Imperial College London

Mathematics: Year 2

Statistical Probability

Xin Wang

September 2, 2020

Abstract

Engineering is a profession founded upon experimentation, the data it produces ,and how it is extracted and analysed.

Probability is the mathematical science of dealing with uncertainty - it provides engineers with the rules for analyzing and understanding ignorance about uncertain situations.

Statistics is the science of reasoning about data in the presence of uncertainty that can arise for numerous reason in the real world such as noise. Practically, data is composed of signal and noise.

Contents

1 Foundational statistics

There are two types of statistics: Descriptive and Inferential statistics.

1.1 Experiments and Sampling

Definitions:

- **Population**: The object of study from which a sample is taken.
- **Sample**: A set of data extracted from a population for investigation. The data must be **representative** and observations must be **independent**.

1.2 Plots

There are a variety of plots that each have its advantages and disadvantages such as:

- Histogram
- Steam-and-Leaf
- Line

Each plots can have a **linear scale** or **non-linear scale**.

1.3 Probability

Methods of inference in inferential statistics rely on probability theory and are strongly related to histograms of data that allows an "idealized" proportional to be defined.

1.3.1 Sample space and Events

Terminology used is derived from Set Theory.

- **Sample space** (S): Set of all possible outcomes of an experiment.
- **Discrete sample space**: Possible outcomes can be written as a finite list.
- **Continuous sample space**: Possible outcomes cannot be written as a finite list.
- **Event**: Any subset of S.

The usual operations of Set Theory are defined as:

- **Union** ($A \cup B$): A or B occurs.
- **Intersection** ($A \cap B$): A and B occurs.
- **Complement** ($S - A$): not-A occurs.

1.3.2 Axioms of probablity

The three rules of probability are the foundation of probability theory:

1.

2 Events, Probability and Sets

2.1 Sets

Set: A collection of unordered, distinct objects (**elements**).

Cardinality: A measure of the number of elements of a set.

Singleton: A set with a single element.

Notations:

- If an element denoted as ω is part of set A, formally defined as "ω belongs to A":
$$\omega \in A$$

- Contents of sets are enclosed in curly braces ...
$$A = \{\omega, 1, A\}$$

A set can contain any mixture of different objects e.g. a set containing integer, letter and other sets. Practically, the type of elements a set can have is usually restricted to a **collection** of possible values. This collection **is a set itself** called the **universal set** (Ω).

Empty set ($\phi = \{\}$): A unique set that contains no elements.

Set relation notation:

- $\subseteq$: A is a **subset** of B - all elements of A are members of B with the possibility that set A is equal (contain the same elements) as set B i.e. $A \subseteq B$

- $\subset$: A is a **proper subset** of B - all elements of A are members of B and set A must be smaller than set B

- $=$: A and B have the identical elements i.e. $A \subseteq B$ and $B \subseteq A$ thus $A = B$

2.2 Set Operations

2.2.1 Union

Union (of sets A and B - $A \cup B$): The set containing all elements in A alone, all elements in B alone **and** all elements shared by both A and B.

$$A \cup B = \{\omega \in \Omega \ : \ \omega \in A \textbf{ or } \omega \in B\}$$

Properties of union:

- Identity law: $A \cup \phi = A$
- Idempotent law: $A \cup A = A$
- Domination law: $A \cup \Omega = \Omega$
- Commutative law: $A \cup B = B \cup A$

2.2.2 Intersection

Intersection (of sets A and B - $A \cap B$): The set containing all elements that belong to both A and B.

$$A \cap B = \{\omega \in \Omega \ : \ \omega \in A \textbf{ and } \omega \in B\}$$

Properties of union:

- Identity law: $A \cap \phi = \phi$
- Idempotent law: $A \cap A = A$
- Domination law: $A \cap \Omega = A$
- Commutative law: $A \cap B = B \cap A$

The combination of Union and Intersection results in the Distributive Law:

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$
$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

Disjoint: If sets A and B share no elements:

$$A \cap B = \phi$$

Partition of Ω: If a collection of sets $A_1, A_2, \ldots, A_n$ are disjoint and the union of the sets equals the **universal set**.

