FTDS // PROBABILITY & DISTRIBUTION

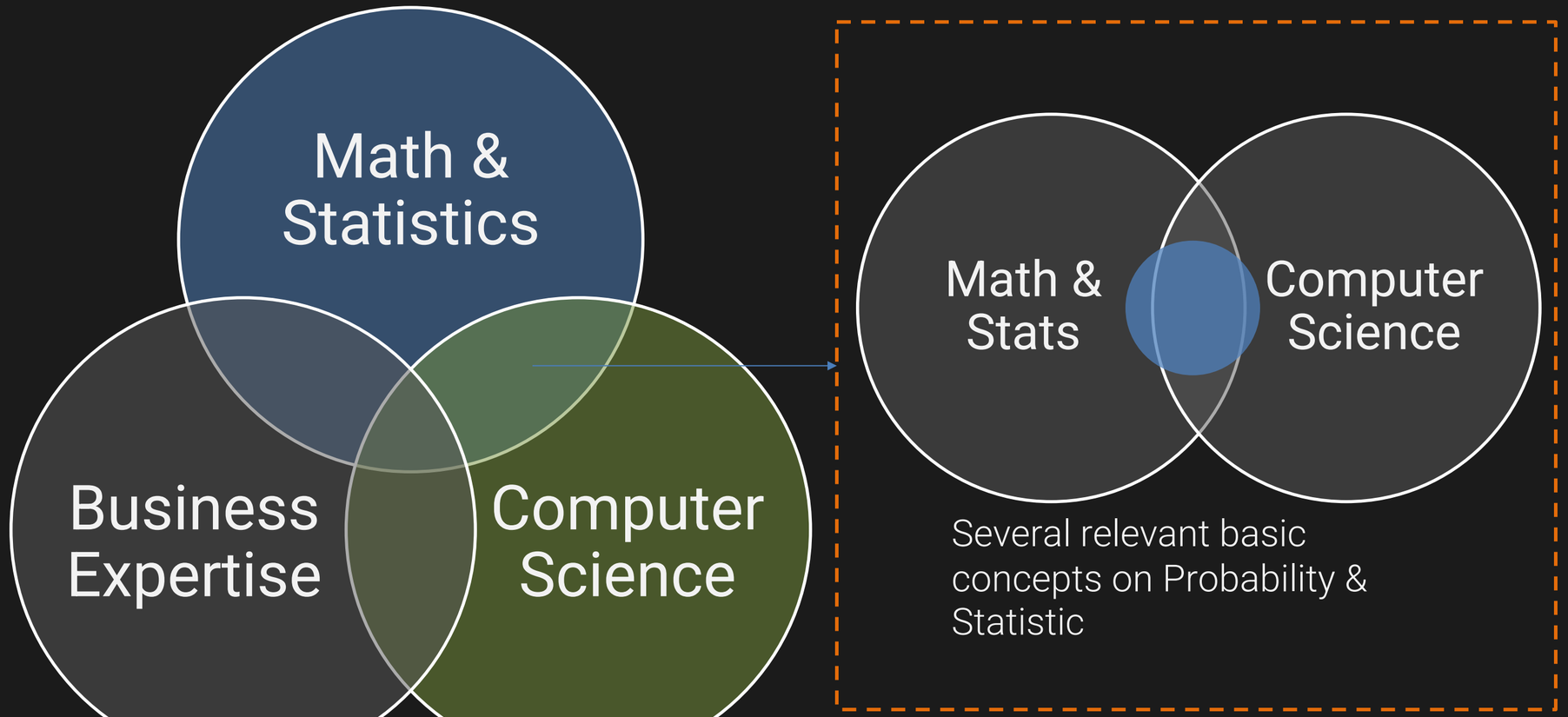
Hacktiv8 DS Curriculum Team

Phase 0 Learning Materials

Hacktiv8 DS
Curriculum
Team

Contents

Learning Scopes



Notes: This is not a substitution of one semester probability or statistical course.

