Probability Calculus

- **Random experiment :** An experiment that can be repeated any number of times under the same set of conditions, and its outcome is known only after the completion of the experiment.
- Eg: tossing of a coin or the rolling of a die.
- A possible outcome of a random experiment is called a simple event (or elementary event) and denoted by ω_i ie, .
- The set of all possible outcomes, $\{\omega 1, \omega 2, \ldots, \omega k\}$, is called the sample space and is denoted as Ω
- , i.e. $\Omega = \{\omega 1, \omega 2, \ldots, \omega k\}$,
- Subsets of Ω are called events and are denoted by capital letters such as A, B,C.
- The set of all simple events that are contained in the event A is denoted $\Omega_{A.}$

- The event $\bar{A}$ refers to the non-occurring of A and is called a composite or complementary event.
- Ω is an event.
- Since it contains all possible outcomes, we say that will always occur and we call it a sure event or certain event.
- On the other hand, if we consider the null set $\emptyset = \{\}$ as an event, then this event can never occur and we call it an impossible event.

Example 6.1.1 (Rolling a die) If a die is rolled once, then the possible outcomes are the number of dots on the upper surface: $1, 2, \ldots, 6$. Therefore, the sample space is the set of simple events $\omega_1 =$ "1", $\omega_2 =$ "2", ..., $\omega_6 =$ "6" and $\Omega = \{\omega_1, \omega_2, \ldots, \omega_6\}$. Any subset of Ω can be used to define an event. For example, an event A may be "an even number of dots on the upper surface of the die". There are three possibilities that this event occurs: ω_2, ω_4, or ω_6. If an odd number shows up, then the composite event $\bar{A}$ occurs instead of A. If an event is defined to observe only one particular number, say $\omega_1 =$ "1", then it is an elementary event. An example of a sure event is "a number which is greater than or equal to 1" because any number between 1 and 6 is greater than or equal to 1. An impossible event is "the number is 7".

Probability

- **Probability** is a numerical measure of how likely an event is to occur.
- For a random experiment, the probability of an event A is defined as
- $P(A) = \frac{\textit{Number of favorable outcomes to } A}{\textit{Total number of possible outcomes}}$

  provided all outcomes are equally likely
- Probability values lie between 0 and 1:
- $P(A)$=0→ the event is impossible
- $P(A)$=1→ the event is certain.
- The sum of probabilities of all possible outcomes is 1.

- If we repeat an experiment many times, the fraction of times an event occurs gives an estimate of its probability.
- ie, Out of 100 tosses, heads occurred 48 times → probability ≈ 0.48
- Mathematically, if we repeat the experiment more and more times, this relative frequency stabilizes.
  That stable value is defined as the probability.

$$P(A) = \lim_{n \to \infty} \frac{n(A)}{n}$$

- where $n(A)$ denotes the number of times an event A occurs out of n times.

Conditional Probability

- Suppose a new medical test is developed to diagnose a particular infection of the blood.
- The test is conducted on blood samples from 100 randomly selected patients.

Table 6.1 Absolute frequencies of test results and infection status

		Infection		Total (row)
		Present	Absent	
Test	Positive (+)	30	10	40
	Negative (–)	15	45	60
	Total (column)	45	55	Total = 100

- There are the following four possible outcomes:
  - The blood sample has an infection and the test diagnoses it, i.e. the test is correctly diagnosing the infection.
  - The blood sample does not have an infection and the test does not diagnose it, i.e. the test is correctly diagnosing that there is no infection.
  - The blood sample has an infection and the test does not diagnose it, i.e. the test is incorrect in stating that there is no infection.
  - The blood sample does not have an infection but the test diagnoses it, i.e. the test is incorrect in stating that there is an infection.