$$\bigcup_{i=1}^{n} A_i = \Omega$$

2.2.3 Complements and Differences

Complements (of a set): The set that contains all elements of Ω that do not belong to A - not A.

$$\overline{A}$$

Properties of union:

- $\overline{(\overline{A})} = A$
- $\overline{\phi} = \Omega$
- $A \cup \overline{A} = \Omega$
- $A \cap \overline{A} = \phi$

Difference (of sets A and B): The set that contains elements of set A that do not belong to set B.

$$A/B = \{\omega \in \Omega \ : \ \omega \in A \text{ and } \omega \notin B\}$$

The two sets A and B are disjoint sets if $A/B = A$. The difference of two sets can be interpreted as:

$$A/B = A \cap \overline{B}$$

and

$$\overline{A} = \frac{\Omega}{A}$$

2.2.4 De Morgan's laws

De Morgan's laws allows the combination of complements, unions and intersections.

$$\overline{(A \cup B)} = \overline{A} \cap \overline{B}$$
$$\overline{(A \cap B)} = \overline{A} \cup \overline{B}$$

2.3 Sample spaces and Events

Consider a **random experiment** where the **outcome** is unknown but all the possible outcomes are described by the **sample space** - a non-empty set S.

Example 1: Find the sample space of a toss of a coin.

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

Often, defining whether a sample space is **discrete** or **continuous** is important and affects the answer.

Example 2: Find the same space of the thickness of a molded plastic container.

- Discrete:
$$S = \{low,\, medium,\, high\}$$

- Continuous:
$$S = \{x : 20 < x < 30\}$$

Events: Subsets of the possible outcomes (sample space).

Example 1: List the possible events that could occur from a coin toss:

With sample space:

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

Possible events are:

$$E = \{H\} \text{ and } E = \{T\}$$

In the context of random events, there are several special events:

- Null event: An empty set - ϕ (Never occurs)
- Elementary event (of S): An event that is a **singleton subset** of S
- Universal event: The sample space S that consists of the union of all elementary events

2.4 Probability axioms

Axioms - A statement/proposition that is regarded as being established.

In order to be able to characterize the uncertainty of an event, a set function P (*Probability function* or *Probability measure*) is define. The set function P takes a set as argument and returns a value.

In order for a set function to be a probability for any event defined $E \subseteq S$, it has to satisfy three conditions:

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

2. $P(S) = 1$

3. If $E \cap F = \phi$ then $P(E \cup F) = P(E) + P(F)$: If two event are disjoint, the probability of either event occurring is the sum of probability of each specified event occurring.

Given disjoint subsets $E_1, E_2, \ldots, E_i \in S$, the probability of either event occurring is the sum of probability of each specified event occurring:

$$P\left(\bigcup_i E_i\right) = \sum_i P(E_i)$$

Understanding the three axioms allows several important relationships to be established:

1. Probability of an event occurring and the probability of the event not occurring must add up to 1:

$$P(E) + P(\overline{E}) = P(S) = 1$$
$$P(E) = 1 - P(\overline{E})$$

   Sometimes it is hard to compute $P(E)$ and easier to rather compute $P(\overline{E})$.

2. Boole's relationship: Union of arbitrary events E and F:

$$E \cup F = E \cup \left(\overline{E} \cap F\right) \Rightarrow P(E \cup F) = P(E) + P(\overline{E} \cap F)$$

   and

$$F = (E \cap F) \cup (\overline{E} \cap F) \Rightarrow P(F) = P(E \cap F) + P(\overline{E} \cap F)$$

   combine to form Boole's relationship:

$$P(E \cup F) = P(E) + P(F) - P(E \cap F)$$

3. Probability of union E and F is less than or equal to sum of individual probabilities of event E and F:

   Since $P(E \cap F) \geq 0$:

$$P(E \cup F) \leq P(E) + P(F)$$

4. The highest possibility between two events will always be less than probability of either event occurring which is less than probability sum of two events occurring.

$$max[P(E), P(F)] \leq P(E \cup F) \leq P(E) + P(F)$$

5. Probability of either events E and F not occurring is less than the probability of both events occurring which is less than the lowest probability between two events.

