Probability and Statistics (BTech CSE)

Anmol Nawani

May 14, 2023

Contents

1 Ungrouped Data

Ungrouped data is data that has not been arranged in any way.So it is just a list of observations

$$x_1, x_2, x_3, ...x_n$$

1.1 Mean

$$\bar{x} = \frac{x_1 + x_2 + x_3 + ... + x_n}{n}$$

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

1.2 Mode

The observation which occurs the highest number of time. So the x_i which has the highest count in the observation list.

1.3 Median

The median is the middle most observations. After ordering the n observations in observation list in either Ascending or Descending order (any order works). The median will be :

- n is even

$$Median = \frac{x_{\frac{n}{2}} + x_{(\frac{n}{2}+1)}}{2}$$

- n is odd

$$Median = x_{\frac{n+1}{2}}$$

1.4 Variance and Standard Deviation

$$Variance = \sigma^2$$

$$Standard\ deviation = \sigma$$

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - Mean)^2}{n}$$

$$\sigma^2 = \frac{\sum_{i=1}^{n} x_i^2}{n} - (Mean)^2$$

1.5 Moments

1.5.1 About some constant A

$$r^{th}\ moment = \frac{1}{n}\Sigma(x_i - A)^r$$

1.5.2 About Mean (Central Moment)

When A = Mean, then the moment is called central moment.

$$\mu_r = \frac{1}{n}\Sigma(x_i - Mean)^r$$

1.5.3 About Zero (Raw Moment)

When A = 0, then the moment is called raw moment.

$$\mu_r^{'} = \frac{1}{n}\Sigma x_i^r$$

2 Grouped Data

Data which is grouped based on the frequency at which it occurs. So if 9 appears 5 times in our observations, we group as x(observation) = 9 and f (frequency) = 5.

x (observations)	f (frequency)
2	5
1	3
4	5
8	9

If we store it in data way, i.e. the observations are of form 10-20, 20-30, 30-40 ... then we will get x_i by doing

$$x_i = \frac{lower\ limit + upper\ limit}{2}$$

i.e,
x_i for 20-30 will be $\frac{20+30}{2}$
So for data

	f (frequency)
0- 20	2
20-40	6
40-60	1
60-80	3

the x_i's will become.

	f_i	x_i
0- 20	2	10
20-40	6	30
40-60	1	50
60-80	3	70

2.1 Mean

$$\bar{x} = \frac{\Sigma f_i x_i}{\Sigma f_i}$$

2.2 Mode

The **modal class** is the record with the row with the highest f_i

$$Mode = l + (\frac{f_1 - f_0}{2f_1 - f_0 - f_2}) \times h$$

In the formula :
l $\rightarrow$ lower limit of modal class
f_1 $\rightarrow$ frequency(f_i) of the modal class
f_0 $\rightarrow$ frequency of the row preceding modal class
f_2 $\rightarrow$ frequency of the row after the modal class
h $\rightarrow$ size of class interval (upper limit - lower limit)

2.3 Median

The median for grouped data is calculated with the help of **cumulative frequency**. The cumulative frequency (cf_i) is given by:

$$cf_i = f_1 + f_2 + f_3 + ... + f_i$$

The **median class** is the class whose cf_i is just greater than or is equal to $\frac{\Sigma f}{2}$

$$Median = l + (\frac{(n/2) - cf}{f}) \times h$$

In the formula :
l → lower limit of the median class
h → size of class interval (upper limit - lower limit)
n → number of observations
cf → cumulative frequency of the median class
f → frequency of the median class

2.4 Variance and Standard Deviation

$$Variance = \sigma^2$$

$$Standard\ deviation = \sigma$$

$$\sigma^2 = \frac{\sum_{i=1}^{n} f_i (x_i - Mean)^2}{\Sigma f_i}$$

$$\sigma^2 = \frac{\sum_{i=1}^{n} f_i x_i^2}{\Sigma f_i} - (Mean)^2$$

2.5 Moments

2.5.1 About some constant A

$$r^{th}\ moment = \frac{1}{\Sigma f_i}[\Sigma f_i (x_i - A)^r]$$

2.5.2 About Mean (Central Moment)

When A = Mean, then the moment is called central moment.

$$\mu_r = \frac{1}{\Sigma f_i}[\Sigma f_i (x_i - Mean)^r]$$

2.5.3 About Zero (Raw Moment)

When A = 0, then the moment is called raw moment.

