IN-CLASS QUIZ 1 SOLUTION

Question 1: *Data collection from real system of interest*

Suppose that you are modelling a laser marking workstation. Jobs arrive at random, wait in a buffer until the laser marking machine is available, get marked, then leave the workstation. Suppose the laser power supply (voltages) can drop after a long period of time, causing the laser output power to drop. This event results in a stoppage during which the laser power supply is replaced. Fill in the description of data collection activity below.

a. .... Clock times are recorded on job arrivals to form the sequence of job interarrival times. Job interarrival times
b. .... The total arrival stream is partitioned into substreams of different types, data collection of interarrival times is carried out separately for each type. Type of job arrivals/ Job types
c. .... The time it takes to finish a job. Given the replacement of laser power supply is modeled separately, this measured period of time excludes any downtime. Laser marking times/ Processing time
d. .... The time interval between two successive replacements of laser power supply minus the total idle time (if any) in that interval. Times between power supply replacements/ Time to failure
e. .... The time from stoppage to the time replacement operation is complete. Power supply replacement times/ Downtime/ Repair time

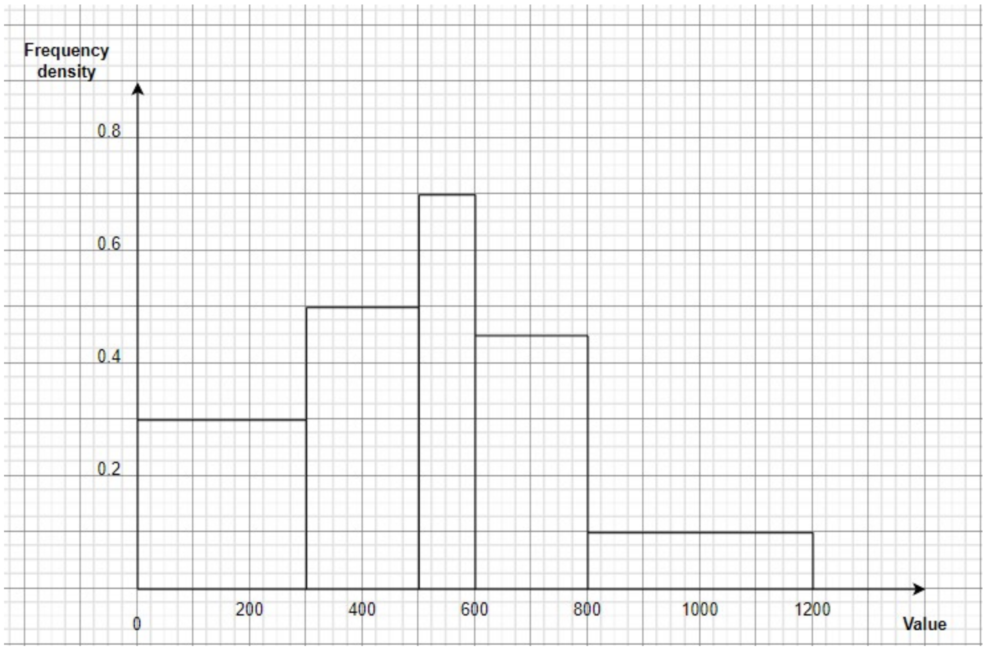
Question 2: *Identification a family of probability distribution to represent the input process*

Select a family of input distribution to represent input in the following scenarios.

a. .... Model the times between arrivals of customers who act independently of one another. Exponential distribution
b. .... Model the number of customer arrivals in a fixed period of time. Poisson distribution
c. .... Model a product testing process for which only the minimum, most likely, and maximum time to test a product are known. Triangular distribution

Question 3: *Histogram (frequency distribution)*

Use the histogram below to complete the frequency table.



Value x	**Frequency**
$0 \leq x \leq 300$	90
$300 \leq x \leq 500$	100
$500 \leq x \leq 600$	70
$600 \leq x \leq 800$	90
$800 \leq x \leq 1200$	40

Given:

$$frequency\ density = \frac{frequency}{class\ width}$$

Question 4: *Parameter estimation from data*

The table below records observations from a sample assumed normally distributed. Determine the sample mean and variance.