$$P(E) + P(F) - 1 \leq P(E \cap F) \leq min[P(E), P(F)]$$

6. Poincare's formula:

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

Example 1: Consider events A and B and the given probabilities:

- $P(A) = P(B) = 0.90$
- $P(A \cap B) = 0.85$

Find $P(A \cup B)$:

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) - P(A \cap B) \\ &= 2 \times 0.9 - 0.85 \\ &= 0.95 \end{aligned}$$

Find $P(\overline{A} \cap B)$:

$$\begin{aligned} P(\overline{A} \cap B) &= P(B) - P(A \cap B) \\ &= 0.05 \end{aligned}$$

2.5 Probability computation

Probability: Considering a sample space S with n number of **equally possible** events, the probability of a event defined as $E \subseteq S$:

$$P(E) = \frac{\text{events in } E}{n}$$

Example 1: Given a 6-faced die with sample space $S = \{1, 2, 3, 4, 5, 6\}$, find the probability of rolling an even number.

1. Find n:

$$P(S) = \frac{1}{6}$$

2. Probability of an even number:

$$P(E) = P(2 \cup 4 \cup 6) = \frac{1}{2}$$

Addition rule of probability:

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

Example 2: Consider drawing a random card from a deck of cards:

- A defined as the event of drawing a heart with probability $P(A) = \frac{13}{52} = \frac{1}{4}$
- B defined as the even of drawing cards $\{J, Q, K\}$ with probability $P(B) = \frac{12}{52} = \frac{3}{13}$
- Since A and B are not disjoint as some face cards are also hearts, $P(A \cap B) = \frac{3}{52}$

The probability of A or B is defined as:

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) - P(A \cap B) \\ &= \frac{1}{4} + \frac{3}{13} + -\frac{3}{52} \\ &= \frac{11}{26} \end{aligned}$$

2.6 Conditional Probability

Conditional probability is the concept of describing the probability of an event **given that another event has already occurred -** this is important as usually events are influenced by other factors.

The common notation defined as:

"the probability that event A occurs given that B has already occurred"

$$P(A|B)$$

For events A and B where $P(B) > 0$, the **conditional probability** $P(A|B)$ is defined as:

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

Due to the conditioning event, the conditional probability considers a reduced subset of the sample space.

- If $B \subseteq A$:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1 \geq P(A)$$

  If B is a subset of A, the probability of an event in B as well in A $[P(A \cap B)]$ occurring is basically the probability of B $[P(B)]$. This means that when finding the probability of A occurring given that B has occurred already and B is a subset of A, the event A has already occurred.

- If $A \cap B = \phi$

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0}{P(B)} = 0 \leq P(A)$$

If there are no common events in A and B, set B does not affect the probabilities of events in A occurring.

Multiplication law of probability:

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

It follows that:

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

The result shows the relationship between conditional probabilities and unconditional probabilities.

Example 1: When throwing a dice, let $A = \{\text{even number}\}$ and $B = \{\text{score} \geq 3\}$:

- $P(A) = \frac{1}{2}$
- $P(B) = \frac{2}{3}$
- $P(A \cap B) = \frac{1}{3}$

Then the difference between $P(A|B)$ and $P(B|A)$ are as follows:

- $P(A|B)$:

$$\begin{aligned} P(A|B) &= \frac{P(A \cap B)}{P(B)} \\ &= \frac{\frac{1}{3}}{\frac{2}{3}} \\ &= \frac{1}{2} \end{aligned}$$

- $P(B|A)$:

$$\begin{aligned} P(B|A) &= \frac{P(B \cap A)}{P(A)} \\ &= \frac{\frac{1}{3}}{\frac{1}{2}} \\ &= \frac{2}{3} \end{aligned}$$

2.6.1 Independence

Given the concept of conditional probability, the concept of **independence** can be defined.

Independence: The occurrence of one event does not change the probability of the occurrence of another event.

Events A and B are said to be independent if:

$$P(A|B) = P(A)$$

For independent events, the multiplication rule results:

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

This results shows that independent events having multiplicative probabilities is a critical concept that leads to **mutual independence**:

$$P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1)P(A_2)\dots P(A_n)$$

Example 1: Given a dice with two events $A = \{1, 2, 3\}$ and $B = \{2, 4\}$, proof the two events are independent.

