Assignment 02

CE605A Probability & Statistics for Civil Engineers 2023-24 I

Shumet Getahun Roll No. 22203262

1. ***Why should the probability associated with an event must always be a positive real number? (5 points)***

For a given certain event "A" the probability of an event is always lies between 0 and 1 ($0 \leq P1 \leq 1$). Probability is always positive number due to the following reasons:

a. *Likelihood interpretation* probability is a method to quantify and forecast random events. Prediction is usually done by positive probability values. Probability of 0 indicates impossible event and 1 represents absolutely happening event.

b. *Comparison* to make expressive comparison between two or more events. The events having higher likelihood have higher probability.

c. *Non negativity* probability represents likelihood of an event occurrence. Negative probability is senseless because negative number implies a number or quantity is less, not likelihood extent and it contradicts with likelihood idea.

d. *Simplicity* positive numbers simplifies mathematical calculation than negative numbers and not create confusion during data analysis and interpretation.

2. ***A sample of 77 individuals working at a particular office was selected and the noise level (dBA) experienced by each individual was determined, yielding the following data ("Acceptable Noise Levels for Construction Site Offices," Building Serv. Engr. Research and Technology, 2009: 87–94).***

55.3 55.3 55.3 55.9 55.9 55.9 55.9 56.1 56.1 56.1 56.1 56.1 56.1 56.8 56.8 57.0 57.0 57.0 57.8 57.8 57.8 57.9 57.9 57.9 58.8 58.8 58.8 59.8 59.8 59.8 62.2 62.2 63.8 63.8 63.8 63.9 63.9 63.9 64.7 64.7

64.7 65.1 65.1 65.1 65.3 65.3 65.3 65.3 67.4 67.4 67.4 67.4 68.7 68.7 68.7 68.7 69.0 70.4 70.4 71.2

71.2 71.2 73.0 73.0 73.1 73.1 74.6 74.6 74.6 74.6 79.3 79.3 79.3 79.3 83.0 83.0 83.0

(a) Can you summarize the data with the various methods that you have learned? (5 points)

Data can be summarized by using histogram, boxplot and violin plots.

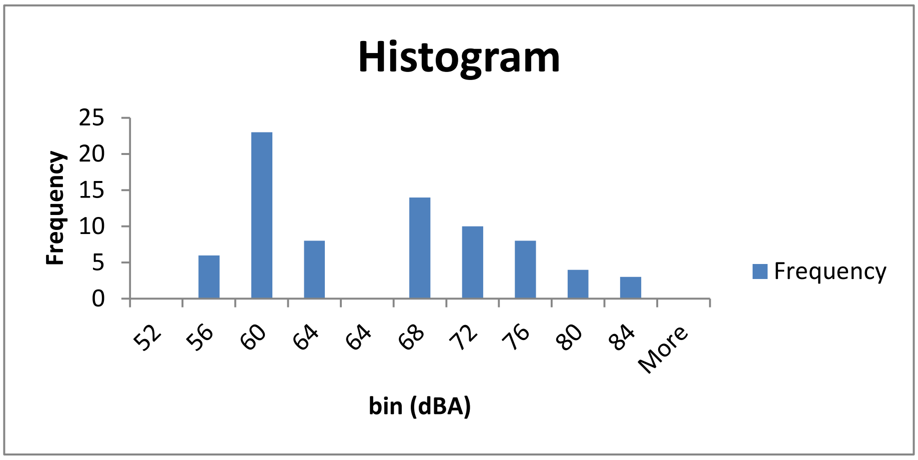

Figure 1: Histogram

Histogram has 8 classes and 4 intervals and it is positively skewed distribution. It starts at 52 and ends at 84. Histogram shows, it is bimodal distribution, the highest frequency is found in 60 and 68.

