CS 108 - Statistics, Sections B

Fall 2019, AUA

Homework No. 04

Due time/date: Section B: 10:32 AM, 27 September, 2019

Note: Please use **R** only in the case the statement of the problem contains (R) at the beginning. Otherwise, show your calculations on the paper. Supplementary Problems will not be graded, but you are very advised to solve them and to discuss later with TA or Instructor.

Problem 1, Theoretical Quantiles

a.

- Find the 27% quantile of the Exponential Distribution $Exp(3)$.
- Find, for any $\alpha \in (0,1)$, the α-quantile of $Exp(1)$ Distribution.
- Find the Median (50% quantile) of the $Exp(1)$ Distribution.
- (Supplementary) Find the Median of the $Bernoulli(0.3)$ Distribution.
- Find the 10% quantile of the Distribution with the PDF

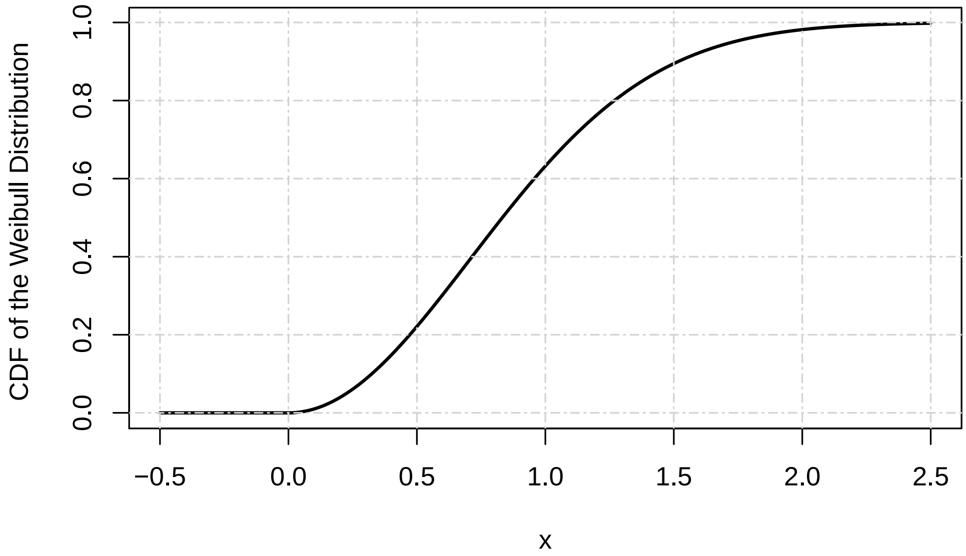
$$f(x) = \begin{cases} 0.5 \cdot \sin(x), & x \in [0, \pi] \\ 0, & otherwise \end{cases}$$

- Which order quantile is 0.7 for the $Unif[-1,1]$ Distribution?

b.

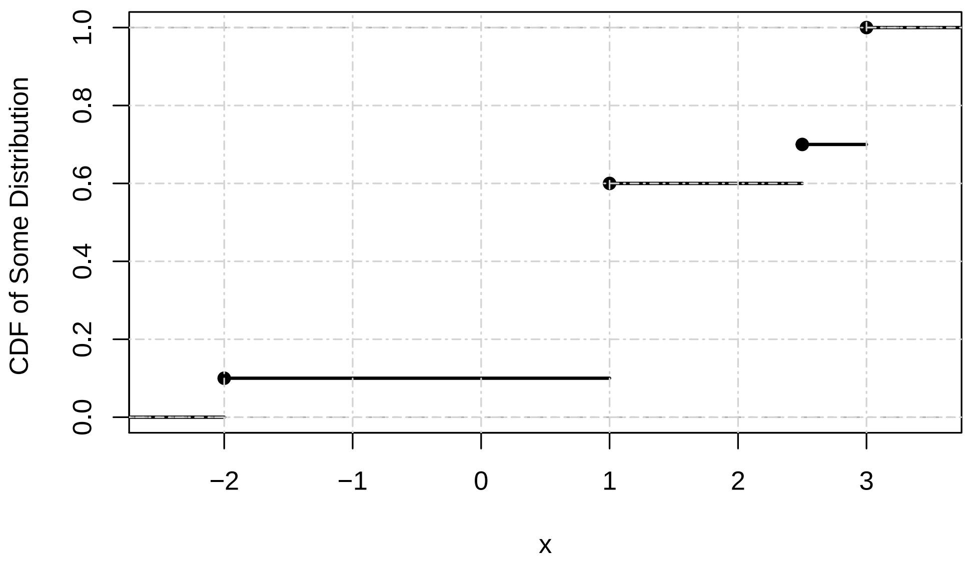
- Below you can find the graph of the CDF of the Weibull Distribution[1] with some parameters.

