# University of Hawai'i

# Probability (M471)

## Lecture Notes

Created by: Pierre-Olivier Parisé

Fall 2023

# A Note to the Reader

These notes were written during the Fall 2023 semester while I was teaching the course *Probability* (M471) at the University of Hawai'i at Mānoa. They were inspired by two main resources:

- *Probability: An Introduction* by Geoffrey Grimmett and Dominic Welsh.
- *Mathematical Statistics with Applications* by Dennis D. Wackerly, William Mendenhall III, and Richard L. Scheaffer.

I aimed to model these notes on a class I wish I had at the undergraduate level. One thing I wished I had seen in my undergraduate probability class is how closely probability is tied to measure theory. This is why, in Chapter A, I provide a gentle introduction to Probability Space and how it is used to construct a probability measure. I tried to make this encounter with measure theory as friendly as possible and hope I have accomplished that goal.

Chapters B and C introduce the basic concepts of probability and discrete random variables. The last four chapters (D, E, F, G) introduce continuous random variables and their vector-valued analogues.

I hope these notes will help readers grasp the essence of probability.

Pierre-Olivier Parisé<br>
Honolulu, O'ahu, Hawai'i<br>
December 2023

# Contents

# Chapter A: Basic Probability

## A.1 Sample Space

Definition A.1 All possible outcomes of an experiment is called the sample space. It is denoted by $S$.

Example A.1

a) Say we flip a coin twice. The result of a flip is head ("h") or is tail ("t"). Then, the possible outcomes are $S = \{(t,t),(t,h),(h,t),(h,h)\}$. Notice that we could also write $S = \{tt, th, ht, hh\}$ to record all the possible outcomes.

b) Suppose we are interested in the number of reactions in a chemical process. Then the sample space will be $S = \{\text{All non-negative integers}\} = \{1, 2, \ldots\}$.

c) If the outcome of an experiment is the height of the waves in Waikiki, then $S = [0, \infty)$ with the measurements in feet. *[Hopefully, it is not negative!]*

Note:

- A sample space is finite if the number of outcomes in the sample space is finite.
- A sample space is discrete (or countable) if the outcomes in $S$ can be listed. So a finite sample space is a discrete sample space.
- Otherwise, the sample space is uncountable.

## A.2 Event Space

Example A.2

a) Taking the experiment from Example A.1.a), if $A = \{(h,h),(h,t)\}$, then $A$ can be seen as the realization of the event that a head occurs on the first coin.

b) Based on the experiment from Example A.1.b), if $A = \{200, 201, 202, 203, 204, 205\}$, then $A$ is the event that the number of reactions in a chemical process is between 200 and 205. If $A_i = \{i\}$, for some $i \geq 0$, then the event "There are 200 or more reactions" is an (infinite) combination of the events $A_i$, for $i \geq 200$, written as $\cup_{i=200}^{\infty} A_i$.

Unfortunately, in the uncountable case, the notion of an event is way more delicate. An intuitive fact would be that the probability of a single real number to occur is zero (in Example A.1.c), that the wave height will be exactly 2 feet is almost impossible!). If we allow the family of events to be all subsets, Banach and Kuratowski proved in 1929 that there is no notion of probability, call it $P$, defined on all subsets of the interval $[0, 1]$ satisfying $P(\{a\}) = 0$ for every

$0 \leq a \leq 1$. For this reason, we have to restrict the family of events and there will be events for which we can't attach a notion of probability.

**DEFINITION A.2** If $S$ is a sample space which is finite, then the event space $\mathcal{A}$ is a family of subsets of $S$ satisfying the following axioms:

a) $\mathcal{A}$ contains $\varnothing$ or $\mathcal{A}$ contains $S$.

b) If $A$ is an event (belongs to $\mathcal{A}$), then its complement $\overline{A}$ is also an event.

c) If $A$ and $B$ are two events, then $A \cup B$ is also an event.

**Note:**

- The elements of an event $A$ are called outcomes.
- An event $A$ occurs if one of its outcomes is observed.
- We will see later how to modify Definition A.2 to incorporate uncountable sample space.

**EXAMPLE A.3** We toss a regular 6-faced dice. We observe the number of dots on the upper face of the die.

a) $S = \{⚀, ⚁, ⚂, ⚃, ⚄, ⚅\}$.

b) $\mathcal{A} = \{\varnothing, \{⚀\}, \{⚁\}, \ldots, \{⚀, ⚁\}, \ldots, \{⚀, ⚁, ⚂\}, \ldots, S\}$ is an event space (all subsets).

c) Let $A = \{⚀, ⚁\}$, $B = \{⚀, ⚂\}$, and $C = \{⚁, ⚃, ⚅\}$. Then, $A \cup B = \{⚀, ⚁, ⚂\}$, $A \cap B = \{⚀\}$, and $\overline{A} = \{⚂, ⚃, ⚄, ⚅\}$ are all events.

d) $B \cap C = \varnothing$, so $B$ and $C$ are mutually exclusive. $A$ and $B$ are not mutually exclusive because $A \cap B = \{⚀\}$. $A$ and $C$ are not mutually exclusive because $A \cap C = \{⚁\}$.

**Note:**

- If $S$ is finite, then the family $\mathcal{A}$ of all subsets of $S$ is an event space. In this case, this will be the default event space and all possible subsets of $S$ will be considered events.
- In particular, every subset with one outcome will be events. They are called atomic events or simple events. They can be used to decompose more complicated events.

## A.3 Axioms of Probability

**EXAMPLE A.4** We throw a 6-faced dice, so $S = \{⚀, ⚁, ⚂, ⚃, ⚄, ⚅\}$. The dice is fair, meaning it is equally likely the die will land on any face. We define a function $P : 2^S \to \mathbb{R}$ such that

$$P(A) = \frac{|A|}{6} \quad (A \subset S).$$

We notice that

a) For any event $A \subset S$, then $|A| \leq 6$ and so $0 \leq P(A) \leq 1$.

b) In particular, we have $P(S) = 1$.

c) Finally, if $A = \{\text{the outcome is divisible by } 3\}$, then $A = \{⚂, ⚅\}$ and

$$P(A) = P(\{⚂\}) + P(\{⚅\}).$$

In other words, if $A$ and $B$ are two events with $A \cap B = \varnothing$, then $P(A \cup B) = P(A) + P(B)$.

We call $P$ a probability measure because it satisfies the three axioms of a probability. $\triangle$

**Definition A.3** Let $S$ be a sample space and $\mathcal{A}$ be an event space. A function $P : \mathcal{A} \to \mathbb{R}$ is a probability measure if

a) [1] For any event $A$, $0 \leq P(A) \leq 1$.

b) [2] We have $P(S) = 1$.

c) [3] If $A$ and $B$ are events that are mutually exclusive (meaning $A \cap B = \varnothing$), then

$$P(A \cup B) = P(A) + P(B).$$

Note:

- The Axioms of Probability were introduced by Andrei Kolmogorov in 1933 in his book *Foundations of the Theory of Probability.*
- A probability space is a triplet $(S, \mathcal{A}, P)$, where $S$ is a sample space, $\mathcal{A}$ is an event space, and $P$ is a probability measure.
- If $A$, $B$, $C$ are three mutually exclusive events, then $P(A \cup B \cup C) = P(A) + P(B) + P(C)$.

**Theorem A.1** If $(S, \mathcal{A}, P)$ is a probability space and $A$ is an event, then $P(\overline{A}) = 1 - P(A)$. In particular, $P(\varnothing) = 0$.

*Proof.* Assume that $(S, \mathcal{A}, P)$ is a probability space and that $A$ is an event. We can write $S = A \cup \overline{A}$, with $A \cap \overline{A} = \varnothing$. From property b) in the definition of a probability measure, we have $P(S) = 1$. But from property c) of a probability measure, we have

$$1 = P(A \cup \overline{A}) = P(A) + P(\overline{A})$$

and therefore $P(\overline{A}) = 1 - P(A)$. Since $\varnothing = \overline{S}$ and $P(S) = 1$, we have $P(\varnothing) = 1 - P(S) = 0$. $\square$

## A.4 Computing Probabilities in the Finite Case

We assume that the event space, in the examples below, is the family of all subsets of the sample space.

**Example A.5** A manufacturer has five seemingly identical computer terminals available for shipping. Unknown to her, two of the five are defective. A particular order calls for two of the terminals and is filled by randomly selecting two of the five that are available. Let $A$ denote the event that the order is filled with two non-defective terminals. Find $P(A)$.

**Solution.**

① Define the sample space. Let $D_1$ and $D_2$ stand for defective terminals and $F_1$, $F_2$ and $F_3$ stand for functioning terminals. Then, the sample space

$$\begin{aligned} S = &\{\{D_1, D_2\}, \{D_2, F_1\}, \{F_1, F_2\}, \{F_2, F_3\}, \{D_1, F_1\}, \{D_2, F_2\}, \{F_1, F_3\}, \\ &\{D_1, F_2\}, \{D_2, F_3\}, \{D_1, F_3\}\}. \end{aligned}$$

[1] Part a) states that the probability of an outcome to be in $A$ is a number $P(E)$ between 0 and 1.

[2] Part b) states that, with probability 1, the outcome will be in the sample space $S$.

[3] Part c) states that, for a sequence of mutually exclusive events the probability of at least one of these events occurring is just the sum of their respective probabilities.

② Define the probability measure. Since all terminals are chosen at random, each terminal is selected equally likely. Therefore, we define the value of the probability measure $P$ on the atomic events:

$$P(\{D_1, D_2\}) = P(\{D_2, F_1\}) = \cdots = P(\{D_1, F_3\}) = \frac{1}{|S|} = \frac{1}{10}.$$

We can check that this definition gives rise to a probability measure.

③ Decompose the event into atomic events. We see that

$$A = \{\{F_1, F_2\}, \{F_2, F_3\}, \{F_1, F_3\}\}.$$

Therefore, using the third property of the probability measure, we get

$$P(A) = P(\{F_1, F_2\}) + P(\{F_2, F_3\}) + P(\{F_1, F_3\}) = \frac{3}{10} = 0.3. \qquad \triangle$$

**Note:**

- When every outcome in a finite sample space $S$ is equally likely to occur, then each atomic event $A$ has $P(A) = 1/|S|$.
- Therefore, we have $P(A) = |A|/|S|$ for any $A \subset S$.
- When $S$ is finite, finding the probability of an event is usually recast as a counting problem.

### Combinatorial Tools for Counting

EXAMPLE A.6 Consider 20 randomly selected people in a room. Ignoring leap years and assuming that there are only 365 possible distinct birthdays that are equiprobable, what is the probability that each person in the 20 has a different birthday?

**Solution.**

① If the person are given an order, starting from 1 and ending at 20, then an outcome will be a list of 20 numbers ranging from 1 to 365. Therefore, using the product rule, we have $|S| = (365)^{20}$.

② Since each birthday are equally likely, then the value of the probability measure $P$ on an atomic event $A$ (a list of 20 numbers ranging from 1 to 365) is

$$P(A) = \frac{1}{(365)^{20}}$$

③ We won't list all the outcomes of $A$, but we will find $|A|$ instead. The first person can have his birthday on one of the 365 days, the second person can have his birthday on one of the 364 remaining days, ..., the 20th person can have his birthday on one of the 346 remaining days. So,

$$|A| = 365 \times 364 \times \cdots \times 346$$

and therefore, using the Property c) of a probability measure:

$$P(A) = \frac{365 \times 364 \times \cdots \times 346}{(365)^{20}} \approx 0.5885. \qquad \triangle$$

**Product Rule:** If $G_1$, $G_2$, ..., $G_N$ are sets of objects with $|G_1|$, $|G_2|$, ..., $|G_N|$ elements, then it is possible to form $|G_1| \times |G_2| \times \cdots \times |G_N|$ lists of objects containing one element from each set.

EXAMPLE A.7 Seven horses identified by the numbers 1, 2, 3, 4, 5, 6, and 7 are in a race and any of the horses have an equal chance to win the race. What is the probability that the horse #3 wins the race?

**Solution.**

① The sample space is formed for all list of 7 horses, representing the positions of the horses at the finish line. There are $P_7^7 = 7!$ such lists.

② Since every horse is equally likely to win, the probability of each possible list is $1/7!$.

③ Let $A$ be the event that the horse #3 wins the race. This means $A = \{(3, a, b, c, d, e, f) : a, b, c, d, e, f$ is the number of the 6 remaining horses$\}$. Since the first horse in every outcome from $A$ is fixed, there are only six remaining horses. Therefore, $|A| = 6!$. Therefore, $P(A) = 6!/7! = 1/7 \approx 0.1429$. △

**Permutation** is an ordered arrangement of $r$ distinct objects taken from $n$ objects. The number of ways of ordering $n$ distinct objects taken $r$ at a time will be designated by the symbol $P_r^n$. We have

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

**Multinomial Coefficients** gives the number of ways of partitioning $n$ distinct objects into $k$ distinct groups containing $n_1$, $n_2$, ..., $n_k$ objects, respectively, where each object appears in exactly one group and $n_1 + n_2 + \cdots + n_k = n$, is

$$\binom{n}{n_1 n_2 \cdots n_k} = \frac{n!}{n_1! n_2! \cdots n_k!}.$$

EXAMPLE A.8 A company wants to select two applicants for a job out of 5. These applicants are ranked according to their competency, 1 is the most competent, and 5 is the most incompetent. This ranking is unknown to the company and the company assume that it is equally likely to select a candidate. Let $A$ denote the event that exactly one of the two best applicants appears in a selection of two applicants. Find $P(A)$.

**Solution.**

① Denote $c_1$, $c_2$, $c_3$, $c_4$, and $c_5$ the ranking of the candidates. Then $S$ is composed of $\binom{5}{2} = 10$ elements. We can list them

$$S = \{\{c_1, c_2\}, \{c_1, c_3\}, \{c_1, c_4\}, \{c_1, c_5\}, \{c_2, c_3\}, \{c_2, c_4\}, \{c_2, c_5\}, \{c_3, c_4\}, \{c_3, c_5\}, \{c_4, c_5\}\}.$$

② Since every candidate is equally likely to be selected, this means that the probability of two candidates to be selected is $1/\binom{5}{2} = 1/10$.

③ Let's count the number of elements in the event $A$. One candidate should be selected from the best 2, and the second one should be selected from the three others, so $|A| = \binom{2}{1}\binom{3}{1} = 2 \times 3 = 6$. Therefore, $P(A) = \frac{6}{10} = 0.6$. △

**Combinations** are an arrangement, where the order of the group of $r$ objects is unimportant. We use the symbol $C_r^n$ or $\binom{n}{r}$. We have

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

**When The Odds are not Equally Likely**

EXAMPLE A.9 The odds are two to one that, when $A$ and $B$ play tennis, $A$ wins. Suppose that $A$ and $B$ play two matches. What is the probability that $A$ wins at least one match?

**Solution.**

① The sample space will be the pairs of letters $A$ and $B$. For example, $(A, A)$ means that $A$ won match one and two. So, the sample space $S$ has 4 elements

$$S = \{(A, A), (A, B), (B, A), (B, B)\}.$$

② The odds are $2:1$ that $A$ wins over $B$. So, this means

$$P(\{(A, A)\}) = \frac{4}{9}, P(\{(A, B)\}) = P(\{(B, A)\}) = \frac{2}{9}, P(\{(B, B)\}) = \frac{1}{9}.$$

③ We see that $A = \{(A, A), (A, B), (B, A)\}$. Therefore,

$$P(A) = P(\{(A, A)\}) + P(\{(A, B)\}) + P(\{(B, A)\}) = \frac{4}{9} + \frac{4}{9} = \frac{8}{9} \approx 0.89. \qquad \triangle$$

THEOREM A.2 Given the probability space $(S, 2^S, P)$ is a probability space, where $S$ is finite, there exists $p_1$, $p_2$, ..., $p_N$, with $N = |S|$, such that

$$P(A_i) = p_i$$

for every atomic even $A_i$, $i = 1, 2, \ldots, N$ and

$$p_1 + p_2 + \cdots + p_N = 1.$$

*Proof.* Let $S = \{s_1, s_2, \ldots, s_N\}$. Therefore, each atomic event is $A_i = \{s_i\}$, for $i = 1, 2, \ldots, N$. From the properties of a probability, we know that $p_i := P(A_i)$ is between 0 and 1, for $i = 1, 2, \ldots, N$. Since $\cup_{i=1}^N A_i = S$, $A_i \cap A_j = \varnothing$ when $i \neq j$ and $P(S) = 1$, we obtain the desire conclusion. $\square$

**Note:**

- The converse is also true. Given $N \geq 1$ numbers $p_1$, $p_2$, ..., $p_N$ between 0 and 1, there is a finite probability space $(S, \mathcal{A}, P)$ such that $|S| = N$ and $P(A_i) = p_i$, for every atomic event $A_i$, $i = 1, 2, \ldots, N$. Moreover, $\mathcal{A}$ can be taken to be the family of all subsets of $S$.
- This means, in the case of a finite sample space $S$, it is not necessary to construct explicitly the probability space.

## A.5 Probability Space For Infinite Sample Spaces

**Example A.10** A coin is tossed infinitely many times so that $S$ is the set of infinite strings of the letter t's (tail) and h's (head). For example, the string $hhthh\cdots$ means that the coin landed on head, head, tail, head, head, etc. Let $A_i$ be the event that the coin lands on head on the $i$-th toss. Then the event $A = \cup_{i=1}^{\infty} A_i$ means the coin lands head for at least one toss.

In the previous example:

- If $A_1$, $A_2$, ... are sets, then the new set $\cup_{i=1}^{\infty} A_i$ is defined to be the set for which at least one element is in some $A_i$, that is

$$\cup_{i=1}^{\infty} A_i = \{x \in S \,:\, \exists i \text{ such that } x \in A_i\}.$$

- If $A_1$, $A_2$, ... are sets, then the new set $\cap_{i=1}^{\infty} A_i$ is defined to be the set of all common elements of every $A_i$, that is

$$\cap_{i=1}^{\infty} A_i = \{x \in S \,:\, \forall i,\, x \in A_i\}.$$

For infinite sample space, the definition of an event space must be modified to allow infinite unions of events.

**Definition A.4** Let $S$ be a sample space and let $\mathcal{A}$ be a family of subsets of $S$. Then $\mathcal{A}$ is called an event space if

a) $\varnothing \in \mathcal{A}$ (or $S \in \mathcal{A}$).

b) If $A$ is an event (so an element of $\mathcal{A}$), then $\overline{A}$ is also an event.

c) If $A_1$, $A_2$, ... are events, then $\cup_{i=1}^{\infty} A_i$ is also an event.

Therefore, for a function to be a probability measure, finite unions must be replaced by infinite unions.

**Definition A.5** Let $S$ be a sample space and let $\mathcal{A}$ be an event space. A function $P : \mathcal{A} \to \mathbb{R}$ is a probability measure if

a) For any event $A$, $0 \leq P(A) \leq 1$;

b) $P(S) = 1$;

c) If $A_1$, $A_2$, ... are mutually disjoint events (meaning that $A_i \cap A_j = \varnothing$, for $i \neq j$), then

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

**Example A.11** Let $S = \{1, 2, 3, \ldots\}$ the set of natural numbers and let $\mathcal{A} = 2^S$. Let $p_1$, $p_2$, ... be a sequence of numbers in $[0, 1]$ such that $\sum_{i=1}^{\infty} p_i = 1$. Definite the function $P$ by

$$P(A) = \sum_{i \in A} p_i \quad (A \subset S).$$

Then $P$ is a probability measure and $(S, \mathcal{A}, P)$ is a probability space. Examples of concrete $p_i$'s would be $p_i = \frac{1}{i(i+1)}$.

**Note:**

- When $S$ is a discrete space, the family $2^S$ can be used to create a probability space $(S, 2^S, P)$, for some probability measure $P$.
- The third axiom is called, in the literature, $\sigma$-additivity of the probability measure.

THEOREM A.3 If $A$ and $B$ are events from a probability space $(S, \mathcal{A}, P)$ and $A \subset B$, then $P(B \cap \overline{A}) = P(B) - P(A)$.

