CAS MA581 PROBABILITY

Spring 2024

Frank Yang
Professor Dhama
MWF 10:10 – 11:00

Contents

1 Probability Basics

1.2 Set Theory

Definition 1.1. A **set** is a collection of elements (usually numbers).

If x is an element of set A, then

$$x \in A$$

otherwise,

$$x \notin A.$$

An empty set is denoted by ϕ.

Definition 1.2. If every element of set A is an element of set B, then A is said to be a **subset** of B: $A \subset B$

Two sets A and B are said to be equal when they have the same elements:

$$A = B \implies A \subset B \text{ and } B \subset A$$

Additional notation:

Definition 1.3.
$\mathbb{R}$ is the set of real numbers
$\mathbb{Q}$ is the set of rationals
$\mathbb{Z}$ is the set of integers
$\mathbb{N}$ is the set of positive integers
$\therefore \mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R}$

Interval notation:
Given $a, b \in \mathbb{R}$:

$$[a, b] = \{x \in \mathbb{R} : a \leq x \leq b\}$$

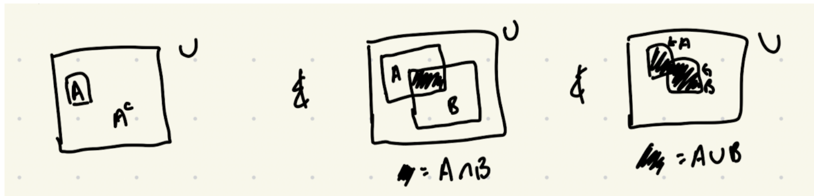
$$[a, b) = \{x \in \mathbb{R} : a \leq x < b\}$$

Set operations (union, intersection, complement):
Let $A, B,$ and E be subsets of U.

- The complement of E is $E^c = \{x \mid x \notin E\}$
- Intersection: $A \cap B = \{x \mid x \in A \text{ and } x \in B\}$
- Union: $A \cup B = \{x \mid x \in A \text{ or } x \in B\}$

Given $A \subset B$, the occurrence of A implies the occurrence of B. Given A^c, A doesn't occur. Given $A \cap B$, A and B occur. Given $A \cup B$, A or B occurs.

Venn diagrams can be used to explain these set operations:



Example 1.4. Given $U = \mathbb{R} = (-\infty, +\infty), E = [0, 3], A = (-\infty, 6), B = (5, +\infty)$, find $E^c, A \cap B$, and $A \cup B$.
$E^c = (-\infty, 0) \cup (3, +\infty)$.
$A \cap B = (5, 6)$.
$A \cup B = (-\infty, +\infty)$.

Theorem 1.5. De-Morgan's Law: Let A and B be subsets of U. Then,

1. $(A \cup B)^c = A^c \cap B^c$
2. $(A \cap B)^c = A^c \cup B^c$

A rigorous proof of Theorem 1.5 proceeds as follows:

$$(A \cup B)^c \subset A^c \cap B^c?$$

$$\text{Let } x \in (A \cup B)^c \implies x \notin A \cup B \implies x \notin A \text{ and } x \notin B \implies$$

$$x \in A^c \text{ and } x \in B^c \implies x \in A^c \cap B^c \therefore (A \cup B)^c \subset A^c \cap B^c$$

$$A^c \cap B^c \subset (A \cup B)^c?$$

$$\text{Let } x \in A^c \cap B^c \implies x \in A^c \text{ and } x \in B^c \implies$$

$$x \notin A \text{ and } x \notin B \implies x \notin A \cup B \implies x \in (A \cup B)^c \therefore A^c \cap B^c \subset (A \cup B)^c$$

$$\therefore (A \cup B)^c = A^c \cap B^c$$

Some other properties: Let A, B, C be subsets of U. Then,

- $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
- $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
- $A \cap B = B \cap A$
- $(\bigcup_{n=1}^{N} A_n)^c = \bigcap_{n=1}^{N} A_n^c$ where $\bigcap_{n=1}^{N} A_n = \{x \mid x \in A_n \text{ for all } n = 1, 2, \ldots, N\}$ and $\bigcup_{n=1}^{N} A_n = \{x \mid x \in A_n \text{ for all } n = 1, 2, \ldots, N\}$ [similarly, $\bigcap_{n=1}^{\infty} A_n = \{x \mid x \in A_n \text{ for all } n \in \mathbb{N}\}$]

Example 1.6. $U = \mathbb{R}$, for each $n \in \mathbb{N} = \{1, 2, \dots\}$. Define $A_n = [\frac{1}{n}, 2 + \frac{1}{n})$.

1. First, try $\bigcap_{n=1}^{3} A_n$ and $\bigcup_{n=1}^{3} A_n$.
   $\bigcap_{n=1}^{3} A_n = A_1 \cap A_2 \cap A_3 = [1,3) \cap [\frac{1}{2}, \frac{5}{2}) \cap [\frac{1}{3}, \frac{7}{3}) = [1, \frac{7}{3})$
   $\bigcup_{n=1}^{3} A_n = [\frac{1}{3}, 3)$

2. Next, try $\bigcap_{n=1}^{\infty} A_n$ and $\bigcup_{n=1}^{\infty} A_n$.
   $\bigcap_{n=1}^{\infty} A_n = [1,2]$ because, if we take $\frac{1}{n}$ to infinity, it converges to 0, thus $2 + \frac{1}{n} = 2$.
   (If $x \in \bigcap_{n=1}^{\infty} A_n \iff$ (if and only if) $x \in [\frac{1}{n}, 2 + \frac{1}{n})$ for all $n \in \mathbb{N} \iff x \geq 1$ and $x \leq 2 \iff x \in [1,2]$.)
   $\bigcup_{n=1}^{\infty} A_n = (0,3)$ because the minimum approaches, but never reaches 0, while the maximum is 3(exclusive) at A_1.
   We note, for each $n \in \mathbb{N}, A_n \subset (0,3) \therefore \bigcup_{n=1}^{\infty} A_n \subset (0,3)$.
   Now, to show $(0,3) \subset \bigcup_{n=1}^{\infty} A_n$:
   Let m be a positive integer such that $x \geq \frac{1}{m}$. If $m = 1$, $x \in [1,3) = A_1$. If $m \geq 2$,then $\frac{1}{m} \leq x < \frac{1}{m-1}$.
   $\therefore \frac{1}{m} \leq x < \frac{1}{m-1} \leq 1 < 2 + \frac{1}{m}$, hence, $x \in A_m = [\frac{1}{m}, 2 + \frac{1}{m}) \implies x \in A_m \subset \bigcup_{n=1}^{\infty} A_n$.

2 Mathematical Probability

2.1 Sample Space and Events

Definition 2.1. A **random experiment** is an experiment whose outcome can't be predicted.

Definition 2.2. The **sample space** is the set of all possible outcomes of the random experiment(denoted by Ω).

Definition 2.3. An **event** is a subset of the sample space. The set of all events is denoted by $\mathcal{F}$.

Example 2.4. Tossing a coin:

$$\Omega = \{H, T\}; \mathcal{F} = \{\phi, \{H\}, \{T\}, \Omega\}$$

If the coin were tossed twice,

$$\Omega = \{HH, TT, TH, HT\}; |\mathcal{F}| = 2^4 = 16$$

Consider a random experiment of tossing a coin three times. Write the event that the total number of heads is two ($E = \{HHT, HTH, THH\}$).
Assumptions: Let Ω be the sample space. We assume that the collection of elements $\mathcal{F}$ satisfies:

1. $\phi \in \mathcal{F}, \Omega \in \mathcal{F}$
2. If $A_1, A_2, \ldots$ are in $\mathcal{F}$, then $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$
3. If $A \in \mathcal{F}, A^c \in \mathcal{F}$

2.2 Axioms of Probability

Let Ω be the sample space, and $\mathcal{F}$ be the collection of events. A probability(or probability measure) $\mathbb{P}$ is the function from $\mathcal{F}$ to $[0,1]$. That is, $\mathbb{P} : \mathcal{F} \longrightarrow [0,1]$ such that the following Kolmogorov axioms are satisfied:

1. $\mathbb{P}(A) \geq 0$
2. If $A_1, \ldots$ is a collection of mutually disjoint events, then $\mathbb{P}(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mathbb{P}(A_n)$
3. $\mathbb{P}(\Omega) = 1$

Example 2.5. Consider the random experiment of rolling a dice.

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

Event A is when we get a number divisible by 2, and Event B is when we get a number divisible by 3.

$$A = \{2, 4, 6\}, B = \{3, 6\}$$

$$A \cup B = \{2, 3, 4, 6\}$$

$\mathbb{P}(A) = \frac{3}{6}; \mathbb{P}(B) = \frac{2}{6}$
$\mathbb{P}(A \cup B) = \frac{4}{6} < \mathbb{P}(A) + \mathbb{P}(B) = \frac{5}{6}$

Example 2.6. Now consider the same experiment from Example 1.11, but where Event A is getting an odd number, and Event B is getting an even number.

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

$\mathbb{P}(A) = \frac{3}{6} = \mathbb{P}(B)$
$\mathbb{P}(A \cup B) = 1 = \mathbb{P}(A) + \mathbb{P}(B)$

The triple $(\Omega, \mathcal{F}, \mathbb{P})$ is a called a probability space.
Properties of Probability:

1. For any event E, $\mathbb{P}(E^c) = 1 - \mathbb{P}(E)$
2. $\mathbb{P}(\phi) = 0$

3. If $A \subset B$, then $\mathbb{P}(A) \leq \mathbb{P}(B)$
4. (The Inclusion-Exclusion Property) Let A and B be two events. Then $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$

Example 2.7. For a community, 60% of people play cricket, 50% play football, and 80% either play cricket or football. What is the probability that a randomly chosen person plays both?
Let A be the event that a randomly chosen person plays cricket, and B be the event that a randomly chosen person plays football. We are given that $\mathbb{P}(A) = 0.6, \mathbb{P}(B) = 0.5, \mathbb{P}(A \cup B) = 0.8$.
$\implies 0.8 = 0.6 + 0.5 - \mathbb{P}(A \cap B) \therefore \mathbb{P}(A \cap B) = 0.3$

2.3 Specifying Probabilities

Definition 2.8. Proposition: (Probabilities for countable sample space) Let Ω be a countable sample space. Then, for each event E we have

1. $\mathbb{P}(E) = \sum_{\omega \in A} \mathbb{P}(\{\omega\})$
2. $\sum_{\omega \in A} \mathbb{P}(\{\omega\}) = 1$

Also, the number of subsets in $\Omega = \{1, \ldots, N\}$ is 2^N.

2.3.1 Classical Probability

Let Ω be a finite sample space with, say N, equally likely outcomes of a random experiment. Here, $\Omega = \{\omega_1, \omega_2, \ldots, \omega_N\}$ and ϕ denotes the probability of each outcome. By the probability axioms,

$$1 = \sum_{i=1}^{N} \mathbb{P}(\{\omega_i\}) = \sum_{i=1}^{N} \phi = N\phi \implies \phi = \frac{1}{N}$$

Result: Let Ω be a finite sample space where each outcome is equally likely. Then, for each event A,

$$\mathbb{P}(A) = \frac{N(A)}{N(\Omega)},$$

where $N(A)$: number of ways that event A can occur, $N(\Omega)$: total n of possible outcomes.

Example 2.9. Consider the random experiment of rolling two balanced dices. Calculate the probability of the event that

1. The sum of the dice is 5.
2. Both dices come up the same number.

Solution: The sample space for this experiment is

$$\Omega = \{(i, j) : 1 \le i \le 6, 1 \le j \le 6\} = \{(1, 1), (1, 2), \dots\}$$

$$\text{so, } N(\Omega) = 6^2 = 36$$

For each event E, we have

$$\mathbb{P}(E) = \frac{N(E)}{N(\Omega)} = \frac{N(E)}{36}$$

1. Let A be the event that the sum of the dice is 5. Therefore, $A = \{(1, 4), (2, 3), (3, 2), (4, 1)\}$. Hence,
$$\mathbb{P}(A) = \frac{N(A)}{36} = \frac{4}{36} = \frac{1}{9}$$
2. Let B be the event that the dices come up the same number. Therefore, $B = \{(1, 1), (2, 2), \dots, (6, 6)\}$.
$$\mathbb{P}(B) = \frac{N(B)}{N(\Omega)} = \frac{6}{36} = \frac{1}{6}$$

2.3.2 Geometric Probability

Consider a sample space that is a subset of $\mathbb{R}^n$. Let A, B be two subsets,

$$A \times B = \{(a, b) : a \in A, b \in B\}$$

$$\mathbb{R}^n = \mathbb{R} \times \mathbb{R} \times \cdots \times \mathbb{R} = \{(x_i, \dots, x_n) \mid x_i \in \mathbb{R}, 1 \le i \le n\}$$

Because of the geometric nature of sample space, the equal likelihood model in this context is called geometric probability.
Let Ω be a subset of $\mathbb{R}^2 = \mathbb{R}\times\mathbb{R}$, and E be a subset of Ω. This could be a random experiment of collecting a point at random from Ω.
Sample space is Ω. E is the event that the point selected is an element of subset E. Since the point is selected at random,

$$\mathbb{P}(E) \propto area(E) \implies \exists \text{ a constant } k \text{ such}$$

$$\mathbb{P}(E) = k \cdot area(E), \text{for all events } E.$$

Now, we determine k: let $E = \Omega$, by certainty axiom, $1 = \mathbb{P}(\Omega) = k \cdot area(\Omega) \implies k = \frac{1}{area(\Omega)}$. Hence, $\mathbb{P}(E) = \frac{area(E)}{area(\Omega)}$.

Note: In higher dimensions, the area can be replaced by volume.

Example 2.10. Suppose that a number is selected at random from $(0, 1)$. Determine the probability that the number obtained is

1. 0.25 or greater
2. between 0.1 and 0.4, inclusive
3. either less than 0.1 or greater than 0.4

Solution: $\Omega = (0, 1)$. Recall for event E, $\mathbb{P}(E) = \frac{length(E)}{length(\Omega)}, \mathbb{P}(E) = length(E) \therefore length(\Omega) = 1$.

1. Event E is the subset $[0.25, 1)$. $\mathbb{P}(E) = length([0.25, 1)) = 0.75$
2. Let E denote the event that the number selected is between 0.1 and 0.4. $E = [0.1, 0.4] \subset= (0, 1), \mathbb{P}(E) = length(E) = 0.3$
3. Here, $E = (0, 0.1) \cup [0.4, 1) \subset (0, 1), \mathbb{P}(E) = \mathbb{P}((0, 0.1)) + ((0.4, 1)) = 0.7$

2.4 Law of Partitions

Definition 2.11. Events $A_1, A_2, \ldots$ form a partition of Ω if

1. mutually exclusive ($A_n \cap A_m = \phi$ for $m \neq n$)
2. exhaustive: $\bigcap_{n \geq 1} A_n = \Omega$

Any event E and its complement form a partition.