Arrivals per period (X_j)	*Frequency (f_j)*
0	*5*
1	*10*
2	*14*
3	*8*
4	*6*
5	*4*
6	*7*
7	*9*
8	*11*

9	10
10	6

Given:

$$\bar{X} = \frac{\sum_{j=1}^{k} f_j X_j}{n}$$

$$S^2 = \frac{\sum_{j=1}^{k} f_j X_j^2 - n\bar{X}^2}{n-1}$$

$\bar{X} = 4.989$

$S^2 = 10.123$ (You may need to refer to my calculation in the Excel file.)

Fill your suggested estimator(s) for each type of distribution in the table below:

Distribution	*Parameter(s)*	*Suggested Estimator(s)*
Normal	μ, σ^2	$\hat{\mu} = \bar{X}$ $\hat{\sigma}^2 = S^2$
Exponential	λ	$\hat{\lambda} = \frac{1}{\bar{X}}$
Poisson	α	$\hat{\alpha} = \bar{X}$

Question 5: *Goodness-of-fit tests: graphical methods and statistical tests*

a. Based on the shape of the histograms in the following figures, predict the type of distribution that each histogram represents.

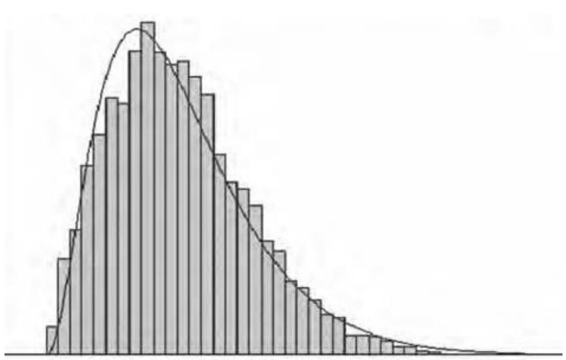


Figure 1

Normal distribution

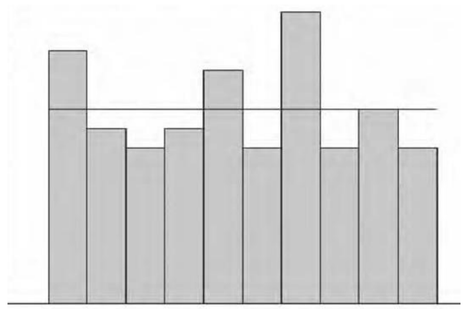


Figure 2

Uniform distribution

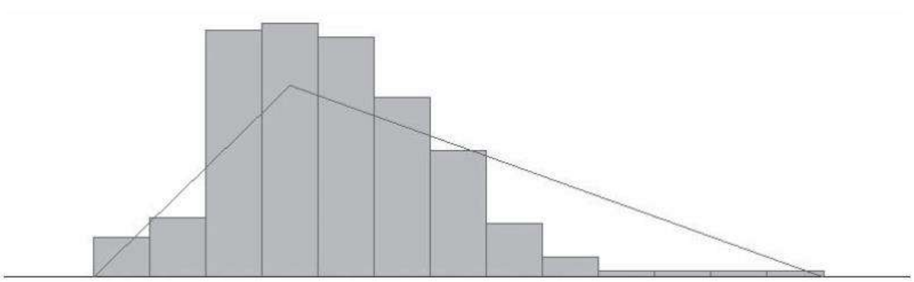
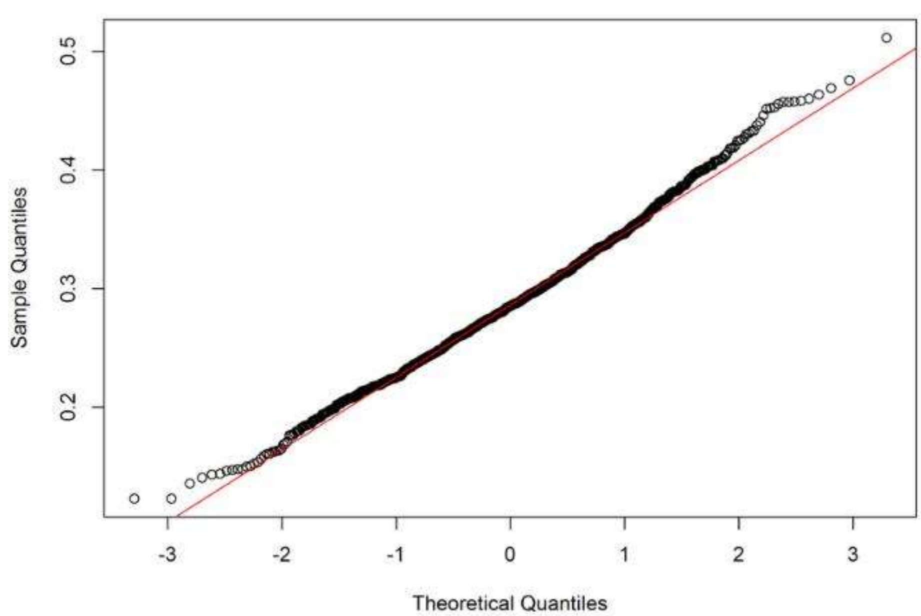