*Proof.* Assume that $(S, \mathcal{A}, P)$ is a probability space and that $A$ and $B$ are two events with $A \subset B$. Then $B = A \cup (B \cap \overline{B})$ and $A \cap B \cap \overline{A} = \varnothing$. Therefore, with $A_1 = A$, $A_2 = B \cap \overline{A}$, and $A_i = \varnothing$, for $i = 3, 4, \ldots$, we obtain

$$P(B) = P(A) + P(B \cap \overline{A}) \quad \Rightarrow \quad P(B) - P(A) = P(B \cap \overline{A}). \qquad \square$$

**Continuity of Probability Measures**

THEOREM A.4 Let $(S, \mathcal{A}, P)$ be a probability space. Let $A_1$, $A_2$, ... be events such that $A_i \subset A_{i+1}$, for every $i \geq 1$. Then

$$P\Big(\bigcup_{i=1}^{\infty} A_i\Big) = \lim_{n\to\infty} P(A_n).$$

*Proof.* Define a list of events as followed: $B_1 = A_1$, $B_2 = A_2 \cap \overline{A}_1$, $B_3 = A_3 \cap \overline{A}_2$, ..., $B_i = A_i \cap \overline{A}_{i-1}$, for $i \geq 2$. Then, $B_i \cap B_j = \varnothing$ for $i \neq j$ and

$$\bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{\infty} B_i.$$

Therefore,

$$P\Big(\bigcup_{i=1}^{\infty} A_i\Big) = P\Big(\bigcup_{i=1}^{\infty} B_i\Big) = \sum_{i=1}^{\infty} P(B_i).$$

But $B_1 = A_1$ so that $P(B_1) = P(A_1)$ and $B_i = A_i \cap \overline{A}_{i-1}$ when $i \geq 2$, so that $P(B_i) = P(A_i) - P(A_{i-1})$, by Theorem A.3. From the definition of series:

$$\begin{aligned}
\sum_{i=1}^{\infty} P(B_i) = \lim_{n\to\infty} \sum_{i=1}^{n} P(B_i) &= \lim_{n\to\infty} P(B_1) + P(B_2) + \ldots + P(B_n) \\
&= \lim_{n\to\infty} P(A_1) + P(A_2) - P(A_1) + \ldots + P(A_n) - P(A_{n-1}) \\
&= \lim_{n\to\infty} P(A_n).
\end{aligned}$$

This ends the proof. $\square$

# Chapter B: Conditional Probability and Independence

## B.1 Conditional Probabilities

**Example B.1** Suppose two dice are tossed and each of the 36 outcomes are equally likely to occur. If the first die landed on a ⚂, then, given this information, what is the probability that the sum of the 2 dice equals 8?

Let $A$ be the event "sum of the two dice equals 8". Let $B$ be the event "the first die landed on ⚂". Then

$$\begin{aligned} A &= \{(\text{⚁}, \text{⚅}), (\text{⚂}, \text{⚄}), (\text{⚃}, \text{⚃}), (\text{⚄}, \text{⚂}), (\text{⚅}, \text{⚁})\} \\ B &= \{(\text{⚂}, \text{⚀}), (\text{⚂}, \text{⚁}), (\text{⚂}, \text{⚂}), (\text{⚂}, \text{⚃}), (\text{⚂}, \text{⚄}), (\text{⚂}, \text{⚅})\}. \end{aligned}$$

Knowing that $B$ has occurred, then the probability for $A$ to occur within $B$ is

$$P(A \text{ knowing } B) = \frac{|A \cap B|}{|B|} = \frac{|\{(\text{⚂}, \text{⚄})\}|}{6} = \frac{1}{6}.$$

If we want to compute the probability from the complete sample space,

$$P(A \text{ knowing } B) = \frac{|A \cap B|/|S|}{|B|/|S|} = \frac{P(A \cap B)}{P(B)}. \qquad \triangle$$

**Definition B.1** Let $(S, \mathcal{A}, P)$ be a probability space and $A$, $B$ be events such that $P(B) > 0$. The conditional probability of $A$ given that $B$ has occurred[1] is denoted by $P(A|B)$ and is given by the following formula:

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

**Example B.2** An urn contains 10 blue balls, 5 red balls, and 10 green balls. A ball is chosen at random from the urn and it is noted that it is not one of the green balls. What is the probability that it is red?

**Solution.**

① The sample space is $S = \{b, r, g\}$, where $b$ stands for blue, $r$ stands for red and $g$ stands for green.

② We have $P(\{b\}) = \frac{10}{25} = 0.4$, $P(\{r\}) = \frac{5}{25} = 0.2$ and $P(\{g\}) = 0.4$.

[1] We also say the probability of $A$ conditional to $B$.

③ Let $A$ be the event "the ball picked is red" and let $B$ be the event "the ball picked is green". The probability we are looking for is $P(A|\overline{B})$. By the formula, we have

$$P(A|\overline{B}) = \frac{P(A \cap \overline{B})}{P(\overline{B})} = \frac{P(A)}{1 - P(B)} = \frac{1/5}{3/5} = \frac{1}{3}. \qquad \triangle$$

**Note:** Instead of applying (B.1), it is sometimes easier to work with the reduced sample space as in Example B.1. *[Show them with the last example.]*

**Corollary B.1** If $(S, \mathcal{A}, P)$ is a probability space and $A, B$ are events, then

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

**Example B.3** Celine is undecided as to whether to take a French or a Chemistry class. The probability of obtaining an $A$ grade if Celine takes French is $1/2$ and if Celine takes Chemistry is $2/3$. If Celine decides to base her course choice on a flip of a fair coin, what is the probability that Celine gets an $A$ in Chemistry?

**Solution.**

① Let $f$ stands for "French class" and $c$ stands for "Chemistry class". We combine this with $A$ which stands for an $A$ grade and $O$ which stands for other grades. So the sample space is $S = \{(f, A), (f, O), (c, A), (c, O)\}$.

② Let $F$ be the event "Celine takes French" and $C$ be the event "Celine takes Chemistry". Then, we have $P(F) = P(C) = 1/2$. We can still recover the probabilities of each atomic event because we know the conditional probabilities!

③ Let $H$ be the event that Celine receives an $A$ grade in whatever class she takes. We try to find the probability of $\{(c, A)\} = H \cap C$. We know that $P(H|C) = 2/3$, so we have

$$P(H \cap C) = P(C)P(H|C) = (1/2)(2/3) = \frac{1}{3}. \qquad \triangle$$

## B.2 Bayes' Formula

**Theorem B.2** If $(S, \mathcal{A}, P)$ is a probability space and $A, B$ are events with $P(B) > 0$, then

$$P(A) = P(A|B)P(B) + P(A|\overline{B})P(\overline{B}). \tag{B.2}$$

*Proof.* We know that $P(A|B) = P(A \cap B)/P(B)$ and $P(A|\overline{B}) = P(A \cap \overline{B})/P(B)$. Therefore,

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

But $(A \cap B) \cup (A \cap \overline{B}) = A$ and $A \cap B$, $A \cap \overline{B}$ are mutually exclusive. This implies

$$P(A \cap B) + P(A \cap \overline{B}) = P(A).$$

This completes the proof. □

**Example B.4** An insurance company believes that people can be divided into two classes: those who are accident prone and those who are not. Their statistics show that:

- If a person is accident-prone, this person will have an accident at some time within a fixed 1-year period with probability .4.
- If a person is non-accident-prone, this person will have an accident at some time within a fixed 1-year period with probability .2.

We assume that 30% of the population is accident prone.

a) What is the probability that a new policyholder will have an accident within a year of purchasing a policy?

b) If a new policyholder has an accident within a year of purchasing a policy, what is the probability that the person is accident prone?

**Solution.**

a) Let $A$ denote the event "new policyholder will have an accident within a year". Let $B$ be the event "new policyholder is accident-prone". Using (B.2), we have

$$\begin{aligned} P(A) &= P(B)P(A|B) + P(\overline{B})P(A|\overline{B}) \\ &= (0.3)(0.4) + (0.7)(0.2) = 0.26. \end{aligned}$$

b) We are looking for the probability $P(B|A)$. Using the definition, we have

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{P(B)P(A|B)}{P(A)} = \frac{(0.3)(0.4)}{(0.26)} = \frac{6}{13} \approx 0.4615. \qquad \triangle$$

**THEOREM B.3** Let $(S, \mathcal{A}, P)$ be a probability space. If $B$ is an event with $P(B) > 0$, then the function $Q : \mathcal{A} \to \mathbb{R}$ defined by

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

is a probability measure.

*Proof.* a) We first show that $0 \leq P(A) \leq 1$. If $A$ and $B$ are events, then $A \cap B \subset B$ and therefore $Q(A) = P(A|B) = P(A \cap B)/P(B) \leq 1$.

b) We then show that $P(S) = 1$. Since $S \cap B = B$, we have $Q(S) = P(S|B) = P(B)/P(B) = 1$.

c) Finally we show that $Q$ is $\sigma$-additivity. Let $A_1, A_2, \ldots$ be mutually exclusive events. Then $A_1 \cap B$, $A_2 \cap B$, $\ldots$ are also mutually exclusive events and $\left(\cup_{i=1}^{\infty} A_i\right) \cap B = \cup_{i=1}^{\infty} (A_i \cap B)$. Therefore,

$$\begin{aligned} Q\Big(\bigcup_{i=1}^{\infty} A_i\Big) = P\Big(\bigcup_{i=1}^{\infty} A_i \Big| B\Big) = \frac{P\Big(\cup_{i=1}^{\infty}(A_i \cap B)\Big)}{P(B)} &= \frac{\sum_{i=1}^{\infty} P(A_i \cap B)}{P(B)} \\ &= \sum_{i=1}^{\infty} \frac{P(A_i \cap B)}{P(B)} \\ &= \sum_{i=1}^{\infty} P(A_i|B) = \sum_{i=1}^{\infty} Q(A_i). \end{aligned}$$

This completes the proof. $\square$

**Example B.5** In a show, you are asked to choose between three doors disposed in front of you at random. Behind one of the door is a big money price and behind the other two, nothing... You choose to open the first door, but before the game host opens the door, he opens at random one of the other two doors, say door 3, to show you there is nothing behind it. He then asks you: Do you want to switch door? What should you do, switch door or not?

**Solution.** Let $A$ be the event "Door 1 has **the price behind it**." and let $B$ be the event "the host shows a door **with nothing behind** it". The probability we are looking for is $P(A|B)$. Notice that $P(A) = 1/3$.

We have

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

We get $P(B|A) = 1/2$ because the host will have the choice to show one of the door with no price behind. Also, we have $P(B|\overline{A}) = 1$ because the host will show the other door with nothing behind it given that nothing is behind door 1. Now, using Bayes' formula,

$$P(B) = P(B|A)P(A) + P(B|\overline{A})P(\overline{A}) = (1/2)(1/3) + (1)(2/3) = 1/2.$$

Therefore,

$$P(A|B) = \frac{(1/2)(1/3)}{(1/2)} = \frac{1}{3}.$$

Since conditional probability is a probability measure, we have $P(\overline{A}|B) = 1 - 1/3 = 2/3$. So there is a 2/3 chance that the price is behind door 2 or door 3 with the new information (not 1/2). △

The Monty Hall's Problem has attracted a lot of attention from mathematicians and, especially a brilliant woman, Marilyn Vos Savant. The following website https://www.statisticshowto.com/probability-and-statistics/monty-hall-problem/#Bayes explains clearly the history behind the problem and its different solutions.

### Partition Theorem

Bayes' Formula generalizes to more than one event.

**Theorem B.4** If $B_1$, $B_2$, ... are mutually exclusive events such that $\cup_{i=1}^{\infty} B_i = S$ and $A$ is an event, then

$$P(A) = \sum_{i=1}^{\infty} P(A|B_i)P(B_i).$$

**Note:** A consequence of this last result is that

$$P(A) = \sum_{i=1}^{N} P(A|B_i)P(B_i)$$

where $B_1$, $B_2$, ..., $B_N$ are mutually exclusive events with $\cup_{i=1}^{N} B_i = S$.

**Example B.6** Suppose that we have 3 cards identical in form. They each have the following characteristics:

**1st card** Both sides are red.

**2st card** Both sides are colored black.

**3rd card** One side is colored red and the other side black.

The three cards are mixed up in a hat, and 1 card is randomly selected and put down on the ground. If the upper side of the chosen card is colored red, what is the probability that the other side is colored black?

**Solution.** The sides of the cards are important. Denote by $RR$ the event that the card is all red, by $BB$ the card is all black, and by $RB$ the card is red-black.

Let $A$ be the event that the upperside of the card is red. Then,

$$\begin{aligned} P(RB|R) = \frac{P(RB \cap R)}{P(R)} &= \frac{P(RB)P(R|RB)}{P(R|RR)P(RR) + P(R|BB)P(BB) + P(R|RB)P(RB)} \\ &= \frac{(1/3)(1/2)}{(1)(1/3) + (0)(1/3) + (1/2)(1/3)} \\ &= \frac{1/6}{1/2} = \frac{1}{3}. \end{aligned}$$

△

### Odd Ratio

Definition B.2 Let $(S, \mathcal{A}, P)$ be a probability space. The odds ratio of an event $A$ is defined by $P(A)/P(\overline{A})$

**Note:**

- If the odds ratio is equal to $\alpha$, then it is common to say that the odds are $\alpha$ to 1 ($\alpha : 1$) in favor of the event $A$.
- If we know that an event $B$ has occured, then the new odds ratio of an event $A$, knowing $B$, is

$$\frac{P(A|B)}{P(\overline{A}|B)} = \frac{P(A)}{P(\overline{A})} \frac{P(B|A)}{P(B|\overline{A})}. \tag{B.3}$$

Example B.7 If coin $X$ is flipped, it comes up heads with probability $1/4$, whereas if coin $Y$ is flipped it comes up heads with probability $3/4$. Suppose that one of these coins is randomly chosen and is flipped twice. If both flips land heads, what is the probability that coin $Y$ was the one flipped?

**Solution.** Intuitively, it should be coin $B$ that was flipped because it has a higher chance to land heads. The odds ratio makes this intuition more mathematically rigorous.

Let $A$ denote the event "coin $Y$ is flipped". Let $B$ be the event "both flips land heads". We have that

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

because $P(A) = P(\overline{A}) = 1/2$. We have $P(B|A) = 9/16$ and $P(\overline{B}|A) = 1/16$. Therefore,

$$\frac{P(B|A)}{P(\overline{B}|A)} = 9,$$

and there is a $9 : 1$ chance that coin $Y$ was flipped. So $P(A|B) = \frac{9}{10}$. △

## B.3 Independent Events

**Definition B.3** Let $(S, \mathcal{A}, P)$ be a probability space and let $A, B$ be two events with $P(A) > 0$ and $P(B) > 0$. We say that $A$ and $B$ are independent if

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

Otherwise, $A$ and $B$ are called dependent.

**Note:**

- So $A$ and $B$ are independent if the information that the event $B$ has occurred does not influence the probability that the event $A$ occurs.
- Since $P(A|B)P(B) = P(A \cap B)$, then $A$ and $B$ are independent when $P(A)P(B) = P(A \cap B)$.

**Example B.8** A card is selected at random from an ordinary deck of 52 playing cards. If $A$ is the event that the selected card is an ace and $B$ is the event that it is a spade, then show that $A$ and $B$ are independent.

**Solution.** Every card is supposed to be equally likely. Then $P(A) = \frac{4}{52} \approx 0.0769$. We have $P(A \cap B) = \frac{1}{52} \approx 0.0192$ and $P(B) = \frac{1}{4} = 0.25$. Therefore,

$$P(A|B) = \frac{1/52}{1/4} = \frac{4}{52} = P(A).$$

So $A$ and $B$ are independent. △

**Example B.9** Suppose that we toss 2 fair dice. Let $A$ denote the event that the sum of the dice is 6 and $B$ denote the event that the first die equals 3. Are $A$ and $B$ independent?

**Solution.** Since the dice is fair, we have $P(A) = \frac{5}{36}$, $P(B) = \frac{6}{36}$ and $P(A \cap B) = \frac{1}{36}$. Therefore,

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{1/36}{6/36} = \frac{1}{6} \neq \frac{5}{36} = P(A).$$

So $A$ and $B$ are dependent. △

**Example B.10** Two fair dice are thrown. Let $A$ denote the event that the sum of the dice is 7. Let $B$ denote the event that the first die equals 4 and let $C$ be the event that the second die equals 3. From Example B.9, we know that $A$ and $B$ are independent. We can also show that $A$ and $C$ are independent. However, $A$ and $B \cap C$ are not independent.

**Note:** If an event $A$ is independent of an event $B$ and of an event $C$, then we can't conclude that $A$ is independent of $B \cap C$.

**Definition B.4** Three events $A$, $B$, $C$ are said to be independent if

a) $P(A)P(B) = P(A \cap B)$;

b) $P(A)P(C) = P(A \cap C)$;

c) $P(B)P(C) = P(B)P(C)$;

d) $P(A \cap B \cap C) = P(A)P(B)P(C)$.

**Note:**

- A group of events $A_1$, $A_2$, …, $A_n$ is said to be independent if for each subgroup of $r$ events $A_{i_1}$, $A_{i_2}$, …, $A_{i_r}$

$$P\Big(\bigcap_{j=1}^{r} A_{i_j}\Big) = \prod_{j=1}^{r} P(A_{i_j}).$$

- An infinite group of events is said to be independent if every finite subgroups are independent.

**Experiments Made Up From Sub-experiments**

EXAMPLE B.11 An experiment consists of continually tossing a coin, where each toss are independent from each other. The coin lands on head with probability $p$ and on tail with probability $1-p$. What is the probability that

a) at least 1 head occurs in the first $n$ tosses;

b) exactly $k$ heads occur in the first $n$ tosses;

c) all tosses result in heads?

**Solution.**

a) Let $A_i$ be the event "the $i$-th toss is head. The probability that at least one head occurs is represented by $A = \cup_{i=1}^{n} A_i$. However, it is easier to compute the complement, which is $\overline{A} = \cap_{i=1}^{n} \overline{A}_i$. By independence, we have

$$P\Big(\bigcap_{i=1}^{n} \overline{A}_i\Big) = P(\overline{A}_1)P(\overline{A}_2)\cdots P(\overline{A}_n) = (1-p)^n.$$

Since $P(A) + P(\overline{A}) = 1$, we get

$$P(A) = 1 - (1-p)^n.$$

b) Let $B$ be the event "$k$ heads occur in the first $n$ tosses". There are $\frac{(n-k)!}{j!(n-k-j)!}$ ways of getting $j$ heads in the last $n-k$ tosses. Therefore,

$$P(B) = \sum_{j=0}^{n-k} \frac{(n-k)!}{j!(n-k-j)!} p^{n-j}(1-p)^j.$$

For example, if $p = 1/2$, $n = 3$ and $k = 1$, then

$$P(B) = \sum_{j=0}^{2} \frac{2!}{j!(2-j)!}(0.5)^{3-j}(0.5)^j = 0.5.$$

c) If $C$ denotes the event "all tosses are head", then we have $C = \cap_{i=1}^{\infty} A_i$. Notice that the complement is $\cup_{i=1}^{\infty} \overline{A}_i$, which is the event "the coin lands on tail at least once". Based on the calculations from a), we have that $P(\cup_{i=1}^{n} \overline{A}_i) = 1 - p^n$. Since $\cup_{i=1}^{n} \overline{A}_i \subset \cup_{i=1}^{n+1} \overline{A}_i$, we can use the continuity property of probability measure and get

$$P(\overline{C}) = P\Big(\bigcup_{i=1}^{\infty} \overline{A}_i\Big) = \lim_{n\to\infty} P\Big(\bigcup_{i=1}^{n} \overline{A}_i\Big) = \lim_{n\to\infty} 1 - p^n = 1.$$

Therefore, $P(C) = 1 - P(\overline{C}) = 0$. So, it's unlikely that all tosses will land on head! *[Ask them if this means impossible.]* $\triangle$

**Note:**

- It is useful to see an event as the completion of successive independent smaller steps, called subexperiments.
- If all the subexperiments are the same, then they are called trials.