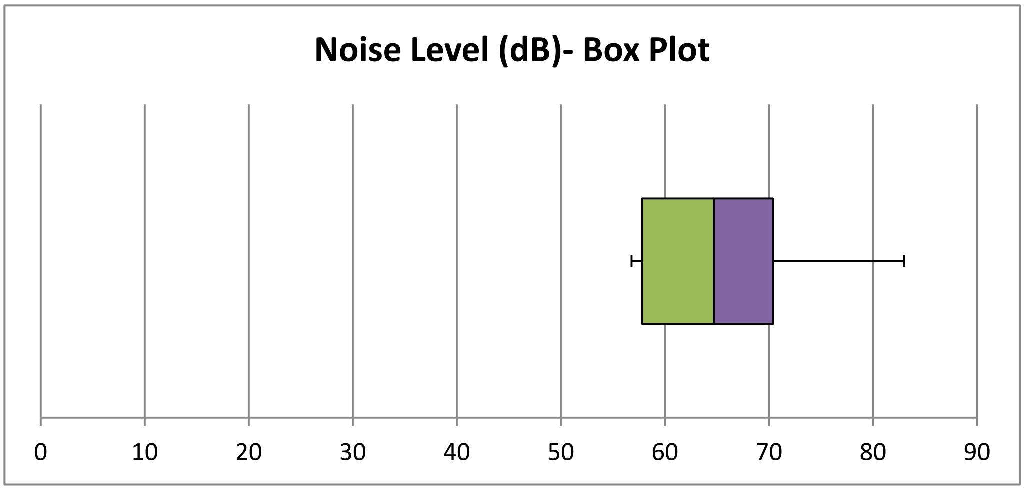

Figure 2: Box Plot

The box plot shows the location of median between 60 and 70. It shows the first and third quartile of the distribution. It also shows the first quartile is below 60 and third quartile is greater than 70. There is no outlier. Upper whisker is wider than lower whisker.

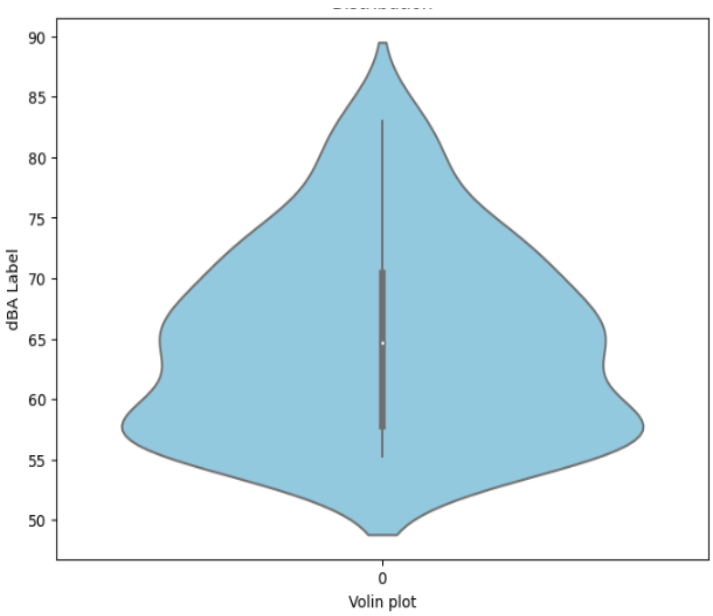

Figure 3: Violin Plot

Fig 3 clearly shows the density of distribution is of data represented by bump shape of violin plot. It also shows the median represented by circle inside box plot. The two edges of rectangle at center represent first and third quartile. Violin plot also presents there are no outliers, since the whisker is inside the plot.

From a given data the following variables are calculated

Median	64.7	
Mode	56.1	
Range	27.7	
Average	64.88	
Largest	83	
Smallest	55.3	
Q1 (lower fourth)	$39 - \frac{(39-1)^{th}}{2} = 20^{th}\ element$	57.8
Q2 (Median)	$\frac{(n+1)^{th}}{2} = \frac{(77+1)^{th}}{2} = 39^{th}\ element$	64.7
Q3 (upper fourth)	$39 + \frac{(77-39)^{th}}{2} = 58^{th}\ element$	70.4

IQR (Inter quartile range) = Q3-Q1 =70.4-57.8=12.6
1.5*IQR = 1.5*12.6=18.9
Lower outlier less than = Q1-1.5*IQR = 57-18.9 = **38.9**
Upper outlier greater than = Q3+1.5*IQR =70.4+18.9 = **89.3**
Therefore there is no outlier since the given range of number is 55.3 to 83.

