McGill U2S3 Lectures

Hy Vu

Probability

Lecture Notes

Prof: Dr. Sajjad

McGill

McGill Probability Lectures (MATH 323)

Hy Vu

Professor: Dr. Alia Sajjad

August 2024

Overview

Mathematics & Statistics (Sci) : Sample space, events, conditional probability, independence of events, Bayes' Theorem. Basic combinatorial probability, random variables, discrete and continuous univariate and multivariate distributions. Independence of random variables. Inequalities, weak law of large numbers, central limit theorem.

Term: FALL 2024

Prerequisites: MATH 141 (Cal II) or equivalent

Contents

1 Basics of Probability

Definition 1.1. **Probability** is the level of possibility of something happening or being true. It's also defined as a measure to quantify uncertainty or even measure of one's belief in the occurrence of a future event.

Example 1.0.1. Your friend tells you that if you get a head on a flip of a coin you win. What is your probability of winning? Well...If it's a **fair** coin then it's 1/2.

Remark 1.1. *We need to emphasize that it's fair in order to give this probability. We will show this "fairness" in term of mathematics later on.*

Kolmorov was a Russian mathematician, who laid thefoundation of modern probability theory. With this, he developped 3 axioms of probability of which we'll talk about. Nevertheless, we'll first a little on some foundational mathematics for probability, that is, set theory.

1.1 Set Theory

Definition 1.2. A **set** is a collection of distinct objects called **elements**. Usually a set is represented as an upper case letter (e.g. A, B, etc.) while it's elements are lower case letters.

Example 1.1.1. $A = \{a_1, a_2, a_3\}$ is a set.

Remark 1.2. *A set is specified by either: listing its elements (if possible) or by stating a property that its elements have to satisfy.*

Definition 1.3. Let A and B be sets then, the **union** of A is B is defined and written as:

$$A \cup B = \{x : x \in A \text{ or } x \in B\} \tag{1.1}$$

and the **intersection** of A and B is defined as:

$$A \cap B = \{x : x \in A \text{ and } x \in B\} \tag{1.2}$$

Example 1.1.2. Let $A = \{1,2,3\}$, $B = \{1,3,5\}$ and $C = \{2,4,6\}$, then $A \cap B = \{1,3\}$. However, when $B \cap C$, we will see that there's no shared element which means this will create a **empty set** symbolized as Ø.

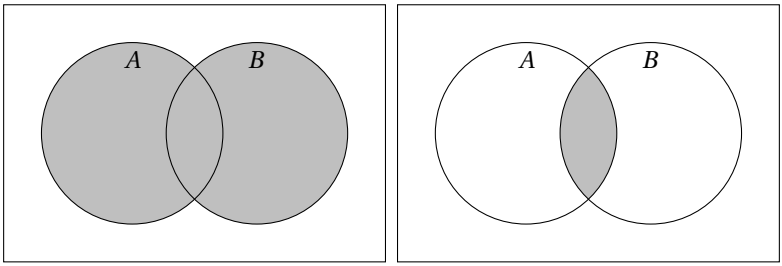


Figure 1.1: Left figure represent set union while the right represent set intersection.

Example 1.1.3. Using the same sets: A, B and C then we can see that $A \cup B = \{1,2,3,5\}$.

Definition 1.4. Let A and B be sets. Then, A is contained or is a **subset** of B if every elements in A is in B. This relation is denoted as

$$A \subset B \tag{1.3}$$

Example 1.1.4. Let $A = \{1,2,3\}, B = \{1,3,5\}, C = \{2,4,6\}$ and $D = \{1,2,3,4,5,6\}$, then $A \subset D, B \subset D$ and $C \subset D$.

Definition 1.5. If a set contain all elements under consideration in a particular context, we call it a **universal set (S)**.[1]

Definition 1.6. If a set consists of no elements, we call it an **empty set** ($\emptyset$).

Definition 1.7. Let $A \subset S$. Then, the **complement** of A, written as A^c or A' is the set of elements in S but not in A. That is,

$$A' = \{a \in S : a \notin A\} \tag{1.4}$$

Example 1.1.5. When rolling a fair dice, the set of all possible outcome with a single throw is $S = \{1,2,3,4,5,6\}$. Let $A = \{1,3,5\}$ then $A' = \{2,4,6\}$.

Remark 1.3. *Sometimes we might observe it's easier to solve $P(A')$ (probability of A'). In which case, we can find that and use the relation between A' and A to deduce $P(A)$.*

[1] Some other notation of universal set can be: U, ξ or $\mathscr{U}$

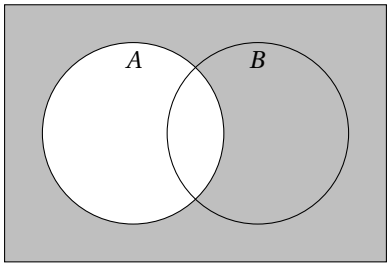


Figure 1.2: Complement of set A

Definition 1.8. When 2 sets A and B are **mutually exclusive**, they have no element in common which means that

$$A \cap B = \emptyset \tag{1.5}$$

Remark 1.4. *Mutually exclusive sets are also known as **disjoint sets**.*

Proposition 1.1. *A and A' are disjoint which means*

$$A \cap A' = \emptyset \tag{1.6}$$

Theorem 1.1. *Let A, B and C be sets then*

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C) \tag{1.7}$$

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C) \tag{1.8}$$

Theorem 1.2. ***(De Morgan's Laws)*** *Let A and B be sets then*

$$(A \cap B)' = A' \cup B' \tag{1.9}$$

$$(A \cup B)' = A' \cap B' \tag{1.10}$$

Now that we got some understanding of set theory, let's establish some foundation for probability.

End of Lecture

1.2 Probability Preliminaries

Definition 1.9. An **experiment** is a process by which an observation is made of which its result is called an **outcome**.

Example 1.2.1. Tossing a coin is an experiment with outcomes of "head" or "tail".

Definition 1.10. A **random experiment** is a process for making an observation(s) whose outcome cannot be predicted with certainty i.e. outcome is random.

Example 1.2.2. When tossing a fair coin, you may know 2 possible outcomes but before the coin land, you do not know which outcome comes up.

Interestingly, one could repeat a random experiment manty times and there's a possibility to observe a different outcome each time.

Definition 1.11. Each repetition of a random experiment is called a **trial** and the outcomes will constitute **events** i.e. an event consists of ≥ 1 outcomes of an experiment.

Definition 1.12. A **simple event** is the outcome of 1 trial while a **compound event** consists of ≥ 2 simple event.

Proposition 1.2. *A compound event can be decomposed into many simple events.*

Example 1.2.3. When tossing a dice many times, the possible simple events are: $E_1 = \{1\}$ (if 1 is observed),..., $E_6 = \{6\}$ (if 6 is observed).
We can create compound events from these simple ones: $E_{\text{even}} = \{2,4,6\}$ (If outcomes are even) and $E_{\text{odd}} = \{1,3,5\}$ (If outcomes are odd).

Definition 1.13. A **sample space (S)** is the set of all possible outcomes of an experiment. [2]

Example 1.2.4. The sample space of tossing a dice is $S = \{1,2,3,4,5,6\}$.

1.2.1 Kolmogorov Axioms

So now that we've got most of a fundamentals, we can revisit the concept we've previously talked about that is the Kolmogorov axioms.

Definition 1.14. An **axiom** is a statement of which we universally assume to be true.

[2] Like universal set, there are other way of notate it. One of the more popular is Ω

Kolmogorov axioms. Consider an experiment with sample space S. To every event $A \subset S$, we associate a number $P(A)$ called **probability of A**, such that the following holds:

1. $P(A) \geq 0$
2. $P(S) = 1$
3. If $E_1, E_2, \ldots$ are events in S s.t. $E_i \cap E_j = \emptyset$ for $i \neq j$, then

$$P(E_1 \cup E_2 \cup \cdots) = P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i) \tag{1.11}$$

From these 3 axioms, we can develop even more important results in probability

Theorem 1.3. *For any event A, $P(A') = 1 - P(A)$.*

Proof. We know from previously that $A \cap A' =$ this would means that $A \cup A' = S$. This means

$$\begin{aligned} P(A \cup A') &= P(S) \\ P(A) + P(A') &= P(S) && \text{(axiom 3)} \\ P(A) + P(A') &= 1 && \text{(axiom 2)} \\ P(A') &= 1 - P(A) \end{aligned}$$

□

Theorem 1.4. $P(\emptyset) = 0$

Proof. We can realize a relation between the null and the universal set that is $S' = \emptyset$. Using theorem 1.3, we'll get

$$\begin{aligned} P(S') &= 1 - P(S) \\ P(\emptyset &= 1 - 1 = 0 \end{aligned}$$

□

Theorem 1.5. *Let A and B then*

$$P(A \cap B') = P(A) - P(A \cap B) \tag{1.12}$$

Proof. Knowing that $A \cap S = A$ and that $A \cap (B \cup B') = (A \cap B) \cup (A \cap B')$ then

$$\begin{aligned} A \cap (B \cup B') &= (A \cap B) \cup (A \cap B') \\ A \cap S &= (A \cap B) \cup (A \cap B') \\ A &= (A \cap B) \cup (A \cap B') \\ P(A) &= P((A \cap B) \cup (A \cap B')) \\ P(A) &= P(A \cap B) + P(A \cap B') \qquad \text{(axiom 3)} \\ \implies P(A \cap B') &= P(A) - P(A \cap B) \end{aligned}$$

□

Theorem 1.6. *If $A \subset B$ then $P(A) \leq P(B)$*

Proof. Since $A \subset B$, we could view B as the "universal set" for A This means $A \cup A' = B$. Then,

$$\begin{aligned} P(B) &= P(A \cup A') \\ &= P(A) + \underbrace{P(A')}_{\geq 0} \qquad \text{(axiom 1 and 3)} \\ \implies P(B) &\geq P(A) \end{aligned}$$

□

Theorem 1.7. *Let A and B be sets then*

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \tag{1.13}$$

Proof. We can realize that $A \cup B$ are actually made up of 3 different sets: $A \cap B'$, $A' \cap B$ and $A \cap B$ which means,

$$\begin{aligned} P(A \cup B) &= P((A \cap B') \cup (A' \cap B) \cup (A \cap B)) \\ &= P(A \cap B') + P(A' \cap B) + P(A \cap B) \qquad \text{(axiom 3)} \\ &= P(A) - P(A \cap B) + P(B) - P(A \cap B) + P(A \cap B) \qquad \text{(Theorem 1.5)} \\ &= P(A) + P(B) - P(A \cap B) \end{aligned}$$

□

End of Lecture

Example 1.2.5. A study of the behaviour of a large number of drug offenders after treatment for drug abuse suggests that the possibility if conviction within two-year period after treatment may depend on the offender's education. Suppose that the probability of the total number of cases found convicted is 0.37 and the probability of the cases with 10 or more years of education is 0.4. Furthermore, the probability that someone will either be convicted or has an education of 10 or more years is 0.67.

Let A = convicted and B = 10 or more years of education. Then, $P(A) = 0.37$ and $P(B) = 0.4$ and $P(A \cup B) = 0.67$.

1. Find the proportion of people who are convicted and have 10 or more years of education
   Answer: This means we're determining $P(A \cap B)$ which is
   $$\begin{aligned} P(A \cap B) &= P(A) + P(B) - PA \cup B) \\ &= 0.37 + 0.4 - 0.67 = \boxed{0.10} \end{aligned}$$
2. What is the probability that someone is convicted and does not have the education of 10 or more years?
   Answer: Essentially, we're finding $P(A \cap B')$ which is
   $$\begin{aligned} P(A \cap B') &= P(A) - P(A \cap B) \\ &= 0.37 - 0.10 = \boxed{0.27} \end{aligned}$$
3. What is the probability that someone is not convicted and has the education of 10 or more years
   Answer: We're trying to determine $P(A' \cap B)$ which is,
   $$\begin{aligned} P(A' \cap B) &= P(B) - P(A \cap B) \\ &= 0.4 - 0.1 = \boxed{0.3} \end{aligned}$$
4. What is the probability that someone is neither convicted nor has the education of 10 years or more?
   Answer: We're trying to find $P(A' \cup B')$ which is,
   $$\begin{aligned} P(A' \cap B') &= P((A \cup B)') \\ &= 1 - P(A \cap B) \\ &= 1 - 0.67 = \boxed{0.33} \end{aligned}$$

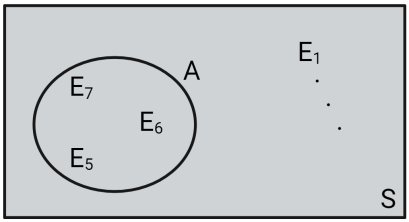
Example 1.2.6. A manufacturer has 5 seemingly identical computer terminals available for shipping. Unknown to her, 2 of the 5 are defective. A particular order calls for 2 of the terminals and is filled by selecting 2 of the terminals.

a. List the sample space
Solution: Defective (D): 2, non-defective (ND): 3, amount of order: 2. This means we have the following: D_1, D_2, ND_2, ND_2 and ND_3. Then, our sample space will be given as

$$S = \{(D_1, D_2), (D_1, ND_1), (D_1, ND_2), (D_1, ND_3), (ND_1, ND_2), (ND_1, ND_3), (ND_2, ND_3), (D_2, ND_1), (D_2, ND_2), (D_2, ND_3)\}$$

b. Let A be the event that the order is filled with 2 non-defective terminals. List the sample space points of A
Solution: $A = \{(ND_1, ND_2), (ND_1, ND_3), (ND_2, ND_4)\}$

c. Construct a Venn diagram illustrating A.
Solution: Suppose that each sample event for A is given as E_5, E_6, E_7 for $(ND_1, ND_2), (ND_1, ND_3), (ND_2, ND_4)$ respectively.



d. Assigns probabilies to simple events in such a way that the information about the experiment is used and Kolmogorov's axioms are met.
Solution: As 2 terminals are chosen randomly out of S, app pairs have the same probabilties of being selected which means that

$$\begin{aligned} E_1 \cup E_2 \cup \cdots \cup E_{10} &= S \\ P(E_1 \cup E_2 \cup \cdots \cup E_{10}) &= P(S) \\ P(E_1) + P(E_2) + \cdots + P(E_{10}) &= 1 \\ 10 \times P(E_i) &= 1 \\ \implies P(E_i) &= \frac{1}{10} \forall i = 1, 2, \ldots, 10 \end{aligned}$$

e. What's $P(A)$?
Solution: $P(A) = \frac{1}{10} + \frac{1}{10} + \frac{1}{10} = \frac{3}{10}$.

Theorem 1.8. *Let S be a finite sample space with N equally likely events and let E be an event be in S. Then,*

$$P(E) = \frac{n}{N} \tag{1.14}$$

where n is the number of outcomes of E i.e. the number of possible ways E can occur, and N is the number of outcomes in S.

Proof. Let's first write E as the union of its simple events as $E = \bigcup_{i=1}^{n} E_i$ then $P(E) = \sum_{i=1}^{n} P(E_i)$. Now, we also know that there are N equally likely events which means $S = \bigcup_{i=1}^{N} E_i$. This means

$$\begin{aligned} P(S) &= \sum_{i=1}^{N} P(E_i) \\ P(S) &= N \cdot P(E_i) \qquad \text{(since events are equally likely.)} \\ 1 &= N \cdot P(E_i) \\ P(E_i) &= \frac{1}{N} \end{aligned}$$

So now, we can use this to find $P(E)$ as

$$P(E) = \sum_{i=1}^{n} P(E_i) = \sum_{i=1}^{n} \frac{1}{N} = \frac{n}{N}$$

$\implies P(E) = \frac{n}{N}$. □

End of Lecture

Example 1.2.7. A balanced coin is tossed three times. What is the probability of observing at least one head? The experiments consists of tossing a coin three times and observing the outcomes. Each simple event will be a sequence of three letters, e.g. HTH. As the coin is balanced all simple events are equally likely.
Solution: Considering all 8 possibilities: $\{HHH, HHT, HTH, THH, THT, TTH, HTT, TTT\}$. So now, the number of sample space is $N = 8$ and the number of outcomes for ≥ 1 head comes up is $n = 7$. Therefore, $P(E) = \frac{7}{8}$.

1.2.2 Counting Rule

Although the above theorem simplifies the problem of finding the probabilities if the events are equally likely and the sample space is finite, finding the number of possible outcomes in a sample space or favourable to an event becomes challenging. In order to address that problem we will study the following counting rules.

Counting Rule 1. Consider 2 sets S_1 and S_2 of which each has n_1 and n_2 elements respectively. Then, the number of ways to form a set from choosing 1 object from S_1 and 1 object from S_2 is

$$n_1 \cdot n_2 \tag{1.15}$$

We can further extend this to k sets. Give k sets: $S_1, S_2, \ldots, S_k$ with number of elements of $n_1, n_2, \ldots, n_k$ respectively. Then, the number of ways to form a set from choosing 1 object from $S_1, S_2, \ldots, S_k$ is

$$n_1 n_2 \cdots n_k \tag{1.16}$$

Example 1.2.8. How many way are there to choose a trouser of 3 different colour with a shirt of 3 different colour? Well, using counting rule 1, we get $3 \times 3 = 9$ ways.

Counting Rule 2. The number of ways to arrange n objects is given as

$$n! = n(n-1)(n-2)\cdots 3 \cdot 2 \cdot 1 \tag{1.17}$$

where $0! = 1! = 1$.

Example 1.2.9. Suppose you have 4 books, how many ways are there to arrange them? Well...using counting rule 2, you get $4! = 4 \times 3 \times 2 \times 1 = 24$ ways to arrange them.

Counting Rule 3. The number of ways to arrange n distinct objects chosen r at a time without replacement such that the order is important is called the **permutation** of n objects taken r at a time. It's given as

$${}^nP_r = {}_nP_r = \frac{n!}{(n-r)!} \tag{1.18}$$

Example 1.2.10. Consider 4 books, what are the possible ways to arrange just 2 out of the 4? Well, using permutation, we will get that $\frac{4!}{(4-2)!} = \frac{24}{2} = 12$ different arrangement.

Counting Rule 4. The number of ways to chose n distinct objects selecting r at a time without replacement such that the order is not important is called the **combination** of n objects taken r at a time. It's given as

$$\binom{n}{r} = \frac{n!}{(n-r)!r!} \tag{1.19}$$

Example 1.2.11. Consider the same 4 books before, now instead of arranging, what are the possible ways to choose 2 out of the 4? Well...using combination, we will get that $\frac{4!}{(4-2)!2!} = \frac{24}{2\times 2} = 6$

The Birthday Problem

Suppose you invite 50 friends to your birthday party. What is the probability that at least two of your friends have the same birthday?