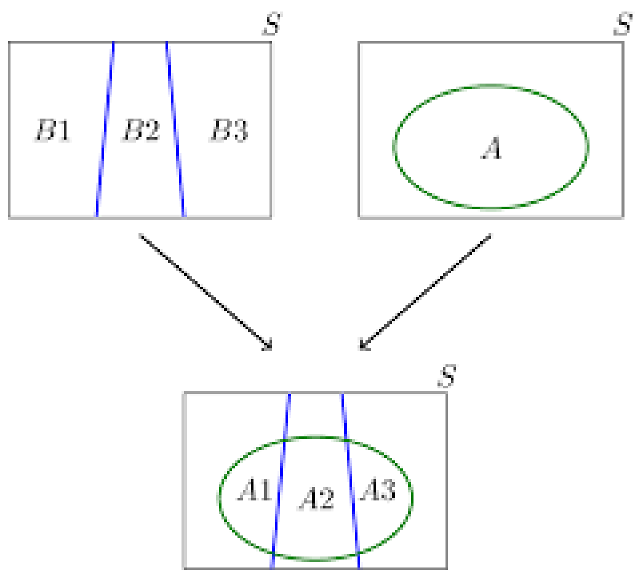


Figure 1: $A = (B_1 \cap A) \cup (B_2 \cap A) \cup (B_3 \cap A)$

Result: From Figure 1, suppose that $B_1, \ldots$ form a partition of S. Then,

1. $\mathbb{P}(A) = \sum_{n=1}^{\infty} \mathbb{P}(B_n \cap A)$, for any event A

2. If E is an event, then $\mathbb{P}(A) = \mathbb{P}(E \cap A) + \mathbb{P}(E^c \cap A) \forall A$

Example 2.12. According to a set of data, 8.1% of institutions of higher education are public schools in the Northeast, 11% are in the Midwest, 16.3% are in the South, and 9.6% are in the West. If a US institution of higher education is selected at random, determine the probability that it is public.
Solution: Let A_1 be the event that the school selected is in the Northeast, A_2 in the Midwest, A_3 in the South, and A_4 in the West. Let B be the event that the school selected is public.
$\mathbb{P}(A_1 \cap B) = 0.081, \mathbb{P}(A_2 \cap B) = 0.11, \mathbb{P}(A_3 \cap B) = 0.163, \mathbb{P}(A_4 \cap B) = 0.096.$
By the law of partitions, $\mathbb{P}(B) = \mathbb{P}(A_1 \cap B) + \mathbb{P}(A_2 \cap B) + \mathbb{P}(A_3 \cap B) + \mathbb{P}(A_4 \cap B) = 0.45.$
45% of institutions of higher education are public.

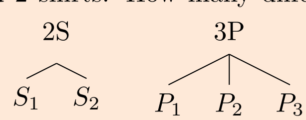
3 Combinatorial Probability

3.1 Basic counting rules

- r actions (choices/experiments) in definitive order
- m_1 possibilities for the first action, m_2 for the second, $\ldots$, m_r for the r^{th} action

Then, there are $m_1, m_2, \ldots, m_r$ possibilities altogether for the r^{th} action.

Example 3.1. Suppose a person A has 3 pants and 2 shirts. How many different pairs of a pant and a shirt can A dress up with?



$S_1P_1, S_1P_2, S_1P_3, S_2P_1, S_2P_2, S_2P_3$ are all of the possible pairs. There are $3 \times 2 = 6$ pairs.

Example 3.2. License plates in Arizona consist of three digits followed by three letters. How many different license plates are possible?
Solution: $10 \times 10 \times 10 \times 26 \times 26 \times 26 = 17,576,000$ combinations.

3.2 Other counting rules

3.2.1 Permutations

A permutation of r objects form a collection of m objects in any *ordered* arrangement of r distinct objects from the m objects.

$$mPr = \frac{m!}{(m-r)!}$$

Example 3.3. How many permutations are there for the word "ROSE"?
Solution: There are $4! = 24$ permutations.
What about for "TOO"? Well, it depends if 'O' and 'O' are distinct. If they're distinct, it would be $3! = 6$ permutations. But, if they aren't distinct, then it would be $\frac{3!}{2!} = 3$ permutations.
For example, find the number of distinct permutations of the letters "BERKELEY", noting that some letters are identical.
Solution: $\frac{8!}{3!} = 8 \times 7 \times 6 \times 5 \times 4$

3.2.2 Combinations

Combinations of r objects from a collection of m objects in any unordered arrangement.

$$mCr = \binom{m}{r} = \frac{m!}{r!(m-r)!}$$

Theorem 3.4. Binomial expansion: For any integer $n \geq 0$,

$$(x+y)^n = \sum_{i=0}^{n} \binom{n}{i} x^{n-i} y^i$$

Example 3.5. A committee of 3 persons is to be constituted from a group of 2 men and 3 women. In how many ways can this be done? How many of these committees would consists of 1 man and 2 women?
Solution: Since the order doesn't matter, combination of 5 different persons taken from 3 at a time,

$$\binom{5}{3} = \frac{5!}{3!2!} = 10.$$

For 1 man and 2 women,

$$\binom{2}{1} \times \binom{3}{2} = 6.$$

Example 3.6. From an ordinary deck of 52 cards, 7 are drawn at random, (1) without replacement, and (2) with replacement. What is the probability that at least one of the cards is a king?
Solution:

1. Let A be the event of interest. Then $\mathbb{P}(A) = 1 - \mathbb{P}(A^c)$. A^C is the event that none of the 7 selected cards is a king.

$$\mathbb{P}(A^C) = \frac{\binom{48}{7}}{\binom{52}{7}}$$

   Hence, $\mathbb{P}(A) = 1 - \frac{\binom{48}{7}}{\binom{52}{7}}$.

2. $\mathbb{P}(A^c) = \frac{48^7}{52^7}$. Hence, $\mathbb{P}(A) = 1 - \frac{48^7}{52^7}$.

Example 3.7. You have 100 towels: 40 are yellow, 60 are white.

1. What is the probability of choosing 5 towels where 3 are white and 2 are yellow?
2. What is the probability that we will have at least 3 white towels?
3. What is the probability that we will have at most 3 white towels?

Solution: Sample space will contain all of the possibilities of selecting the 5 towels. $|\Omega| =$ the number of elements in $\Omega = \binom{100}{5}$.

1. Let A be the event of choosing 3 white towels and 2 yellow towels.

$$(A) = \binom{60}{3} \times \binom{40}{2}$$

   Now, since the sample space is finite, classical probability model applies here:

$$\mathbb{P}(A) = \frac{\binom{60}{3}\binom{40}{2}}{\binom{100}{5}}$$

2. $\mathbb{P}(B) = \mathbb{P}(3W+2Y) + \mathbb{P}(4W+1Y) + \mathbb{P}(5W+0Y) = \frac{\binom{60}{3}\binom{40}{2}}{\binom{100}{5}} + \frac{\binom{60}{4}\binom{40}{1}}{\binom{100}{5}} + \frac{\binom{60}{5}\binom{40}{0}}{\binom{100}{5}}$
3. $\mathbb{P}(C) = \mathbb{P}(0W + 5Y) + \mathbb{P}(1W + 4Y) + \mathbb{P}(2W + 3Y) + \mathbb{P}(3W + 2Y)$

4 Conditional Probability and Independence

4.1 Conditional Probability

The probability of E given F, or $\mathbb{P}(E \mid F) = \frac{\text{number of elements in } E \cap F}{\text{number of elements in } F}$.

Example 4.1. Consider the experiment of tossing a balanced coin three times:

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

$$|\Omega| = 2^3 = 8$$

Let E be the event that at least two heads appear. Let F be the event that the first coin shows tails. Then,

$$E = \{HHH, HHT, HTH, THH\}$$

$$F = \{THH, THT, TTH, TTH\}$$

$$\mathbb{P}(E) = \frac{1}{2}, \mathbb{P}(F) = \frac{1}{2}, \mathbb{P}(E \cap F) = \frac{1}{8}$$

Question: Suppose the first coin shows tails(F occurs). What is the probability of the occurrence of event E?
The probability of E considering F as the sample space is $\frac{1}{4}$. AKA, $\mathbb{P}(E \mid F) = \frac{1}{4}$.

Definition 4.2. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be the probability space. Let $E, F \in \mathcal{F}$ be events such that $\mathbb{P}(F) > 0$. The conditional probability of E given F, denoted by $\mathbb{P}(E \mid F)$ is defined by

$$\mathbb{P}(E \mid F) = \frac{\mathbb{P}(E \cap F)}{\mathbb{P}(F)}$$

Example 4.3. Roll two fair dices. What is the probability that the numbers on the dice add up to 8, given that the dice show different numbers?
Solution: Let A be the event that numbers on dice add up to 8. Let B be the event that the numbers on dice show different numbers. So, $\Omega = \{(i, j) : 1 \leq i, h \leq 6\}, |\Omega| = 6^2 = 36$

$$A = \{(2,6), (6,2), (3,5), (5,3), (4,4)\}, |A| = 5$$

$$B = \{(i, j) : 1 \leq i \leq 6, 1 \leq j \leq 6, i \neq j\}, |B| = 30$$

$$A \cap B = \{(2,6), (6,2), (3,5), (5,3)\}, \mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \implies \frac{\frac{4}{36}}{\frac{30}{36}} = \frac{4}{30}$$

4.2 General Multiplication Rule

Proposition: Let A and B be events with $\mathbb{P}(A) > 0$. Then,

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \times \mathbb{P}(B \mid A).$$

More generally, if $A_1, A_2, \ldots, A_N$ are events with $\mathbb{P}(A_1 \cap \cdots \cap A_{N-1}) > 0$.

$$\mathbb{P}(\bigcap_{n=1}^{N} A_n) = \mathbb{P}(A_1) \times \mathbb{P}(A_2 \mid A_1) \times \cdots \times \mathbb{P}(A_N \mid A_1 \cap \cdots \cap A_{N-1})$$

Example 4.4. For the 107th Congress, 18.7% of members are senators and 50% of the senators are democrats. What is the probability that a randomly selected member of the 107th Congress is a democratic senator?
Solution: Let D be the event that the member selected is a democrat. Let S be the event that the member selected is a senator.

$$\mathbb{P}(S) = 0.187, \mathbb{P}(D \mid S) = 0.5$$

By the multiplication rule,

$$\mathbb{P}(D \cap S) = \mathbb{P}(S) \times \mathbb{P}(D \mid S) = 0.187 \times 0.5 = 0.094$$

Example 4.5. An urn contains 5 red and 4 blue balls. A total of 4 balls is drawn sequentially (without replacement). What is the probability that colors of balls alternate?
Solution: Let R_1, B_i be events that the i^{th} ball is red, blue respectively. That means we want

$$\mathbb{P}(\{R_1 \cap B_2 \cap R_3 \cap B_4\} \cup \{B_1 \cap R_2 \cap B_3 \cap R_4\}) = \mathbb{P}(R_1 \cap B_2 \cap R_3 \cap B_4) + \mathbb{P}(B_1 \cap R_2 \cap B_3 \cap R_4)$$

By the general multiplication rule,

$$\mathbb{P}(R_1 \cap B_2 \cap R_3 \cap B_4) = \mathbb{P}(R_1) \times \mathbb{P}(B_2 \mid R_1) \times \mathbb{P}(R_3 \mid R_1 \cap B_2) \times \mathbb{P}(B_4 \mid R_1 \cap B_2 \cap R_3)$$

$$= \frac{5}{9} \times \frac{4}{8} \times \frac{4}{7} \times \frac{3}{6}$$

Similarly,

$$\mathbb{P}(B_1 \cap R_2 \cap B_3 \cap R_4) = \frac{4}{9} \times \frac{5}{8} \times \frac{3}{7} \times \frac{4}{6}$$

Add these two together for the final probability.

4.3 The Law of Total Probability

Proposition: Suppose that $A_1, A_2, \ldots$ form a partition $(A_i \cap A_j = \phi, i \neq j \& \bigcap_{n\geq 1} A_n = \Omega)$ of sample space Ω. Then, for each event B,

$$\mathbb{P}(B) = \sum_{n=1}^{\infty} \mathbb{P}(A_n)\mathbb{P}(B \mid A_n).$$

In particular, if E is an event, then

$$\mathbb{P}(B) = \mathbb{P}(E)\mathbb{P}(B \mid E) + \mathbb{P}(E^c)\mathbb{P}(B \mid E^c)$$

for each event B.

Example 4.6. 54% of US men and 33% of US women believe in aliens. Also, 48% of US adults are men. What percentage of US adults believe in aliens?
Solution: Assume that a US adult is selected at random and A: adult selected believes in aliens, $M : \{$ adult selected is a man $\}$.
$\mathbb{P}(M) = 0.48; \mathbb{P}(N^c) = 0.52; \mathbb{P}(A \mid M) = 0.54; \mathbb{P}(A \mid M^c) = 0.33$. Apply the law of total probability:

$$\mathbb{P}(A) = \mathbb{P}(M) \times \mathbb{P}(A \mid M) + \mathbb{P}(M^c) \times \mathbb{P}(A \mid M^c) = 0.4308$$

43.08% of US adults believe in aliens.

4.4 Independent Events

In general, $\mathbb{P}(A \mid B) \neq \mathbb{P}(A)$. But, if two events A and B are independent, then the occurrence of A does not affect B, or

$$\mathbb{P}(A \mid B) = \mathbb{P}(A).$$

Example 4.7. Consider the experiment of randomly selecting one card from a deck of 52. Let K : event a King is selected. H : event a heart is selected.

$$\mathbb{P}(K \mid H) = \frac{1}{13} = 0.077$$

$$\mathbb{P}(K) = \frac{4}{52} = 0.077$$

Therefore, they are independent.

Definition 4.8. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a sample space; And $A, B \in \mathcal{F}$ with $\mathbb{P}(A) > 0$. We say that event B is independent of event A if the occurrence of A doesn't affect the probability that event B occurs, i.e.

$$\mathbb{P}(B \mid A) = \mathbb{P}(B), \mathbb{P}(A) > 0.$$

Two events are also said to be independent events if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \times \mathbb{P}(B).$$

Mutually exclusive(disjoint) and independence are two different things. Mutually exclusive means two sets do not overlap. Independence means one event happening does not affect the outcome of the other.

4.5 Bayes' Theorem

Example 4.9. One bag contains 3 red and 4 black balls while the other bag contains 5 red and 6 black balls. One ball is drawn at random from one of the bags. What is the probability that the color of the ball is red?
Solution: Let E_1 be the event of choosing from bag 1, and E_2 be the event of choosing from bag 2, and A be the event of drawing a red ball. Then, $\mathbb{P}(E_1) = \mathbb{P}(E_2) = \frac{1}{2}$. Also, $\mathbb{P}(A \mid E_1) = \frac{3}{7}$, and $\mathbb{P}(A \mid E_2) = \frac{5}{11}$. Hence, by the law of total probability,

$$\mathbb{P}(A) = \sum \mathbb{P}(E_n)\mathbb{P}(A \mid E_n)$$

The probability of drawing a ball from bag 2, being given that the ball is red,

$$\mathbb{P}(E_2 \mid A) = \frac{\mathbb{P}(E_2 \cap A)}{\mathbb{P}(A)} = \frac{\mathbb{P}(E_2)\mathbb{P}(A \mid E_2)}{\mathbb{P}(E_1)\mathbb{P}(A \mid E_1) + \mathbb{P}(E_2)\mathbb{P}(A \mid E_2)}$$

Theorem 4.10. Suppose that $A_1, A_2, \ldots$ form a partition of the sample space. Then, for each event B,

$$\mathbb{P}(A_j \mid B) = \frac{\mathbb{P}(A_j)\mathbb{P}(B \mid A_j)}{\sum_{n \geq 1} \mathbb{P}(A_n)\mathbb{P}(B \mid A_n)}$$

Example 4.11. Machine A, B, and C respectively manufacture 25%, 35%, and 40% of the bolts. Of their outputs, 5, 4, and 2 percent are respectively defective bolts. A bolt is drawn at random from the product and is found to be defective. What is the probability that it is manufactured by machine B?
Solution: Let E be the event that the bolt is defective. And B_1, B_2, B_3 be the events that the bolt is manufactured by machines A, B, and C respectively. $\mathbb{P}(B_1) = 0.25, \mathbb{P}(B_2) = 0.35, \mathbb{P}(B_3) = 0.4$. If we want $\mathbb{P}(B_2 \mid E)$, then $\mathbb{P}(E \mid B_1) = 0.05, \mathbb{P}(E \mid B_2) = 0.04, \mathbb{P}(E \mid B_3) = 0.02$. Hence, by Bayes's theorem,

$$\mathbb{P}(B_2 \mid E) = \frac{\mathbb{P}(B_2)\mathbb{P}(E \mid B_2)}{\sum_{i=1}^{3} \mathbb{P}(B_i)\mathbb{P}(E \mid B_i)}$$

5 Random Variables and their Distribution

Definition 5.1. Random Variable: Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A function $x : \Omega \to R$ is called a random variable.