- $P(A) = \frac{1}{2}$
- $P(B) = \frac{1}{3}$
- $P(A \cap B) = \frac{1}{6}$

Process:

1. Find $P(A|B)$:

$$\begin{aligned} P(A|B) &= \frac{P(A \cap B)}{P(B)} \\ &= \frac{\frac{1}{6}}{\frac{1}{3}} \\ &= \frac{1}{2} \\ &= P(A) \end{aligned}$$

2. Find $P(B|A)$:

$$\begin{aligned} P(B|A) &= \frac{P(B \cap A)}{P(A)} \\ &= \frac{\frac{1}{6}}{\frac{1}{2}} \\ &= \frac{1}{3} \\ &= P(B) \end{aligned}$$

3. Conclude $P(A|B) = P(A)$ and $P(B|A) = P(B)$ thus A and B are independent.

2.7 Probability tables

Information regarding the probability of pairs of events and their complements can be conveniently represented as a **probability table**.

	A	$\overline{A}$	
B	$P(A \cap B)$	$P(\overline{A} \cap B)$	$P(B)$
$\overline{B}$	$P(A \cap \overline{B})$	$P(\overline{A} \cap \overline{B})$	$P(\overline{B})$
	$P(A)$	$P(\overline{A})$	1

Figure 1

The table represents disjoint unions i.e. $P(A) + P(\overline{A}) = 1$ or $P(A) = P(A \cap B) + P(A \cap \overline{B})$.

Example 1: Given 100 girders, faults of type A and B are checked.

- 2 have both types of faults - $A \cap B$
- 6 have type A fault - $A \cap \overline{B}$
- 4 have type B fault - $\overline{A} \cap B$

	A	$\overline{A}$	6
B	2	4	6
$\overline{B}$	6	?	$100 - 6$
	8	$100 - 8$	100

Figure 2

Conditional probability calculations are straight forward:

- $P(A|B) = \frac{1}{3}$
- $P(B|A) = \frac{2}{8} \frac{1}{4}$

- Probability of one fault:

$$
\begin{aligned}
P((A \cap \overline{B}) \cup (\overline{A} \cap B)) &= P(A \cap \overline{B}) + P(\overline{A} \cap B) \\
&= \frac{6}{100} + \frac{4}{100} \\
&= \frac{1}{10}
\end{aligned}
$$

2.8 Total probability

Previous chapters mentioned that a collection of disjoint sets $A_1, A_2, \ldots, A_n$ forms a **partition** of S if $A_i \cap A_j = \phi$ thus:

$$S = \bigcup_{i=1}^{k} A_i$$

Therefore for any event B:

$$B = (B \cap A_1) \cup \cdots \cup (B \cap A_k) \Rightarrow P(B) = P(B \cap A_1) + \cdots + P(B \cap A_k)$$

Each event can be written in terms of the multiplication law to yield:

$$P(B) = P(B|A_1)P(A_1) + \cdots + P(B|A_k)P(A_k) = \sum_{i=1}^{k} P(B|A_i)P(A_i)$$

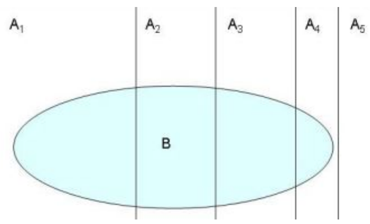


Figure 3

For events $A_1, A_2, \ldots, A_k$ such that $A_i \cap A_j = \phi$:
- For all i and j
- $i \neq j$
- $\bigcup_{i=1}^{k} A_i = S$

the theorem of total probability states:

$$P(B) = \sum_{i=1}^{k} P(B|A_i)P(A_i)$$

Example 1: Consider a factory with 3 machines X, Y and Z that is used to produce a specific component and each that a different defect probability:

- X: Produces 50% of components with 3% defective.
- Y: Produces 30% of components with 4% defective.
- Z: Produces 20% of components with 5% defective.

Compute the probability that a randomly selected item is a defect.