We don’t want to be overloaded, nor do we have the capacity to crammed it in a week

Important Concepts

- Probability and Events
- Probability distribution
- Central tendency
- Variability
- Relationship between variables
- Hypothesis testing
- Statistical significance

Probability Basics

Probability is a value between 0 and 1 that indicates the likelihood of a particular event, with 0 meaning that the event is impossible, and 1 meaning that the event is inevitable.

$$Probability = \frac{Number\ of\ sample\ points}{Total\ number\ of\ sample\ points\ in\ sample\ space} = P(A)$$

Event	P(A)
Head on fair coin flip	1/2
One side of dice	1/6
Get a Queen in a deck of card	1/13
Get a Spade in a deck of card	1/4

Complement and Bias

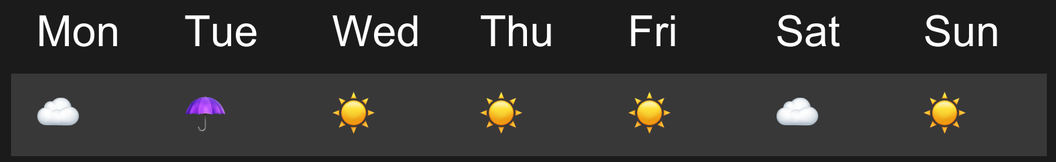
The *complement* of an event is the set of *sample points* that do ***not*** result in the event.

$$P(A') = 1 - P(A)$$

Event	P(A')
Head on fair coin flip	1/2
One side of dice	5/6
Get a Queen in a deck of card	12/13
Get a Spade in a deck of card	3/4

