Probability, Conditional Probability and Bayes theorem

Axioms of Probability

Probability is a measure of the likelihood of an event to occur. Many events cannot be predicted with total certainty. We can predict only the chance of an event to occur i.e. how likely they are to happen, using it.

Probability is a number that is assigned to each member of a collection of events from a random experiment that satisfies the following properties:

If S is the sample space and E is any event in a random experiment,

(1) $P(S) = 1$

(2) $0 \leq P(E) \leq 1$

(3) For two events E_1 and E_2 with $E_1 \cap E_2 = \emptyset$

$$P(E_1 \cup E_2) = P(E_1) + P(E_2)$$

Probability

Probability of event to happen P(E) = $\dfrac{\textbf{Number of favorable outcomes}}{\textbf{Total Number of outcomes}}$

1) There are 6 pillows in a bed, 3 are red, 2 are yellow and 1 is blue. What is the probability of picking a yellow pillow?

2) There is a container full of colored bottles, red, blue, green and orange. Some of the bottles are picked out and displaced. Sumit did this 1000 times and got the following results:

- No. of blue bottles picked out: 300
- No. of red bottles: 200
- No. of green bottles: 450
- No. of orange bottles: 50

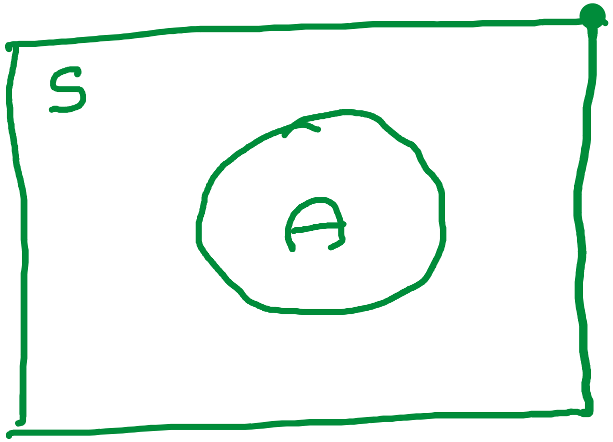
a) What is the probability that Sumit will pick a green bottle?
b) If there are 100 bottles in the container, how many of them are likely to be green?

Probability

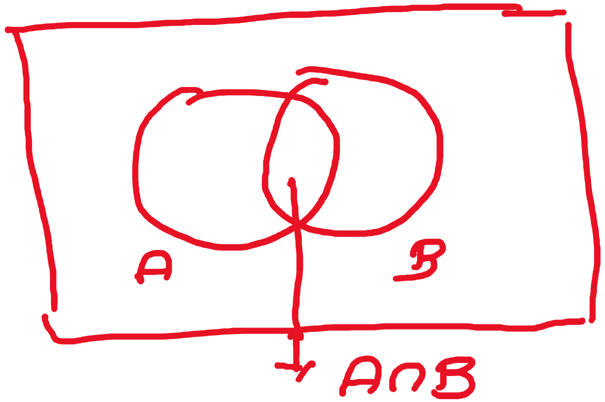
$$A \cup \bar{A} = S$$

$$P(A \cup \bar{A}) = P(S)$$

$$P(A) + P(\bar{A}) = 1$$



$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$



Example 1

Two coins are tossed, find the probability that two heads are obtained. **Note:** Each coin has two possible outcomes H (heads) and T (Tails).

Solution

The sample space S is given by.

S = {(H,T),(H,H),(T,H),(T,T)}

Let E be the event "two heads are obtained".

E = {(H,H)}

We use the formula of the classical probability.

P(E) = n(E) / n(S) = 1 / 4

Example 2

A die is rolled and a coin is tossed, find the probability that the die shows an odd number and the coin shows a head.

Solution

Let H be the head and T be the tail of the coin. The sample space S of the experiment described in question 5 is as follows

S = { (1,H),(2,H),(3,H),(4,H),(5,H),(6,H)

(1,T),(2,T),(3,T),(4,T),(5,T),(6,T)}

Let E be the event "the die shows an odd number and the coin shows a head". Event E may be described as follows

E={(1,H),(3,H),(5,H)}

The probability P(E) is given by

P(E) = n(E) / n(S) = 3 / 12 = 1 / 4

Conditional Probability

The probability of event B given that event A has occurred P(B|A) or, the probability of event A given that event B has occurred P(A|B)

Conditional Probability Formula

$$\underbrace{P(A \mid B)}_{\text{Probability of } A \text{ given } B} = \frac{\overbrace{P(A \cap B)}^{\text{Probability of } A \text{ and } B}}{\underbrace{P(B)}_{\text{Probability of } B}}$$

Conditional Probability

Conditional Probability

Let A and B be events in a sample space S. If $P(A) \neq 0$, then the conditional probability of B given A, denoted by $P(B|A)$, is

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

Useful variations of the formula.

