Notes on Probability and Statistics

Table of Contents

Further reading: **https://greenteapress.com/thinkstats/thinkstats.pdf**

- **Experiment** is any procedure that can be infinitely repeated and has a well-defined set of possible outcomes, known as the sample space (S).

- **Event (E)** is a subset of the sample space of an experiment. i.e., E ⊆ S

- **Probability (P)** is the likelihood of an event occurring and is given as:.

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

- **Mutually Exclusive Events :**
  If two events are mutually exclusive then the probability of both the events occurring at the same time is equal to zero. i.e. $P(A \cap B) = 0$.
  For example, while tossing of a coin, coming up heads or tails are two mutually Exclusive Events.

- **Mutually Exhaustive Events :** When two events 'A' and 'B' are exhaustive, it means that one of them must occur. i.e. $P(A \cup B) = 1$.
  For example, while tossing of a coin, coming up heads or tails are two mutually Exhaustive Events.

- **Independent Events :** Two events are independent if the occurrence of one does not change the probability of the other occurring.
  If two events 'A' and 'B' are independent, then : $P(A \cap B) = P(A).P(B)$

- **Conditional Probability :** is defined as the likelihood of an event or outcome occurring, based on the occurrence of a previous event or outcome.
  P(Event A will occur given that Event B has already occurred) = $P(A|B) = \frac{P(A \cap B)}{P(B)}$
  Two events A and B are mutually independent if $P(A|B) = P(A)$.
  also, $P(A|B) \neq P(B|A)$.

- **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 : $P(A^c) = 1 - P(A)$

$A \cap B$
A and B

$A \cap B = \varnothing$
A and B is the empty set

$A \cup B$
A or B

- **Bayes Theorem :** Conditional probability of each of a set of possible causes (B), given an observed outcome (A) is given as:

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

E.g. A man speaks truth ¾ times. He draws a card and reports 'King'. What is the probability of it actually being a king?

P(Man speak truth)= P(T) =¾, P(drawing a king)=P(K)= 4/52 = ¼

P(drawing a king given that the man speaks truth) = P(K|T)

$$P(K|T) = \frac{P(T|K).p(K)}{P(T)} = \frac{(3/4).(1/13)}{(1/13).(3/4) + (12/13).(1/4)} = 0.2$$

- **Types of a Random variable :**

Discrete random variable: A random variable with a finite or countable number of possible values. e.g. random variable of the outcome when a dice is thrown. It can take integer values in range 1 to 6.

Continuous random variable: A random variable that can take an infinite number of possible values. E.g. a random variable representing the height of students in a class.

- **Probability Distribution Function (PDF)** is a function that is used to give the probability of all the possible values that a random variable can take.

If the random variable is discrete, then it is called **Probability Mass Function**.

For the continuous random variable, it is called **Probability Density Function**.

- **Cumulative Distribution Function (CDF):** returns the probability that a random variable will take a value less than or equal to x.
- **Expectation** of a discrete random variable(X) having a Probability mass function P(x) is the weighted average of possible values that X can take.

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

Note :For a uniformly distributed random variable, Expectation is equal to the mean.

Properties of Expectation :

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

- **Variance** is a statistical measurement that is used to determine the spread of numbers in a data set with respect to the average value or the mean. It is given as :

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

Where, σ is the **standard deviation**. Therefore,

$$Variance = (Standard\ Deviation)^2$$

For a random variable *X,* **Variance** can be given as :

$$Var(X) = E(X^2) - [E(X)]^2$$

Properties of variance:

1. $Var(k.X) = k^2.Var(X)$

2. Assuming that the samples were collected independently.

   $$Var(X_1 + X_2 + X_3 + ...) = Var(X_1) + Var(X_2) + Var(X_3) + ...$$

3. $Var(X + c) = Var(X)$

- **Covariance** is the variance of two quantities with respect to each other. It is a measure of how much two random variables vary together.

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

- **Discrete probability distributions:**

1. **Bernoulli distribution:** The distribution of a random variable which takes a binary, boolean output: 1 with probability p, and 0 with probability (1-p).