Figure 3

Triangular distribution

b. What do you conclude about the following Q-Q plot? (What distribution you can safely assume it to represent? Explain why?)



FYI: One way to evaluate the fit of a selected distribution is to use histogram. However, this method is not useful when there is insufficient data (≤ 30). A quantile-quantile (Q-Q) plot is an effective tool for assessing distribution fit, one that does not suffer from these problems. In most cases, Q-Q plot is used to assess whether or not normal distribution fits a data set. If the data follows normal distribution, the points in a Q-Q plot will lie on a straight diagonal line.

From the Q-Q plot, it can be safely assumed that the set of studied data follows normal distribution, because points lie along the straight diagonal line with only minor deviations along each of the tails.

c. Fill in the blanks:

<table>
<tr><td></td><td>K-S test</td><td>Chi-square test</td></tr>
<tr><td>Hypothesis</td><td colspan="2">H_0: Random variable X conforms to the distributional assumption with the parameter(s) given by the parameter estimate(s).
H_1: Random variables X does not conform.</td></tr>
<tr><td>Test statistics
(Explain the terms)</td><td>$D = \max(D^+, D^-)$ or
$D = \max\ |\mathrm{F(x)} - \mathrm{S_N}(x)|$
$D^+ = max_{1\leq i\leq M}\left\{\frac{1}{N} - R_i\right\}$
$D^- = max_{1\leq i\leq M}\left\{R_i - \frac{i-1}{N}\right\}$</td><td>$X_0^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$
k: class intervals
O_i: observed frequency in the i^{th} class interval
E_i: expected frequency in that class interval, $E_i = np_i$ where p_i is the theoretical, hypothesized probability associated with the i^{th} class interval.</td></tr>
<tr><td>Critical value with alpha = 0.01</td><td>(Testing with 20 numbers)
$D_{0.01,20} = 0.356$</td><td>(Testing with 50 numbers, 5 intervals)
$X^2_{\alpha,k-s-1} = X^2_{0.01,3} = 11.34$
(where s is the number of parameters, consider Poisson the hypothesized distribution)</td></tr>
</table>

You may need to take note on how to write hypothesis test conclusion!

Question 6: *Calculate covariance and correlation of 2 random variables*

Consider the available data on lead time and sales for the last five years. Estimate the correlation and covariance of the two variables.

Lead time	Sales
3	300
4	460
5	620
6	550
5	550
6	760
3	350
7	900
4	530
5	590

Given:

$$\widehat{cov} = \frac{1}{n-1}\left(\sum_{j-1}^{n} X_{1j}X_{2j} - n\bar{X}_1\bar{X}_2\right)$$

$$corr(X_1, X_2) = \frac{cov(X_1, X_2)}{\sigma_1\sigma_2}$$

$$\widehat{cov} = 215.778$$

$$corr(X_1, X_2) = 0.924$$

IN-CLASS QUIZ 2 SOLUTION

Question 1: *Data collection from real system of interest*

Suppose that you are modelling a filling process in a packaging line. The process experiences random operation-independent delays due to machine malfunctions and adjustment. It is observed that the times between a repair completion and the next failure are exponentially distributed with a mean of 60 hours. Whereas its repair times are observed to be uniformly distributed between 2 hours to 3 hours. The process needs an adjustment after every 280 departures from the workstation. In addition, the adjustment times are triangularly distributed with a minimum of 45 minutes, a maximum of 90 minutes, and a most likely period of 60 minutes.