Let's solve this for n friends. Let's ignore all gaps year i.e. only 365 possible birthdays, we first write the sample space of $S = \{(\text{Jan1},\ldots,\text{Jan1}),\ldots,(\text{Dec31},\ldots,\text{Dec31})\}$. This means the total number of points in the sample space is 365^n

Now, if we try to find an event that 2 friends have the same birthdays, we would have to make lots of interesection of events which is quite difficult so we'll change how we approach such problems. If E is the event that 2 friends have the same birthday then E^c is the event that no 2 friends have the same birthday. this means

$$P(E) = 1 - P(E^c) = 1 - \frac{n'}{N}$$

where n' is the number of outcome in E^c and N is the number of outcomes in the sample space.
To determine n' we can use the permutation, for the first friend, you get 365 choices, for the second friends you get 364 etc. which means

$$n' = (365)(364)\cdots(365-(n-1)) = {}^{365}P_n$$

This also means

$$P(E^c) = \frac{n'}{N} = \frac{(365)(364)\cdots(365-(n-1))}{365^n} = \frac{{}^{365}P_n}{365^n}$$

End of Lecture

Now, let $n = 2$, we will get that

$$\begin{aligned} P(E) = 1 - P(E^c) &= 1 - \frac{^{365}P_2}{365^2} \\ &= 1 - \frac{365!}{(365-2)!(365)^2} \\ &= 1 - \frac{364 \cdot 365}{365^2} \\ &= 1 - \frac{364}{365} = \frac{1}{365} \end{aligned}$$

This makes sense because the probability that within 2 friends and they have the same birthday is only 1 out of 365 possible choices. So now, we can answer the original question of 50 friends which is $n = 50$ thus

$$P(E) = 1 - \frac{^{365}P_{50}}{365^{50}} \approx 0.9704$$

$\implies$ There's a 97% chance that within a group of friends of 50 people, there are 2 person with the same birthday.

Let's look at another problem.

Example 1.2.12. A student taking an exam is directed to answer 7 out of 10 questions. What is the probability that the student chooses two questions from the first five and attempts all last the five?

Solution: First, let E be the even of that the student chose the 2 questions from the first five and attempts the last five. Now, let n be the number of outcomes in E which can be given as

$$n = \binom{5}{2}\binom{5}{5}$$

This is because you've partition the questions into 2 parts, the first five will be only chosen 2 while the last five is all chosen. Now, let N be the total number of outcomes; then, we will choose 7 questions out of 10 which means

$$N = \binom{10}{7}$$

Finally, you can tell that the for a random trial, the probability of selecting any of these questions are equally likely which means that

$$P(E) = \frac{n}{N} = \frac{\binom{5}{2}\binom{5}{5}}{\binom{10}{7}} = \frac{10}{120} = \frac{1}{12} \approx 0.083$$

$\implies$ There's a 8.3% chance that a student will select 2 questions from the first five while complete the last five.

We can now generalize this question further to give:

Generlization: Let M be the total number of questions and only m amount is selected. A student is directed to choose x amount from the first a and $m-x$ from the last $M-a$.

$$\underbrace{1,2,\ldots,a}_{x},\underbrace{a+1,\ldots,M}_{m-x}$$

Let E be the event corresponding to said direction above. Then,

$$P(E) = \binom{a}{x}\binom{M-a}{m-x}\binom{M}{m}^{-1} \tag{1.20}$$

Let's look at more of other examples.

Example 1.2.13. Suppose that there are N fish in a lake of which a are tagged and $N-a$ are untagged. You choose a random sample of n fish without replacement. What is the probability of getting x tagged fish in your sample?

Solution: Let E be the event where there's x tagged fish in the sample. Now, for any trial, the probability of picking 1 fish would be the same as any other which means that $P(E) = \frac{m}{M}$ where m is the number of outcomes in E while M is the total number of outcomes. We can begin to define them

- The question ask for x amount of tagged fished which means we'll be choosing x from a. Additionally, the sample of fish is n which means the amount of untagged fish is $n-x$ i.e. choosing $n-x$ from $N-a$. Thus,

$$m = \binom{a}{x}\binom{N-a}{n-x}$$

- The total amount of outcome is simply choosing n fish from the total population of N. Thus,

$$M = \binom{N}{n}$$

So now, the probability of getting x tagged fish (consequently also $n-x$ untagged fish) in the sample is

$$P(E) = \binom{a}{x}\binom{N-a}{n-x}\binom{N}{n}^{-1}$$

1.2.3 Practice Problems

Problem 1. An airline has 6 flights from New York to California and 7 flights from from California to Hawaii. If the flights are to be made on separate days, how many different flight arrangement can the airline offer from New York to Hawaii?

Solution: Since the flights are made on separate days, the first day there would be 6 possible flight from NY to CA while on the second day, there would be 7 possible flight from CA to HI. Thus, $6 \times 7 = 42$ possible flight arrangment from NY to HI.

Problem 2. A personnel director for a corporation has hired 10 new engineers. If 3 (distinctly different) positions are open at a Cleveland plant, in how many ways can she fill the positions?

Solution: Since the 3 positions are distinctly different $\Longrightarrow$ the order matter. Now, from the 10 engineer, once you pick 1 engineer the total engineer that can be picked for the next position decreased to 9. With this, we can use permutation to calculate,

$${}^{10}P_3 = \frac{10!}{(10-3)!} = \frac{10!}{7!} = 720$$

$\Longrightarrow$ There are 720 different ways she can fill the positions.

Problem 3. An assembly operation in a manufacturing plant requires three steps that can be performed in any sequence. How many different ways can the assembly be performed?

End of Lecture

Solution: Since there are 3 steps and the order does not matter thus we can use factorial. Thus, 3! = 6 differemt ways that the operation can be performed.

1.3 Conditional Probability

Example 1.3.1. Consider a box of eight identical balls - 2 red and six green. You choose 2 balls at random (one-by-one) without replacement. What is the probability that the second ball is red?

Solution: Well to know this, you need to know what the first ball is e.g. suppose that the first ball is green, then the probability of second red is 2/7. On the other hand, if the first ball is red, $P(\text{second red}) = 1/7$. This is because you're taking the ball out without replacement.

As you can see from above, the probability of the second red ball changes depending on the condition of the first ball. To address this we introduce the conditional probability.

Definition 1.15. Let A and B be 2 events, such that $P(A) \neq 0$. Then,

$$P(B \text{ given that } A \text{ has occured}) = \frac{P(A \cap B)}{P(A)} \tag{1.21}$$

Another way to write the above equation is

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)} \tag{1.22}$$

Remark 1.5. *$B \mid A$ is not a new event, it's the same event in context of more knowledge.*

Example 1.3.2. Suppose you can compute the probability of rain on a particular day by finding the fraction of days in which rain occured throughout history (Unconditional probability). This probability will change if let's say we also include atmospheric pressure in the context (conditional probability)

Example 1.3.3. Consider tossing a fair die one. Define the following events:

- A: an even number is observed.
- B: 2 is observed.

Then, $P(B|A) = \frac{P(A\cap B)}{P(A)} = \frac{1/6}{1/2} = \frac{1}{3}$

Now we need to check that it follows Kolmorov's axiom.

1. $P(B \mid A) \geq 0$

   Proof. By axiom 1, we have that $P(A\cap B) \geq 0$ and $P(A) > 0 \implies \frac{P(A\cap B)}{P(A)} \geq 0$. □

2. $P(S \mid A) = 1$

   Proof.
   $$P(S \mid A) = \frac{P(S \cap A)}{P(A)} = \frac{P(A)}{P(A)} = 1$$
   $\therefore P(S \mid A) = 1$ □

3. $P\left(\bigcup_{i=1}^{\infty} B_i \mid A\right) = \sum_{i=1}^{\infty} P(B_i \mid A)$ where $B_i \cap B_j = \emptyset$ for some $i \neq j$.

   Proof. Knowing that $B_i \cap B_j = \emptyset$ for some $i \neq j$, we can tell that B_i are disjoints. Then,
   $$\begin{aligned} P\left(\bigcup_{i=1}^{\infty} B_i \mid A\right) &= \frac{P\left((B_1 \cup B_2 \cup \cdots) \cap A\right)}{P(A)} \\ &= \frac{P\left((B_1 \cap A) \cup (B_2 \cap A) \cup \cdots\right)}{P(A)} \\ &= \frac{P(B_1 \cap A) + P(B_2 \cap A) + \cdots}{P(A)} \\ &= \frac{P(B_1 \cap A)}{P(A)} + \frac{P(B_2 \cap A)}{P(A)} + \cdots \\ &= \sum_{i=1}^{\infty} P(B_i \mid A) \end{aligned}$$
   Thus $P\left(\bigcup_{i=1}^{\infty} B_i \mid A\right) = \sum_{i=1}^{\infty} P(B_i \mid A)$ □

Example 1.3.4. Suppose that a research funding organization reviews there search proposal at two different stages. The probability that a proposal passes the first stage is 0.2. Of those research proposals that pass stage 1, 80% pass stage two as well. What is the probability that a research proposal

passes both stages?

Solution: Let A be stage 1 and B be stage 2. Essentially, you're trying to find $P(A \cap B)$. Now, we know that that $P(A) = 0.2$. For B, we're given a probability knowing that the proposal passed first stage $\implies P(B \mid A) = 80\% = 0.8$. Thus,

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)} \iff P(A \cap B) = P(B \mid A)P(A) = (0.2)(0.8) = 0.16$$

Theorem 1.9. ***(Multiplication Rule).*** *From the definition of conditional probability,* [3]

$$P(A \cap B) = P(B \mid A)P(A) = P(A \mid B)P(B) \tag{1.23}$$

Remark 1.6. *In word problems, the phrase "of those that" typically implies conditionally probability.*

Theorem 1.10. ***(Exending Multiplication Rule).*** *Let A_1, A_2, A_3 be three events. Then,*

$$P(A_1 \cap A_2 \cap A_3) = P(A_1) \cdot P(A_2 \mid A_1) \cdot P(A_3 \mid A_1 \cap A_2) \tag{1.24}$$

Similarly, Let $A_1, A_2, \ldots, A_n$ be a sequence of n events. Then,

$$P\left(\bigcap_{i=1}^{n} A_i\right) = P(A_1) \cdot P(A_2 \mid A_1) \cdot P(A_3 \mid A_1 \cap A_2) \cdot \ldots \cdot P(A_n \mid A_1 \cap A_2 \cap \cdots \cap A_{n-1}) \tag{1.25}$$

Proof. We first solve for the conditional probability of 3 events. By the definition of conditional probability, we will get that

$$\begin{aligned} P(A_1 \cap A_2 \cap A_3) &= P(A_3 \mid A_1 \cap A_2)P(A_1 \cap A_2) \\ &= P(A_3 \mid A_1 \cap A_2) \cdot P(A_2 \mid A_1) \cdot P(A_1) \end{aligned}$$

Now, we will look at the extension to n events.

$$\begin{aligned} P\left(\bigcap_{i=1}^{n} A_i\right) &= P(A_1 \cap A_2 \cap \cdots \cap A_n) \\ &= P(A_n \mid A_1 \cap A_2 \cap \cdots \cap A_{n-1}) \cdot P(A_1 \cap \cdots \cap A_{n-1}) \end{aligned}$$

[3] Proof wouldn't be necessary since this is just your everyday algebraic manipulation.

End of Lecture

$$= P(A_n \mid A_1 \cap A_2 \cap \cdots \cap A_{n-1}) \cdot P(A_{n-1} \mid A_1 \cap A_2 \cap \cdots \cap A_{n-2}) \cdot P(A_1 \cap \cdots \cap A_{n-1})$$

$$\vdots$$

$$= P(A_1) \cdot P(A_2 \mid A_1) \cdot P(A_3 \mid A_1 \cap A_2) \cdot \ldots \cdot P(A_n \mid A_1 \cap A_2 \cap \cdots \cap A_{n-1})$$

□

Example 1.3.5. Consider the "identical balls in the box" problem, we have 2 red and six green balls. Draw five balls at random, without replacing. What is the probability that the first ball is green, the second is red, the third is green, the fourth is green and the fifth 5th is red?

Solution: Let us define the following events:

- G_1: first ball is green.
- R_2: second ball is red.
- G_3: third ball is green.
- G_4: fourth ball is green.
- R_5: fifth ball is red.

We're trying to find $P(G_1 \cap R_2 \cap G_3 \cap G_4 \cap R_5)$. There are 2 ways to approach this, you could solve this logically by deducing that when you pick 1 ball, the sample space decrease by 1 for the next choice e.g. if pick a green ball, the entire sample space decrease by 7 for the next choice while the sample space of specifically green ball will decrease to 5 if the next choice is green. Thus,

$$P(G_1 \cap R_2 \cap G_3 \cap G_4 \cap R_5) = \frac{6}{8} \cdot \frac{2}{7} \cdot \frac{5}{6} \cdot \frac{4}{5} \cdot \frac{1}{4} = \frac{1}{28}$$

Another way is using the above multiplication rule of conditional probability.

$$\begin{aligned} P(G_1 \cap R_2 \cap G_3 \cap G_4 \cap R_5) &= P(G_1) \cdot P(R_2 \mid G_1) \cdot P(G_3 \mid G_1 \cap G_2) \\ &\quad \cdot P(G_4 \mid G_1 \cap G_2 \cap G_3) \cdot P(G_5 \mid G_1 \cap G_2 \cap G_3 \cap G_4) \\ &= \frac{6}{8} \cdot \frac{2}{7} \cdot \frac{5}{6} \cdot \frac{4}{5} \cdot \frac{1}{4} = \frac{1}{28} \end{aligned}$$

1.3.1 Special Cases of Conditional Probability

We will now look at 3 special cases in conditional probability.

1. If A and B are disjoints events i.e. $A \cap B = \emptyset$. Then,

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)} = \frac{0}{P(A)} = 0 \tag{1.26}$$

2. Let A and B be events such that $A \subseteq B$. Then,

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)} = \frac{P(A)}{P(A)} = 1 \tag{1.27}$$

3. For some integer k, let the subsets $B_1, B_2, \ldots, B_k$ be such that

   i. $S = B_1 \cup B_2 \cup \cdots \cup B_k$.

   ii. $B_i \cap B_j = \emptyset$ for some $i \neq j$

   Then, the collection of sets $\{B_1, B_2, \ldots, B_k\}$ is said to be a **partition** of S.
   Additionally if $A \subseteq S$ and $\{B_1, B_2, \ldots, B_k\}$ is a partition of S. Then

$$A = A \cap S = A \cap (B_1 \cup B_2 \cup \cdots \cup B_k) = \bigcup_{i=1}^{k} A \cap B_i \tag{1.28}$$

   where $(A \cap B_i) \cap (A \cap B_j) = \emptyset$ for some $i \neq j$.

Theorem 1.11. ***(Law of Total Probability).*** *Let $\{B_1, B_2 \ldots, B_k\}$ be a partition of S such that $P(B_i) > 0$ where $i = 1, 2, \ldots, k$. Then, for any event A in S,*

$$P(A) = \sum_{i=1}^{k} P(A \mid B_i) P(B_i) \tag{1.29}$$

Proof. Remember that we can represent A as the union of intersection of A with the partition. Then,

$$\begin{aligned} A &= \bigcup_{i=1}^{k} A \cap B_i \\ P(A) &= P\left(\bigcup_{i=1}^{k} A \cap B_i \right) \\ &= P\left((A \cap B_1) \cup (A \cap B_2) \cup \cdots \cup (A \cap B_k) \right) \end{aligned}$$

$$
\begin{aligned}
&= P(A \cap B_1) + P(A \cap B_2) + \cdots + P(A \cap B_k) \\
&= P(A \mid B_1)P(B_1) + P(A \mid B_2)P(B_2) + \cdots + P(A \mid B_k)P(B_k) \\
&= \sum_{i=1}^{k} P(A \mid B_i)P(B_i)
\end{aligned}
$$

□

Example 1.3.6. Consider "Identical Balls in a Box Problem". In this example we will use a different setting. Assume that there are three identical boxes with 10 balls each. Box 1 has 7 red and 3 green balls, box 2 has 6 red and 4 green balls, and box 3 has 5 red and 5 green balls. A box is chosen at random and a ball is picked. What is the probability that the ball is red?

Solution: Let us define the following events:

- R: Choosing a red ball.
- B_1: Ball chosen from box 1.
- B_2: Ball chosen from box 2.
- B_3: Ball chosen from box 3.

Notice that the probability of ball chosen from 3 box are equall thus $P(B_i) = \frac{1}{3}$ for $i = 1,2,3$. We can thus use the law of total probability to get

$$
\begin{aligned}
R &= R \cap S \\
R &= R \cap (B_1 \cup B_2 \cup B_2) \\
R &= (R \cap B_1) \cup (R \cap B_2) \cup (R \cap B_3) \\
P(R) &= P((R \cap B_1) \cup (R \cap B_2) \cup (R \cap B_3)) \\
&= P(R \cap B_1) + P(R \cap B_2) + P(R \cap B_3) \\
&= P(R \mid B_1)P(B_1) + P(R \mid B_2)P(B_2) + P(R \mid B_3)P(B_3) \\
&= \frac{7}{10} \cdot \frac{1}{3} + \frac{6}{10} \cdot \frac{1}{3} + \frac{5}{10} \cdot \frac{1}{3} = \frac{18}{30}
\end{aligned}
$$

Now, continuing from the above question. Suppose that we observe the chosen ball is red. What is the probability that box 1 was chosen? That is $P(B_1 \mid R)$?

Solution: We can use the definition of conditional probability, that is

$$
P(B_1 \mid R) = \frac{P(A \cup B)}{P(R)}
$$

$$= \frac{P(R \mid B_1)P(B_1)}{P(R)}$$
$$= \frac{\frac{7}{10} \cdot \frac{1}{3}}{\frac{18}{30}} = \frac{7}{10} \cdot \frac{1}{3} \cdot \frac{30}{18} = \frac{7}{18}$$

1.3.2 Bayes' Theorem

Theorem 1.12. ***(Bayes' Theorem).*** *Let $\{B_1, B_2, \ldots, B_k\}$ be a partition of S such that $P(B_i) > 0$ for some $i = 1, 2, \ldots, k$. Then, for any event $A \subseteq S$,*

$$P(B_i \mid A) = \frac{P(A \mid B_i)P(B_i)}{\sum_{i=1}^{k} P(A \mid B_i)P(B_i)} \tag{1.30}$$

Note: In word problems, if you were asked to find the reverse of a conditional probability result, it's a Bayes' theorem problem. Contrarily, if you were given some conditional probabilities $P(A \mid B_i)$ along with some unconditional $P(B_i)$, and you're required to find $P(A)$, it's a total probability problem.

Proof. (of Bayes' Theorem). Knowing that A can be written as the union of $A \cap B_i$ for some $i = 1, 2, \ldots, k$ (from above), we can use the law of total probability. Then,

$$\begin{aligned} P(B_i \mid A) &= \frac{P(A \cap B_i)}{P(A)} \\ &= \frac{P(A \mid B_i)P(B_i)}{P(A)} && \text{by the multiplication rule} \\ &= \frac{P(A \mid B_i)P(B_i)}{\sum_{i=1}^{k} P(A \mid B_i)P(B_i)} && \text{by the law of total probability} \end{aligned}$$

Thus $P(B_i \mid A) = \dfrac{P(A \mid B_i)P(B_i)}{\sum_{i=1}^{k} P(A \mid B_i)P(B_i)}$ □

Example 1.3.7. A diagnostic test for a disease is such that it (correctly) detects the disease in 90% of the individuals who actually have the disease. Also, if a person does not have the disease, the test will report that he or she does not have it with probability 0.9.

Only 1% of the population has the disease in question. If a person is chosen at random from the population and the diagnostic test indicates that they have the disease, what is the conditional probability that the person does, in fact,has the disease?

Solution: let us define the following events:

End of Lecture

- pos: Positive for the diagnostic test (the test detect the disease) $\implies$ pos^c : negative for the diagonistic test.
- D: People have the disease $\implies$ D^c : people do not have the disease.