Example 5.2. Toss a fair coin three times.

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

X : number of H's. So, $X(HHH) = 3, X(HHT) = 2, X(TTT) = 0, X : \Omega \to \{0, 1, 2, 3\}$, so X is a discrete random variable.
Notation: for $x \in \mathbb{R}, \{X = x\} = \{\omega \in \Omega \mid X(\omega) = x\}$

$$\{X = 2\} = \{\omega : x(\omega) = 2\} = \{HHT, HTH, THH\}$$

$$\{X = 1\} = \{\omega \in \Omega \mid X(\omega) = 1\} = \{HTT, THT, TTH\}$$

$$\{X = 0\} = \{TTT\}; \{X = 3\} = \{HHH\}$$

$$\mathbb{P}(X = 3) = \mathbb{P}(\omega : X(\omega) = 3) = \frac{1}{8}$$

$$\mathbb{P}(X = 1) = \mathbb{P}(X = 2) = \frac{3}{8}$$

More formally:

$$\mathbb{P}(x = x) = \begin{cases} \frac{1}{8}; x = 0 \\ \frac{1}{8}; x = 3 \\ \frac{3}{8}; x = 1 \\ \frac{3}{8}; x = 2 \\ 0 \text{ if } x \notin \{0, 1, 2, 3\} \end{cases}$$

Note: $\sum_{x \in \mathbb{R}} \mathbb{P}(X = x) = 1$

5.1 Discrete Random Variables

A real-valued function on the sample space of a random experiment whose possible values (range) can be assumed to be a countable set.

Definition 5.3. Probability Mass Function: Let X be a discrete random variables (r.v.). Then the probability mass function (PMF) of X, denoted by p_X, is a function

$$p_X : \mathbb{R} \to [0,1] \text{ s.t. } p_X(x) = \mathbb{P}(X = x).$$

Proposition(Basic properties of a PMF):

1. $p_X(x) \geq 0 \forall x \in \mathbb{R}$
2. $\sum_{x \in \mathbb{R}} p_X(x) = 1$
3. $\{x \in \mathbb{R} \mid p_X(x) \neq 0\}$ is countable

Example 5.4. Consider the random experiment of tossing a coin three times. Let X denote the total number of heads obtained in three tosses of the coin. Determine $\mathbb{P}(1 \leq X \leq 3)$.
Solution: Possible values of X are $0, 1, 2, 3$. Hence, if $x \notin \{0, 1, 2, 3\}, (X = x) = (\omega \mid X(\omega) = x) = p$.

$$\{1 \leq x < 3\} = \{\omega \in \Omega \mid 1 \leq X(\omega) < 3\} = \{\omega \mid X(\omega) = 1\} \cup \{\omega \mid X(\omega) = 2\} =$$

$$\{X = 1\} \cup \{X = 2\}$$

Hence, $\{1 \leq X < 3\}$ is the union of two mutually exclusive events, i.e. $\{X = 1\}$ and $\{X = 2\}$.

$$\mathbb{P}(1 \leq X < 3) = \mathbb{P}(X = 1) + \mathbb{P}(X = 2) = p_X(1) + p_X(2) = \frac{3}{8} + \frac{3}{8} = \frac{3}{4}$$

In general, for a discrete random variable, and $A \subset \mathbb{R}$, $\mathbb{P}(X \in A) = \sum_{x \in A} p_X(x)$.

5.1.1 Bernoulli random variables

1. An experiment with two outcomes ($\Omega = \{S, F\}$)
2. Independent trials – i.e. Bernoulli trials
3. Fixed probabilities of success – i.e. $p \in [0,1]$

Let x be the number of success in a single Bernoulli trial. Thus, x is a discrete random variable taking values of $\{0,1\}$ – i.e. $x(F) = 0$ and $x(S) = 1$.
For PMF,

$$p_x(1) = \mathbb{P}(x = 1) = \mathbb{P}(S) = p$$

$$p_x(0) = \mathbb{P}(x = 0) = \mathbb{P}(F) = 1 - p$$

Hence,

$$p_x = \begin{cases} p; \text{if } x = 1 \\ 1 - p; \text{if } x = 0 \\ 0; \text{otherwise} \end{cases}$$

Here, the random variable whose RMF is given by the above piecewise is called a Bernoulli random variable, and it is denote by $X \sim Ber(p)$.

5.1.2 Binomial random variables

To define a binomial random variable, we start with three Bernoulli trials, with probability of success p.
Sample space: $\Omega = \{SSS, SSF, \ldots, FFF\} = 8$
Let x be the number of success in 3 trials. So, the range of $x = \{0, 1, 2, 3\}$.
The probability of the number of successes is as follows:

$$\mathbb{P}(x = 0) = \mathbb{P}(\{FFF\}) = \mathbb{P}(F) \times \mathbb{P}(F) \times \mathbb{P}(F) = (1 - p)^3$$

$$\mathbb{P}(x = 1) = \mathbb{P}(\{SFF, FSF, FFS\}) = \mathbb{P}(\{SFF\}) + \mathbb{P}(\{FSF\}) + \mathbb{P}(\{FFS\}) =$$

$$3(\mathbb{P}(S)\mathbb{P}(F)\mathbb{P}(F)) = 3p(1 - p)^2$$

Now, let's find the probability of x successes in n Bernoulli trials. Clearly, x successes can be obtained in

$$nCx = \frac{n!}{x!(n-x)!} \text{ ways.}$$

In each of these ways, the probability of x successes is

$$= \mathbb{P}(x) \times \mathbb{P}(n - x) = p^x \times (1 - p)^{n-x}.$$

Thus, the probability of x successes in n Bernoulli trials is

$$\binom{n}{x} p^x (1 - p)^{n-x}; x = 0, 1, \ldots, n$$

Example 5.5. If a fair coin is tossed 10 times, find the probability of

1. exactly six heads
2. at least six heads
3. at most six heads

Solution: Here, Bernoulli trials are repeated tosses. X: number of heads in an experiment of 10 trials. $\therefore X$ has the binomial with $n = 10, p = \frac{1}{2}$. $\therefore PMF$ of X is $p_X(x) = \binom{n}{x} p^x (1-p)^{n-x} \implies P(X = x) p_X(x) = \binom{10}{x} (\frac{1}{2})^{10}$

1. $\mathbb{P}(X = 6) = \mathbb{P}\{\omega : X(\omega) = 6\} = p_X(6) \implies \mathbb{P}(X = 6) = \binom{10}{6} (\frac{1}{2})^{10}$
2. $\mathbb{P}(X \geq 6) = \mathbb{P}\{\omega : X(\omega) \geq 6\} = \sum_{x=6}^{10} p_X(x) = \sum_{x=6}^{10} \binom{10}{x} (\frac{1}{2})^{10}$
3. $\mathbb{P}(X \leq 6) = \sum_{x=0}^{6} p_X(x) = \sum_{x=0}^{6} \binom{10}{x} (\frac{1}{2})^{10}$

5.1.3 Poisson random variables

A discrete random variable X taking values in $\{0, 1, \dots\}$ is called a Poisson random variable ($X \sim Pois(\lambda)$) if its PMF has the form

$$p_X(x) = \mathbb{P}(X = x) = \begin{cases} e^{-\lambda} \frac{\lambda^x}{x!}, x = 0, 1, \dots \\ 0, \text{ otherwise} \end{cases}$$

Here, the parameter λ of Poisson random variable represents its average value.

Example 5.6. Let the number of patients arriving in an emergency room from 6 to 7 pm have a Poisson distribution with parameter $\lambda = 6.9$. Determine the probability that, on a given day, the number of patients arriving at the emergency room between 6 to 7 will be

1. exactly four
2. at least two
3. between four and ten, inclusive

Solution: The PMF of X is $p_X(x) = e^{-6.9} \frac{6.9^x}{x!}, x = 0, 1, 2, \dots$.

1. $\mathbb{P}(X = 4) = e^{-6.9} \frac{6.9^4}{4!} = 0.095$
2. $\mathbb{P}(X \geq 2) = 1 - \mathbb{P}(X < 2) = 1 - \sum_{x<2} p_X(x) = 1 - p_X(0) - p_X(1) = 0.992$
3. $\mathbb{P}(4 \leq X \leq 10) = \sum_{4 \leq x \leq 10} p_X(x) = 0.821$

5.1.4 Geometric random variable

Consider repeated Bernoulli trials with success probability p. Let X be the number of trials up to and including the first success. Then PMF of the random variable X is

$$p_X(x) = \mathbb{P}\{X = x\} = p(1-p)^{x-1}; x = 1, 2, \ldots.$$

Definition 5.7. A random variable X taking values $\{1, 2, \ldots\}$ is called a geometric random variable if its PMF is given by

$$p_X(x) = \begin{cases} p(1-p)^{x-1}, x = 1, 2, \ldots \\ 0, \text{ otherwise} \end{cases}$$

So, geometric distribution describes the waiting time for the first success.
Proposition: Let $X \sim g(p)$. Then, $\mathbb{P}(X > n) = (1-p)^n$. So, the probability of the event that first first success takes at least $(n+1)$ trials in $(10p)^n$.

5.1.5 Lack-of-memory property

A random variable is said to satisfy the lack-of-memory property if

$$\mathbb{P}(X = n + k \mid X > n) = \mathbb{P}(X = k); n, k \in \mathbb{N}.$$

Proposition: Geometric random variable is the only discrete random variable that has the lack-of-memory property.

5.1.6 Negative Binomial random variable

Consider repeated Bernoulli trials with success probability p. Let X denote the number of trials up to and including the r^{th} success. Then,

$$\mathbb{P}(X = x) = \binom{x-1}{r-1} p^r (1-p)^{x-r}; x = r, r+1, \ldots$$

5.2 Functions of discrete random variables

Example 5.8. Let X be a discrete random variable. Then $Y = X^2$ is also a discrete random variable. The PMF of X is

x	-1	0	1
$p_X(x)$	0.1	0.7	0.2

Then, find the PMF of y. For this, possible values of y are $\{0, 1\}$, or $\mathbb{P}(y = 0)$ & $\mathbb{P}(y = 1)$.

$$\mathbb{P}(y = 0) = \mathbb{P}(x^2 = 0) = \mathbb{P}(x = 0) = 0.7$$

$$\mathbb{P}(y = 1) = \mathbb{P}(x^2 = 1) = \mathbb{P}(X \in \{-1, 1\}) = p_X(-1) + p_X(1) = 0.1 + 0.2 = 0.3$$

$$\mathbb{P}(y = k) = 0 \text{ if } k \neq 0, 1$$

The PMF of y is

y	0	1
$p_X(x)$	0.7	0.3

Proposition: Let x be a discrete random variable and let g be a real-valued function defined on the range of X. Then, the PMF of $y = g(X)$ is

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

where $g^{-1}\{y\} = \{x \mid g(x) = y\}$; preimage of y under map g.
Note: If g is one-to-one, then $p_Y(y) = p_X(x) = p_X(g^{-1}(y))$.

Example 5.9. Let $X \sim g(p)$ (geometric random variable), $p_X(x) = p(1-p)^{x-1}; x = 1, 2, \ldots$. Find the distribution of $y = x^2$.
Solution: The range of X is $\{1, 2, \ldots\}$, hence the range of Y is $\{1, 4, 9, \ldots\}$. It is clear that $g = x^2$ is one-to-one on the range of X. Then, $y = g(x) \implies x = g^{-1}(y) = \sqrt{y}$. Hence, for $y \in \{1, 4, 9, \ldots\}$, we have $p_X(g^{-1}(y)) = p_X(\sqrt{y}) = p(1-p)^{\sqrt{y}-1}$.

6 Jointly discrete random variables

Let X and Y be two discrete random variables defined on the same probability space Ω. Then, the joint probability mass function (joint PMF) of X and Y (denoted $p_{X,Y}$) is a function

$$p_{X,Y} : \mathbb{R}^2 \to [0, 1]$$

such that $p_{X,Y}(x, y) = \mathbb{P}(X = x, Y = y)$.

$$\{X = x, Y = y\} = \{X = x\} \cap \{Y = y\}, \{\omega \in \Omega \mid X(\omega) = x, Y(\omega) = y\}.$$

Note: If any one of the event $\{X = x\}$ and $\{Y = y\}$ doesn't occur; then $\{X = x, Y = y\} = \phi$. As with the single variable case, we should have that,

1. $p_{X,Y}(x, y) \geq 0$
2. $\sum_{X,Y} p_{X,Y}(x, y) = 1$
3. $\{(x, y) \in \mathbb{R}^2 \mid p_{X,Y}(x, y) \neq 0\}$ is a countable set

Example 6.1. Consider 50 homes currently for sale, where the number of bedrooms and number of bathrooms are given in the table: $X \sim$ number of bedrooms, $Y \sim$ number of bathrooms:

$X\|Y$	1	2	3
2	10	7	0
3	5	10	2
4	3	8	5

Suppose that one of these 50 homes is selected at random. Find $p_{X,Y}(2, 1)$ and $\mathbb{P}(X = 2)$.

Solution: $p_{X,Y}(2, 1) = \mathbb{P}(X = 2, Y = 1\} = \frac{10}{50}$.

$\mathbb{P}\{X = 2\} = \mathbb{P}(X = 2, Y = 1\} + \mathbb{P}\{X = 2, Y = 2\} + \mathbb{P}\{X = 2, Y = 3\} = \frac{10}{50} + \frac{7}{50} + \frac{0}{50} = \frac{17}{50}$.

Proposition: Let X, Y be discrete random variables. Then,

$$p_X(x) = \sum_y p_{X,Y}(x, y), x \in \mathbb{R}$$

and

$$p_Y(y) = \sum_x p_{X,Y}(x, y), y \in \mathbb{R}.$$

Example 6.2. Let the joint PMF of X and Y be given by

$$p_{X,Y}(x,y) = \begin{cases} p^2(1-p)^{x+y-2}; x, y \in \mathbb{N} \\ = 0; o/w \end{cases}$$

Find and identify the marginal PMFs of X and Y.
Solution:

$$p_X(x) = \sum_y p_{X,Y}(x,y) = \sum_y p^2(1-p)^{x+y-2} = p^2 \sum_{y=1}^{\infty} (1-p)^{x+y-2} =$$

$$p^2(1-p)^{x-2} \sum_{y=1}^{\infty} (1-p)^y = p^2(1-p)^{x-2} \frac{1-p}{1-(1-p)} = p(1-p)^{x-1} \implies$$

$$p_X(x) = \begin{cases} p(1-p)^{x-1}; x, y \in \mathbb{N} \\ = 0; o/w \end{cases}$$

$X \sim g(p)$; geometric random variable with parameter p. Similarly, $Y \sim g(p)$.

6.1 Fundamental Probability Formula

Remember $p_{X,Y}(x,y) = \mathbb{P}\{X = x, Y = y\} = \mathbb{P}\{X \in A, Y \in B\} \implies A = \{x\}, B = \{y\}$. For any subset $A \subset \mathbb{R}^2$, we have

$$\mathbb{P}((X,Y) \in A) = \sum_{(x,y) \in A} \sum p_{X,Y}(x,y).$$

Example 6.3. Suppose that one of the 50 homes is selected at random. X is the number of bedrooms, Y is the number of bathrooms.