Based on the description above, fill in the summary of data collection below.

Name	Type	Up Time	Up Time Units	Count	Down Time	Down Time Units
Random Failures	Time	EXPO(60)	Hours		UNIF(120,180)	Minutes
Adjustment	Count			280	TRIA(0.75,1,1.5)	Hours

Question 2: *Identification a family of probability distribution to represent the input process*

Select a family of input distribution to represent input in the following scenarios.

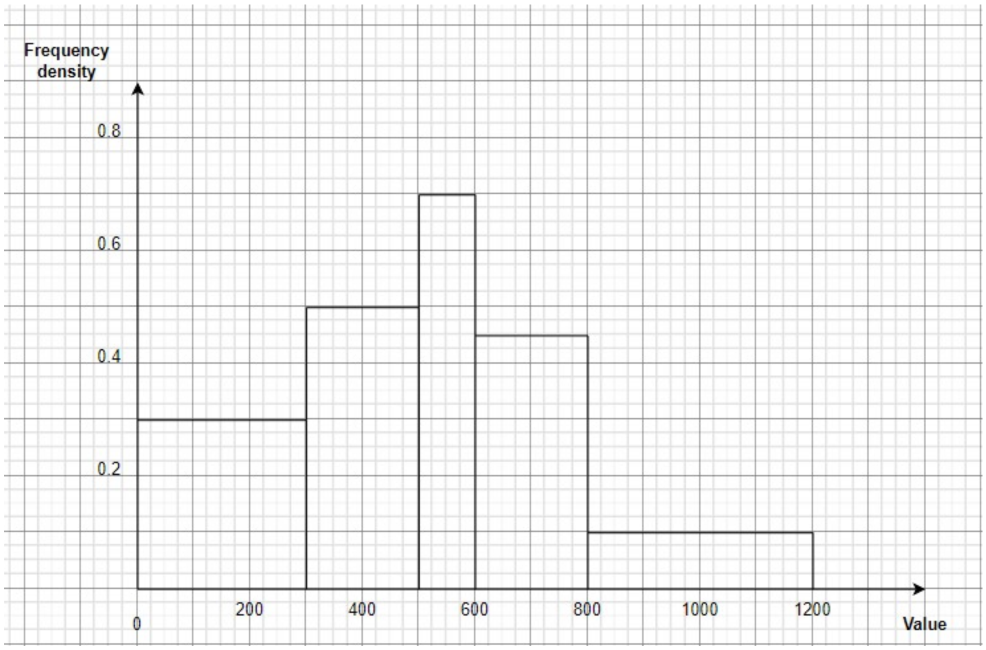
d. .... Model the number of defective light bulbs found in a sample of 100 randomly selected ones. Binomial

e. .... Model the number of customer phone calls an office receives during the time from 11.00am to 12.00am. Poisson

f. .... Model a process that takes a minimum of 1.5 hours, a maximum of 3 hours, and a mode of 2 hours, to test a computer chip. Triangular

Question 3: *Histogram (frequency distribution)*

The histogram below shows the monthly renting-cost distribution (in $) of a group of people in Ho Chi Minh City. Complete the table and answer the questions.



Value x	**Frequency**
$0 \leq x \leq 300$	90
$300 \leq x \leq 500$	100
$500 \leq x \leq 600$	70
$600 \leq x \leq 800$	90
$800 \leq x \leq 1200$	40

a. How many people are paying up to $500 monthly? 190

b. How many people are paying equal or more than $500 monthly? 200

c. What is the percentage of people paying between $500 and $800 monthly?

$$\frac{160}{390} = 41\%$$

Given:

$$frequency\ density = \frac{frequency}{class\ width}$$

Question 4: *Parameter estimation from data*

The table below records observations from a sample assumed normally distributed. Determine the sample mean and variance.