$$\mu_r^{'} = \frac{1}{\Sigma f_i}[\Sigma f_i x_i^r]$$

3 Relation between Mean, Median and Mode

$$3Median = 2Mean + Mode$$

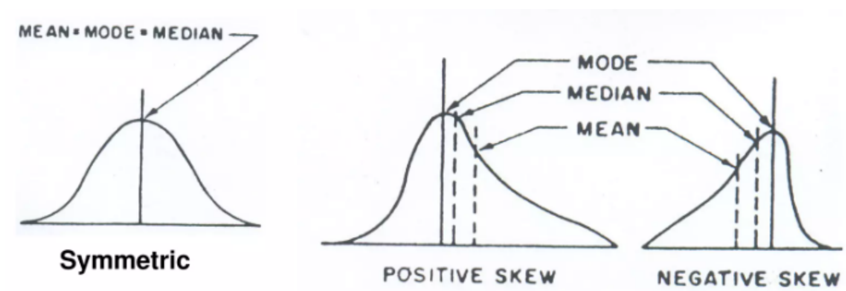
4 Relation between raw and central moments

$$\mu_0 = \mu_0^{'} = 1$$
$$\mu_1 = 0$$
$$\mu_2 = \mu_2^{'} - \mu_1^{'2}$$
$$\mu_3 = \mu_3^{'} - 3\mu_1^{'}\mu_2^{'} + 2\mu_1^{'3}$$
$$\mu_4 = \mu_4^{'} - 4\mu_3^{'}\mu_1^{'} + 6\mu_2^{'}\mu_1^{'2} - 3\mu_1^{'4}$$

5 Skewness and Kurtosis

5.1 Skewness

- If Mean > Mode, then skewness is positive
- If Mean = Mode, then skewness is zero (graph is symmetric)
- If Mean < Mode, then skewness is zero



5.1.1 Pearson's coefficient of skewness

The pearson's coefficient of skewness is denoted by S_{KP}

$$S_{KP} = \frac{Mean - Mode}{Standard\ Deviation}$$

- If S_{KP} is zero then distribution is symmetrical
- If S_{KP} is positive then distribution is positively skewed
- If S_{KP} is negative then distribution is negatively skewed

5.1.2 Moment based coefficient of skewness

The moment based coefficient of skewness is denoted by β_1. The μ here is central moment.

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3}$$

The drawback of using β_1 as a coefficient of skewness is that it **can only tell if distribution is symmetrical or not** ,when $\beta_1 = 0$. It can't tell us the direction of skewness, i.e positive or negative.

- If β_1 is zero, then distribution is symmetrical

5.1.3 Karl Pearson's γ_1

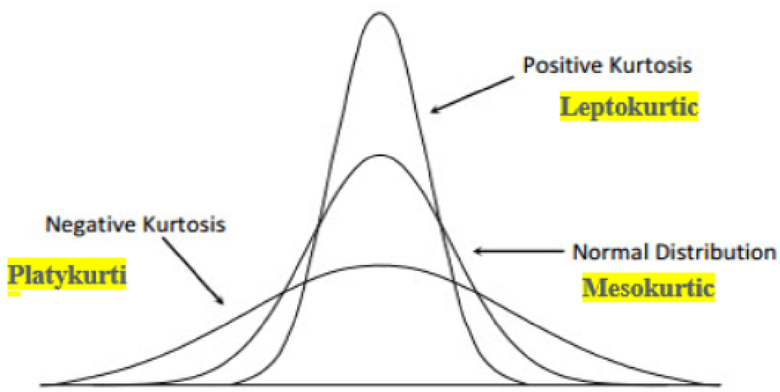
To remove the drawback of the β_1 , we can derive Karl Pearson's γ_1