# Chapter C: Random Variables

In the definitions and theorems in this chapter, we assume a probability space $(S, \mathcal{A}, P)$ is given.

## C.1 Discrete Random Variables

Loosely speaking, a discrete random variable is a map $X : S \to \mathbb{R}$.

**Example C.1** Suppose a fair coin is tossed 3 times in a row. Then our sample space $S$ are all the possible triplets of letters $t$ or $h$ (for tail and head respectively). The event space is $\mathcal{A} = 2^S$, and every outcome have an equal probability, so that $P(A) = 1/8$ for every atomic event $A$.

Let $X$ denotes "the number of heads appearing", then $X$ is a discrete random variable.

The possible values of $X$, called the image of $X$, is $\text{im}X = \{0, 1, 2, 3\}$.

**Definition C.1** A discrete random variable $X$ on a probability space $(S, \mathcal{A}, P)$ is defined to be a mapping $X : S \to \mathbb{R}$ such that

a) the set $\operatorname{Im} X$ is a countable subset of $\mathbb{R}$;

b) $X^{-1}(\{x\}) := \{s \in S \,:\, X(s) = x\}$ is an event for every $x \in \mathbb{R}$.

**Note:**

- To simplify the notation, we will use the notation $\{X = x\}$ to denote the set $X^{-1}(\{x\})$.
- We also generalize this notation to include pre-images of intervals. For $x \in \mathbb{R}$,

$$\{X \leq x\} := \{s \in S \,:\, X(s) \leq x\}$$

and similarly for $\{X \geq x\}$.

**Example C.2** Let $(S, \mathcal{A}, P)$ be a probability space in which

$$S = \{1, 2, 3, 4, 5, 6\}, \quad \mathcal{A} = \{\varnothing, \{2, 4, 6\}, \{1, 3, 5\}, S\}.$$

Let $U$ and $V$ be defined by

$$U(s) = s, \quad V(s) = \begin{cases} 1 & \text{if } s \text{ is even,} \\ 0 & \text{if } s \text{ is odd} \end{cases}$$

for $s \in S$.

a) Is $U$ a discrete random variable?

b) Is $V$ a discrete random variable?

**Solution.**

a) The set $\operatorname{Im} U = \{1, 2, 3, 4, 5, 6, \}$, which is discrete. The problem is the set $U^{-1}(\{u\})$ for some real values of $u$. If $u = 1$, then $U^{-1}(\{1\}) = \{1\}$ which is not in the event space. Therefore, $U^{-1}(\{1\})$ is not an event and $U$ is not a discrete random variable.

b) The set $\operatorname{Im} V = \{0, 1\}$, which is discrete. If $v \neq 0, 1$, then $V^{-1}(\{v\}) = \varnothing$ because there is no $s$ such that $V(s) = v$. If $v = 0$, then $V^{-1}(\{0\}) = \{1, 3, 5\}$ which is an event from the event space. If $v = 1$, then $V^{-1}(\{1\}) = \{2, 4, 6\}$ which is an event from the event space. Therefore, $V$ is a discrete random variable.

**Theorem C.1** If $X$ and $Y$ are two discrete random variable, then the mapping $Z : S \to \mathbb{R}$ defined by $Z(s) = X(s) + Y(s)$ is a discrete random variable.

*Proof.* First, since $X$ and $Y$ are discrete random variable, then $\operatorname{Im} Z$ will be discrete necessarily. Let $z \in \mathbb{R}$. If $\{Z = z\} = \varnothing$ and since $\varnothing$ is an event, $\{Z = z\}$ is an event. So assume that $\{Z = z\} \neq \varnothing$. We can show that

$$\{Z = z\} = \{s \in S \,:\, X(s) + Y(s) = z\} = \bigcup_{y \in \operatorname{Im} Y} \Big(\{X = z - y\} \cap \{Y = y\}\Big).$$

Since $\operatorname{Im} Y$ is discrete, the union on the left hand side is an infinite union of countable sets. We know that $\{X = z - y\}$ and $\{Y = y\}$ are events, for any choices of $y$ and $z$. Therefore $\{X = z - y\} \cap \{Y = y\}$ is an event. Since the union of countably many events is still an event, we conclude that $\{Z = z\}$ is an event. $\square$

## C.2 Probability Mass Functions

Since the values of $X$ depends on data from a probability space, it is possible to assign a probability to each element in $\operatorname{Im} X$. For instance, in Example C.1, the values of the discrete random variable $X$ : "Number of heads observed after 3 tosses" can be assigned the following probabilities:

$$P(\{X = 1\}) := P(\{s \in S \,:\, X(s) = 1\}) = P(\{htt, tht, tth\}) = \frac{3}{8}.$$

Similar calculations will give the probabilities of the other values of $X$ (see the next example). The notations will be simplified a little bit as followed:

- $P(X = x)$ for $P(X^{-1}\{x\})$.
- Also, we will write $P(X \leq x)$ for $P(\{X \leq x\})$.

**Definition C.2** Let $X$ be a discrete random variable. The <u>probability mass function</u> (abbreviated <u>pmf</u>) $p_X$ of $X$ is the function defined by

$$p_X(x) = P(X = x).$$

**Example C.3** In Example C.1, the pmf is

$$p_X(0) = \frac{1}{8}, \quad p_X(1) = \frac{3}{8} = p_X(2), \textit{ and } \quad p_X(3) = \frac{1}{8}.$$

Notice that $p_X(0) + p_X(1) + p_X(2) + p_X(3) = 1$.

**Note:**

- For a discrete random variable $X$, $\operatorname{Im} X$ is countable, so we can write $\operatorname{Im} X = \{x_1, x_2, \ldots\}$ in a list.
- Also, for any $x \notin \operatorname{Im} X$, we have $p_X(x) = 0$.
- In this case, we see that

$$\sum_{i=1}^{\infty} p_X(x_i) = \sum_{x \in \operatorname{Im} X} p_X(x) = P(S) = 1.$$

**Theorem C.2** Let $T = \{t_1, t_2, \ldots, t_N\}$ be a set of distinct real numbers and let $\{\pi_1, \pi_2, \ldots, \pi_N\}$ be a collection of real numbers satisfying

$$\pi_j \geq 0, \quad \forall j, \text{ and } \quad \sum_{j=1}^{N} \pi_j = 1.$$

Then there exists a probability space $(T, \mathcal{B}, Q)$ and a discrete random variable $X : T \to \mathbb{R}$ such that the pmf of $X$ is given by

$$p_X(s) = \begin{cases} \pi_j, & s = s_j \\ 0, & s \neq s_j. \end{cases}$$

*Proof.* Define $\mathcal{B} = 2^T$ and define the probability measure $Q$ to be $Q(\{A\}) = \sum_{j=1}^{n} \pi_{i_j}$ for $A = \{t_{i_1}, t_{i_2}, \ldots, t_{i_n}\}$. Then $(T, \mathcal{B}, Q)$ is a probability space from the assumptions on the numbers $\pi_j$ and the fact that $T$ is finite. Now, let $X : T \to \mathbb{R}$ be $X(t) = t$. Then, we have $\operatorname{Im} X = T$, $p_X(x) = 0$ if $x \notin T$ and if $x \in \operatorname{Im} X$, say $x = t_j$ for some $j$, then

$$p_X(x) = P(X = t_j) = P(\{t_j\}) = \pi_j. \qquad \square$$

.

**Note:** With this theorem, it is enough to say "Let $X$ be a random variable taking the values $t_j$ with probability $\pi_j$, for $j = 1, 2, 3, \ldots, N$" and forget about the probability space.

**Example C.4** Let $S = \{0, 1, 2, \ldots\}$ and let $X : S \to \mathbb{R}$ be a discrete random variable with $\operatorname{Im} X = \{0, 1, 2, 3, 4, \ldots\}$. Define the function $p : \mathbb{R} \to \mathbb{R}$ by $p(x) = c2^x/x!$, for $x = 0, 1, 2, \ldots$, and $p(x) = 0$ for $x \neq 0, 1, 2, \ldots$.

a) For what value of $c$ is the function $p$ a pmf?

b) Find $P(X = 0)$.

c) Find $P(X > 2)$.

**Solution.**

a) We must have $\sum_{x \in \operatorname{Im} X} p(x) = 1$. But

$$\sum_{x=0}^{\infty} \frac{2^x}{x!} = e^2$$

and therefore

$$1 = \sum_{x=0}^{\infty} p(x) = ce^2 \quad \Rightarrow \quad c = e^{-2}.$$

b) We have $P(X = 0) = p(0) = e^{-2}2^0/0! = e^{-2} \approx 0.1353$.

c) We have $P(X > 2) = 1 - P(X \leq 2)$. But,

$$P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2) = e^{-2}\Big(\frac{2^0}{0!} + \frac{2^1}{1!} + \frac{2^2}{2!}\Big) \approx 0.6767$$

and therefore $P(X > 2) = 1 - 0.6767 = 0.3233$. △

## C.3 Functions of Discrete Random Variables

Let $X$ be a discrete random variable and let $g : \mathbb{R} \to \mathbb{R}$ be a function. Then we can show that the function $Y : S \to \mathbb{R}$ defined by

$$Y(s) = g(X(s))$$

is a discrete random variable[1]. We usually write $Y = g(X)$.

**Example C.5** Let $X$ be a discrete random variable.

a) Let $g(x) = ax + b$. Then $Y = g(X) = aX + b$. In this case, we have

$$P(Y = y) = P(aX + b = y) = P(X = (y - b)/a) = p_X((y - b)/a).$$

b) Let $g(x) = x^2$. Then $Y = g(X) = X^2$. In this case, for $y > 0$,

$$\begin{aligned} P(Y = y) = P(X^2 = y) &= P(\{s : X(s) = \sqrt{y}\} \cup \{s : X(s) = -\sqrt{y}\}) \\ &= P(X = \sqrt{y}) + P(X = -\sqrt{y}) \\ &= p_X(\sqrt{y}) + p_X(-\sqrt{y}). \end{aligned}$$

**Theorem C.3** Let $X$ be a discrete random variable and $g : \mathbb{R} \to \mathbb{R}$ be a function. Then the pmf of $Y$ is

$$p_Y(y) = \sum_{x \in g^{-1}(y)} P(X = x)$$

for $y \in \mathbb{R}$.

*Proof.* By definition, for a given $y \in \operatorname{Im} Y$, we have

$$p_Y(y) = P(Y = y) = P(g(X) = y).$$

But,

$$\{s \in S : g(X(s)) = y\} = \{x \in \operatorname{Im} X : g(x) = y\} = g^{-1}(\{y\}) \cap \operatorname{Im} X.$$

Since $\overline{\operatorname{Im} X}$ (the complement of $\operatorname{Im} X$) does not contribute to the value of the probability, we can therefore write

$$p_Y(y) = P(\{x \in \operatorname{Im} X : g(x) = y\}) = P(g^{-1}(\{y\}) \cap \operatorname{Im} X) = \sum_{x \in g^{-1}(y)} P(X = x). \qquad \square$$

[1]Trick: $\operatorname{Im} Y$ is discrete because $\operatorname{Im} X$ is. Also, if $x \in \mathbb{R}$, then $Y^{-1}(\{x\}) = X^{-1}(g^{-1}(\{x\})) = X^{-1}(\operatorname{Im} X \cap g^{-1}(\{x\})) \cup X^{-1}(\overline{\operatorname{Im} X} \cap g^{-1}(\{x\})) = X^{-1}(\operatorname{Im} X \cap g^{-1}(\{x\})) \cup \varnothing = X^{-1}(\operatorname{Im} X \cap g^{-1}(\{x\})) = \cup_{a \in A} X^{-1}(\{a\})$, where $A = \operatorname{Im} X \cap X^{-1}(\{x\})$ is a countable set.

## C.4 Expectation and Variance

### Expected Value

**Example C.6** Consider a fair 6-faced die and the following game. After tossing the die, if the face lands on an even number, then you win 2 US dollars. But if the face lands on an even number, then you loose 1 US dollar. Would you like to play this game?

**Solution.** Each outcome from {⚀, ⚂, ⚄} will result in a lost of 1 dollar and each outcome from the set {⚁, ⚃, ⚅} will result in a win of 2 dollars. Therefore, out of the six possibilities, we will win on average:

$$\frac{-1+2-1+2-1+2}{6} = \frac{(2)(3)}{6} + \frac{(-1)(3)}{6} = (2)\left(\frac{1}{2}\right) + (-1)\left(\frac{1}{2}\right) = \frac{1}{2}.$$

If we let $S =$ {⚀, ⚁, ⚂, ⚃, ⚄, ⚅} and $P(A) = \frac{1}{6}$ for every atomic event $A$ and if $X$ is defined in the following way:

$$X(s) = \begin{cases} 2 & \text{if } s \text{ is even,} \\ -1 & \text{if } s \text{ is odd.} \end{cases}$$

Then, the above calculations can be rewritten in the following way:

$$2P(X = 2) + (-1)P(X = -1). \qquad \triangle$$

The above expression is called the expected value, mean or expectation of the discrete random variable $X$.

**Definition C.3** If $X$ is a discrete random variable, then the expectation or mean of $X$ is denoted by $\mathrm{Exp}(X)$ or by $\mu_X$ and defined by

$$\mathrm{Exp}(X) = \sum_{x \in \mathrm{Im}\, X} xP(X = x)$$

whenever this sum converges absolutely, meaning $\sum_{x \in \mathrm{Im}\, X} |xP(X = x)| < \infty$.

**Note:** Using the pmf of $X$, the expected value can be rewritten as

$$\mathrm{Exp}(X) = \sum_{x \in \mathrm{Im}\, X} xp_X(x).$$

Also, when the context is clear, we will denote the expectation of $X$ simply by $\mu$.

**Theorem C.4** If $X$ is a discrete random variable and $g : \mathbb{R} \to \mathbb{R}$, then

$$\mathrm{Exp}(g(X)) = \sum_{x \in \mathrm{Im}\, X} g(x)P(X = x),$$

whenever the sum converges absolutely.

*Proof.* From the definition of the expected value, we have

$$\begin{aligned}
\mathrm{Exp}(g(X)) &= \sum_{y \in \operatorname{Im} g(X)} y P(g(X) = y) \\
&= \sum_{y \in \operatorname{Im} g(X)} y \sum_{x \in \operatorname{Im} X : g(x) = y} P(X = x) \\
&= \sum_{y \in \operatorname{Im} g(X)} \sum_{x \in \operatorname{Im} X : g(x) = y} g(x) P(X = x) \\
&= \sum_{x \in \operatorname{Im} X} g(x) P(X = x).
\end{aligned}$$

□

An application of the previous theorem to the expectation gives the following result.

**Corollary C.5** Let $X$ be a discrete random variable and let $a, b \in \mathbb{R}$.

a) If $P(X \geq 0) = 1$ and $\mathrm{Exp}(X) = 0$, then $P(X = 0) = 1$.

b) $\mathrm{Exp}(aX + b) = a\mathrm{Exp}(X) + b$.

*Proof.* a) Assume that $P(X \geq 0) = 1$ and $\mathrm{Exp}(X) = 0$. Notice that

$$P(X < 0) = 1 - P(X \geq 0) = 0.$$

Therefore, the values of $X$ should all be positive or zero. By definition of $\mathrm{Exp}(X)$, we have that $\sum_{x \in \operatorname{Im} X} x P(X = x) = 0$. Since every $x \in \operatorname{Im} X$ is positive, we must have $xP(X = x) = 0$ for $x > 0$ and therefore $P(X = x) = 0$ in this case. Now,

$$1 = P(X \geq 0) = P(X = 0) + P(X > 0) = P(X = 0) + 0 \quad \Rightarrow \quad P(X = 0) = 1.$$

b) This is a consequence of Theorem C.4. □

### Variance

Another important statistics of a discrete random variable to know about is the variance.

**Definition C.4** The <u>variance</u> of a discrete random variable $X$, denoted by $\mathrm{var}(X)$ is defined by

$$\mathrm{var}(X) = \mathrm{Exp}\big([X - \mu]^2\big),$$

where $\mu := \mathrm{Exp}(X)$.

**<u>Note:</u>** The <u>standard deviation</u> of a random variable is $\sqrt{\mathrm{var}(X)}$, usually denoted by $\sigma_X$.

Here is an easier expression to compute the variance of a discrete random variable.

**Theorem C.6** Let $X$ be a discrete random variable. Then,

$$\mathrm{var}(X) = \mathrm{Exp}(X^2) - \mu^2,$$

where $\mu := \mathrm{Exp}(X)$.

*Proof.* Let $\mu := \text{Exp}(X)$. From Theorem C.4 with $g(x) = (x-\mu)^2$, we have

$$\begin{aligned}
\text{var}(X) &= \sum_{x \in \text{Im}\, X} g(x) P(X=x) = \sum_{x \in \text{Im}\, X} (x-\mu)^2 P(X=x) \\
&= \sum_{x \in \text{Im}\, X} \big(x^2 - 2x\mu + \mu^2\big) P(X=x) \\
&= \sum_{x \in \text{Im}\, X} x^2 P(X=x) - 2\mu \sum_{x \in \text{Im}\, X} x P(X=x) + \mu^2 P(X=x) \\
&= \sum_{x \in \text{Im}\, X} x^2 P(X=x) - 2\mu \sum_{x \in \text{Im}\, X} x P(X=x) + \mu^2 \sum_{x \in \text{Im}\, X} P(X=x) \\
&= \text{Exp}(X^2) - 2\mu^2 + \mu^2 = \text{Exp}(X^2) - \mu^2
\end{aligned}$$

□

EXAMPLE C.7 The manager of an industrial plant is planning to buy a new machine of either type $a$ or type $b$. If $t$ denotes the number of hours of daily operation, the number of daily repairs $Y_1$ required to maintain a machine of type $a$ is a random variable with mean and variance both equal to $t/10$. The number of daily repairs $Y_2$ for a machine of type $b$ is a random variable with mean and variance both equal to $3t/25$. The daily cost of operating $a$ is $C_a(t) = 10t + 30Y_1^2$; for $b$ it is $C_b(t) = 8t + 30Y_2^2$. Assume that the repairs take negligible time and that each night the machines are tuned so that they operate essentially like new machines at the start of the next day. Which machine minimizes the expected daily cost if a workday consists of

a) 10 hours.

b) 20 hours.

**Solution.** Using linearity, the expected value of the discrete random variable $C_a(t)$ is

$$\text{Exp}(C_a(t)) = \text{Exp}(10t) + \text{Exp}(30Y_1^2) = 10t + 30\text{Exp}(Y_1^2).$$

From Theorem C.6, we have

$$\text{var}(Y_1) = \text{Exp}(Y_1^2) - \mu_1^2 \quad \Rightarrow \quad \text{Exp}(Y_1^2) = \text{var}(Y_1) + \mu_1^2,$$

where $\mu_1 = \text{Exp}(Y_1)$. Plugging in the values of $\text{var}(Y_1)$ and $\text{Exp}(Y_1)$ into the above equation, we get

$$\text{Exp}(Y_1^2) = \frac{t}{10} + \frac{t^2}{100},$$

and hence

$$\text{Exp}(C_a(t)) = 10t + 30\Big(\frac{t}{10} + \frac{t^2}{100}\Big) = 13t + 0.3t^2.$$

Similar calculations give

$$\text{Exp}(C_b(t)) = 11.6t + 0.432t^2.$$

a) In this scenario, $t = 10$. Therefore, $\text{Exp}(C_a) = 160$ and $\text{Exp}(C_b) = 159.2$. Machine $b$ will be less expensive to run.

b) In this scenario, $t = 20$. Therefore, $\text{Exp}(C_a) = 380$ and $\text{Exp}(C_b) = 404.8$. Machine $a$ will be less expensive to run. △

## C.5 Conditional Expectation and the Partition Theorem

When a condition is added, then the additional information will influence the probability $P(X = x)$ and therefore directly the expectation of $X$. We therefore introduce the conditional expectation.