Table 6.2 Relative frequencies of patients and test

		Infection		Total (row)
		Present (IP)	Absent (IA)	
Test	Positive (+)	0.30	0.10	0.40
	Negative (–)	0.15	0.45	0.60
	Total (column)	0.45	0.55	Total = 1

- It can be seen that the probability that a test is positive is $P(T+) = 0.30 + 0.10 = 0.40$ and the probability that an infection is present is $P(I\,P) = 0.30 + 0.15 = 0.45$.
- If one already knows that the test is positive and wants to determine the probability that the infection is indeed present, then this can be achieved by the respective **conditional probability** $P(I\,P|T+)$ which is

$$P(I\,P|T+) = \frac{P(I\,P \cap T+)}{P(T+)} = \frac{0.3}{0.4} = 0.75.$$

- Assume that we have prior information that *A* has already occurred. Now we want to calculate the probability of B.

$$P(B|A) = \frac{n_{AB}/k}{n_A/k} = \frac{P(A \cap B)}{P(A)}.$$

- Let *P*(*A*) > 0. Then the **conditional probability** of event *B* occurring, given that event *A* has already occurred, is

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

- The roles of *A* and *B* can be interchanged to define *P*(*A*|*B*) as follows. Let *P*(*B*) > 0. The conditional probability of *A* given *B* is

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

Tutorials

Exercise 4.1 A newspaper asks two of its staff to review the coffee quality at different trendy cafés. The coffee can be rated on a scale from 1 (miserable) to 10 (excellent). The results of the two coffee enthusiasts X and Y are as follows:

Café i	x_i	y_i
1	3	6
2	8	7
3	7	10
4	9	8
5	5	4

(a) Calculate and interpret Spearman's rank correlation coefficient.

(b) Does Spearman's R differ depending on whether ranks are assigned in a decreasing or increasing order?

Exercise 4.2 A total of 150 customers of a petrol station are asked about their satisfaction with their car and motorbike insurance. The results are summarized below:

	Satisfied	Unsatisfied	Total
Car	33	25	58
Car (diesel engine)	29	31	60
Motorbike	12	20	32
Total	74	76	150

(a) Determine and interpret Pearson's χ^2 statistic, Cramer's V, and C_{corr}.
(b) Combine the categories "car" and "car (diesel engine)" and produce the corresponding 2×2 table. Calculate χ^2 as efficiently as possible and give a meaningful interpretation of the odds ratio.
(c) Compare the results from (a) and (b).

Random Variable

Definition 7.1.1 Let Ω represent the sample space of a random experiment, and let $\mathcal{R}$ be the set of real numbers. A random variable is a function X which assigns to each element $\omega \in \Omega$ one and only one number $X(\omega) = x$, $x \in \mathcal{R}$, i.e.

$$X : \Omega \to \mathcal{R}. \tag{7.1}$$

- A random variable is just a rule that takes the outcome of an experiment and turns it into a number.

Sample space:

$$\Omega = \{H, T\}$$

- Eg: Coin toss

Define a random variable X as:

- X(H) = 1
- X(T) = 0

So:

- Outcome = Head → Number = 1
- Outcome = Tail → Number = 0

- X converts non-numerical outcomes into numbers.

- Example 2: Throwing a die

Sample space:

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

Define a random variable X as:

$$X(\omega) = \text{number shown on the die}$$

So:

- If $\omega = 4$, then $X(\omega) = 4$
- If $\omega = 6$, then $X(\omega) = 6$

Cumulative Distribution Function (CDF)

Definition 7.2.1 The **cumulative distribution function (CDF)** of a random variable X is defined as

$$F(x) = P(X \leq x). \tag{7.2}$$

ie, The CDF of a random variable X tells us the probability that X is less than or equal to a given value.

Let:

$$X = \text{number on the die}$$

Then:

- $F(3) = P(X \leq 3) = P(1, 2, 3) = \frac{3}{6} = 0.5$
- $F(6) = P(X \leq 6) = 1$

- $F(x)$ is a monotonically non-decreasing function (if $x_1 \leq x_2$, it follows that $F(x_1) \leq F(x_2)$),
- $\lim_{x \to -\infty} F(x) = 0$ (the lower limit of F is 0),
- $\lim_{x \to +\infty} F(x) = 1$ (the upper limit of F is 1),
- $F(x)$ is continuous from the right, and
- $0 \leq F(x) \leq 1$ for all $x \in \mathcal{R}$.

ie, CDF always:

- starts at 0
- ends at 1
- never decreases

Probability Mass Function (PMF) — *Discrete case*

- When do we use **PMF**?
- When the random variable takes **countable values** (e.g., 0,1,2,…)

Definition 7.2.4 Let X be a discrete random variable which takes k different values. The **probability mass function** (PMF) of X is given by

$$f(X) = P(X = x_i) = p_i \quad \text{for each} \quad i = 1, 2, \ldots, k. \tag{7.5}$$

It is required that the probabilities p_i satisfy the following conditions:

(1) $0 \leq p_i \leq 1$,
(2) $\sum_{i=1}^{k} p_i = 1$.