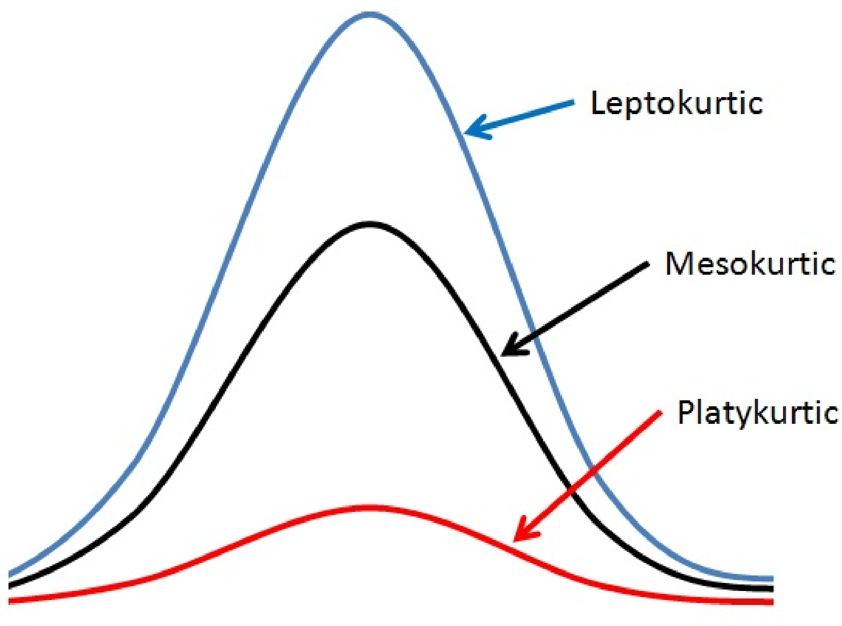
$$\gamma_1 = \sqrt{\beta_1}$$
$$\gamma_1 = \frac{\mu_3}{\mu_2^{3/2}}$$

- If μ_3 is positive, the distribution has positive skewness
- If μ_3 is negative, the distribution has negative skewness
- If μ_3 is zero, the distribution is symmetrical

5.2 Kurtosis

Kurtosis is the measure of the peak and the curve and the "fatness" of the curve.





The kurtosis is calculated using β_2

$$\beta_2 = \frac{\mu_4}{\mu_2^2}$$

The value of β_2 tell's us about the type of curve

- Leptokurtic (High Peak) when $\beta_2 > 3$
- Mesokurtic (Normal Peak) when $\beta_2 = 3$

- Platykurtic (Low Peak) when $\beta_2 < 3$

5.2.1 Karl Pearson's γ_2

γ_2 is defined as:

$$\gamma_2 = \beta_2 - 3$$

- Leptokurtic when $\gamma_2 > 0$
- Mesokurtic when $\gamma_2 = 0$
- Platykurtic when $\gamma_2 < 0$

6 Basic Probability

6.1 Conditional Probability

If some event B has already occured, then the probability of the event A is:

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

$P(A \mid B)$ is read as A given B. So we are given that B has occured and this is probability of now A occuring.

6.2 Law of Total Probability

The law of total probability is used to find probability of some event A that has been partitioned into several different places/parts.

$$P(A) = P(A|B_1)P(B_1)+P(A|B_2)P(B_2)+P(A|B_3)P(B_3)+...+P(A|B_i)P(B_i)$$

$$P(A) = \Sigma P(A|B_i)P(B_i)$$

Example, Suppose we have 2 bags with marbles

- Bag 1 : 7 red marbles and 3 green marbles
- Bag 2 : 2 red marbles and 8 green marbles

Now we select one bag at random (i.e, the probability of choosing any of the two bags is equal so 0.5). If we draw a marble, what is the probability that it is a green marble?

Sol. The green marbles are in parts in bag 1 and bag 2.
Let G be the event of green marble.
Let B_1 be the event of choosing the bag 1
Let B-2 be the event of choosing the bag 2

Then, $P(G|B_1) = \frac{3}{7+3}$ and $P(G|B_2) = \frac{8}{2+8}$
Now, we can use the law of total probability to get

$$P(G) = P(G|B_1)P(B_1) + P(G|B_2)P(B_2)$$

Example 2, Suppose a there are 3 forests in a park.

- Forest A occupies 50% of land and 20% plants in it are poisonous
- Forest B occupies 30% of land and 40% plants in it are poisonous
- Forest C occupies 20% of land and 70% plants in it are poisonous

What is the probability of a random plant from the park being poisonous.

Sol. Since probability is equal across whole area of the park. Event A is plant being from Forest A, Event B is plant being from Forest B and Event C is plant being from Forest C. If event P is plant being poisonous, then using law of total probability,

$$P(P) = P(P|A)P(A) + P(P|B)P(B) + P(P|C)P(C)$$

And we know P(A) = 0.5, P(B) = 0.3 and P(C) = 0.2. Also P(P|A) = 0.20, P(P|B) = 0.40 and P(P|C) = 0.70

6.3 Some basic identities

- Probabilities follow law of inclusion and exclusion

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

- DeMorgan's Theorem

$$P(\overline{A \cap B}) = P(\overline{A} \cup \overline{B})$$
$$P(\overline{A \cup B}) = P(\overline{A} \cap \overline{B})$$