What we're trying to find is $P\left(\mathrm{D} \mid \mathrm{pos}\right)$. Then,

$$\begin{aligned} P\left(\mathrm{D} \mid \mathrm{pos}\right) &= \frac{P(\mathrm{pos} \mid \mathrm{D})P(\mathrm{D})}{P(\mathrm{pos})} \\ &= \frac{P(\mathrm{pos} \mid \mathrm{D})P(\mathrm{D})}{P(\mathrm{pos} \mid \mathrm{D})P(\mathrm{D}) + P(\mathrm{pos} \mid \mathrm{D}^c)P(\mathrm{D}^c)} \\ &= \frac{0.9(0.1)}{(0.9)(0.1) + (1-0.9)(1-0.1)} = \frac{1}{12} \end{aligned}$$

Definition 1.16. The probability that the test detects the disease given that the patient has the disease is called the **sensitivity of the test** ($P(\mathrm{pos} \mid \mathrm{D})$). Meanwhile, the probability that the test indicates no disease given that the patient is diseae free is called the **specificity of the test** ($P(\mathrm{pos}^c \mid \mathrm{D}^c)$).

Definition 1.17. The **positive predictive value** ($P(\mathrm{D} \mid \mathrm{pos})$) of the test is the probability that the patient has the disease given that the test indicates that the disease is present.

Example 1.3.8. The sensitivity of diagnostic tests for COVID-19 could be as low as 0.7. When COVID-19 is likely, a single negative test should not rule it out Canadian health network

Example 1.3.9. This study looked at 731 people tested at a community testing site in San Francisco during the Omicron wave in January 2022. Study participants were tested using anterior nares swabs on the AbbottBinaxNOW assay versus RT-PCR. Overall prevalence was 40.5%, and antigen test sensitivity was 0.65 overall.

Now assume that the specificity of the test is 0.95. What is the positive predictive value of this test?

Solution: Essentially, we're trying to find $P(\mathrm{D} \mid \mathrm{pos})$. We're given the following:

- $P(\mathrm{pos} \mid D) = 0.65$.
- $P(\mathrm{pos}^c \mid D^c) = 0.95$
- $P(D) = 0.405$.

Then,

$$P(D \mid \text{pos}) = \frac{P(\text{pos} \mid D)P(D)}{P(\text{pos})} = \frac{P(\text{pos} \mid D)P(D)}{P(\text{pos} \mid D)P(D) + P(\text{pos} \mid D^c)P(D^c)}$$
$$= \frac{P(\text{pos} \mid D)P(D)}{P(\text{pos} \mid D)P(D) + [1 - P(\text{pos}^c \mid D^c)][1 - P(D)]}$$
$$= \frac{(0.65)(0.405)}{(0.65)(0.405) + (1 - 0.95)(1 - 0.405)} \approx 0.89$$

You might ask yourself, why is the predictive value so high even though the sensitivity to low? Well...it's because the prevalence of the disease is high (compared to previous example where prevalence is only 1%, predictive value is low).

1.4 Independence

Now, sometimes, there are instances where information that an event has occurred does not affect the probability of the other event i.e. Knowing that A occurred does not change $P(B)$.

Example 1.4.1. Consider our Balls in a Box problem. You draw a ball from the box containing 2 red and 6 green identical balls. This first ball is red. Before you draw the second ball, you put this ball back. Now your box has two red and six green balls again (as you have put the ball back, that you had drawn). What is the probability that the second ball drawn is red given that the first ball was red?

Solution: Let R_1 be the even of drawing the first ball is red and R_2 be drawing the second one is red. Then,

$$P(R_2 \mid R_1) = \frac{2}{8}$$

This is because we put back the ball for the second drawing $\implies$ the sample size will not decrease.

Remark 1.7. *We call this kind of experimentation:* ***sampling with replacement.***

Example 1.4.2. Consider rolling a fair die twice. All six outcomes are equally likely under the assumption of fairness. What is the probability of having two 6?

Solution: There's not reduction of sample size nor rolling 1 side will affect the other in the next run. Let A be the event of having 6 in the first throw and B for having 6 in the second, Then,

$$P(A \cap B) = P(B \mid A) \cdot P(A) = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36}$$

Definition 1.18. Let A and B be two events. A and B are said to be **independent** if and only if

$$P(B \mid A) = P(B) \tag{1.31}$$

If A is independent of B, we denote it as $A \perp B$. If $A \perp B \implies B \perp A$.

Theorem 1.13. *2 events A and B are independent if and only if*

$$P(A \cap B) = P(A) \cdot P(B) \tag{1.32}$$

Proof. Because this is an "if and only if" statement, we need to prove it in both direction.
($\implies$) Let $A \perp B$. Then,

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

$\therefore (A \cap B) = P(A) \cdot P(B)$
($\impliedby$) Let $A \perp B$. Then,

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

$\therefore P(A) \cdot P(B) = (A \cap B)$. □

Definition 1.19. The set of events $\{A_1, A_2, \ldots, A_n\}$ are said to be **mutually independent** if for any subset $\{A_1, A_2, \ldots, A_k\}$ of these events,

$$P\left(\bigcap_{i=1}^{k} A_i\right) = \prod_{i=1}^{k} P(A_i) \tag{1.33}$$

In other words, $P(A_1 \cap \cdots \cap A_k) = P(A_1) \cdots P(A_k)$.

Definition 1.20. The set of events $\{A_1, A_2, \ldots, A_n\}$ is said to be **pairwise independent** if for any 2 events A_i and A_j

$$P(A_i \cap A_j) = P(A_i) \cdot P(A_j) \tag{1.34}$$

for some $i \neq j$

Question. Does pairwise independence implies mutual independence?
Answer. No not necessarily.

Example 1.4.3. Consider flipping a fair coin. Let H be the event where the coin lands head while T where the coin lands tail. Define S to be the event where after you flip 2 times have a head and tail (regardless of order). Then, $P(H) = P(T) = P(S) = \frac{1}{2}$. We can see that $P(H \cap T) = P(H)P(T) = \frac{1}{4}$ and similarly $P(H \cap S) = P(H)P(S) = \frac{1}{2}$ and similar for $P(T \cap S) = \frac{1}{2}$. Thus H, T, S are pairwise independent. However, if we look at $P(H \cap T \cap S) = \frac{1}{2} \cdot \frac{1}{2} \cdot 1 = \frac{1}{4} \neq \left(\frac{1}{2}\right)^3 = \frac{1}{8} \implies$ These events are not mutually independent.

Remark 1.8. *Essentially, the equation (1.33) does not imply that the events $A_1, A_2, \ldots, A_k$ are mutually independent.*

Proposition 1.3. *If $B \perp A$, then the followings are true:*

1. $A \perp B^c$
2. $B \perp A^c$
3. $A^c \perp B^c$

Proof. Suppose that $A \perp B$. Then,

1. We will show that $P(A \cap B^c) = P(B^c)P(A)$ to show independency.

$$\begin{aligned} P(A \cap B^c) &= P(B^c \mid A)P(A) \\ &= (1 - P(B \mid A))P(A) \\ &= (1 - P(B))P(A) = P(B^c)P(A) \end{aligned}$$

   $\implies A \perp B^c$.

2. The proof will follow the same as above but now it's the complement of A instead of B.

3. We will show that $P(A^c \cap B^c) = P(A^c)P(B^c)$. Then,

$$\begin{aligned} P(A^c \cap B^c) &= P(B^c \mid A^c)P(A^c) \\ &= [1 - P(B \mid A^c)]P(A^c) \\ &= [1 - P(B)]P(A^c) \qquad \text{by property 2 above.} \\ &= P(B^c)P(A^c) \end{aligned}$$

   $\implies A^c \perp B^c$

End of Lecture

Thus we've proven the above proposition. □

Example 1.4.4. Toss a fair coin repeatedly until you observe the first head at which point you stop. Let A be the event that a head occurs at sixth toss. What is $P(A)$?

Solution: We can see that the event A can be given as $T \cap T \cap T \cap T \cap T \cap H$. Knowing that this is a fair coin and the the tosses are independent from each other.

$$\begin{aligned} P(A) &= P(T \cap T \cap T \cap T \cap T \cap H) \\ &= P(T)P(T)P(T)P(T)P(T)P(H) \\ &= \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{64} \end{aligned}$$

1.4.1 Independence and Mutually Exclusive

Consider 2 mutually exclusive events A and B. By the definition alone, we can tell that A and B has nothing in common and they cannot occur simultaneously. **Are A and B independent?** Well...no.

Theorem 1.14. *Let A and B be two events, with $P(A)$ and $P(B) \neq 0$. If A and B are mutually exclusive, then they're not independent.*

Proof. If A and B are mutually exclusive then $P(A \cap B) = P(\emptyset) = 0 \neq P(A)P(B)$. Thus, they're not independent. □

Theorem 1.15. *Two mutually exclusive events A and B are independent if and only if either $P(A)$ or $P(B) = 0$.*

Proof. Because A and B are mutually exclusive, $P(A \cap B) = 0$. For the events to be independent $P(A \cap B) = P(A)P(B)$. Thus $P(B)P(A) = 0 \implies$ either $P(A)$ or $P(B) = 0$. □

So now, with all of what we've learn, when you're finding the probability of the union of 2 events A and B, you need to identify whether they're mutually exclusive, independent or dependent (conditional).

1. Mutually exclusive:

$$P(A \cup B) = P(A) + P(B) + \underbrace{P(A \cap B)}_{=0} = P(A) + P(B)$$

2. Independent:

$$P(A \cup B) = P(A) + P(B) + \underbrace{P(A \cap B)}_{=P(A)P(B)} = P(A) + P(B) + P(A)P(B)$$

3. Conditional:

$$P(A \cup B) = P(A) + P(B) + \underbrace{P(A \cap B)}_{=P(A|B)P(B)} = P(A) + P(B) + P(A \mid B)P(B)$$

Or

$$P(A \cup B) = P(A) + P(B) + \underbrace{P(A \cap B)}_{=P(B|A)P(A)} = P(A) + P(B) + P(B \mid B)P(A)$$

Example 1.4.5. Consider the union of several independent events $A_1, A_2, \ldots, A_n$. Then,

$$\begin{aligned} P(A_1 \cup A_2 \cup \cdots \cup A_n) &= 1 - P(A_1 \cup A_2 \cup \cdots \cup A_n)^c \\ &= 1 - P(A_1^c \cap A_2^c \cap \cdots \cap A_n^c) \\ &= 1 - [(1 - P(A_1)) \cdots (1 - P(A_n))] \\ &= 1 - \prod_{i=1}^{k} 1 - P(A_i) \end{aligned}$$

1.4.2 Independence and Sampling

We've also seen that when sampling in different ways, the dependency of events varies:

- Sampling with replacement: The outcomes will be independent.
- Sampling without replacement: The outcomes will be dependent.
- Sampling without replacement ($n << N$): When the sample size is small as compared to the actual population, independence can be assumed.

Example 1.4.6. Suppose that in very large city 20% of people have a certain genetic mutation. If 10 people are examined.

a. What is the probability that exactly 2 have the mutation?

b. What is the probability that at least 2 will have the mutation?

Solution: We can first assume independence. Let us defined M_i to be the event where the i^{th} individual have the genetic mutation for some $i = 1,2,\ldots,10$. Let X be the number people that have the mutation. Then,

a. To find the probability of having exactly 2 mutations, we first find the probability of a specific configuration (in this case person 1 and 2 is mutated) that result in 2 mutated and 8 non-mutated. Then, we will add up all the possibility of other kinds of configuration (e.g. person 1 and person 3, person 4 and person 5, etc.). Define E to the event of having the first and second person having the mutation. Then,

$$\begin{aligned} P(E) &= P(M_1)P(M_2)P(M_3^c)\cdots P(M_{10}^c) \\ &= 0.2\cdot 0.2\cdot 0.8\cdot \ldots \cdot 0.8 \\ &= (0.2)^2(0.8)^8 \end{aligned}$$

Now, the probability of having exactly 2 person that have the mutation would be: [4]

$$P(X=2) = \binom{10}{2} P(E) = \binom{10}{2}(0.2)^2(0.8)^8$$

b. Now that we got the standard form from A, we can solve find at least 2 people having the mutation through the following:

$$\begin{aligned} P(X \geq 2) &= 1 - P(X<2) \\ &= 1 - [P(X=0) + P(x=1)] \\ &= 1 - \left[\binom{10}{0}(0.2)^0(0.8)^{10} + \binom{10}{1}(0.2)^1(0.8)^9\right] \\ &= 1 - \binom{10}{0}(0.2)^0(0.8)^{10} - \binom{10}{1}(0.2)^1(0.8)^9 \end{aligned}$$

Or another way to do this is

$$\begin{aligned} P(X \geq 2) &= P(X=2) + P(X=3) + \cdots + P(X=10) \\ &= \binom{10}{2}(0.2)^2(0.8)^8 + \binom{10}{3}(0.2)^3(0.8)^7 + \cdots + \binom{10}{10}(0.2)^{10}(0.8)^0 \\ &= \sum_{i=2}^{10}\binom{10}{k}(0.2)^i(0.8)^{10-i} \end{aligned}$$

[4] We used $\binom{10}{2}$ because we're getting all the configuration of choosing 2 from 10 people.

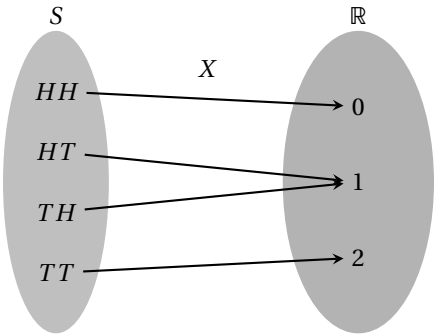
This perfectly leads us to the concept of the next chapter and that is random variable.

2 Random Variables and Distributions

We've previously learned that a variable is just a symbol that can take on any specific value e.g. $x = 1$ or $a = 100$. But in the formal context of probability, its meaning would be slightly different.

Definition 2.1. A **random variable** is a real-valued function defined over a sample space i.e. it's a function that would map events/outcomes in the sample space onto a real number. [1]

Example 2.0.1. Toss a fair coin twice. this will yield the sample space $S = \{HH, HT, TH, TT\}$. We define the random variable X as the number of tails from the coin toss. Then, we can see the following mapping:



Remark 2.1. *The upper case X represents the random variable while the lower case x represents a particular value that X takes.*

Example 2.0.2. Using the same example as above, we have X as the random variable while $x = 0, 1, 2$.

Definition 2.2. A **probability distribution** is a function or table that represent the value that a random variable can take and its likelyhood (probability).

[1] It assigns each events an real value.

End of Lecture

Example 2.0.3. Consider rolling 2 fair dice. The sample space is $S = \{s = (i, j) : 1 \leq i \leq 6, 1 \leq j \leq 6\}$, where each of the 36 points in S is assigned equal probability of $p(s) = 1/36$. The random variable X can be defined as the sum of the values on the 2 dice, i.e.,

$$X(s) = X((i, j)) = i + j$$

We can first define the set possible value of random variable X as $R_X = \{2, 3, \ldots, 12\}$, we call R_X the *range* or *support*. With R_x, we can construct a table of probability distribution of X as

x	$P(X = x)$	Sample points
2	1/36	(1,1)
3	2/36	(1,2), (2,1)
4	3/36	(2,2), (3,1), (1,3)
5	4/36	(1,4), (4,1), (2,3), (3,2)
⋮	⋮	⋮
12	1/36	(6,6)

Here's more example like above:

1. Throw a fair die 10 times. Let X be the number of sixes in 10 throws. The possible values of x are $R_X = \{0, 1, 2, \ldots, 10\}$.

2. Throw a fair die until a six occurs. Let Y be the number of throws until the first six appears. Then $R_Y = \{1, 2, 3, 4, \ldots\}$.

3. Let Z be the time between 2 eruptions of Italy's Stromboli volcano. Then $R_Z = [0, \infty[$.

We can see from the above example that certain random variable are define in with a definite value e.g. sum of dice, number of dice etc. However, there are random variables that have to defined as intervals e.g. time, temperature, height etc.

This relates directly how random variables are segregated: discrete and continuous random variables.

2.1 Discrete Random Variable

Definition 2.3. A **discrete random variable** are random variable with range that is finite or at least countably infinite.

Example 2.1.1. Let X be a discrete random variable with range $R_X = \{x_1, x_2, \ldots\}$. Then, $x_1, x_2, \ldots$ are the possible values that X can take.
We can represent its probability distribution $P(X = x_i)$ (with $i = 1, 2, \ldots$) in the form of a function, table or even a graph. If it's a function, we tend to write it as $p(x) = P(X = x) \forall x$ that X can take.

2.1.1 Cumulative Distribution Function

A method to describe the probability distribution of a discrete random variable is through the *cumulative distribution function*.

Definition 2.4. The **cumulative distribution function (CDF)** of a discrete random variable X with domain R is defined as

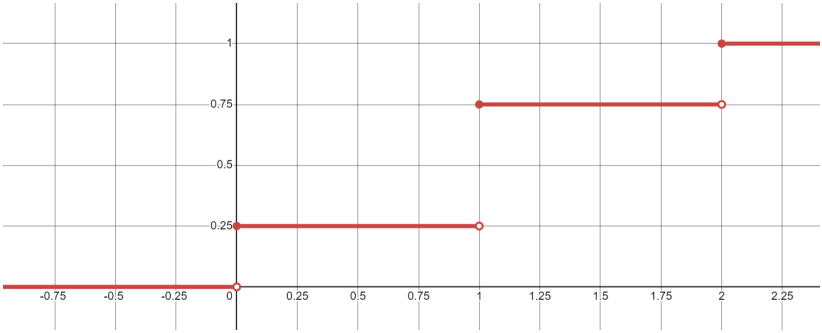
$$F_X(x) = P(X \leq x), \forall x \in R \tag{2.1}$$

The domain of $F_X(x)$ is $\mathbb{R}$.

Example 2.1.2. Toss two fair coins. The sample space S is $\{HH, HT, TH, TT\}$. The random variable X can be defined as the total number tails that can be observed. Then, the range or the support of X is $R_X = \{0, 1, 2\}$. We can thus define the CDF of x as

$$F_X(x) \begin{cases} P(X \leq x) = 0 & x < 0 \\ P(X \leq x) = 1/4 & 0 \leq x < 1 \\ P(X \leq x) = 3/4 & 1 \leq x < 2 \\ P(X \leq x) = 1 & x \geq 2 \end{cases}$$

and the graph of this function is given as



Note: The domain of CDF is $\mathbb{R}$ while the range of CDF is $[0, 1]$.

End of Lecture

In-Depth Look at CDF

What you'd realized about CDF are the followings:

- It's non-decreasing i.e. for any $x \leq y, F_X(x) \leq F_X(y)$.
- $\lim_{x \to +\infty} F_X(x) = 1$
- $\lim_{x \to -\infty} F_X(x) = 0$
- All probabilities in X can be stated as a term of $F_X(x) \implies$ CDF of X can complete its distribution.
- $F_X(x)$ is right continuous which is expressed as

$$\lim_{x \to x_k^+} F_X(x) = F_X(x_k), \text{ for } x_k \leq x \leq x_{k+1} \tag{2.2}$$

For a discrete random variable X:

- It must have a finite ($R_X = \{x_1, x_2, x_3, \ldots, x_n\}$) or countably infinite support ($R_X = \{x_1, x_2, \ldots\}$)
- Furthermore, $x_1 < x_2 < x_3 < \cdots$ where x_1 is the least element.
- The CDF will always start at 0, then it jumps at each point of the support while staying smooth for any points in-between support values.