$X \mid Y$	1	2	3
2	10	7	0
3	5	10	2
4	3	8	5

Determine the probability that the home obtained

1. has the same number of bedrooms and bathrooms
2. has more bedrooms than bathrooms.

Solution:

1. $\mathbb{P}(X = Y) = \sum_{x=y} \sum p_{X,Y}(x,y), A = \{(x,y) \in \mathbb{R}^2 \mid x = y\}$
   $= p_{X,Y}(3,3) + p_{X,Y}(2,2) = \frac{2}{50} + \frac{7}{50} = \frac{9}{50}$.
2. $\mathbb{P}(X > Y) = \sum_{x>y} \sum p_{X,Y}(x,y), A = \{(x,y) \in \mathbb{R}^2 \mid x > y\}$
   $= p_{x,y}(2,1) + p_{x,y}(3,1) + p_{x,y}(3,2) + \dots$

6.3 Conditional Probability Mass Function

Recall $\mathbb{P}(F \mid E) = \frac{\mathbb{P}(F \cap E)}{\mathbb{P}(E)}$. Assume X, Y are two discrete random variables on the same probability space Ω. Then, for $E = \{X = x\}, F = \{Y = y\}$,

$$\mathbb{P}(Y = y \mid X = x) = \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(X = x)} = \frac{p_{X,Y}(x,y)}{p_X(x)},$$

hence, if $p_X(x) > 0$, the conditional probability mass function of Y given $X = x$, denoted by $p_{y|x}$,

$$p_{y|x} = \frac{p_{X,Y}(x,y)}{p_X(x)}, y \in \mathbb{R}.$$

In fact, $p_{y|x}$ is a probability mass function.

Example 6.4. Let X, Y denote the lifetimes of the two components observed at discrete timed units. The joint PMF of X and Y is

$$p_{X,Y}(x,y) = p^2(1-p)^{x+y-2}; x, y \in \mathbb{N}.$$

Find the conditional PMF of Y given $X = x$:

$$p_{Y|X}(y \mid x) = \frac{p_{X,Y}(x,y)}{p_X(x)} = \frac{p^2(1-p)^{x+y-2}}{p(1-p)^{x-1}} = p(1-p)^{y-1}.$$

So, $p_{Y|X} \sim g(p)$ – geometric.

Note: $\mathbb{P}(Y \in A \mid X = x) = \sum_{y \in A} p_{Y|X}(y \mid x)$, for any subset $A \subset \mathbb{R}$.

Example 6.5. Suppose that one of the 50 homes from Example 6.3 is selected at random. Given that a randomly selected home has 3 bedrooms, what the the probability that it has almost three bathrooms? (See tables from previous examples).
Solution: We want $\mathbb{P}(Y \leq 3 \mid X = 3) = \sum_{y \leq 3} p_{Y|X}(y \mid 3) = p_{Y|X}(1 \mid 3) + p_{Y|X}(2 \mid 3) + p_{Y|X}(3 \mid 3) = \frac{5}{50} + \frac{10}{50} + \frac{2}{50} = \frac{17}{50}$.

6.4 Independent random variables

Recall: two events E, F are independent if

$$\mathbb{P}(E \mid F) = \mathbb{P}(E); \mathbb{P}(F) > 0.$$

$$\mathbb{P}(E \cap F) = \mathbb{P}(E) \cdot \mathbb{P}(F).$$

In the context of random variables, events are considered (for example) $\{X = x\}$.

Definition 6.6. Two random variables X, Y defined on the same probability space Ω are said to be **independent** if

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A) \cdot \mathbb{P}(Y \in B), \text{ for any subsets } A, B.$$

Proposition: If X, Y are independent random variables, then $g(X), h(Y)$ are also independent; here, g, h are some real-valued functions.
Proof: $\mathbb{P}(g(X) \in A, h(Y) \in B) = \mathbb{P}(X \in g^{-1}(A), Y \in h^{-1}(B)) = \mathbb{P}(X \in g^{-1}(A)) \cdot \mathbb{P}(Y \in h^{-1}(B)) = \mathbb{P}(g(X) \in A) \cdot \mathbb{P}(h(Y) \in B)$. Hence $g(X), h(Y)$ are independent.

Example 6.7. If X, Y are independent, and

$$z := x^4 + e^X + \sin X;$$

$$s := log|Y| + e^Y \cos Y + Y^{175} + 3;$$

$\implies$ z and s are also independent.

Proposition: If X, Y are two discrete random variables on the same probability space, then X, Y are independent if and only if

$$p_{X,Y}(x, y) = p_X(x) \cdot p_Y(y), \text{ for all } x, y \in \mathbb{R}.$$

Proof: Suppose X, Y are independent. We consider $A = \{x\}, B = \{y\}; x, y \in \mathbb{R}$. So, $p_{X,Y}(x, y) = \mathbb{P}(X = x, Y = y) = \mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A) \cdot \mathbb{P}(Y \in B) = \mathbb{P}(X = x) \cdot \mathbb{P}(Y = y) = p_X(x) \cdot_Y (y) = p_X(x) \cdot_Y (y)$.

Example 6.8. Let X, Y be two discrete random variables whose joint PMF is given by

$$p_{X,Y}(x, y) = p^2(1-p)^{x+y-2}; x, y \in \mathbb{N}, 0 < p < 1.$$

Determine whether X, Y are independent random variables.
Solution: Check if

$$p_X(x) = \begin{cases} p(1-p)^{x-1}, \text{ if } x \in \mathbb{N} \\ 0, o/w \end{cases}, (p_X(x) = \sum_y p_{X,Y}(x, y))$$

Then do the same for Y. This result in a $p_{X,Y}(x, y) = p_X(x) p_Y(y) \implies X, Y$ are independent random variables.

In short, if X, Y are independent random variables, then

$$p_{Y|X}(y \mid x) = \frac{p_{X,Y}(x, y)}{p_X(x)} = \frac{p_X(x) p_Y(y)}{p_X(x)} = p_Y(y).$$

6.5 Functions of Two or More Discrete Random Variables

We want to study the distribution of $z = g(X,Y)$; for given random variables X, Y. For example: $g(x,y) = x+y$, or $g(x,y) = xy$, or $g(x,y) = min(x,y)$.
Proposition: Let $z = g(X,Y)$. Then, the PMF of the random variable z is

$$p_Z(z) = \sum\sum_{(x,y)\in g^{-1}(\{z\})} p_{X,Y}(x,y); g^{-1}(\{z\}) = \{(x,y) \mid g(x,y) = z\}.$$

Proof:

$$p_Z(z) = \mathbb{P}(Z = z) = \mathbb{P}(g(X,Y) = z) = \mathbb{P}((X,Y) \in g^{-1}(\{z\})) = \sum\sum_{(x,y)\in g^{-1}(\{z\})} p_{X,Y}(x,y)$$

Example 6.9. Let $X, Y \sim g(p)$: X, Y are independent. Determine and identify the PMF of the random variable $z = X + Y$.
Solution: Since X, Y are independent random variables, the joint PMF of X, Y is

$$p_{X,Y}(x,y) = \begin{cases} p^2(1-p)^{x+y-2}; x, y \in \mathbb{N} \\ 0; o/w \end{cases}.$$

Range of $z = \{2, 3, \dots\}$. From the previous proposition, $p_Z(z) = p_{X+Y}(z) = \sum\sum_{x+y=z} p_{X,Y}(x,y)$.

$$= \sum_{x=1}^{z-1} \sum_{y=z-x} p_{X,Y}(x,y) = \sum_{x=1}^{z-1} \sum_{y=z-x} p^2(1-p)^{x+y-2} = \sum_{x=1}^{z-1} p^2(1-p)^{z-2}$$

$$= p^2(1-p)^{z-2} \sum_{x=1}^{z-1} 1 = (z-1)p^2(1-p)^{z-2}; z = 2, 3, \dots$$

$Z \sim NB(r,p), r = 2 \implies X + Y \sim NB(2,p)$, if X, Y are independent geometric random variables.

Example 6.10. X, Y are two independent geometric random variables with parameter p. $X, Y \sim g(p)$. $z = min\{X, Y\}$. Find the PMF of Z.
Solution: Range of Z is the set of all positive integers, $\mathbb{N}$. By the proposition,

$$p_Z(z) = \sum_{min\{X,Y\}=Z}\sum p_{X,Y}(x,y) = p_{X,Y}(z,z) + \sum_{y=z+1}^{\infty} p_{X,Y}(z,y) + \sum_{z=z+1}^{\infty} p_{X,Y}(x,z)$$

$$= p^2(1-p)^{2z-2} + \sum_{y=z+1}^{\infty} p^2(1-p)^{z+y-2} + \sum_{x=z+1}^{\infty} p^2(1-p)^{x+z-2}$$

$$= p^2(1-p)^{2z-2} + 2p^2(1-p)^{z-2}\sum_{\omega=z+1}^{\infty}(1-p)^{\omega} = p^2(1-p)^{2z-2} + 2p^2(1-p)^{z-2}\frac{(1-p)^{z+1}}{1-(1-p)}$$

$$= (2p-p^2)(1-(2p-p^2))^{z-1}, \text{ if } z \in \mathbb{N} = q(1-q)^x \implies$$

$$z = min\{X, Y\} \sim \text{geometric}(q), \text{ where } q = 2p - p^2; X \sim g(p), p(1-p)^{x-1}.$$

Second method: To find the distribution of $z = min\{X, Y\}$, we note $Z > z$ if and only if $X > Z$ and $Y > Z$.

$$\implies \mathbb{P}(Z > z) = \mathbb{P}(X > z)\mathbb{P}(Y > z) = (1-p)^z(1-p)^z = (1-p)^{2z} = \lfloor (1-p)^2 \rfloor^z$$

$$= (1-(2p-p^2)^z) \implies Z \sim g(2p-p^2).$$

Let X, Y be two discrete random variables with join PMF $p_{X,Y}(x,y)$. Set $Z = X + Y$. Then,

$$p_Z(z) = \sum_{x+y=z}\sum p_{X,Y}(x,y) = \sum_x \sum_{y=z-x} p_{X,Y}(x,y) = \sum_x p_{X,Y}(x, z-x), z \in \mathbb{R}$$

$$\implies p_Z(z) = \sum_x p_{X,Y}(x, z-x).$$

In particular, if X, Y are two independent random variables, then using $p_{X,Y}(x,y) = p_X(x)p_Y(y), p_Z(z) = \sum_x p_X(x)p_Y(z-x)$.

Example 6.11. 1. If X, Y are two independent variables, $X \sim P(\lambda)$, $Y \sim P(\mu)$. Let $z = X + Y$. Find the PMF of z. Claim: $z \sim P(\lambda + \mu)$.
We start: $p_Z(z) = \mathbb{P}(Z = z) = \sum_x p_X(x)p_Y(z-x) = \sum_{x=0} e^{-\lambda}\frac{\lambda^x}{x!}e^{-\mu}\frac{\mu^{z-x}}{(z-x)!}; z - x = y \geq 0, z \geq x \geq 0 = e^{-\lambda-\mu}\sum_{x=0}^{z}\frac{\lambda^x}{x!}\frac{\mu^{z-x}}{(z-x)!} = \frac{e^{-\lambda-\mu}}{z!}\sum_{x=0}^{z}\binom{z}{x}\lambda^x\mu^{z-x} = \frac{e^{-\lambda-\mu}}{z!}(\lambda+\mu)^z$.
$\implies z \sim P(\lambda + \mu)$.
Note: $X \sim B(n_1, p), Y \sim B(n_2, p)., X + Y \sim B(n_1 + n_2, p)$.

7 Expected Value of Discrete Random Variables

Definition 7.1. The **expected value** of a discrete random variable, denoted by $\mathbb{E}[X] = \sum_x x p_X(x) = \sum_x x \mathbb{P}(X = x)$. This definition makes sense if the right side is absolutely summable, i.e. $\sum_x p_X(x) < \infty$. We say that X has a finite expectation.

Example 7.2. Following are some examples of this concept:

1. Let $X \sim \text{Bernoulli}(p)$, then $\mathbb{E}[X] = 1 \cdot p + 0 \cdot (1-p) = p$.

2. Let $X \sim B(n,p); X = \{0, \dots, \}$. $\mathbb{E}[X] = \sum_x x p_X(x) = \sum_{x=0}^n x \binom{n}{x} p^x (1-p)^{n-x} = \sum_{x=1}^n \frac{n!}{(x-1)!(n-x)!} p^x (1-p)^{n-x} = np \sum_{x=1}^n \frac{(n-1)!}{(x-1)!(n-x)!} p^{x-1} (1-p)^{n-x} = np \sum_{x=1}^n \binom{n-1}{x-1} p^{x-1} (1-p)^{n-x} = np \sum_{j=1}^{n-1} \binom{n-1}{j} p^j (1-p)^{(n-1-j)} = np$.

3. Let $X \sim \text{Poisson}(\lambda)$: $\mathbb{E}[X] = \sum_x x p_X(x) = \sum_{x=0}^\infty x e^{-\lambda} \frac{\lambda^x}{x!} = e^{-\lambda} \sum_{x=1}^\infty \frac{\lambda^x}{(x-1)!} = \lambda e^{-\lambda} \sum_{x=1}^\infty \frac{\lambda^{x-1}}{(x-1)!} = \lambda e^{-\lambda} \sum_{k=0}^\infty \frac{\lambda^k}{k!} = \lambda e^{-\lambda} \cdot e^\lambda = \lambda$. So, $\mathbb{E}[X] = \lambda$.

In general, if g is a real valued function defined on the range of random variable X, then $\mathbb{E}[g(X)] = \sum_x g(x) p_X(x)$. In particular, if $g(x) = x$,

$$\mathbb{E}[X] = \sum_x x p_X(x)$$

$$\mathbb{E}[g(X,Y)] = \sum_x \sum_y g(x,y) p_{X,Y}(x,y), \text{ provided } \sum_{x,y} (g(x,y) \mid p_{X,Y}(x,y)) < \infty.$$

7.1 Fundamental Properties of Expected Value

1. If $X =$ constant c,

$$\mathbb{E}[X] = \mathbb{E}[c] = \sum_x c \cdot p_X(x) = c \sum_x p_X(x) = c.$$

2. If c is a constant, then

$$\mathbb{E}[cX] = c \cdot \mathbb{E}[X].$$

3. $\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$.

4. $\mathbb{E}[a + bX] = a + b\mathbb{E}[X]$, where $a, b \in \mathbb{R}$ and X is a discrete random variable.

5. $\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]; a, b \in \mathbb{R}$.