[1]See Wiki for Weibull Distrib



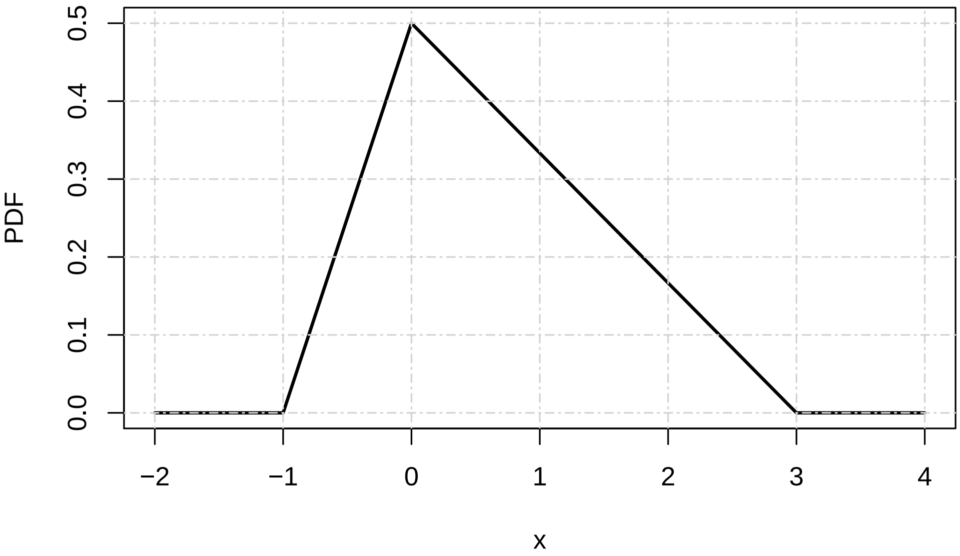
Find the approximate values of the Median and the $20\%, 70\%$ quantiles of that Distribution. Explain your reasoning.

- Below you can find the graph of the CDF of some Distribution.



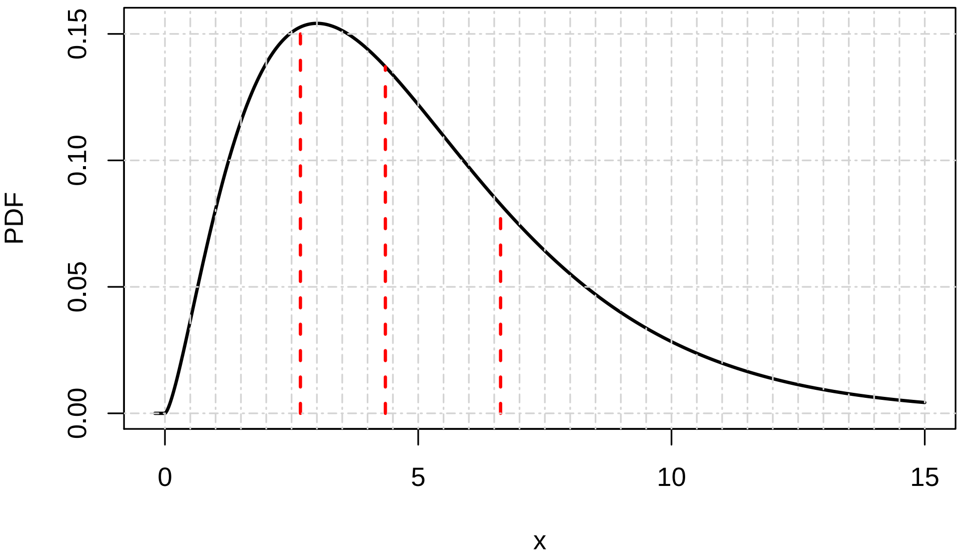
Find the approximate values of the quantiles of order $20\%, 40\%, 60\%, 90\%$.

- Below you can find the graph of the PDF of some Distribution:



Find the exact values of the Median and the $20\%, 70\%$ quantiles of that Distribution. Show your calculations.