**Definition C.5** Let $X$ be a discrete random variable and $B$ be an event with $P(B) > 0$. The conditional expectation of $X$ given $B$ is denoted by $E(X|B)$ and is defined by

$$E(X|B) = \sum_{x \in \operatorname{Im} X} xP(X = x|B),$$

whenever this sum converges absolutely.

We therefore have an analogous result to the Partition Theorem, but for the conditional expectation.

**Theorem C.7** Let $X$ be a discrete random variable and $B_1$, $B_2$, ... be mutually exclusive events such that $\cup_{i=1}^{\infty} B_i = S$ and $P(B_i) > 0$ for each $i$. Then

$$E(X) = \sum_{i=1}^{\infty} E(X|B_i)P(B_i),$$

whenever the sum converges absolutely.

*Proof.* We can partition $S$ as $S = \cup_{i=1}^{\infty} B_i$, where $B_i \cap B_j = \varnothing$, when $i \neq j$. Therefore,

$$\begin{aligned} E(X) = \sum_{x \in \operatorname{Im} X} xP(X = x) &= \sum_{x \in \operatorname{Im} X} x\Big(\sum_{j=1}^{\infty} P(X = x|B_j)P(B_j)\Big) \\ &= \sum_{j=1}^{\infty} \sum_{x \in X} xP(X = x|B_j)P(B_j) \\ &= \sum_{j=1}^{\infty} E(X|B_j)P(B_j). \end{aligned}$$

□

## C.6 Examples of Discrete Random Variables

Let $X$ be a discrete random variable.

### Bernouilli distribution

A discrete random variable $X$ has the Bernouilli distribution with parameter $p \in [0, 1]$ if $\operatorname{Im} X = \{0, 1\}$ and

$$P(X = 1) = p \quad \text{and} \quad P(X = 0) = 1 - p.$$

Used Scenarios: The Bernouilli distribution is usually used to model experiment in which the outcome is "success" or "failure".

### Binomial Distribution

Let $n$ be an integer and $q \in [0,1]$. $X$ has the binomial distribution with parameters $n$ and $q$ if $\operatorname{Im} X = \{0, 1, 2, \ldots, n\}$ and

$$P(X = k) = \frac{n!}{k!(n-k)!} q^k (1-q)^{n-k}, \quad k = 0, 1, 2, \ldots, n.$$

Used Scenarios: Experiments where the goal is to obtain a certain number of successes in $n$ trials.

EXAMPLE C.8 There are $n = 6$ machines to test if they are working properly or not. According to a recent survey, a machine is working properly in 75% of the time. What is the probability that 4 machines are working properly.

**Solution.** We have $q = 0.75$ and $n = 6$. Let $X$ be the discrete random variable given the number of machines that are working properly. Then $X \sim Bi(6, 0.75)$. Therefore,

$$P(X = 4) = \binom{6}{4} (0.75)^4 (0.25)^2 = \frac{6!}{4!2!} (0.75)^4 (0.25)^2 \approx 0.2966. \qquad \triangle$$

.

### Poisson Distribution

Let $\lambda > 0$. $X$ has the Poisson distribution if $\operatorname{Im} X = \{0, 1, 2, \ldots\}$ and

$$p_X(k) = \frac{1}{k!} \lambda^k e^{-\lambda}, \quad k = 0, 1, 2, \ldots.$$

Used Scenarios: Experiments where the goal is to obtain a certain number of successes in $n$ trials, with $n$ large.

**Note:** The parameter $\lambda$ usually refers to the expected number of successes in an experiment.

EXAMPLE C.9 Consider an experiment that consists of counting the number of $\alpha$-particles given off in a 1-second interval by 1 gram of radioactive material. If we know from past experience that, on the average, 3.2 such $\alpha$-particles are given off, what is a good approximation to the probability that no more than 2 $\alpha$-particles will appear?

**Solution.** We think of a the surface of the material as a composition of a high number $n$ of particular, that has $3.2/n$ chance of given off. We therefore can approximate the desire probability by a Poisson distribution with parameter $\lambda = nq = 3.2$. Then,

$$P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2) = \frac{e^{-3.2} 3.2^0}{0!} + \frac{e^{-3.2} 3.2^1}{1!} + \frac{e^{-3.2} 3.2^2}{2!} \approx 0.3799.$$

$\triangle$

THEOREM C.8 Let $X$ be a discrete random variable which follows a binomial distribution with parameters $n$ and $q$ and let $\lambda = nq$. Then

$$p_X(k) \approx \frac{1}{k!} \lambda^k e^{-\lambda}, \quad k = 0, 1, 2, \ldots,$$

when $n$ is large enough and $q$ is small enough.

### Negative Binomial Distribution

Let $q \in (0,1)$ and $n \geq 0$ be an integer. Then $X$ has the negative binomial distribution with parameters $q$ and $n$ if $\operatorname{Im} X = \{n, n+1, n+2, \ldots\}$ and

$$p_X(k) = \frac{(k-1)!}{(n-1)!(k-n)!} q^n (1-q)^{k-n}, \quad k = n, n+1, n+2, \ldots$$

Used-case Scenarios: Experiments where the goal is to find the probability of having the $n$-th success after $k$ trials.

EXAMPLE C.10 A geological study indicates that an exploratory oil well drilled in a particular region should strike oil with probability 0.2. Find the probability that the third oil strike comes on the fifth well drilled.

**Solution.** Let $X$ be the number of strikes needed to obtain a third oil strike. In this case, we have $q = 0.2$ and $n = 3$. We are searching for $P(X = 5)$. Then

$$P(X = 5) = \frac{4!}{2!2!}(0.2)^3(0.8)^2 = 0.03072. \qquad \triangle$$

.

### Geometric Distribution

Let $q \in (0,1)$. Then $X$ has the geometric distribution with parameter $q$ if $\operatorname{Im} X = \{1, 2, \ldots\}$ and

$$p_X(k) = (1-q)^{k-1} q, \quad k = 1, 2, 3, \ldots.$$

Used-case Scenarios: Experiments where the goal is to find the probability of the first success to occur within $k$ tries.

EXAMPLE C.11 An urn contains 10 red balls and 20 blue balls. Some balls are randomly selected, one at a time, until a red one is obtained. If we assume that each selected ball is replaced before the next one is drawn, what is the probability that

a) exactly 3 draws are needed?

b) at least 6 draws are needed.

**Solution.** Let $X$ be the discrete random variable counting the number of time needed to get a red ball. The random variable $X$ follows a geometric distribution with parameter $q$, giving the probability of selecting a red ball.

Since the ball is replaced in the urn, the probability of selecting a red ball is always the same, that is $1/3$. Therefore, $q = 1/3$.

a) Let $k = 3$, so that $P(X = k) = (1 - 1/3)^2(1/3) = 4/27$.

b) What is $P(X \geq 6)$? Using the complement, this is $1 - P(X < 6)$. Therefore,

$$P(X \geq 6) = 1 - P(X = 1) - P(X = 2) - P(X = 3) - P(X = 4) - P(X = 5) \approx 0.8683.$$

$\triangle$

### Summary

The table below is a summary of the expected value and variance of each of the examples presented in this section.

| Distribution | Expected Value | Variance |
|---|---|---|
| $B(q)$ | $q$ | $q(1-q)$ |
| $B(n,q)$ | $nq$ | $nq(1-q)$ |
| $\mathcal{P}(\lambda)$ | $\lambda$ | $\lambda$ |
| $NB(n,q)$ | $n/q$ | $n(1-q)/q^2$ |
| $G(q)$ | $1/q$ | $(1-q)/q^2$ |

Table C.1: Table of Mean and Variance of different distributions

# Chapter D: Continuous Random Variables

In this chapter, we assume that a probability space $(S, \mathcal{A}, P)$ is given.

## D.1 Distribution Function

Example D.1 Let $X$ be a discrete random variable with $\operatorname{Im} X = \{0, 1, 2, 3, 4, 5\}$ and let $p_X$ be its probability mass function given by $p_X(x) = 1/6$, if $x = 0, 1, 2, 3, 4, 5$.

a) What is $P(X \leq -1)$?

b) What is $P(X \leq 2)$?

c) What is $P(X \leq 3.5)$?

**Solution.**

a) Since $X$ takes only non-negative values, $\{X \leq -1\} = \varnothing$ and therefore $P(X \leq -1) = 0$.

b) We have

$$P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2) = p_X(0) + p_X(1) + p_X(2) = \frac{3}{6} = \frac{1}{2}.$$

c) We have

$$P(X \leq 3.5) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) = \frac{4}{6} = \frac{2}{3}.$$

In the case of a discrete random variable, from Problem **??**, we know that $\{X \leq x\}$ is an event, whenever $x \in \mathbb{R}$. To include more random variable where $\operatorname{Im} X$ is not necessarily discrete, we will have to require that $\{X \leq x\}$ is measurable because we will be interested in this event rather that $\{X = x\}$.

Definition D.1 A map $X : S \to \mathbb{R}$ is a random variable if $\{X \leq x\}$ is an event, for every $x \in \mathbb{R}$.

**Note:** From Problem **??**, when $\operatorname{Im} X$ is discrete, then this above definition also describes the discrete random variables.

Definition D.2 If $X$ is a random variable, the distribution function of $X$ is the function $F_X : \mathbb{R} \to [0, 1]$ defined by

$$F_X(x) = P(X \leq x).$$

**Example D.2** Three coins are flipped and let $X$ be the number of heads obtained. We know that $P(X=0)=1/8$, $P(X=1)=P(X=2)=3/8$, and $P(X=3)=1/8$. Find the distribution function of $X$.

**Solution.** Let $\lfloor x \rfloor$ be the larger integer smaller than $x$. For example $\lfloor 3.5 \rfloor = 3$ and $\lfloor -4.5 \rfloor = -5$.

When $x<0$, then $F_X(x)=P(X\leq x)=0$. For $x\geq 0$, we get

$$F_X(x)=\sum_{k=0}^{\lfloor x \rfloor} p_X(k).$$

The graph is illustrated in the Figure D.1. $\triangle$

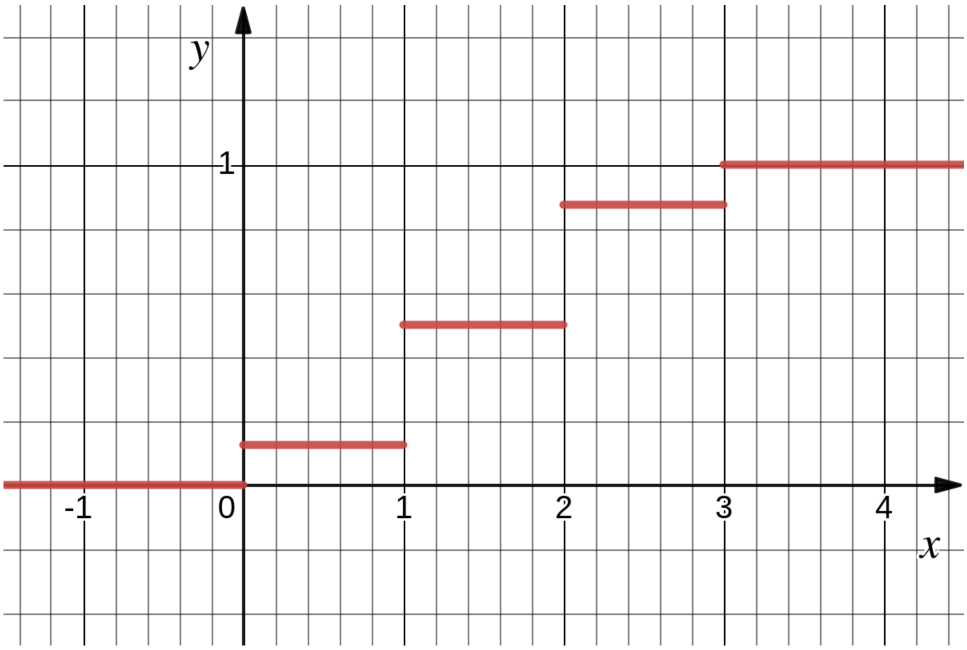



Figure D.1: Distribution function from Example D.2

**Theorem D.1** Let $X$ be a random variable. Then its distribution function $F_X$ satisfies the following properties:

a) $F_X$ is non-decreasing, meaning $x_1\leq x_2 \Rightarrow F_X(x_1)\leq F_X(x_2)$;

b) The set $\{a<X\leq b\}$ is an event and $P(a<X\leq b)=F_X(b)-F_X(a)$.

*Proof.* Let $X$ be as in the statement.

a) Assume that $x_1\leq x_2$. Then, $\{X\leq x_1\}\subset\{X\leq x_2\}$ and therefore $P(X\leq x_1)\leq P(X\leq x_2)$. With the notations introduced for the distribution function, this means $F_X(x_1)\leq F_X(x_2)$.

b) The set $\{a<X\leq b\}=\{X\leq b\}\cap\overline{\{X\leq a\}}$. Therefore, it is an event. By the property of the measure $P$, we get

$$P(a<X\leq b)=P(X\leq b)-P(X\leq a)=F_X(b)-F_X(a).$$ $\square$

## D.2 Continuous Random Variable

**Definition D.3** A random variable $X$ is <u>continuous</u> if its distribution function $F_X$ may be written in the form

$$F_X(x) = \int_{-\infty}^{x} f_X(u)\, du \quad (x \in \mathbb{R}),$$

for a map $f_X : \mathbb{R} \to [0, \infty)$.

**Note:**

- The function $f_X$ is called the <u>probability density function</u> (pdf for short) of $X$.
- It is customary to give the pdf of a continuous random variable instead of the distribution function.
- $f_X(x)$ does not represent a probability and its values may even exceed 1.
- In fact, $f_X(x)$ is a measure of probability in the following sense. If $\Delta x$ is small and positive, then the probability that $X$ is near $x$ is

  $$P(x \leq X \leq x + \Delta x) = F(x + \Delta x) - F(x) = \int_{x}^{x+\Delta x} f_X(u)\, du \approx f_X(x)\Delta x.$$

  Therefore, the true analogy is between $f_X(x)\Delta x$ and $p_X(x)$, for small $\Delta x$.
- Also, a technical detail that we won't get into is the following. To make sense of the above integral, the function $f$ should be integrable. There is different ways of making sense of the notion of integrability. We will assume we are dealing with Rieman integration.

**Theorem D.2** If $X$ is a continuous random variable with density function $f_X$, then

a) $P(X = x) = 0$, $\forall x \in \mathbb{R}$;

b) $P(a < X \leq b) = \displaystyle\int_a^b f_X(x)\, dx$, for any $a, b \in \mathbb{R}$ with $a \leq b$.

c) $\displaystyle\int_{-\infty}^{\infty} f_X(x)\, dx = 1$.

*Proof.* Let $X$ be as in the Theorem. Therefore, $F_X(x) = \displaystyle\int_{-\infty}^{x} f_X(u)\, du$.

a) Let $x \in \mathbb{R}$. For each positive integer $n$, set $A_n := \{x - 1/n < X \leq x\}$. Then, we see that the sequence $\overline{A}_n$ forms an increasing sequence of events. By the continuity of the probability measure $P$, we have

$$P\Big(\bigcup_{n=1}^{\infty} \overline{A}_n\Big) = \lim_{n\to\infty} P(\overline{A}_n)$$

which implies, using de Morgan's laws and the fact that $P(\overline{A}) = 1 - P(A)$,

$$1 - P\Big(\bigcap_{n=1}^{\infty} A_n\Big) = 1 - \lim_{n\to\infty} P(A_n).$$

Therefore,

$$P\Big(\bigcap_{n=1}^{\infty} A_n\Big) = \lim_{n\to\infty} P(X \leq x) - P(X \leq x - 1/n).$$

Since $\lim_{n\to\infty} x - 1/n = x$, we see that $\bigcap_{n=1}^{\infty} A_n = \{X = x\}$. Hence

$$P(X = x) = \lim_{n\to\infty} \Big(F_X(x) - F_X(x - 1/n)\Big) = \lim_{n\to\infty} \int_{x-1/n}^{x} f_X(u)\, du = 0.$$

b) Since $P(X = a) = 0$, we have

$$P(a \leq X \leq b) = P(X = a) + P(a < X \leq b) = P(a < X \leq b).$$

Therefore,

$$P(a < X \leq b) = F_X(b) - F_X(a) = \int_a^b f_X(x)\,dx.$$

c) We have

$$\int_{-\infty}^{\infty} f_X(x)\,dx = \lim_{x\to\infty} \int_{-\infty}^{x} f_X(u)\,du = \lim_{x\to\infty} F_X(x).$$

By the continuity of the probability measure $P$, for any sequence $x_n \to \infty$, the sets $\{X \leq x_n\}$ is an increasing sequence with $\bigcup_{n=1}^{\infty}\{X \leq x_n\} = S$ and hence

$$\lim_{x_n\to\infty} P(X \leq x_n) = P\Big(\bigcup_{n=1}^{\infty}\{X \leq x_n\}\Big) = P(S) = 1.$$

Therefore, since this is true for any sequence converging to $\infty$:

$$\int_{-\infty}^{\infty} f_X(x)\,dx = \lim_{x\to\infty} F_X(x) = \lim_{x\to\infty} P(X \leq x) = 1. \qquad \square$$

**Example D.3** A random variable $X$ has density function

$$f(x) = \begin{cases} 2x & \text{if } 0 < x < 1 \\ 0 & \text{otherwise.} \end{cases}$$

a) Find the distribution function $F_X$ of $X$.

b) Find $P(0.25 \leq X \leq 0.75)$.

**Solution.**

a) When $x \leq 0$, then $F_X(x) = 0$. For $0 < x < 1$, we then have

$$F_X(x) = \int_{-\infty}^{x} f_X(u)\,du = \int_0^x 2u\,du = x^2.$$

When $x \geq 1$, then

$$F_X(x) = \int_{-\infty}^{x} f_X(u)\,du = \int_0^1 2u\,du = 1.$$

b) Using the expression of $F_X$, we have

$$P(0.25 \leq X \leq 0.75) = F_X(0.75) - F_X(0.25) = \frac{9}{16} - \frac{1}{16} = \frac{1}{2}. \qquad \triangle$$

## D.3 Uniform Distribution

A random variable $X$ is said to have an uniform distribution with parameters $a$ and $b$, with $a < b$ if its density function is given by

$$f_X(x) = \begin{cases} 1/(b-a) & \text{if } a < x < b \\ 0 & \text{otherwise.} \end{cases}$$

We usually write $X \sim U(a, b)$.

The distribution function $F_X$ is then

$$F_X(x) = \int_{-\infty}^{x} f_X(u)\,du = \begin{cases} 0 & \text{if } x \le a \\ \frac{x-a}{b-a} & \text{if } a < x < b \\ 1 & \text{if } b \le x \end{cases}$$

Some phenomena in physical, management, and biological sciences can be modeled on a uniform distribution. The assumption are usually that the process follows a Poisson distribution and we know one occurrence has happened within a time interval $(0, t)$.

**Example D.4** Arrivals of customers at a checkout counter follow a Poisson distribution. It is known that, during a given 30-minute period, one customer arrives at the counter. Find the probability that a customer arrives during the last 5 minutes of the 30-minute period.

**Solution.** We assume that it is equally likely that a person arrives at any given minute within the 30-min period. If $X$ is the minute when a customer shows up, then we assume that $X \sim U(0, 30)$. Therefore,

$$f_X(x) = \frac{1}{30-0} \quad (0 \le x \le 30).$$

We then have

$$P(25 \le X \le 30) = \int_{25}^{30} f_X(x)\,dx = \frac{30-25}{30} = \frac{1}{6}. \qquad \triangle$$

**Example D.5** A parachutist lands at a random point on a line between markers $A$ and $B$.

a) Find the probability that she is closer to $A$ than to $B$.

b) Find the probability that her distance to $A$ is more than three times her distance to $B$.

**Solution.**

a) We assume that any given position between $A$ and $B$ is equally likely to occur, so that if $X$ is the location between $A$ and $B$, then $X \sim U(0, 1)$, where $0 = A$ and $1 = B$.