b. Compare and contrast the variability of the data as obtained from variance, mean absolute deviation and median absolute deviation. Compute the quantities and describe why (or why not) they are different

Calculated data from a given distribution

$Mean\ = \frac{\sum_1^{77} xi}{n}$ = **64.89** Variance = $\frac{\sum_1^{77}(Xi-\bar{X})^2}{n-1} = \frac{4627}{76}$ =60.88

Standard deviation from variance=$\sqrt{variance}$ = $\sqrt{60.88} = \mathbf{7.083}$

Mean absolute deviation = $\frac{\sum_1^{77}|Xi-\bar{X}|}{n} = \frac{490.33}{77}$ =**6.368**

Median absolute deviation = $\frac{\sum_1^{77}|Xi-\tilde{x}|}{n} = \frac{489.4}{77}$ =**6.3588**

Median absolute deviation is lower than Mean absolute and Variance standard deviation because it depends on median and not sensitive to outliers. Median depends on the sequencing order of numbers not outliers or larger numbers. Despite there is no outliers in given case, the variance deviation is higher than median Absolute deviation and Mean absolute deviation due to its higher sensitivity to outliers. The extreme outlier has higher effect on the variance standard deviation to be much larger than the mean absolute deviation.

3. ***The conditional probability is given as***

$$P(A/B) = \frac{P(A \cap B)}{P(B)}$$

(a) ***What does the denominator and numerator mean in the right hand side of the equation? (3 points)***

P(A∩B) is probability of A intersection B, which is the represents the probability of event A and B occur at the same time.

P(B) is prior probability (conditioning probability), probability of occurrence of event B.

(b) ***Can you extend the above equation for the P (A/ B, C, D)? Give reasons and derivation for your answer.*** **(3 points)**

In this case there are three conditional probabilities (B,C,D) that determines the occurrence of event A, instead of only one event in case (a). Therefore by substituting these three events in place of B event, we can drive equation for a given probability P(A/B,C,D).

In this case we can substitute event B, C, D instead of B to drive equation for a given case. P (B, C, D) is the joint probability of events B, C, and D occurring together. This can be expressed in terms of intersection of events B, C and D. P (B, C, D)=P(B∩C∩D), represents the intersection of event B,C and D simultaneously.

$$P(A/B) = \frac{P(A \cap B)}{P(B)}$$

(devore, 2016)

By substituting event B, C and D in place of event B, the following equation is driven.

$$P(A/B,C,D) = \frac{P(A \cap (B, C, D))}{P(B,C,D)} = \frac{P(A \cap (B \cap C \cap D))}{P(B,C,D)}$$

P (A∩ (B∩C∩D)) = P (A∩B∩C∩D) using associative property of intersection of events (simultaneous occurrence of events). P (A/ B, C, D) is the conditional probability of event A given that events B, C, and D have occurred. Using all above parameters the following equation is derived.

$$P(A\ /B,C,D) = \frac{P\big(A \cap (B \cap C \cap D)\big)}{P(B,C,D)} = \frac{P(A \cap B \cap C \cap D)}{P(B \cap C \cap\ D)}$$

$$\boldsymbol{P(A\ /B,C,D) = \frac{P(A \cap B \cap C \cap D)}{P(B \cap C \cap D)}}$$

c. Blue Cab operates 15% of the taxis in a certain city, and the other 85%. After a nighttime hit- and-run accident involving a taxi, an eyewitness said the vehicle was blue. Suppose, though, that under night vision conditions, only 80% of individuals can correctly distinguish between a blue and a green vehicle. What is the (posterior) probability that the taxi at fault was blue? In answering, be sure to indicate which probability rules you.