Relationship between PDF and CDF

- Example :Die ,

$$p(x) = \frac{1}{6}, \quad x = 1, 2, 3, 4, 5, 6$$

- Ie,PMF tells us What is P(X = 4)? Whre as CDF tells us What is P(X ≤ 4)?

$$F(x) = \sum_{t \leq x} p(t)$$

- CDF is the **sum of PMF values** up to x.

CDF of a Continuous random variable

Definition 7.2.2 A random variable X is said to be **continuous** if there is a function $f(x)$ such that for all $x \in \mathcal{R}$

$$F(x) = \int_{-\infty}^{x} f(t)\mathrm{d}t \tag{7.3}$$

holds. $F(x)$ is the cumulative distribution function (CDF) of X, and $f(x)$ is the probability density function (PDF) of x and $\frac{\mathrm{d}}{\mathrm{d}x}F(x) = f(x)$ for all x that are continuity points of f.

Theorem 7.2.1 *For a function $f(x)$ to be a* ***probability density function (PDF)*** *of X, it needs to satisfy the following conditions:*

(1) $f(x) \geq 0$ for all $x \in \mathcal{R}$,
(2) $\int_{-\infty}^{\infty} f(x)\mathrm{d}x = 1$.

Theorem 7.2.2 *Let X be a random variable with CDF $F(x)$. If $x_1 < x_2$, where x_1 and x_2 are known constants, $P(x_1 \leq X \leq x_2) = F(x_2) - F(x_1) = \int_{x_1}^{x_2} f(x)\mathrm{d}x$.*

Theorem 7.2.3 *The probability of a continuous random variable taking a particular value x_0 is zero:*

$$P(X = x_0) = 0. \tag{7.4}$$

Probability Density Function

- A PDF does NOT give probability at a point.
  It gives probability density, i.e., how concentrated the probability is around a point.
- Definition:

A function $f(x)$ is a PDF if:

1. $f(x) \geq 0$
2. $\int_{-\infty}^{\infty} f(x)\, dx = 1$
3. $P(a \leq X \leq b) = \int_a^b f(x)\, dx$

- **Example: Height of people**
  - Let X = height (continuous).
  - Probability that height is **exactly 170 cm** = 0
  - Probability that height is between **169 and 171 cm** > 0
- PDF tells us **how likely heights are around 170 cm**, not exactly at 170.
- Relationship between PDF and CDF.

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\, dt$$

$$f(x) = \frac{d}{dx} F(x)$$

- CDF is fundamental, PDF is the derivative of the CDF (when it exists)

Probability Distributions – Standard Discrete Distributions

1.Discrete Uniform Distribution

- The discrete uniform distribution assumes that all possible outcomes have equal probability of occurrence.
- **Definition 8.1.1** A discrete random variable X with k possible outcomes x1, x2, . . . ,xk is said to follow a discrete **uniform** distribution if the probability mass function (PMF) of X is given by

$$P(X = x_i) = \frac{1}{k}, \quad \forall i = 1, 2, \ldots, k.$$

- If the outcomes are the natural numbers xi = i (i = 1, 2, . . . , k), the mean and variance of X are obtained as

$$\mathrm{E}(X) = \frac{k+1}{2},$$
$$\mathrm{Var}(X) = \frac{1}{12}(k^2 - 1).$$

Example 8.1.1 If we roll a fair die, the outcomes "1", "2", ..., "6" have equal probability of occurring, and hence, the random variable X "number of dots observed on the upper surface of the die" has a uniform discrete distribution with PMF

$$P(X = i) = \frac{1}{6}, \quad \text{for all} \quad i = 1, 2, \ldots, 6.$$

The mean and variance of X are

$$\mathrm{E}(X) = \frac{6+1}{2} = 3.5,$$

$$\mathrm{Var}(X) = \frac{1}{12}(6^2 - 1) = 35/12.$$

2.Degenerate Distribution

- **Definition 8.1.2** A random variable X has a **degenerate distribution** at a, if a is the only possible outcome with P(X = a) = 1. The CDF in such a case is given by

$$F(x) = \begin{cases} 0 & \text{if } x < a \\ 1 & \text{if } x \geq a. \end{cases}$$

