Probability

Ignacio Sada Sólomon

Winter 2020

Abstract

Sample space, events, conditional probability, independence of events, Bayes' Theorem. Basic combinatorial probability, random variables, discrete and continuous univariate and multivariate distributions. Moment generating functions Independence of random variables. Chebyshev's inequality,central limit theorem, weak law of large numbers (if time allows).

Contents

1 Fundamentals

1.1 Set Theory

Definition 1.1. A **set** is a collection of objects, called elements.

Example 1.1. The set of natural numbers $\mathbb{N} = \{1, 2, 3, \dots\}$, which is both infinite and countable.

Example 1.2. The set of all McGill students.

Definition 1.2. Let X be a set. A **subset** is another set A such that every element of A is in X as well:

$$A = \{x : x \in X\} \quad \text{or} \quad x \in A \implies x \in X$$

Remark. The empty set is denoted $\varnothing = \{\}$

Definition 1.3. An **intersection** between the sets A and B is defined as

$$A \cap B = x \in A \quad \text{and} \quad x \in B$$

Definition 1.4. A **union** between the sets A and B is defined as

$$A \cup B = x \in A \quad \text{or} \quad x \in B$$

Definition 1.5. Let X be a universal set with a subset A. Then the complement of A with respect to X is

$$A^c = X \setminus A = \{x : x \in X \quad \text{and} \quad x \notin A\}$$

Example 1.3. $\mathcal{P}(B) = \{A : A \subseteq B\}$ (all subsets of B)

Let $B = \{1, 2, 3\}$, then

$$\mathcal{P}(B) = \left\{ \underbrace{\{\}, \{1, 2, 3\}}_{\text{trivial subsets}}, \underbrace{\overbrace{\{1\}, \{2\}, \{3\}}^{\text{singletons}}, \overbrace{\{1, 2\}, \{1, 3\}, \{2, 3\}}^{\text{doubles}}}_{\text{proper subsets}} \right\}$$

Definition 1.6. Two sets A and B are denoted as **disjoint** whenever they share no elements, i.e. they have nothing in common:

$$A \cap B = \varnothing$$

Example 1.4. Let $\Omega = \mathbb{N} = \{0, 1, 2, 3, 4, \dots\}$ and $A \subseteq N = \{n \in \mathbb{N} : n \text{ is even}\}$. Then,

$$A^c = \{n \in \mathbb{N} : n \text{ is odd}\}$$

Here are some important set properties:

- $A \cap \varnothing = \varnothing$
- $A \cup \varnothing = A$
- $A \cup B = B \cup A$
- $A \cap B = B \cap A$

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

Exercise 1.1. Prove that

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

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

Note. Consider $A \setminus B = \{x \in A : x \notin B\}$. Then we define the **exclusive or** as

$$A \setminus B \cup B \setminus A = A \triangle B$$

2 Probability

2.1 Introduction

Probability is the branch of applied mathematics that deals with random events, which are non-deterministic in nature, unlike calculus for instance, which deals with deterministic events.

Example 2.1. Toss a coin. We won't know the outcome of the toss, but we do know that the outcome will be either *heads* or *tails*.

Example 2.2. Roll a die. We now have several outcomes, depending on the number of faces on the die.

Definition 2.1. Given a random experiment, the set of all possible outcomes is denoted Ω, and is called the **sample space**.

Example 2.3.

- The sample space for a coin toss is $\Omega = \{H, T\}$

- The sample space for a rolled 6-face die is $\Omega = \{1, 2, 3, 4, 5, 6\}$

- The sample space for a coin toss where we require heads once is $\Omega = \{H, TH, TTH, \dots\}$

Definition 2.2. Let Ω denote the set of all possible outcomes of a random experiment.

  2.2.1. An **elementary event** is a subset of Ω with a singular element (i.e. one single outcome) and is also an element of $\mathcal{P}(\Omega)$.

  2.2.2. A **compound event** is any subset of Ω (including the elementary events).

  2.2.3. An **event** is any element of Ω.

  2.2.4. $\varnothing$ denotes **impossible events**.

  2.2.5. A **complimentary event** given an event A is denoted A^c.

  2.2.6. Given 2 events A and B, where $A \cap B = \varnothing$, we denote them as **disjoint events**.

Definition 2.3. Let Ω be the sample space attached to a random experiment. A **probability** is a function P such that

$$P : \mathcal{P}(\Omega) \to [0, 1]$$

Note that $P(\Omega) = 1$. Given a sequence $A_1, \dots, A_n, \dots$, of pairwise disjoint events ($A_i \cap A_j = \varnothing \quad \forall i \neq j$), we have that

$$P\left(\bigcup_{i=1} A_i\right) = \sum_{i=1} P(A_i)$$

Remark. In general, if Ω is countable, then it is enough to define P on the elementary event, as every other event can be written as a union.

Example 2.4. Toss a coin such that $\Omega = \{H, T\}$. A probability on Ω is completely given by $p \in [0, 1]$, and the assignment of $P(\{H\}) = p$, and $P(\{T\}) = 1 - p$.

Remark. If $p = 1/2$ in the previous example, then $P(\{H\}) = P(\{T\}) = 1/2$, implying that *the coin is fair/balanced.*

Example 2.5. Throw a die such that $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6\}$. A probability on Ω is given by 6 non-negative integers p_i where $i = 1, 2, 3, 4, 5, 6$, such that

$$p_i = P(\{w_i\}) \geq 0$$

And so

$$\sum_{i=1}^{6} p_i = 1$$

Therefore, if we consider the die to be fair, we have $p_i = 1/6 \quad \forall i \in \{1, 2, 3, 4, 5, 6\}$.

Let $A = \{\omega_1, \omega_3\}$ Then

$$
\begin{aligned}
A = \{\omega_1\} \cup \{\omega_3\} &\implies P(A) \\
&= P(\{\omega_1\}) + P(\{\omega_3\}) \\
&= 1/6 + 1/6 \\
&= 1/3
\end{aligned}
$$

Exercise 2.1. Consider a die such that the probability $P(\{\omega_i\})$ is proportional to k such that

$$P\big(\{\omega_i\}\big) = ck$$

Therefore:

$$
\begin{aligned}
1 &= \sum_{k=1}^{6} P(\{\omega_k\}) \\
&= \sum_{k=1}^{6} ck \\
&= c \sum_{k=1}^{6} k \\
&= c \frac{n(n+1)}{2} \bigg|_{n=6}
\end{aligned}
$$

$$= c\frac{6 \times 7}{2}$$

$$= 21c$$

$$\therefore c = \frac{1}{21}$$

What would be the probability of an even number?

Let $A = \{\omega_2, \omega_4, \omega_6\}$, and so

$$P(A) = P(\{\omega_2\}) + P(\{\omega_4\}) + P(\{\omega_6\})$$
$$= \frac{2 + 4 + 6}{21}$$
$$= \frac{12}{21}$$

What would be the probability of an odd number?

Simple:

$$P(B) = 1 - P(A)$$
$$= \frac{9}{21}$$

Example 2.6. Toss a coin until heads appears. We then have:

$$\Omega = \{\omega_1 = H, \omega_2 = TH, \omega_3 = TTH, \ldots, \omega_n = \underbrace{TTT}_{n-1} H, \ldots\}$$

$$\therefore P(\{\omega_n\}) = c \left(\frac{1}{3}\right)^n$$

We now let $1 = c \left(\sum_{n=1}^{\infty} \left(\frac{1}{3}\right)^n\right)$ such that

$$c = \frac{1}{\sum_{n=1}^{\infty} \left(\frac{1}{3}\right)^n}$$
$$= \frac{1}{\frac{1}{3}\left(\frac{1}{1-1/3}\right)}$$
$$= \frac{1}{\frac{1}{3}\left(\frac{3}{2}\right)}$$
$$= \frac{1}{\frac{1}{2}}$$
$$= 2$$

Thus, $P(\{\omega_n\}) = 2 \left(\frac{1}{3}\right)^n \implies H$ appears in an even number of trials. Then, what is $P(A)$? Recall that $A = \{\omega_1, \omega_2, \ldots, \omega_n, \ldots\}$ so

$$
\begin{aligned}
P(A) &= \sum_{n=1}^{\infty} P\left(\{\omega_n\}\right) \\
&= \sum_{n=1}^{\infty} 2 \left(\frac{1}{3}\right)^{2n} \\
&= 2 \sum_{n=1}^{\infty} \left(\frac{1}{3^2}\right)^{2n} \\
&= 2 \cdot \frac{1}{9} \left(\frac{1}{1 - \frac{1}{9}}\right) \\
&= \frac{1}{4}
\end{aligned}
$$

Theorem 2.1. *Let A, B be subsets of a set Ω. The probability function $P : \Omega \to [0,1]$ has the following properties:*

1) $P(\varnothing) = 0$

2) $P(A^c) = 1 - P(A)$

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

Proof.

1)

$$
\begin{aligned}
\Omega = \Omega \cup \varnothing &\iff \Omega \cap \varnothing = \varnothing \\
&\iff P(\Omega) = P(\Omega \cup \varnothing) \\
&\iff P(\Omega) + P(\Omega) \\
&\iff 1 = 1 + P(\varnothing) \\
&\iff P(\varnothing) = 0
\end{aligned}
$$

2)

$$
\begin{aligned}
\Omega = A \cup A^c &\iff A \cap A^c = \varnothing \\
&\iff 1 = P(\Omega) = P(A) + P(A^c) \\
&\iff P(A^c) = 1 - P(A)
\end{aligned}
$$

3)

$$A \cup B = (A \setminus B) \cup B \iff (A \setminus B) \cap B = \varnothing \implies P(A \cup B) = P(A \setminus B) + P(B)$$

$$\therefore A = (A \setminus B) \cup (A \cap B) \iff (A \setminus B) \cap (A \cap B) = \varnothing \implies P(A) = P(A \setminus B) + P(A \cap B)$$

$$\iff P(A \setminus B) = P(A) - P(A \cap B)$$

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

■

Exercise 2.2. Show that $P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$

Definition 2.4. Let Ω_1, Ω_2 be two finite sample spaces.