Probability of the complement and the event itself must add up to 1

Bias makes one outcome more likely than another



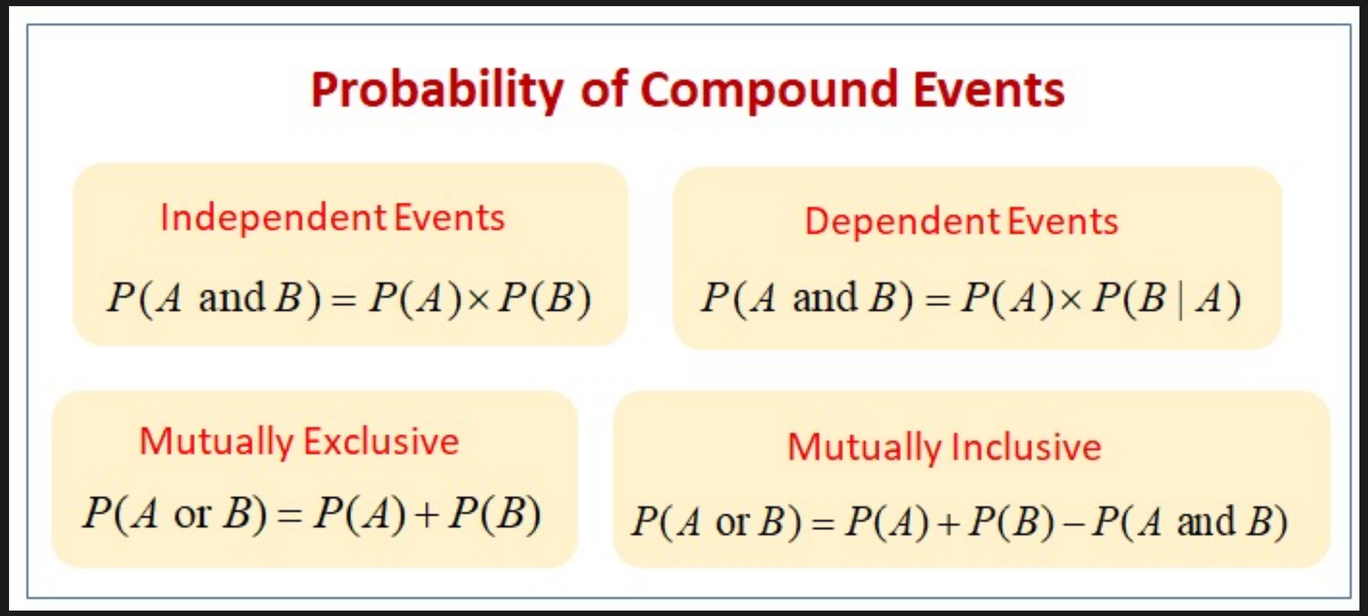
Probability of sunny = P(A) = 4/7

Probability of cloudy = P(B) = 2/7

Probability of raining = P(C) = 1/7

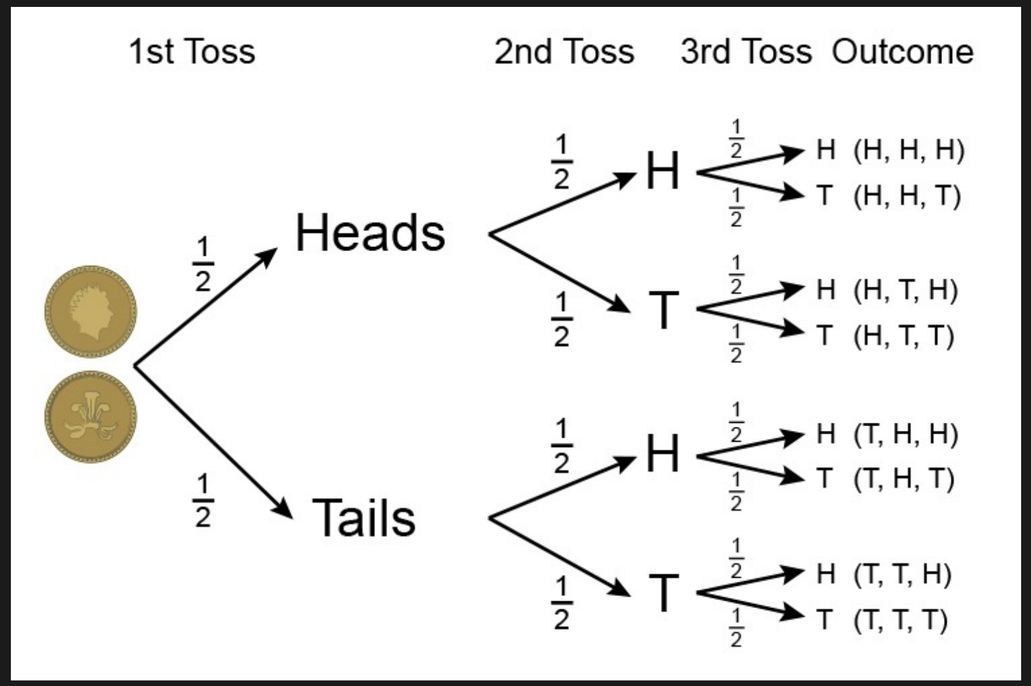
Compound Events

Involves the probability of more than one outcome. Another way to view compound events is as *a combination of two or more simple events.*



Probability Trees

Probability trees are useful for calculating combined probabilities for sequences of events. It helps you to map out the probabilities of many possibilities graphically, without the use of complicated probability formulas.



Binomial Distribution

The binomial distribution is used when there are exactly two mutually exclusive outcomes of a trial. These outcomes are appropriately labeled "success" and "failure".

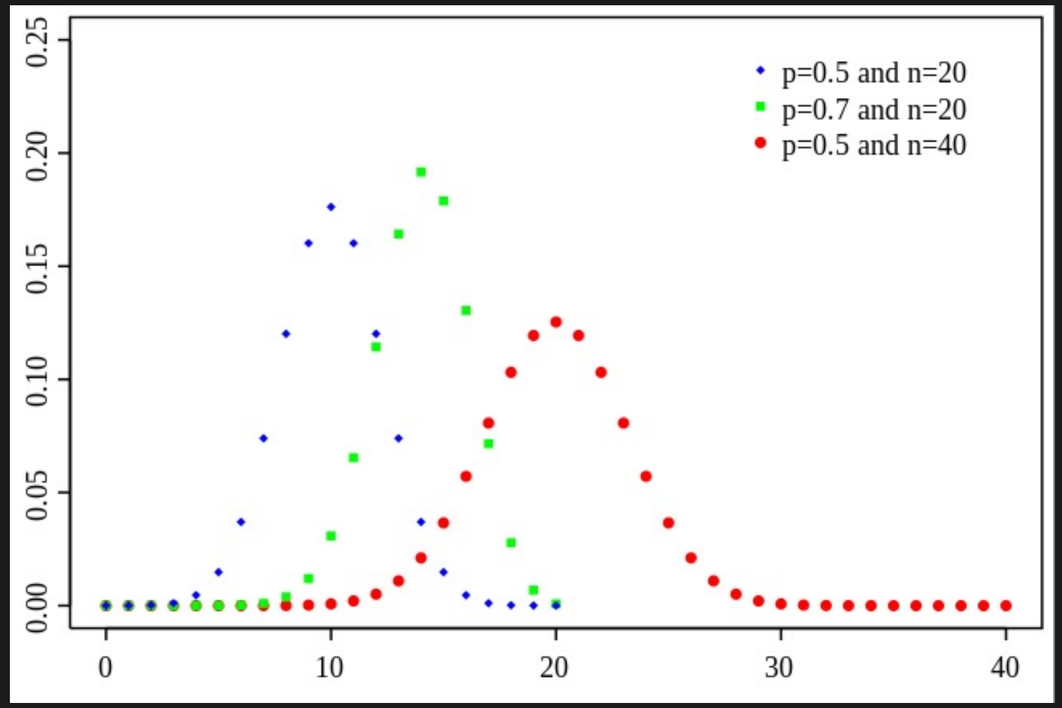
Mean (Expected Value) $\mu = np$

Total value we can expect given probability of a value occurring and the number of trials.