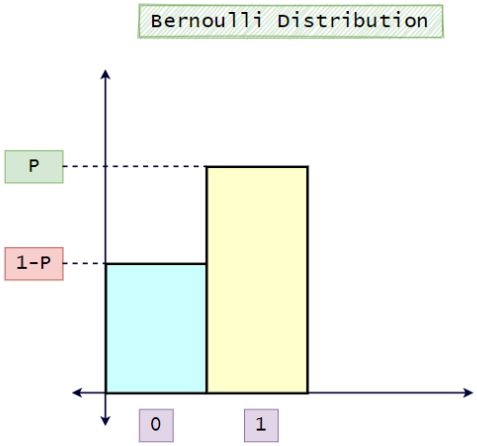
   Let X follows the Bernoulli distribution, i.e. $X \sim Bern(p)$

   then,

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

   and

$$E(X) = p$$
$$Var(X) = p(1-p)$$



2. **Binomial distribution:** It is the probabiity distribution of getting *x* successes in *n* Bernoulli trials.

   Let X follows the Binomial distribution, i.e. $X \sim B(n, p)$

   where, n = number of trials , p = probability of success

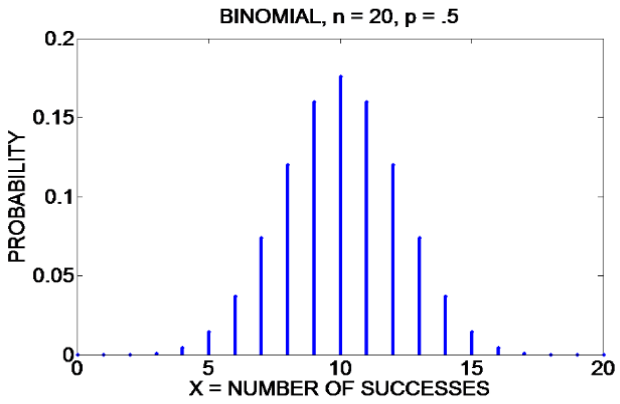
   then,

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

   and

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

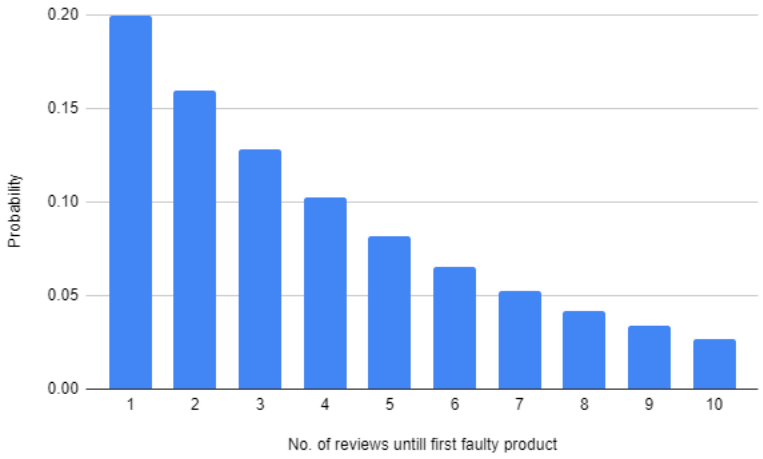
E.g. Probability Distribution of number of heads in 20 coin flips.



3. **Geometric distribution:** It is the probability distribution of number of Bernoulli trials needed to get one success.

   Its Probability mass function is given as : $P(k) = (1-p)^{k-1}.p$

   E.g. The probability of getting a faulty product after reviewing *k* non-faulty products.



4. **Poisson distribution:** It gives the probability of an event happening a certain number of times (x) within a given interval.

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

Expectation, mean and variance are all equal to λ.

For example, the Probability distribution of no. of people visiting a website in one hour.

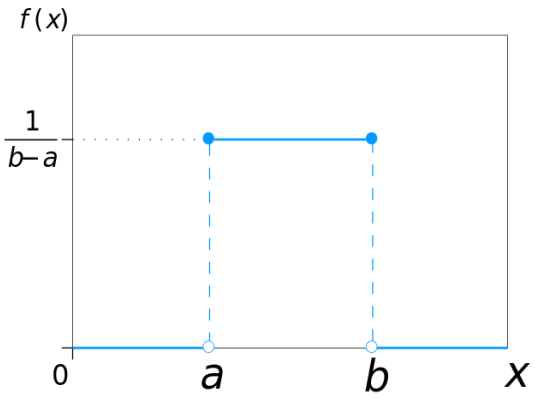
- **Continuous Probability distributions:**
  1. **Uniform distribution :** uniform distribution refers to a type of probability distribution in which all outcomes are equally likely.