- Some other Identity

$$P(\overline{A} \cap B) + P(A \cap B) = P(B)$$
$$P(A \cap \overline{B}) + P(A \cap B) = P(A)$$

7 Probability Function

It is a mathematical function that gives probability of occurance of different possible outcomes. We use variables to represent these possible outcomes called **random variables**. These are represented by capital letters. Example, X, Y, etc. We use these random variables as:

Suppose X is flipping two coins.

$$X = \{HH, HT, TT, TH\}$$

We can represent it as,

$$X = \{0, 1, 2, 3\}$$

Now we can write a probability function $P(X = x)$ for flipping two coins as :

x	$P(X = x)$
0	0.25
1	0.25
2	0.25
3	0.25

Another example is throwing two dice and our random variable X is sum of those two dice.

x	$P(X = x)$
2	1/36
3	2/36
4	3/36
5	4/36
6	5/36
7	6/36
8	5/36
9	4/36
10	3/36
11	2/36
12	1/36

7.1 Types of probability functions (Continious and Discrete random variables)

Based on the range of the Random variables, probability function has two different names.

- For discrete random variables it is called Probability Distribution function.
- For continious random variables it is called Probability Density function.

8 Proability Mass Function

If we can get a function such that,

$$f(x) = P(X = x)$$

then $f(x)$ is called a **Probability Mass Function** (PMF).

8.1 Properties of Probability Mass Function

Suppose a PMF

$$f(x) = P(X = x)$$

Then,

8.1.1 For discrete variables

$$\Sigma f(x) = 1$$

$$E(X^n) = \Sigma x^n f(x)$$

For $E(X)$, the summation is over all possible values of x.

$$Mean = E(X) = \Sigma x f(x)$$

$$Variance = E(X^2) - (E(X))^2 = \Sigma x^2 f(x) - (\Sigma x f(x))^2$$

To get probabilities

$$P(a \leq X \leq b) = \sum_a^b f(x)$$

$$P(a < X \le b) = (\sum_{a}^{b} f(x)) - f(a)$$

$$P(a \le X < b) = (\sum_{a}^{b} f(x)) - f(b)$$

Basically, we just add all $f(x)$ values from range of samples we need.

8.1.2 For continious variables

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

$$E(X^n) = \int_{-\infty}^{\infty} x^n f(x)dx$$

We only consider integral from the possible values of x. Else we assume 0.

$$Mean = E(X) = \int_{-\infty}^{\infty} xf(x)dx$$

$$Variance = E(X^2) - (E(X))^2 = \int_{-\infty}^{\infty} x^2 f(x)dx - (\int_{-\infty}^{\infty} xf(x)dx)^2$$

To get probability from a to b (inclusive and exclusive doesn't matter in continious).

$$P(a < X < b) = \int_{a}^{b} f(x)dx$$

8.2 Some properties of mean and variance

- Mean

$$E(aX) = aE(X)$$

$$E(a) = a$$

\[E(X + Y) = E(X) + E(Y)]

- Variance

If

$$V(X) = E(X^2) - (E(X))^2$$

Then

$$V(aX) = a^2 V(X)$$

$$V(a) = 0$$

9 Moment Generating Function

The moment generating function is given by

$$M(t) = E(e^{tX})$$

9.1 For discrete

$$M(t) = \sum_{0}^{\infty} e^{tx} f(x)$$

9.2 For continious

$$M(t) = \int_{-\infty}^{\infty} e^{tx} f(x) dx$$

9.3 Calculations of Moments (E(X)) using MGF

$$E(X^n) = (\frac{d^n}{dt^n} M(t))_{t=0}$$

10 Binomial Distribution

The use of a binomial distribution is to calculate a known probability repeated n number of times, i.e, doing **n** number of trials. A binomial distribution deals with discrete random variables.

$$X = \{0, 1, 2, ...n\}$$

where **n** is the number of trials.

$$P(X = x) = \ ^nC_x \ (p)^x (q)^{n-x}$$

Here

$$n \rightarrow number\ of\ trials$$

$$x \rightarrow number\ of\ successes$$

$$p \rightarrow probability\ of\ success$$

$$q \rightarrow probability\ of\ failure$$

$$p = 1 - q$$

- Mean

$$Mean = np$$

- Variance

$$Variance = npq$$

- Moment Generating Function

$$M(t) = (q + pe^t)^n$$

10.1 Additive Property of Binomial Distribution

For an independent variable X. The binomial distribution is represented as

$$X\ B(n, p)$$

Here,

$$n \to number\ of\ trials$$

$$p \to probability\ of\ success$$

- Property

If given,

$$X_1 \sim B(n_1, p)$$

$$X_2 \sim B(n_2, p)$$

Then,

$$X_1 + X_2 \sim B(n_1 + n_2, p)$$

- **NOTE**

If

$$X_1 \sim B(n_1, p_1)$$

$$X_2 \sim B(n_2, p_2)$$

Then $X_1 + X_2$ is not a binomial distribution.

