Professional Certificate in Machine Learning and Artificial Intelligence

Module 3: Probability for Machine Learning

Quick Reference Guide

Learning outcomes

- Categorise examples of probability versus statistics.
- Calculate absolute, conditional and total probabilities.
- Classify independent versus dependent events.
- Recognise equations that use Bayes' rule correctly.
- Run simulations using random number generating libraries in Python and NumPy.
- Distinguish between discrete and continuous random variables.
- Identify and compute probabilities and values related to binomial distribution.
- Define the central limit theorem function values.
- Predict the impact of changes in variables on changes in distributions.

Difference between probability and statistics

Statistics	Probability
Statistics is a branch of mathematics which helps in going from the data to fit a model that explains that data. Data → Model	Probability is the mathematical framework of understanding the data or the properties of the data that the model can produce. Model → Data

Probability theory example

Coin flips

Models of probability

There are some basic and interesting models in probability theory that help in building intuition for more elaborate models.

1. Fair coin model

If we assume that we're dealing with a fair coin, then we should not expect more heads than tails. In the fair coin model, we expect the frequency of heads and tails to be approximately the same or between 0.5 or 50 percent. It doesn't imply that we get heads and tails alternately. It simply means that we should expect a balanced distribution of heads or tails as we flip the coin over and over again.

2. Loaded coin model

It is a model that describes a coin with varying probabilities for heads and tails.
We could have a loaded coin for which

- the frequency of heads is 0.6 and the frequency of tails is 0.4
- the frequency of heads is 0.75 and the frequency of tails is 0.25
- the frequency of heads is 0.99 and the frequency of tails is 0.01

The law for total probability works here. This means the probabilities of the possible outcomes add up to 1.

3. Two flips of a coin

If we flip two coins consecutively, what is the probability of getting heads?
We can solve this problem by building a table of possible outcomes. This works only for small problems, but it helps in building some intuition about probability.

Coin 1	Coin 2
Heads	Heads
Heads	Tails
Tails	Heads
Tails	Tails

The probability of getting at least one heads is 0.75.The probability of getting exactly one heads is 0.25.

Bayes' rule

Bayes' rule is the mathematical foundation for what we call statistical inference. Statistical inference means that with Bayes' rule, we can find the predictive probability or the probability of something happening based on our understanding or assumptions of what happened in the past.

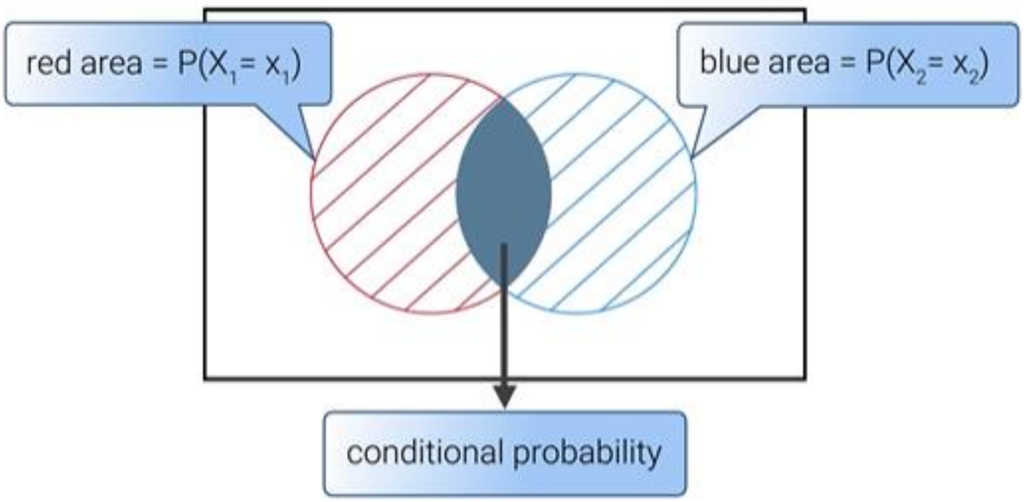
Example

Let's think of probabilities in terms of areas. The square here is the space of all possible variables of all possible occurrences, i.e. variables X_1 and X_2.

Red area — It is equal to the probability that the variable X_1 has taken some value which is x_1.

Blue area — It is the probability that a second random variable, X_2, has taken a second value x_2.

Intersection of the red and blue areas is called 'Conditional Probability'.



The probability that X takes some value, say x2, given X1 has taken x1

$$P(X_2 = x_2 \,|\, X_1 = x_1) \;=\; P(X_1, X_2)/P(X_1)$$

So, given that X1 happened, what is the probability that X2 happened? And that conditional probability here is the intersection in relation to the bigger area. Thus, geometrically, this is what we call conditional probability.

We arrive at Bayes' rule after combining the two formulas of **conditional probability**.