Answer

P (B) = probability of Blue Cab operates = 0.15

P (I/B) = probability of eyewitness distinguish a blue vehicle correctly = 0.8

P (I'/B) = probability of eyewitness distinguish a blue vehicle not correctly =1- P (I/B) = 1-0.8= 0.2

P (G) = probability of Green Cab operates = 0.85

P (I/G) = probability of eyewitness distinguish a green vehicle correctly =1- P (I'/B) = 1-0.2= 0.8

P (I'/G) = probability of eyewitness distinguish a green vehicle not correctly =1- P (I/G) = 1-0.8= 0.2

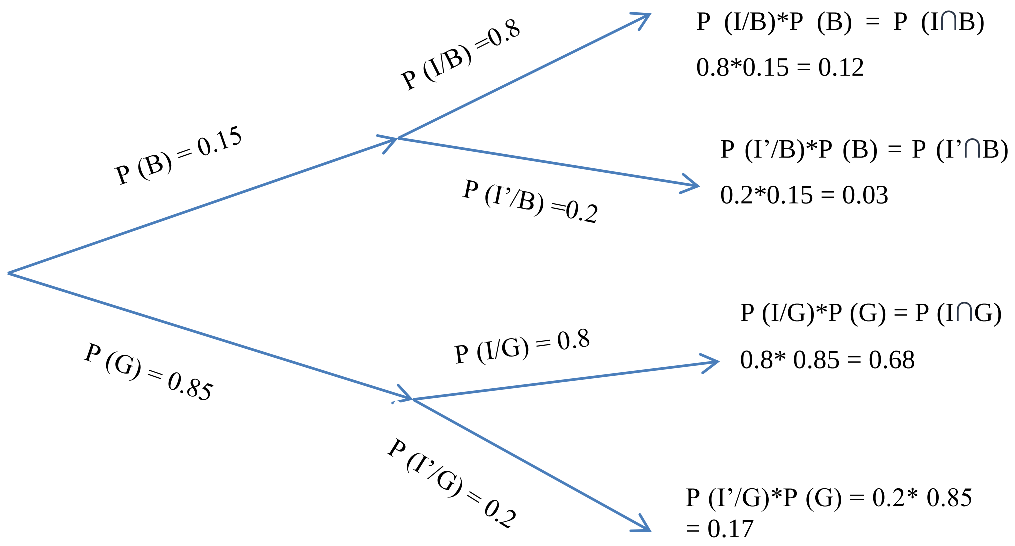

Figure 4 Tree Diagram of Cab Operation

Probability that the taxi at fault was blue can be solved using Multiplication rule for intersection and total probability rule is applied for denominator. It is based on Bayes theorem.

$$P(B/I) = \frac{P(B\cap I)}{P(I)} = \frac{P(I/B)*P(B)}{P(I)} = \frac{P(I/B)*P(B)}{P(I/B)*P(B)+P(I/G)*P(G)} = \frac{0.8*0.15}{0.8*0.15+0.2*0.85} = \frac{0.12}{0.12+0.68} = 0.15$$

Probability of the blue taxi fault was witnessed correctly is **0.15**

$$P(B/I') = \frac{P(B\cap I')}{P(I')} = \frac{P(I'/B)*P(B)}{P(I')} = \frac{P(I'/B)*P(B)}{P(I'/B)*P(B)+P(I'/G)*P(G)} = \frac{0.2*0.15}{0.2*0.15+0.2*0.85} = \frac{0.03}{0.03+0.17} = 0.15$$

Probability of the blue taxi fault was witnessed not correctly is **0.15**

4. ***The number of pumps in use at both a six-pump station and a four-pump station will be determined. Give the possible values for the number of stations having exactly two pumps in use***

Answer

The possibility of pumps in use can be determined based on events in both station 1 and station 2. Different possible values are arranges as flows in a table.

		station 2 (4 pumps)				
		0	**1**	**2**	**3**	**4**
station 1 (6 pumps)	**0**	[0 0]	[0 1]	[0 2]	[0 3]	[0 4]
	1	[1 0]	[1 1]	[1 2]	[1 3]	[1 4]
	2	[2 0]	[2 1]	[2 2]	[2 3]	[2 4]
	3	[3 0]	[3 1]	[3 2]	[3 3]	[3 4]
	4	[4 0]	[4 1]	[4 2]	[4 3]	[4 4]
	5	[5 0]	[5 1]	[5 2]	[5 3]	[5 4]
	6	[6 0]	[6 1]	[6 2]	[6 3]	[6 4]