j	X_j	j	X_j
1	50	11	55
2	41	12	50
3	32	13	54
4	33	14	38

5	42	15	46
6	55	16	24
7	46	17	47
8	37	18	49
9	58	19	51
`10	59	`20	60

Given:

$$\bar{X} = \frac{\sum_{j=1}^{k} X_j}{n}$$

$$S^2 = \frac{\sum_{j=1}^{k} X_j^2 - n\bar{X}^2}{n-1}$$

$$\bar{X} = 46.35$$

$$S^2 = 96.56$$

Assuming that the data in the above table is exponential distributed, what is your suggested estimator of its parameter?

$$\hat{\lambda} = \frac{1}{\bar{X}} = 0.0216$$

Question 5: *Goodness-of-fit tests: graphical methods and statistical tests*

Use Chi-square test for testing the hypothesis that the random sample of random variable X, given below, follows **Poisson** distribution.

Arrivals per period (X_j)	*Frequency (f_j)*
0	16
1	15
2	11
3	13
4	13
5	8
6	9
7	6
8	5
9	4

Given:

$$\bar{X} = \frac{\sum_{j=1}^{k} f_j X_j}{n}$$

And given the pmf for the Poisson distribution:

$$p(x) = \begin{cases} \frac{e^{-\alpha}\alpha^{x}}{x!}, & x = 0,1,\ldots \\ 0, & otherwise \end{cases}$$

You may use the guide of solution below.

A guide of solution:

Parameter estimation:

$$\hat{\alpha} = \bar{X} = \frac{\sum_{j=1}^{k} f_j X_j}{n} = 5.282$$

x_i	O_i	p_i	$E_i = np_i$	$\frac{(O_i - E_i)^2}{E_i}$
0	*16*	*0.033*	*14.7*	*18.129*
1	*15*	*0.113*		
2	*11*	*0.193*	*19.3*	*3.563*
3	*13*	*0.219*	*21.9*	*3.592*
4	*13*	*0.186*	*18.6*	*1.677*
5	*8*	*0.126*	*12.6*	*1.701*
6	*9*	*0.072*	*7.2*	*0.473*
7	*6*	*0.035*	*5.5*	*16.316*
8	*5*	*0.015*		
9	*4*	*0.006*		

Chi-square test	
Hypothesis	$H_0: X \sim Poisson(\hat{\alpha})$ $H_1: X\ ! \sim Poisson(\hat{\alpha})$
Test statistics	$X_0^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} = 45.450$
Critical value with confidence interval of 95%	$X_{0.05,5}^2 = 11.1$

Conclusion	$X_0^2 > X_{0.05,5}^2$ Reject the null hypothesis.

Question 6: *Covariance and correlation of 2 random variables*

Fill in the blanks to test your understanding of covariance and correlation coefficient.

a. The covariance of two random variables indicates how they .... together.
   vary

b. Let $\rho(X, Y)$ be the correlation coefficient of two random variables X and Y, then:
   - .... $\leq \rho(X, Y) \leq$ ....
     $-1 \leq \rho(X, Y) \leq 1$
   - If X and Y are independent, they are uncorrelated and $\rho(X, Y) =$ ....
     $\rho(X, Y) = 0$
   - Let $Y = aX + b$, then:
     - If $a > 0$, $\rho(X, Y) =$ ....
       $\rho(X, Y) = 1$
     - If $a < 0$, $\rho(X, Y) =$ ....
       $\rho(X, Y) = -1$

c. What is your comment on the correlation coefficient of two random variables illustrated in each of the following figures?

From left to right:
- Figure 1: There is no correlation between two random variables. $\rho(X, Y) = 0$
- Figure 2: Two random variables are strictly negatively autocorrelated. $\rho(X, Y) = -1$
- Figure 3: Random variables are positively autocorrelated. $0 < \rho(X, Y) < 1$
- Figure 4: Random variables are negatively autocorrelated. $0 > \rho(X, Y) > -1$

d. Express the relationship between covariance and correlation coefficient in mathematical language.

$$\rho(X, Y) = \frac{cov(X_1, X_2)}{\sigma_1, \sigma_2}$$