- $P(X_2 \mid X_1) = P(X_1, X_2)/P(X_1)$
- $P(X_1 \mid X_2) = P(X_1, X_2)/P(X_2)$

Naive Bayes' rule: $P(X_2 \mid X_1) = [P(X_1 \mid X_2)\ P(X_2)]/P(X_1)$

Bayes' rule application

Bayes' rule is a way of inverting probabilities. Algebraically, Bayes' rule is expressed like this:

$P(A \mid B) = [P(B \mid A) \times P(A)]/P(B)$

Inverse probabilities

- Probabilities that can be expressed in terms of each other
- One of the first instances of inference using probability theory

Probability Distributions

The basic question probability distributions help us address is, what is the probability of an outcome of a random variable? Data is the realisation of a random variable.

Example:

Heads or tails in the coin flip.

The Central Limit Theorem helps us to model the probability distribution when the number of outcomes is too large or infinite.

A lot of random variables are not discrete like people's height, weight or body temperature. In such cases, we need a function representation of the probability distribution, like a constant function, or a linear function, or even an exponential function. One of the most common continuous probability distributions is the normal or Gaussian distribution.

Binomial Distribution

This is an important concept from probability theory, which will lead to another important distribution called the normal distribution

Binomial distribution → Normal distribution

Comes about when we start counting and taking averages

Discrete variables

Continuous variables

If we flip a coin **five times**, what is the probability of getting exactly **three heads**?
We can answer this by using **binomial coefficients**.

Binomial coefficient = C_k^n

Where:

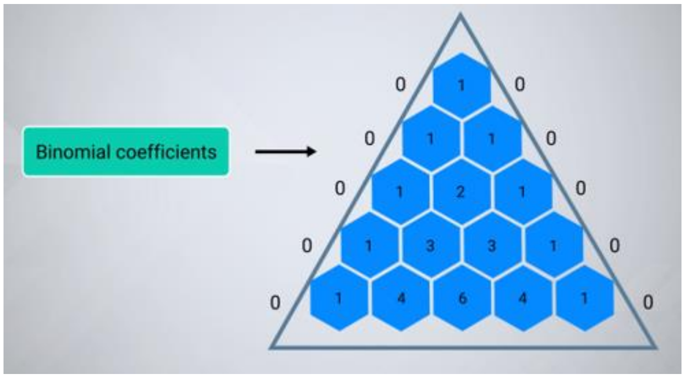
k = number of heads

n = number of tosses

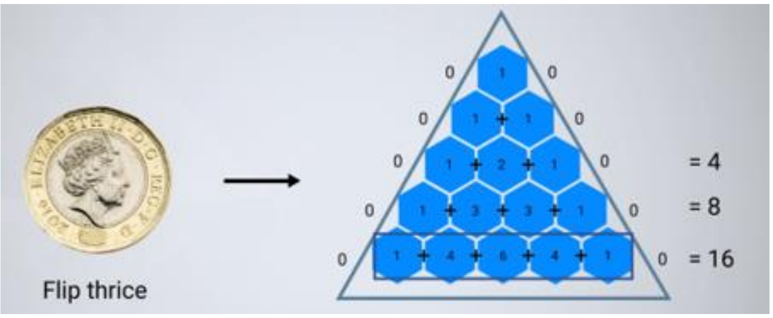
$= n!/k!\,(n-\ k)!$

Pascal's Triangle

We can create Binomial Coefficients by a procedure called Pascal's triangle.



Each line in the triangle represents the number of tosses. If you add up all the numbers in each row, you end up with all the possible outcomes.



We can derive a formula here:

P(#Heads) = k) in ‘n’ Tosses = $C_R^M/2^n$

There is a more geometric way of representing this distribution in the form of a histogram, which will lead us next to the Central Limit Theorem and the normal distribution.

Central Limit Theorem

The **Central Limit Theorem** is a fundamental result from probability theory, which is applicable and useful in various statistical and modelling problems.

Example

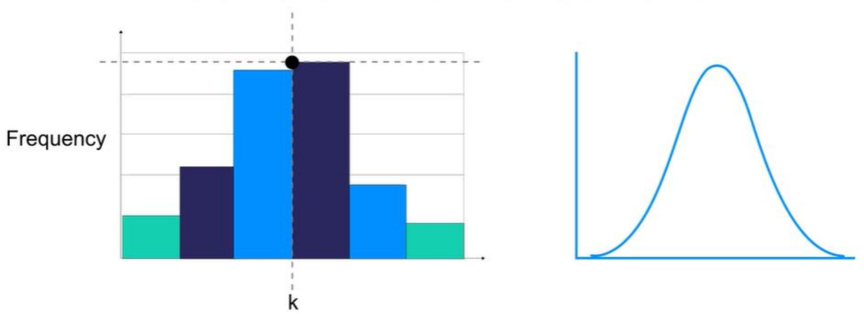
In a series of ‘n’ coin tosses,

Fair coin	$\theta = \frac{1}{2}$
Loaded coin	$\theta \neq \frac{1}{2}$

$$f(k) = \boxed{C_k^n}\theta^k\,(1-\theta)^{n-k}$$

Binomial coefficient

We can derive a histogram based on the above expression. After rescaling the limit of the histogram, an approximate Gaussian curve can be drawn. This is also known as the **normal distribution**.



Histogram Gaussian distribution

Central limit theorem (CLT) is the connection between probability results from the binomial distribution and the normal distribution.