Standard Deviation $\sigma = \sqrt{np(1-p)}$

Measure of the amount of variation or dispersion of a set of values.

We will revisit the Mean and Standard Deviation concept on the next lectures.



Cumulative Distribution Function

In probability theory and statistics, the cumulative distribution function (CDF) of a real-valued random variable X, or just distribution function of X, evaluated at x, is the probability that X will take a value less than or equal to x.

In the case of a scalar continuous distribution, it gives the area under the probability density function from minus infinity to x. Cumulative distribution functions are also used to specify the distribution of multivariate random variables.

Max value of CDF is always one and min values is always 0

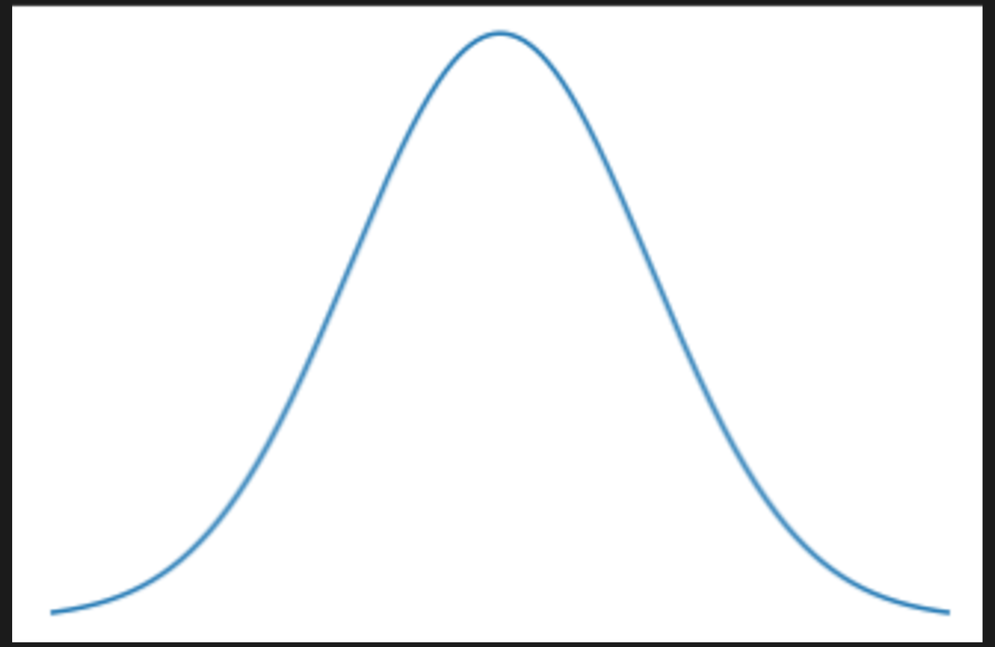
$$F_X(x) = P[X \leq x]$$

Probability Distribution Function

In probability theory, a probability density function (PDF), or density of a continuous random variable, is a function whose value at any given sample (or point) in the sample space (the set of possible values taken by the random variable) can be interpreted as providing a relative likelihood that the value of the random variable would equal that sample.