1) $\Omega_1 \times \Omega_2 = \{(\omega_1, \omega_2) : \omega_1 \in \Omega_1, \omega_2 \in \Omega_2\}$

2) $|\Omega_1 \times \Omega_2| = |\Omega_1| \cdot |\Omega_2|$

Example 2.7. Roll a die twice. What is the probability that the sum of the number obtained is 7? We will define the set of events where the sum of both rolls is 7 as A:

$$\Omega = \Omega_1 \times \Omega_2 \implies |\Omega| = |\Omega_1| \cdot |\Omega_2| = 36$$

$$\Omega_1 = \Omega_2 = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6\}$$

Thus,

$$A = \{(\omega_1, \omega_6), (\omega_2, \omega_5), (\omega_3, \omega_4), (\omega_4, \omega_3), (\omega_5, \omega_2), (\omega_6, \omega_1)\}$$

$$\therefore P(A) = \frac{|A|}{|\Omega|} = \frac{6}{36} = \frac{1}{6}$$

Exercise 2.3. Find the probability such that the sum of the number obtained in both rolls is even.
(*Ans: 1/2*)

Definition 2.5. A **permutation** of r elements chosen from n elements is equivalent to throwing successively without replacement r elements from a urn which contains n elements. The total number of combinations for a permutation is

$$C_r^n = \binom{n}{r} = \frac{P_r^n}{r!} = \frac{n!}{r!(n-r)!}$$

Note that:

$$P_r^n = n(n-1)(n-2) \dots (n-r+1) = \frac{n!}{(n-r)!}$$

is the number of possible combinations.

Example 2.8. A urn contains 4 balls; 1 green ball, 1 blue ball, 1 red ball, and 1 yellow ball. Draw successively without replacement 3 balls from the urn.

Example 2.9. A hand is a subset of 5 cards from a deck of 52 cards.

$$C_r^n = C_5^{52} = \frac{52!}{5!(47!)} = \frac{48 \times 49 \times 50 \times 51 \times 52}{1 \times 2 \times 3 \times 4 \times 5}$$

What is the probability that a selected (or random) hand will contain at least one Jack? Let $A =$ "At least 1 Jack" such that $A^c = $ "No jack". Then,

$$P(A^c) = \frac{C_5^{48}}{C_5^{52}} \qquad\qquad P(A) = 1 - P(A^c)$$

If we let $A_i =$ "At least i Jacks" for $i = 1, 2, 3, 4$, then

$$P(A) = P\left(\bigcup_{i=1}^4 A_i\right)$$

$$= \sum_{i=1}^4 P(A_i) \qquad\qquad P(A_i) = \frac{C_i^4 \times C_{5-i}^{48}}{C_5^{52}}$$

Remark. C_r^n is the number of subsets with r elements of a set S which has n elements.

Consider this alternative solution:

$$P(A) = 1 - \overbrace{P(A^c)}^{\text{No jack}}$$

$$\therefore P(A^c) = \frac{C_5^{48}}{C_5^{52}} \implies P(A) = 1 - \frac{C_5^{48}}{C_5^{52}}$$

Exercise 2.4. The letters of the word "ANANAS" are written on 6 marbles. Select 3 marbles successively, without replacement, from the initial 6 marbles to form a 3 letter word.

Theorem 2.2. *The Binomial Theorem:*

$$(a + b)^n = \sum_{k=0}^n C_k^n a^k b^{n-k}$$

Example 2.10. $(a + b)^4 = b^4 + 4ab^3 + 6a^2b^2 + 4a^3b + a^4$

Example 2.11. Find the coefficient of x^7 in the expansion of $(2 + 3x^2)^6$
(*Ans: $C_4^5 3^4 2^3$*)

We can apply this concept to find the *power set of a fine set.* Let Ω be a finite set. $\mathcal{P}(\Omega)$ is the power set of Ω. If $|\Omega| = n$, then $|\mathcal{P}(\Omega)|$ is

$$
\begin{aligned}
|\mathcal{P}(\Omega)| &= C_0^n + C_1^n + C_2^n + \cdots + C_k^n + \cdots + C_n^n \\
&= \sum_{k=0}^{n} C_k^n \\
&= \sum_{k=0}^{n} (1)^k (1)^{n-k} C_k^n \\
&= (1+1)^n \\
&= \boxed{2^n}
\end{aligned}
$$

Example 2.12.

$$
\begin{aligned}
(a+b)^4 &= \sum_{k=0}^{4} C_k^4 a^k b^{4-k} \\
&= b^4 + 4ab^3 + 6a^2 b^2 + 4a^3 b + a^4
\end{aligned}
$$

2.2   Conditional Probability

Consider rolling a fair die. The probability that we obtain 2 as a result is $\frac{1}{6}$. However, the probability of obtaining 2 from only even numbers is $\frac{1}{3}$.

Definition 2.6. Let Ω be a parent set of events, and let us consider subsets of events $A \subset \Omega$ and $B \subset \Omega$. Then, the **conditional probability** of A given B is

$$
P(A \mid B) = \frac{P(A \cap B)}{P(B)}
$$
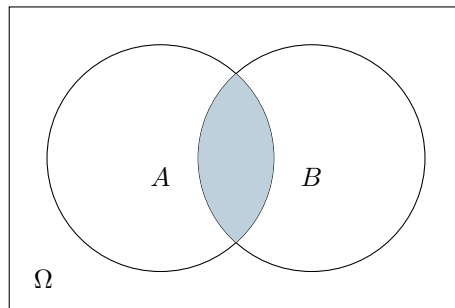
Remark. The map $A \mapsto P(A \mid B)$ is a probability on $\mathcal{P}(\Omega)$.

Example 2.13. Throw a fair die. We define the following events:

- A = result is even.
- B = either 1, 2, or 5
- C = either 1 or 2

Thus:

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

Hence:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{1}{3} \qquad P(A \mid C) = \frac{P(A \cap C)}{P(C)} = \frac{1}{2} \qquad P(B \mid A) = \frac{P(A \cap B)}{P(A)} = \frac{1}{3}$$
$$P(C \mid A) = \frac{P(A \cap C)}{P(A)} = \frac{1}{3} \qquad P(B \mid C) = \frac{P(B \cap C)}{P(C)} = 1 \qquad P(C \mid B) = \frac{P(B \cap C)}{P(B)} = \frac{2}{3}$$

Definition 2.7. Two events A and B are said to be **independent** if at least one of the following statements is true:

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

Example 2.14. In Example 2.13:

- A and B are not independent $\left(P(A \cap B) = \frac{1}{6}, \quad P(A)P(B) = \frac{1}{4}\right)$

- A and C are independent $\left(P(A \cap C) = \frac{1}{6}, \quad P(A)P(C) = \frac{1}{6}\right)$

- B and C are not independent $\left(P(B \cap C) = \frac{1}{3}, \quad P(B)P(C) = \frac{1}{6}\right)$

Example 2.15. Toss a coin (which is such that $P(H) = p \in (0,1)$)twice. What is the probability that we will obtain T twice?

$$P(\underbrace{\{T\} \times \{T\}}_{TT})$$

By design, the outcome obtained on the first toss is *independent from the outcome obtained in the second toss:*

$$P(TT) = P(\{T\}) \times P(\{T\})$$
$$= (1 - p)^2$$

Example 2.16. Toss a coin such that $P(\{H\}) = p \in (0,1)$ until H appears.

Solution.

We have:

$$\Omega = \{H, TH, TTH, TTTH, \ldots, \underbrace{T\ldots T}_{n-1}H\}$$

$$\therefore P(\omega_n) = \big(P(T)\big)^{n-1} \times \big(P(H)\big)$$
$$= (1-p)^{n-1} \times (p)$$

We now recall infinite series expansion to evaluate $P(\omega_n)$:

$$\sum_{n=1}^{\infty}(1-p)^{n-1}(p) = p\sum_{n=1}^{\infty}(1-o)^{n-1}$$
$$= p\sum_{n=0}^{\infty}(1-p)^{n}$$
$$= p\left(\frac{1}{1-(1-p)}\right)$$
$$= 1$$

◎

Theorem 2.3. *If A and B are independent $(A \perp\!\!\!\perp B)$, then so are A^C and B^C. Consequently:*

$$A^C \perp\!\!\!\perp B \qquad\qquad\qquad B^C \perp\!\!\!\perp A$$

Proof.

$$P(A^C \cap B) = P(B \setminus A)$$
$$= P(B) - P(A \cap B)$$
$$= P(B) - P(A)P(B)$$
$$= P(B) - (1 - P(A))$$
$$= P(B)P(A^C)$$

B^C and A are similarly shown to be independent. ∎

Example 2.17. Let $P(A) = 0.7$, $P(B) = 0.8$, A and B are independent $(A \perp\!\!\!\perp B)$. Find $P(A^C \cap B^C)$.

$$P(A^C \cap B^C) = P(A^C)P(B^C)$$
$$= (0.3)(0.2)$$
$$\boxed{= 0.06}$$

◎

Example 2.18. Let $P(A) = 0.7$, $P(B) = 0.8$, A and B are independent $(A \perp\!\!\!\perp B)$. Find $P(A \cup B)$.

Solution.

$$(A \cup B)^C = A^C \cap B^C$$
$$= P(A) + P(B) - P(A \cap B)$$
$$= P(A) + P(B) - P(A)P(B)$$
$$= 0.7 + 0.8 - 0.56$$
$$\boxed{= 0.94}$$

◎

Exercise 2.5. A die is such that $P(\{\omega_2\}) = p$, $P(\{\omega_4\}) = q$, and $p + q \in (0, 1)$. Roll the die until 2 or 4 appears. What is the probability that 2 appears first?