Process:

1. D denotes the event concerned:
$$D = (D \cap X) \cup (D \cap Y) \cup (D \cap Z)$$

2. Apply the law of total probability:
$$\begin{aligned} P(D) &= P(D|X)P(X) + P(D|Y)P(Y) + P(D|Z)P(Z) \\ &= 0.03(0.5) + 0.04(0.3) + 0.05(0.2) \\ &= 0.037 \end{aligned}$$

2.9 Bayes Theorem

Bayes Theorem is also commonly known as the rule for switching conditional probabilities.

For events $A_1, \ldots, A_n$ that form a partition of S and any other event B, the multiplication rule for conditional probability states:

$$P(A_k \cap B) = P(A_k)P(B|A_k)$$

Therefore:

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

The term in the denominator can be expressed as terms of conditional probabilities via the theorem of total probability.

Bayes Theorem:
For events $A_1, \ldots, A_n$ that form a partition of S and any other event B:

$$P(A_k|B) = \frac{P(A_k)P(B|A_k)}{P(A_1)P(B|A_1) + \cdots + P(A_n)P(B|A_n)}$$

Bayes Theorem provides the ability to determine the probability that a particular event A **occurred** given that event B has **already occurred**.

Example 1: Building on the previous example, suppose a defective component is found among the output of a factory, find the probability that it came from each of the machines X, Y and Z. $P(D)$ is already calculated to be 0.037.

- Find $P(X|D)$:

$$P(X|D) = \frac{P(D|X)P(X)}{P(D)} = \frac{0.03(0.5)}{0.037} = 0.4054$$

- Find $P(Y|D)$:

$$P(Y|D) = \frac{P(D|Y)P(Y)}{P(D)} = \frac{0.04(0.3)}{0.037} = 0.3243$$

- Find $P(Z|D)$:

$$P(Z|D) = \frac{P(D|Z)P(Z)}{P(D)} = \frac{0.05(0.2)}{0.037} = 0.2703$$

3 Random variables and Probability distributions

3.1 Random Variables

Random variables are fundamental in probability and statistics.

Random variable X (on a sample space S): A measurable function that assigns a real numbered value to each outcome of S.

Range (of a random variable): The collection of values the random variable can take on.

Notation: **Uppercase letters** for random variables and **lowercase letters** to denote particular values a random variable can take on e.g. random variable X takes on a specific value x - $X = x$.

The expression $X = x$ refers to the set of all elements in S assigned the value of x by the random variable X - since random variable refers also to the elements in S:

$$P(X = x)$$

meaning "the probability that the random variable X takes the value x".

The range of the random variables determine whether it is a **discrete random variable** or a **continuous random variable**.

- **Discrete random variable**: Range of random variable can be counted e.g. coin toss or dice roll.
- **Continuous random variable**: A random variable whose function is continuous everywhere - takes on all values in a given interval of numbers.

3.2 Discrete Random Variables

Probability distribution (for a discrete random variable X): A collection of probabilities assigned by the random variable to its range and ca be represented in some form that provides the probabilities corresponding to each x such as as a table and a graph.

For a discrete random variable X with range x_1, x_2,...:

$$p_i = f_X(x_i) = P(X = x_i) \text{ where } i = 1, 2, \dots$$

from the above equation, the following must be true:

- $p_i \geq 0$
- $\sum_{i=1}^{\infty} p_i = 1$

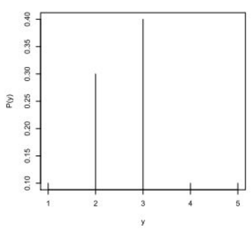
Probability mass distribution: $P(X = x)$ described as a function of x

Cumulative distribution function $F_X(x)$ (of a discrete random variable X): The probability that a random variable taking a value **less than** or **equal to** x.

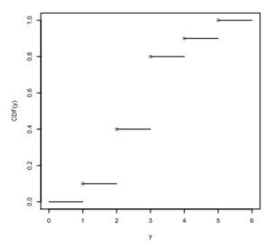
$$F_X(x_j) = P(X \leq x_j) = \sum_{i=1}^{j} p_j = p_1 + p_2 + \cdots + p_j$$

Example 1: Consider the discrete random variable Y with range $\{1, 2, ..., 5\}$ and corresponding probability distribution:

y	1	2	3	4	5
$f_Y(y)$	0.1	0.3	0.4	0.1	0.1



(a) Probability mass function



(b) Cumulative distribution function

3.3 Theoretical Mean and Variance

The following two numbers **mean** and **variance** are often used to summarize a probability distribution for a random variable X. Note that these two numbers do not uniquely define a distribution, some distributions will have the same mean and variance.