- Below you can find the graph of the PDF of some Distribution (x-gridline stepsize is 0.5):



The red dashed lines divide the area under the cure into 4 equal parts. Find the approximate values of the quantiles of order $10\%, 25\%, 50\%$ and 75%. Explain your reasoning.

c. (R)

- Calculate all Deciles (quantiles of order $10\%, 20\%, 30\%, ..., 90\%$) of the Standard Normal Distribution;
- Construct a sequence of quantile levels $\alpha = 0.01, 0.02, ..., 0.99$. Plot the $\alpha-$Quantiles z_α of the Standard Normal Distribution vs α. Then, on the same graph, and using another color, plot the $(1-\alpha)$-level Quantiles of the same Distribution. Explain the symmetry (if you have plotted correclty, of course ☺).
- Find a symmetric interval $[a, b]$ such that for $X \sim \mathcal{N}(0,1)$,

$$\mathbb{P}(X \notin [a, b]) = 0.99$$

d. (Supplementary)

- Prove that if z_α is the α-quantile of the Standard Normal Distribution, then

$$z_\alpha = -z_{1-\alpha}, \qquad \forall \alpha \in (0,1).$$

- Generalize the above for α-quantiles of $\mathcal{N}(\mu, \sigma^2)$;
- Generalise the above for any Distribution symmetric about 0.

Problem 2, Q - Q Plot

a.

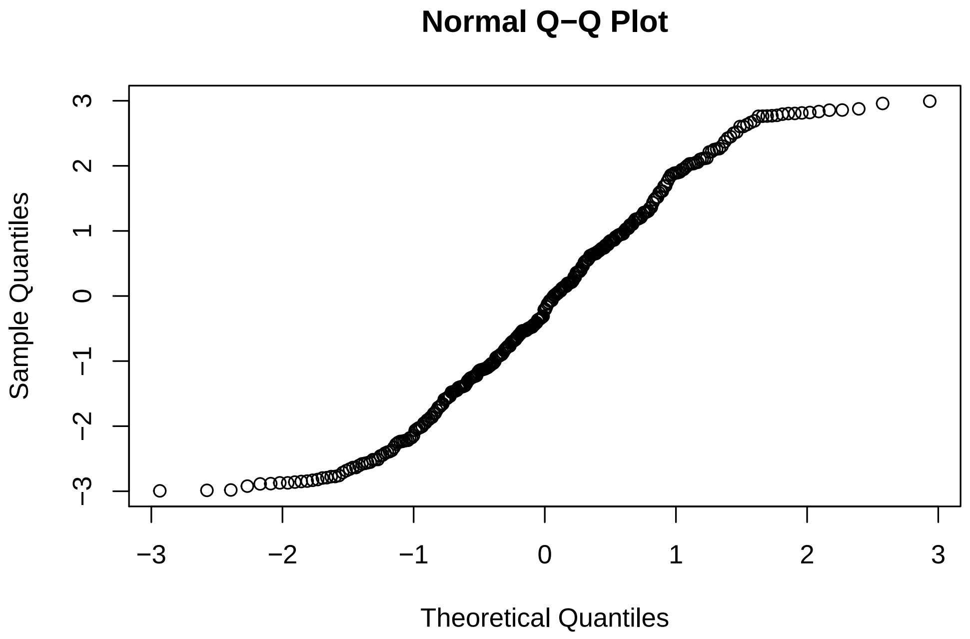
- Using the quantiles of order $\alpha = 0.2, 0.4$ and 0.6, draw the Q-Q Plot for the following Datasets:

$$x: \ -4, 2, 1, 3, 2, -1, 1, 1, 13, \qquad y: \ 5, 5, 2, -1, -1, 3, 5.$$

- Using the quantiles of order $\alpha = 0.2, 0.4$ and 0.6, draw the Q-Q Plot of the above x vs the quantiles of $Unif[0,1]$.
- Express the Quantiles of $Unif[a, b]$ in terms of the Quantiles of the Standard Uniform Distribution, i.e., Quantiles of $Unif[0,1]$.
- If I will generate a sample from $Unif[-3, 12]$ and another one from $Unif[0,1]$, and draw the Q-Q Plot of that two sample, what will be the approximate picture, where the points will lie?

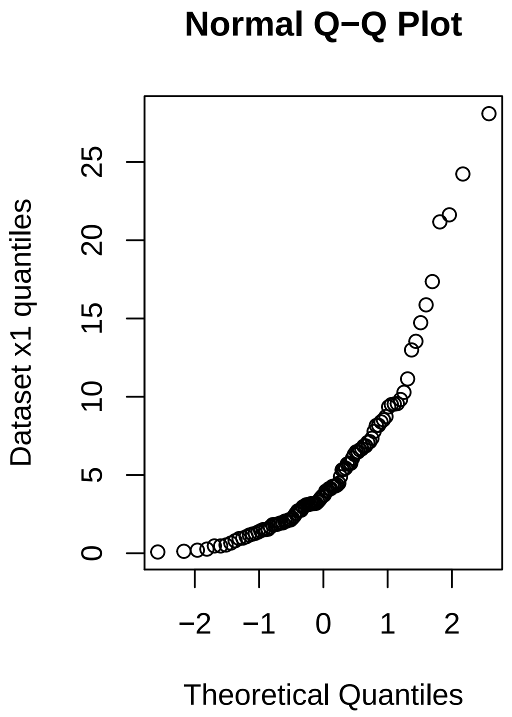
b.