$(Ans: \frac{p}{p+q})$

2.3 Baye's Rule

Example 2.19. On a given exam, we are given the following distribution: Here,

	Success (S)	Failure (Fl)	Total
Male (M)	0.3	0.1	0.4
Female (Fm)	0.4	0.2	0.6
Total	0.7	0.3	1

- $P(M) = 0.4$
- $P(Fm) = 0.6$
- $P(S \mid M) = \frac{3}{4}$

- $P(S \cap M) = 0.3$
- $P(S \cap Fm) = 0.4$
- $P(S \mid Fm) = \frac{2}{3}$

Thus,

$$P(S) = P(M \cap S) + P(Fm \cap S)$$

$$= P(S \mid M)P(M) + P(S \mid Fm)P(Fm)$$

$$= \left(\frac{3}{4} \times \frac{4}{10} \right) + \left(\frac{2}{3} \times \frac{6}{10} \right)$$

$$\boxed{= \frac{7}{10}}$$

Definition 2.8. Let Ω be a sample space. A **finite partition** of Ω is given by events $A_1, A_2, A_3, \ldots, A_n$, which are pairwise disjoint, and also

$$\bigcup_{n \in \mathbb{N}^*} A_n = \Omega$$
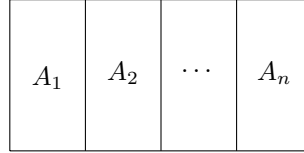
Theorem 2.4. *Baye's Theorem*

Given a partition $A_1, A_2, \ldots, A_n$ of Ω, suppose that $P(A_1), P(A_2), \ldots, P(A_n)$ are known. Suppose that for an event B, the following are known: $P(B \mid A_1), P(B \mid A_2), \ldots, P(B \mid A_n)$. Thus:

1.

$$P(B) = \sum_{i=1}^{n} P(B \mid A_i)P(A_i)$$

2. Let $k \in \{1, 2, \ldots, n\}$ be fixed. Then:

$$P(A_k \mid B) = \frac{P(B \mid A_k)P(A_k)}{\sum_{i=1}^{n} P(B \mid A_i)P(A_i)}$$

Proof.

Let $B = \Omega$, then:

$$B = B \bigcap \left(\bigcup_{n \in N \setminus \{0\}} A_n \right)$$

$$= \bigcup_{n \in N \setminus \{0\}} \left(B \bigcap A_n \right)$$

$$\therefore P(B) = \sum_{k=1}^{n} P(B \bigcap A_k)$$

13

$$\therefore P(A_i \mid B) = \frac{P(B \bigcap A_i)}{P(B)}$$
$$= \frac{P(B \mid A_i)P(A_i)}{P(B)}$$

∎

Example 2.20. We have 2 urns. Urn 1 has 3 red balls and 2 blue balls, and urn 2 has 5 red balls and 6 blue balls. We now select a ball from urn 1, place it in urn 2, and select a ball from urn 2. What is the probability that said ball is red?

Solution.

Let A be the event where the ball from urn 2 is red, and B where the ball from urn 1 is blue. By Baye's Theorem,

1. $P(A) = \frac{2}{5} \cdot \frac{5}{12} + \frac{1}{2} \cdot \frac{3}{9} = \frac{7}{15}$

2. $P(B \mid A) = \frac{5}{12}$

◎

Example 2.21. On a weather forecast model we have that:

  a) If today is rainy, then tomorrow is rainy with a probability of 0.6.

  b) If today is sunny, then tomorrow is sunny with a probability of 0.7.

Suppose Monday is rainy. What is the probability that Wednesday is rainy?

Solution.

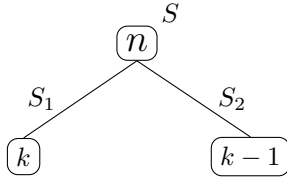Consider the following tree structure representing the probabilities of the forecast model:
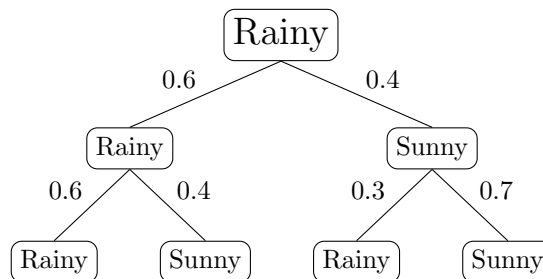
Let $B =$ "Wednesday is rainy", $A_1 =$ "Tuesday is rainy", and $A_2 =$ "Tuesday is sunny." Then,

$$P(A_1) = 0.6 \implies P(B \mid A_1) = 0.6$$

$$P(A_2) = 0.4 \implies P(B \mid A_2) = 0.3$$

$$\therefore P(B) = P(B \mid A_1)P(A_1) + P(B \mid A_2)P(A_2)$$

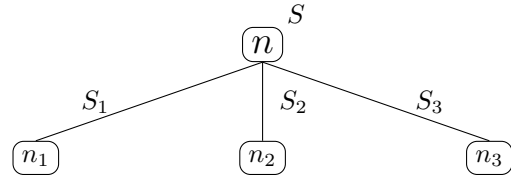$$= (0.6)(0.6) + (0.3)(0.4)$$

$$\boxed{= 0.48}$$

2.4 Multinomial Coefficients

We begin by recalling that *binomial coefficients* are such that a combination with C_k^n is clearly defined
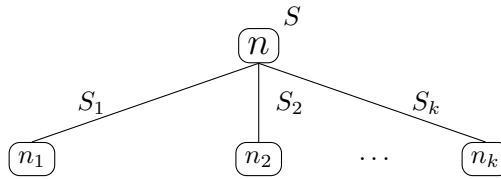

Partition of a set into two subsets.

Partition of a set into three subsets.

We consider the above cases to generalize by induction. In the case where we have 3 partitions, we have that $n = n_1 = n_2 = n_3$, and so the total number of partitions is

$$C_{n_1}^n = C_{n_3}^{n-n_1}$$

$$= \frac{n!}{(n-n_3)!n!}$$

$$= \frac{n!}{n_2!(n-n_1-n_3)!}$$

$$= \frac{n!}{n_1!n_2!n_3!}$$

Now suppose that we partition a set S with n elements into subsets $S_1, S_2, \ldots, S_k$ with respective

$n_1, n_2, \ldots, n_k$ elements ($\sum_i n_i = n$), so the total number of partitions is given by

$$\boxed{\binom{n}{n, n_1, \ldots, n_k} = \frac{n}{n_1! n_2! \ldots n_k!}}$$

Example 2.22. 10 employees are split in $k = 3$ ways for 3 different tasks. There are 3 people assigned to task 1, 2 people assigned to task 2, and 5 people assigned to task 3. Suppose that the assignment is done randomly. What is the probability that Peter and John do not belong to the same team?

Solution.

Recall that
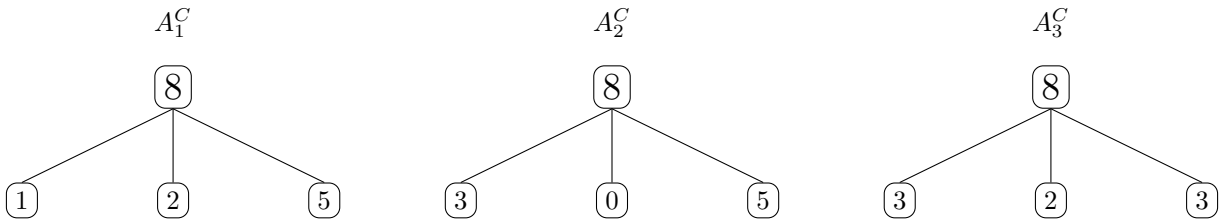
$$C_{n_1}^n = C_{n_2}^{n-n_1} = C_{n_3}^{n-n_1-n_2}$$

Hence:

$$
\begin{aligned}
C_{n_1}^n &= \frac{n!}{n_1! \cancel{(n-n_1)!}} \cdot \frac{\cancel{(n-n_1)!}}{n_2! \cancel{(n-n_1-n_2)!}} \cdot \frac{\cancel{(n-n_1-n_2)!}}{n_3!} \\
&= \frac{n!}{n_1! n_2! n_3!} \\
&= \binom{n}{n_1, n_2, n_3}
\end{aligned}
$$

Generally, if there are k tasks and n employees, with n_1 employees going to task i, then:

$$\sum_{i=1}^{k} n_i = n$$

We thus consider the cases where both Peter and John are on the same team:

Hence,

$$P(A^C) = \frac{A_1^C + A_2^C + A_3^C}{\left(\frac{10!}{3!2!5!}\right)}$$

$$\therefore P(A) = 1 - P(A^C) = 1 - \frac{A_1^C + A_2^C + A_3^C}{\left(\frac{10!}{3!2!5!}\right)}$$

Theorem 2.5. *Multinomial Theorem*

$$(a_1 + a_2 + \cdots + a_k)^n = \sum_{n_1 + n_2 + \cdots + n_k = n} \binom{n}{n_1, n_2, \ldots, n_k} \prod_{i=1}^{k} a_i^{n_i}$$

where $\binom{n}{n_1, n_2, \ldots, n_k} = \frac{n!}{n_1! n_2! \ldots n_k!}$

Exercise 2.6. Consider $(1 + k + 2k^5)^7$. Find the coefficient of k^5.

Example 2.23.

$$(a + b + c)^3 \implies \begin{array}{ccc} \boxed{(3,0,0)} & (1,2,0) & \boxed{(0,2,1)} \\ (2,1,0) & (1,0,2) & (0,0,3) \\ (2,0,1) & (1,1,1) & (0,3,0) \end{array}$$

3 Discrete Random Variables

We let Ω be a sample space. We assign a number to each probable element such that the sum of all elements is one, so for k elements with corresponding probabilities p_i for $i = 1, 2, \ldots, k$, we have

$$\sum_{i=0}^{k} p_i = 1$$

Definition 3.1. A **random variable** (r, σ) is a function $X : \Sigma \to \mathbb{R}$ such that:

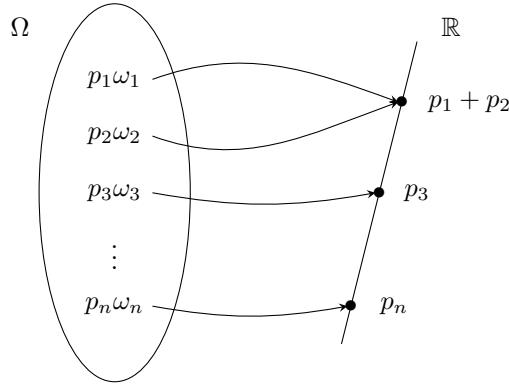$$\forall \omega \in \Omega : \quad X(\omega) \in \mathbb{R}$$

Example 3.1. Flip a coin, $\Omega = \{T, H\}$. Define $X(T) = -5$, $X(H) = 10$.