6. If $X \leq Y$ then $\mathbb{E}[X] \leq \mathbb{E}[Y]$.

7. If X, Y are independent random variables, then

$$\mathbb{E}[XY] = \mathbb{E}[X] \cdot \mathbb{E}[Y].$$

In general, if $X_1, \dots, X_n$ are independent, then

$$\mathbb{E}[\prod_{i=1}^{n} X_i] = \prod_{i=1}^{n} \mathbb{E}[X_i]$$

and always (i.e. independent or not)

$$\mathbb{E}[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} \mathbb{E}[X_i].$$

Example 7.3. Consider the following examples:

1. $S = X + \sin(Y) + e^Z$

$$\mathbb{E}[S] = \mathbb{E}[X] + \mathbb{E}[\sin(Y)] + \mathbb{E}[e^Z].$$

2. If $X \sim Bin(n, p)$, then $\mathbb{E}[X] = np$. If Y_i are Bernoulli random variables with parameter p, then $X = \sum_{i=1}^{n} Y_i$.

$$\mathbb{E}[X] = \mathbb{E}[\sum_{i=1}^{n} Y_i] = \sum_{i=1}^{n} \mathbb{E}[Y_i] = \sum_{i=1}^{n} p = np.$$

3. Let $X \sim Bin(n, p)$. Then,

$$\mathbb{E}[e^X] = \sum_{x=0}^{n} e^x p_X(x) = \sum_{x=0}^{n} e^x \binom{n}{x} p^x (1-p)^{n-x}$$

$$= \sum_{x=0}^{n} (ep)^x \binom{n}{x} p^x (1-p)^{n-x} = (ep + 1 - p)^n$$

Recall that for a random variable X,

$$\mathbb{E}[X] = \sum_{x} x \cdot p_X(x).$$

To develop this definition, the expectation is interpreted a long-run average value of the random variable in repeated trials. That is, for a random variable X, let $X_1, \dots, X_n$ are n observed values of X, then

$$\frac{X_1 + \dots + X_n}{n} \approx \mathbb{E}[X], \text{ for large } n.$$

Here,

$$\mathbb{E}[X] \approx \frac{X_1 + \dots + X_n}{n} = \frac{1}{n} \sum_{k=1}^{m} n(\{X = x_k\}) x_k = \sum_{x=1}^{n} x_k \frac{n(\{X = x_k\})}{n} = \sum_{k=1}^{n} x_k \mathbb{P}(X = x_k)$$

$$= \sum_{k=1}^{n} x_k p_X(x_k); \frac{n(E)}{n} \approx \mathbb{P}(E) = \sum_x x p_X(x).$$

7.1.1 Tail Probabilities

We can use tail probabilities to obtain expected values: Let $X \geq 0$ be an integer-values random variable. Then X has finite expectation (i.e. $\mathbb{E}[X] < \infty$) if and only if $\sum_{n=0}^{\infty} \mathbb{P}(X > n) < \infty$. And here, $\mathbb{E}[X] = \sum_{n=0}^{\infty} \mathbb{P}(X > n)$.

Example 7.4. Let $X \sim g(p)$; geometric random variable with parameter p. Check that $\mathbb{E}[X] = \sum_x x p_X(x)$.
Solution: $\mathbb{P}(X > n) = (1-p)^n$. Hence, $\mathbb{E}[X] = \sum_{n=0}^{\infty} \mathbb{P}(X > n) = \sum_{n=0}^{\infty} (1-p)^n = \frac{1}{1-(1-p)} = \frac{1}{p}$.

7.2 Mean and Variance

Recall: For a given function g, $\mathbb{E}[g(X)] = \sum_x g(x) p_X(x)$ if $g(x) = x^r$; $r \geq 1$ is an integer. Also, $\mathbb{E}[X^r] = \sum_x x^r p_X(x)$; provided $\sum_x |x|^r p_X(x) < \infty$.
In this case, $\mathbb{E}[X^r]$ is known as the r^{th} moment of random variable X. We let $\mu_X = \mathbb{E}[X]$ be the mean of X (the expectation of X).

Definition 7.5. Let X be a random variable with $\mathbb{E}[X^2] < \infty$. Then the variance of X, denoted by $var(X)$, is defined by

$$var(X) = \mathbb{E}[(X - \mu_X)^2]; \mu_X = \mathbb{E}[X].$$

$var(X) = \mathbb{E}[(X - \mu_X)^2]$ is a measure of variation of possible values of random variable X. So, if $var(X)$ is large, then on average, X is far from its mean.
Proposition: Let X be a random variable. Then, $var(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2; \mathbb{E}[X^2] < \infty$.
Summary: Moments of X are $\mathbb{E}[X^r]; r \geq 1$. $var(X) = \mathbb{E}[X - \mathbb{E}[X]^2] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$. As a note, higher moments imply lower moments.

7.2.1 Properties of Variance

1. $\sigma_X^2 = var(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]$.

2. Standard deviation (denoted by σ_X) $= \sqrt{\sigma_X^2} = \sqrt{var(X)}$.

3. A random variable X has zero variance if it is a constant random variable. In fact, $var(X) = var(c) = \mathbb{E}[(c - \mathbb{E}[c])^2] = \mathbb{E}[c - c]^2 = 0$.

4. $var(cX) = c^2 var(X)$ where $c \in \mathbb{R}$ and $var(X) < \infty$, since

$$var(cX) = \mathbb{E}[(cX - \mathbb{E}[cX])^2] = \mathbb{E}[c^2(X - \mathbb{E}[X])^2] = c^2\mathbb{E}[(X - \mathbb{E}[X])^2] = c^2 var(X).$$

5. $var(c + X) = var(X)$.

Consequently, $var(aX+b) = a^2 var(X)$.

Example 7.6. Following are some examples:

1. $var(9X+5) = 81var(X)$.
2. $X \sim Ber(p); \mathbb{E}[X] = p$. $var(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = 0^2(1-p) + 1^2 p - (0 \cdot (1-p) + 1 \cdot p)^2 = p - p^2 = p(1-p)$.

7.3 Standardized Random Variable

Let X be a random variable such that $0 < var(X) < \infty$. Then, the standardized random variable (denoted by X^*) is defined by

$$X^* = \frac{X - \mathbb{E}[X]}{\sqrt{var(X)}} = \frac{X - \mu_X}{\sigma_X}.$$

Note: Mean of a standardized random variable is zero:

$$\mathbb{E}[X^*] = \mathbb{E}[\frac{X-\mu_X}{\sigma_X}] = \frac{1}{\sigma_X}\mathbb{E}[X-\mu_X] = \frac{1}{\sigma_X}(\mathbb{E}[X] - \mathbb{E}[\mu_X]) = \frac{1}{\sigma_X}(\mu_X - \mu_X) = 0.$$

Also, variance of a standardized random variable X^* is 1:

$$var(X^*) = var(\frac{X-\mu_X}{\sigma_X}) = \frac{1}{\sigma_X^2} var(X - \mu_X) = \frac{1}{\sigma_X^2} var(X) = \frac{1}{\sigma_X^2} \cdot \sigma_X^2 = 1.$$

Thus, X^* has mean 0 and variance 1, which is why it is called the standardized random variable.
Recall that $var(X) = \mathbb{E}[(X-\mu_X)^2] = \sum_x (X-\mu_X)^2 p_X(x); \mu_X = \mathbb{E}[X]$.
Proposition: Chebyshev's Inequality: Let X be a random variable with $var(X) < \infty$. Then, for each $\delta > 0$,

$$\mathbb{P}(|X - \mu_X| \geq \delta) \leq \frac{var(X)}{\delta^2}.$$

Note: Chebyshev's inequality also provides a method for obtaining bounds for probabilites of a random variable based on its mean of variance.

Example 7.7. Let $X \sim Bin(n,p)$. Then, $\mathbb{E}[X] = np; var(X) = np(1-p)$. Can you bound the probability that X differs from its expected value more than 3 units? (i.e. $\mathbb{P}(|X - \mu_X| \geq 3$?)
Solution: By Chebyshev's inequality, $\mathbb{P}(|X - \mu_X| \geq 3) \leq \frac{var(X)}{3^2} = \frac{var(X)}{9} \implies \mathbb{P}(|X - \mu_X| \geq 3) \leq \frac{np(1-p)}{9}$.

7.4 Covariance and Correlation

Recall: $\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$. $\mathbb{E}[XY] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$ if X, Y are independent. Can you write $var(X+Y) = var(X) + var(Y)$?

We start by studying the variance of the sum of two random variables.

$$\mu_{X+Y} = \mathbb{E}[X+Y] = \mu_X + \mu_Y = \mathbb{E}[X] + \mathbb{E}[Y].$$

Now, we know $var(X) = \mathbb{E}[(X-\mu_X)^2]$. We consider $((X+Y)-\mu_{X+Y})^2 = ((X+Y)-(\mu_X+\mu_Y))^2 = ((X-\mu_X)+Y-\mu_Y))^2 = (X-\mu_X)^2 + (Y-\mu_Y)^2 + 2(X-\mu_X)(Y-\mu_Y)$. Taking expectations on both sides, we have $\mathbb{E}[(X+Y)-\mu_{X+Y}]^2 = \mathbb{E}[(X-\mu_X)^2] + \mathbb{E}[(Y-\mu_Y)^2] + 2\mathbb{E}[(X-\mu_X)\cdot(Y-\mu_Y)]$.
$\implies var(X+Y) = var(X) + var(Y) + 2\mathbb{E}[(X-\mu_X)(Y-\mu_Y)]$, therefore no.
That extra part at the end, the $2\mathbb{E}[(X-\mu_X)(Y-\mu_Y)]$, is known as the **covariance**.

Definition 7.8. Let X, Y be two random variables such that $\sigma_X^2, \sigma_Y^2 < \infty$. The covariance of X, Y (denoted by $cov(X,Y)$) is

$$cov(X,Y) = \mathbb{E}[(X-\mu_X)(Y-\mu_Y)].$$

Note: If $X = Y$, $cov(X,Y) = var(X) = \mathbb{E}[(X-\mu_X)^2]$.

7.4.1 Some Properties of Covariance

Let X, Y, Z be random variables with finite variables. Then,

1. $cov(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\cdot\mathbb{E}[Y]$.
2. $cov(X,Y) = cov(Y,X)$.
3. $cov(cX,Y) = c\cdot cov(X,Y)$.
4. $cov(X+Y,Z) = cov(X,Z) + cov(Y,Z)$.
   More generally, $var(\sum_{k=1}^m X_k) = \sum_{k=1}^m var(X_k) + 2\sum_{i<j}\sum cov(X_i, Y_i)$.

Proposition: $var(X+Y) = var(X) + var(Y) + 2cov(X,Y)$.

Example 7.9. As an example, expand $cov(2X+3, -5Y+6)$.

$$= cov(2X, -5Y) = -10cov(X,Y).$$

Important Property: Let X, Y be a independent random variables. then, $cov(X,Y) = 0$. However, $cov(X,Y) = 0$ does not imply that X, Y are independent. Using this, however, the $var(X+Y) = var(X) + var(Y)$.

Definition 7.10. Does the correlation coefficient say something about how much and how X, Y are related? Consider standardized random variables, $X^* = \frac{X-\mu_X}{\sigma_X}$ and $Y^* = \frac{Y-\mu_Y}{\sigma_Y}$. The correlation coefficient between X, Y is

$$\rho(X,Y) = cov(X^*,Y^*) = \frac{cov(X,Y)}{\sqrt{var(X)}\sqrt{var(Y)}}.$$

7.4.2 Properties of Correlation Coefficient

Let X, Y be random variables with finite non-zero variance. Then,

1. $|\rho(X, Y)| \leq 1$.
2. If X, Y are independent, then $\rho(X, Y) = 0$.
3. $\rho(X, Y) = 1$ iff $Y = a + bX$, with $a \in \mathbb{R}, b > 0$.
4. $\rho(X, Y) = -1$ iff $Y = a + bX$, with $a \in \mathbb{R}, b < 0$.
5. $\rho(X, Y) > 0$ means that X, Y are positively correlated.
6. $\rho(X, Y) < 0$ means that X, Y are negatively correlated.

Example 7.11. If $var(X) = 0.3376, var(Y) = 0.56, cov(X, Y) = 0.224$, then $\rho(X, Y) = \frac{0.224}{\sqrt{var(X)} \cdot \sqrt{var(Y)}} = 0.575 \implies X, Y$ are positively correlated.

7.5 Conditional Expectation

$$\mathbb{P}(B) = \sum P(B \mid A_n) \cdot \mathbb{P}(A_n)$$

where A_n's were the partitioned events.

Example 7.12. Let N = number of customers arriving at a bar on a given day, and S_i = amount of time it takes each customer to be served. We are interested in knowing the total amount of time to be served all customers: $Y_N = \sum_{i=1}^{N} S_i$. S_i are identically distributed. $\mathbb{E}[S_i] = \mu$. $\mathbb{E}[Y_n] = \mathbb{E}[\sum_{i=1}^{n} S_i] = \mathbb{E}[S_1] + \mathbb{E}[S_2] + \cdots + \mathbb{E}[S_n] = n \cdot \mu$. $\mathbb{E}[Y_N] = \mathbb{E}[\sum_{i=1}^{N} S_i] = \mu \mathbb{E}[N]$.

$$\mathbb{E}[X] = \sum x \cdot_X (x).$$

$$\mathbb{E}[g(X)] = \sum g(x) p_X(x).$$

$$\mathbb{E}[X + Y] = \sum (X + Y) p_{X,Y}(x, y).$$

Recall that for two random variables X, Y, the conditional PMF of Y given $X = x$; $p_{Y|X}(y \mid x) = \frac{p_{X,Y}(x,y)}{p_X(x)}, p_X(x) > 0$.

Definition 7.13. Let X, Y be two random variables. If $p_X(x) > 0$, we define the conditional expectation of Y given $X = x$ as $\mathbb{E}[Y \mid X = x] = \sum_y y \cdot p_{Y|X}(y \mid x)$.

7.5.1 Properties

1. $\mathbb{E}[cY \mid X = x] = c\mathbb{E}[Y \mid X = x]$.
2. $\mathbb{E}[Y + Z \mid X = x] = \mathbb{E}[Y \mid X = x] + \mathbb{E}[Z \mid X = x]$.

3. $\mathbb{E}[a + bY \mid X = x] = a + b\mathbb{E}[Y \mid X = x]$.

Example 7.14. Let $X \sim P(\lambda), Y \sim P(\mu)$. Find $\mathbb{E}[X \mid X + Y = Z]$, for each non-negative integer z.
Solution: First, we need to find the distribution of $Z = X + Y$.

$$\therefore X \sim P(\lambda); Y \sim P(\mu);\ \text{independent}$$

$$Z = X + Y \sim P(\lambda + \mu).$$

Now, we want to obtain $\mathbb{E}[X \mid Z = z] = \sum x \cdot p_{X|Z}(x \mid z)$. Now, we find conditional PMF of X given $Z = z$. So, $p_{X|Z}(z \mid z) = \frac{p_{X,Z}(x,z)}{p_Z(z)} = \frac{P(X=x,Z=z)}{P(Z=z)} = \frac{P(X=x,Y=z-x)}{P(Z=z)} = \frac{P(X=x)P(Y=z-x)}{P(Z=z)} = \frac{\frac{e^{-\lambda}\lambda^x}{x!}\frac{e^{-\lambda}\lambda^{z-x}}{(z-x)!}}{\frac{e^{-(\lambda+\mu)}(\lambda+\mu)^z}{z!}} = \binom{z}{x}\frac{\lambda^x\mu^{z-x}}{(\lambda+\mu)^z} = \binom{z}{x}\frac{\lambda^x}{(\lambda+\mu)^x} \cdot \frac{\mu^{z-x}}{(\lambda+\mu)^{z-x}} = \binom{z}{x}(\frac{\lambda}{\lambda+\mu})^x(1 - \frac{\lambda}{\lambda+\mu})^{z-x}$.
So, $X \mid Z = x \sim Bin(z, \frac{\lambda}{\lambda+\mu})$. We know $\mathbb{E}[Bin(n,p)] = np \implies \mathbb{E}[X \mid Z = z] = z \cdot \frac{\lambda}{\lambda+\mu} = \frac{z\lambda}{\lambda+\mu}$.

We define

$$\psi(X) = \mathbb{E}[Y \mid X]$$

and $\mathbb{E}[Y \mid X]$ is called the conditional expectation of Y given X.

7.6 Law of Total Expectation

Let X, Y be two discrete random variables with $\mathbb{E}[Y] < \infty$. Then,

$$\mathbb{E}[Y] = \sum_x \mathbb{E}[Y \mid X = x] \cdot p_X(x) = \mathbb{E}[\psi(X)] - \mathbb{E}[\mathbb{E}[Y \mid X]].$$

$$\implies \mathbb{E}[\mathbb{E}[Y \mid X]] = \mathbb{E}[Y].$$

In fact, we know

$$\mathbb{E}[g(X,Y)] = \sum_{x,y} g(x,y) p_{X,Y}(x,y).$$

If we consider $g(x,y) = y$, then $\mathbb{E}[Y] = \sum\sum_{(x,y)} y \cdot_{X,Y}(x,y) = \sum\sum_{(x,y)} y \cdot_X x \cdot_{Y|X}(y \mid x) = \sum_x p_X x(\sum_y y \cdot_{Y|X}(y \mid x) = \sum_x p_X x \mathbb{E}[Y \mid X = x] \implies \mathbb{E}[Y] = \sum_x [Y \mid X = x] \cdot p_X(x)$.

