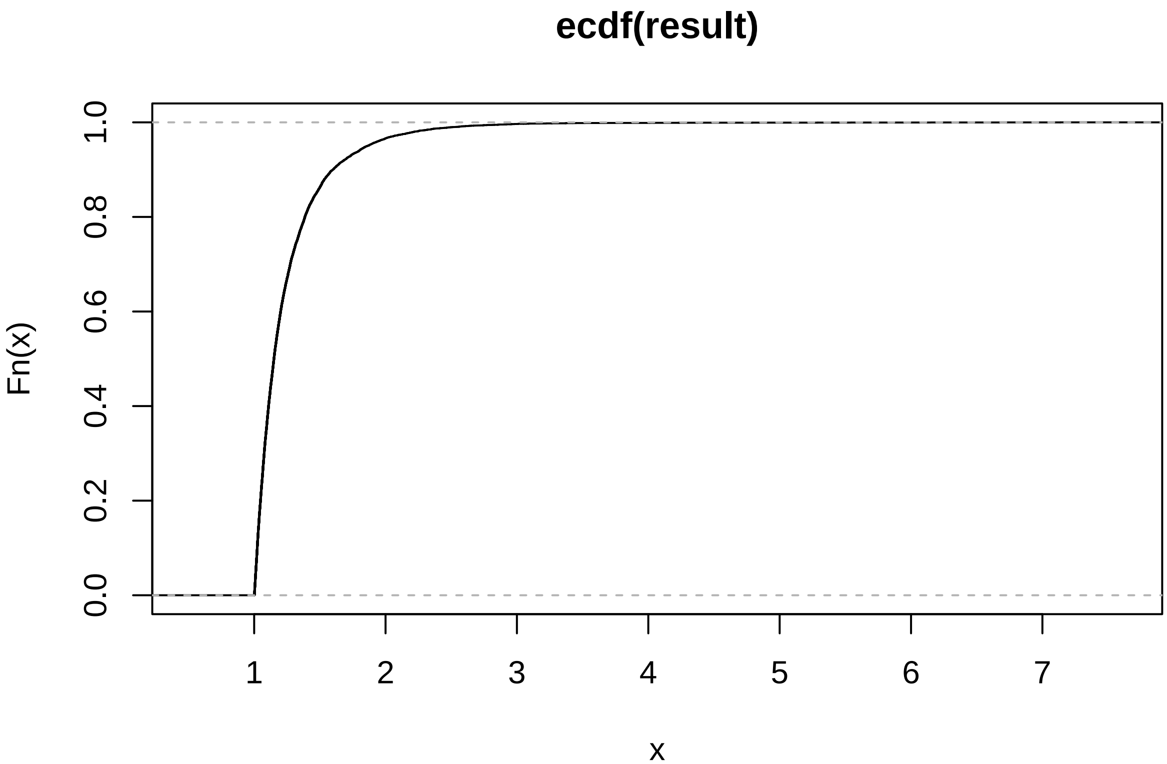
```
Pareto <- function(n){

  u <- runif(n)
  return(exp(log(1) - (log(1-u)/5)))

}
```

(c) Write R code to generate 10,000 random numbers and draw a EDF plot.

```
result <- Pareto(10000)
plot(ecdf(result))
```



ecdf(result)
Fn(x)
0.0
0.2
0.4
0.6
0.8
1.0
1
2
3
4
5
6
7
x