Example 3.2. Toss a die and add the numbers obtained. We have:

$$\Omega = \left\{ (\omega_i, \omega_j) : \begin{array}{c} i = 1, 2, 3, 4, 5, 6 \\ j = 1, 2, 3, 4, 5, 6 \end{array} \right\}$$

$$\therefore X : \Omega \to \mathbb{R} \implies X(\omega_1, \omega_2) = i + j$$

Definition 3.2. A **discrete random variable** is a random variable $X : \Omega \to \mathbb{R}$ that is said to be discrete if its *range* $X(\Omega)$ is countable. Note that it can be either finite or infinite.

Definition 3.3. Let X be a discrete random variable on a sample space Ω, and let P be the probability on $\mathcal{P}(\Omega)$. The **probability function** of X is the function denoted $P_X : \mathbb{R} \to [0, 1]$ such that

$$P_X(X) = \underline{P}(X = t)$$

Remark. $\{X = t\}$ is an event such that $X^{-1}(\{t\})$. For a discrete random variable, if $t \notin X(\Omega)$, then $\{X = t\} = \varnothing$, meaning that $P_X(X) = 0$. If $t \in X(\Omega)$, then we have that

$$P_X(t) = \sum_{X(\Omega)=t} P(\{\omega\})$$

Example 3.3. Let the sample space of a coin toss be $\Omega = \{H, T\}$. For a fair coin, we earn 10 credits for heads but lose 5 credits for tails. Find the probability. What is the probability of losing 5 credits?

Solution.

We have the following probability table:

Credits	-5	10
$P_X(t)$	$\frac{1}{2}$	$\frac{1}{2}$

Therefore,

$$\left.\begin{array}{l} X(T) = -5 \\ X(H) = 10 \end{array}\right\} P_X(-5) = P(X = -5)$$
$$= P(T)$$
$$= \frac{1}{2}$$

⊡

Example 3.4. Suppose that an item selected at random from a production line can have up to two types of defects:

$$D_1 = \text{"Item has defect of type 1."}$$
$$D_2 = \text{"Item has defect of type 2."}$$

Then, we assign $P(D_1) = 0.7$, $P(D_2) = 0.5$, and $P(D_1 \cap D_2) = 0.3$. Let X be the possible number of defects that an item selected at random can have such that: $X(\Omega) = \{0, 1, 2\}$. This is the support for P_X. Hence,

$$P_X(2) = P(X = 2)$$
$$= P(D_1 \cap D_2)$$
$$= 0.3$$

$$P_X(1) = P(X = 1)$$
$$= P(D_1 \triangle D_2)$$
$$= P\big((D_1 \cap D_2^C) \cup (D_1^C \cap D_2)\big)$$
$$= P(D_1 \cap D_2^C) + P(D_1 \cap D_2^C)$$
$$= P(D_1) - P(D_1 \cap D_2) + P(D_2) - P(D_1 \cap D_2)$$
$$= 0.6$$

$$P_X(0) = P(D_1^C \cap D_2^C)$$
$$= P\big((D_1 \cup D_2)^C\big)$$
$$= 1 - P(D_1 \cup D_2)$$
$$= 1 - 0.9$$
$$= \frac{1}{10}$$

3.1 Cumulative Distribution Function

Definition 3.4. Let X be a random variable (discrete or not). The **cumulative distribution function** of X is the function:

$$F_X = \begin{cases} R \to [0, 1] \\ x \mapsto P\big(\{t_1 \le t_2\}\big) \end{cases}$$
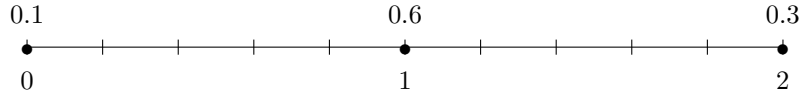
It has the following properties:

(1)
$$\lim_{t \to -\infty} F_X(t) = 0$$
$$\lim_{t \to +\infty} F_X(t) = 1$$

(2) F_X is *nondecreasing*, meaning

$$t_1 \le t_2 \implies F_X(t_1) \le F_X(t_2)$$

(3) F_X is always right continuous, meaning

$$\lim_{t \to t_0^+} F_X(t) = F_X(t_0)$$

(4) For a discrete random variable, F_X is a *step-function*.

Example 3.5. Given the following probability table, find F_X.
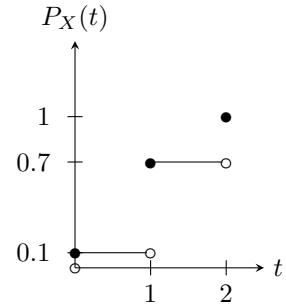
t	0	1	2
$P_X(t)$	0.1	0.6	0.3

Solution.

We have the following range:

$$F_X = \begin{cases} 0 & t \in (-\infty, 0) \\ 0.1 & t \in [0,1) \\ 0.7 & t \in [1,2) \\ 1 & t \in [2,+\infty) \end{cases}$$

Remark. For a discrete random variable, we have that

$$P_X(t) = 0 \iff F_X \text{ is continuous at } t$$

$$F_X(t) \text{ is discontinuous at } t_0 \iff P_X(t_0) = F_X(t_0) - \lim_{t \to t_0-} F_X(t)$$

◻

We will now list properties for the probability function. Let X be a random variable, and P_X its probability function.

(1) $P_X(t) \in [0,1] \quad \forall x \in \mathbb{R}$

(2) $\{x \in \mathbb{R} : P(t) \neq 0\}$ is discrete.

(3) $\sum_t P(t) = 1$ (this is the sum for $x \in \{t \in \mathbb{R} : P_X(t) \neq 0\}$, which can be finite or an infinite series)

Proposition 3.1. *Let $P : \mathbb{R} \to [0,1]$ be a function such that*

(1) $\{t \in \mathbb{R} : P(t) \neq 0\}$ is discrete.

(2) $\sum_t P(t) = 1$

Then there exists a discrete random variable X such that P is the probability function of X:

$$P = P_X$$

Example 3.6. Let P be a function such that

$$P(t) = \begin{cases} 0 & t \leq 0 \in \mathbb{Z} \\ \frac{k}{4t^2 + 2t} & t \in \mathbb{N}^* = \mathbb{N} \setminus \{0\} \end{cases}$$

Find k such that the function P is the probability function of some random variable X.

Solution.

It is enough to find $k > 0$ such that

$$\sum_{n=1}^{+\infty} \frac{k}{4n^2 + 2n} = 1$$

Thus:

$$\begin{aligned} k &= \frac{1}{\sum_{n=1}^{+\infty} 4n^2 + 2n} \\ &= \sum_{n=1}^{+\infty} \left(\frac{1}{2n} - \frac{1}{2n+1} \right) \\ &= \frac{1}{2} - \frac{1}{3} + \frac{1}{4} - \frac{1}{5} \end{aligned}$$

◻

3.2   Expected Value

Definition 3.5. Let X be a random variable, and P_X be its probability function. The **expected value** of X (if it exists) is defined as:

$$\mathbb{E}(X) = \sum_t t P_X(t)$$

Remark. The expected value of X is well-defined if $\sum_t |t| P_X(t)$ is a convergent infinite series or is finite.

Example 3.7. Consider the following probability table: Then the expected value $\mathbb{E}(X)$ is:

t	0	1	2
$P_X(t)$	0.1	0.6	0.3

$$\begin{aligned} \mathbb{E}(X) &= \sum_t t P_X(t) \\ &= 0 \times 0.1 + 1 \times 0.6 + 2 \times 0.3 \\ &= 1.2 \end{aligned}$$

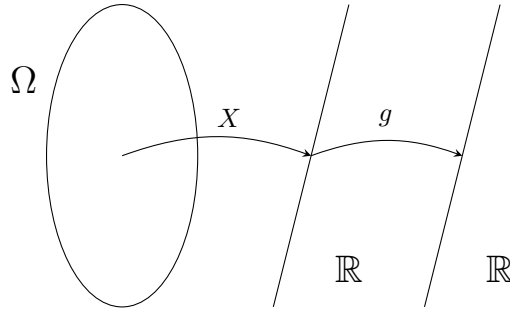**Example 3.8.** Let X be a discrete random variable such that

$$P_X = \begin{cases} 0 & n \in \mathbb{N} \setminus \{0\} \\ \frac{k}{n^2+n} & n \in \mathbb{N} \end{cases}$$

Then:

$$k = \frac{1}{\sum_{n=1}^{\infty} \frac{1}{n^2+n}} \implies \sum_{n=1}^{N} \left[\frac{1}{n} - \frac{1}{n+1} \right] = 1 - \frac{1}{2} + \frac{1}{2} - \frac{1}{3} + \cdots - \frac{1}{N+1}$$

$$\implies \sum_{t} |t| P_X(t) = \sum_{n=1}^{\infty} \frac{1}{n+1} \text{ diverges}$$

$$\therefore \mathbb{E}(X) \text{ is not defined.}$$

Proposition 3.2. *Let X be a discrete random variable, and P_X its probability function. Let $g : \mathbb{R} \to \mathbb{R}$ be a function, and $Y = g \circ X = g(X)$. Then Y is a discrete random variable as well.*

Corollary 3.3.

$$\mathbb{E}(Y) = \mathbb{E}\big(g(X)\big) = \sum_{t} g(t) P_X(t)$$

Example 3.9. Consider the following probability table:

t	0	1	2
$P_X(t)$	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{2}{3}$

Set $g(X) = X^2 = Y$, so the probability table becomes: By proposition (3.2) we have that

s	0	4
$P_X(s)$	$\frac{1}{2}$	$\frac{1}{2}$

$$\mathbb{E}(Y) = \mathbb{E}(X^2)$$

$$= \sum_t t^2 P_X(t)$$

$$= (-2)^2 + 0^2 \cdot \frac{1}{2} + (2)^2 \cdot \frac{1}{3}$$

$$= 4\left(\frac{1}{6} + \frac{1}{3}\right)$$

$$= 2$$

Recall that if X is a discrete random variable, then X is uniquely defined by its probability function:

$$\mathbb{E}\big(f(t_0)\big) = \sum_{t_0} f(t_0)P_X(t_0)$$

We now present some properties of the expected value:

(1) If $f : \mathbb{R} \to \mathbb{R}$ is a function such that $f(t_0) = k \quad \forall t_0 \in \mathbb{R}$ (k is a constant), then

$$\big(f(X)\big) = \mathbb{E}(k) = k$$

(2) Let f, g be 2 functions. Then

$$\mathbb{E}\big(f(X) + g(X)\big) = \mathbb{E}\big(f(X)\big) + \mathbb{E}\big(g(X)\big)$$

(3) For $\alpha \in \mathbb{R}$,

$$\mathbb{E}\big(af(X)\big) = \alpha\mathbb{E}\big(f(x)\big)$$

Proof.