Example 7.15. Let N = number of customers arriving at a bank on a given day (N is a random variable). S_i = amount of time it takes each customer to be served. Then, we are interested in knowing the total amount of time to serve all customers, i.e. $Y_N = \sum_{i=1}^{N} S_i$. Aim: $\mathbb{E}[Y_N]$.
Solution: We assume that N is independent to each S_i and S_i are identically distributed. $\mathbb{E}[S_i] = \mu$ for all i. We have $\mathbb{E}[Y_N \mid N = 0] = \mathbb{E}[Y_0 \mid N = 0] = \mathbb{E}[0 \mid N = 0) = 0$. For $n \geq 1$, $\mathbb{E}[Y_N \mid N = n] = \mathbb{E}[Y_n \mid N = n] = \mathbb{E}[Y_n](\because S_i's$ are independent of $N)= \mathbb{E}[\sum_{i=1}^{n} S_i] = \sum_{i=1}^{n} \mathbb{E}[S_i] = n \cdot \mathbb{E}[S_i]$.. Hence, $\mathbb{E}[Y_N \mid N] = \mathbb{E}[S_1] \cdot N$. Hence, by the Law of Total Explanation, $\mathbb{E}[Y_N] = \mathbb{E}[\mathbb{E}[Y_N \mid N]] = \mathbb{E}[N \cdot \mathbb{E}[S_i]] = \mathbb{E}[S_i]\mathbb{E}[N] \implies \mathbb{E}[Y_N] = \mathbb{E}[n]\mathbb{E}[S_i]$.

7.7 Conditional Variance

$$var(X) = \mathbb{E}[(X - \mu_X)^2]; \mu_X = \mathbb{E}[X].$$

$$var(Y \mid X = x) = \mathbb{E}[(Y - \mathbb{E}[Y \mid X = x])^2 \mid X = x]$$

$$= \mathbb{E}[Y^2 \mid X = x] - (\mathbb{E}[Y \mid X = x])^2.$$

8 Continuous Random Variables & their Distributions

8.1 Introducing Continuous Random Variables

Example 8.1. Suppose that a number is selected at random from the interval $(0, 1)$. Let X denote the number obtained. Find $\mathbb{P}(X = x)$ for $x \in \mathbb{R}$.
Solution: Here, the sample space $\Omega = (0, 1)$. For any event E,

$$\mathbb{P}(E) = \frac{|E|}{|\Omega|} = |E|; E \subset \Omega.$$

Random variable is $X(x) = x$ if $x \notin (0, 1) \implies \mathbb{P}\{X = x\} = 0$. If $x \in (0, 1)$: $\{X = x\} = \{\omega \mid X(\omega) = x\} = \{x \mid X(x) = x\} = \{x\}$. Hence, $\mathbb{P}\{X = x\} = |\{x\}| = 0$. Hence, $\forall x \in \mathbb{R}, \mathbb{P}\{X = x\} = 0$.

Definition 8.2. A random variable X is called a **continuous random variable** if $\mathbb{P}(X = x) = 0 \; \forall \; x \in \mathbb{R}$. Note: "Continuous" for these random variables is because the range of such random variables form a continuum of real numbers (e.g. in previous example, the range of X was $(0, 1)$).

If X is a discrete random variable,

$$\mathbb{P}(X \in A) = \sum_x \mathbb{P}(X = x).$$

In case of a continuous random variable,

$$\mathbb{P}(X = x) = 0.$$

8.2 Cumulative Distribution Functions

Definition 8.3. Let X be a random variable (discrete/continuous). The **cumulative distribution function** (CDF) of X, denoted F_X, is the real-valued f^n

$$F_X : \mathbb{R} \to [0,1] \; s.t. \; F_X(x) = \mathbb{P}(X \leq x); \; x \in \mathbb{R}.$$

Note: CDF applies to any random variable, discrete or continuous.

Example 8.4. Toss a coin three times, and let X be the number of heads. Determine the CDF of X and compare it with the PMF. Note, X is a discrete random variable. So,

$$X(\omega) = \begin{cases} 0;\ \omega = TTT \\ 1;\ \omega \in \{HTT, THT, TTH\} \\ 2;\ \omega \in \{HHT, HTH, THH\} \\ 3;\ \omega = HHH \end{cases}.$$

Solution: The PMF of X is

$$p_X(x) = \mathbb{P}(X = x) = \begin{cases} \frac{1}{8};\ x = 0 \\ \frac{3}{8};\ x = 1 \\ \frac{3}{8};\ x = 2 \\ \frac{1}{8};\ x = 3 \\ 0;\ else \end{cases}.$$

Now, for CDF: $F_X(x) = \mathbb{P}(X \leq x)\ \forall\ x \in \mathbb{R}$.
Suppose $x < 0$(e.g. $x = -0.7$):

$$\mathbb{P}(X \leq x) = 0.$$

Suppose $x = 0$:

$$\mathbb{P}(X \leq x) = \mathbb{P}(X \leq 0) = \mathbb{P}(X = 0) = \frac{1}{8}.$$

Suppose $0 \leq x < 1$:

$$\mathbb{P}(X \leq x) = \mathbb{P}(X = 0) = \frac{1}{8}.$$

Suppose $1 \leq x < 2$:

$$\mathbb{P}(X \leq x) = \mathbb{P}(X = 0) + \mathbb{P}(X = 1) = \frac{1}{8} + \frac{3}{8} = \frac{1}{2}.$$

Suppose $2 \leq x < 3$:

$$\mathbb{P}(X \leq x) = \mathbb{P}(X = 0) + \mathbb{P}(X = 1) + \mathbb{P}(X = 2) = \frac{7}{8}.$$

Suppose $x \geq 3$:

$$\mathbb{P}(X \leq 3) = 1.$$

Thus,

$$F_X(x) = \begin{cases} 0;\ x < 0 \\ \frac{1}{8};\ 0 \leq x < 1 \\ \frac{1}{2};\ 1 \leq x < 2 \\ \frac{7}{8};\ 2 \leq x < 3 \\ 1;\ 3 \leq x \end{cases}.$$

Example 8.5. Let X denote a number selected at random from $(0,1)$. Find the CDF of X.
Solution: The cases of $x < 0$ and $x \geq 1$ are easy. The only problem is when $0 \leq x < 1$. To figure this out, $F_X(x) = \mathbb{P}(X \leq x) = \mathbb{P}\{\omega : X(\omega) \leq x\} = |(0,x)| = x$. Therefore,

$$F_X(x) = \begin{cases} 0; \ x < 0 \\ x; \ 0 \leq x < 1 \\ 1; \ x \geq 1 \end{cases} .$$

Proposition: The CDF F_X of random variable X satisfies:

1. F_X is non-decreasing, i.e. if $t_1 < t_2, F_X(t_1) \leq F_X(t_2)$.
2. F_X is a right-continuous function, i.e. for every $t \in \mathbb{R}$, $F_X(t^+) = F_X(t)$. Also, for every $t \in \mathbb{R}$, $F_X(t^-)$ exists.
3. $\lim_{x\to\infty} F_X(x) = 1$, and $\lim_{x\to-\infty} F_X(x) = 0$.

8.2.1 Computing Probability using CDF

Let X be a random variable with CDF F_X. Let $a, b \in \mathbb{R}, a < b$.

1. $\mathbb{P}(X \leq a) = F_X(a)$.
2. $\mathbb{P}(a < X \leq b) = F_X(b) - F_X(a)$.
3. $\mathbb{P}(a < X \leq b) = F_X(b) - F_X(a);\ a < b$.
4. $\mathbb{P}(X < a) = F_X(a^-)$.
5. $\mathbb{P}(X = a) = F_X(a) - F_X(a^-)$.
6. $\mathbb{P}(a \leq X \leq b) = F_X(b) - F_X(a^-)$.
7. $\mathbb{P}(X \geq a) = 1 - F(a^-)$.

Proposition: A random variable is continuous iff its CDF is an everywhere continuous function.

8.3 Probability Density Functions

Definition 8.6. Let X be a continuous random variable. A non-negative function

$$f_X : \mathbb{R} \to [0, \infty)$$

is said to be a **probability density function** (PDF) of X, if for all $a < b$,

$$\mathbb{P}(a \leq X \leq b) = \int_a^b f_X(x)\,dx.$$

Proposition: Let X be a continuous random variable. If $F'_X(x)$ exists and is continuous, then X has a PDF, and it is given by

$$f_X(x) = \begin{cases} F'_X(x), & \text{if } f_X \text{ is differentiable at } x \\ 0, \; o/w \end{cases}.$$

Proposition: A random variable X has a PDF iff there exists a non-negative function f defined on $\mathbb{R}$ such that

$$F_X(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^{x} f(t)\,dt, x \in \mathbb{R}.$$

In this case, f is a PDF of X and $f(x) = F'_X(x)$.

8.3.1 Basic Properties of PDF

Suppose X is a continuous random variable with PDF f_X and CDF F_X.

1. Since f_X is a density of probability (rather than probability), f_X may be ≥ 1.
2. $\int_{-\infty}^{\infty} f_X(x)\,dx = 1$ ($\because \mathbb{P}(X < \infty) = \int_{-\infty}^{\infty} f_X(x)\,dx = 1$).
3. $\mathbb{P}(a < X \leq b) = \mathbb{P}(a \leq X \leq b) = \mathbb{P}(a \leq X < b)$.

8.3.2 Fundamental Probability Formula

Proposition: Suppose that X is a continuous random variable with PDF f_X. Then, for any subset $A \subset \mathbb{R}$:

$$\mathbb{P}(X \in A) = \int_A f_X(x)\,dx.$$

For a discrete case, $\mathbb{P}(X \in A) = \sum_{x \in A} p_X(x)$.

Example 8.7. Let

$$f_X(x) = \begin{cases} 0; \ x < 2 \\ \frac{c}{x^2}; x \geq 2 \end{cases}.$$

1. Find c such that $f_X(x)$ is PDF.

2. Compute $\mathbb{P}(X \geq 5)$.

3. Find $\mathbb{P}(X > 0)$.

Solution:

1. $\int_{-\infty}^{\infty} f_X(x)\,dx \implies \int_{-\infty}^{2} f_X(x)\,dx + \int_{2}^{\infty} f_X(x)\,dx = 1 \implies 0 + \int_{2}^{\infty} \frac{c}{x^2}\,dx = 1 \implies c = 2$. So, for $c = 2$, $f_X(x)$ is a PDF of random variable X.

2. From FPF, we know $\mathbb{P}(X \in A) = \int_A f_X(x)\,dx$. Here, in this case, $A = [5, \infty) \subset \mathbb{R}$.
$$\mathbb{P}(X \in [5, \infty)) = \mathbb{P}(X \geq 5) = \int_5^{\infty} f_X(x)\,dx = \int_5^{\infty} \frac{2}{x^2}\,dx = \frac{2}{5}.$$
So, $\mathbb{P}(X \geq 5) = \mathbb{P}(X > 5) = \frac{2}{5}$ ($\because X$ is a continuous random variable).

3. $\mathbb{P}(X > 0) = \int_0^{\infty} f_X(x)\,dx = 1$.

Example 8.8. Let's now find the CDF from the PDF from the previous example. We know $\mathbb{P}(X \leq x) = f_X(x) = \int_{-\infty}^{\infty} f_X(x)\,dx$. So, $F_X(x) = \int_{-\infty}^{x} f_X(y)\,dy = 0$, if $x \leq 2$.
For $x > 2$,
$$F_X(x) = \int_{-\infty}^{x} f_X(y)\,dx = \int_{2}^{x} f_X(y)\,dy = (1 - \frac{2}{x}).$$
So,
$$F_X(x) = \begin{cases} 0; \ x \leq 2 \\ 1 - \frac{2}{x}; \ x > 2 \end{cases}.$$
$$f_X(x) = \begin{cases} 0; \ x < 2 \\ \frac{2}{x^2}; \ x \geq 2 \end{cases}.$$

8.4 Uniform and Exponential Random Variables

Definition 8.9. A continuous random variable X is called a **uniform random variable** over (a, b) if its value is equally likely to lie anywhere in this interval. We write

$$X \sim U(a, b)$$

and its PDF is

$$f_X(x) = \begin{cases} 0; \ x \notin (a, b) \\ \frac{1}{b-a}; \ x \in (a, b) \end{cases} .$$

Example 8.10. A commuter train arrives at a station every half an hour. Let X be the amount of time in minutes that we need to wait for the train to arrive.

1. $X \sim U(0, 30)$.
$$f_X(x) = \begin{cases} 0; \ x \notin (0, 30) \\ \frac{1}{30}; \ x \in (0, 30) \end{cases}$$
2. $\mathbb{P}(5 \leq X \leq 15)$ = probability that waiting time is in between 5 and 15 = $\int_5^{15} f_X(x)\, dx = \int_5^{15} \frac{1}{30}\, dx = \frac{1}{30} \cdot 10 = \frac{1}{3}$.
3. $\mathbb{P}(X \leq 10) = \frac{10}{30} = \frac{1}{3}$.
4. $\mathbb{P}(X > 10) = 1 - \mathbb{P}(X \leq 10) = 1 - \frac{1}{3} = \frac{2}{3}$.

8.4.1 Exponential Random Variables

Definition 8.11. A continuous random variable X is called an **exponential random variable** if it has PDF

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}; \ x > 0 \\ 0; \ o/w \end{cases}$$

where $\lambda > 0$. We write $X \sim Exp(\lambda)$. The CDF of X is given by

$$F_X(x) = \int_{-\infty}^{x} f_X(y)\, dy = \begin{cases} 1 - e^{-\lambda x}; \ x \geq 0 \\ 0; \ x < 0 \end{cases} .$$

Example 8.12. The time until the first patient arrives at the emergency room has an exponential distribution with $\lambda = 6.9$. Determine the probability that beginning at 6 PM on any given day, the first patient arrives

1. between 6:15 PM and 6:30 PM.

2. before 7 PM.

Solution: Let random variable X denote the time in minutes until the first patient arrives. $X \sim Exp(6.9)$.

$$f_X(x) = \begin{cases} 6.9e^{-6.9x};\ x > 0 \\ 0;\ else. \end{cases}$$

1. $\mathbb{P}(\frac{1}{4} \leq X \leq \frac{1}{2}) = \int_{\frac{1}{4}}^{\frac{1}{2}} f_X(x)\,dx = \int_{\frac{1}{4}}^{\frac{1}{2}} 6.9e^{-6.9x}\,dx = 0.146$. There is a 14.6% chance that the event happens in that time period.

2. $\mathbb{P}(X < 1) = \int_{-\infty}^{1} f_X(x)\,dx = 0.99$.

Note: Exponential random variables are the continuous analogue of discrete geometric random variables.

8.4.2 Lack of Memory Property

$\mathbb{P}(X > t + s \mid X > s) = \mathbb{P}(X > t); s, t \geq 0$. For example, given $s = 3, t = 2$, $\mathbb{P}(X > 5 \mid X > 3) = \mathbb{P}(X > 2)$. The process doesn't remember the first 3 minutes and it starts to count time from beginning. Indeed, $\mathbb{P}(X > s + t \mid X > s) = \frac{\mathbb{P}(X>s+t, X>s)}{\mathbb{P}(X>s)} = \frac{\mathbb{P}(X>s+t)}{\mathbb{P}(X>s)} = \frac{e^{-\lambda(t+s)}}{e^{-\lambda s}} = e^{-\lambda t} = \mathbb{P}(X > t)$.