- **Mean**: A measure of the center of the probability distribution.
- **Variance**: A measure of the dispersion (variability) in the distribution.

E is the **expectation operator** defined as the computation that finds the sum of function X weighted by the corresponding probability.

Mean or **Theoretical mean** (of a discrete random variable X) - $E(X)$ or μ: Weighted average of the possible values of X where the weights are equal to the probabilities.

$$\mu = E(X) = \sum_x x\, f(x)$$

If $f(x)$ is considered a probability mass function of a loading on a long, thin beam, $E(X)$ is the point at which the beam balances. This reinforces the point that $E(X)$ describes the center of the distribution of X.

Properties:

- $E(aX + b) = aE(X) + b$ for any $a, b \in \Re$
- $E[g(X)] = \sum_x g(x) f_X(x)$ for any function $g(x)$

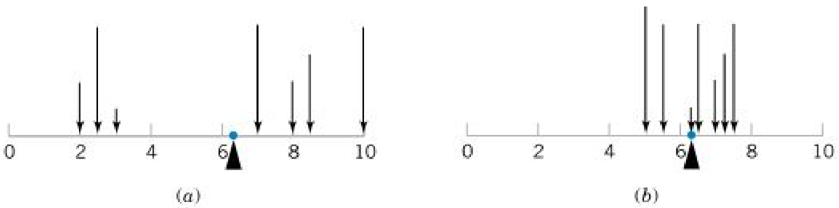


Figure 5: Triangle representing the graphical representation of Mean

Example 1:

Variance or **Theoretical variance** (of X) - σ^2 or $V(X)$: A measure of dispersion/scatter of the distribution.

$$\sigma^2 = V(X) = E(X-\mu)^2 = \sum_x (x-\mu)^2 f(x) = \sum_x x^2 f(x) - \mu^2$$

or

$$E(X^2) - E(X)^2$$

where $V(X) \geq 0$
Standard deviation (of X) - σ:

$$\sigma = \sqrt{\sigma^2}$$

Some distributions will have the same mean but different variances.

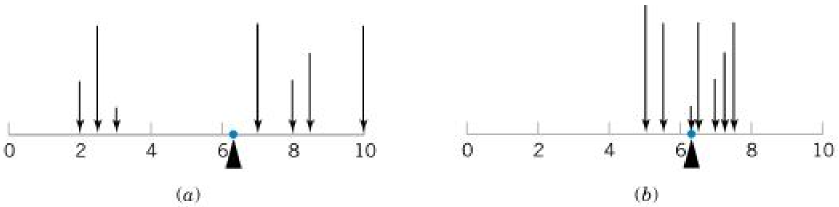


Figure 6: (a) and (b) showing different variance

Properties:

- $V(aX + b) = a^2 V(X)$ for any $a, b \in \Re$

Example 2: Given the discrete random variable Y with range $1, 2, \ldots, 5$ and corresponding probability distribution, find the variance and standard deviation.

y	1	2	3	4	5
$f_Y(y)$	0.1	0.3	0.4	0.1	0.1

Process:

1. Find the Mean μ:

$$\begin{aligned} \mu &= E(Y) \\ &= 1(0.1) + 2(0.3) + 3(0.4) + 4(0.1) + 5(0.1) \\ &= 2.8 \end{aligned}$$

2. Find the variance:

$$\begin{aligned} V(Y) &= \sum_y (y - \mu)^2 f_Y(y) \\ &= (1 - 2.8)^2(0.1) + (2 - 2.8)^2(0.3) + (3 - 2.8)^2(0.4) + (4 - 2.8)^2(0.1) + (5 - 2.8)^2(0.1) \\ &= 1.16 \end{aligned}$$

3. Standard deviation:

$$\begin{aligned} \sigma &= \sqrt{1.16} \\ &\approx 1.8 \end{aligned}$$

Example 3:

3.4 Important discrete distribution

3.4.1 Discrete uniform distribution

The simplest discrete random variable is one that assumes only a finite number of possible values, each with equal probability.