Left as an exercise to the reader. 5 ∎

Definition 3.6. Let X be a random variable, and let $\mu_X = \mathbb{E}(X)$. The **variance** of X is defined as

$$\mathbb{V} = \mathbb{E}\left[(X - \mu_X)^2\right]$$

Example 3.10. Consider the following probability table:

t	0	1	2
$P_X(t)$	0.1	0.6	0.3

Here, $\mu_X = (0)(0.1) + (1)(0.6) + (2)(0.3) = 1.2$, and so:

$$\begin{aligned}
\mathbb{V}(X) &= \mathbb{E}\left[(X - \mu_X)^2\right] \\
&= \mathbb{E}\left[(X - 1.2)^2\right] \\
&= \sum_{t_0}(t_0 - 1.2)^2 P_X(t_0) \\
&= (1.2)^2(0.1) + (0.2)^2(0.6) + (0.8)^2(0.3) \\
&= 0.36
\end{aligned}$$

Proposition 3.4.
$$\mathbb{V}(X) = \mathbb{E}(X^2) - \left(\mathbb{E}(X)\right)^2$$

Proof.

$$\begin{aligned}
\mathbb{V}(X) &= \mathbb{E}\left((X - \mu_X)^2\right) \\
&= \mathbb{E}\left[X^2 - 2\mu_X X + \mu_X^2\right] \\
&= \mathbb{E}(X^2) - 2\mu_X \mathbb{E}(X) + \mathbb{E}(\mu_X^2) \\
&= \mathbb{E}(X^2) - 2\mu_X^2 + \mu_X^2 \\
&= \mathbb{E}(X^2) - \left(\mathbb{E}(X)\right)^2
\end{aligned}$$

■

Example 3.11.

$$\begin{aligned}
\mathbb{E}(X^2) &= (0)^2(0.1) + (1)^2(0.6) + (2)^2(0.3) \\
&= (0.6) + (1.2) \\
&= 1.8 \\
\therefore \mathbb{V}(X) &= 1.8 - (1.2)^2 \\
&= 0.36
\end{aligned}$$

Exercise 3.1. Discrete Uniform Distribution

Let k be a fixed positive number. Let $t_1, t_2, \ldots, t_k \in \mathbb{R}$ be a uniform discrete distribution, with support $\{X_1, X_2, \ldots, X_k\}$, where X_i is a random variable with a probability function:

$$P_{X_i}(t_0) = \begin{cases} \frac{1}{k} & t_0 \in \{t_1, t_2, \ldots, t_k\} \\ 0 & t_0 \notin \{t_1, t_2, \ldots, t_k\} \end{cases}$$

Prove that

(1) $\mathbb{E}(X) = \bar{t_0} = \frac{1}{k} \sum_{i=1}^{k} (t_0)_i$

(2) $\mathbb{V}(X) = \frac{1}{k} \sum_{i=1}^{k} \left((t_0)_i - \bar{t_0} \right)^2$

Here are some properties for the variance $\mathbb{V}(X)$:

(1) $\mathbb{V}(X + \beta) = \mathbb{V}(X)$

(2) $\mathbb{V}(\alpha X) = \alpha^2 \mathbb{V}(x)$

3.3 Types of Discrete Random Variables

3.3.1 Bernoulli Random Variables

We consider an experiment with two outcomes, such as a coin toss with may not be fair. Let $\Omega = \{H, T\}$ such that

$$P(H) = p \in (0, 1)$$

without loss of generality. Then, define a discrete random variable X such that $X(H) = 1$ and $X(T) = 0$. Then, the probability of X is given by the following probability table:

X	1	0
$P_X(t)$	p	$1-p$

Thus, X has a **Bernoulli Distribution**, with parameter p such that

$$X \sim \text{Ber}(p)$$

Exercise 3.2. Show that if $X \sim \text{Ber}(p)$, then $Y = 1 - X \sim \text{Ber}(1-p)$.

Exercise 3.3. Consider a random variable X with a probability function with $a, b \in \mathbb{R}$ fixed such that the following probability table holds:

X^*	a	b
$P_X(X^*)$	p	$1-p$

Show that $Y = \frac{X-b}{a-b} \sim \text{Ber}(p)$ where $X = (a-b)Y + b$ and $Y \sim Ber(p)$.

Proposition 3.5. *Let a discrete random variable be* $X \sim Ber(p)$, *then:*

(1) $\mathbb{E}(X) = p$

(2) $\mathbb{V}(X) = p(1-p)$

Proof.

(1) $\mathbb{E}(X) = 1(p) + 0(1-p) = p$

(2) $\mathbb{E}(X^2) = 1(p) + 0(1-p) = p \implies \mathbb{V}(X) = p - p^2 = p(1-p)$

■

Remark.

$$\mathbb{E}(X) = \mathbb{E}\big((a - bY + b)\big)$$
$$= (a - b)\mathbb{E}(Y) + b$$
$$= (a - b)p + b$$
$$\mathbb{V}(X) = \mathbb{V}\big((a - b)Y + b\big)$$
$$= V\big((a - b)Y\big)$$
$$= (a - b)^2\mathbb{V}(Y)$$
$$= (a - b)^2 p(1 - p)$$

3.3.2 Binomial Distribution

Assume that e have a Bernoulli trial experiment with two possible events A and B. Here:

$$P(A) = p \in (0, 1)$$
$$P(B) = 1 - p$$

We repeat the trial experiment $k \in \mathbb{N}$ times, and let $X = \{1, 2, \ldots, k\}$ be the number of events A observed. The random variable X is then said to have a **Binomial distribution** with parameters k, p, denoted:

$$X \sim \text{Bin}(k, p)$$

Proposition 3.6. *For a Bernoulli trial with* $k \in \mathbb{N}$ *repetitions such that* $n \in \mathbb{N} \setminus (k, \infty)$ *and* $P(A) = p \in (0, 1)$:

$$P_X(X = n) = C_n^k p^n (1 - p)^{k - n}$$

Proof.

We start by assuming that the trials are independent from each other. Then,

$$
\begin{aligned}
P_X(0) &= P(X = 0) \\
&= P(\underbrace{BBB \ldots B}_{k \text{ times}}) \\
&= (1 - p)^k \\
P_X(1) &= P(X = 1) \\
&= P\big((A\underbrace{BB \ldots B}_{k-1 \text{ times}}) \cap (BA\underbrace{B \ldots B}_{k-2 \text{ times}}) \cap \cdots \cap (\underbrace{BB \ldots B}_{k-1 \text{ times}}A)\big) \\
&= kp(1 - p)^{k-1} \\
P_X(2) &= P(X = 2) \\
&= P\big((AA\underbrace{B \ldots B}_{k-2 \text{ times}}) \cap (BAA\underbrace{B \ldots B}_{k-3 \text{ times}}) \cap \cdots \cap (\underbrace{B \ldots B}_{k-2 \text{ times}}AA)\big) \\
&= C_2^k p^2 (1 - p)^{k-2} \\
&\;\;\vdots \\
P_X(n) &= P(X = n) \\
&= C_n^k p^n (1 - p)^{k-n}
\end{aligned}
$$

∎

Proposition 3.7. *If $X \sim Bin(k, p)$ then*

   (1) $\mathbb{E}(X) = kp$

   (2) $\mathbb{V}(X) = kp(1 - p)$

Proof.

   (1)

$$
\begin{aligned}
\mathbb{E}(X) &= \sum_{m=1}^{k} m P_X(m) \\
&= \sum_{m=1}^{k} m C_m^k p^m (1 - p)^{k-m}
\end{aligned}
$$

28

Recall that $xC_x^y = x\frac{y!}{x!(y-k)!} = yC_{x-1}^{y-1}$, so

$$E(X) = \sum_{m=1}^{k} C_{m-1}^{k-1}p^m(1-p)^{k-m}$$

$$= k\sum_{\ell=0}^{k}-1C_{\ell}^{k-1}p^{\ell+1}(1-p)^{k-1-\ell}$$

$$= kp\sum_{\ell=0}^{k-1}C_{\ell}^{k-1}p^{\ell}(1-p)^{(k-1)-\ell}$$

$$= kp\big(p+(1-p)\big)^{k-1}$$

$$= kp$$

(2) Remark that $X^2 = X(X-1) + X$. Thus,

$$\mathbb{V}(X) = \mathbb{E}(X^2) - (kp)^2$$

$$= E\big(X(X-1)\big) + kp - k^2p^2$$

$$\therefore E\big(X(X-1)\big) = \sum_{m=0}^{k} m(m-1)C_m^k p^m(1-p)^{k-m}$$

$$= \sum_{m=2} m(m-1)C_m^k p^m(1-p)^{k-m}$$

$$\therefore m(m-1)C_m^k = m(m-1)\frac{k!}{m!(k-m)!}$$

$$= k(k-1)C_{m-1}^{k-2}$$

$$\therefore \mathbb{E}\big(X(X-1)\big) = k(k-1)\sum_{m=2}^{k} C_{m-2}^{k-2}p^m(1-p)^{k-m}$$

Let $b = m - 2$, then:

$$\mathbb{E}\big(X(X-1)\big) = k(k-1)\sum_{b=0}^{k-2} C_b^{k-2}p^{b+2}(1-p)^{k-2-b}$$

$$= k(k-1)p^2\sum_{b=0}^{k-2} C_b^{k-2}p^b(1-p)^{k-2}$$

$$= k(k-1)p^2$$

$$\therefore \mathbb{V}(X) = k(k-1)p^2 + kp - (kp)^2$$

$$= -kp^2 + kp$$

$$= kp(1-p)$$

∎

Exercise 3.4. Let $X \sim \text{Bin}(k, p)$. Find $\mathbb{E}(X^3)$.