     The PDF of a continuous uniform distribution is given as:

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

Mean = $E(X) = \frac{a+b}{2}$

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



  2. **Normal distribution:** This distribution is very common in nature and has a symmetric bell shaped curve. It is also called as Gaussian distribution.

     PDF of Normal distribution is given as

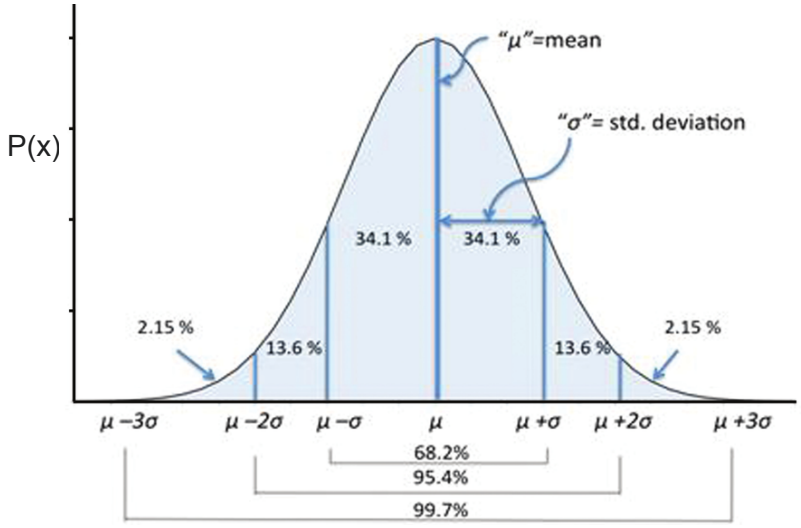
$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

Where μ is mean and σ is standard deviation.

Properties of Normal distribution :

- ➢ Symmetric about mean and has a bell shaped distribution.
- ➢ Mean (μ) = mode = median
- ➢ **Empirical rule :**

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



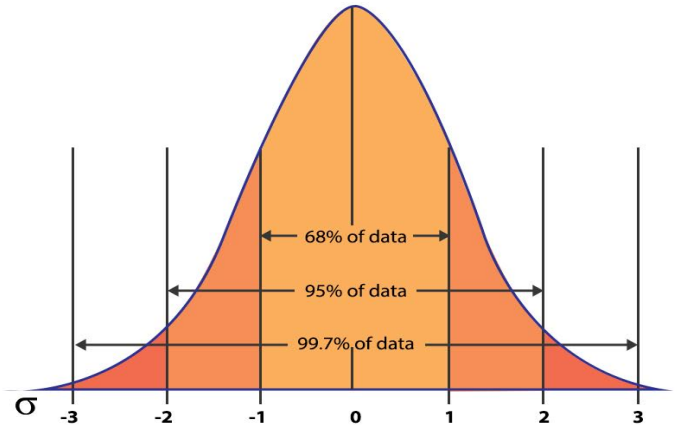
3. **Standard Normal distribution:** It is a special case of Normal distribution when the mean is 0 and the standard deviation is 1.

   Any normal distribution can be standardized by converting its values into z-scores.

   z- score of a value *x* in normal distribution is given by :

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

   z-score tells us about the distance from mean in terms of standard deviation.