Discrete uniform distribution: A random variable is defined by which each of the n values in its range e.g. $x_1, x_2, \ldots, x_n$ with equal probability thus:

$$f(x_i) = \frac{1}{n}$$

Properties:

- Probability Mass Function:

$$f_X(x) = \frac{1}{k} \text{ for } x = 1, 2, \ldots, k$$

  Equally distributed with even spacing.

- Cumulative Distribution Function:

$$F_X(x) = \sum_{i=1}^{x} \frac{1}{k} = \frac{x}{k} \text{ for } x = 1, 2, \ldots, k$$

- Mean:

$$E(X) = \frac{k+1}{2} \text{ where } k \text{ is the number of elements in the distribution}$$

- Variance:

$$V(X) = E(X^2) - E(X)^2 = \frac{k^2 - 1}{12}$$

3.4.2 Binomial distribution

Many experiments/trails have two exclusive outcomes i.e. H or T and defective or not. Such trials are called **Bernoulli trial**.

Key characteristics of a Bernoulli Trail:

- Dichotomous outcomes e.g. labels *"success"* or *"failure"*
- n identical trails
- Trails are **independent -** one event does not affect the other event
- Constant probability of success p

Note: Factorial notation is used in Binomial Distribution: $n!$ - Product of integers less than or equal to n:

$$n! = n(n-1)(n-2)\dots(2)(1)$$

Combinatorial notation defined as:

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}$$

where $\binom{n}{x}$ is equal to the *total number* of different sequences of trials that contain:

- x successes
- $n - x$ failures

The random variable X equals the number of trials that result in a success can be labelled as a **binomial random variable** where parameters $)0 < p < 1$ and $n = 1, 2\dots,$ the probability mass function of X:

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x} \text{ where } x = 0, 1, \dots, n$$

The common notation defined:

$$X \sim Bin(n, p)$$

where:

- X - The random variable is binomial
- n - Number of trials
- p - Probability of success

Example 1: A fair coin is tossed 6 times with sample space $S = \{H, T\}$. Note this is a binomial experiment with $n = 6$ and $p = q = \frac{1}{2}$.

1. Probability of exactly two H: $x = 2$ and $n - x = 4$:

$$\begin{aligned} f_X(2; n = 6, p = \frac{1}{2}) &= \binom{6}{2} \left(\frac{1}{2}\right)^4 \left(\frac{1}{2}\right)^2 \\ &= \frac{6!}{2!(6-2)!} \left(\frac{1}{4}\right) \left(\frac{1}{16}\right) \\ &\approx 0.23 \end{aligned}$$

2. E denotes event where **at least** 4 H is achieved i.e. 4, 5 and 6 H's, find

the probability.

$$
\begin{aligned}
P(E) &= \sum_{x\in\{4,5,6\}} f_X\left(x; n=6, p=\frac{1}{2}\right) \\
&= \binom{6}{4}\left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right)^2 + \binom{6}{5}\left(\frac{1}{2}\right)^5\left(\frac{1}{2}\right) + \binom{6}{6}\left(\frac{1}{2}\right)^6\left(\frac{1}{2}\right)^0 \\
&= \frac{15}{64} + \frac{6}{64} + \frac{1}{64} \\
&= \frac{11}{32} \\
&\approx 0.34
\end{aligned}
$$

3. Find the probability of getting no heads i.e. all failures:

$$
\begin{aligned}
(1-6)^6 &= \left(\frac{1}{2}\right)^6 \\
&= \frac{1}{64}
\end{aligned}
$$

If X is a binomial random variable with parameters p and n:

- Mean: $\mu = E(X) = np$
- Variance: $\sigma^2 = V(X) = np(1-p)$

3.4.3 Geometric distribution

3.4.4 Poisson distribution

4 Systems and Component reliability

5 Jointly distributed random variables

6 Law of large numbers and Central limit theorem

7 Statistics