$\left(\textit{Hint: } X^3 = \frac{X(X-1)(X-2)}{3X^2 - 2X}\right)$

Remark. $X \sim \text{Bin}(k, p) \iff Y = k - X \sim \text{Bin}(k, 1 - p)$

Example 3.12. An oil firm undergoes 10 explorations. The probability of a succesful exploration is 0.1.

$$\text{cost} = \begin{cases} \$20,000 & \text{Equipments} \\ \$30,000 & \text{for each succesful exploration} \\ \$15,000 & \text{for each failed exploration} \end{cases}$$

Find the mean and the variance of the (total) cost for the firm.

Solution.

Set a random variable $X = $ "number of succesful explorations" such that $X \sim \text{Bin}(10, 0.1)$. Let $Y = \text{cost}_{\text{Tot}}$.
Then:

$$
\begin{aligned}
Y &= 2(10^4) + 3(10^4) + 15(10^3)(10 - X) \\
&= 15,000X + 170,000 \\
\therefore \mathbb{E}(Y) &= 15,000\mathbb{E}(X) + 170,000 \\
&= 15,000kp + 170,000 \\
&= 185,000 \quad (kp = 1) \\
\therefore \mathbb{V}(Y) &= \left(15(10^3)\right)^2 \mathbb{V}(X) \\
&= 225(10^6)kp(1 - p) \\
&= 2025(10^5)
\end{aligned}
$$

Note. Variance refers to the *spread* of the data given.

3.3.3 Geometric Distribution

Recall the *Geometric series:*

$$\sum_{k=0}^{\infty} x^k = \frac{1}{1-x} \quad \forall |x| < 1$$

Consider a Bernoulli trial. We have two events, A with probability $P(A) = p$, and B with probability $P(B) = 1 - p$. We repeat the trial until the *first* instance of A occurs.

Example 3.13. Roll a die unil 2 appears. Let $X =$ " Number of trials until success (assuming that trials are independent).

$$X = \{1, 2, 3, \dots\} = \mathbb{N}^* = \mathbb{N} \setminus \{0\}$$

The probability function of X is

$$
\begin{aligned}
p(k) &= P(X = k) \\
&= P\big((\underbrace{BBB \dots B}_{k-1}, A)\big) \\
&= (1-p)^{k-1} p \qquad k \in \mathbb{N}^*
\end{aligned}
$$

X is said to have a geometric distribution with parameter p, denoted

$$X \sim \mathrm{Geo}(p)$$

Exercise 3.5. Memoryless Property

Given $k, \ell \in \mathbb{N}^*$ with $X \sim \mathrm{Geo}$ (p) prove that $P(t > k + \ell : \; > \ell) = P(X > k)$

Example 3.14. Consider

$$
\begin{aligned}
P(X > n_0) &= P(X = n_0 + 1) + P(X = n_0 + 2) + \dots \\
&= \sum_{k=n_0+1}^{\infty} P(X = k) \\
&= \sum_{k=n_0+1}^{\infty} (1-p)^{k-1} p \\
&= p \sum_{k=n_0+1}^{\infty} (1-p)^{k-1} \\
&= p \sum_{\ell=n_0}^{\infty} (1-p)^{\ell}
\end{aligned}
$$

Let $\ell = k - 1$,

$$P(X > n_0) = p(1-p)^{n_0} \sum_{\ell=n_0}^{\infty} (1-p)^{\ell - n_0}$$

Let $m = \ell - n_0$,

$$= p(1-p)^{n_0} \sum_{m=0}^{\infty} (1-p)^m$$

$$= (1-p)^{n_0}$$

The cumulative distribution function of X is

$$F_X(n_0) = P(X \le n_0)$$

$$= 1 - P(X > n_0)$$

$$= 1 - (1-p)^{n_0}$$

Proposition 3.8. *Assume that we have* $X \sim Geo(p)$*, then:*

(1) $\mathbb{E}(X) = \frac{1}{p}$

(2) $\mathbb{V}(X) = \frac{1}{p}\left(\frac{1}{p} - 1\right)$

Proof.

Recall that

$$\forall |x| < 1: \quad \frac{1}{1-x} = \sum_{k=0}^{\infty} x^k \quad \Longrightarrow$$

Thus,

$$\mathbb{E}(X) = \sum_{k=1}^{\infty} kP(X = k)$$

$$= \sum_{k=1}^{\infty} k(1-p)^{k-1}p$$

$$= p\frac{1}{\left(1 - (1-p)\right)^2}$$

$$= \frac{p}{p^2}$$

$$= \frac{1}{p}$$

Now

$$\mathbb{V}(X) = \mathbb{E}(X^2) - \frac{1}{p^2}$$

Recall that $\mathbb{E}(X^2) = \sum_{k=1}^{\infty} k^2(1-p)^{k-1}p$ and $X^2 = X(X-1) + X$, so

$$\mathbb{V}(X) = E\big(X(X-1)\big) + \frac{1}{p} - \frac{1}{p^2}$$

Remark that

$$\begin{aligned}
\mathbb{E}\big(X(X-1)\big) &= \sum_{k=2}^{\infty} k(k-1)(1-p)^{k-1}p \\
&= \frac{2(1-p)}{p^2} \\
\therefore \mathbb{V}(X) &= \frac{2}{p^2} - \frac{2}{p} + \frac{1}{p} - \frac{1}{p^2} \\
&= \frac{1}{p^2} - \frac{1}{p} \\
&= \frac{1}{p}\left(\frac{1}{p} - 1\right)
\end{aligned}$$

$\blacksquare$

3.3.4 Hypergeometric Distribution

Consider a sample space Ω of population ω with a subset $S \subset \Omega$ of size σ. We take a random sample of size m and let X be the number of subjects from the subset S included in the random sample.

Example 3.15. An urn contains 6 red balls, and 4 blue balls. We select a random sample of 7 balls. Let $X =$"number of blue balls in the sample" such that $\omega = 10$, $\sigma = 4$, and $m = 7$. Thus, $\omega - \sigma = 6$.

Note. $X \leq \sigma$ and $X \leq m$, meaning that $X \leq \min(m, \sigma)$. Also,

$$(m - X \leq m)\bigcap(m - X \leq \omega - \sigma) \implies m \geq \max(0, m + \sigma - \omega)$$

Thus, X is a random integer such that

$$\max(0, m + \sigma - \omega) \leq X \leq \min(m, \sigma)$$

The probability function of X is found by letting $k \in \mathbb{Z}$ such that

$$\max(0, m + \sigma - \omega) \leq k \leq \min(m, \sigma)$$

Thus:

$$P_X(k) = P(X = k)$$
$$= \frac{C_k^\sigma C_{m-k}^{\omega-\sigma}}{C_m^\omega}$$

X is said to have a **hypergeometric distribution** with parameters ω, m, σ and denoted:

$$X \sim \text{HyGeo}(\omega, m, \sigma)$$

Proposition 3.9. *If $X \sim HyGeo(\omega, m, \sigma)$, then*

(1) $\mathbb{E}(X) = m\frac{\sigma}{\omega}$

(2) $\mathbb{V}(X) = m\left(\frac{\sigma}{\omega}\right)\left(1 - \frac{\sigma}{\omega}\right)\left(\frac{\omega-m}{\omega-1}\right)$

3.3.5   Poisson Distribution

Recall that

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} \qquad \forall x \in \mathbb{R}$$

We will come up with a probability function $P(X)$ such that

(1) $0 \le P(X) \le 1 \quad \forall X$

(2) $\sum_{\omega \in \Omega} P(\omega) \in 1$

Thus, we define:

$$P(\omega) = \begin{cases} 0 & \omega \notin \mathbb{N} \\ e^{-t}\left(\frac{t^\omega}{\omega!}\right) & \omega \in \mathbb{N} \end{cases}$$

Note that $0 \le P(X) \le 1 \quad \forall X$, the set $\{t \in \mathbb{R} : P(t) \neq 0\}$ is discrete, and

$$\sum_t P(t) = 1$$

Therefore, P is the probability function of some discrete random variable X which is said to have a **Poisson distribution** with a parameter t, denoted

$$X \sim \text{Poisson(t)}$$

For the next two examples, assume a discrete random variable $X \sim \text{Poisson}(t = 3.2)$, where X represents the number of calls required at a call center during a given day.

Example 3.16. What is the probability that the call center receives more than 2 calls?

Solution.

$$P(X > 2) = 1 - P(X \leq 2)$$
$$= 1 - \big(P(0) + P(1) + P(2)\big)$$
$$= 1 - \left(e^{-t}\left(\frac{t^0}{0!} + \frac{t^1}{1!} + \frac{t^2}{2!}\right)\right)$$
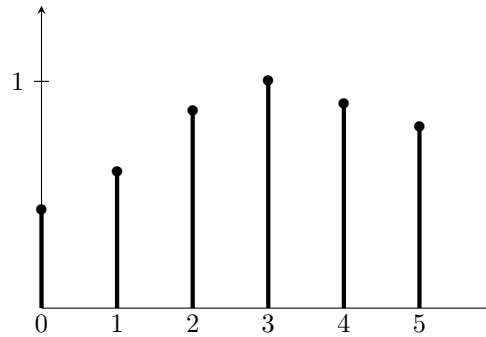
Example 3.17. What is the most likely number of calls received?

(i.e. for which n in $P(n) = e^{-t}\frac{t^n}{n!}$ is maximal. We measure)

$$\frac{P(n+1)}{P(n)} = \frac{t}{n+1}$$

Thus,

$$\frac{t}{n+1} \geq 1 \iff n+1 < t$$
$$\iff n < t - 1 = 2.2$$
$$\frac{t}{n+1} < 1 \iff n > 2.2$$

Thus, the maximal value is $n = 3$ such that $P(3) = e^{-t}\frac{t^3}{3!}$

Exercise 3.6. Let X_n for $n = 1, 2$ be the number of cars going through the entrance n in a parking lot on any given day. Assume that X_1 and X_2 are independent such that the events $\{X_1 = k\}$ and $\{X_2 = \ell\}$ are independent $\forall k, \ell$. Prove that

$$X_i \sim \text{Poisson}(t_i), \quad i = 1, 2 \implies X = X_1 + X_2$$

Solution.