   For her to be closer to $A$, this means that $X < 1/2$. Therefore, the probability is

$$P(X < 1/2) = P(X \le 1/2) = \int_0^{1/2} dx = 1/2.$$

b) Let $x$ be the distance of the parachutist to the point $A$. Then the distance of the parachutist to $B$ is $1 - x$. We want to find when $x \ge 3(1-x)$, which is equivalent to

$$4x \ge 3 \iff x \ge \frac{3}{4}.$$

   Therefore, the probability is

$$P(X \ge 3/4) = 1 - P(X < 3/4) = 1 - 3/4 = \frac{1}{4}. \qquad \triangle$$

## D.4 Functions of Random Variables

**Example D.6** Let $X$ be a continuous random variable and let $g(x) = 2x + 3$. Find the distribution function of $Y = g(X)$ and its density function $f_Y$.

**Solution.** By definition, $F_Y(y) = P(Y \leq y)$. Since $Y = g(X)$, we have

$$\{Y \leq y\} = \{2X + 3 \leq y\} = \{X \leq (1/2)(y - 3)\}$$

and hence

$$F_Y(y) = P(X \leq (1/2)(y-3)) = F_X((1/2)(y-3)).$$

Therefore, $Y$ is a continuous random variable and

$$f_Y(y) = \frac{d}{dy}\Big(F_X((1/2)(y-3))\Big) = f_X\Big(\frac{y-3}{2}\Big)\frac{d}{dy}\Big(\frac{y-3}{2}\Big) = f_X\Big(\frac{y-3}{2}\Big)\Big(\frac{1}{2}\Big).$$

Using the notation $g^{-1}(y) = \frac{y-3}{2}$, it is possible to rewrite $f_Y(y)$ as

$$f_Y(y) = f_X(g^{-1}(y))\frac{d}{dy}[g^{-1}(y)]. \qquad \triangle$$

**Theorem D.3** Let $X$ be a continuous random variable with density function $f_X$, and let $g : \mathbb{R} \to \mathbb{R}$ be a strictly increasing and differentiable function. Then $Y = g(X)$ is a continuous random variable with density function

$$f_Y(y) = f_X(g^{-1}(y))\frac{d}{dy}[g^{-1}(y)] \quad (y \in \mathbb{R})$$

where $g^{-1}$ is the inverse function of $g$.

*Proof.* We first find the distribution function of $Y$. Notice that a result from Calculus II (or Introduction to Real Analysis) implies that the inverse function of $g$, $g^{-1}$, exists and $g^{-1}$ is also increasing. We then have

$$Y \leq y \iff g(X) \leq y \iff X \leq g^{-1}(y).$$

This implies that

$$F_Y(y) = P(Y \leq y) = P(X \leq g^{-1}(y)) = F_X(g^{-1}(y))$$

and differentiating both sides, we get

$$f_Y(y) = f_X(g^{-1}(y))\frac{d}{dy}[g^{-1}(y)]. \qquad \square$$

**Notes:**

- If $g$ is strictly decreasing and differentiable, then the density function $f_Y$ is

$$f_Y(y) = -f_X(g^{-1}(y))\frac{d}{dy}[g^{-1}(y)] \quad (y \in \mathbb{R}).$$

- In the cases where the above result does not apply, we have to treat each case on their own.

**Example D.7** Let $X$ be a continuous random variable. If $Y = X^2$, find the distribution function of $Y$ and its density function.

**Solution.** For $y \leq 0$, we have $F_Y(y) = P(Y \leq y) = 0$, because $Y$ takes only positive values. Assume that $y > 0$. Then,

$$P(Y \leq y) = P(X^2 \leq y) = P(|X| \leq \sqrt{y}) = P(-\sqrt{y} \leq X \leq \sqrt{y})$$

and from the properties of the distribution function, we have

$$F_Y(y) = F_X(\sqrt{y}) - F_Y(-\sqrt{y}).$$

Then, differentiating with respect to $y$ gives

$$f_Y(y) = f_X(\sqrt{y})\frac{1}{2\sqrt{y}} + f_X(-\sqrt{y})\frac{1}{2\sqrt{y}} = \frac{f_X(\sqrt{y}) + f_X(-\sqrt{y})}{2\sqrt{y}}. \qquad \triangle$$

## D.5 Expectations of Continuous Random Variables

**Definition D.4** If $X$ is a continuous random variable with density $f_X$, then the expectation of $X$ is denoted by $\mathrm{Exp}(X)$ and is defined by

$$\mathrm{Exp}(X) = \int_{-\infty}^{\infty} x f_X(x)\, dx,$$

whenever this integral converges absolutely, meaning that $\int_{-\infty}^{\infty} |x f_X(x)|\, dx < \infty$.

**Example D.8** Find the expectation of a continuous random variable $X \sim U(-1, 1)$.

**Solution.** By definition, we have

$$\mathrm{Exp}(X) = \int_{-1}^{1} x \frac{1}{1-(-1)}\, dx = \frac{1}{2}(1^2 - (-1)^2)/2 = 0. \qquad \triangle$$

To be able to derive a useful formula for the variance of a random variable $X$, we need the following result.

**Theorem D.4** If $X$ is a continuous random variable with density $f_X$ and if $g : \mathbb{R} \to [0, \infty)$ is a map such that $Y = g(X)$ is a continuous random variable. Then,

$$\mathrm{Exp}(Y) = \int_{-\infty}^{\infty} g(x) f_X(x)\, dx,$$

whenever this integral converges absolutely.

To prove this result, we need the following interesting result.

**Lemma D.5** If $X$ is a continuous random variable taking only non-negative values, then

$$\mathrm{Exp}(X) = \int_0^{\infty} (1 - F_X(x))\, dx.$$

*Proof.* By definition of the expectation,

$$\mathrm{Exp}(X) = \int_{-\infty}^{\infty} x f_X(x)\, dx.$$

Since $X$ takes only positive values, we have that $F_X(x) = 0$ for any $x \leq 0$. Therefore, $f_X(x) = 0$ for any $x \leq 0$ and

$$\mathrm{Exp}(X) = \int_{0}^{\infty} x f_X(x)\, dx.$$

Using the fact that $x = \displaystyle\int_0^x dt$, we have

$$\mathrm{Exp}(X) = \int_0^\infty \int_0^t f_X(x)\, dt dx.$$

From Fubini's Theorem, we get

$$\mathrm{Exp}(X) = \int_0^\infty \int_t^\infty f_X(x)\, dx dt = \int_0^\infty \Big(1 - 1 + \int_t^\infty f_X(x)\, dx\Big)\, dt.$$

Using the fact that $1 = \int_0^\infty f_X(x)\, dx$, we find that

$$1 - 1 + \int_t^\infty f_X(x)\, dx = 1 - \int_0^\infty f_X(x)\, dx + \int_t^\infty f_X(x)\, dx = 1 - \int_0^t f_X(x)\, dx = 1 - F_X(t).$$

Hence,

$$\mathrm{Exp}(X) = \int_0^\infty (1 - F_X(t))\, dt. \qquad \square$$

*Proof of Theorem D.4.* Since $g$ takes only non-negative values, the continuous random variable $Y = g(X)$ takes only non-negative values. Lemma D.5 then implies

$$\mathrm{Exp}(Y) = \int_0^\infty (1 - F_Y(u))\, du = \int_0^\infty 1 - P(g(X) \leq u)\, du.$$

Recall that $P(Y < \infty) = \int_{-\infty}^\infty f_Y(t)\, dt = 1$, so that

$$\mathrm{Exp}(Y) = \int_0^\infty P(g(X) > u)\, du = \int_0^\infty \Big(\int_B f_X(x)\, dx\Big) du,$$

where $B = \{x \,:\, g(x) > u\}$. Using Fubini's Theorem, we get

$$\mathrm{Exp}(Y) = \int_0^\infty \int_0^{g(x)} du\, f_X(x)\, dx = \int_0^\infty g(x) f_X(x)\, dx. \qquad \square$$

**Note:** Theorem D.4 remains true if we have a function $g : \mathbb{R} \to \mathbb{R}$. In this case, we have to write $g = g_+ - g_-$, where $g_+$ and $g_-$ are defined as followed:

$$g_+(x) = \max\{g(x), 0\} \quad \text{and} \quad g_-(x) = \max\{-g(x), 0\}$$

and apply Lemma D.5 to $g_+$ and $g_-$.

As we did in the discrete case, we define the variance as the expectation of $(X - \mathrm{Exp}(X))^2$.

Definition D.5 Let $X$ be a continuous random variable, then the variance of $X$ is denoted by $\mathrm{Var}(X)$ and is defined by

$$\mathrm{Var}(X) = \mathrm{Exp}((X - \mu)^2),$$

where $\mu := \mathrm{Exp}(X)$.

Using Theorem D.4 with $g(x) = (x-\mu)^2$, we find that

$$\mathrm{Var}(X) = \int_{-\infty}^{\infty} (x-\mu)^2 f_X(x)\,dx = \mathrm{Exp}(X^2) - \mu^2.$$

EXAMPLE D.9 Find the variance of $X \sim U(a,b)$.

**Solution.** From the above formula, we have $\mathrm{Var}(X) = \mathrm{Exp}(X^2) - \mu^2$, where $\mu = \mathrm{Exp}(X)$. We have

$$\mathrm{Exp}(X) = \int_{-\infty}^{\infty} x f_X(x)\,dx = \int_a^b \frac{x}{b-a}\,dx = \frac{a+b}{2},$$

and

$$\mathrm{Exp}(X^2) = \int_{-\infty}^{\infty} x^2 f_X(x)\,dx = \int_a^b \frac{x^2}{b-a}\,dx = \frac{a^2+ab+b^2}{3}.$$

Hence

$$\mathrm{Var}(X) = \frac{a^2+ab+b^2}{3} - \frac{a^2+2ab+b^2}{4} = \frac{(b-a)^2}{12}. \qquad \triangle$$

## D.6 Other Examples of Continuous Random Variables

### Normal Distribution

A random variable $X$ is said to have a normal distribution with parameters $\mu$ and $\sigma$ if its density function is given by

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right).$$

We usually write $X \sim N(\mu, \sigma)$. We have $\mathrm{Exp}(X) = \mu$ and $\mathrm{Var}(X) = \sigma^2$.

The distribution function $F_X$ can not be given in a closed form. We usually use a table that contains approximations of the value of $F_X$.

The normal distribution is one of the most important distribution out there. We will see later that a lot of phenomena can be approximated by a normal distribution. We first need the following result.

THEOREM D.6 Let $X \sim N(\mu, \sigma)$ be a continuous random variable and let $Z = (X-\mu)/\sigma$. Then $Z \sim N(0,1)$.

*Proof.* With $g(x) = (x-\mu)/\sigma$ so that $g^{-1}(z) = \sigma z + \mu$. Therefore, the density function of $Z$ is

$$f_Z(z) = f_X(\sigma z + \mu)\sigma = \frac{\sigma}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(\sigma z + \mu - \mu)^2\right) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right). \qquad \square$$

EXAMPLE D.10 The achievement scores of college entrance examination are normally distributed with mean 75 and standard deviation 10. What fraction of the scores lies between 80 and 90?

**Solution.** We have $\mu = 75$ and $\sigma = 10$. We create the new random variable $Z = (X-\mu)/\sigma$, so that $Z \sim N(0,1)$ from Theorem D.6. Therefore,

$$80 \le X \le 90 \iff \frac{80-75}{\sigma} \le Z \le \frac{90-75}{10} \iff 0.5 \le Z \le 1.5.$$

The probability we are looking for is $P(80 \leq X \leq 90) = P(0.5 \leq Z \leq 1.5)$. Using the table, we have

$$P(0.5 \leq Z \leq 1.5) = 0.93319 - 0.69146 = 0.24173. \qquad \triangle$$

## Exponential Distribution

A random variable $X$ is said to have an exponential distribution with parameter $\lambda > 0$ if its density function is given by

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x > 0 \\ 0 & \text{if } x \leq 0 \end{cases}$$

We usually write $X \sim \text{Poisson}(\lambda)$. We have $\text{Exp}(X) = 1/\lambda$ and $\text{Var}(X) = 1/\lambda^2$.

The distribution function $F_X$ is then

$$F_X(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ 1 - e^{-\lambda x} & \text{if } x > 0. \end{cases}$$

## Beta Distribution

A random variable $X$ is said to have a beta distribution with parameters $\alpha, \beta > 0$ if the density function $X$ is

$$f_X(x) = \begin{cases} \dfrac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)} & , 0 \leq x \leq 1 \\ 0 & \text{elsewhere,} \end{cases}$$

where

$$B(\alpha,\beta) = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}\,dt = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}.$$

We have $\text{Exp}(X) = \alpha/(\alpha+\beta)$ and $\text{Var}(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$.

# Chapter E: Random Vectors

In this chapter, we assume that a probability space $(S, \mathcal{A}, P)$ is given.

## E.1 Bivariate Distributions

Given two random variables $X$ and $Y$ defined on the probability space $(S, \mathcal{A}, P)$, we can think of them acting together as a random vector $(X, Y)$ taking values in $\mathbb{R}^2$.

**Example E.1** Let $S = \{$⚀, ⚁, ⚂, ⚃, ⚄, ⚅$\}$ with $\mathcal{A} = \mathcal{P}(S)$ be the probability space associated to throwing a fair dice. Assume that we throw two such die and consider the following random variable:

1) $X$: "The number of dots on the first dice".
2) $Y$: "The number of dots on the second dice".

Assuming the two dice are independent, what is the probability that $X \leq 2$ and $Y \leq 2$?

**Solution.** Denote this probability by $p$. There is $1/6$ chance of obtaining all the different faces. If $A = \{X \leq 2\} \cap \{Y \leq 2\}$, then

$$\begin{aligned} p &= P(A) \\ &= P(X = 1, Y = 1) + P(X = 1, Y = 2) + P(X = 2, Y = 1) + P(X = 2, Y = 2) \\ &= \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} \\ &= \frac{4}{36} = \frac{1}{9}. \end{aligned}$$

We denote $p$ by $F_{X,Y}(2, 2)$ and this is called the joint distribution function of $X$ and $Y$ at $(X, Y) = (2, 2)$. $\triangle$

**Notations:**

- the set $\{s : X(s) \leq x, Y(s) \leq y\} = \{X \leq x\} \cap \{Y \leq y\}$ will be denoted by $\{X \leq x, Y \leq y\}$.
- the probability of $\{X \leq x, Y \leq y\}$ is denoted by $P(X \leq x, Y \leq y)$.

**Definition E.1** The joint distribution function of a pair of random variables $X$, $Y$ is the mapping $F_{X,Y} : \mathbb{R}^2 \to [0, 1]$ given by

$$F_{X,Y}(x, y) = P(X \leq x, Y \leq y).$$

Basic Properties:

1) $\lim_{x\to-\infty}\lim_{y\to-\infty} F_{X,Y}(x,y) = \lim_{y\to-\infty}\lim_{x\to-\infty} F_{X,Y}(x,y) = 0.$

2) $\lim_{x\to\infty}\lim_{y\to\infty} F_{X,Y}(x,y) = \lim_{y\to\infty}\lim_{x\to\infty} F_{X,Y}(x,y) = 1.$

3) Increasing: $F(x_1,y_1) \leq F(x_2,y_2)$, for any $x_1 \leq x_2$ and $y_1 \leq y_2$.

4) $F_X(x) = \lim_{y\to\infty} F_{X,Y}(x,y)$. [Proof: Notice that $\{Y < \infty\} = S$, so that

$$F_X(x) = P(X \leq x) = P(X \leq x, Y < \infty) = \lim_{y\to\infty} F_{X,Y}(x,y).]$$

5) $F_Y(y) = \lim_{x\to\infty} F_{X,Y}(x,y)$.

**Definition E.2** If $X$ and $Y$ are random variables, then the above limits in items 4 and 5 are called the marginal distributions of the random vector $(X,Y)$.

**Example E.2** Assume that $X$ and $Y$ are random variables with joint distribution function

$$F_{X,Y} = \begin{cases} 1 - e^{-x} - e^{-y} + e^{-x-y} & \text{if } x, y \geq 0. \\ 0 & \text{otherwise.} \end{cases}$$

Find the marginals of this joint distribution.

**Solution.** We first let $y \to \infty$, so that, if $x \geq 0$, then

$$\lim_{y\to\infty} F_{X,Y}(x,y) = \lim_{y\to\infty} 1 - e^{-x} - e^{-y} + e^{-x-y} = 1 - e^{-x}.$$

When $x < 0$, then $F_{X,Y}(x,y) = 0$. Therefore, this gives us $F_X(x) = \max\{0, 1 - e^{-x}\}$, which is the distribution function of the exponential distribution with parameter $\lambda = 1$. Similarly, $F_Y(y) = \max\{0, 1 - e^{-y}\}$. △

Notice that, in the last example,

$$F_{X,Y}(x,y) = F_X(x)F_Y(y)$$

meaning that

$$P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y).$$

When this happens, we say that $X$ and $Y$ are independent.

**Definition E.3** Let $X$ and $Y$ be two random variables with joint distribution function $F_{X,Y}$. Then $X$ and $Y$ are independent if for every $x, y \in \mathbb{R}$,

$$F_{X,Y}(x,y) = F_X(x)F_Y(y)$$

where $F_X$ and $F_Y$ are the marginals of $X$ and $Y$ respectively.

**Remark.** When $X$ and $Y$ are not independent, we say that they are dependent.

## E.2 Continuous Random Vectors

**Definition E.4** Two random variables $X, Y$ are (jointly) continuous if there is an integrable function $f : \mathbb{R} \times \mathbb{R} \to [0, \infty)$ such that

$$F_{X,Y}(x,y) = \iint_R f(u,v)\, dA$$

for every $x, y \in \mathbb{R}$, where $R = (-\infty, x] \times (-\infty, y]$.

**Remark:**

- When $X$ and $Y$ are jointly continuous, the function $f(x,y)$ in the integral is usually called the joint probability density function and usually denoted by $f_{X,Y}$.
- Recall from multivariable calculus that, in cartesian coordinates,

$$\iint_R f(u,v)\,dA = \int_{-\infty}^{x}\int_{-\infty}^{y} f(u,v)\,dvdu.$$

- What we mean by an integrable function is a function $f : \mathbb{R}\times\mathbb{R} \to [0,\infty)$ such that

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dxdy < \infty.$$

EXAMPLE E.3 Suppose that a radioactive particle is randomly located in a square with sides of unit length. That is, if two regions within the unit square and of equal area are considered, the particle is equally likely to be in either region. Let $X$ and $Y$ denote the coordinates of the particle's location. A reasonable model for the joint probability density function of $X$ and $Y$ is

$$f_{X,Y}(x,y) = \begin{cases} 1 & \text{if } (x,y) \in [0,1]\times[0,1] \\ 0 & \text{elsewhere.} \end{cases}$$

a) Sketch the probability density function.

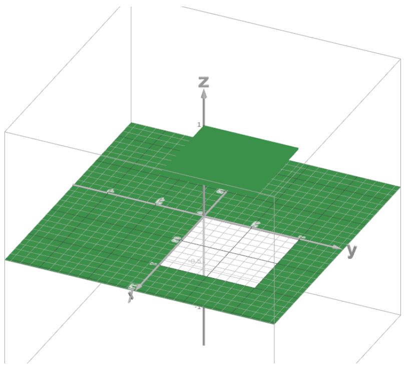
b) Find $F(0.2, 0.4)$.

c) Find $P(0.1 \le X \le 0.3, 0 \le Y \le 0.5)$.

d) Find $P(X+Y \le 0.5)$.