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

Example

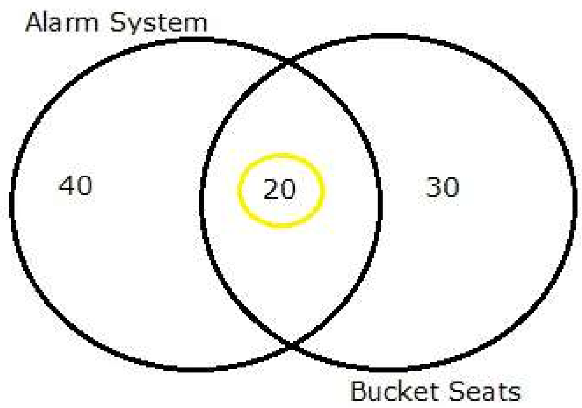
In a group of 100 sports car buyers, 40 bought alarm systems, 30 purchased bucket seats, and 20 purchased an alarm system and bucket seats. If a car buyer chosen at random bought an alarm system, what is the probability they also bought bucket seats?

Step 1: Figure out P(A). It's given in the question as 40%, or 0.4.

Step 2: Figure out P(A∩B). This is the intersection of A and B: both happening together. It's given in the question 20 out of 100 buyers, or 0.2.

Step 3: Insert your answers into the formula:
$P(B|A) = P(A \cap B) / P(A) = 0.2 / 0.4 = 0.5$.

The probability that a buyer bought bucket seats, given that they purchased an alarm system, is 50%.



Venn diagram for 90 buyers, showing that 20 alarm buyers also purchased bucket seats.

Bayes' Theorem

Bayes' Theorem is a way of finding a probability when we know certain other probabilities.

The formula is:

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

Which tells us:	how often A happens *given that B happens*, written **P(A|B)**,
When we know:	how often B happens *given that A happens*, written **P(B|A)**
	and how likely A is on its own, written **P(A)**
	and how likely B is on its own, written **P(B)**

Multiplication Law

$$P(A \cap B) = P(A \mid B)P(B) = P(B \mid A)P(A)$$

Example -: A bag 'A' contains 2 w and 3 red balls, a bag 'B' contains 4 w and 5 black balls. A bag is selected randomly and a ball is drawn from it. Drawn ball is observed to be white. Find the probability that bag 'B' was selected.

Sol: Bag A Bag B

2 w, 3R 4 w, 5B

Let A_1 be the event that bag 'A' is selected and A_2 be the event that bag B is selected.

$P(A_1) = P(A_2) = 1/2$

Let 'A' be the event that a white ball is drawn from the selected bag.

$\Rightarrow P(A/A_1) = 2/5, \; P(A/A_2) = \frac{4}{9}$

$P(A) = P(A_1) . P(A/A_1) + P(A_2) . P(A/A_2)$

$= \frac{1}{2}.\frac{2}{5} + \frac{1}{2}.\frac{4}{9} = \frac{1}{2}\left(\frac{2}{5} + \frac{4}{9}\right) = \frac{38}{90}$

Finally, $P(A_2/A) = \frac{P(A_2).P(A/A_2)}{P(A)} = \frac{(1/2).(4/9)}{(38/90)} = \frac{90.4}{18.38} = \frac{10}{19}$

Random Variables

We now introduce a new term….

Instead of saying that the possible outcomes are 1,2,3,4,5 or 6, we say that ***random variable*** X can take values {1,2,3,4,5,6}.

A random variable is an expression whose value is the outcome of a particular experiment.

The random variables can be either *discrete* or *continuous.*

It's a convention to use the upper-case letters (X, Y) for the names of the random variables and the lower case letters (x,y) for their possible particular values.

Random Variables

A **discrete** random variable is a random variable with a finite (or countably infinite) range.
A **continuous** random variable is a random variable with an interval (either finite or infinite) of real numbers for its range.