Let $X \in \{0, 1, 2, \dots\} = \mathbb{N}$ with

$$\begin{aligned}
P(0) &= P\left(X_1 = 0 \bigcap X_2 = 0\right) \\
&= P(X_1 = 0) \cdot P(X_2 = 0) \\
&= e^{-t_1} \cdot e^{-t_2} \\
&= e^{-(t_1 + t_2)} \\
P(1) &= P\left[\left(X_1 = 1 \bigcap X_2 = 0\right) \bigcup \left(X_1 = 0 \bigcap X_2 = 1\right)\right] \\
&= P\left(X_1 = 1 \cap X_2 = 0\right) + P(X_1 = 0 \cap X_2 = 1) \\
&= e^{-t_1} t_1 e^{-t_2} + e^{-t_1} e^{-t_2} t_2 \\
&= (t_1 + t_2) e^{-(t_1 + t_2)}
\end{aligned}$$

Proposition 3.10. *If $X \sim Poisson(t)$, then*

    (1) $\mathbb{E}(X) = t$

    (2) $\mathbb{V}(X) = t$

Proof.

    (1)

$$\begin{aligned}
\mathbb{E}(X) &= \sum_{n=0}^{\infty} n e^{-t} \frac{t^n}{n!} \\
&= \sum_{n=1}^{\infty} e^{-t} \frac{t^n}{(n-1)!} \\
&= t e^{-t} \sum_{n=1}^{\infty} \frac{t^{n-1}}{(n-1)!}
\end{aligned}$$

    Let $(k = n - 1)$ such that:

$$\begin{aligned}
\mathbb{E}(X) &= t e^{-t} \sum_{k=0}^{\infty} \frac{t^k}{k!} \\
&= t
\end{aligned}$$

(2)

$$\mathbb{V}(X) = \mathbb{E}\big(X(X-1)\big) + \mathbb{E}(X) - \big(\mathbb{E}(X)\big)^2$$
$$= \mathbb{E}\big(X(X-1)\big) + t - t^2$$
$$= t$$

∎

Exercise 3.7. Find $\mathbb{E}(X^3)$

3.4 Moment Generating Function

Definition 3.7. Given a random variable X, a **moment generating function** of X is the function defined as

$$m_X(t) = \mathbb{E}(e^{tX})$$

More precisely, for discrete random variable we have

$$m_X(t) = \sum_X e^{tX} P_X(k)$$

Here are some of its properties:

(1) m_X is always defined for $t = 0$:

$$m_X(0) = \mathbb{E}\left(e^{0X}\right)$$
$$= \mathbb{E}(1)$$
$$= 1$$

(2) The domain of m_X is always an interval which contains zero.

(3) m_X is a unique identifier of a distribution:

$$m_X \equiv m_Y$$

Example 3.18. Consider the following probability table:

X	0	1	2
$P_X(X^*)$	0.1	0.6	0.3

The moment generating function mgf of X is

$$m_X(t) = \mathbb{E}(e^{tX})$$
$$= \sum_X e^{tX} p_X(X)$$
$$= 0.1(e^{t0}) + 0.6(e^{t1}) + 0.3(e^{t2}) \quad \forall t \in \mathbb{R}$$

Example 3.19. Consider the following probability table:

X	-5	-2	0	4
$P_X(X^*)$	$\frac{1}{7}$	$\frac{4}{7}$	$\frac{2}{7}$	$\frac{1}{7}$

$$m_X(t) = \mathbb{E}(e^{tX})$$
$$= \mathbb{E}\left[\sum_{n=0}^{\infty} \frac{(tX)^n}{n!}\right]$$
$$= \mathbb{E}\left[\sum_{n=0}^{\infty} X^n \frac{t^n}{n!}\right]$$

Treading very carefully, we have:

$$m_X(t) = \sum_{n=0}^{\infty} \frac{\mathbb{E}(X^n)}{n!} t^n \quad \forall t \in \mathbb{R}$$

Here, $m_X(t)$ is the *Maclaurin Series* of mgf. It follows that:

$$\underbrace{\mathbb{E}(X^n)}_{n^{th} \text{ moment of } X} = \frac{d^n}{dt^n}(m_X(t))\Big|_{t=0}$$

Proposition 3.11. *Let X be a random variable such that the domain D of the mgf contains an open interval centered at 0, then*

$$\mathbb{E}(X^n) = \frac{d^n}{dt^n}(m_x(t))\Big|_{t=0}$$

Example 3.20. Consider the following probability table:

X	0	1	2
$P_X(X^*)$	$\frac{1}{10}$	$\frac{6}{10}$	$\frac{3}{10}$

Thus:

$$m_X(t) = \frac{1}{10}e^{t0} + \frac{6}{10}e^{t1} + \frac{3}{10}e^{t2}$$

$$\therefore D = \mathbb{R} = (-\infty, +\infty)$$

Hence, 0 is an interior point. It follows that:

$$\mathbb{E}(X^n) = \frac{d^n}{dt^n}m_X(t)\Big|_{t=0}$$

$$\therefore \frac{d}{dt}m_X(t) = \frac{6}{10}e^t + \frac{6}{10}e^{2t}$$

$$\frac{d}{dt}m_X(0) = \frac{6}{10} + \frac{6}{10} = \frac{6}{5} = \mathbb{E}(X)$$

$$\frac{d^2}{dt^2}m_X(0) = \frac{6}{10} + 2 \cdot \frac{6}{10} = \frac{9}{5} = \mathbb{E}(X^2)$$

Therefore:

$$\mathbb{V}(X) = \mathbb{E}\big(X(X-1)\big) + \frac{6}{5} - \frac{9}{5}$$

We can now analytically redefine the *Moment Generating Function mgf* such that: $m_X(t) = \mathbb{E}(e^{tX})$ is defined from $(-\varepsilon, \varepsilon)$ for $\varepsilon > 0$. Then,

$$\mathbb{E}(X^n) = \frac{d^n}{dt^n}\big(m_X(t)\big)\Big|_{t=0}$$

Example 3.21. *mgf* of a Binomial

Let $X \sim \text{Bin}(n,p)$ such that

$$P(X = k) = C_k^n p^k (1-p)^k \qquad k = 0, 1, 2, \ldots, n$$

Thus:

$$m_X(t) = \mathbb{E}(e^{tX})$$

$$= \sum_{k=0}^{n} e^{tk}C_k^n p^k (1-p)^{n-k}$$

$$= \sum_{k=0}^{n} C_k^n (pe^t)^k (1-p)^{n-k}$$

$$= \big(pe^t + (1-p)\big)^n \quad \forall t \in \mathbb{R}$$

$$\therefore \frac{d}{dt}m_X(t) = npe^t(pe^t + (1-p))^{n-1}$$

Let $t = 0$, so

$$\frac{d}{dt} m_X(0) = np = \mathbb{E}(X)$$

Now,

$$\frac{d^2}{dt^2} m_X(t) = npe^t(pe^t + (1-p))^{n-1} + n(n-1)p^2 e^{2t}(pe^t + (1-p))^{n-2}$$

$$\therefore \frac{d^2}{dt^2} m_X(0) = np + (n)(n-1)p^2 = \mathbb{E}(X^2)$$

We may now find the variance:

$$\mathbb{V}(X) = \mathbb{E}(X^2) - \left(\mathbb{E}(X)\right)^2$$

$$= np + (n)(n-1)p^2 - (np)^2$$

$$= np + (n^2 - n)p^2 - (np)^2$$

$$= np - \cancel{(np)^2} - np^2 - \cancel{(np)^2}$$

$$= np(1-p)$$

Exercise 3.8. Find $\mathbb{E}(X^3)$ and compare it with your previous answer.

Example 3.22. Start with a fortune $m = 100$. Toss a coin such that $P(H) = p \in (0,1)$. If the result is H, the fortune doubles. If the result is T, the fortune is halved. What will be the expected fortune after n tosses? Let us denote the random variables Y as the fortune, and X the number of H after n tosses. Then,

$$X \sim \text{Bin}(n, p)$$

$$Y = m2^X \left(\frac{1}{2}\right)^{n-X}$$

Hence,

$$Y = \frac{100}{2^n} 2^{2X}$$

$$= \frac{100}{2^n} 4^X$$

$$\therefore \mathbb{E}(Y) = \frac{100}{2^n} \mathbb{E}(4^X)$$
$$= \frac{100}{2^n} \mathbb{E}(e^{\ln(4)X})$$
$$= \frac{100}{2^n} m_X(t = \ln(4))$$
$$= \frac{100}{2^n} (4p + 1 - p)^n$$
$$= \frac{100}{2^n} (3p + 1)^n$$

Example 3.23. *mgf* of a Geometric random variable

Let $X \sim \text{Geometric}(p)$, and recall that $X \in \{1, 2, \dots\} = \mathbb{N} \setminus \{0\}$. Thus,

$$P(X = k) = (1 - p)^{k-1}(p)$$
$$\therefore m_X(t) = \mathbb{E}(e^{tX})$$
$$= \sum_{k=1}^{\infty} e^{tk} P(X = k)$$
$$= \sum_{k=1}^{\infty} e^{tk} p(1 - p)^{k-1}$$

Let $k = n + 1$, then

$$m_X(t) = \sum_{n=0}^{\infty} e^{t(n+1)} (1 - p)^n p$$
$$= pe^t \sum_{n=0}^{+\infty} \left[(1 - p)e^t \right]^n$$
$$= \frac{pe^t}{1 - (1 - p)e^t} \qquad (\text{if } (1 - p)e^t < 1)$$

Therefore,

$$(1 - p)e^t < 1 \iff t < -\ln(1 - p)$$
$$\therefore t < -\ln(1 - p) \iff m_X(t) = \frac{pe^t}{1 - (1 - p)e^t}$$

Example 3.24. *mgf* **of a Poisson random variable**

Let $X \sim \text{Poisson}(\lambda)$. Thus, we have that

$$P(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}, \qquad k \in \mathbb{N}$$

$$\therefore m_X(t) = \mathbb{E}\left(e^{tX}\right)$$
$$= \sum_{k=0}^{+\infty} e^{tk} e^{-\lambda}\frac{\lambda^k}{k!}$$
$$= e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda e^t)^k}{k!}$$

