1. According to Murphy, Machine Learning (ML) are a set of methods that can automatically detect patterns from the underlying data to derive meaningful conclusions and directions. The uncovered patterns can then be used to predict future data, or to perform other kinds of decision making under uncertainty. Currently, the onslaught of exponential data growth and collection calls for automated methods of data analysis.

According to Tibshirani, Witten et al, Statistical Learning can help us understand the causality between the dependent and independent variables. Statistical learning can be defined as a set of approaches to estimate a function like Y = f(X) + ε, where ε is the underlying error and X can be singular or multiple independent variables whose mathematical operations can help us identify the causal connection with Y a.k.a the dependent variable. ML as stated above can help us make sense of the voluminous data by discerning the relevant patterns to derive meaningful insights.

1. **Binomial Distribution** -> **Application**: This distribution can help us get a definitive idea regarding how many patients may develop a negative side-effect to a prescribed medication.

Diagram

Description automatically generated with medium confidence where (n,k) is n!/(n-k)!

**Bernoulli Distribution** -> **Application**: In a game of football, Bernoulli distribution can be used to determine the success rate of shot taken towards goal. Bernoulli distribution is a generalized version of Binomial distribution.

Table

Description automatically generated

**Poisson Distribution** -> **Application**: Web hosting companies can use Poisson distribution to estimate hourly traffic to a website and thus allocate bandwidth accordingly.

Text

Description automatically generated

**Uniform Distribution**: **Application** -> Raffle tickets. The event’s organising committee tends to select a particular seat out of thousands of seats and reward the person sitting on it with a prize. The people participating in the event buy numbered raffle tickets and each of them possesses an equal chance of winning. This is because the probability of a seat being chosen by the organisers as the winner is equal to the remaining seats.

Table

Description automatically generated

**Normal Distribution**: **Application** -> The most used statistical distribution. It can be used to see the distribution or underlying patterns for things ranges from heights of adults to blood pressures across age groups.

Diagram

Description automatically generated

**Student’s T Distribution**: **Application** -> This is used to compare and determine if the mean value of two samples is statistically different. This can be used to do an analysis like determining the BMI of different demographics and hypothesize the possible causes.

Text, letter

Description automatically generated

**Chi-Squared Distribution**: **Application** -> This can be used to gauge public perception or political preference of a population subset.

Diagram

Description automatically generated with medium confidence

**Gamma Distribution**: **Application** -> To identify factors that can significantly alter or affect the hospital stay of patients after certain surgical procedures.

A picture containing diagram

Description automatically generated

P.S. – I don’t exactly understand the need nor the differences between the F(x) and f(x) notations.

**Beta Distribution**: **Application** -> This distribution is useful for simulating cloud patterns over a region for specific periods of time.

Table

Description automatically generated

**Pareto Distribution**: **Application** -> This can be used to model change in population over time.

Table

Description automatically generated

**Random Variable (RV)** – A random variable assigns a numerical value to each outcome of a chance event. Random variables are not the same as the events they quantify.

There are two types of random variables – Continuous Random Variables and Discrete Random Variables.