Definition 2.5. The CDF jumps at each point in the support of a discrete random variable X and the height of said jump is defined as

$$F_X(x_k) - F_X(x_{k-1}) = P(X = x_k) \tag{2.3}$$

Example 2.1.3. Suppose using the same CDF in example 2.1.2, we wants to find the height between the support of 2 and 1. Then using the above definition, we get that

$$\begin{aligned} F_X(2) - F_X(1) &= P(X = 2) \\ &= P(X \geq 2) - P(X \geq 1) = 1 - \frac{3}{4} = \frac{1}{4} \end{aligned}$$

Theorem 2.1. $P(a < x \leq b) = F_X(b) - F_x(a), \ \forall a < b$

Proof. Suppose that for some arbitrary $a, b : a < b$. Then,

$$\begin{aligned} P(a < x \leq b) &= P(X = a+1) + P(X = a+2) + \cdots + P(X = b) \\ &= \Big[P(X = 0) + P(X = 2) + \cdots + P(X = a) + P(X = a+1) + \cdots \\ &\qquad + P(X = b)\Big] - \Big[P(X = 0) + P(X = 1) + \cdots + P(X = a)\Big] \\ &= P(X \leq b) - P(X \leq a) \\ &= F_X(b) - F_x(a) \end{aligned}$$

$\therefore P(a < x \leq b) = F_X(b) - F_x(a),\ \forall a < b.$ □

2.1.2 Probability Mass Function

Definition 2.6. The **probability mass function (PMF)** of a discrete random variable X with $R_X = \{x_1, x_2, \ldots\}$ is given as

$$P_X(x_k) = P(X = x_k) \tag{2.4}$$

where $P_X(x_k) = P(X = x_k) > 0$ if $x_k \in R_x$ and $P_X(x_k) = 0$ if $x_k \notin R_X$

The reason we have PMF is that it will give you exact probabilities like you have see as a probability distribution [table] i.e. PMF gives determines uniquely the probability distribution of $X \implies$ CDF can be determined by PMF and even vice versa. The last statement above can be formulated as following

Definition 2.7. For any given $P_X(x_k) = P(X = x_k)$ for X with $R_X = \{x_1, x_2, \ldots\}$. Then,

$$F_X(x_i) = P(X \leq x_i) = \sum_{x \leq x_i} P(X = x) \tag{2.5}$$

2.1.3 Discrete Uniform Distribution

We'll now be looking at some *discrete distribution* of which many real life scenarios are modelled after.

Definition 2.8. A discrete random varaible X is said to have a **discrete uniform distribution** if for $x_1, x_2, \ldots, x_N \in \mathbb{R}$,

$$P_X(x) = P(X = x) = \frac{1}{N} \tag{2.6}$$

where $X = x_1, x_2, \ldots, x_N$ and N is the total number of points in the support of X (i.e. $N = |R_X|$).

Remark 2.2. *We can see that the equation of uniform discrete distribution is expressed in the form of a PME*

Example 2.1.4. Throw a fair die. The possible outcomes are $\{1,2,3,4,5,6\}$. The random variable takes the values: $X = 1,2,3,4,5,6$. Then,

$$P(X = x_i) = \frac{1}{6}$$

for all $x_i \in \{1,2,3,4,5,6\}$.

Select a phone number randomly from the list of contacts in your cellphone. Let X be the last digit of the randomly selected phone number. Then, the possible values of X are $X = 1,2,\ldots,9$. Then, the probability that the randomly selected number has the last digit x_i is

$$P(X = x_i) = \frac{1}{10}$$

2.1.4 Bernoulli Distribution

Definition 2.9. A random variable X is said to have a **Bernoulli distribution** if X can take only 2 possible values and that is: $X = 0,1$ where $X = 0$ is representative of a failed experiment while $X = 1$ is the success. The probability distribution Bernoulli random variable is given as

$$P(X = x) = \begin{cases} p^x(1-p)^{1-x} & \text{if } x \in \{0,1\} \\ 0 & \text{if otherwise} \end{cases} \tag{2.7}$$

where p is the probability of success.

Definition 2.10. An random experiment where there's only 2 outcomes is called a **Bernoulli trial**.

Remark 2.3. *If X has a distribution given above, we say that X is Bernoulli distributed with parameter p and is writen as*

$$X \sim \text{Bernoulli}(P) \tag{2.8}$$

It does not just stop at failure and success, a Bernoulli random variable can also be related to the occurrence (or non occurrence) of a certain event E. If event E occurs, then $X = 1$; otherwise $X = 0$.

Example 2.1.5. Consider the following examples:

- Making a job application (accepted vs rejected).
- Testing a light bulb (defected or not defected).
- Getting tested for COVID-19 (infected or not infected).

End of Lecture

2.1.5 Binomial Distribution

Definition 2.11. A **binomial experiment** is an experiment if it satisfies the following criteria:

1. It consists of n independent Bernoulli trials.
2. The probability of success p remains constant from trial to trial.
3. We are interested in x success out of n trials, where $x = 1, 2, \ldots, n$.

Definition 2.12. Let X be the random variable that counts the number of successes in n Bernoulli trials, where the probability of success remains constant from trial to trial. Furthermore, the trials are independent. Then X is a **Binomial random variable** and it's probability distribution is called the **Binomial Distribution**. We say that X is the binomial random variable with parameters n and p, which is denoted as

$$X \sim \text{Binom}(n, p) \tag{2.9}$$

Example 2.1.6. According to tables by the *National Centre for Health Sciences in Vital Statistics for United States*, a person at the age of 20 years has a probability of 0.8 of being alive at the age of 65 years. Suppose three people of age 20 years are chosen at random (all people in the population of 20 years old the US are equally likely to be chosen).

Then, we can let A_i to be the event that the i^{th} person chosen to be alive at age 65 while A_i^c is the i^{th} person chosen to be dead at age 65. We are given that a person at the age of 20 years has the probability of 0.8 of being alive at the age of 65 years. And, we've seen that the Bernoulli random variable is the model that can take 2 possible value, in this case, alive vs not alive. Now, let 0 to be the value corresponds to not alive while 1 to be the value corresponds to being alive at 65. Then, we get the following Bernoulli probability distribution distribution:

x	0	1
$P(X = x)$	0.2	0.8

Suppose that we observe 3 people to see if they're alive or not, this can be considered as a binomial experiment as it satisfies all the criteria above. The number of possible ways that this can occur in 3 people are: $\binom{3}{x}$ where x is the number of people alive out of the 3. So, we will get the following Binomial distribution:

x	$P(X=x)=\binom{3}{x}(0.8)^x(0.2)^{3-x}$
0	$P(X=0)=0.008$
1	$P(X=1)=0.096$
2	$P(X=2)=0.384$
3	$P(X=3)=0.512$

Theorem 2.2. *For a binomial random variable X*

$$P(X=x)=\binom{n}{x}p^x(1-p)^{n-x} \tag{2.10}$$

where $x=0,1,2,\ldots,n$

Proof. Consider an experiment that consists of n Bernoulli trials, with the probability of success p. Each sample point in the sample space consists of a sequence of x successes and $n-x$ failures. A sample point with first x trials resulting in the successes and the last $n-x$ trials resulting in the failures is

$$\underbrace{SSS\ldots S}_{x}\underbrace{FF\ldots F}_{n-x}$$

Essentially, this represents the intersection of n independent events. As the probability of success S remains constant from trial to trial, and thus

$$P(\underbrace{S\cap S\cap S\cap\cdots\cap S}_{x})P(\underbrace{F\cap F\cap\cdots\cap F}_{n-x})$$

which can be written as

$$\underbrace{p\cdot p\cdot p\cdot\ldots\cdot p}_{x}\underbrace{((1-p)\cdot\ldots\cdot(1-p))}_{n-x}=p^x(1-p)^{n-x}$$

This result corresponds to one particular sequence. There are $\binom{n}{x}$ such sequences. Therefore, the event of obtaining x successes in n trials is the union of all $\binom{n}{x}$ possible outcomes all of which are disjoint with probability $p^x(1-p)^{n-x}$. Therefore,

$$P(X=x)=\binom{n}{x}p^x(1-p)^{n-x}$$

for $x=0,1,2,\ldots,n$ and $0<p<1$. □

Example 2.1.7. Suppose that a lot of 5000 electrical fuses contains 5% defectives. If a sample of 5 fuses is tested, find the probability of observing at

least one defective.

Solution: First, we're given a choice between defective and non-defective with defective has a chance of 5% $\implies$ this can be a Bernoulli trial. We're then needed to in a sample of 5 fuses $\implies$ 5 trials. Thus, let X be the random variable, then, $X \sim \text{Binom}(5, 0.05)$. To be specific,

$$P(X = x) = \binom{5}{x}(0.05)^x(0.95)^{5-x}$$

where x is the number of defectives. Now, we're asked for at least 1 defective. Therefore,

$$\begin{aligned} P(X \geq 1) &= 1 - P(X < 1) \\ &= 1 - P(X = 0) \\ &= 1 - \binom{5}{0}(0.05)^0(0.95)^{5-0} \approx \boxed{0.2262} \end{aligned}$$

2.1.6 Geometric Distribution

Example 2.1.8. Suppose that 30% of the applicants for a certain industrial job possess advanced training in computer programming. Applicants are interviewed sequentially and are selected at random from the pool. Suppose that the job pool is sufficiently large so that the independence can be assumed. Find the probability that the first applicant with advanced training in programming is found on the 5^{th} interview.

Solution: Basically, the event we're trying to find is $E = F \cap F \cap F \cap F \cap S$ where F is failure and S is success. In this case, success denotes by the applicant with advanced training in programming. Now, because we're looking at a specific sequence, we cannot really use binomial distribution (it does not consider a particular sequence) even if it satisfies all conditions. Nevertheless, with the law of probability and the event space, we can deduce the following:

$$\begin{aligned} & E = F \cap F \cap F \cap F \cap S \\ \implies & P(E = P(F \cap F \cap F \cap F \cap S) \\ & = P(F)P(F)P(F)P(F)P(S) && \text{since they're independent} \\ & = (1-p)^4 p^1 = (0.7)^4(0.3) = \boxed{0.072} \end{aligned}$$

Now, suppose that we want to find the probability that the first applicant with advanced training in programming is found on the x^{th} interview. The first applicant with the required characteristics can be found on the first trial ($x = 1$), or on the second ($x = 2$), or on the third ($x = 3$) and so on. Then, we get that

$$\begin{aligned} P(X = x) &= P(\underbrace{F \cap F \cap F \cap \cdots \cap F}_{x-1} \cap S) \\ &= \underbrace{P(F) \cdot \ldots \cdot P(F)}_{x-1} \cdot P(S) \\ &= (1-p)^{x-1} p = (0.7)^{x-1}(0.3) \end{aligned}$$

where $x \in \{1, 2, \ldots\}$

Definition 2.13. A random variable X is said to have **geometric distribution** ($X \sim \text{Geom}(p)$) if

$$P_X(x) = P(X = x) = (1-p)^{x-1} p \tag{2.11}$$

where $x = 1, 2, 3, \ldots$ and $0 < p < 1$ i.e. it's the probability that the first success occur on x^{th} trials.

Remark 2.4. *The proof that this is the first success occur on x^{th} trials is similar (even same) to the answer of the second part of example 2.1.8*

Remark 2.5. *Note the followings:*

- *The binomial random variable gives the number of successes in fixed number of trials.*
- *The geometric random variable gives the trial at which the first success occurs, where the number of trials is not fixed.*

Example 2.1.9. You roll a fair die repeatedly until a 6 observed. If X is the total number of times that you roll the die until you get a 6, find $P(X = k)$, for $k = 1, 2, 3, \ldots$.

Solution: Let X be the random variable, then $X \sim \text{Geom}(p)$. Now, the probability of getting a 6 on the k^{th} trial is given as $1/6$ (since it's a fair die). Then,

$$\begin{aligned} P(X = k) &= (1-p)^{k-1} p \\ &= \left(1 - \frac{1}{6}\right)^{k-1} \cdot \frac{1}{6} \end{aligned}$$

$$= \left(\frac{5}{6}\right)^{k-1} \cdot \frac{1}{6} = \boxed{\frac{5^{k-1}}{6^k}}$$

Example 2.1.10. You ask people outside a polling station who they voted for until you find someone that voted for the Liberal candidate in a recent election. The probability that a randomly chosen person votes a liberal candidate is 0.1. What is the probability that the first person who voted Liberals is the fifth person you interviewed.

Solution: Let X be the random variable, then $X \sim \text{Geom}(p)$. Therefore,

$$\begin{aligned} P(X = 5) &= (1 - 0.1)^{5-1}(0.1) \\ &= (0.9)^4(0.1) = \boxed{0.06561} \end{aligned}$$

We can twist this question around to make it into a binomial distribution: You ask 20 people outside a polling station whether they voted for Liberal candidate in a recent election.The probability that a randomly chosen person votes a liberal candidate is 0.1. What is the probability that the five of them voted for Liberals.

Solution: Let X be the random variable, then $X \sim \text{Binom}(n, p) = \text{Binom}(20, 0.1)$. Therefore,

$$\begin{aligned} P(X = 5) &= \binom{20}{5}(0.1)^5(0.9)^{20-5} \\ &= \binom{20}{5}(0.1)^5(0.9)^{15} = \boxed{0.03192} \end{aligned}$$

2.1.7 Negative Binomial Distribution

Example 2.1.11. Suppose that 30% of the applicants for a certain industrial job possess advanced training in computer programming. Applicants are interviewed sequentially and are selected at random from the pool. Find the probability that the **third** applicant with advanced training in programming is found on the fifth interview.

Solution: Here, like before, the settings are quite similar to that of the binomial distribution, but we cannot use it as we require a specific sequence once again.

Now, let's analyze the question, out of the three applicants with advanced training, two can be in any of the first to fourth interview. On the other hand, the third applicant with advanced training will definitely be in the fifth interview. The probability of success (have advanced training) is given as $p = 0.3$. Then, define the following event space:

- E = the event that third success occurs on the fifth trial.
- E_1 = the event that 2 success occurs in the first 4 trials.
- E_2 = the event that third success occurs on the fifth trial.

Then,

$$\begin{aligned} & E = E_1 \cap E_2 \\ \implies & P(E) = P(E_1 \cap E_2) \\ & = P(E_1) \cdot P(E_2) && \text{since they're independent} \\ & = \binom{4}{2}(0.3)^2(0.7)^2 \cdot (0.3) \\ & = \boxed{0.0794} \end{aligned}$$

This model is known as the *negative binomial distribution.*

Definition 2.14. The **negative binomial random variable** X gives the trial on which r^{th} success is achieved in a sequence of independent trials, denoted as

$$X \sim \text{NB}(r, p) \tag{2.12}$$

Each trial can result in either a success (S) or a failure (F). The probability of success remains constant from trial to trial.

Theorem 2.3. *Let $X \sim NB(r, p)$. Then,*

$$P(X = x) = \binom{x-1}{r-1} p^r (1-p)^{x-r} \tag{2.13}$$

where $x = r, r+1, \ldots$

Proof. The event $\{X = x\}$ (r^{th} success on x^{th} trial) can be written as $\{X = x\} = \{r-1 \text{ successes in } x-1 \text{ trials}\} \cap \{r^{\text{th}} \text{ success on } x^{\text{th}} \text{ trial}\}$. Then,

$$P(X = x) = P(E_1 \cap E_2)$$

$$
\begin{aligned}
&= P(E_1) \cdot P(E_2) \\
&= \binom{x-1}{r-1} p^{r-1} (1-p)^{x-1-(r-1)} p \\
&= \binom{x-1}{r-1} p^{r} (1-p)^{x-r}
\end{aligned}
$$

where $x = r, r+1, \ldots$. □

Example 2.1.12. An oil company conducts a geological study that indicates that an exploratory oil well should have a 20% chance of striking oil. What is the probability that the first strike comes on the third well drilled?

Solution: Let X be the random variable. Then, $X \sim \text{NB}(r, p) = \text{NB}(1, 0.2)$. Therefore,

$$
\begin{aligned}
P(X = 3) &= \binom{3-1}{1-1} (0.2)^1 (0.8)^{3-1} \\
&= (0.8)^2 (0.2) = \boxed{0.128}
\end{aligned}
$$

In fact, you can also solve this using geometric distribution as we're trying to find the first success.

Now, Using the same information above, what is the probability that the third strike comes on the seventh well drilled?

Solution: Like before, $X \sim \text{NB}(r, p) = \text{NB}(3, 0.2)$. Therefore,

$$
\begin{aligned}
P(X = 7) &= \binom{7-1}{3-1} (0.2)^3 (0.8)^{7-3} \\
&= \binom{6}{2} (0.2)^3 (0.8)^4 \approx \boxed{0.0491}
\end{aligned}
$$

2.1.8 Poisson Distribution

Let us suppose that we are interested in modeling the occurrence of events during a fixed time interval or the number of successes in a very large number of trials, such as

- The number of students in zoom office hours.

- The number of text messages you receive during fixed time interval.
- Number of accidents on a particular intersection.

This type of situation can be modeled using the *Poisson distribution*

Definition 2.15. A discrete random variable X is said to have **Poisson distribution** with parameter λ if it has the following probability distribution

$$P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!} \tag{2.14}$$

where $x = 0, 1, 2, \ldots$ (number of occurences in a unit time interval), λ is the average number of arrivals, and $e = 2.7182\ldots$. If X follows the poisson distribution, we denote it as

$$X \sim \text{Pois}(\lambda) \tag{2.15}$$

Example 2.1.13. Show that the probabilities assigned by the Poisson probability distribution satisfy that $\sum_{x=0}^{\infty} p(x) = 1 \ \forall x$.

Proof. Suppose that $p(x)$ follows Poisson probability distribution. Then,

$$\begin{aligned}
\sum_{x=0}^{\infty} p(x) = \sum_{x=0}^{\infty} \frac{\lambda^x e^{-1\lambda}}{x!} &= P(X=0) + P(X=1) + \cdots \\
&= \frac{\lambda^0 e^{-1\lambda}}{0!} + \frac{\lambda^1 e^{-1\lambda}}{1!} + \cdots \\
&= e^{-\lambda} \underbrace{\left(\frac{\lambda^0}{0!} + \frac{\lambda^1}{1!} + \frac{\lambda^2}{2!} + \cdots \right)}_{\text{Maclaurin's series for } e^{\lambda}} \\
&= e^{-\lambda} e^{\lambda} = 1
\end{aligned}$$

Thus, $\sum_{x=0}^{\infty} p(x) = 1 \ \forall x$ □

Theorem 2.4. *Let $X \sim Binom(n, p)$. Then, for $np = \lambda$, we have*

$$\lim_{\substack{n\to\infty \\ p\to 0}} P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!} \tag{2.16}$$

Proof. Suppose that $X \sim \text{Binom}(n, p)$ and $np = \lambda$. Then,

$$\lim_{\substack{n\to\infty \\ p\to 0}} P(X = x) = \lim_{\substack{n\to\infty \\ p\to 0}} \binom{n}{x} p^x (1-p)^{n-x}$$

$$
\begin{aligned}
&= \lim_{n\to\infty} \binom{n}{x} \left(\frac{\lambda}{n}\right)^x \left(1-\frac{\lambda}{n}\right)^{n-x} \\
&= \lim_{n\to\infty} \frac{n!}{x!(n-x)!} \left(\frac{\lambda}{n}\right)^x \left(1-\frac{\lambda}{n}\right)^{n-x} \\
&= \lim_{n\to\infty} \frac{n(n-1)(n-2)\cdots(3)(2)(1)}{x!(n-x)(n-x-1)\cdots(3)(2)(1)} \frac{\lambda^x}{n^x} \left(1-\frac{\lambda}{n}\right)^n \left(1-\frac{\lambda}{n}\right)^{-x} \\
&= \lim_{n\to\infty} \frac{\lambda^x}{x!} \left(n(n-1)\cdots(n-(x-1))\right) \frac{1}{n^x} \left(1-\frac{\lambda}{n}\right)^n \left(1-\frac{\lambda}{n}\right)^{-x} \\
&= \lim_{n\to\infty} \frac{\lambda^x}{x!} \underbrace{\left(\frac{n}{n}\left(1-\frac{1}{n}\right)\cdots\left(1-\frac{x-1}{n}\right)\right)}_{=1\cdot 1\cdot\ldots\cdot 1=1} \left(1-\frac{\lambda}{n}\right)^n \left(1-\frac{\lambda}{n}\right)^{-x} \\
&= \lim_{n\to\infty} \frac{\lambda^x}{x!} \underbrace{\left(1-\frac{\lambda}{n}\right)^n}_{=e^{-\lambda}} \underbrace{\left(1-\frac{\lambda}{n}\right)^{-x}}_{=1} \quad (x \text{ is constant}) \\
&= \frac{\lambda^x e^{-\lambda}}{x!}
\end{aligned}
$$

Thus, $\lim_{\substack{n\to\infty \\ p\to 0}} P(X=x) = \frac{\lambda^x e^{-\lambda}}{x!}$ □

This theorem is basically the Poisson approximation to the binomial distribution.

