SCALER

PROB STATS CHEAT SHEET

01. Basic Terminology

1. Experiment

Any procedure that can be **infinitely repeated** and has a **well-defined set of possible outcomes.**

- Sample Space(S):

These well defined set of outcomes is known as sample sample

2. Event (E)

A **subset of the sample space** of an experiment.

$$E \subseteq S$$

3. Probability (P)

The likelihood of an event occurring

$$p(A) = \frac{no.\ of\ favourable\ outcomes\ to\ A}{Total\ no.\ of\ possible\ outcomes}$$

4. Mutually Exclusive Events

If two events are mutually exclusive then the **probability of both events** occurring at the **same time** is equal to **zero**.

$$p(A \cap B) = 0$$

5. Independent Events

The occurrence of one **does not change the probability** of the other occurring.

$$P(A \cup B) = 1$$

6. Conditional Probability

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

7. Basic rules of probability

Addition rule	$P(A \cup B) = P(A) + P(B) - P(A \cap B)$
Multiplication Rule	$P(A \cap B) = P(A\|B).P(B) = P(B\|A).P(A)$
Complement Rule	$P(A^C) = 1 - P(A)$

8. Bayes Theorem

$$P(A|B) = \frac{P(B|A).P(A)}{P(B)}$$

9. Expectation

For a **discrete random variable(X)** having a **Probability mass function P(x).**

$$E(X) = \sum_{i=1}^{n} x_i.P(x_i)$$

10. Properties of Expectation

$$E(aX) = a.E(X)$$
$$E(X+b) = E(X) + b$$
$$E(aX+B) = a.E(X) + b$$

11. Variance

The spread of numbers in a data set w.r.t mean

$$Var(X) = \sigma^2 = \frac{1}{n}\sum_{1}^{n}(x_i - \bar{x})^2$$

12. Properties of variance

$$Var(k.X) = k^2.Var(X)$$
$$Var(X_1 + X_2 + X_3 + \dots) = Var(X_1) + Var(X_2) + Var(X_3)\dots$$
$$Var(X+c) = Var(X)$$

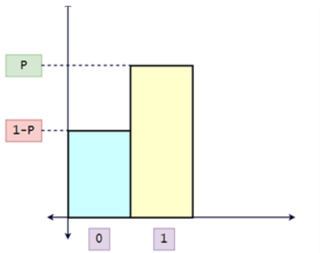
13. Covariance

Variance of two quantities w.r.t each other.

$$Cov(X,Y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

02. Discrete probability distributions:

1. Bernoulli distribution

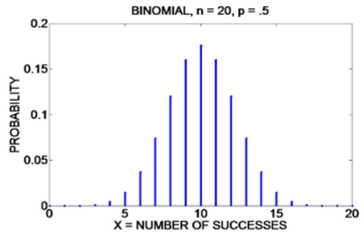


$$P(X=x) = \begin{cases} 1-P, & X=0 \\ P, & X=1 \end{cases}$$

$$E[X] = P$$

$$Var(X) = P(1-P)$$

2. Binomial distribution

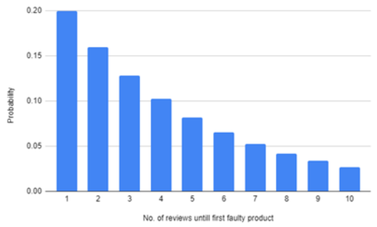


$$P(X=x) = {}^{n}C_x.p^x.(1-p)^{n-x}$$

$$E[X] = n.p$$

$$Var(X) = n.p.(1-p)$$

3. Geometric distribution



$$P(k) = (1-P)^k - 1.P$$

$$E(X) = \frac{1}{p}\ ,\ Var(X) = \frac{1-P}{P^2}$$

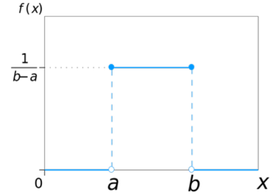
4. Poisson distribution

$$P(X=x) = \frac{\lambda e^{-}x}{x!}$$

$$E(X) = Var(X) = \lambda$$

03. Continuous Probability Distributions

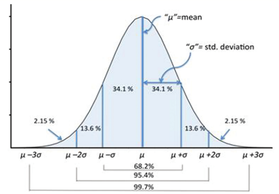
1. Uniform distribution



$$f(x) = \begin{cases} \frac{1}{b-a} & for\ a \leqslant x \leqslant b \\ 0 & for\ x < a\ or\ x < b \end{cases}$$

$$E(X) = \frac{a+b}{2}\ ,\ Variance(\sigma^2) = \frac{(b-a)^2}{12}$$

2. Normal distribution



$$PDF = f(x) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

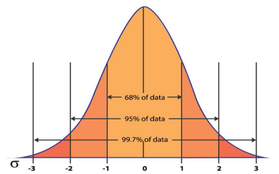
Properties:

Symmetric about mean and has a **bell-shaped** distribution.

Empirical rule:

- 68% of values lie within 1 standard deviation from the mean.
- 95% of values lie within 2 standard deviations from the mean.
- 99.7% of values lie within 3 standard deviations.

3. Standard Normal distribution



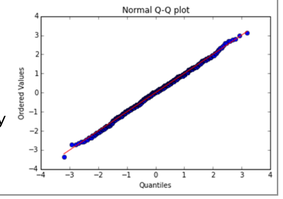
$$z - score = \frac{x-\mu}{\sigma}$$

Special case of Normal distribution when the mean=0 and standard deviation=1.

04. Quantile

Quantile-Quantile(Q-Q) plot

- Comparing two probability distributions by plotting their quantiles against each other.
- The points in a Q-Q plot roughly lie on the line y = x if the two distributions are similar.



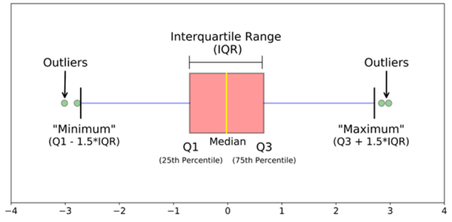
Inter Quartile Range (IQR)

A measure of the middle 50% of a data set.
IQR = Q3 - Q1
Q3 =third quantile value,
Q1 = first quantile value
Used for the purpose of

- Detecting outliers.
- All the points having values greater than or less than are considered to be outliers.

Boxplot

- A standardized way of displaying the distribution of data based on a five-number summary.
- These are “minimum”, first quartile [Q1], median, third quartile [Q3], and “maximum”.



05. Sample VS Population

Sample vs population

Population Mean	$\mu = \frac{1}{N}\sum_{1}^{N} x_i$
Population Variance	$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2$
Sample Mean	$\bar{x} = \frac{1}{n}\sum_{1}^{n} x_i$
Sample Variance	$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$

Standard error

the spread of sample means around the population mean

$$Standard\ \ error = \frac{\sigma}{\sqrt{n}}$$