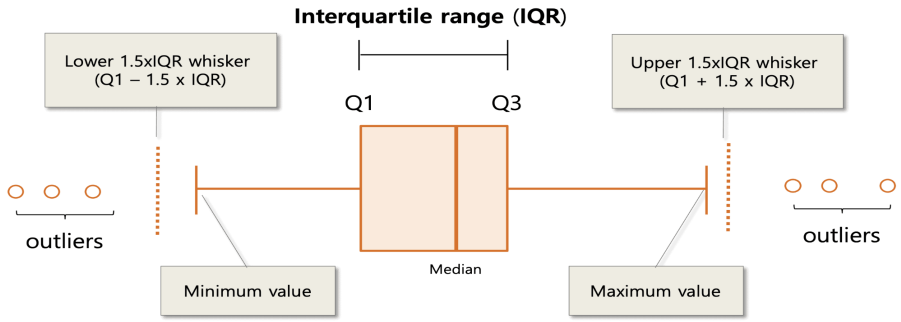
- **Outlier :** Any data point which is far away from the rest of the data points.
- **Inter Quartile Range (IQR)** is a measure of the middle 50% of a data set.

  IQR = Q3 - Q1

  where, Q3 is the third quantile value and Q1 is the first quantile value of the data.
- IQR is used for the purpose of **detecting outliers** in the data**.**

  All the points having value greater than ($Q3 \ + 1.5 \times IQR$) **or** less than ($Q1 - 1.5 \times IQR$) are considered to be the outliers.

- **Boxplot** is 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”.



- A **Sample** is an analytic subset of a larger population. If the sample is well selected, the sample will be generalizable to the 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 = $\overline{x} = \frac{1}{n}\sum_{1}^{n} x_i$, *Sample Variance* = $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$

The reason we use *(n-1)* rather than *n* in the denominator is so that the sample variance will be an unbiased estimator of the population variance.

As the sample size is increased, the sample mean gets closer to the population mean and the variance in the means of samples decreases.

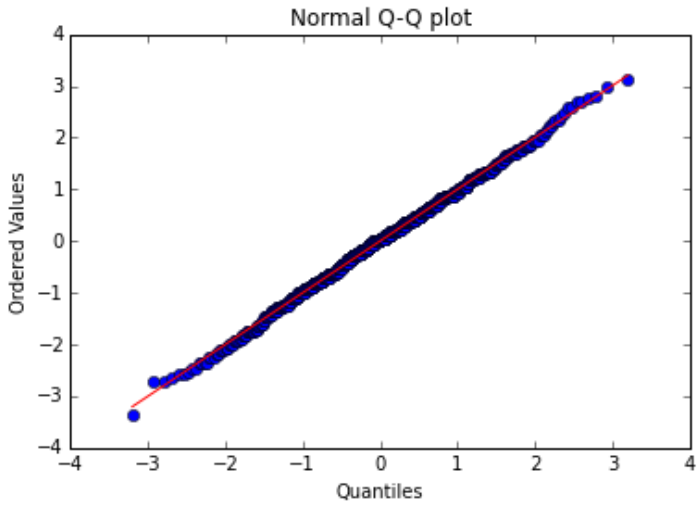
- **Standard error** is the spread of sample means around the population mean.

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

As the sample size increases, Standard error decreases.

- **Central Limit Theorem (CLT)** states that 'the distribution of sample means is Gaussian, no matter what the shape of the original distribution is.
  The assumption of CLT is that the population mean and population standard deviation should be finite and sample size should be >=30 (commonly accepted number is 30).

- **Quantile-Quantile (Q-Q) plot:** is a graphical way for comparing two probability distributions by plotting their quantiles against each other.
  If the two distributions being compared are similar, the points in the Q–Q plot will approximately lie on the line *y* = *x*. Given below is the Q-Q plot of two similar distributions.



- **Hypothesis testing** is a method of statistical inference used to decide whether the data at hand sufficiently support a particular hypothesis.
  A hypothesis test is the means by which we generate a **test statistic** that directs us to either reject or not reject the null hypothesis.
  If **p value** is lower than **significance level**; then we reject the **null hypothesis**, else we fail to reject the null hypothesis.

  For e.g., Testing whether a coin is fair or not. Say we conducted our experiment and got 65 heads out of 100 tosses.

  The **Null hypothesis (**H_0**)** represents the assumption that is made about the data sample whereas the **alternative hypothesis (**H_a**)** represents a counterpoint .

  Null hypothesis **(**H_0**) :** the coin is fair (p=0.5)

  Alternative hypothesis **(**H_a**) :** the coin is not fair (p>0.5)

  By observing the alternative hypothesis we can say that it is a right tailed test.