Example 8.4. (continued)

3. Given that the patient doesn't arrive by 6:15 PM, determine the probability that they arrive by 6:45 PM.

Solution: $X \sim Exp(6.9)$. We want $\mathbb{P}(X \leq \frac{3}{4} \mid X > \frac{1}{4})$. So, $\mathbb{P}(X \leq \frac{3}{4} \mid X > \frac{1}{4}) = 1 - \mathbb{P}(X > \frac{3}{4} \mid X > \frac{1}{4}) = 1 - \mathbb{P}(X > \frac{1}{4} + \frac{2}{4} \mid X > \frac{1}{4}) = 1 - \mathbb{P}(X > \frac{2}{4}) = 1 - e^{-6.9\frac{1}{2}} = 0.968$.

8.5 Normal(Gaussian) Random Variable

De-Moivre: bell-shaped curve.

Definition 8.5. A continuous random variable X is called a normal random variable (Gaussian) if it has a PDF given as

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}};\ -\infty < x < \infty$$

where μ and $\sigma > 0$ are real numbers. We say X is normally distributed with parameters μ and σ^2.

$$X \sim N(\mu, \sigma^2).$$

PDF of normal random variable with parameters μ and σ^2 is centered at μ and its spread depends of σ. This means μ is called the location parameter, and σ is called the scale parameter. $\mathbb{E}[X] = \mu, var(X) = \sigma^2$.
By FPF, we have $A \subset \mathbb{R}$: $\mathbb{P}(X \in A) = \int_A \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x\mu)^2}{2\sigma^2}}\, dx$.
So, if $X \sim N(\mu, \sigma^2)$, then $Z = \frac{X-\mu}{\sigma} \sim N(0, 1)$.
Indeed, given $Z \in \mathbb{R}$,

$$F_Z(z) = \mathbb{P}(Z \leq z) = \mathbb{P}(\frac{X-\mu}{\sigma} \leq z) = \mathbb{P}(X \leq \mu + \sigma z) = \int_{-\infty}^{(\mu+\sigma z)} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx.$$

If $\frac{x-\mu}{\sigma} = t$,

$$F_Z(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}\, dt.$$

Hence,

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}};\ -\infty < z < \infty \implies Z \sim N(0, 1).$$

Definition 8.6. A normal random variable with parameters 0 and 1 is called a standard normal random variable, and its PDF and CDF are given by

$$PDF:\ \phi(Z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}};\ -\infty < z < \infty.$$

$$CDF:\ \Phi(Z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}\, dt.$$

So, if $X \sim N(\mu, \sigma^2)$,

$$\mathbb{P}(a < X < b) = \mathbb{P}(\frac{a-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \frac{b-\mu}{\sigma}) = \mathbb{P}(\frac{a-\mu}{\sigma} < Z < \frac{b-\mu}{\sigma})$$

$$= \Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma}).$$

And,

$$\Phi(-z) = 1 - \Phi(z).$$

Example 8.7. Let $X \sim N(2,4)$ and $\Phi(2) = 0.9772, \Phi(\frac{1}{2}) = 0.6915$. a) Find $\mathbb{P}(1 \leq X \leq 6)$.
Solution:

$$\mathbb{P}(1 \leq X \leq 6) = \mathbb{P}(\frac{1-2}{2} \leq Z \leq \frac{6-2}{2}) = \mathbb{P}(-\frac{1}{2} \leq Z \leq 2) = \mathbb{P}(Z \leq 2) - \mathbb{P}(Z \leq -\frac{1}{2})$$

$$= \Phi(2) - \Phi(\frac{1}{2}) = \Phi(2) - (1 - \Phi(\frac{1}{2})) = 0.6687.$$

b) Find X_p such that $\mathbb{P}(X \geq X_p) = 0.5$.

$$\mathbb{P}(Z \geq \frac{X_p - 2}{2}) = 0.5 \implies 1 - \mathbb{P}(Z \leq \frac{X_p - 2}{2}) = 0.5 \implies \mathbb{P}(Z \leq \frac{X_p - 2}{2}) = 0.5$$

$$\implies \frac{X_p - 2}{2} = 0 \implies X_p = 2.$$

8.6 Other Important Continuous Random Variables

8.6.1 Gamma Random Variable

The gamma random variable is the continuous variable version of a Negative Binomial random variable.

Definition 8.8. A continuous random variable X is called a **gamma random variable** if it has a PDF of the form

$$f_X(x) = \begin{cases} 0; \ x \leq 0 \\ \frac{\lambda^\alpha x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)}; \ x > 0 \end{cases}$$

$X \sim \Gamma(\alpha, \lambda)$, and its CDF is given by

$$F_X(x) = 1 - e^{-\lambda x} \sum_{j=0}^{\alpha-1} \frac{(\lambda x)^j}{j!}; \ x \geq 0.$$

$\Gamma(x)$ is the gamma function, where

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} \, dx; a > 0.$$

Example 8.9. The time until the third patient arrives has the gamma distribution with parameters $\alpha = 3, \lambda = 6.9$. Determine the probability that, beginning at 6 PM, the third patient arrives between 6:15 and 6:30 PM.
Solution: $X \sim \Gamma(3, 6.9)$. We are asked to find $\mathbb{P}(\frac{1}{4} < X < \frac{1}{2}) = \int_{\frac{1}{4}}^{\frac{1}{2}} f_X(x)\,dx$. If we use CDF, $\mathbb{P}(\frac{1}{4} < X < \frac{1}{2}) = F_X(\frac{1}{2}) - F_X(\frac{1}{4})$.

$$F_X(x) = 1 - e^{-6.9x} \sum_{j=1}^{2} \frac{(6.9x)^j}{j} = 0.420.$$

Special cases:

1. If $x = 1$, $\Gamma(1, \lambda) = Exp(\lambda)$.
2. If $\alpha = \frac{v}{2}$ and $\lambda = \frac{1}{2}$, then $\Gamma(\frac{v}{2}, \frac{1}{2}) = X^2(v)$.

8.7 Functions of continuous random variables

g : real-valued function. X : discrete random variable. $Y = g(X)$ will be discrete.
But, if X : continuous random variable, then $Y = g(X)$ could be either discrete or continuous. There are a few ways to find the distribution of Y.

8.7.1 Method 1 – the CDF Method

If g is monotone, then

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

Example 8.10. Let $X \sim N(\mu, \sigma^2)$ and $Y = 4X - 2$. So, $g(X) = Y = 4X - 2$. Hence, $f_Y(y) = f_X(\frac{y+2}{4}) = \mathbb{P}(X \leq \frac{y+2}{4}) = \int_{-\infty}^{\frac{y+2}{4}} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\,dx.$

Example 8.11. Let $X \sim N(0,1)$ and $Y = X^2$. We note here that $g(x) = x^2$ and the range of random variable Y is $[0, \infty)$ (Note: g is NOT monotone).

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(X^2 \leq y) = \mathbb{P}(-\sqrt{y} \leq X \leq \sqrt{y}) = F_X(\sqrt{y}) - F_X(-\sqrt{y}).$$

So, for $y > 0$, the PDF of Y is

$$f_Y(y) = F_Y'(y) = F_X'(\sqrt{y})\frac{1}{2\sqrt{y}} - F_X'(-\sqrt{y})(\frac{-1}{2\sqrt{y}}) = \frac{1}{2\sqrt{y}}f_X(\sqrt{y}) + \frac{1}{2\sqrt{y}}f_X(-\sqrt{y})$$

$$= \frac{1}{\sqrt{y}}f_X(\sqrt{y}) = \frac{1}{\sqrt{y}}\frac{1}{\sqrt{2\pi}}e^{-\frac{y}{2}}.$$

So, $X \sim N(0,1), f_X(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}, x \in \mathbb{R}$. $Y = X^2$; $f_Y(y) = \frac{1}{\sqrt{y}}\frac{1}{\sqrt{2\pi}}e^{\frac{-y}{2}}$. Evidently, even if $Y = X^2, f_Y \neq (f_X)^2$.
Here,

$$f_Y(y) = \frac{(\frac{1}{2})^{\frac{1}{2}}}{\sqrt{\pi}}y^{\frac{1}{2}-1}e^{-\frac{y}{2}} = \frac{(\frac{1}{2})^{\frac{1}{2}}}{\Gamma(\frac{1}{2})}y^{\frac{1}{2}-1}e^{-\frac{y}{2}}. \ (\sqrt{\pi} = \Gamma(\frac{1}{2}))$$

So, $Y \sim \Gamma(\frac{1}{2}, \frac{1}{2})$: Y has the chi-square distribution with 1 degree of freedom. $X \sim N(0,1); Y \sim \chi^2(1)$.

8.7.2 Method 2 – the PDF Method

Let X be a continuous random variable with CDF F_X, and g be a real-valued, strictly monotone and differentiable on the range of X. Then, the PDF of $Y = g(X)$ is

$$f_Y(y) = \frac{f_X(x)}{|g'(x)|}, \text{where } x = g^{-1}(y).$$

Example 8.12. Let $X \sim N(\mu, \sigma^2)$ and $Y = a - 3X$. Thus, $g(x) = y = a - 3x \implies g'(x) = -3$.
Using PDF formula,

$$f_Y(y) = \frac{f_X(x)}{|g'(x)|} = \frac{1}{3}f_X(x) = \frac{1}{3}f_X(\frac{a-y}{3}).$$

Since $X \sim N(\mu, \sigma^2)$,

$$f_Y(y) = \frac{1}{3}\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2}[\frac{\frac{a-y}{3}-\mu}{\sigma}]^2} = \frac{1}{\sqrt{2\pi(3\sigma^2)}}e^{-\frac{1}{2}[\frac{y-(a-3\mu)}{3\sigma}]^2}.$$

Example 8.13. If $X \sim U(-\frac{\pi}{2}, \frac{\pi}{2})$, find a PDF of $Y = \tan(X)$.

$$f_X(x) = \begin{cases} \frac{1}{\pi};\ -\frac{\pi}{2} < x < \frac{\pi}{2} \\ 0;\ o/w \end{cases}$$

Here, range of Y is $(-\infty, \infty)$. By the PDF method,

$$f_Y(y) = \frac{1}{|g'(x)|} f_X(x) = \frac{1}{\sec^2 x} \cdot \frac{1}{\pi}.$$

So,

$$f_Y(y) = \frac{1}{\sec^2(\tan^{-1}(y))} \cdot \frac{1}{\pi} = \frac{1}{\pi} \cdot \frac{1}{(1+y^2)}.$$

Definition 8.14. A continuous random variable X is called a **Cauchy random variable** if its PDF is given by

$$f_X(x) = \frac{1}{\pi(1+x^2)};\ -\infty < x < \infty.$$

Note: this distribution will have some unusual properties.

9 Jointly Continuous Random Variables

Definition 9.1. Let X, Y be two random variables. Then, the joint CDF of X, Y is

$$F_{X,Y}(x,y) = \mathbb{P}(X \le x, Y \le y);\ x, y \in \mathbb{R}$$

$$F_{X,Y} : \mathbb{R}^2 \to [0,1].$$

9.1 Joint CDFs

If $f_X(x) = \mathbb{P}(X \le x)$, then

$$F_X(x) = \lim_{y\to\infty} F_{X,Y}(x,y) = F_{X,Y}(x,\infty), x \in \mathbb{R}.$$

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

9.2 Joint PDFs

Definition 9.2. Let X, Y be random variables. A non-negative function $f_{X,Y}$ is a **joint PDF** if

$$\mathbb{P}(a \leq X \leq b, c \leq Y \leq d) = \int_a^b \int_c^d f_{X,Y}(x, y)\, dy\, dx.$$

So intuitively,

$$\mathbb{P}(x \leq X \leq x + \Delta x, y \leq Y \leq y + \Delta y) \approx f_{X,Y}(x, y)\Delta x \Delta y.$$

Joint PDF is partial derivatives of joint CDF:
Proposition: Let X, Y be continuous random variables with join CDF $F_{X,Y}$ and joint PDF $f_{X,Y}$. Then,

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

at the continuity points of partials (if they exist).
Proposition: Let X, Y be defined on some sample space with joint PDF $f_{X,Y}(x, y)$. Then,

$$F_{X,Y}(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(s, t)\, ds\, dt; (x, y) \in \mathbb{R}^2$$

and

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

for all points where the mixed derivative exists.

9.3 Properties of Joint Density Functions

For joint PDF $f_{X,Y}(x, y)$:

1. $f_{X,Y}(x, y) \geq 0 \ \forall (x, y) \in \mathbb{R}^2$.
2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dx\, dy = 1$.

9.3.1 Fundamental Probability Formula (FPF)

Suppose that X, Y are two random variables with a joint PDF. Then, for any $A \subset \mathbb{R}^2$,

$$\mathbb{P}((x, y) \in A) = \iint_A f_{X,Y}(x, y)\, dx\, dy.$$

If $(x, y) \in (-\infty, x] \cdot (-\infty, y]$,

$$F_{X,Y} = \mathbb{P}\{-\infty < X \leq x, -\infty < Y \leq y\} = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(s, t)\, dt\, ds.$$

9.4 Marginal and Conditional PDFs

Proposition: Let X, Y be two continuous random variables with joint PDF. Then,

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

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

By differentiating it,

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

Example 9.3. Let X, Y denote lifetimes of electrical components A and B. The joint PDF of X, Y is

$$f_{X,Y}(x,y) = \lambda\mu e^{-(\lambda x + \mu y)};\ x, y > 0.$$

1. Determine the probability that both components are functioning at time t.
2. Determine the probability that A is the first component to fail.
3. Determine the probability that B is the first component to fail.

Solution:

1. Both components functioning at time t means that $X > t, Y > t$. $\mathbb{P}(X > t, Y > t) = \int_t^\infty \int_t^\infty f_{X,Y}(x,y)\,dx\,dy = \int_t^\infty \int_t^\infty \lambda\mu e^{-(\lambda x+\mu y)}\,dx\,dy = e^{-\mu t}e^{-\lambda t}$.
2. The event that component A is the first to fail, i.e. $\{X < Y\}$. $\mathbb{P}(X < Y) = \iint_{x<y} f_{X,Y}(x,y)\,dx\,dy = \int_{x=0}^\infty (\int_{y=x}^\infty \lambda\mu e^{-\lambda x+\mu y)}\,dy)\,dx = \frac{\lambda}{\lambda+\mu}$.
3. Similarly, $\mathbb{P}(Y < X) = \frac{\mu}{\mu+\lambda}$.

9.4.1 Conditional PDFs

Let X, Y be two continuous random variables with joint PDF $f_{X,Y}$.

$$f_{Y|X}(y \mid x) = \frac{f_{X,Y}(x,y)}{f_X(x)};\ f_X(x) > 0, y \in \mathbb{R}.$$

If $f_X(x) = 0$, then $f_{Y|X}(y \mid x) = 0$.

9.4.2 Conditional CDFs

$$F_{Y|X}(y \mid x) = \mathbb{P}(Y \leq y \mid X = x) = \int_{-\infty}^{y} f_{Y|X}(s \mid x)\,ds.$$

So, $\mathbb{P}(Y \in A \mid X = x) = \int_A f_{Y|X}(y \mid x)\,dy$.