**Solution.**

a) Using Desmos 3D online app, the joint density function looks like this:




b) By definition, we have

$$F(0.2,0.4) = \iint_R f_{X,Y}(u,v)\,dA,$$

where $R = (-\infty, x]\times(-\infty, y]$. Since the function is 0 outside $[0,1]\times[0,1]$, we can restrict the integral to $R\cap[0,1]\times[0,1] = [0,0.2]\times[0,0.4]$. Therefore

$$F(0.2,0.4) = \int_0^{0.4}\int_0^{0.2} 1\,dxdy = (0.4)(0.2) = 0.08.$$

c) Here, we will use the fact that $\{0.1 \le X \le 0.3, 0 \le Y \le 0.5\}$ is equal to

$$\{X \le 0.3, Y \le 0.5\} \cap \overline{\{X \le 0.1, Y \le 0.5\} \cup \{X \le 0.3, Y \le 0\}}.$$

Using the identity $P(A \cap \overline{B}) = P(A) - P(B)$, when $B \subset A$ and with $A = \{X \leq 0.3, Y \leq 0.5\}$ and $B = \{X \leq 0.1, Y \leq 0.5\} \cup \{X \leq 0.3, Y \leq 0\}$, we find that

$$P(0.1 \leq X \leq 0.3, 0 \leq Y \leq 0.5) = P(A) - P(B).$$

Using the fact that $P(C \cup D) = P(C) + P(D) - P(C \cap D)$, with $C = \{X \leq 0.1, Y \leq 0.5\}$ and $D = \{X \leq 0.3, Y \leq 0\}$, we find that

$$P(B) = P(X \leq 0.1, Y \leq 0.5) + P(X \leq 0.3, Y \leq 0) - P(X \leq 0.1, Y \leq 0).$$

Combining everything together, we get

$$\begin{aligned} P(0.1 \leq X \leq 0.3, 0 \leq Y \leq 0.5) &= P(X \leq 0.3, Y \leq 0.5) - P(X \leq 0.1, Y \leq 0.5) \\ &\quad - P(X \leq 0.3, Y \leq 0) + P(X \leq 0.1, Y \leq 0) \\ &= F(0.3, 0.5) - F(0.1, 0.5) - F(0.3, 0) + F(0.1, 0) \\ &= \int_{0.1}^{0.3} \int_0^{0.5} 1 \, dy dx = (0.2)(0.5) = 0.1. \end{aligned}$$

d) Let $R = \{(X, Y) : X + Y \leq 0.5\}$. Then, because $f_{X,Y} = 0$ outside of $[0, 1] \times [0, 1]$, we can write the region as

$$R = \{(X, Y) : 0 \leq X \leq 0.5, 0 \leq Y \leq 0.5 - X\}.$$

Therefore, we get

$$\begin{aligned} P((X, Y) \in R) = P(0 \leq X \leq 0.5, 0 \leq Y \leq 0.5 - X) &= \int_0^{0.5} \int_0^{0.5-x} 1 \, dy dx \\ &= \int_0^{0.5} (0.5 - x) \, dx = 0.025. \end{aligned}$$

Notice that

$$P((X, Y) \in R) = \iint_R f_{X,Y}(x, y) \, dA. \qquad \triangle$$

**Remark:**

- If $R = [x_0, y_0] \times [x_1, y_1]$ is a rectangle in the plane, then

$$P((X, Y) \in R) = P(x_0 \leq X \leq x_1, y_0 \leq Y \leq y_1) = \iint_R f(u, v) \, dA.$$

- More generally, if $R$ is any region in the plane, then

$$P((X, Y) \in R) = \iint_R f(u, v) \, dA.$$

- When $(X, Y)$ are jointly continuous with density $F_{X,Y}$, then we can recover the joint probability mass function $f_{X,Y}$ in the following way:

$$f_{X,Y}(x, y) = \frac{\partial^2}{\partial x \partial y} F(x, y).$$

- The density functions of $X$ and $Y$ can be recovered in the following ways:
  1) $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) \, dy$;
  2) $f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) \, dx$.

**Example E.4** Let $(X, Y)$ be a random vector with probability density function

$$f_{X,Y}(x, y) = \begin{cases} 2x & \text{if } (x, y) \in [0,1] \times [0,1] \\ 0 & \text{elsewhere.} \end{cases}$$

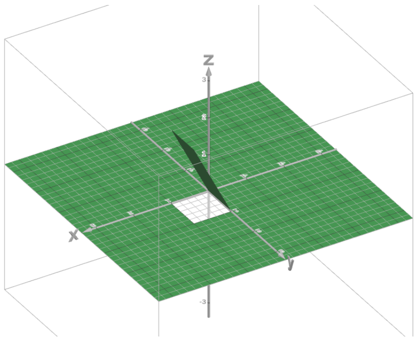
a) Sketch $f_{X,Y}$.

b) Find the marginals of $X$ and $Y$.

c) What can you conclude on $f_{X,Y}$?

**Solution.**

a) The graph of the function should look like this:




b) By definition $F_X(x) = \lim_{y\to\infty} F_{X,Y}(x, y)$. Since $X$ and $Y$ are jointly continuous, we have

$$F_X(x) = \lim_{y\to\infty} \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(u, v)\, dvdu.$$

When $x < 0$, we have $F_X(x) = 0$. For $0 \le x \le 1$, we have $f_{X,Y}(x) = 2x$, if $y$ is big enough, so that

$$F_X(x) = \int_0^x \int_0^1 2u\, dvdu = x^2.$$

When $x > 1$ and $y$ is sufficiently big (in fact bigger than 1), we see that $F_X(x) = 1$. Therefore,

$$F_X(x) = \begin{cases} 0 & x < 0 \\ x^2 & 0 \le x \le 1 \\ 1 & x > 1 \end{cases}$$

or we can also rewrite $F_X(x) = \Big(\max\{0, \min\{x, 1\}\}\Big)^2$.

With similar calculations, we obtain

$$F_Y(y) = \begin{cases} 0 & y < 0 \\ x & 0 \le y \le 1 \\ 1 & y > 1 \end{cases}$$

or we can write $F_Y(y) = \max\{0, \min\{y, 1\}\}$.

c) Recall that $f_X = \frac{d}{dx} F_X$. Therefore, when $x < 0$ and $x > 1$, we have $f_X = 0$. When $0 < x < 1$, we have $f_X(x) = 2x$. Similarly, we have $f_Y = \frac{d}{dy} F_Y$. Therefore, when $y < 0$ and $y > 1$, we have $f_Y = 0$. When $0 < y < 1$, we have $f_Y(y) = 1$. Remarkably, we get

$$f_{X,Y}(x, y) = f_X(x) f_Y(y)!!$$

△

**Theorem E.1** Jointly continuous random variables $X$ and $Y$ are independent if and only if there are two functions $g, h : \mathbb{R} \to [0, \infty)$ such that

$$f_{X,Y}(x, y) = g(x)h(y) \quad \forall x, y \in \mathbb{R}.$$

**Remark:**

- The functions $g$ and $h$ are called the marginal probability density functions of the random vector $(X, Y)$.
- The function $g$ is equal to $f_X$, the probability density function of $X$.
- The function $h$ is equal to $f_Y$, the probability density function of $Y$.

**Example E.5** Let $X$ and $Y$ be two random variables having joint probability density function

$$f_{X,Y}(x, y) = \begin{cases} 2e^{-x-y} & \text{if } 0 < x < y \\ 0 & \text{elsewhere.} \end{cases}$$

Are $X$ and $Y$ independent?

**Solution.** The marginal density function of $X$ is given by

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dy.$$

When $x < 0$, then $f_{X,Y}(x, y) = 0$ and so $f_X(x) = 0$. Let $x > 0$. Then, $f_{X,Y}(x, y) = 2e^{-x-y}$ when $y > x$ and

$$f_X(x) = \int_x^{\infty} 2e^{-x-y}\, dy = 2e^{-2x}.$$

Thus, $f_X(x) = 0$ if $x < 0$ and $f_X(x) = 2e^{-2x}$, if $x \geq 0$.

The marginal density function of $Y$ is given by

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dx.$$

When $y < 0$, then $f_{X,Y}(x, y) = 0$ and so $f_Y(y) = 0$. Let $y > 0$. Then $f_{X,Y}(x, y) = 2e^{-x-y}$ when $0 < x < y$, so that

$$f_Y(y) = \int_0^y 2e^{-x-y}\, dy = 2e^{-y}(1 - e^{-y}).$$

Therefore, $f_Y(y) = 0$ when $y < 0$ and $f_Y(y) = 2e^{-y}(1 - e^{-y})$ when $y \geq 0$.

We see that

$$f_{X,Y}(x, y) \neq f_X(x) f_Y(y)$$

and therefore, $X$ and $Y$ are not independent. △

### Conditional Density Functions

When we don't have any information on $X$, the marginal density function $f_Y$ of $Y$ is the average

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dx.$$

But, if we have information on values of $X$, we may consider $P(Y \leq y | X = x)$. However, since $P(X = x) = 0$ in the continuous case, we can't use formula $P(A|B) = P(A \cap B)/P(B)$ to get an expression of this probability. Instead, we compute $P(Y \leq y | x \leq X \leq X + \Delta x)$, for a small $\Delta x$ and then divide by $\Delta x$. Doing so, we get that

$$\begin{aligned} F_{Y|X=x}(y) = \lim_{\Delta x \to 0} P(Y \leq y | x \leq X \leq x + \Delta x) &= \lim_{\Delta x \to 0} \frac{P(Y \leq y, x \leq X \leq x + \Delta x)}{P(x \leq X \leq x + \Delta x)} \\ &= \lim_{\Delta x \to 0} \frac{\int_{-\infty}^{y} \frac{1}{\Delta x} \int_{x}^{x+\Delta x} f_{X,Y}(u,v)\, du dv}{\frac{1}{\Delta x} \int_{x}^{x+\Delta x} f_X(u)\, du} \\ &= \int_{-\infty}^{y} \frac{f_{X,Y}(x,v)}{f_X(x)}\, dv. \end{aligned}$$

DEFINITION E.5 If $X$ and $Y$ are jointly continuous random variables, then the conditional density function of $Y$ given that $X = x$ is denoted by $f_{Y|X}(\cdot|x)$ and is defined by

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$$

for any $y \in \mathbb{R}$ and $x$ satisfying $f_X(x) > 0$.

**Remark:**

- In the continuous case, we define

$$P(Y \leq y | X = x) := \int_{-\infty}^{y} f_{Y|X}(v|x)\, dv.$$

- Similarly, conditioning with respect to the event $\{Y = y\}$, then the conditional density function of $X$ given that $Y = y$ is

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(x)}.$$

- The formulas of the conditional density are very similar to the definition of $P(A|B)$.

EXAMPLE E.6 Find $f_{X|Y}$ if $X$ and $Y$ have joint probability density function

$$f_{X,Y}(x,y) = \begin{cases} 2e^{-x-y} & \text{if } 0 < x < y < \infty \\ 0 & \text{otherwise.} \end{cases}$$

**Solution.** From Example E.5, the density of $Y$ is

$$f_Y(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dx = 2e^{-y}(1 - e^{-y}).$$

For $y < 0$, we have $f_{X|Y}(x|y) = 0$. For $y > 0$, we have

$$f_{X|Y}(x|y) = \frac{2e^{-x-y}}{2e^{-y}(1 - e^{-y})} = \frac{e^{-x}}{1 - e^{-y}}. \qquad \triangle$$

## E.4 Important Measurements

Expectation is defined for a random variable. Therefore, we will compute the expectation of a random variable $Z = g(X, Y)$, that is

$$Z(s) = g(X(s), Y(s)) \quad (s \in S),$$

for some function $g : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$.

**Example E.7** Compute the expected value of the random variable $Z = X + Y$, if $X$ and $Y$ have joint probability density function $f_{X,Y}$.

**Solution.** By definition, the expectation of $Z$ is

$$\mathrm{Exp}(Z) = \int_{-\infty}^{\infty} z f_Z(z)\, dz.$$

However, we have

$$P(Z \leq z) = P(X + Y \leq z) = \iint_A f_{X,Y}(x, y)\, dA$$

where $A = \{(x, y) \, : \, x + y \leq z\}$. We therefore see that

$$F_Z(z) = \int_{-\infty}^{\infty} \int_{-\infty}^{z-x} f_{X,Y}(x, y)\, dydx = \int_{-\infty}^{z} \int_{-\infty}^{\infty} f_{X,Y}(u, v - u)\, dudv$$

after the transformation $u = x$ and $v = x + y$. Therefore, differentiating, we obtain

$$f_Z(z) = \int_{-\infty}^{\infty} f_{X,Y}(u, z - u)\, du.$$

Replacing in the formula of the expectation, we find that

$$\mathrm{Exp}(Z) = \int_{-\infty}^{\infty} z \int_{-\infty}^{\infty} f_{X,Y}(u, z - u)\, dudz = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} z f_{X,Y}(u, z - u)\, dudz.$$

Now, reset $u = x$ and $y = z - u$, so that $z = x + y$ and therefore

$$\mathrm{Exp}(Z) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x + y) f_{X,Y}(x, y)\, dxdy.. \qquad \triangle$$

**Theorem E.2** If $X$ and $Y$ are random variables that are jointly continuous and if $Z = g(X, Y)$ is a continuous random variable for some function $g : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, then

$$\mathrm{Exp}(Z) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f_{X,Y}(x, y)\, dxdy$$

whenever this integral exists.

**Properties:**

- $\mathrm{Exp}(aX + bY) = a\mathrm{Exp}(X) + b\mathrm{Exp}(Y)$.
- If $X$ and $Y$ are independent random variable, then

$$\mathrm{Exp}(XY) = \mathrm{Exp}(X)\mathrm{Exp}(Y).$$

**Example E.8** Let $X, Y$ be two uniformly distributed on the unit disc, so that

$$f_{X,Y}(x,y) = \begin{cases} \pi^{-1} & \text{if } x^2+y^2 \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Find $\mathrm{Exp}(\sqrt{X^2+Y^2})$.

**Solution.** Here, $Z = \sqrt{X^2+Y^2}$, so that $g(x,y) = \sqrt{x^2+y^2}$. Applying the formula of the expectation, we find

$$\mathrm{Exp}(Z) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \sqrt{x^2+y^2} f_{X,Y}(x,y)\, dxdy.$$

Since $f_{X,Y}(x,y) = 0$ when $x^2+y^2 > 1$, then

$$\mathrm{Exp}(Z) = \iint_D \frac{\sqrt{x^2+y^2}}{\pi}\, dA$$

where $D = \{(x,y) : x^2+y^2 \leq 1\}$. Using polar coordinates $x = r\cos\theta$ and $y = r\sin\theta$, we get

$$\mathrm{Exp}(X) = \int_0^{2\pi}\int_0^1 \frac{r^3}{\pi}\, drd\theta = \frac{1}{2}. \qquad \triangle$$

## Covariance

Recall that if $X$ and $Y$ are independent, then

$$\mathrm{Exp}(XY) = \mathrm{Exp}(X)\mathrm{Exp}(Y).$$

An interpretation of the last identity is that no information of $X$ and $Y$ mixed up the expectation of $XY$. We would like a measurement of how a random variable $X$ affects the outcome of another variable $Y$. In other words, we want to measure how $X$ and $Y$ are correlated!

**Definition E.6** The covariance of the random variables $X$ and $Y$ is the quantity denoted $\mathrm{Cov}(X,Y)$ and given by

$$\mathrm{Cov}(X,Y) = \mathrm{Exp}((X-\mu_X)(Y-\mu_Y)),$$

where $\mu_X = \mathrm{Exp}(X)$ and $\mu_Y = \mathrm{Exp}(Y)$.

**Remarks:**

- The $\mathrm{Cov}(X,Y)$ depends upon the scale of measurement. This is why, in practice, we normalize to obtain the correlation (coefficient):

$$\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sigma_X\sigma_Y},$$

  where $\sigma_X = \sqrt{\mathrm{Var}(X)}$ and $\sigma_Y = \sqrt{\mathrm{Var}(Y)}$. In this case, $\rho \in [-1,1]$.

- When $\rho(X,Y) \neq 0$ it means there is a linear dependence between $X$ and $Y$, getting stronger as $\rho$ gets closer and closer to $-1$ or $1$. We have two different types of correlation between $X$ and $Y$:

  1) $\rho(X,Y) > 0$, then when $X$ increases, then $Y$ increases.

  2) $\rho(X,Y) < 0$, then when $X$ increases, then $Y$ decreases.

**Theorem E.3** If $X$ and $Y$ are random variables with means $\mu_X$ and $\mu_Y$, respectively, then

$$\mathrm{Cov}(X,Y) = \mathrm{Exp}(XY) - \mathrm{Exp}(X)\mathrm{Exp}(Y).$$

**Example E.9** Assume that $X$ and $Y$ are uniformly distributed on the triangle with vertices $(-1,0)$, $(0,1)$, and $(1,0)$.

a) Find $\mathrm{Cov}(X,Y)$.

b) Are $X$ and $Y$ independent?

**Solution.**

a) The joint density function of $X$ and $Y$ are

$$f_{X,Y}(x,y) = \begin{cases} 1 & \text{if } y-1 \leq x \leq 1-y,\ 0 \leq y \leq 1 \\ 0 & \text{elsewhere} \end{cases}$$

Using the formula for the expected value of the random variable $Z = XY$, we get

$$\begin{aligned}\mathrm{Exp}(XY) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy f_{X,Y}(x,y)\,dxdy \\ &= \int_0^1 \int_{y-1}^{1-y} xy\,dxdy = 0.\end{aligned}$$

After some calculations, we find that

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy = \begin{cases} 1+x & \text{if } -1 < x < 0 \\ 1-x & \text{if } 0 < x < 1 \\ 0 & \text{elsewhere} \end{cases}$$

and

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx = \begin{cases} 2-2y & \text{if } 0 < y < 1 \\ 0 & \text{elsewhere.} \end{cases}$$

Therefore,

$$\mathrm{Exp}(X) = \int_{-\infty}^{\infty} x f_X(x)\,dx = 0$$

and

$$\mathrm{Exp}(Y) = \int_{-\infty}^{\infty} y f_Y(y)\,dy = \frac{1}{3}.$$

Hence

$$\mathrm{Cov}(X,Y) = \mathrm{Exp}(XY) - \mathrm{Exp}(X)\mathrm{Exp}(Y) = 0 - (0)\left(\frac{1}{3}\right) = 0.$$

b) In this case, we see that $f_{X,Y}(x,y) \neq f_X(x)f_Y(y)$ and therefore the random variables $X$ and $Y$ are not independent eventhough their covariance is 0!

**Remarks:**

- If $X$ and $Y$ are independent, then $\mathrm{Cov}(X,Y) = 0$.
- But, the fact that $\mathrm{Cov}(X,Y) = 0$ does not necessarily imply that the random variables are independent!
- If $X = Y$, then $\mathrm{Cov}(X,X) = \mathrm{Var}(X)$.

## Uniform distribution

If $X$ and $Y$ are two jointly random variables, then they have a uniform distribution on the square $[a,b] \times [c,d]$ if

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{(b-a)(d-c)} & \text{if } (x,y) \in [a,b] \times [c,d] \\ 0 & \text{elsewhere.} \end{cases}$$

For a general region $D$, two jointly random variables $X$ and $Y$ have the uniform distribution on the region $D$ if

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{\text{Area}(D)} & \text{if } (x,y) \in D \\ 0 & \text{elsewhere.} \end{cases}$$

The probability of the random vector $(X,Y)$ to be in a region $B \subset D$ is given by

$$P((X,Y) \in B) = \frac{1}{\text{Area}(D)} \iint_B dA = \frac{\text{Area}(B)}{\text{Area}(D)}.$$

## Bivariate Normal Distribution

Given $\rho \in (-1,1)$, two jointly random variables $X$ and $Y$ has a standard bivariate normal distribution if their joint density function is

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\Big( - \frac{1}{2(1-\rho^2)}(x^2 - 2\rho xy + y^2)\Big) \quad \forall x, y \in \mathbb{R}.$$

We can show that

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dxdy = 1.$$

and

1) $f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$, for any $x \in \mathbb{R}$;

2) $f_Y(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2}$, for any $y \in \mathbb{R}$.

This means $X$ and $Y$ has the stardard normal distribution.