Remark 2.6. *When the value of n in a binomial distribution is large and the value of p is very small, the binomial distribution can be approximated by a Poisson distribution. If $n \geq 100$ and $np \leq 10$, then Poisson provides a good approximation to the binomial distribution.*

Example 2.1.14. Suppose that sections of textile of length 1 cm have a flaw in them with the probability 0.01. If 1000 such sections are examined what is the probability that

a. Exactly 10 will have a flaw?

b. At least 50 will have a flaw?

Solution: Now, we can evidently calculate this with binomial distribution however it will takes a really long time. So let's check if the condition is met to use Poisson distribution instead. First, $n = 1000 > 100$ and $\lambda = np = 0.01 \cdot 1000 = 10$, thus, we can use it. Then,

a. If exactly 10 will have flaw then, $P(X = 10) = \frac{\lambda^{10} e^{-\lambda}}{10!} \approx \boxed{0.125}$

b. Try it yourself!

Example 2.1.15. Suppose that you receive two text messages per hour on average.

a. Find the probability of no test messages in a given hour.

b. Find the probability of at least three text messages in a given hours.

Solution: This situation can be modeled by Poisson distribution. Let X be a random variable that gives the number of text messages per hour. We have that $\lambda = 2 \implies X \sim \text{Pois}(2)$. For a, we need to find $P(X = 0)$:

$$P(X = 0) = \frac{2^0 e^{-2}}{0!} = e^{-2} \approx \boxed{0.1353}$$

For b, we have to find $P(X \geq 3)$, which is:

$$\begin{aligned} P(X \geq 3) &= 1 - P(X < 3) \\ &= 1 - P(X = 0) - P(X = 1) - P(X = 2) \\ &= 1 - \frac{2^0 e^{-2}}{0!} - \frac{2^1 e^{-2}}{1!} - \frac{2^2 e^{-2}}{2!} \approx \boxed{0.3233} \end{aligned}$$

Example 2.1.16. Industrial accidents occur according to a Poisson process with an average of three accidents per month. What is the probability of ten accidents during last two months?

Solution: This situation can be modeled by Poisson distribution. Let X be a random variable that gives the number of accidents per month. We have that $\lambda = 3 \implies X \sim \text{Pois}(3)$. Since we're look at the last 2 months, we will remodified our average rate of arrivales into $\lambda' = 2\lambda = 2(3) = 6$. Then, $X \sim \text{Pois}(6)$. We're trying to find $P(X = 10)$. Then,

$$P(X = 10) = \frac{6^{10} e^{-6}}{10!} \approx \boxed{0.0413}$$

2.1.9 Hypergeometric Distribution

We've previously seen this kind of example, but let's investigate it again and formalize a definition.

Example 2.1.17. You have a box that contains a green balls and b red balls. Let $a+b=N$. You choose n balls at random (without replacement). Let X be the number of green balls in your sample. Then,

$$P(X=x)=\frac{\binom{a}{x}\binom{b}{n-x}}{\binom{N}{n}}=\frac{\binom{a}{x}\binom{N-a}{n-x}}{\binom{N}{n}}$$

Definition 2.16. A random varialbe is said to have **hypergeometric distribution** if it has the following probability distribution,

$$P(X=x)=\frac{\binom{a}{x}\binom{N-a}{n-x}}{\binom{N}{n}} \tag{2.17}$$

where N is the population size, n is the sample size, a is the number of possible successes in the population, and $x=0,1,2,\ldots,\min(a,n)$. If a random variable X is a hypergeometric random variable with parameters N, n, and a, we denote it as

$$X\sim \text{Hypergeom}(N,n,a) \tag{2.18}$$

2.2 Expectation and Variance of Discrete Random Variable

The more general definition of expection is as follows: expected value of a random variable X is the measure of centrality of X. It is the weighted mean or average of all possible values of the random variable X.

2.2.1 Expectation

Definition 2.17. Let X be a discrete random variable with the possible values $\{x_1,x_2,\ldots\}$ (finite or countably infinite). The **expected value** of X, denoted as $E(X)$, is defined as

$$E(X)=\sum x_k P(X=x_k) \tag{2.19}$$

Note the followings:

- $E(X)$ is also know as the population mean of X.
- $E(X)$ is also denoted as μ_X.
- $E(X)$ is the weighted average of X, where the weight are probabilities at the possible values of X.

- $E(X)$ can be thought of as a long run average of X.
- $E(X)$ is a theoretical average. It's rarely realized in practice.
- $E(X)$ exists if $E|X|$ converges.

Proposition 2.1. *If a mathematical expectation exists and let a be a constant, it will satisfy the following properties.*

1. $E(a) = a$
2. $E(aX) = aE(X)$
3. $E(X_1 + X_2 + \cdots + X_n) = E(X_1) + E(X_2) + \cdots + E(X_n)$
4. $E(a_1X_1 + a_2X_2 + \cdots + a_nX_n) = a_1E(X_1) + a_2E(X_2) + \cdots + a_nE(X_n)$, *where a_i are constant* $\forall\, i \in \{1, 2, \ldots, n\}$

Example 2.2.1. The expected value of rolling a fair die would be :

$$E(X) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + \cdots + 6 \cdot \frac{1}{6} = \boxed{3.5}$$

Tossing a fair coin would be:

$$E(X) = 0 \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} = \boxed{\frac{1}{2}}$$

Example 2.2.2. Consider the PMF $f(x) = \frac{c}{x^2}$ of the random variable X, where $x = 1, 2, \ldots$. Find the expected value of X.

Solution: Consider the pmf $f(x) = \frac{c}{x^2}$ of the random variable X, where $x = 1, 2, \ldots$. Then,

$$\begin{aligned} E(X) = \sum_{n=1}^{\infty} xP(X = x) &= \sum_{n=1}^{\infty} x \cdot \frac{c}{x^2} \\ &= \sum_{n=1}^{\infty} \frac{c}{x} = c \cdot \underbrace{\sum_{n=1}^{\infty} \frac{1}{x}}_{\text{diverges}} \end{aligned}$$

In this case, the entire $E(X)$ diverges $\forall x \implies E(X)$ does not exists.

Example 2.2.3. An insurance company issues a one-year $1000 policy insuring against an occurrence A that historically happens to 2 out of every 100 owners of the policy. Administrative fees are $15 per policy and are not part of the company's "profit". How much should the company charge for the policy if it requires that the expected profit per policy be $50.

Solution: Let C be amount the company should charge the owners and X_1, X_2 be the company profits. Then,

- $X_1 = C - 15$ if the A does not happen (A^c).
- $X_2 = C - 15 - 100$ if the A happens.

We're given that $P(A) = 0.02 \implies P(A^c) = 0.98$. We're also given the expected value as $E(X_1 + X_2) = 50$. Then,

$$\begin{aligned} E(X_1 + X_2) &= E(X_1) + E(X_2) \\ 50 &= X_1 P(A^c) + X_2 P(A) \\ 50 &= (C - 15)(0.98) + (C - 15 - 100)(0.02) \\ \implies C &= \boxed{\$85} \end{aligned}$$

2.2.2 Variance

Now, the mean of a random variable is not sufficient to characterize its distribution. We need a measure of spread as well. Hence, we define the following:

Definition 2.18. The **variance** of a random variable X, with mean $E(X) = \mu_X$, is defined as

$$\text{Var}(X) = E[(X - \mu_X)^2] = E[(X - E(X))^2] \tag{2.20}$$

Note the followings relating to variance of a random variable X:

- $\text{Var}(X)$ is can be denoted as σ_X^2.
- $\text{Var}(X)$ is the expected value of $(X - \mu_X)^2$.
- As $\text{Var}(X) \uparrow$, the values of X are spread further from the mean.
- A probability distribution can be completely specified by μ_X and σ_X^2, the parameters of the distribution which are constant for a given population.

- As $(X-\mu_X)^2 \geq 0, \text{Var}(X) \geq 0$.
- The units of $\text{Var}(X)$ are the same as that of X^2. Therefore, we prefer to use the positive square root of $\text{Var}(X)$, which is the **standard deviation**: $\sqrt{\text{Var}(X)} = SD(X) = \sigma_X$, which has the same unit as X.
- $\text{Var}(cX) = c^2\text{Var}(X)$.

Theorem 2.5. *Let X be a random variable. Then,*

$$Var(X) = E[X^2] - [E(X)]^2 \tag{2.21}$$

Proof. Let X be a random variable. Then,

$$\begin{aligned}
\text{Var}(X) &= E[(X-E(X))^2] \\
&= E[X^2 - 2XE(X) + [E(X)]^2] \\
&= E[X^2] - 2E[XE(X)] + E[E(X)^2] \\
&= E[X^2] - 2E[X]E[X] + E[X]^2 \\
&= E[X^2] - 2E[X]^2 + E[X]^2 \\
&= E[X^2] - E[X]^2
\end{aligned}$$

Thus, $\text{Var}(X) = E[X^2] - E[X]^2$. □

Now, we've previous calculated the expected value for rolling a fair die, **what about calculating its variance?** Well...using theorem 2.5, We can solve for the variance as:

$$\begin{aligned}
\text{Var}(X) &= E[X^2] - (E[X])^2 \\
&= \sum_{x=1}^{6} x^2 P(X=x) - (3.5)^2 \\
&= \frac{91}{6} - \frac{49}{4} \approx \boxed{2.92}
\end{aligned}$$

We can also calculate the standard deviation to be $\sigma_X = \sqrt{2.92} \approx \boxed{1.71}$.

Example 2.2.4. Suppose that X be a random variable such that $P(X=x) = \frac{1}{n}$ for $x = 1, 2, \ldots, n$ i.e. X has uniform distribution on first n positive integers. Calculate its expectation and variance.

Solution: Its expectation will be calculated as:

$$E(X) = \sum_{x=1}^{n} xP(X=x)$$

$$= \sum_{x=1}^{n} x \cdot \frac{1}{n}$$
$$= \frac{1}{n} \sum_{x=1}^{n} x = \frac{1}{n} \cdot \frac{n(n+1)}{2} = \boxed{\frac{n+1}{2}}$$

We can then calculate its variance as:

$$\begin{aligned}
\text{Var}(X) &= E[X^2] - (E[X])^2 \\
&= \sum_{x=1}^{n} x^2 \cdot \frac{1}{n} - \left(\frac{n+1}{2}\right)^2 \\
&= \frac{1}{n} \sum_{x=1}^{n} x^2 - \left(\frac{n+1}{2}\right)^2 \\
&= \frac{1}{n} \cdot \frac{n(n+1)(2n+1)}{6} - \left(\frac{n+1}{2}\right)^2 \\
&= \frac{(n+1)(2n+1)}{6} - \left(\frac{n+1}{2}\right)^2 = \boxed{\frac{n^2-1}{12}}
\end{aligned}$$

2.2.3 Moments of a Distribution

To simply put, *moments of a distribution* are quantitative measures that gives some information on the shape, spread, and characteristics of a probability distribution.

Definition 2.19. Let X be a discrete random variable then the **moments** of X around the origin are:

- $E(X)$ is the first moment of X about the origin.
- $E(X^2)$ is the second moment of X about the origin.

  ⋮
- $E(X^k)$ is the k^{th} moment of X about the origin

where $E(X^r) < \infty$ for $r = 1, 2, \ldots, k$.

Remark 2.7. *The above definition is the moment of X about the origin. However, there's also moments of X about the mean. To be specific, the k^{th} moment of X about its mean μ_X is defines as $E(X - E[X])^k$ i.e. $E(X - E[X])$ is the first moment of X about its mean.*

Basically, we can say that the mean of the random variable X, $E(X)$ is the first moment of X about the origin. On the other hand, the variance of the random variable X, $E(X-E[X])^2$ is the second moment of X about its mean.

Definition 2.20. Let X be a random variable. Then, the ***r^{th}* factorial moment** of X is defined as

$$E(X_r) = E[X(X-1)\cdots(X-(r-1))] \tag{2.22}$$

where $E(X)$ is the first factorial moment of X and $E[X(X-1)]$ is the second factorial moment of X.

Interestingly, there's a way to obtain the powered (classical) moments from the factorial moments by solving a system of linear equation e.g. to obtain $E[X^2]$, we will solve from the second factorial moment of X:

$$\begin{aligned} E[X(X-1)] &= E[X^2 - X] \\ &= E(X^2) - E(X) \\ \implies E(X^2) &= E[X(X-1)] + E(X) \end{aligned}$$

Now that we've equipped ourselves with the knowledge of mean and variance, we can calculate for each of the distribution we've studied.

2.2.4 Mean and Variance of the Uniform Distribution

Theorem 2.6. *Let X be a random variable with discrete uniform distribution where $p_X(x) = P(X = x) = \frac{1}{N}$ for $X = x_1, x_2, \ldots, x_N$. Then,*

$$E(X) = \frac{1}{N}\sum_{k=1}^{N} x_k = \overline{x} \tag{2.23}$$

and

$$Var(X) = \frac{1}{N}\sum_{k=1}^{N} x_k^2 - \overline{x}^2 \tag{2.24}$$

Proof. Suppose that X be a random variable with discrete uniform distribution where $p_X(x) = P(X = x) = \frac{1}{N}$ for $X = x_1, x_2, \ldots, x_N$. Then,

$$E(X) = \sum_{k=1}^{N} x_k P(X = x_k) \qquad \text{Var}(X) = E[X^2] - [E(X)]^2$$

$$= \sum_{k=1}^{N} x_k \cdot \frac{1}{N} \qquad = \sum_{k=1}^{N} x_k^2 P(X = x_k) - \overline{x}^2$$

$$= \frac{1}{N} \sum_{k=1}^{N} x_k = \overline{x} \qquad = \frac{1}{N} \sum_{k=1}^{N} x_k^2 - \overline{x}^2$$

Therefore, we've proven it. □

2.2.5 Mean and Variance of the Bernoulli Distribution

Theorem 2.7. *Let X be a random variable with Bernoulli distribution where $P(X = x) = \begin{cases} 1-p & , x = 0 \\ p & , x = 1 \end{cases}$. Then,*

$$E(X) = p \tag{2.25}$$

and

$$Var(X) = p(1-p) \tag{2.26}$$

Proof. Suppose that X be a random variable with Bernoulli distribution where $P(X = x) = \begin{cases} 1-p & , x = 0 \\ p & , x = 1 \end{cases}$. Then,

$$E(X) = E(X = 0) + E(X = 1) \qquad \text{Var}(X) = E[X^2] - [E(X)]^2$$

$$= 0(1-p) + 1(p) \qquad = [0^2(1-p) + 1^2(p)] - p^2$$

$$= p \qquad = p - p^2 = p(1-p)$$

Therefore, we've proven it. □

2.2.6 Mean and Variance of the Binomial Distribution

Theorem 2.8. *Let X be a random variable with binomial distribution where $P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$. Then,*

$$E(X) = np \tag{2.27}$$

and

$$Var(X) = np(1-p) \tag{2.28}$$

Proof. Suppose X be a random variable with binomial distribution where $P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$. Then,

$$E(X) = \sum_{x=0}^{n} x \cdot \binom{n}{x} p^x (1-p)^{n-x}$$

$$
\begin{aligned}
&= \sum_{x=0}^{n} x \cdot \frac{n!}{x!(n-x)!} \cdot p^x (1-p)^{n-x} \\
&= \sum_{x=0}^{n} \frac{n(n-1)!}{(x-1)!(n-x)!} \cdot p^x (1-p)^{n-x} \\
&= \sum_{x=0}^{n} \frac{n(n-1)!}{(x-1)!((n-1)-(x-1))!} \cdot p p^{x-1} (1-p)^{((n-1)-(x-1))} \\
&= np \sum_{x-1=0}^{n} \binom{n-1}{x-1} p^{x-1} (1-p)^{((n-1)-(x-1))} = np
\end{aligned}
$$

Now, to find the variance, we have the following: $\text{Var}(X) = E[X^2] - [E(X)]^2$. We will start calculating $E[X^2]$ which we can use the factorial moment of X as before:

$$E[X^2] = E[X(X-1)] + E(X)$$

$$
\begin{aligned}
\implies E[X(X-1)] &= \sum_{x=0}^{n} x(x-1) P(X=x) \\
&= \sum_{x=2}^{n} x(x-1) \binom{n}{x} p^x (1-p)^{n-x} \\
&= \sum_{x=2}^{n} x(x-1) \cdot \frac{n!}{x!(n-x)!} \cdot p^x (1-p)^{n-x} \\
&= \sum_{x=2}^{n} x(x-1) \cdot \frac{n(n-1)(n-2)!}{x!(n-x)!} \cdot p^x (1-p)^{n-x} \\
&= \sum_{x=2}^{n} \frac{n(n-1)(n-2)!}{(x-2)!(n-x)!} \cdot p^2 p^{(x-2)} (1-p)^{n-x} \\
&= \sum_{x=2}^{n} \frac{n(n-1)(n-2)!}{(x-2)!((n-2)-(x-2))!} \cdot p^2 p^{(x-2)} (1-p)^{((n-2)-(x-2))} \\
&= n(n-1)p^2 \sum_{x-2=0}^{n} \binom{n-2}{x-2} p^{(x-2)} (1-p)^{((n-2)-(x-2))} \\
&= n(n-1)p^2
\end{aligned}
$$

Then, we can calculate the variance as follows:

$$
\begin{aligned}
\text{Var}(X) &= E[X^2] - [E(X)]^2 \\
&= E[X(X-1)] + E(X) - [E(X)]^2 \\
&= n(n-1)p^2 + np - np^2 = np(1-p)
\end{aligned}
$$

Therefore, we've proven it. □

2.2.7 Mean and Variance of Other Discrete Distribution

Theorem 2.9. *Let X be a random variable with Poisson distribution where $P(X=x)=\frac{\lambda^x e^{\lambda}}{x!}$. Then,*

$$E(X)=\lambda \tag{2.29}$$

and

$$Var(X)=\lambda \tag{2.30}$$

Theorem 2.10. *Let X be a random variable with geometric distribution where $P(X=x)=(1-p)^{x-1}p$. Then,*

$$E(X)=\frac{1}{p} \tag{2.31}$$

and

$$Var(X)=\frac{1-p}{p^2} \tag{2.32}$$

Theorem 2.11. *Let X be a random variable with negative binomial distribution where $P(X=x)=\binom{x-1}{r-1}p^r(1-p)^{x-r}$. Then,*

$$E(X)=\frac{r}{p} \tag{2.33}$$

and

$$Var(X)=\frac{r(1-p)}{p^2} \tag{2.34}$$

Theorem 2.12. *Let X be a random variable with hypergeometric distribution where $P(X=x)=\frac{\binom{a}{x}\binom{N-a}{n-x}}{\binom{N}{n}}$. Then,*

$$E(X)=n\cdot\frac{a}{N} \tag{2.35}$$

and

$$Var(X)=n\cdot\frac{a}{N}\left(1-\frac{a}{N}\right)\cdot\frac{N-n}{N-1} \tag{2.36}$$

End of Lecture

2.3 Continuous Random Variable

Now, we know that in discrete random variable, the outcomes in the sample space is finite or even countably infinite. Now, we're going to look at continuous random variable where the amount of outcomes are infinite and uncountable.