10.2 Using a binomial distribution

We can use binomial distribution to easily calculate probability of multiple trials, if probability of one trial is known. Example, the probability of a duplet (both dice have same number) when two dice are thrown is $\frac{6}{36}$. Suppose now we want to know the probability of a 3 duplets if a pair of dice is thrown 5 times. So in this case :

$$number\ of\ trials\ (n) = 5$$

$$number\ of\ duplets\ we\ want\ probability\ for\ (x) = 3$$

$$probability\ of\ duplet\ (p) = \frac{6}{36}$$

$$q = 1 - p = 1 - \frac{6}{36}$$

So using binomial distribution,

$$P(probability\ of\ 3\ duplets) = P(X = 3) = {}^5C_3 \left(\frac{6}{36}\right)^3 \left(\frac{30}{36}\right)^{5-3}$$

11 Poisson Distribution

A case of the binomial distribution where **n** is indefinitely large and **p** is very small and $\lambda = np$ is finite.

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}\ if\ x = 0, 1, 2.....$$

$$P(X = x) = 0\ otherwise$$

$$\lambda = np$$

- Mean

$$Mean = \lambda$$

- Variance

$$Variance = \lambda$$

- Moment Generating Funtion

$$M(t) = e^{\lambda(e^t - 1)}$$

11.1 Additive property

If X_1, X_2, $X_{3..X_n}$ follow poisson distribution with λ_1, λ_2, $\lambda_{3....\lambda n}$
Then,

$$X_1 + X_2 + X_3... + X_n \sim \lambda_1 + \lambda_2 + \lambda_3 + ... + \lambda_n$$

12 Exponential Distribution

A continuous random distribution which has probability mass function

$$f(x) = \lambda e^{-\lambda x} \ , \ when \ x \geq 0$$

$$f(x) = 0 \ , \ otherwise$$

$$where \ \lambda > 0$$

- Mean

$$Mean = \frac{1}{\lambda}$$

- Variance

$$Variance = \frac{1}{\lambda^2}$$

- Moment Generating Function

$$M(t) = \frac{\lambda}{\lambda - t}$$

12.1 Memory Less Property

$$P[X > (s+t) \mid X > t] = P(X > s)$$

13 Normal Distribution

Suppose for a probability funtion with random variable X, having mean μ and variance σ^2. We denote normal distribution using $X \sim N(\mu, \sigma)$
The probability mass funtion is

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)$$

$$-\infty < x < \infty$$

$$-\infty < \mu < \infty$$

$$\sigma > 0$$

Here, $exp(x) = e^x$

- Moment Generating Funtion

$$M(t) = exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right)$$

13.1 Odd Moments

$$E(X^{2n+1}) = 0 \ , \ n = 0, 1, 2, ...$$

13.2 Even Moments

$$E(X^{2n}) = 1.3.5....(2n-3)(2n-1)\sigma^{2n} \ , \ n = 0, 1, 2, ...$$

13.3 Properties

- In a normal distribution

$$Mean = Mode = Median$$

- For normal distribution, mean deviation about mean is

$$\sigma\sqrt{\frac{2}{\pi}}$$

13.4 Additive property

Suppose for distributions X_1, X_2, X_3 ... X_n with means μ_1 , μ_2 , μ_3 ... μ_n and standard deviation ${\sigma_1}^2$, ${\sigma_2}^2$, ${\sigma_3}^2$ ..... ${\sigma_n}^2$ respectively.

Then $X_1 + X_2 + X_3$ will have mean $\boldsymbol{(\mu_1 + \mu_2 + \mu_3 + ... + \mu_n)}$ and standard deviation $\boldsymbol{({\sigma_1}^2 + {\sigma_2}^2 + {\sigma_3}^2 + ... + {\sigma_n}^2)}$

- Additive Case

Given,

$$X_1 \sim N(\mu_1, \sigma_1)$$

$$X_2 \sim N(\mu_2, \sigma_2)$$

Then,

$$aX_1 + bX_2 \sim N\left(a\mu_1 + b\mu_2, \sqrt{a^2\sigma_1^2 + b^2\sigma_2^2}\right)$$