9.5 Independent Continuous Random Variables

X, Y are independent if for any two subsets $A, B \subset \mathbb{R}$,

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A) \cdot \mathbb{P}(Y \in B).$$

For discrete random variables, if they are independent, $p_{X,Y}(x, y) = p_X(x) \cdot p_Y(y)$. Thus, for continuous random variables,

$$f_{X,Y} = f_X(x) \cdot f_Y(y).$$

Equivalently, if $f_X(x) > 0$,

$$f_{Y|X}(y \mid x) = f_Y(y);\ \forall\, y \in \mathbb{R}.$$

Example 9.4. Let $(X, Y) \sim$ Uniform in unit disk.

$$f_{X,Y}(x, y) = \begin{cases} \frac{1}{\pi};\ x^2 + y^2 = 1 \\ 0;\ o/w \end{cases}.$$

The marginal PDF of X:

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dy = \begin{cases} 0;\ x \notin [-1, 1] \\ \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \frac{1}{\pi}\, dy;\ x \in (-1, 1) \end{cases}$$

$$= \begin{cases} 0;\ x \notin [-1, 1] \\ \frac{2}{\pi}\sqrt{1 - x^2};\ x \in (-1, 1) \end{cases}.$$

Similarly, the marginal PDF for y is

$$f_Y(y) = \begin{cases} 0;\ y \notin [-1, 1] \\ \frac{2}{\pi}\sqrt{1 - y^2};\ y \in (-1, 1) \end{cases}.$$

The conditional PMF on $X = x$ is

$$f_{Y|X=x}(y \mid x) = \frac{f_{X,Y}(x, y)}{f_X(x)} = \begin{cases} \frac{1}{\pi} \frac{1}{\frac{2}{\pi}\sqrt{1-x^2}};\ |y| \leq \sqrt{1 - x^2} \\ 0;\ o/w \end{cases}$$

$$= \begin{cases} \frac{1}{2\sqrt{1-x^2}};\ |y| \leq \sqrt{1 - x^2} \\ 0;\ o/w \end{cases}.$$

So, $Y \mid X = x \sim Unif(-\sqrt{1 - x^2}, \sqrt{1 - x^2})$. As an example:

$$\mathbb{P}(X^2 + Y^2 \leq \frac{1}{4}) = \iint_{x^2+y^2\leq\frac{1}{4}} \frac{1}{\pi}\, dx\, dy = \frac{1}{\pi} \cdot \frac{\pi}{4} = \frac{1}{4}.$$

9.6 Functions of Two or More Continuous Random Variables

Example 9.5. Let $X_1, X_2 \sim Exp(\lambda)$, independently and identically distributed. What is $\mathbb{P}(X_1 + X_2 \leq t)$? (Intuitively, $X_1 + X_2 \sim \Gamma(2, \lambda)$.)
Solution: Due to independence, if $Y = X_1 + X_2$,

$$f_Y(y) = f_{X_1,X_2}(x_1, x_2) = \begin{cases} \lambda^2 e^{-\lambda x_1} e^{-\lambda x_2};\ x_1, x_2 > 0 \\ 0;\ o/w \end{cases} \qquad \Longrightarrow$$

$$F_Y(y) = \mathbb{P}(X_1 + X_2 \leq t) = \iint_A f_{X_1,X_2}(x_1, x_2)\, dx_1\, dx_2$$

$$= \int_{x_2=0}^{t} \int_{x_1=0}^{t-x_2} \lambda^2 e^{-\lambda x_1} e^{-\lambda x_2}\, dx_1\, dx_2 = \lambda^2 \int_0^t e^{-\lambda x_2} \Big(\int_0^{t-x_2} e^{-\lambda x_1}\, dx_1 \Big)\, dx_2$$

$$= 1 - e^{-\lambda t} - \lambda t e^{-\lambda t}.$$

So, the PDF of $Y = X_1 + X_2$ is

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \begin{cases} 0;\ y < 0 \\ \lambda^2 y e^{-\lambda y};\ y \geq 0. \end{cases}$$

Example 9.6. $X_i \sim Exp(\lambda_i), i = 1, 2, \ldots, m$. X_i can be considered as the lifetimes of m components. Thus, the lifetime of the system is $X = min\{X_1, \ldots, X_m\}$.

Proposition: Let X, Y be continuous random variables with a joint PDF. Then a PDF of random variables $X + Y$ can be obtained from either of these two:

$$f_{X+Y}(z) = \int_{-\infty}^{\infty} f_{X,Y}(x, z - x)\, dx;\ z \in \mathbb{R}.$$

$$f_{X+Y}(z) = \int_{-\infty}^{\infty} f_{X,Y}(z - y, y)\, dy;\ z \in \mathbb{R}.$$

If X, Y are independent,

$$f_{X+Y}(z) = \int_{-\infty}^{\infty} f_{X,Y}(x, z - x)\, dx = \int_{-\infty}^{\infty} f_X(x) f_Y(z - x)\, dx = f_X \cdot f_Y.$$

Example 9.7. Sum of two independent Gamma random variables: Let $X \sim \Gamma(a, \lambda)$; $Y \sim \Gamma(b, \lambda)$; iid. Then, $Z = X + Y$.
Recall that $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x}\, dx; \alpha > 0$.
Solution: If $X \sim \Gamma(a, \lambda)$,

$$f_X(x) = \frac{\lambda^a}{\Gamma(a)} x^{\alpha-1} e^{-\lambda x};\ x > 0.$$

Since X, Y are independent,

$$f_{X+Y}(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\, dx = f_X \cdot f_Y$$

$$= \int_{x=0}^{z} \frac{\lambda^a}{\Gamma(a)} e^{-\lambda x} x^{a-1} \frac{\lambda^b}{\Gamma(b)} (z-x)^{b-1} e^{-\lambda(z-x)}\, dx$$

$$= \frac{\lambda^{a+b}}{\Gamma(a)\Gamma(b)} e^{-\lambda z} \int_0^z x^{a-1} (z-x)^{b-1}\, dx ...$$

We use the substitution $u = \frac{x}{z}$:

$$= \frac{\lambda^{a+b}}{\Gamma(a)\Gamma(b)} e^{-\lambda z} \int_0^1 (uz)^{a-1} (z-uz)^{b-1} z\, du$$

$$= z^{a+b-1} \frac{\lambda^{a+b}}{\Gamma(a)\Gamma(b)} e^{-\lambda z} \int_0^1 u^{a-1} (1-u)^{b-1}\, du ...$$

Now, using the fact that $B(a,b) = \int_0^1 u^{a-1}(1-u)^{b-1}\, du = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$,

$$= \frac{\lambda^{a+b}}{\Gamma(a+b)} e^{-\lambda z} z^{a+b-1}.$$

This is the PDF of $\Gamma(a+b, \lambda)$. $\therefore X + Y \sim \Gamma(a+b, \lambda)$.

Proposition:

1. If $X_i \sim \Gamma(a_i, \lambda)$ is independent for $i = 1, \ldots, m$:

$$\sum_{i=1}^{m} X_i \sim \Gamma(\sum_{i=1}^{m} a_i, \lambda).$$

2. If $X_i \sim N(\mu_i, \sigma_i^2)$ is independent for $i = 1, \ldots, m$:

$$\sum_{i=1}^{m} X_i \sim N(\sum_{i=1}^{m} \mu_i, \sum_{i=1}^{m} \sigma_i^2).$$

9.7 Multivariate Transformation Theorem

Recall that in 1 dimension, if $Y = g(X)$, then $f_Y(y) = \frac{f_X(x)}{|g'(x)|}$; $x = g^{-1}(y)$.
In multiple dimensions, $U = g(X, Y)$, and $V = h(X, Y)$. This helps define

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

Proposition: Let $(X, Y) \sim f_{X,Y}(x, y)$ and $U = g(X, Y)$ and $V = h(X, Y)$, where g, h are two real-valued functions, such that

$$J(x, y) = \begin{vmatrix} \frac{\partial g}{\partial x}(x, y) & \frac{\partial g}{\partial y}(x, y) \\ \frac{\partial h}{\partial x}(x, y) & \frac{\partial h}{\partial y}(x, y) \end{vmatrix} \neq 0.$$

Then,

$$f_{U,V}(u, v) = \frac{f_{X,Y}(x, y)}{|J(x, y)|}$$

where (x, y) is the unique point in the range of (x, y) such that $g(x, y) = u$ and $h(x, y) = v$.

Example 9.8. Let $X, Y \sim Exp(\lambda)$; *iid*. Let $S = \frac{X}{X+Y}$ and $T = X + Y$. Find $f_{S,T}(s, t)$.
Solution: By independence,

$$f_{X,Y}(x, y) = \lambda^2 e^{-\lambda x} e^{-\lambda y};\ x, y > 0.$$

Here, $S = g(X, Y) = \frac{X}{X+Y}$ and $T = h(X, Y) = X + Y$.

$$J(x, y) = \begin{vmatrix} \frac{\partial g}{\partial x}(x, y) & \frac{\partial g}{\partial y}(x, y) \\ \frac{\partial h}{\partial x}(x, y) & \frac{\partial h}{\partial y}(x, y) \end{vmatrix} = \begin{vmatrix} \frac{y}{(x+y)^2} & \frac{-x}{(x+y)^2} \\ 1 & 1 \end{vmatrix} = \frac{1}{x+y} = \frac{1}{t}.$$

So, rearranging the equations gives $S = \frac{X}{T} \implies X = ST$. Therefore, $T = TS + Y \implies T - TS = Y \implies T(1 - S) = Y$. Using these, we can find the answer by

$$f_{S,T}(s, t) = \frac{f_{X,Y}(x, y)}{|J(x, y)|} = t\lambda^2 e^{-\lambda x} e^{-\lambda y} = t\lambda^2 e^{-\lambda(st)} e^{-\lambda(t(1-s))}$$

$$= t\lambda^2 e^{-\lambda t};\ t > 0, 0 < s < 1.$$

Example 9.9. Let $(X, Y) \sim f_{X,Y}(x, y)$. Find the PDF of $Z = X + Y$.
Solution: Let $Z = X + Y$ and $U = X$.
Should get

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

10 Expected Value of Continuous Random Variables

10.1 Expected Value of a Continuous Random Variable

Recall that in the discrete case, $\mathbb{E}[X] = \sum_x x \cdot \mathbb{P}(X = x)$. Now, let X be a continuous random variable with PDF $f_X(x)$. The expectation of X is defined by

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x \cdot f_X(x)\, dx.$$

Example 10.1. Following are some examples:

1. $X \sim U(a,b)$.

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x)\, dx = \int_a^b x \cdot \frac{1}{b-a}\, dx = \frac{a+b}{2}.$$

2. $X \sim Exp(\lambda)$.

$$\mathbb{E}[X] = \int_0^{\infty} x \cdot \lambda e^{-x}\, dx = \frac{1}{\lambda}$$

3. $X \sim N(\mu, \sigma^2)$, then $\mathbb{E}[X] = \mu$.

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x \cdot \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} x \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx.$$

Let $\frac{x-\mu}{\sigma} = y \implies dx = \sigma\, dy$

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\mu + \sigma y) e^{-\frac{y^2}{2}}\, dy = \mu \int_{-\infty}^{\infty} e^{-\frac{y^2}{2}}\, dy + \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y e^{-\frac{y^2}{2}}\, dy$$

$$= \frac{\mu}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{y^2}{2}}\, dy = \mu.$$

10.2 Basic Properties of Expected Value

1. $\mathbb{E}[c] = c$.

2. $\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$.

3. If X, Y are independent, then $\mathbb{E}[XY] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$.

4. Tail probabilities: If $X \geq 0$ and is continuous, then $\mathbb{E}[X] = \int_0^{\infty} \mathbb{P}(X > y)\, dy$.

Example 10.2. Let $X \sim Exp(\lambda)$. Then, $\mathbb{P}(X > y) = e^{-\lambda y} \implies \mathbb{E}[X] = \frac{1}{\lambda}$.

10.3 Variance and Covariance

$var(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \int_{-\infty}^{\infty} x^2 f_X(x)\, dx - (\int_{-\infty}^{\infty} x f_X(x)\, dx)^2$.
$cov(X,Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$.

Example 10.3. Let $X \sim N(0,1)$. Find $var(X)$.
Solution: $var(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \mathbb{E}[X^2]$.

$$\mathbb{E}[X^2] = \int_{-\infty}^{\infty} x^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\, dx = \frac{2}{\sqrt{2\pi}} \int_0^{\infty} x^2 e^{-\frac{x^2}{2}}\, dx = \frac{2}{\sqrt{2\pi}} \int_0^{\infty} x(xe^{-\frac{x^2}{2}})\, dx.$$

$$= 1.$$

10.4 Conditional Expectation

$\mathbb{E}[Y \mid X = x] = \int_{-\infty}^{\infty} y f_{Y|X}(y \mid x)\, dx$.

10.5 Generating functions and Limit Theorems

Definition 10.4. The moment generating function (MGF) of X is defined as

$$M_X(t) = \mathbb{E}[e^{tx}],\ t \in \mathbb{R}.$$

If X is a continuous random variable,

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

Note: $M_X(t)$ is defined for all $t \in \mathbb{R}$ for which the random variable e^{tx} has finite expectation.

Example 10.5. Given $X \sim Bin(n,p)$, find its MGF.

$$p_X(x) = \binom{n}{x} p^x (1-p)^{n-x},\ x = 0, 1, \ldots, n.$$

$$M_X(t) = \sum_x e^{tx} p_X(x) = \sum_{x=0}^{n} e^{tx} \binom{n}{x} p^x (1-p)^{n-x} = (pe^t + 1 - p)^n.$$

Example 10.6. Let $X \sim Exp(\lambda)$. Find the domain of MGF of X.

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

The domain of MGF for $Exp(\lambda)$ is $\{t \in \mathbb{R} \mid t < \lambda\}$.

10.6 Generating Moments

The r^{th} moment is given by $\mathbb{E}[X^r]$; $r \in \mathbb{N}$.
Here, let X be a random variable. Then, $M_X(t) = \mathbb{E}[e^{tX}]$. Now,

$$\frac{d}{dt} M_X(t) = \frac{d}{dt} \mathbb{E}[e^{tX}] = \mathbb{E}[\frac{\partial}{\partial t} e^{tX}] = \mathbb{E}[Xe^{tX}] \implies$$

$$\frac{d^r}{dt^r}M_X(t) = \mathbb{E}[X^r e^{tX}].$$

At $t = 0$,

$$M_X'(0) = \mathbb{E}[X^r].$$

So,

$$\mathbb{E}[X] = M_X'(0).$$

$$\mathbb{E}[X^2] = M_X''(0).$$

Example 10.7. Let $X \sim N(\mu, \sigma^2)$.

$$M_X(t) = e^{\mu t + \frac{\sigma^2 t^2}{2}} \implies \mathbb{E}[X] = M_X'(0) = \mu \text{ and } \mathbb{E}[X^2] = M_X''(0) = \sigma^2.$$

11 Limit Theorems

Definition 11.1. Let $X_1, \ldots, X_n$ be *iid* random variables with common finite mean μ. Then, the **Law of Large Numbers** tells us that $\forall\, \epsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}(|\frac{X_1 + \cdots + X_n}{n} - \mu| < \epsilon) = 1.$$

Definition 11.2. Let $X_1, X_2, \ldots$ be *iid* random variables from an *unknown* distribution with finite mean and finite variance. Let $S_n = X_1 + \cdots + X_n$ and $Y_n = \frac{S_n - \mathbb{E}[Sn]}{\sqrt{var(S_n)}} = \frac{S_n - n\mu}{\sqrt{n\sigma^2}}$.
What happens to the distribution of Y_n as $n \to \infty$?

$$\lim_{n\to\infty} \mathbb{P}(Y_n \leq t) = \mathbb{P}(Z \leq t);\ Z \sim N(0, 1).$$

The **Central Limit Theorem** tells us that as $n \to \infty$, the distribution of Y_n becomes Normal.

$$\lim_{n\to\infty} \mathbb{P}(\frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \leq t) = \mathbb{P}(Z \leq t).$$