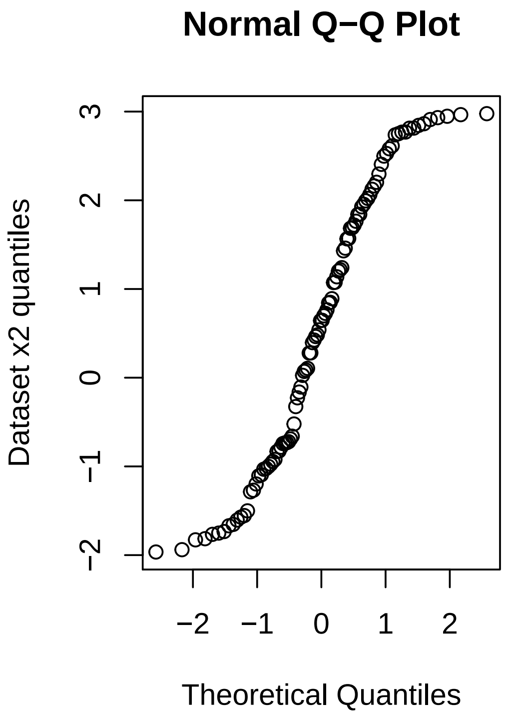
- I have generated a random sample from one of the Distributions $Unif[-3, 3]$ or $Exp(3)$, but forgot from which one. But I have the Q-Q Plot of my sample vs the Standard Normal Distribution:

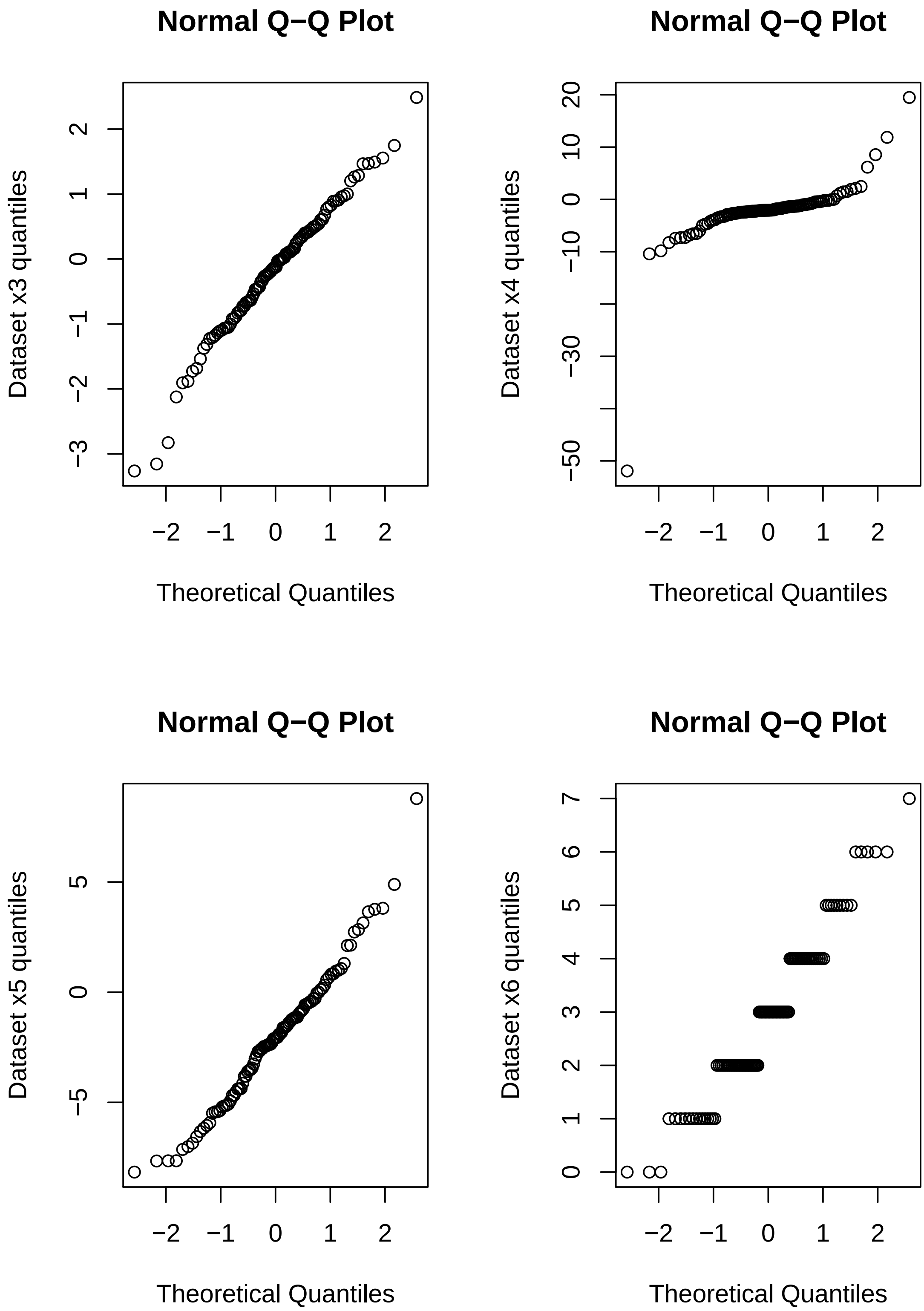


Help me to identify from which Distribution was my sample. Give your reasoning (so that next time I will be able to identify by myself ◡̈).

- Here are the Q-Q Plots of some random samples vs Standard Normal Distribution.







Which of these Dataset is likely to be from the Normal Distribution?

(Supplementary) Why is the last (bottom-right) Q-Q Plot different from the others? Explain!

c. (R)

- Generate a sample of size 200 from $Exp(3)$ and another one of size 400 from $Exp(0.2)$, and draw their Q-Q Plot.
- Write a function `qqexp(x)` and `qqunif(x)` so that they will draw the Q-Q Plot of the Dataset x vs the theoretical Quantiles from $Exp(1)$ and $Unif[0,1]$, respectively.

d. Q-Q Plot of AMZN daily returns (R)

Here we want to see if the weakly returns of the Amazon Stock can be modelled using a Normal Distribution. Daily returns are usually close to zero, sometimes positive (when the price increases), and sometimes negative (if the price decreases).

- Navigate to finance.yahoo.com and search for the Amazon ticker `AMZN`. Navigate to `Historical Data`, change the time period to `1Y` (1 year), choose daily frequency, hit `Apply`, and then `Download Data`. You will have the file of daily prices `AMZN.csv`.
- Read, using the `read.csv(file.choose())` command that `.csv` file. Separate in a new variable the `Adj.Close` (Adjusted Close Prices) variable.
- Calculate daily returns using the Adjusted Close Prices.
- Plot the Histogram of that daily returns.
- Draw the Q-Q Plot of that daily returns vs Standard Normal Distribution, using the `qqnorm` command. Add the `qqline` to the graph.
- Explain and make conclusions - will it be reasonable to model daily returns by using a Normal Distribution?
- (Supplementary) I want to know the possible price for Amazon Stock for tomorrow. Suggest me a method to generate the possible value of the tomorrow's return, and I will calculate tomorrow's possible price.

e. (Supplementary)

- Express the Quantiles of $\mathcal{N}(\mu, \sigma^2)$ in terms of the Quantiles of Standard Normal Distribution.

Problem 3, Covariance and Correlation

a.

- We have the following observations for two variables x and y:

$$x : 0, 1, 2, 1, 0, \qquad y : -3, 2, 1, 2, 1$$

  - Give the Scatterplot for this data;
  - Calculate the sample covariance and correlation coefficient for x, y.

- Prove that for any 2D dataset (x, y),
  1. $cov(\alpha \cdot x, \beta \cdot y) = \alpha\beta \cdot cov(x, y)$;
  2. $var(x + y) = var(x) + 2 \cdot cov(x, y) + var(y)$;

3. if x and y are uncorrelated, then $var(x+y) = var(x) + var(y)$

- Construct a bivariate dataset with uncorrelated variables. The number of datapoints need to be larger than 2. Give the scatterplot of your dataset.

b.

Let $X \sim \mathcal{N}(0,1)$, and $Y = X^2$.

- Calculate $Cov(X,Y)$ (theoretical Covariance). Are X and Y independent? Explain!
- (**R**) Generate a sample x of size 2500 from $\mathcal{N}(0,1)$. Calculate $y = x^2$, and the $cov(x,y)$. Is it close to the theoretical Covariance from the previous part?
- (**R**) Find out which denominator is using **R** when calculating the sample covariance, n or $n-1$. Write your own function to calculate the Sample Covariance using the other denominator.