The general form of the bivariate distribution is

$$f_{X,Y}(x,y) = \frac{e^{-Q/2}}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}$$

where

$$Q = \frac{1}{1-\rho^2}\Big[\frac{(x-\mu_X)^2}{\sigma_X^2} - 2\rho\frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} + \frac{(y-\mu_Y)^2}{\sigma_Y^2}\Big],$$

where $\mu_X = \text{Exp}(X)$, $\mu_Y = \text{Exp}(Y)$, $\sigma_X = \sqrt{\text{Var}(X)}$, $\sigma_Y = \sqrt{\text{Var}(Y)}$, and $\rho \in (-1,1)$.

# Chapter F: Moments

In this chapter, we assume that a probability space $(S, \mathcal{A}, P)$ is given.

## F.1 Moments

**Definition F.1** If $X$ is a random variable, then the $k$-th moment, with $k \geq 0$ an integer, is given by

$$M_k := \mathrm{Exp}(X^k),$$

if the expectation exists.

When $X$ is a continuous random variable with pdf $f_X$, then

$$M_k = \int_{-\infty}^{\infty} x^k f_X(x)\, dx.$$

**Example F.1** If $X \sim U(a, b)$, then

$$M_k = \int_a^b x^k\, dx = \frac{b^{k+1} - a^{k+1}}{k+1}. \qquad \triangle$$

**Definition F.2** If $X$ is a random variable with mean $\mu_X$, then the $k$-th centered moment, with $k \geq 0$ an integer is given by

$$CM_k := \mathrm{Exp}((X - \mu_X)^k),$$

if the expected value exists.

The variance of $X$ is the second centered moment of a random variable. Notice also that the $\mathrm{Var}(X)$ can be rewritten as a polynomial expression involving only $M_k$:

$$\mathrm{Var}(X) = \mathrm{Exp}(X^2) - (\mathrm{Exp}(X))^2 = M_2 - (M_1)^2.$$

## F.2 Moment Generating Function

Notice that, if all the moments of a continuous random variable $X$ exist, then

$$\sum_{k=0}^{\infty} \frac{t^k}{k!} \mathrm{Exp}(X^k),$$

may exists for a certain value of $t > 0$ (and therefore for any value less than $t$). If this happens, then

$$\sum_{k=0}^{\infty} \frac{t^k}{k!} \mathrm{Exp}(X^k) = \int_{-\infty}^{\infty} \sum_{k=0}^{\infty} \frac{t^k x^k}{k!} f_X(x)\, dx = \int_{-\infty}^{\infty} e^{tx} f_X(x)\, dx = \mathrm{Exp}(e^{tX}).$$

The expectation on the right-hand side is called the **moment generating function** of $X$ and is denoted by $M_X(t)$.

EXAMPLE F.2 If $X$ has the exponential distribution with parameter $\lambda$, then $f_X(x) = \lambda e^{-\lambda x}$ for $x \geq 0$ so that

$$M_X(t) = \int_0^\infty e^{tx} \lambda e^{-\lambda x}\, dx.$$

If $t < \lambda$, then $(t-\lambda)x < 0$ for any $x > 0$. Therefore, the above integral does exists with $t < \lambda$ and

$$M_X(t) = \frac{\lambda}{\lambda - t}.$$

When $t > \lambda$, then

$$M_X(t) = \infty.$$

The moment generating function $M_X$ only exists when $t < \lambda$. △

EXAMPLE F.3 Ìf $X$ has the normal distribution with mean 0 and variance 1, then

$$M_X(t) = \int_{-\infty}^\infty e^{tx} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\, dx = e^{\frac{1}{2}t^2}.$$

Using the fact that $e^t = 1 + t + t^2/2 + t^3/3! + \ldots$, we obtain

$$M_X(t) = 1 + \frac{1}{2}t^2 + \frac{1}{8}t^4 + \frac{1}{8 \cdot 3!}t^6 + \ldots.$$

△

The moment generating function is really useful to find the moments of a random variable.

THEOREM F.1 If $M_X(t)$ exists in a neighborhood of 0, then, for $k = 1, 2, \ldots$

$$\mathrm{Exp}(X^k) = M_X^{(k)}(0).$$

EXAMPLE F.4 Let $X$ be a random variable with the exponentital distribution of parameter $\lambda > 0$. Compute $M_3$.

**Solution.** From the previous examples, we have $M_X(t) = \frac{\lambda}{\lambda - t}$, for $t < \lambda$. We have

$$M_X'(t) = \frac{\lambda}{(\lambda - t)^2}, \quad M_X''(t) = \frac{2\lambda}{(\lambda - t)^3} \quad \text{and} \quad M_X^{(3)}(t) = \frac{6\lambda}{(\lambda - t)^4}.$$

Therefore,

$$M_3 = M_X^{(3)}(0) = \frac{6\lambda}{(\lambda - 0)^4} = 6\lambda^{-3}.$$

△

THEOREM F.2 If $X$ and $Y$ are independent random variables, then $X + Y$ has moment generating function

$$M_{X+Y}(t) = M_X(t) M_Y(t).$$

*Proof.* Since $X$ and $Y$ are independent, we have $f_{X,Y}(x,y) = f_X(x) f_Y(y)$. Hence,

$$\begin{aligned} M_{X+Y}(t) = \mathrm{Exp}(e^{t(X+Y)}) &= \int_{-\infty}^\infty \int_{-\infty}^\infty e^{t(x+y)} f_{X,Y}(x,y)\, dxdy \\ &= \int_{-\infty}^\infty \int_{-\infty}^\infty e^{tx} f_X(x) e^{ty} f_Y(y)\, dxdy \\ &= \mathrm{Exp}(e^{tX}) \mathrm{Exp}(e^{tY}). \end{aligned}$$

□

**Theorem F.3** Let $X$ be a continuous random variable. Assume the generating function $M_X$ satisfies $M_X(t) = \mathrm{Exp}(e^{tX})$ for all $t$ with $-\delta < t \leq \delta$ and for some $\delta > 0$. Then

a) There is a unique distribution with moment generating function $M_X$.

b) Furthermore, under these conditions, we have that $\mathrm{Exp}(X^k) < \infty$ for any $k = 1, 2, \ldots$ and
$$M_X(t) = \sum_{k=0}^{\infty} \frac{t^k}{k!} \mathrm{Exp}(X^k) \quad |t| < \delta.$$

**Example F.5** Let $X$ and $Y$ be independent random variables, $X \sim N(\mu_X, \sigma_X)$ and $Y \sim N(\mu_Y, \sigma_Y)$. Show that their sum $Z = X + Y$ has the normal distribution with parameters $\mu_X + \mu_Y$ and $\sigma_X^2 + \sigma_Y^2$.

**Solution.** Let $z = (x - \mu_X)/\sigma_X$, then
$$\begin{aligned} M_X(t) &= \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sqrt{2\pi\sigma_X^2}} e^{-\frac{1}{2\sigma_X^2}(x-\mu_X)^2}\, dx \\ &= \int_{-\infty}^{\infty} e^{t(\sigma_X z + \mu_X)} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}\, dz \\ &= e^{t\mu_X} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{t\sigma_X z} e^{-\frac{1}{2}z^2}\, dz \\ &= e^{t\mu_X + \frac{1}{2}t^2\sigma_X^2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z - t\sigma_X)^2}\, dz \end{aligned}$$

and hence $M_X(t) = \exp(t\mu_X + \frac{1}{2}t^2\sigma_X^2)$. Similarly, we have $M_Y(t) = \exp(t\mu_Y + \frac{1}{2}t^2\sigma_Y^2)$. Therefore, by Theorem F.2,
$$M_{X+Y}(t) = M_X(t)M_Y(t) = \exp\Big(t(\mu_X + \mu_Y) + \frac{1}{2}t^2(\sigma_X^2 + \sigma_Y^2)\Big).$$

By uniqueness of the moment generating function, we see that
$$X + Y \sim N(\mu_x + \mu_Y, \sigma_X^2 + \sigma_Y^2). \qquad \triangle$$

# Chapter G: Two Main Laws

We assume a probability space $(S, \mathcal{A}, P)$ is given.

## G.1 Law of Averages

Definition G.1 If $x_1, x_2, \ldots, x_n$ are real numbers, then the average is defined by

$$\sigma_n := \frac{x_1 + x_2 + \cdots + x_n}{n}.$$

Example G.1 We throw a fair die 10, 20, 40, and 60 times and record the result (check www.rolladie.net). Let $X_j$ be the random variable giving the number of dots of the face on which the die landed. Compute the average of the $X_j$'s. What do you observe? What number do you think the average should be close to?

In general, in probability, we would like a theorem telling us something like "if we repeat an experiment many times, then the average of the results approaches the underlying mean value."

Definition G.2 We say that the sequence $X_1, X_2, \ldots$ of random variables converges in mean square to the random variable $X$ if

$$\mathrm{Exp}\Big((X_n - X)^2\Big) \to 0 \quad \text{as } n \to \infty.$$

If this holds, then we write "$X_n \to X$ in mean-square as $n \to \infty$".

Theorem G.1 [Mean-Square Law of Averages]

Let $X_1, X_2, \ldots, X_n, \ldots$ be a list of uncorrelated random variables, each having mean $\mu$ and variance $\sigma^2$. Then,

$$\frac{X_1 + X_2 + \ldots + X_n}{n} \to \mu \quad \text{(in mean-square.)}$$

*Proof.* Write $S_n = X_1 + X_2 + \ldots + X_n$. Then, We have

$$\mathrm{Exp}(S_n/n) = \frac{1}{n}\mathrm{Exp}(X_1 + X_2 + \cdots + X_n) = \frac{1}{n} n\mu = \mu.$$

and

$$\mathrm{Exp}\Big((\sigma_n - \mu)^2\Big) = \mathrm{Exp}\left(\frac{(S_n - n\mu)^2}{n^2}\right) = \frac{1}{n^2}\mathrm{Var}(S_n).$$

The random variables are uncorrelated, meaning $\mathrm{Cov}(X_i, X_j) = 0$, for any $i \neq j$. Therefore,

$$\mathrm{Var}(X_1 + X_2 + \cdots + X_n) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + \cdots + \mathrm{Var}(X_n) = n\sigma^2.$$

Hence,

$$\mathrm{Exp}\Big((\sigma_n - \mu)^2\Big) = \frac{\sigma^2}{n} \to 0$$

as $n \to \infty$. Therefore, $\sigma_n \to \mu$ in mean-square as $n \to \infty$. $\square$

## Convergence in Probability

**Definition G.3** A sequence of random variable $X_1, X_2, \ldots$ converges in probability to a random variable $X$ if for any $\varepsilon > 0$,

$$\lim_{n\to\infty} P(|X_n - X| > \varepsilon) = 0.$$

**Example G.2** Let $X_1, X_2, \ldots$ be a sequence of discrete random variables. Assume that the probability mass function of $X_n$ is given by

$$P(X_n = 0) = 1 - \frac{1}{n} \quad \text{and} \quad P(X_n = n) = \frac{1}{n} \quad (n \geq 1).$$

Show that $X_n \to 0$ in probability as $n \to \infty$, but $X_n$ does not converge to 0 in mean square as $n \to \infty$.

**Solution.** We start with the convergence in probability. For $\varepsilon > 0$ and for any $n > \varepsilon$, we have

$$P(|X_n - 0| > \varepsilon) = P(X_n = n) = \frac{1}{n} \to 0 \quad (n \to \infty).$$

This implies that $X_n \to 0$ in probability when $n \to \infty$.

As for the mean square convergence, we have

$$\text{Exp}((X_n - 0)^2) = E(X_n^2) = (0)\left(1 - \frac{1}{n}\right) + (n^2)\left(\frac{1}{n}\right) = n \to \infty.$$

Hence, $X_n$ does not converge in mean square to 0.

**Remark:** The last example shows that convergence in probability does not imply convergence in mean.

On the other hand, convergence in mean does imply convergence in probability. This is Theorem G.3. To show this result, we need an intermediate result called Chebyshev's inequality.

**Theorem G.2** If $X$ is a random variable and $\text{Exp}(X^2) < \infty$, then for any $t > 0$

$$P(|X| \geq t) \leq \frac{1}{t^2}\text{Exp}(X^2).$$

Here is a consequence of Chebyshev's inequality.

**Theorem G.3** If a sequence of random variables $X_1, X_2, \ldots$ converges in mean-square to a random variable $X$, then it also converges in probability to $X$.

*Proof.* Assume that $\text{Exp}((X_n - X)^2) \to 0$, as $n \to \infty$. Given $\varepsilon > 0$, from Chebyshev's inequality, we get that

$$P(|X_n - X| > \varepsilon) \leq \frac{1}{\varepsilon^2}\text{Exp}((X_n - X)^2)$$

Since $\varepsilon$ is fixed and $\lim_{n\to\infty} \text{Exp}((X_n - X)^2) = 0$, we get

$$\lim_{n\to\infty} P(|X_n - X| > \varepsilon) = 0. \qquad \square$$

## G.2 Central Limit Theorem

Let $X_1$, $X_2$, $\ldots$ be independent random variables with the same means $\mu$ and variance $\sigma^2$. Letting $S_n = X_1 + X_2 + \ldots + X_n$, we have

$$\mathrm{Exp}(S_n) = n\mu \quad \text{and} \quad \mathrm{Var}(S_n) = n\sigma^2.$$

Then, define

$$Z_n = \frac{S_n - \mathrm{Exp}(S_n)}{\sqrt{\mathrm{Var}(S_n)}} = \frac{S_n - n\mu}{\sqrt{n}\sigma}.$$

Then, $\mathrm{Exp}(Z_n) = 0$ and $\mathrm{Var}(Z_n) = 1$. These are called the standardized version of the sum $S_n$.

Remarkably, the sequence of random variables $Z_n$ settles to a limit as $n \to \infty$. Even more impressive is the fact the distribution of the limit can be identified explicitly!

**Theorem G.4** Let $X_1$, $X_2$, $\ldots$ be independent and identically distributed random variables, each with mean $\mu$ and non-zero variable $\sigma^2$. The random variables $Z_n$ satisfies the following:

$$P(Z_n \leq z) \to \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-u^2/2}\, du \quad \forall z \in \mathbb{R}.$$

We won't prove this Theorem in this class. It requires a difficult result called the *Continuity Theorem*. We will therefore take for granted the Central Limit Theorem and use it in the next example.

**Example G.3** An unknown fraction $p$ of the population are jedi knights. It is desired to estimate $p$ with error not exceeding 0.005 by asking a sample of individuals. How large a sample is needed?

**Solution.** We use the Central Limit Theorem. Let $X_i$ be the random variable following a Bernouilli distribution, meaning $X_i = 0$ if the $i$-th person answers no and $X_i = 1$ if that same person answers yes. We have

$$\mathrm{Exp}(X_i) = p \quad \text{and} \quad \mathrm{Var}(X_i) = p(1-p).$$

Set $S_n = X_1 + X_2 + \ldots + X_n$, where $n$ is the size of the sample of individuals. We decide to estimate $p$ with $\hat{p} = S_n/n$. Our goal is to find $n$ such that $|\hat{p} - p| < 0.005$.

The good practice is to reach our goal with a certain degree of certainty. Therefore, we want to find $n$ such

$$P(|\hat{p} - p| > 0.005) < \alpha$$

where $\alpha$ is called the $p$-value or the tolerance value. In practice, we use $\alpha = 0.05$. Therefore, we want to find $n$ such that

$$P(|\hat{p} - p| \leq 0.005) < 0.95.$$

We can rewrite the previous probability as

$$P(|\hat{p} - p| \leq 0.005) = P\left(\frac{|S_n - np|}{\sqrt{np(1-p)}} \leq 0.005\frac{\sqrt{n}}{\sqrt{p(1-p)}}\right) = P\left(|Z_n| \leq 0.005\frac{\sqrt{n}}{\sqrt{p(1-p)}}\right),$$

where $Z_n = \frac{S_n - \mathrm{Exp}(S_n)}{\sqrt{\mathrm{Var}(S_n)}}$. However, $p(1-p)$ is unknown and the number $0.005\frac{\sqrt{n}}{\sqrt{p(1-p)}}$ is useless. We can see that $p(1-p) \leq \frac{1}{4}$ and therefore

$$P\left(|Z_n| \leq 0.005\frac{\sqrt{n}}{\sqrt{p(1-p)}}\right) \geq P(|Z_n| \leq 0.005\sqrt{4n}).$$

From the Central Limit Theorem, if $n$ is large enough, then the distribution of $Z_n$ approaches the distribution of a $Z \sim N(0,1)$. Therefore

$$P\left(|Z_n| \leq 0.005\frac{\sqrt{n}}{\sqrt{p(1-p)}}\right) \gtrapprox P(|Z| \leq 0.005\sqrt{4n}) = P(Z \leq 0.005\sqrt{4n}) - P(Z \leq -0.005\sqrt{4n}).$$

For that probability to be bigger than 0.95, we have to have $0.005\sqrt{4n} \geq 1.96$, which means $n \gtrapprox 40,000$. $\triangle$

# Appendix L: Logic and Proofs

## L.1 Mathematical Statements

A statement is a sentence (written in words, mathematical symbols, or a combination of the two) that is either true or false.[1]

**Example L.1**

a) $4 + 11 = 15$.
This is a statement and it is true.

b) $x > 5$.
This is not a statement. Grammatically, it is a complete sentence, written in mathematical symbols, with a subject ($x$) and a predicate (*is greater than* 5). The sentence, however, is neither true or false because the value of $x$ is not specified.

c) If $x = 5$, then $x > 0$.
This is a statement and it is true.

d) There exists a positive integer $n$ such that $n > 2$.
This is a statement and it is true.

e) Is the number 20 an even number?
This is not a statement. A question is neither true or false.

A proof is a piece of writing that demonstrates that a particular statement is true. A statement that we prove to be true is often called a theorem. A statement that we assume without proof is an axiom. A definition is an agreement between the writer (or professor) and the reader (or the student) as to the meaning of a word or phrase. A definition needs no proof.

## L.2 Logic and Mathematical Language

In the section on set theory, you will have the chance to practice the methods of proof presented below.

### Negation

If $P$ is a statement, then it has a truth value: true or false. The negation of a statement $P$ is defined as *it is not the case that* $P$. The negation of a statement $P$ will be abbreviated by *not* $P$ or $\neg P$.

[1]Most of the material presented here is from the really good notes retrieved online at https://sites.math.washington.edu/~conroy/m300-general/ConroyTaggartIMR.pdf. Some passages might entirely be copied or modified slightly from this resource.

Example L.2 Consider the statement $P$: "2 is an even integer". The negation of $P$ is "It is not the case that 2 is an even integer" that may be rewritten as "2 is not an even integer". We may even go further and rewrite $\neg P$ as follows: "2 is an odd integer". Notice that $P$ is true and $\neg P$ is false.

**Conjunction and Disjunction**

Let's consider two statements $P$ and $Q$.

- The conjunction of $P$ and $Q$ is the statement "$P$ and $Q$". It is denoted by $P \wedge Q$ and it is true only when $P$ and $Q$ are true; otherwise it is false.
- The disjunction of $P$ and $Q$ is the statement "$P$ or $Q$". It is denoted by $P \vee Q$ and it is true when one of the two statements is true.

We can use a truth table to illustrate the conjunction and disjunction of two statements $P$ and $Q$ as shown below.

| $P$ | $Q$ | $P \wedge Q$ |
|---|---|---|
| $T$ | $T$ | $T$ |
| $T$ | $F$ | $F$ |
| $F$ | $T$ | $F$ |
| $F$ | $F$ | $F$ |

(a) Truth table for $P \wedge Q$

| $P$ | $Q$ | $P \vee Q$ |
|---|---|---|
| $T$ | $T$ | $T$ |
| $T$ | $F$ | $T$ |
| $F$ | $T$ | $T$ |
| $F$ | $F$ | $F$ |

(b) Truth table for $P \vee Q$

Example L.3

a) Consider the statement $P$: "2 is a positive integer" and the statement $Q$: "$-4$ is a negative integer". The statement $P \wedge Q$ is true because $P$ and $Q$ are true. But $P \wedge (\neg Q)$ is not true because $\neg Q$: "$-4$ is not a negative integer" is false.

b) Consider the same statements from part a). The statement $P \vee Q$ is true because the integer 2 is a positive integer and only one of the statements $P$, $Q$ needs to be true. The statement $(\neg P) \vee Q$ is also true because the statement $Q$ is true ($-4$ is a negative number). But the statement $(\neg P) \vee (\neg Q)$ is not true because $\neg P$ and $\neg Q$ are both false.