Random Variables

Discrete

$X = \{1, 2, 3, 4\}$

→ No of students

→ No of bits transmitted

Continuous

$X \in (1, 3)$

→ pressure

→ Temperature

→ Time

→ Voltage

Consider

$$S = \{HTT, THT, TTH, HHT, HTH, THH, TTT, HHH\}$$

X = R.V : number of heads

$X = x$, $x = 0, 1, 2, 3$

$X : S \longrightarrow \mathbb{R}$

S → Sample space

$\mathbb{R}$ → $\{0, 1, 2, 3\}$

$X = x$	0	1	2	3
$P(X = x)$	$1/8$	$3/8$	$3/8$	$1/8$

↓

probability distribution

$$X = \{0, 1, 2, 3\}$$

↓

Random Variable

Discrete

Random Variables

Discrete

$P(x)$

Validation :-

(1) $0 \leq P(x) \leq 1$

2) $\sum P(x) = 1$

Probability distribution function

Continuous

$f(x)$

1) $0 \leq f(x) \leq 1$

2) $\int f(x)\,dx = 1$

probability density function

Mathematical expectation, also known as the **expected value**, which is the summation of all possible values from a **random variable**. It is also known as the product of the probability of an event occurring, denoted by P(x), and the value corresponding with the observed occurrence of the event.

Expectation of a random Variable

$x_1 \to p_1$

$x_2 \to p_2$

$x_3 \to p_3$

$\vdots$

$x_n \to p_n$

then

$$x_1 p_1 + x_2 p_2 + x_3 p_3 + \cdots + x_n p_n$$

$$= \sum x_n p_n$$

$$\text{or } \sum x P(x)$$

is called "Mathematical Expectation" of a random variable x.

Mean of Random Variables

The expected value, or **mathematical expectation E(X) of a random variable X** is the long-run average value of X that would emerge after a very large number of observations. We often denote the expected value as µX , or µ if there is no confusion. **µX = E(X) is also referred to the mean of the random variable X, or the mean of the probability distribution of X**. In the case of a finite population, the expected value is the population mean.

Mathematical expectation

$$E(X) = \sum x\, p(x) \quad \text{if } X \text{ is discrete}$$

$$\int x\, f(x)\, dx \quad \text{if } x \text{ is conti}$$

μ

called as mean of a random variable

Consider a university with 15000 students and let X be the number of courses for which a randomly selected student is registered. The probability distribution of X is as follows:

x	1	2	3	4	5
No. of students	300	900	2850	4500	6450
$f(x)$	0.02	0.06	0.19	0.30	0.43

The average number of courses per student, or the average value of X in the population, results from computing the total number of courses taken by all students, and then dividing by the number of students in the population.

The mean, or average value of the random variable X, is therefore

$$\mu = \frac{1(300)+2(900)+3(2850)+4(4500)+5(6450)}{15000} = 4.06$$

Since

$$\frac{300}{15000} = 0.02 = f(1)$$
$$\frac{45900}{15000} = 0.06 = f(2),$$

and so on, an alternative expression for the mean is

$$\begin{aligned}\mu &= 1 \cdot f(1) + 2 \cdot f(2) + 3 \cdot f(3) + 4 \cdot f(4) + 5 \cdot f(5) \\ &= 1(0.02) + 2(0.06) + 3(0.19) + 4(0.30) + 5(0.43) \\ &= 0.02 + 0.12 + 0.57 + 1.20 + 2.15 \\ &= 4.06\end{aligned}$$