- Further, E(X) = a and Var(X) = 0.
- The degenerate distribution indicates that there is only one possible fixed outcome, and therefore, no randomness is involved.
- It follows that we need at least two different possible outcomes to have randomness in the observations of a random variable or random experiment.
- The Bernoulli distribution is such a distribution where there are only two outcomes, e.g. success and failure or male and female. These outcomes are usually denoted by the values “0” and “1”.

3.Bernoulli Distribution

- **Definition 8.1.3** A random variable X has a Bernoulli distribution if the PMF of X is given as

$$P(X = x) = \begin{cases} p & \text{if } x = 1 \\ 1 - p & \text{if } x = 0. \end{cases}$$

- The cumulative distribution function (CDF) of X is

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 - p & \text{if } 0 \le x < 1 \\ 1 & \text{if } x \ge 1. \end{cases}$$

- The mean (expectation) and variance of a Bernoulli random variable are calculated as
- The mean (expectation) and variance of a Bernoulli random variable are calculated as

$$\mathrm{E}(X) = 1 \cdot p + 0 \cdot (1 - p) = p \tag{8.4}$$

and

$$\mathrm{Var}(X) = (1 - p)^2 \cdot p + (0 - p)^2 \cdot (1 - p) = p(1 - p), \tag{8.5}$$

respectively.

- A Bernoulli distribution is useful when there are only two possible outcomes, and our interest lies in any of the two outcomes, e.g. whether a customer buys a certain product or not, or whether a hurricane hits an island or not.

- The outcome of an event A is usually coded as 1 which occurs with probability p.
- If the event of interest does not occur, i.e. the complementary event A’ occurs, the outcome is coded as 0 which occurs with probability 1 – p.
- So p is the probability that the event of interest A occurs.

Example 8.1.2 A company organizes a raffle at an end-of-year function. There are 300 lottery tickets in total, and 50 of them are marked as winning tickets. The event A of interest is "ticket wins" (coded as $X = 1$), and the probability p of having a winning ticket is *a priori* (i.e. before any lottery ticket has been drawn)

$$P(X = 1) = \frac{50}{300} = \frac{1}{6} = p \quad \text{and} \quad P(X = 0) = \frac{250}{300} = \frac{5}{6} = 1 - p.$$

According to (8.4) and (8.5), the mean (expectation) and variance of X are

$$E(X) = \frac{1}{6} \quad \text{and} \quad \text{Var}(X) = \frac{1}{6} \cdot \frac{5}{6} = \frac{5}{36} \text{ respectively.}$$

4. Binomial Distribution

- Consider n independent trials or repetitions of a Bernoulli experiment. In each trial or repetition, we may observe either A or A'. At the end of the experiment, we have thus observed A between 0 and n times.
- Suppose we are interested in the probability of A occurring k times, then the binomial distribution is useful.
- **Definition 8.1.4** A discrete random variable X is said to follow a binomial distribution with parameters n and p if its PMF is given by

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k} \quad (k = 0, 1, \ldots, n).$$

- We also write X ~ B(n; p).
- The mean and variance of a binomial random variable X are given by

$$\mathrm{E}(X) = np,$$
$$\mathrm{Var}(X) = np(1-p).$$

- A Bernoulli random variable is therefore B(1; p) distributed.