We can take the negation of a conjunction and of a disjunction.

Example L.4 A friend tells you the conditions to come to his party. He tells you that you must wear green clothes only AND bring a one-page explanation of why you are at his party. A person that wants to go to your friend's party must satisfies *both* conditions. Anyone who is wearing a non-green piece of clothe will not be allowed at the party. Also, anyone who did not write the one-page essay will not be allowed at the party. Therefore, anyone who does not wear a green outfit or anyone who did not write the one-page essay will not come to the party.

Conclusion: The negation of $P \wedge Q$ is $(\neg P) \vee (\neg Q)$.

Example L.5 Your friend decides to be more welcoming. He tells you the conditions to come to his party remains the same, but only one of them must be meet. In other words, you may wear green clothes only OR bring a one-page explanation of why you are at his party. A person that wants to go to your friend's party must satisfy one of the two conditions. But if the person is not dressed in green clothes and does not bring the one-page essay, then unfortunately, that person will not be allowed to join the party. In other words, if both conditions are not respected by a person, then that person will not be allowed to join the party.

Conclusion: The negation of $P \vee Q$ is $(\neg P) \wedge (\neg Q)$.

### Conditional

A lot of statements we will encounter are of the form "If $P$, then $Q$". These statements are called conditional statements. We will use the following notation "$P \Rightarrow Q$" to denote a conditional statement.

The truth value of the statement $P \Rightarrow Q$ depends on the truth values of $P$ and $Q$.

**Example L.6** We think of $P \Rightarrow Q$ as an agreement. Joe makes a deal with his parents. Let $P$: "Joe did the dishes after dinner" and $Q$: "Joe got \$5". The agreement is

$$P \Rightarrow Q : \text{ If Joe did the dishes, then he got \$5.}$$

Joe is not required to do the dishes (it is a compulsory act for his love for his family, but also for his love for money). In the case that Joe did the dishes ($P$ is true) and got paid ($Q$ is true), the agreement is met ($P \Rightarrow Q$ is true). In the case that Joe didn't do the dishes ($P$ is false) and didn't get paid ($Q$ is false), the agreement is met ($P \Rightarrow Q$ is true). Since Joe is not required to wash the dishes, his parents may choose to give him \$5 for some other reason. That is, in the case Joe did not do the dishes ($P$ is false) and got \$5 anyway ($Q$ is true), the agreement is still met ($P \Rightarrow Q$ is true). The only instance in which the agreement is not met ($P \Rightarrow Q$ is false) is in the case that Joe did wash the dishes ($P$ is true), but did not get the money from his parents ($Q$ is false).

To summarize, the statement $P \Rightarrow Q$ is true unless $P$ is true and $Q$ is false, like it is shown in the table below.

| $P$ | $Q$ | $P \Rightarrow Q$ |
|---|---|---|
| $T$ | $T$ | $T$ |
| $T$ | $F$ | $T$ |
| $F$ | $T$ | $F$ |
| $F$ | $F$ | $T$ |

Truth table for the conditional

To explain the negation of the conditional, we use Joe's story from the previous example. Joe claims that his parents broke their verbal contract, while the parents deny Joe's claim. In other words, Joe's parents say that $P \Rightarrow Q$ is true, while Joe says that $\neg(P \Rightarrow Q)$ is true. If you were Joe's lawyer, what evidence would you have to provide to win the case? You would need to show that Joe washed the dishes and did not get paid. That is, you would need to show that $P \wedge (\neg Q)$ is true.

Conclusion: The negation of the conditional statement $P \Rightarrow Q$ is the statement $P \wedge (\neg Q)$.

### Converse and Contrapositive

**Definition L.1** The converse of $P \Rightarrow Q$ is the statement $Q \Rightarrow P$. The contrapositive of $P \Rightarrow Q$ is the statement $(\neg Q) \Rightarrow (\neg P)$.

**Example L.7** Consider the statements $P$: "Valérie's cat is hungry" and $Q$: "Valérie's cat meows".

- The implication $P \Rightarrow Q$ reads as "If Valérie's cat is hungry, then the cat meows".

- The converse $Q \Rightarrow P$ of $P \Rightarrow Q$ reads as "If Valérie's cat meows, then the cat is hungry". Notice that in this contexte, $P \Rightarrow Q$ does not have the same truth value of $Q \Rightarrow P$. For instance, the cat might moews because it wants to be pet.
- The contrapositive of $P \Rightarrow Q$ reads as "If Valérie's cat does not meow, then the cat is not hungry". You can check that $P \Rightarrow Q$ has the same truth value of $(\neg Q) \Rightarrow (\neg P)$.

### Equivalent statement

DEFINITION L.2 The statement $P$ if and only if $Q$, written $P \iff Q$, is readily the statement

$$(P \Rightarrow Q) \wedge (Q \Rightarrow P).$$

### Quantifiers

Suppose $n$ is an integer and $P(n)$ is a statement about $n$.

- If $P(n)$ is true for at least one integer $n$, then we say "There exists $n$ such that $P(n)$". This type of statement is called an existence statement and the symbol $\exists$ is used as a shortcut for the "there exists" part.
- If $P(n)$ is true no matter what value $n$ takes, then we say "For all $n$, $P(n)$". This type of statement is called a universal statement and the symbol $\forall$ is used as a shortcut for the "For all" part.

These statements are called quantified statements.

EXAMPLE L.8 Assume throughout this example that $n$ is an integer.

a) "There exists $n$ such that $n > 0$". In this statement, $P(n)$ is "$n > 0$. The statement $P(10)$ is true because $10 > 0$, therefore the statement "$\exists n$ such that $n > 0$" is true because $P(n)$ is true for at least one integer $n$.

b) "For all $n$, $n > 0$". The statement $P(-1)$ is false since $-1$ is not greater than 0. Therefore, the statement "$\forall n$, $P(n)$" is false because $P(n)$ is *not* true for every integer $n$.

c) "$\exists n$ such that $|n| < 0$". This statement is false because there is no integer $n$ with $|n| < 0$; the absolute value turns every integer into a positive or zero integer.

Here are the ways to negate a quantified statement:

- The negation of "$\exists n$ such that $P(n)$" is "$\forall n$, $\neg(P(n))$".
- The negation of "$\forall n$, $P(n)$" is "$\exists n$ such that $\neg(P(n))$".

## L.3 Methods of Proof

We will cover some methods to prove mathematical statements. The two we will cover are direct proofs of a conditional statement and proofs by contradiction.

### Direct Proof

There are many ways to proof a conditional statement. The method covered is called "direct proof". If one of the other ways is needed later on in the semester, then I will explain it to you on the spot. This is an agreement between you and me ;).

The "direct proof" method works as followed. We assume the hypothesis (the statement just after the "if") and use definitions, logic, and previously proved results to reach the desired conclusion (the statement after the "then").

Example L.9 Prove the following statement: If $a$ and $b$ are even integers, then $a + b$ is an even integer.

**Solution.** Suppose that $a$ and $b$ are even integers. In other words, this means $a$ and $b$ are multiples of 2: There exists an integer $n$ such that $a = 2n$ and there exists an integer $m$ such that $b = 2m$. Then

$$a + b = 2n + 2m = 2(n + m).$$

This implies that $a + b$ is a multiple of 2 and therefore it is an even integer. $\triangle$

### Proof by Contradiction

In a proof using the method called contradiction, the fact that a statement and its negation have opposite truth values is used. Therefore, to prove that $P$ is true, we suppose instead that the statement $\neg P$ is true and apply logic, definitions, and previous results to arrive at a conclusion known to be false. Then this will imply $\neg P$ must be false and thus $P$ must be true.

Example L.10 No integer is both even and odd.

**Solution.** Suppose that there is an integer $n$ that is both even and odd (the negation of the statement "$\forall n$, $n$ is neither even or odd", which is equivalent to the statement in the example). Since $n$ is assumed even, $n = 2k$ for some integer $k$. But $n$ is also assumed odd, so $n = 2l + 1$. Therefore, since $n = n$, we have

$$2k = 2l + 1 \Rightarrow 2k - 2l = 1 \Rightarrow 2(k - l) = 1.$$

Since $k - l$ is an integer, the last equation means that 1 is a multiple of 2 (or that 1 is divisible by 2), which is clearly false! Therefore the assumption that there is an integer $n$ that is both even and odd must be false and it turns out that no integer is both even and odd. $\triangle$

### Proof of An Equivalent Statement

To prove the statement "$P \iff Q$", it must be shown that $P \Rightarrow Q$ is true *and* $Q \Rightarrow P$ is true.

### Proof of An Existential Statement

To prove a statement of the form "there exists an $n$ such that $P(n)$", the technique used is "construction". This means the object $n$ will be found and be demonstrated that $P(n)$ is true for this choice of $n$.

### Proof of A Universal Statement

The proof of a statement of the form "for all objects $n$, $P(n)$" is rather more subtle. It is really hard to deal with all objects $n$ at once. Instead, we think of an equivalent way to interpret a universal statement. In fact, the statement "for all objects $n$, $P(n)$" is equivalent to the statement "If $n$ is such an object, then $P(n)$". For example, the statement "For all integers $n$, $|n| \geq 0$" has the same meaning as "If $n$ is an integer, then $|n| \geq 0$".

Therefore, to prove a universal statement, we first select a single *arbitrary* object and proved that the conclusion is true for that object. It is really important that the object chosen was arbitrary.

**Example L.11** Prove the following statement: For all odd integers $n$ and $m$, $nm$ is odd.

**Solution.** Suppose that $n$ is an odd integer and that $m$ is an odd integer. This means there exist integers $k$ and $l$ such that $n = 2k + 1$ and $m = 2l + 1$. Then

$$nm = (2k + 1)(2l + 1) = 4kl + 2k + 2l + 1 = 2(2kl + k + l) + 1.$$

Therefore, the product $nm$ takes the form of an odd integer. △

# Appendix S: Set Theory

## S.1 Terminology

A set is presented with a list of its elements surrounded by curly brackets. For instance, $\{1, 2, 3, 4\}$ is a way to present a set. A set may be presented as a family of objects satisfying a certain statement $P(n)$. For instance, $\{n : n \text{ is an odd integer}\}$ would describe the set of odd integers $\{\ldots, -3, -1, 1, 3, \ldots\}$.

Uppercase letters are usually used to denote sets and lowercase letters are used to denote an arbitrary element of a set. For example, $A = \{1, 2, 3, 4\}$. The notation $a \in A$ is used to mean "the element $a$ belongs to the set $A$". The negation of $a \in A$ will be denoted by $x \notin A$. This means the element $a$ does not belong to the set $A$.

A reference set or universal set is a set $U$ containing all elements under consideration. For example, a sample space $S$ would be a universal set for an experiment. All the definitions below is based on the assumption of the existence of a universal set $U$.

**Definition S.1** The null set is the set $\varnothing$ containing no element.

**Definition S.2** If $A$ and $B$ are two sets, then $A$ is a subset of $B$ if all the elements of the set $A$ belongs to the set $B$. We denote this by $A \subset B$. The family of all subsets, called the power set, of a set $A$ is denoted by $2^A$.

**Note:** To prove that a set $A$ is a subset of $B$, the statement "$\forall x \in A, x \in B$" should hold. In other words, we have to prove that the statement "$x \in A \Rightarrow x \in B$.

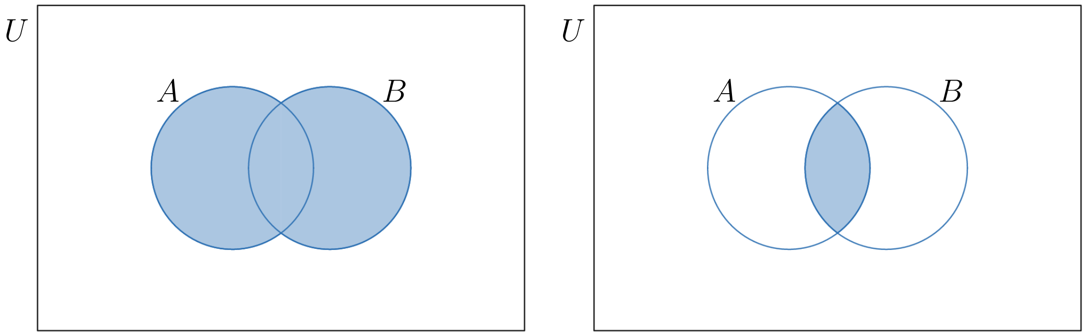
**Definition S.3** If $A$ and $B$ are two sets, then we say that $A$ is equal to $B$, denoted by $A = B$, if $A \subset B$ and $B \subset A$.

In other words, all the elements of $A$ are the same as the elements of $B$.

**Definition S.4** If a set $A$ has a finite number of elements, then $\#A$ denotes the number of elements in $A$.

## S.2 Operations With Sets

**Definition S.5** a) The set $A \cup B$ is the union of $A$ and $B$. It is the set of elements from $A$ and from $B$. Note: $A \cup S = S$.
b) The set $A \cap B$ is the set of all elements that are common to $A$ and $B$. It is called the intersection of $A$ and $B$. Note: $A \cap S = A$.

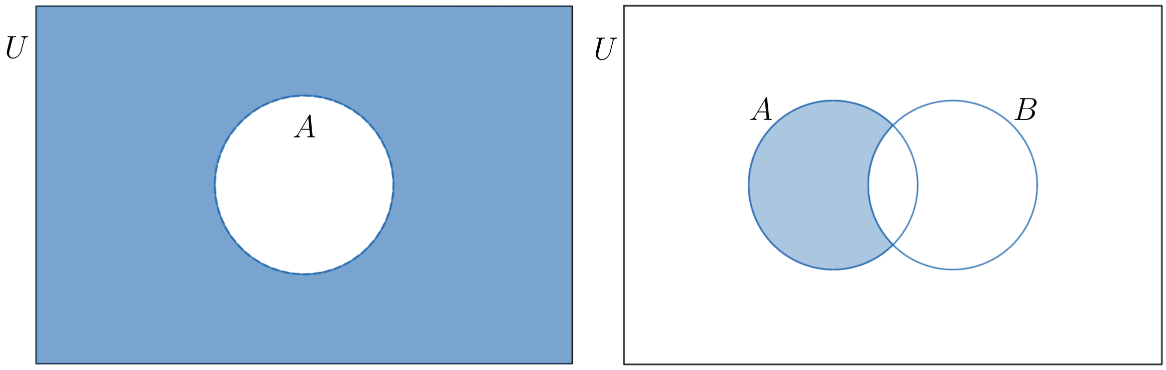



(a) Union (b) Intersection

**Notes:**

- Using the mathematical language from Appendix L, we can rewrite the union of $A$ and $B$ as followed: $A \cup B = \{x \in U \,:\, x \in A \text{ or } x \in B\}$.
- Similarly, we can rewrite the intersection of $A$ and $B$ as followed: $A \cap B = \{x \in U \,:\, x \in A \text{ and } x \in B\}$.

DEFINITION S.6 a) The complement of a set is the new set $\overline{A}$ of all elements in $U$ that are not in $A$. Note that $A \cup \overline{A} = U$.
b) The set difference of two sets $A$ and $B$ is the set $A \cap \overline{B}$. In other words, it is the set of elements that are in $A$, but not in $B$.




(a) Complement (b) Difference

**Notes:**

- We can rewrite the complement of a set $A$ as followed: $\overline{A} = \{x \in U \,:\, x \notin A\}$.
- We can rewrite the set difference of $A$ and $B$ as followed: $A \cap \overline{B} = \{x \in U \,:\, x \in A \text{ and } x \notin B\}$.

DEFINITION S.7 Two sets, $A$ and $B$, are disjoint or mutually exclusive if $A \cap B = \varnothing$.

EXAMPLE S.1 Let $U = \{1, 2, 3, 4, 5\}$. The sets $A = \{1, 2\}$ and $B = \{3\}$ are mutually exclusive because they have nothing in common, meaning $A \cap B = \varnothing$.

## S.3 Important Laws for Set Algebra

**Theorem S.1**

a) Commutative laws:

$$A \cup B = B \cup A \quad \text{and} \quad A \cap B = B \cap A.$$

b) The distributive laws:

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C) \tag{S.1}$$
$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C). \tag{S.2}$$

*Proof.* We will prove one of the commutative laws and one of the distributive laws. The other formulas are left as an exercise.

a) To prove the equation $A \cup B = B \cup A$, we have to show (i)$A \cup B \subset B \cup A$ and; (ii)$B \cup A \subset A \cup B$.

   (i) Assume $x \in A \cup B$. Then, $x \in A$ or $x \in B$ by definition of the union of sets. But, this is the same thing as writing $x \in B$ or $x \in A$ (the order of the presentation is unimportant). Therefore, the element $x$ belongs to $B \cup A$ by definition of the union of $B$ and $A$.

   (ii) Now, assume $x \in B \cup A$. Then $x \in B$ or $x \in A$ from the definition of $B \cup A$. Since the order of the presentation is unimportant, $x \in A$ or $x \in B$. Therefore, $x \in A \cup B$ by definition of the union of $A$ and $B$.

   Let's wrap this up. We just proved that $A \cup B \subset B \cup A$ and that $B \cup A \subset A \cup B$. From the definition of equality of sets, this means $A \cup B = B \cup A$.

b) We prove the equation $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.

   (i) We start by proving that $A \cap (B \cup C) \subset (A \cap B) \cup (A \cap C)$. Assume $x \in A \cap (B \cup C)$. Then, by definition of the intersection of two sets, this means $x \in A$ and $x \in B \cup C$. By definition of the union of two sets, $x \in B \cup C$ implies that $x \in B$ or $x \in C$. Since $x$ belongs to $A$ in both cases, then if $x$ belongs to $B$, we conclude that $x$ belongs to $A \cap B$, but if $x$ belongs to $C$, we conclude that $x$ belongs to $A \cap C$. Therefore, $x$ belongs to $A \cap B$ or to $A \cap C$. By definition of the union, the element $x$ belongs to $(A \cap B) \cup (A \cap C)$.

   (ii) Now we prove that $(A \cap B) \cup (A \cap C) \subset A \cap (B \cup C)$. Assume $x \in (A \cap B) \cup (A \cap C)$. Then, this means that $x$ belongs to $A \cap B$ or to $A \cap C$. The fact $x \in A \cap B$ implies that $x \in A$ and $x \in B$. The second fact that $x \in A \cap C$ implies that $x \in A$ and $x \in C$. Since $x$ belongs to $A$ in both cases, we see that $x \in A$ and $x$ belongs to $B$ or to $C$. In other words, the element $x$ belongs to $A$ and to $B \cup C$, which means that $x \in A \cap (B \cup C)$.

   From (i) and (ii), we can then conclude that $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.

□

**Theorem S.2** De Morgan's laws:

$$\overline{A \cap B} = \overline{A} \cup \overline{B} \quad \text{and} \quad \overline{A \cup B} = \overline{A} \cap \overline{B}.$$

*Proof.* We only prove the first equation and leave the proof of the second one as an exercise.

Suppose $x \in \overline{A \cap B}$. From the definition of the complement of a set, $x \notin A \cap B$. this means that $x$ does not belong to $A \cap B$. We have to negate the definition of intersection. We have $x \in A \cap B$ when $x \in A$ *and* $x \in B$. The negation is $x \notin A$ or $x \notin B$. Therefore, $x \in \overline{A}$ or $x \in \overline{B}$. From the definition of the union of two sets, we see that $x \in \overline{A} \cup \overline{B}$.

Now, assume $x \in \overline{A} \cup \overline{B}$. This means that $x \notin A$ or $x \notin B$. From the last paragraph, this last statement is the negation of $x \in A \cap B$. Therefore, $x \notin A \cap B$. In other words, $x$ belongs to $\overline{A \cap B}$.

We can then conclude that $\overline{A \cap B} = \overline{A} \cup \overline{B}$. $\square$