This series converges $\forall t$, hence:

$$m_X(t) = e^{-\lambda} e^{\lambda e^t}$$
$$= e^{\lambda(e^t - 1)}$$

So,

$$m_X(t) = e^{\lambda(e^t - 1)} \qquad \forall t \in \mathbb{R}$$

$$\therefore \dot{m}_X(t) = \lambda e^t e^{\lambda(e^t - 1)}$$
$$\dot{m}_X(0) = \lambda = \mathbb{E}(X)$$
$$\therefore \ddot{m}_X(t) = \lambda e^t e^{\lambda(e^t - 1)} + \lambda^2 e^2 t e^{\lambda(e^t - 1)}$$
$$\ddot{m}_X(0) = \lambda + \lambda^2$$
$$= \mathbb{E}(X^2)$$

$$\therefore \mathbb{V}(X) = \lambda + \lambda^2 - (\lambda)^2 = \lambda$$

Exercise 3.9. Find $\mathbb{E}(X^3)$

Example 3.25. The number of customers making a certain payment in a store is represented by the random variable X. Let $X \sim \text{Poisson}(\lambda = 2)$. We have the following distribution of payments per customer:

Customer	Payment
1st	\$100.00
2nd	\$50.00
3th	\$25.00
$\vdots$	$\vdots$
X^{th}	\$100.00$\left(\frac{1}{2}\right)^{X-1}$

Therefore,

$$\mathbb{E}(\text{Cost}) = \mathbb{E}\left(100\left(\frac{1}{2}\right)^{X-1}\right)$$

$$= \mathbb{E}\left(50(2)^{-X}\right)$$

$$= 50 \cdot \mathbb{E}\left(2^{-X}\right)$$

$$= 50 \cdot \mathbb{E}\left(\left(\frac{1}{2}\right)^{X}\right)$$

$$= 50 \cdot \mathbb{E}\left(e^{-X\ln(2)}\right)$$

Thus, the *mgf* of X is

$$m_X(t) = e^{\lambda(e^t - 1)} \quad \forall t \in \mathbb{R}$$

$$= \mathbb{E}(e^{tX})$$

$$= 50 m_X(t = -\ln(2))$$

$$= 50 e^{\lambda(\frac{1}{2} - 1)}$$

$$= 50 e^{-\frac{\lambda}{2}}$$

We know that $\lambda = 2$, so

$$= \frac{50}{e}$$

4 Continuous Random Variables

4.1 Chebyshev Inequality

Theorem 4.1. *Let X be a random variable such that $\mathbb{E}(X) = \mu$ and $\mathbb{V}(X) = \sigma^2$. Then,*

$$P\big(|X - \mu| \geq k\sigma\big) \leq \frac{1}{k^2}$$
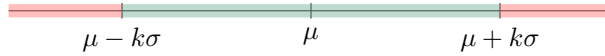
For instance, let $k = 2$ such that

$$P\big(|X - \mu| \geq 2\sigma\big) \leq \frac{1}{4}$$

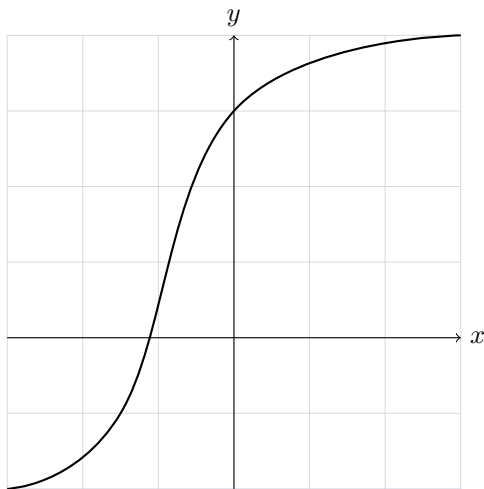Hence, more than 75% of the value of X are within 2σ from μ.

Remark. Let A, B, C be events such that $A \perp\!\!\!\perp B$, $B \perp\!\!\!\perp C$, and $A \perp\!\!\!\perp C$. Then,
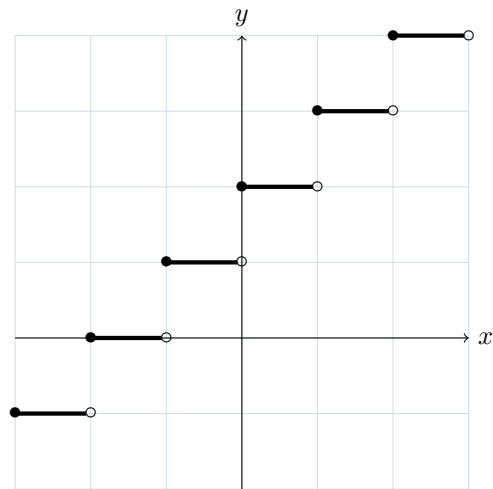
$$P(A \cap B \cap C) = P(A)P(B)P(C)$$

Definition 4.1. Recall the cumulative distribution function (*cdf*) of a random variable X is

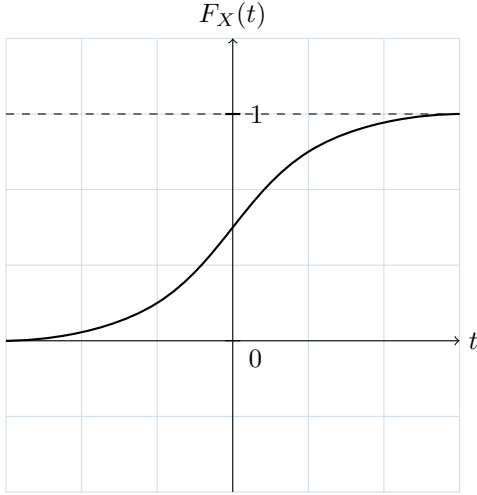$$F_X(t) = P(X \leq t)$$

For example:

A continuous function.

A discrete function.

A **continuous random variable** X is a random variable for which the *cdf* F_X is a continuous function, in addition to the fact that F_X is *non-decreasing*.

$$\lim_{t \to -\infty} F_X(t) = 0$$
$$\lim_{t \to +\infty} F_X(t) = 1$$

$\left.\vphantom{\begin{array}{c}a\\b\end{array}}\right\}$ F_X is non-decreasing

Proposition 4.2. *If F_X is the cumulative distribution function (cdf) of a continuous random variable X, then F_X is differentiable everywhere, except at certain subsets of points.*

Let us define

$$f_X(t) = \frac{d}{dt} F_X(t) = \dot{F}_X(t)$$

We can denote the following properties for f_X:

(1)

$$f_X(t) \geq 0 \qquad \forall t$$

(2)

$$
\begin{aligned}
\int_{-\infty}^{+\infty} f_X(t)dt &= \lim_{\substack{a \to -\infty \\ b \to +\infty}} \int_a^b f_X(t)dt \\
&= \lim_{\substack{a \to -\infty \\ b \to +\infty}} \int_a^b \dot{F}_X(t)dt \\
&= \lim_{\substack{a \to -\infty \\ b \to +\infty}} (F_X(b) - F_X(a)) \\
&= 1 - 0 \\
&= 1
\end{aligned}
$$

Definition 4.2. If X is a continuous variable, then

$$f_X(t) = \dot{F}_X(t)$$

is called the **probability density function** (pdf) of X.

Example 4.1. Set $f(t) = \frac{k}{1+t^2}$ whenever $t \in \mathbb{R}$. For what values of k is $f(t)$ the pdf of a continuous random variable?

Solution.

We must have $k \geq 0$, and also

$$\int_{-\infty}^{+\infty} \frac{k}{1+t^2} dt = 1$$

$$\therefore k \int_{-\infty}^{+\infty} \frac{1}{1+t^2} dt = 1$$

$$\therefore f_X(\pi) = \underbrace{\left(\frac{1}{\pi}\right)}_{\substack{Normalizing \\ constant}} \cdot \underbrace{\left(\frac{1}{1+t^2}\right)}_{Kernel} \quad \forall \underbrace{t \in \mathbb{R}}_{Support}$$

The above is the probability distribution function (*pdf*) of the **Cauchy Distribution**. The cumulative distribution function (*cdf*) of the Cauchy Distribution is

$$-F_X(t) = \int_{-\infty}^{t} f_X(s) ds$$

$$= \frac{1}{\pi} \int_{-\infty}^{t} \frac{1}{1+s^2} ds$$

$$= \frac{1}{\pi} \arctan(t) + \frac{\pi}{2}$$

Example 4.2. The following plot represents the probability distribution function (*pdf*) of a continuous random variable X. Find $P\left(-\frac{1}{2} \leq X \leq \frac{1}{4}\right)$
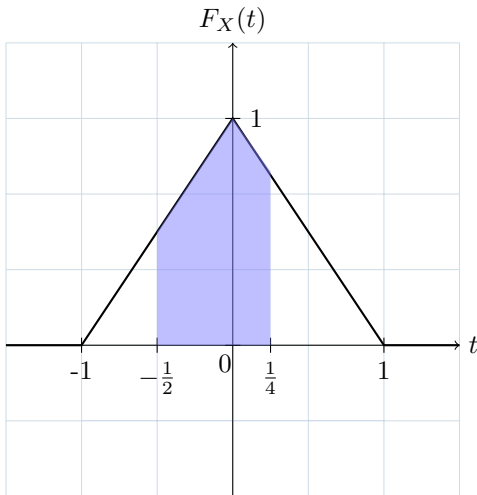
We have

$$P\left(-\frac{1}{2} \leq X \leq \frac{1}{4}\right) = P\left(X < \frac{1}{4}\right) - P\left(X < -\frac{1}{2}\right)$$

$$= F_X\left(\frac{1}{4}\right) - F_X\left(-\frac{1}{2}\right)$$

$$= \int_{-\frac{1}{2}}^{\frac{1}{4}} f_X(s) ds$$

$$= 1 - \left(\frac{1}{8} + \frac{9}{32}\right)$$

$$= \frac{19}{32}$$