Example 8.1.4 Consider an unfair coin where the probability of observing a tail (T) is $p(T) = 0.6$. Let us denote tails by "1" and heads by "0". Suppose the coin is tossed three times. In total, there are the $2^3 = 8$ following possible outcomes:

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

Note that the first outcome, viz. (1, 1, 1) leads to $x = 3$, the next 3 outcomes, viz., (1, 1, 0), (1, 0, 1), (0, 1, 1) obtained by $(= \binom{3}{2})$ lead to $x = 2$, the next 3 outcomes, viz., (1, 0, 0), ((0, 1, 0), (0, 0, 1) obtained by $(= \binom{3}{1})$ lead to $x = 1$, and the last outcome, viz. (0, 0, 0) obtained by $(= \binom{3}{0})$ leads to $x = 0$. We can, for example, calculate

$$P(X=2) = \binom{3}{2} 0.6^2 (1-0.6)^1 = 0.432 \quad (\text{or} \quad 43.2\,\%).$$

Further, the mean and variance of X are

$$\mathrm{E}(X) = np = 3 \cdot 0.6 = 1.8, \quad \text{and} \quad \mathrm{Var}(X) = np(1-p) = 3 \cdot 0.6 \cdot 0.4 = 0.72.$$

5. Poisson Distribution

- **Definition 8.1.5** A discrete random variable X is said to follows Poisson distribution with parameter λ > 0 if its PMF is given by

$$P(X = x) = \frac{\lambda^x}{x!} \exp(-\lambda) \quad (x \lambda= 0, 1, 2, \ldots).$$

We also write $X \sim Po(\lambda)$. The mean and variance of a Poisson random variable are identical:

$$\mathrm{E}(X) = \mathrm{Var}(X) = \lambda.$$

Example 8.1.6 Suppose a country experiences $X = 4$ tropical storms on average per year. Then the probability of suffering from only two tropical storms is obtained by using the Poisson distribution as

$$P(X = 2) = \frac{\lambda^x}{x!}\exp(-\lambda) = \frac{4^2}{2!}\exp(-4) = 0.146525.$$

Standard Continuous Distributions

1. Continuous Uniform Distributions

Definition 8.2.1 A continuous random variable X is said to follow a (continuous) **uniform distribution** in the interval $[a, b]$, i.e. $X \sim U(a, b)$, if its probability density function (PDF) is given by

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{if} \quad a \leq x \leq b \quad (a < b) \\ 0 & \text{otherwise.} \end{cases}$$

The mean and variance of $X \sim U(a, b)$ are

$$\mathrm{E}(X) = \frac{a+b}{2} \quad \text{and} \quad \mathrm{Var}(X) = \frac{(b-a)^2}{12},$$

respectively.

2. Normal Distribution

Definition 8.2.2 A random variable X is said to follow a **normal distribution** with parameters μ and σ^2 if its PDF is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right); \qquad -\infty < x < \infty, -\infty < \mu < \infty, \sigma^2 > 0. \tag{8.15}$$

We write $X \sim N(\mu, \sigma^2)$. The mean and variance of X are

$$\mathrm{E}(X) = \mu; \quad \text{and} \quad \mathrm{Var}(X) = \sigma^2,$$

respectively. If $\mu = 0$ and $\sigma^2 = 1$, then X is said to follow a ***standard* normal distribution**, $X \sim N(0, 1)$. The PDF of a standard normal distribution is given by

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2}); \quad -\infty < x < \infty.$$

Exponential Distribution

Definition 8.2.3 A random variable X is said to follow an exponential distribution with parameter $\lambda > 0$ if its PDF is given by

$$f(x) = \begin{cases} \lambda \exp(-\lambda x) & \text{if } \; x \geq 0 \\ 0 & \text{otherwise.} \end{cases} \tag{8.23}$$

We write $X \sim Exp(\lambda)$. The mean and variance of an exponentially distributed random variable X are

$$\mathrm{E}(X) = \frac{1}{\lambda} \quad \text{and} \quad \mathrm{Var}(X) = \frac{1}{\lambda^2},$$

respectively. The CDF of the exponential distribution is given as

$$F(x) = \begin{cases} 1 - \exp(-\lambda x) & \text{if } \; x \geq 0 \\ 0 & \text{otherwise.} \end{cases} \tag{8.24}$$

Note, that $P(X > x) = 1 - F(x) = \exp(-\lambda x)$ $(x \geq 0)$. An interesting property of the exponential distribution is its **memorylessness**: if time t has already been reached, the probability of reaching a time greater than $t + \Delta$ does not depend on t. This can be written as

$$P(X > t + \Delta | X > t) = P(X > \Delta) \quad t, \Delta > 0.$$

Point Estimation