Variance of Random Variables

$$\boxed{Var(x) = E(x-\mu)^2}$$

σ^2

$$= E(x^2 - 2\mu x + \mu^2)$$

$$= E(x^2) - 2\mu E(x) + \mu^2$$

$E(x) \downarrow \mu$

$$= E(x^2) - 2\mu^2 + \mu^2$$

$$= E(x^2) - \mu^2$$

$$= E(x^2) - [E(x)]^2$$

$$E(x^2) - \underbrace{[E(x)]^2}_{\text{mean}}$$

$\sum x^2 p(x)$ discrete

$\sum x^2 f(x) dx \rightarrow$ continuous

Probability Distributions

Probability Distributions

- Discrete
  - Binomial

    $$P(x) = {}^{n}C_{x}\, p^{x} q^{n-x}$$

    $x = 0, 1, 2, \ldots n$
  - Poisson

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

    $x = 0, 1, 2 \ldots$
- Continuous
  - Normal or Gaussian

Binomial Distributions

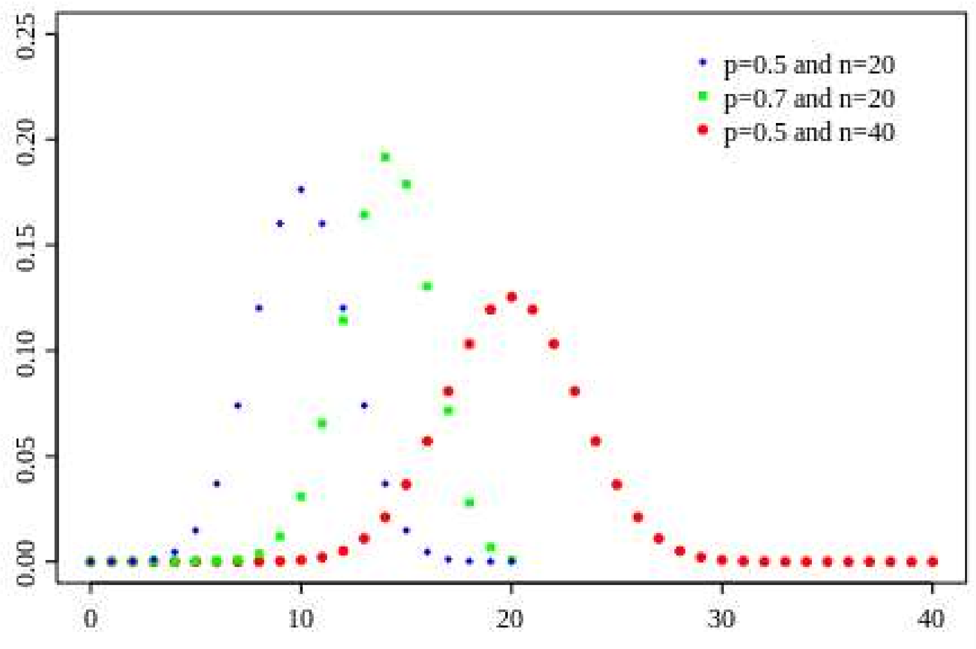
A **binomial distribution** can be thought of as simply the probability of a SUCCESS or FAILURE outcome in an experiment or survey that is repeated multiple times. The binomial is a type of distribution that has two possible outcomes.

We have a binomial experiment if ALL of the following four conditions are satisfied:

- The experiment consists of n identical trials.
- Each trial results in one of the two outcomes, called success and failure.
- The probability of success, denoted p, remains the same from trial to trial.
- The n trials are independent.

Binomial Distribution Formula

$$P(X) = {}_{n}C_{x}\, p^{x}(1-p)^{n-x}$$



https://www.mathsisfun.com/data/binomial-distribution.html

Total 'n' trials for ex: 10

4 ← 'x' are successful with prob. p

$(p.p.p.p - - - x \text{ times})\ (q.q.q - - - n-x \text{ times})$

4 $\qquad$ 10 − 4

$p^x \ q^{n-x}$

$\downarrow \quad {}^nC_x$

ie

$${}^nC_x \, p^x q^{n-x}$$

where $x = 0, 1, 2, \ldots n$

Poisson Distributions

A **Poisson Process** is a model for a series of discrete event where the *average time* between events is known, but the exact timing of events is random. The arrival of an event is independent of the event before (waiting time between events is memoryless).

Poisson Distribution a discrete probability distribution that expresses the probability of a given number of events occurring in a fixed interval of time or space if these events occur with a known constant mean rate and independently of the time since the last event