Definition 2.21. A random variable X with CDF $F_X(x)$ is said to be **continuous** if $F_X(x)$ is a **continuous function** $\forall x \in \mathbb{R}$.

Example 2.3.1. For a continuous random variable $P(X \leq a) = P(X < a)$. This is because,

$$\begin{aligned} P(X = x) &= P(X \leq x) - P(X < x) \\ &= F_X(x) - \lim_{y \to x^-} F_X(y) \\ &= F_X(x) - F_X(x) = 0 \qquad \text{since } F_X(x) \text{ is continuous} \end{aligned}$$

Hence, single points have probability 0 $\implies$ $\boxed{P(X \leq a) = P(X < a)}$.

Proposition 2.2. *Let $F_X(x)$ be a continuous function for a continuous random variable X. Then,*

$$P(x_1 \leq X \leq x_2) = P(x_1 < X < x_2) \tag{2.37}$$
$$= P(x_1 \leq X < x_2) \tag{2.38}$$
$$= P(x_1 < X \leq x_2) \tag{2.39}$$

Proof. Follows immediately from example 2.3.1. □

Now, remember that for a discrete random variable, we can determine its probability distribution using either PMF or CDF. However, for continuous random variables, we cannot use PMF since $P(X = x) = 0$ thus we use the CDF. Now, since a continuous random variable doers not have a positive mass at any single point, we will define a *probability density function* for a continuous variable X.

2.3.1 Probability Density Function

Definition 2.22. Let X be a continuous random variable with the CDF, $F_X(x)$, then the function $f(x)$ is called a **probability density function (PDF)** of X if

$$F_X(x) = \int_{-\infty}^{x} f_X(y)\,dy \tag{2.40}$$

i.e. when integrate PDF over $(-\infty, x]$ it will give the CDF.

Furthermore, a PDF should satisfy the following:

- If $f_X(x)$ is a PDF then $f_X(x) \geq \forall x \in (-\infty, \infty)$
- $\int_{-\infty}^{\infty} f_X(x)dx = 1$.
- We can obtain CDF from PDF and vice versa. Obtaining PDF from CDF, we simply use the fundamental theorem of calculus:

$$F'_X(x) = \frac{d}{dx}\int_{-\infty}^{x} f_X(y)dy = f_X(x)$$

Remark 2.8. *Any non-negative function can be used as a PDF.*

Example 2.3.2. Suppose that we have a function f that satisfies the first property of a PDF but when integrate it over the entire real line, it does not equal to 1 i.e.

$$\int_{-\infty}^{\infty} g(x)dx = C \neq 1$$

To get a PDF from g, we can normalize it as follows: Define $f(x) = \frac{g(x)}{C}$. Then,

$$\int_{-\infty}^{\infty} f(x)dx = \int_{-\infty}^{\infty} \frac{g(x)}{C}dx = 1$$

Example 2.3.3. Given $f(y) = cy^2, 0 \leq y \leq 2$ and $f(y) = 0$ elsewhere. Find the value of c for which $f(y)$ is a valid density function.

Solution: To have c as a value that make $f(y)$ a pdf, we will first integrate $f(y)$ over the entire real line and set it to equal to 1 (by definition of a PDF). Then,

$$\begin{aligned}
\int_{-\infty}^{\infty} f(y)dy &= 1 \\
\implies \int_{-\infty}^{0} f(y)dy + \int_{0}^{2} f(y)dy + \int_{2}^{\infty} f(y)dy &= 1 \\
\int_{-\infty}^{0} 0dy + \int_{0}^{2} f(y)dy + \int_{2}^{\infty} 0dy &= 1 \\
\int_{0}^{2} f(y)dy &= 1 \\
\int_{0}^{2} cy^2dy &= 1
\end{aligned}$$

$$c\int_0^2 y^2\,dy = 1$$

$$c\cdot\frac{8}{3} = 1 \iff \boxed{c = \frac{3}{8}}$$

Thus, $c = \frac{3}{8}$ for $f(y)$ to be a valid PDF.

Remark 2.9. *The PDF does not give the probabilities and can be greater than 1. However, when integrate the PDF over a given interval, it will give the probability.*

Example 2.3.4. Let X be a continuous random variable whose PDF is $f_X(x) = 3x^2, 0 < x < 1$. Then, $f(0.8) = 3 \times 0.8^2 = 1.92 < 1$, this is not a probability. Suppose that we need to find a probability over the interval $[0,1]$ for example, then we will integrate it.

In other words, the probability of a continuous random variable is the area under the curve of $f_X(x)$ (the PDF).

2.3.2 Expected Value and Variance of Continuous Random Variable

Definition 2.23. The **expected value of a continuous random variable** X is defined as

$$E(X) = \int_{-\infty}^{\infty} x f(x)\,dx \tag{2.41}$$

where $f(x)$ is the PDF. The $E(X)$ exists if $E(|X|) < \infty$.

Definition 2.24. The **variance of a continuous random variable** X is defined as

$$\text{Var}(X) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx \tag{2.42}$$

or

$$\text{Var}(X) = \int_{-\infty}^{\infty} x^2 f(x)\,dx - [E(X)]^2 \tag{2.43}$$

where $\mu = E(X)$ which is the expected value of X.

Moments of a Continuous Random Variable

Definition 2.25. The ***k^{th}* moment about the origin** of a continuous random variable X is defined as

$$E(X^k) = \int_{-\infty}^{\infty} x^k f(x)\,dx \tag{2.44}$$

Similarly, the **k^{th} moment about the mean** of a continuous random variable X is defined as

$$E[(X-\mu)^k] = \int_{-\infty}^{\infty} (x-\mu)^k f(x)\,dx \tag{2.45}$$

Remark 2.10. *The properties of expectation and variance for discrete random variables also hold true for continuous random variables.*

Example 2.3.5. Suppose X is a continuous random variable with the following PDF:

$$f(x) = \begin{cases} 3x^2 & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

Find the $E(X)$ and $\text{Var}(X)$.

Solution: We can determine the $E(X)$ and $\text{Var}(X)$ using the definitions above.

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} x f(x)\,dx \\ &= \int_{-\infty}^{0} x f(x)\,dx + \int_{0}^{1} x f(x)\,dx + \int_{1}^{\infty} x f(x)\,dx \\ &= \int_{-\infty}^{0} x0\,dx + \int_{0}^{1} x f(x)\,dx + \int_{1}^{\infty} x0\,dx \\ &= \int_{0}^{1} x f(x)\,dx \\ &= \int_{0}^{1} x3x^2\,dx = 3\left|\frac{x^4}{4}\right|_0^1 = \boxed{\frac{3}{4}} \end{aligned}$$

To calculate $\text{Var}(X)$, we will first calculate $E(X^2)$, which is

$$\begin{aligned} E(X^2) &= \int_{-\infty}^{\infty} x^2 f(x)\,dx \\ &= \int_{-\infty}^{0} x^2 f(x)\,dx + \int_{0}^{1} x^2 f(x)\,dx + \int_{1}^{\infty} x^2 f(x)\,dx \\ &= \int_{-\infty}^{0} x^2 0\,dx + \int_{0}^{1} x^2 f(x)\,dx + \int_{1}^{\infty} x^2 0\,dx \\ &= \int_{0}^{1} x^2 f(x)\,dx \\ &= \int_{0}^{1} x^2 3x^2\,dx = \frac{3}{5} \end{aligned}$$

End of Lecture

Then,

$$\begin{aligned}\text{Var}(X) &= E(X^2) - [E(X)]^2 \\ &= \frac{3}{5} - \left(\frac{3}{4}\right) = \boxed{\frac{3}{80}}\end{aligned}$$

We will now look at some special continuous distributions

2.3.3 Continuous Uniform Distributions

Definition 2.26. A continuous random variable X is said to have a **uniform distribution** over the interval $[a,b]$, shown as $X \sim$ Unif(a,b), if its PDF is given by

$$f_X(x) \begin{cases} \frac{1}{b-a} & a < x < b \\ 0 & x < a \text{ or } b > x \end{cases} \tag{2.46}$$

The support of $X \sim$ Unif(a,b) or $X \sim U(a,b)$ is (a,b) itself. If this distribution has $a = 0$ and $b = 1$, then it's called a **standard uniform distribution**.

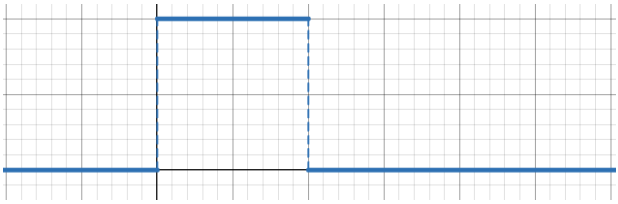


Figure 2.1: Continuous uniform distribution illustration.

Example 2.3.6. Let X a continuous random variable with standard uniform distribution. Then, its PDF is written as

$$f_X(x) \begin{cases} \frac{1}{1-0} = 1 & 0 < x < 1 \\ 0 & x < 0 \text{ or } 1 > x \end{cases}$$

Suppose that we want to calculate $P(0.2 < X < 0.4)$, then, we can calculate them as follows:

$$\begin{aligned}P(0.2 < X < 0.4) &= \int_{0.2}^{0.4} f_X(x)dx \\ &= \int_{0.2}^{0.4} 1dx = \boxed{0.2}\end{aligned}$$

Thus, $P(0.2 < X < 0.4) = 0.2$

Theorem 2.13. *The mean of a continuous uniform random variable defined over $a < x < b$ is*

$$E[X] = \frac{a+b}{2} \tag{2.47}$$

Proof. Let X be continuous uniform random variable. Then,

$$\begin{aligned} E[X] = \int_a^b x f_X(x)\,dx &= \int_a^b x \cdot \frac{1}{b-a}\,dx \\ &= \frac{1}{b-a}\int_a^b x\,dx \\ &= \frac{1}{b-a}\Big|\frac{x^2}{2}\Big|_a^b \\ &= \boxed{\frac{a+b}{2}} \end{aligned}$$

Therefore, we've shown that $E[X] = \frac{a+b}{2}$. □

Theorem 2.14. *The mean of a continuous uniform random variable defined over $a < x < b$ is*

$$Var[X] = \frac{(b-a)^2}{12} \tag{2.48}$$

Proof. Let X be continuous uniform random variable. To calculate the variance, we will first determine $E[X^2]$ over the support of X:

$$\begin{aligned} E[X^2] = \int_a^b x^2 f_X(x)\,dx &= \int_a^b x^2 \cdot \frac{1}{b-a}\,dx \\ &= \frac{1}{b-a}\int_a^b x^2\,dx \\ &= \frac{1}{b-a}\Big|\frac{x^3}{3}\Big|_a^b \\ &= \frac{a^2+ab+b^2}{3} \end{aligned}$$

Then, the variance $\mathrm{Var}[X]$ is given as:

$$\begin{aligned} \mathrm{Var}[X] &= E[X^2] - (E[X])^2 \\ &= \frac{a^2+ab+b^2}{3} - \left(\frac{a+b}{2}\right)^2 \\ &= \frac{a^2+ab+b^2}{3} - \frac{a^2+2ab+b^2}{4} \end{aligned}$$

$$= \frac{4a^2 + 4ab + 4b^2 - 3a^2 - 6ab - 3b^2}{12}$$
$$= \frac{a^2 - 2ab + b^2}{12} = \boxed{\frac{(b-a)^2}{12}}$$

□

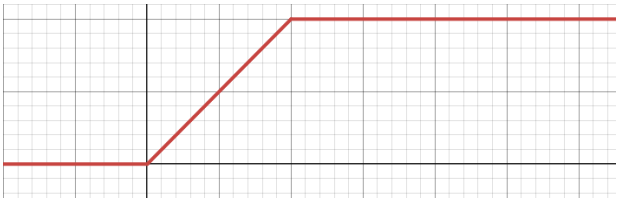


Figure 2.2: CDF of uniform continuous distribution.

Theorem 2.15. *Let X be a continuous uniform random variable. Then, its CDF will be defined as*

$$F_X(x) = \begin{cases} 0 & x \leq a \\ \frac{x-a}{b-a} & a < x < b \\ 1 & x \geq b \end{cases} \tag{2.49}$$

Proof. Let X be a continuous uniform random variable. Then, consider 3 interval that we will determine the CDF: $(-\infty, a], (a, b), [b, \infty)$. First, by the definition PDF of X, then, CDF at $(-\infty, a]$ will be 0 while that at $[b, \infty)$ is 1 (2 limits of CDF). Now, the CDF over the normal interval (a, b) will be found like usual:

$$\begin{aligned} F_X(x) &= \int_{-\infty}^{x} f_X(x)\,dx \\ &= \int_{-\infty}^{a} f_X(x)\,dx + \int_{a}^{x} f_X(x)\,dx \\ &= \int_{-\infty}^{a} 0\,dx + \int_{a}^{x} \frac{1}{b-a}\,dx \\ &= \frac{1}{b-a}\int_{a}^{x} dx = \boxed{\frac{x-a}{b-a}} \end{aligned}$$

Thus putting them together, we get the following CDF:

$$F_X(x) = \begin{cases} 0 & x \leq a \\ \frac{x-a}{b-a} & a < x < b \\ 1 & x \geq b \end{cases}$$

□

Corollary 2.1. *Let X be a continuous standard uniform random variable. Then, its CDF will be defined as:*

$$F_X(x) = \begin{cases} 0 & x \leq 0 \\ x & 0 < x < 1 \\ 1 & x \geq 1 \end{cases} \tag{2.50}$$

2.3.4 Gamma Distributions

Definition 2.27. A **Gamma function ($\Gamma(\alpha)$)** is function defined as the following:

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx \tag{2.51}$$

and it holds for the followings:

- $\Gamma(\alpha) = (\alpha - 1) \times \Gamma(\alpha - 1)$
- If $n \in \mathbb{Z}^+$, $\Gamma(n) = (n-1)!$
- $\Gamma(1/2) = \sqrt{\pi}$

Definition 2.28. A continuous random variable follows a **gamma distribution** with parameters $\alpha > 0$ and $\beta > 0$ if its PDF is:

$$f_X(x) = \begin{cases} \frac{1}{\Gamma(\alpha)} \frac{1}{\beta^\alpha} x^{\alpha-1} e^{-x/\beta} & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases} \tag{2.52}$$

We denote $X \sim \Gamma(\alpha, \beta)$ if X have gamma distribution.

Proof. We will show that the above function for gamma distribution is is truly a PDF. First, $f(x) \geq 0$ by definition $\implies$ integration over the positive real interval would yield 1 if it's a PDF:

$$\begin{aligned} \int_0^\infty f(x) dx &= \int_0^\infty \frac{1}{\Gamma(\alpha)} \frac{1}{\beta^\alpha} x^{\alpha-1} e^{-x/\beta} dx \\ &= \frac{1}{\Gamma(\alpha)} \frac{1}{\beta^\alpha} \int_0^\infty x^{\alpha-1} e^{-x/\beta} dx \\ &\text{Let } x/\beta = y \implies dx/\beta = dy \\ &= \frac{1}{\Gamma(\alpha)} \frac{1}{\beta^\alpha} \int_0^\infty (\beta y)^{\alpha-1} e^{-y} dy \end{aligned}$$

$$= \frac{1}{\Gamma(\alpha)} \int_0^\infty (y)^{\alpha-1} e^{-y} dy$$
$$= \frac{1}{\Gamma(\alpha)} \cdot \Gamma(\alpha) = 1$$

Thus, $f(x)$ is a PDF. □

Theorem 2.16. *Let $X \sim \Gamma(\alpha, \beta)$. Then,*

$$E[X] = \alpha\beta \tag{2.53}$$
$$Var[X] = \alpha\beta^2 \tag{2.54}$$

Proof. Suppose that $X \sim \Gamma(\alpha, \beta)$. Then,

$$\begin{aligned} E[X] = \int_0^\infty x f(x) dx &= \int_0^\infty x \frac{1}{\Gamma(\alpha)} \frac{1}{\beta^\alpha} x^{\alpha-1} e^{-x/\beta} dx \\ &= \frac{1}{\Gamma(\alpha)} \frac{1}{\beta^\alpha} \int_0^\infty x x^{\alpha-1} e^{-x/\beta} dx \\ &= \frac{1}{\Gamma(\alpha)} \frac{1}{\beta^\alpha} \int_0^\infty x^\alpha e^{-x/\beta} dx \\ &\text{Let } x/\beta = y \implies dx/\beta = dy \\ &= \frac{1}{\Gamma(\alpha)} \frac{1}{\beta^\alpha} \int_0^\infty (\beta y)^\alpha e^{-y} dy \\ &= \frac{1}{\Gamma(\alpha)} \frac{1}{\beta} \int_0^\infty (y)^\alpha e^{-y} dy \\ &= \beta \frac{1}{\Gamma(\alpha)} \cdot \Gamma(\alpha+1) \\ &= \beta\alpha \frac{1}{\Gamma(\alpha)} \cdot \Gamma(\alpha) = \boxed{\alpha\beta} \end{aligned}$$

To find Var$[X]$, we will first determine, $E[X^2]$:

$$\begin{aligned} E[X] = \int_0^\infty x^2 f(x) dx &= \int_0^\infty x^2 \frac{1}{\Gamma(\alpha)} \frac{1}{\beta^\alpha} x^{\alpha-1} e^{-x/\beta} dx \\ &= \frac{1}{\Gamma(\alpha)} \frac{1}{\beta^\alpha} \int_0^\infty x^2 x^{\alpha-1} e^{-x/\beta} dx \\ &= \frac{1}{\Gamma(\alpha)} \frac{1}{\beta^\alpha} \int_0^\infty x^{\alpha+1} e^{-x/\beta} dx \\ &\text{Let } x/\beta = y \implies dx/\beta = dy \\ &= \frac{1}{\Gamma(\alpha)} \frac{1}{\beta^\alpha} \int_0^\infty (\beta y)^{\alpha+1} e^{-y} dy \end{aligned}$$

$$= \frac{1}{\Gamma(\alpha)}\beta^2 \int_0^\infty (y)^{\alpha+1} e^{-y} dy$$
$$= \frac{1}{\Gamma(\alpha)}\beta^2 \Gamma(\alpha+2) = \alpha(\alpha+1)\beta^2$$

Then, the variance is calculated as

$$\begin{aligned} \text{Var}[X] &= E[X^2] - (E[X])^2 \\ &= \alpha(\alpha+1)\beta^2 - \alpha^2\beta^2 \\ &= \alpha^2\beta^2 + \alpha\beta^2 - \alpha^2\beta^2 = \boxed{\alpha\beta^2} \end{aligned}$$

Thus, $E[X] = \alpha\beta$ and $\text{Var}[X] = \alpha\beta^2$. □

Remark 2.11. *The gamma distribution does not have a closed form because the integral from* 0 *to x, of the PDF, it will not have a closed form.*

2.3.5 Chi-Square Distributions

Definition 2.29. Let $X \sim \Gamma(\alpha, \beta)$ such that $\alpha = \nu/2$ and $\beta = 2$ where $\nu \in \mathbb{Z}^+$. Then, X is said to have **Chi-square (χ^2) distribution**, denoted as: $X \sim \chi^2(\nu)$, and its PDF is given as:

$$f_X(x) = \begin{cases} \frac{1}{\Gamma(\nu/2)} \frac{1}{2^{\nu/2}} x^{\nu/2-1} e^{-x/2} & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases} \tag{2.55}$$

Additionally, ν is also called the **number of degrees of freedom**.