The possibility that two pumps in use are three combinations, [0,2], [1, 1] and [2 0] expressed by 0,1, 2

5. ***An insurance company offers its policyholders a number of different premium payment options. For a randomly selected policyholder, let X be the number of months between successive payments.***

The cdf of X is as follows:

$$F(x) = \begin{cases} 0 & x < 1 \\ 0.3 & 1 \leq x < 3 \\ 0.4 & 3 \leq x < 4 \\ 0.45 & 4 \leq x < 6 \\ 0.6 & 6 \leq x < 12 \\ 1 & 12 \leq x \end{cases}$$

(c) ***What is the pmf of X?*** **(5 points)**

$P\ (a \leq X \leq b) = F\ (b) - F(a-)$

P (1) = P(x=1) =F (1) – F (0) = 0.3-0 = **0.3**

P (3) = P(x=3) =F (3) – F (2) = 0.4 -0.3= **0.1**

P (4) = P(x=4) =F (4) – F (3) = 0.45-0.4= **0.05**

P (6) = P(x=6) =F (6) – F (5) = 0.6-0.45= **0.15**

P (12≤ x) = F (12) – F (11) = 1-0.6= **0.4**

(d) ***Using just the cdf, compute P (3 ≤ X ≤ 6) and P (4 ≤ X).*** **(3 points)**

P (3≤x≤6) = F (6) – F (2) = (0.6-0.3) = **0.30**

P(4 ≤ X) = F (12) – F (3) = 1-0.4=**0.6**

6. ***Twenty percent of all telephones of a certain type are submitted for service while under warranty. Of these, 60 % can be repaired, whereas the other 40 % must be replaced with new units. If a company purchases ten of these telephones, what is the probability that exactly two will end up being replaced under warranty? (5 points)***

Answer

Total number of telephones to be purchased, N=10

P (W) = Probability of telephones under warranty from total number of being purchased =20%

P (M) =Probability of being repaired = 60%

P(R) = Probability of being replaced = 40%

Let P (N) is the probability of a phone is submitted for service while under warranty and is replaced with a new telephone.

Probability of replacing a product which is under warranty is the intersection of P (W) and P(R)

$$P(N) = P(W) \cap P(R) = P(W)*P(R) = 0.2*0.4 = 0.08$$

Probability of two telephones end up replaced, p(x) is calculated using binominal equation. X is 2 and p=0.08.

$$p(x) = \binom{n}{x} * p^x * (1-p)^x = \binom{10}{2} * 0.08^2 * (1-0.08)^8 = \mathbf{0.1478}$$

Bonus

The probability theory is encapsulated as a quantitative system by Kolmogorov's axioms. However, there are other ways of encapsulating probability. Can you identify those systems and compare them with Kolmogorov's axioms? (maximum 500 words) (**10 points**)

Answer

Despite Kolmogorov's axioms theory is popular, there are other alternative theories encapsulating probability theory. One of the theories, Cox's theorem is a mathematical result that establishes the correspondence between reasonable reasoning and probability under specific assumptions. Cox's theorem allowing countable addition and the ability to make uncountable many probability claims, which modifies the Kolmogorov's theory. Cox's theorem is more flexible than komogrovs theory in terms of analyzing more uncertain events since it does not necessarily depend on specific events that reduce subjectivity. (Alexander Terenin, 2015)

A fuzzy probability theorem is a mathematical model that extends the conventional evenly distributed probability space to a complex distribution of fuzzy probability. It is based on fuzzy sets and is used to deal with random fuzzy events, uncertainties in various fields such as formal inference, cognitive linguistics and computational intelligence. It is applied on artificial intelligence such as fuzzy expert systems, fuzzy multi criteria decision analysis. (Čunderlíková, 2020)

Reference

Alexander Terenin, D. D. (2015). Rigorizing and Extending the Cox-Jaynes Derivation of Probability: Implications for Statistical Practice.

Čunderlíková, K. (2020). Martingale Convergence Theorem for the Conditional Intuitionistic Fuzzy Probability.

devore, J. L. (2016). *Probability and statistics for Engineering and Science*.