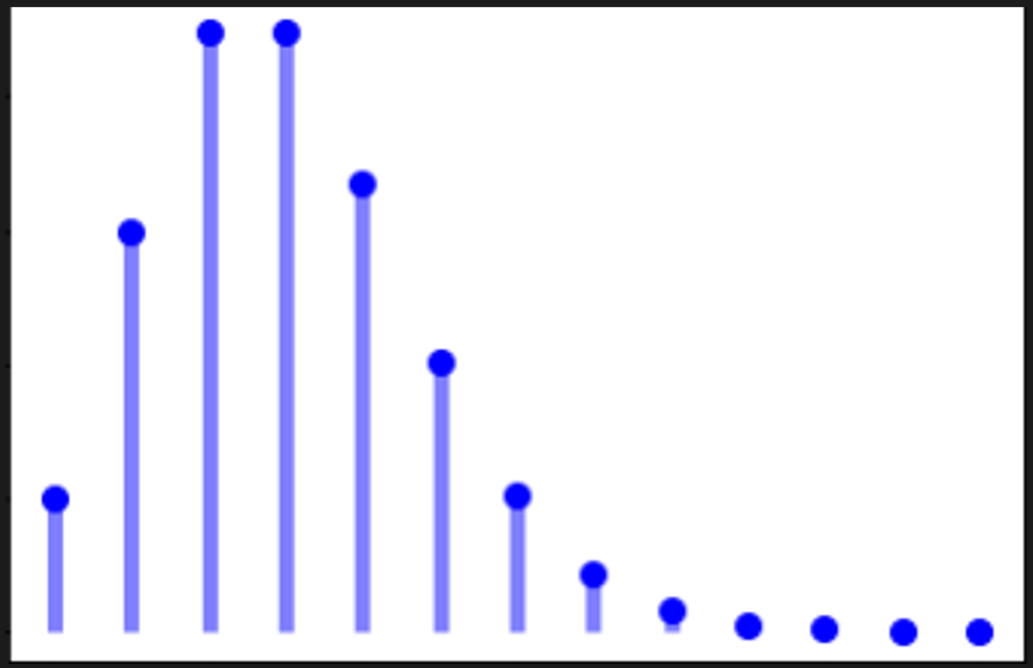
Properties of PDF:

1. Area under PDF is always accumulated to one
2. Integrate the pdf to get the cdf.

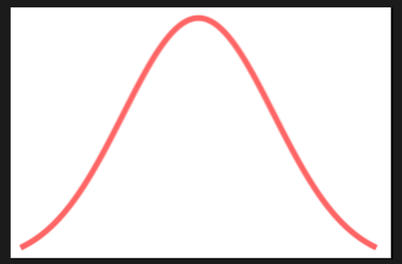


Probability Mass Function

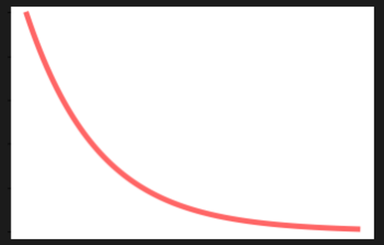
In probability and statistics, a probability mass function is a function that gives the probability that a discrete random variable is exactly equal to some value. Sometimes it is also known as the discrete density function. It can be regarded as PDF in term of discrete variable.



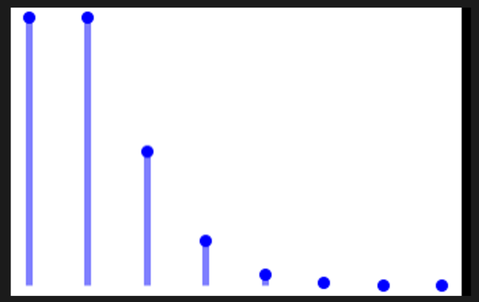
Common Distribution Functions



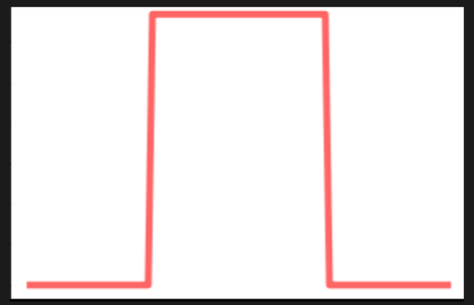
Gaussian Distribution



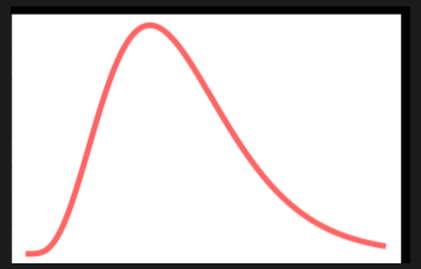
Exponential Distribution



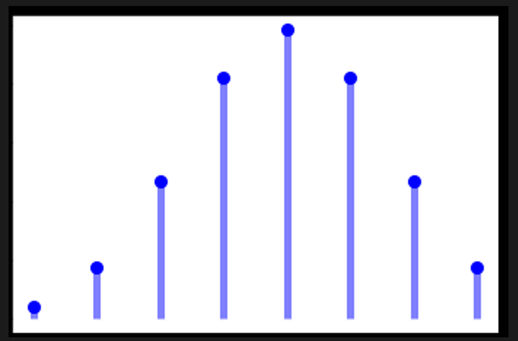
Poisson Distribution



Uniform Distribution



Gamma Distribution



Binomial Distribution

External References

Colab Link

Visit Here