Remark 2.12. *Since χ^2 distribution is typically used to develop hypothesis tests and confidence intervals, and rarely for modeling applications.*

Theorem 2.17. *Let $X \sim \chi^2(\nu)$. Then,*

$$E[X] = \nu \tag{2.56}$$
$$Var[X] = 2\nu \tag{2.57}$$

2.3.6 Exponential Distributions

Definition 2.30. Let $X \sim \Gamma(\alpha, \beta)$ such that $\alpha = 1$ and $\beta > 0$. Then, X is said to have **exponential distribution**, denoted as $X \sim \text{Exp}(\beta)$ with PDF:

$$f_X(x) = \begin{cases} \frac{1}{\beta} \cdot e^{-x/\beta} & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases} \tag{2.58}$$

End of Lecture

Theorem 2.18. *Let $X \sim Exp(\beta)$. Then,*

$$E[X] = \beta \tag{2.59}$$

$$Var[X] = \beta^2 \tag{2.60}$$

Theorem 2.19. *Let $X \sim Exp(\beta)$. Then, its CDF is given as:*

$$F_X(x) = \begin{cases} 0 & if\, x < 0 \\ 1 - e^{-x/\beta} & if\, x > 0 \end{cases} \tag{2.61}$$

Example 2.3.7. A manufacturing plant uses a specific bulk product. The amount of product used in one day can be modeled by an exponential distribution with $\beta = 4$ (measurements in tons). Find the probability that the plant will use more than 4 tons on a given day.

Solution: Let X be the amount of product use in a day such that $X \sim \text{Exp}(\beta)$. Then, we're asked to find $P(X > 4)$, which we can use the CDF from theorem 2.19:

$$\begin{aligned} P(X > 4) &= 1 - P(X \leq 4) \\ &= 1 - (1 - e^{-4/4}) \\ &= 1 - (1 - e^{-1}) = e^{-1} \approx \boxed{0.3679} \end{aligned}$$

Example 2.3.8. The magnitude of earthquakes recorded in a region of North America can be modeled as having an exponential distribution with mean 2.4, as measured on the Richter scale. Find the probability that an earthquake striking this region will:

1. exceed 3.0 on the Richter scale.
2. Fall between 2.0 and 3.0 on the Richter scale.

Solution: Let X the be magnited of earthquakes recorded in a region of NA such that $X \sim \text{Exp}(\beta)$. We're given $E[X] = 2.4 \implies \beta = 2.4$. Then, we're asked to find the followings:

1. $P(X > 3.0)$. Like before, we can change this to $1 - P(X \geq 3)$ and then use the CDF to calculate to get $\boxed{0.6873}$.
2. $P(2.0 < X < 3.0)$. We can find it as the followings:

$$\begin{aligned} P(2.0 < X < 3.0) &= P(X < 3.0) - P(X < 2.0) \\ &\approx 0.7135 - 0.5654 = \boxed{0.1481} \end{aligned}$$

Example 2.3.9. Historical evidence indicates that times between fatal accidents on scheduled American domestic passenger flights have an approximately exponential distribution. Assume that the mean time between accidents is 44 days. If one of the accidents occurred on July 1 of a randomly selected year in the study period, what is the probability that another accident occurred that same month? What is the variance of the times between accidents?

Solution: Let Y time between fatal airplane accidents such that $Y \sim \text{Exp}(\beta)$. We're given $E[Y] = 44 \implies \beta = 44$. We're asked to find $P(Y \leq 30)$ (another accident occured in the same month i.e. within the 30 days period before August). We can easily find this using CDF which will give $\boxed{0.4943}$. To find the variance, we can simply use equation (2.60), $\text{Var}[X] = \beta^2 = 44^2 = \boxed{1936}$.

Theorem 2.20. ***(Memoryless Property of Exponential Distribution).*** *If $X \sim Exp(\beta)$. Then,*

$$P(X \geq a + x \mid X \geq a) = P(X \geq x) \tag{2.62}$$

Proof. First, consider the 2 event, $X \geq a + x$ and $X \geq a$, we'd realized that $(X \geq a + x) \subset (X \geq a) \implies (X \geq a + x) \cap (X \geq) = (X \geq a + x)$. Then,

$$\begin{aligned}
P(X \geq a + x \mid X \geq a) &= \frac{P((X \geq a + x) \cap (X \geq))}{P(X \geq a)} \\
&= \frac{P(X \geq a + x)}{P(X \geq a)} \\
&= \frac{1 - P(X \leq a + x)}{1 - P(X \leq a)} \\
&= \frac{1 - (1 - e^{-x-a/\beta})}{1 - (1 - e^{-\alpha/\beta})} \\
&= e^{-x/\beta} \\
&= 1 - (1 - e^{-x/\beta}) = 1 - P(X \leq x) = P(X \geq x)
\end{aligned}$$

Thus, we've shown $P(X \geq a + x \mid X \geq a) = P(X \geq x)$. □

The memory-less property tells you about the conditional behavior of the exponential random variable. What it says is that if X is exponentially distributed with parameter β and you know that $X > a$ then the probability that X is also greater than some value $x + a$ is the same as $X \geq x$.
This means that the conditional probability $P(X \geq a + x \mid X \geq a)$ does not depend upon a. Hence, what happened earlier does not affect what will happen next if $X \text{exp}(\beta)$.

2.3.7 (Standard) Normal Distribution

The normal distribution is one of the most commonly used distributions.It plays a very important role in conducting statistical inference. A lot of real life phenomenon generate the distributions that are bell shaped (at least approximately).

Definition 2.31. A continuous random variable is said to have **normal distribution** with parameter μ and σ^2 if it has the following PDF:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{2.63}$$

where $-\infty < x < \infty$. If X has a normal distribution, we denote as $X \sim N(\mu, \sigma^2)$.

We cannot find closed-form expression for the integral of the Normal PDF. Therefore, we need numerical integration techniques to solve it. However, the probabilities and quantiles for a normal random variable can be found using online calculators, R and the probability table (which would be provided in exam).

Now, let's look at some properties of the normal distributions:

- μ (the mean) locates the center of the distribution.
- The shape of the distribution is determined by σ, the standard deviation.
- $\sigma \uparrow \implies$ curve's height$\downarrow$ and spread$\uparrow$. On the contrary. $\sigma \downarrow \implies$ curve's height$\uparrow$ and spread$\downarrow$.
- The normal distribution is perfectly symmetric about μ.
- We need to obtain the numerical value of μ and σ in order to graph the curve.

Remark 2.13. $\sim 99\%$ *of data for the normal distribution will fall between the interval* $[-3\sigma, 3\sigma]$.

Standard Normal Distribution

Definition 2.32. Let $X \sim N(\mu, \sigma^2)$ such that $\mu = 0$ and $\sigma^2 = 1$. Then, X is said to have **standard normal distribution**, denoted as $X \sim N(0,1)$. We can

convert the normal random variable X to the **standard random variable** Z through the following equation:

$$Z = \frac{X - \mu}{\sigma} \qquad X = \mu + \sigma Z \tag{2.64}$$

The process of transforming a normal to a standard normal random variable is called **standardization**.

The table below is that of the standard normal distribution with $P(Y \leq y)$.

Entries in table are $P(Y \leq y)$

y	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

Example 2.3.10. The intelligence quotient determined by the Stanford-Binet Intelligence Quotient Test is a continuous random variable that can be modeled by a normal distribution. Assume that for American population the IQs are normally distributed with $\mu = 100$ and $\sigma = 10$.

- What is the probability that a randomly selected American has an IQ below 90?
- What is the probability that a randomly selected American has an IQ above 120?
- What is the IQ of a randomly chosen American if their IQ is more than the IQ of 80% of the Americans.

Solution: Let X be a continuous random variable that denote the IQ of the American population ($X \sim N(\mu, \sigma^2)$). Then,

- We're asked to find $P(X < 90)$. To find this, we will first standardize the random variable then find the entries from the table provided above.

$$\begin{aligned} P(X < 90) &= P\left(\frac{X-\mu}{\sigma} < \frac{90-100}{10}\right) \\ &= P(Z < -1) \\ &= P(Z > 1) && \text{(by symmetry)} \\ &= 1 - P(Z \leq 1) \\ &= 1 - 0.8413 = \boxed{0.1587} \end{aligned}$$

- We're asked to find $P(X > 120)$. Then,

$$\begin{aligned} P(X > 120) &= P\left(\frac{X-\mu}{\sigma} > \frac{120-100}{10}\right) \\ &= P(Z > 2) \\ &= 1 - P(Z \leq 2) \\ &= 1 - 0.9772 = \boxed{0.0228} \end{aligned}$$

- We're asked to find $P(X < x) = 0.8$. Then,

$$P(X < x) = 0.8$$

$$\begin{aligned} P(X \le x) &= 0.8 \\ P\left(\frac{X-\mu}{\sigma} < \frac{x-\mu}{\sigma}\right) &= 0.8 \\ P(Z \le z) &= 0.8 \\ \implies z &\approx 0.845 \end{aligned}$$

Then, we will convert this standard normal value back to the normal value as $x = 100 + 10(0.845) = \boxed{108.45}$.

Example 2.3.11. Wires manufactured for use in a computer system are specified to have resistances between 0.12 and 0.14 ohms. The actual measured resistances of the wires produced by company A have a normal probability distribution with mean 0.13 ohm and standard deviation 0.005 ohm.

1. What is the probability that a randomly selected wire from company A's production will meet the specifications.

2. If four of these wires are used in each computer system and all are selected from company A, what is the probability that all four in a randomly selected system will meet the specifications?

Solution: Let X be the continuous normal random variable denote the resistance of the wires by company A. Then,

1. We're asked to find $P(0.12 < X < 0.14)$. Then,

$$\begin{aligned} P(0.12 < X < 0.14) &= P(X \le 0.12) - P(X \le 0.14) \\ &= P\left(\frac{X-\mu}{\sigma} \le \frac{0.14-0.13}{0.005}\right) - P\left(\frac{X-\mu}{\sigma} \le \frac{0.12-0.13}{0.005}\right) \\ &= P(Z \le 2) - P(Z \le -2) \\ &= P(Z \le 2) - 1 + P(Z \le 2) = 2P(Z \le 2) - 1 = \boxed{0.994} \end{aligned}$$

2. In this case, we can use binomial distribution to find the probability that 4 wires in the system meets qualification. Let p be the probability of success where $p = 0.994$. Let Y be the number of wires that meet qualification. Then,

$$P(Y = 4) = \binom{4}{4}(0.994)^4(1-0.994)^0 \approx \boxed{0.9762}$$

Example 2.3.12. A soft-drink machine can be regulated so that it discharges an average of μ ounces per cup. If the ounces of fill are normally distributed with standard deviation 0.3 ounce, give the setting for μ so that 8-ounce cups will overflow only 1% of the time.

Solution: Let X be the continuous normal random variable denote the amount of soft-drink discharge. we're trying to find μ, given that $P(X > 8) = 0.01$. Then,

$$\begin{aligned} P(X > 8) &= 0.01 \\ P\left(\frac{X-\mu}{\sigma} > \frac{8-\mu}{\sigma}\right) &= 0.01 \\ P(Z > z) &= 0.01 \\ P(Z \leq z) &= 0.99 \\ \implies z &= 2.33 \end{aligned}$$

Now, since $z = \frac{x-\mu}{\sigma} \implies \mu = z\sigma - x = 8 - 2.33(0.3) = \boxed{7.3 \text{ ounces}}$.

3 Functions of Random Variables

In some situations we are interested in a function of a random variable where we know the PDF or CDF of our random variable. The function of a random variable is another random variable which is a transformation of X into a new variable.

Example 3.0.1. We observe V and know its probability distribution, where V is the velocity of a particle (and therefore is a random variable). However, we are interested in the probability distribution of its Kinetic Energy $K.E = 1/2mV^2 2$ (which is a random variable as it is the function of a random variable)

Definition 3.1. The process of using a function of a random variable in place of a random variable itself is known as **transformation**.

Transformation is especially important for many reasons and here are some of the highlights:

1. It's convenient to use a transformed random variable than using the original one.
2. The transformation e.g. $\ln X$ and $\sqrt{X}$ help to reduce skewness in the data.
3. Transformations can used to achieve homoscedasticity.

There are 3 commonly used methods to find the probability of a function of random variables:

1. The method of Distribution Functions
2. The method of Transformations
3. The method of moment generating functions.

3.1 The Method of Distribution Functions

Remark 3.1. *This methods is also known as the CDF method*

Method: let X be a continuous random variable and $Y = f(X)$. Then, Y is also a continuous random variable. To find the distribution of Y it consists of 2 steps for this method: find CDF of Y and differentiate the CDF to find its PDF.

Example 3.1.1. Let X be a random variable with the PDF:

$$f_X(x) = \begin{cases} 2e^{-2x} & \text{for } x > 0 \\ 0 & \text{otherwise} \end{cases}$$

Find $Y = \sqrt{X}$.

Solution: Notice that X has a PDF that is of exponential distribution i.e. $X \sim \text{Exp}(\beta)$ where $2 = \frac{1}{\beta} \implies \beta = \frac{1}{2}$. Notice also that $X \geq 0$ and also by construction it must be the case that $Y \geq 0$[1] So now, we find the CDF of Y:

$$\begin{aligned} P(Y \leq y) &= P(\sqrt{X} \leq y) = P(X \leq y^2) \\ &\implies F_Y(y) = F_X(y^2) \\ &= \int_0^{y^2} 2e^{-2x}dx = e^{-2x}\Big|_0^{y^2} = 1 - e^{-2y^2} \end{aligned}$$

Now, we will differentiate this to get the PDF:

$$\begin{aligned} \frac{d}{dy}F_Y(y) &= \frac{d}{dy}F_X(y^2) \\ f_Y(y) &= \frac{d}{dy}\left[1 - e^{-2y^2}\right] = -e^{-2y^2}(-4y) = \boxed{4ye^{-2y^2}} \end{aligned}$$

And thus, we get the PDF of Y as:

$$f_Y(y) = \begin{cases} 4ye^{-2y^2} & \text{for } y > 0 \\ 0 & \text{otherwise} \end{cases}$$

Example 3.1.2. The amount of sugar produced by a manufacturing plant is a random variable denoted by Y with the PDF:

$$f_Y(y) = \begin{cases} 2y & 0 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Let U be the profit given by $U = 2Y - 1$, find the PDF of U.

[1] Any number below square root that yield a real number must be positive.

Solution: We will first find the CDF of U:

$$P(U \leq u) = P(2Y - 1 \leq u) = P\left(Y \leq \frac{u+1}{2}\right)$$

$$\implies F_U(u) = F_Y\left(\frac{u+1}{2}\right) = \int_0^{\frac{u+1}{2}} 2y\,dy = y^2\Big|_0^{\frac{u+1}{2}} = \left(\frac{u+1}{2}\right)^2$$

Notice that we will now have to determine the support for this random variable which can be done using that of y. That is, $0 \leq y \leq 1 \implies 0 \leq \frac{u+1}{2} \leq 0 \implies -1 \leq u \leq 2$. Now, we we've obtained the CDF, we will integrate it to yield the PDF:

$$\frac{d}{du}F_U(u) = \frac{d}{du}F_Y\left(\frac{u+1}{2}\right)$$

$$= \frac{d}{du}\left[\left(\frac{u+1}{2}\right)^2\right] = \frac{2}{9}(u+1)$$

Thus, we obtain the PDF of U as:

$$f_U(u) = \begin{cases} \frac{2}{9}(u+1) & -1 \leq u \leq 2 \\ 0 & \text{otherwise} \end{cases} \tag{3.1}$$

End of Lecture

3.2 The Method of Transformations

Remark 3.2. *This method is also called the "Change of Variable Technique".*

Method: Let X be a continuous variable with CDF $F_X(x)$ and PDF $f_X(x)$. Let $Y = g(X)$ for some function g. Then:

1. If g is strictly increasing[2] then we proceed as follows:
   - Write the CDF of Y in terms of the CDF of X.

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

   - Differentiate the newly found CDF with respect to y.

     $$f_Y(y) = \frac{d}{dx}F_Y(y) = \frac{d}{dy}F_X(g^{-1}(y)) = F_X'(g^{-1}(y))\left[\frac{d}{dx}g^{-1}(y)\right]$$

     $$= F_X'(g^{-1}(y)) \cdot \frac{1}{\frac{dy}{dx}}$$

[2] Which means if $x_1 > x_2 \implies g(x_1) > g(x_2)$.

2. If g is a decreasing function[3]. Then, we will follows similar steps of as above but will yield a PDF:

$$f_Y(y) = F_X(g^{-1}(y)) \cdot -\frac{1}{\frac{dy}{dx}}$$

Thus, combining these two scenarios, we have the general formula for either increasing or decreasing PDF of Y as:

$$f_Y(y) = F'_X(g^{-1}(y)) \cdot \frac{1}{\left|\frac{dy}{dx}\right|} \tag{3.2}$$

Example 3.2.1. Let $Y = \sqrt{X}$ where the PDf of X is:

$$f_X(x) = \begin{cases} 2e^{-2x} & \text{for } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Find the PDF of Y.

Solution: This is an increasing function thus we follows scenario 1, with step 1 giving us: $F_Y(y) = F_X(y^2)$ and thus the PDF is given as:

$$\begin{aligned} f_Y(y) &= F_X(y^2) \cdot \frac{d}{dx}(y^2) \\ &= 2e^{-2y^2} \cdot 2y = \boxed{4ye^{-2y^2}} \end{aligned}$$

Example 3.2.2. Use the transformation method to find the PDF of $Z = \frac{X-\mu}{\sigma}$ where

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

Solution: First, we will write the CDF of Z in term of X:

$$F_Z(Z) = P(Z \leq z) = P(X \leq \mu + \sigma z) = F_X(\mu + \sigma z)$$

Second, we will differentiate it with respect of z:

$$\begin{aligned} f_Z(z) &= F'_X(\mu + \sigma z) \cdot \frac{d}{dz}(\mu + \sigma z) \\ &= \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma} e^{-\frac{1}{2}\left(\frac{\mu+\sigma z-\mu}{\sigma}\right)^2} \cdot \sigma = \boxed{\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}} \end{aligned}$$

[3] which means if $x_1 < x_2 \implies g(x_1) > g(x_2)$.

where $-\infty < z < \infty$. Remember that we've said that $Z \sim N(0,1)$ and indeed, if we were to plug these number in for σ and μ, we would get what we've just calculated using transformation.

Example 3.2.3. (Important Result). Let $Z \sim N(0,1)$, then, show that $Z^2 \sim \chi^2(1)$.