Poisson Distribution Formula

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

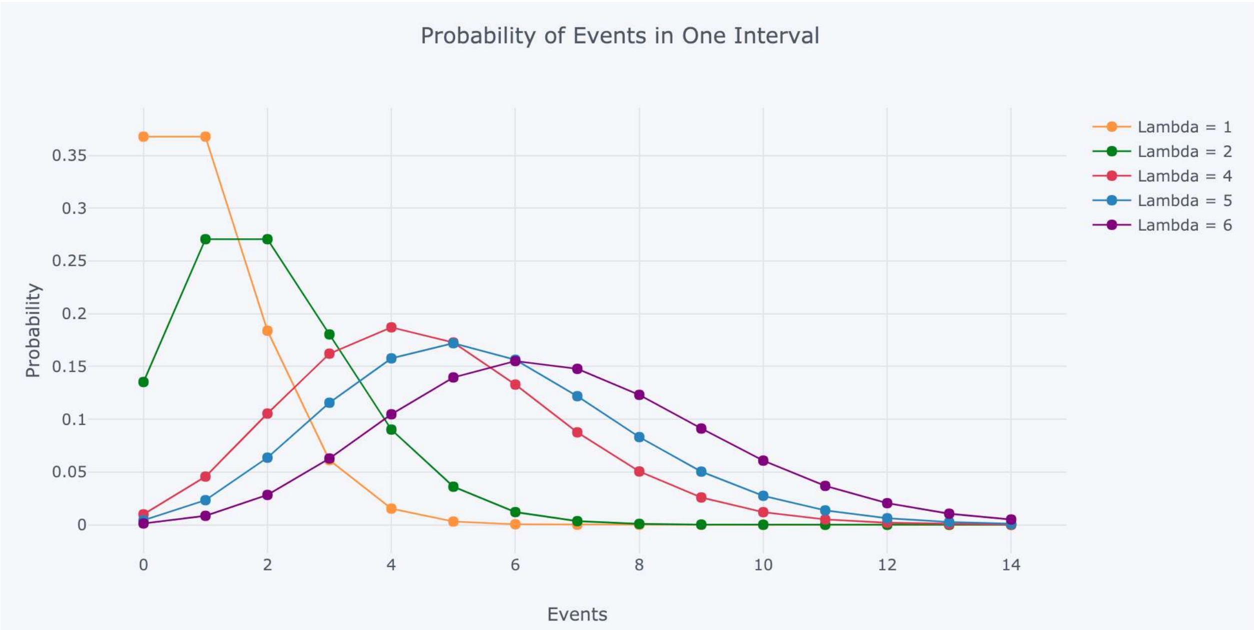
where

x = 0, 1, 2, 3, …

λ = mean number of occurrences in the interval

e = Euler's constant $\approx$ 2.71828

1. Events are **independent** of each other. The occurrence of one event does not affect the probability another event will occur.
2. The average rate (events per time period) is constant.
3. Two events cannot occur at the same time.



Normal Distributions

Normal distribution, also known as the **Gaussian distribution**, is a probability **distribution** that is symmetric about the mean, showing that data near the mean are more frequent in occurrence than data far from the mean

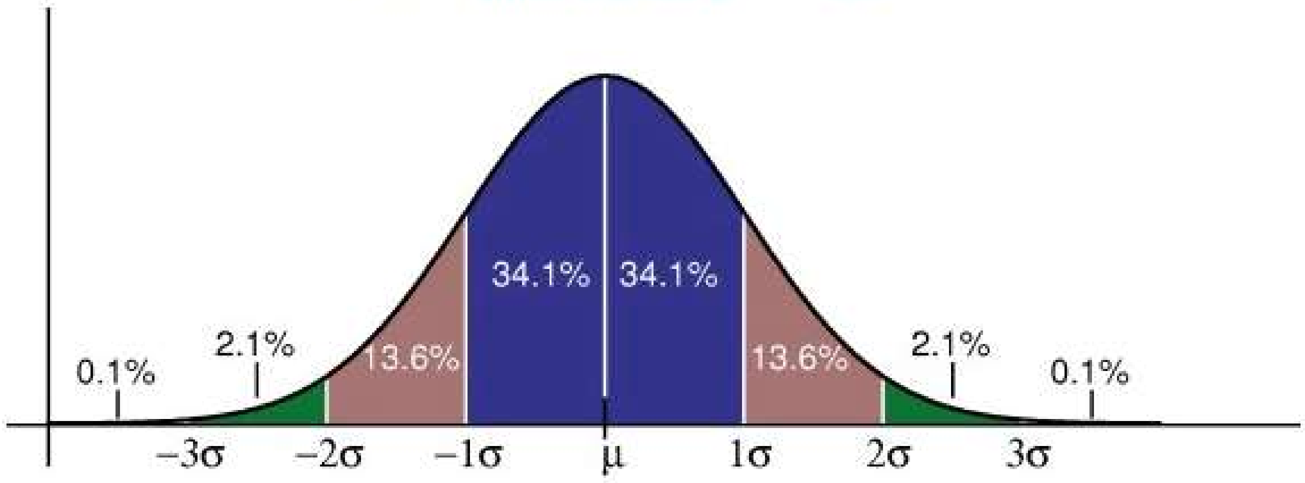
A normal distribution has some interesting properties: it has a bell shape, the mean and median are equal, and 68% of the data falls within 1 standard deviation.

Many groups follow this type of pattern. That’s why it’s widely used in business, statistics and in government bodies like the FDA:

- Heights of people.
- Measurement errors.
- Blood pressure.
- Points on a test.
- IQ scores.
- Salaries

The empirical rule tells you what percentage of your data falls within a certain number of standard deviations from the mean:

- 68% of the data falls within one standard deviation of the mean.
- 95% of the data falls within two standard deviations of the mean.
- 99.7% of the data falls within three standard deviations of the mean.



Thanks