Proof. First, let $Y = Z^2$ then its CDF will be defined as: [4]

$$\begin{aligned} F_Y(y) = P(Y \leq y) &= P(Z^2 \leq y) \\ &= P(-\sqrt{y} \leq Z \leq \sqrt{y}) \\ &= P(Z \leq \sqrt{y}) - P(Z \leq -\sqrt{y}) \\ &= F_Z(\sqrt{y}) - F_Z(-\sqrt{y}) \end{aligned}$$

Now, we will differentiate it with respect to Y:

$$\begin{aligned} f_Y(y) &= F'_Z(\sqrt{y}) \cdot \frac{d}{dy}(\sqrt{y}) - F'_Z(-\sqrt{y}) \cdot \frac{d}{dy}(-\sqrt{y}) \\ &= \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\sqrt{y}^2} \cdot \frac{1}{2} y^{-\frac{1}{2}}\right) - \left(-\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\sqrt{y}^2} \cdot -\frac{1}{2} y^{-\frac{1}{2}}\right) \\ &= \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\sqrt{y}^2} \cdot \frac{1}{2} y^{-\frac{1}{2}}\right) + \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\sqrt{y}^2} \cdot \frac{1}{2} y^{-\frac{1}{2}}\right) \\ &= \frac{1}{2} y^{-\frac{1}{2}} \cdot 2\left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\sqrt{y}^2}\right) \\ &= \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sqrt{y}} e^{-\frac{1}{2}y} \end{aligned}$$

For $y \geq 0$ since $Y = Z^2$ where $-\infty < z < \infty$. Now, we can rewrite the above as follows:

$$\begin{aligned} \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sqrt{y}} e^{-\frac{1}{2}y} &= \frac{1}{\sqrt{\pi}} \frac{1}{2^{1/2}} \cdot y^{1/2-1} e^{-y/2} \\ &= \frac{1}{\Gamma(1/2)} \frac{1}{2^{1/2}} \cdot y^{1/2-1} e^{-y/2} \end{aligned}$$

$\implies \nu = 1$ and $\therefore Y = Z^2 \sim \chi^2(1)$. □

Theorem 3.1. *Let X be a continuous random variable with CDF $F_X(x)$. Then, the function $F_X(X)$ is uniformly distributed i.e.*

$$F_X(X) \sim \mathit{Unif}(0,1) \tag{3.3}$$

[4] $Z^2 \leq y \implies |Z| \leq \sqrt{y} \implies -\sqrt{y} \leq Z \leq \sqrt{y}$

Remark 3.3. *$F_X(X)$ is not a cumulative probability where as $F_X(x)$ is which means $F_X(X) \neq P(X \leq x)$. $F_X(X)$ is a transformation*

Example 3.2.4. Let X be a continuous random variable with PDF given below:

$$f_X(x) = \begin{cases} 3x^2 & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

Show that $F_X(X) \sim \text{Unif}(0,1)$.

Proof. First, we need to determine $F_X(x)$ which is

$$F_X(x) = \int_0^x 3t^2\,dt = t^3\Big|_0^x = x^3$$

Then, $F_X(X) = X^3$ (a new function of X). Let $U = X^3$ (a transformation). We will now show that $U \sim \text{Unif}(0,1)$. First, we will find its CDF:

$$F_U(u) = P(U \leq u) = P(X^3 \leq u) = P(X \leq \sqrt[3]{u}) = F_X(u^{1/3})$$

Now, we will differentiate it with respect to u:

$$\begin{aligned} f_U(u) &= F_X'(u^{1/3}) \cdot \frac{d}{du} u^{1/3} \\ &= 3(u^{1/3})^2 \cdot \frac{1}{3} u^{-2/3} \\ &= u^{2/3} \cdot u^{-2/3} = 1 \end{aligned}$$

Thus, $f_U(u) = 1$ and $\therefore U = F_X(X) \sim \text{Unif}(0,1)$. □

Proof. (Theorem 3.1). Let X be a continuous random variable and $P(X \leq x) = F_X(x)$. We get a new function of X from the CDF by replacing x by X. Let $U = F_X(X)$. We want to find PDF of U. First, we have the following:

$$\begin{aligned} F_U(u) = P(U \leq u) = P(F_X(X) \leq u) &= P(X \leq F_X^{-1}(u)) \\ &= F_X(F_X^{-1}(u)) = u \end{aligned}$$

We will then integrate this:

$$f_U(u) = \frac{d}{du} u = 1$$

This is the PDf of a uniform random variable over the interval $[0,1]$ and $\therefore U = F_X(X) \sim \text{Unif}(0,1)$. □

End of Lecture

3.3 The Method of Moment Generating Function

Before talking about the method of moment generating function, let's first talk about the moment generating function itself.

3.3.1 Moment Generating Function

A moment generating function is a function of real number t. It's a another way of describing probability distribution.

Definition 3.2. A **moment generating function (mgf)** of the (distribution of the) random variable X is the function of a real parameter t defined by

$$M_X(t) = E\left[e^{tX}\right] \tag{3.4}$$

for all $t \in \mathbb{R}$ for which the expectation $E[e^{tX}]$ is well-defined.

From the definition of mgf, we have:

1. For the discrete random variables,

$$M_X(t) = E[e^{tX}] = \sum_{\text{all } x} e^{tx} P(X = x) \tag{3.5}$$

where $P(X = x)$ is the PMF of X.

2. For continuous random variables,

$$M_X(t) = E[e^{tX}] = \int_{-\infty}^{\infty} e^{tx} f_X(x)\,dx \tag{3.6}$$

where $f_X(X)$ is the PDF of X.

N.B. the PDF of PMF of the random variable X can be obtained from its mgf and v.v.

Result

Let X be a continuous random variable. Then, $M_X(t) = E[e^{tX}] = \int_{-\infty}^{\infty} e^{tx} f_X(x)\,dx$ and thus

$$\begin{aligned} M'_X(t) &= \frac{d}{dt}\int_{-\infty}^{\infty} e^{tx} f_X(x)\,dx \\ &= \int_{-\infty}^{\infty} e^{tx} x f_X(x)\,dx \end{aligned}$$

Notice that the first derivative of $M_X(t)$ evaluate at 0 will yield the first moment about the origin i.e. the mean:

$$M'_X(t)\Big|_{t=0} = \int_{-\infty}^{\infty} e^{0x} x f_X(x) \int_{-\infty}^{\infty} x f_X(x) = E[X]$$

Similarly, the second derivative at 0 will give $E[X^2]$

$$M''_X(t)\Big|_{t=0} = E[X^2]$$

we can thus generalize this into the following equation:

$$\frac{d^k}{dt^k} M_X(t)\Big|_{t=0} = M_X^{(k)}(t)\Big|_{t=0} = E[X^k] \tag{3.7}$$

Example 3.3.1. Let $X \sim \text{Binom}(n, p)$. Find its mgf and using the mgf to derive its expectation, variance and $E[X^2]$.

Solution: To find the mgf, we use the above equations for discrete random variable:

$$\begin{aligned} M_X(t) = E[e^{tX}] &= \sum_{x=0}^{n} e^{tx} \binom{n}{x} p^x (1-p)^{n-x} \\ &= \sum_{x=0}^{n} \binom{n}{x} (pe^t)^x (1-p)^{n-x} \\ &= \boxed{\left(pe^t + 1 - p\right)^n} \end{aligned}$$

Now, let's determine the expectation using this mgf:

$$\begin{aligned} E[X] &= M'_X(t)\Big|_{t=0} \\ &= \frac{d}{dt}\left(pe^t + 1 - p\right)^n \\ &= n\left(pe^t + 1 - p\right)^{n-1} pe^t\Big|_{t=0} \\ &= n(p + 1 - p)^{n-1} p = \boxed{np} \end{aligned}$$

and

$$\begin{aligned} E[X^2] &= M''_X(t)\Big|_{t=0} \\ &= n^2 p^2 - np^2 + np \end{aligned}$$

Thus, the variance is given as

$$\begin{aligned}\text{Var}[X] &= E[X^2] - (E[X])^2 \\ &= n^2p^2 - np^2 + np - n^2p^2 = \boxed{np(1-p)}\end{aligned}$$

Exercise 1. Let $X \sim \text{Poiss}(\lambda)$ where $P(X=x) = \frac{e^{\lambda}\lambda^x}{x!}$ and $x \in \{0,1,2,\ldots\}$. Find $M_X(t)$ and use the found mgf to determine its $E[X]$, $E[X^2]$ and $\text{Var}[X]$.
Hint: Maclaurin's series expansion of the exponential function.

Exercise 2. Let $X \sim \Gamma(\alpha,\beta)$. Show that $M_X(t) = \frac{1}{(1-\beta t)^{\alpha}}$ where $1-\beta t>0$
Hint: Set up the integral like Gamma function.

3.3.2 Moment Generating Function: Normal Distribution

To find the mgf of normal distribution, we first start with $Z \sim N(0,1)$. Then, mgf will be defined as:

$$\begin{aligned}M_Z(t) = E[e^{tZ}] &= \int_{-\infty}^{\infty} e^{tz} f_Z(z)\,dz \\ &= \int_{-\infty}^{\infty} e^{tz} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}\,dz \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tz-\frac{1}{2}z^2}\,dz \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(z-t)^2+\frac{1}{2}t^2}\,dz \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(z-t)^2} \underbrace{e^{\frac{1}{2}t^2}}_{\text{constant}}\,dz \\ &= e^{\frac{1}{2}t^2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}{}^{-\frac{1}{2}(z-t)^2}\,dz \\ &= e^{\frac{1}{2}t^2}\end{aligned}$$

Theorem 3.2. *Let X be a random variable with mgf $M_X(t)$ and $Y = aX+b$. Then,*

$$M_Y(t) = e^{bt} M_X(at) \tag{3.8}$$

Proof. Using the definition, we get:

$$M_Y(t) = E\left[e^{Yt}\right] = E\left[e^{(aX+b)t}\right] = E\left[e^{aXt+bt}\right]$$

End of Lecture

$$= E\left[e^{aXt}e^{bt}\right]$$
$$= e^{bt}E\left[e^{aXt}\right] = e^{bt}M_X(at)$$

$\therefore M_Y(t) = e^{bt}M_X(at)$. □

Theorem 3.3. ***(mgf of Normal Distribution).*** *Let $X \sim N(\mu,\sigma^2)$ and $X = \mu + \sigma Z$ where $Z \sim N(0,1)$. Then,*

$$M_X(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2} \tag{3.9}$$

4 Multivariate Probability Distributions

A multivariate distribution describes the joint behaviour or two or more random variables.

Example 4.0.1. Consider tossing 2 fair coins. Let Y_1 be the random variable that denote the number of head in the first coin while Y_2 be the random variable that denote the number of head in the second coin. It's evident that both would have $\{0,1\}$ as support. Then, you'd have the sample space as: $\{(0,0)(1,0)(0,1)(1,1)\}$. Then, we can find the probability of this sample space as

$$P(Y_1 = y_1 \cap Y_2 = y_2)$$

where $y_1 \in \{0,1\}$ and $y_2 \in \{0,1\}$. This probability basically tell us the probability of having simultaneous occurrence of 2 events i.e. joint behaviour of Y_1 and Y_2. This is type of probability called the **joint probability distribution**. In fact, if we were to write out its distribution, we'd get:

y_1	y_2	$P(Y_1 = y_1 \cap Y_2 = y_2)$
1	0	1/4
0	1	1/4
0	0	1/4
1	1	1/4

4.1 Joint Probability Distributions

Definition 4.1. Let Y_1 and Y_2 be discrete random variables. The **joint (or bivariate) probability function** for Y_1 and Y_2 is given by:

$$p(y_1, y_2) = P(Y_1 = y_1 \cap Y_2 = y_2) = P(Y_1 = y_1, Y_2 = y_2) \tag{4.1}$$

where $y_1, y_2 \in \mathbb{R}$.

Theorem 4.1. *If Y_1 and Y_2 are discrete random variables with joint pmf $p(y_1, y_2)$. Then,*

- $p(y_1, y_2) \geq 0 \; \forall y_1, y_2$.
- $\sum_{y_1, y_2} p(y_1, y_2) = 1$

Once the joint pmf has been determined, it becomes straight forward to compute the probabilities of events involving Y_1 and Y_2.

Example 4.1.1. Roll 2 fair dice and defined:

- Y_1 : the number appearing on die 1
- Y_2 : the number appearing on die 2.

Then, $p(y_1, y_2) = 1/36$ since there are 36 different combination of (y_1, y_2) which all have equal chances of coming up (fair). This means $P(Y_1 = 2, Y_2 = 3) = 1/36$. We can also determine the joint probability over a range of value, let's say $2 \le Y_1 \le 3$ and $1 \le Y_2 \le 2$. Then,

$$\begin{aligned} P(2 \le Y_1 \le 3, 1 \le Y_2 \le 2) &= p(2,1) + p(2,2) + p(3,1) + p(3,2) \\ &= 4 \times \frac{1}{36} = \boxed{\frac{1}{9}} \end{aligned}$$

Definition 4.2. If Y_1 and Y_2 are jointly discrete random variable. Then, the **joint CDF** of Y_1 and Y_2 is the function given as

$$\begin{aligned} F_{Y_1,Y_2}(y_1, y_2) &= P(Y_1 \le y_1, Y_2 \le y_2) && (4.2) \\ &= \sum_{t_1 \le y_1} \sum_{t_2 \le y_2} P_{Y_1,Y_2}(t_1, t_2) && (4.3) \end{aligned}$$

where $P_{Y_1,Y_2}(y_1, y_2) = P(Y_1 = y_1, Y_2 = y_2)$

Example 4.1.2. Consider the same rolling 2 fair dice as example 4.1.1. Then, Find $F_{Y_1,Y_2}(2,3)$.

Solution: Using the definition of joint CDF as above, we get that

$$\begin{aligned} F_{Y_1,Y_2}(2,3) &= P(Y_1 \le 2, Y_2 \le 3) \\ &= p(1,1) + p(1,2) + p(1,3) + p(2,1) + p(2,2) + p(2,3) \\ &= 6 \times \frac{1}{36} = \boxed{\frac{1}{6}} \end{aligned}$$

Definition 4.3. Let Y_1 and Y_2 be 2 jointly continuous random variable. Then, their **joint CDF** $F_{Y_1,Y_2}(y_1, y_2)$ is continuous in both y_1 and y_2; it also satisfies that

$$F_{Y_1,Y_2}(y_1, y_2) = P(Y_1 \le y_1, Y_2 \le y_2)$$

$$= \int_{-\infty}^{y_1} \int_{-\infty}^{y_2} f_{Y_1,Y_2}(t_1,t_2)\,dt_2\,dt_1$$

where $f_{Y_1,Y_2}(t_1,t_2)$ is the **joint PDF**. Geometrically and more generally, you can interpret $P(Y_1,Y_2)$ as the volume underneath the surface of which you've defined using the PDF. i.e.

$$P(Y_1,Y_2) \in \iint_A f_{Y_1,Y_2}(y_1,y_2)\,dy_2\,dy_1 \tag{4.4}$$

Example 4.1.3. Consider the joint PDF:

$$f(Y_1,Y_2) = \begin{cases} 1, & 0 \le y_1 \le 1 \quad 0 \le y_2 \le 1 \\ 0, & \text{otherwise} \end{cases}$$

Find $F(0.2,0.4)$.

Solution: Using the definition of CDF and PDF of joint continuous distribution, we get:

$$\begin{aligned} F(0.2,0.4) &= P(Y_1 \le 0.2, Y_2 \le 0.4) \\ &= \int_{-\infty}^{0.2} \int_{-\infty}^{0.4} f_{Y_1,Y_2}(t_1,t_2)\,dt_2\,dt_1 \\ &= \int_{0}^{0.2} \int_{0}^{0.4} f_{Y_1,Y_2}(t_1,t_2)\,dt_2\,dt_1 \\ &= \int_{0}^{0.2} \int_{0}^{0.4} 1\,dt_2\,dt_1 \\ &= \int_{0}^{0.2} 0.4\,dt_1 = \boxed{0.08} \end{aligned}$$

4.2 Marginal Probability Distribution

Definition 4.4. Let Y_1 and Y_2 be jointly discrete random variable with joint pmf $p(y_1,y_2)$. Then, the **marginal pmfs** of Y_1 and Y_2 are defined as:

$$P(Y_1 = y_1) = \sum_{y_2} p(y_1,y_2) \qquad P(Y_2 = y_2) = \sum_{y_1} p(y_1,y_2) \tag{4.5}$$

Definition 4.5. Let Y_1 and Y_2 be jointly continuous random variable with joint PDF $f_{Y_1,Y_2}(y_1,y_2)$. Then, the **marginal PDFs** of Y_1 and Y_2 are defined as:

$$f_{Y_1}(y_1) = \int_{-\infty}^{\infty} f_{Y_1,Y_2}(y_1,y_2)\,dy_2 \qquad f_{Y_2}(y_2) = \int_{-\infty}^{\infty} f_{Y_1,Y_2}(y_1,y_2)\,dy_1 \tag{4.6}$$

Remark 4.1. *So basically, marginal probability is the probability where we fix 1 variable as constant while varies the other i.e. you'd able to see the effect of the variable the vary since the other constant is fixed.*

Example 4.2.1. Consider the joint PDF:

$$f(y_1, y_2) = \begin{cases} 2y_1, & 0 \le y_1 \le 1 \quad 0 \le y_2 \le 1 \\ 0, & \text{otherwise} \end{cases}$$

Find the marginal PDF of Y_1 and Y_2.

Solution: Using the definition of marginal PDF, we will begin with finding that of Y_1:

$$\begin{aligned} f_{Y_1}(y_1) &= \int_{-\infty}^{\infty} f(y_1, y_2) dy_2 \\ &= \int_{-\infty}^{0} f(y_1, y_2) dy_2 + \int_{0}^{1} f(y_1, y_2) dy_2 + \int_{0}^{\infty} f(y_1, y_2) dy_2 \\ &= \int_{-\infty}^{0} 0 dy_2 + \int_{0}^{1} 2y_1 dy_2 + \int_{0}^{\infty} 0 dy_2 \\ &= 2y_1 \int_{0}^{1} dy_2 = 2y_1(1-0) = 2y_1 \end{aligned}$$

Now, consider the "otherwise" case also, we will get the marginal PDF of Y_1 as:

$$f_{Y_1}(y_1) = \begin{cases} 2y_1, & 0 \le y_1 \le 1 \\ 0, & \text{otherwise} \end{cases}$$

Similarly, we can follow the same process to find marginal PDF of Y_1:

$$\begin{aligned} f_{Y_2}(y_2) &= \int_{-\infty}^{\infty} f(y_1, y_2) dy_1 \\ &= \int_{-\infty}^{0} f(y_1, y_2) dy_1 + \int_{0}^{1} f(y_1, y_2) dy_1 + \int_{0}^{\infty} f(y_1, y_2) dy_1 \\ &= \int_{-\infty}^{0} 0 dy_1 + \int_{0}^{1} 2y_1 dy_1 + \int_{0}^{\infty} 0 dy_1 \\ &= \frac{2y_1^2}{2}\Big|_0^1 = 1^2 - 0^2 = 1 \end{aligned}$$

Now, consider the "otherwise" case also, we will get the marginal PDF of Y_2 as:

$$f_{Y_2}(y_2) = \begin{cases} 1, & 0 \le y_2 \le 1 \\ 0, & \text{otherwise} \end{cases}$$

4.3 Conditional Probability Distribution

Definition 4.6. Let Y_1 and Y_2 be jointly discrete random variable. Then, the **conditional probability** of Y_1 given that Y_2 has happen is defined as:

$$P_{Y_1|Y_2=y_2}(y_1 \mid y_2) = \frac{P(Y_1 = y_1 \cap Y_2 = y_2)}{P(Y_2 = y_2)} \tag{4.7}$$

Definition 4.7. Let Y_1 and Y_2 be jointly continuous random variable. Then, the **conditional probability** of Y_1 given that Y_2 has happen is defined as:

$$f_{Y_1|Y_2=y_2}(y_1 \mid y_2) = \frac{f(y_1, y_2)}{f_{Y_2}(y_2)} \tag{4.8}$$

Example 4.3.1. Consider the same example as example 4.2.1. where the joint PDF is

$$f(Y_1, Y_2) = \begin{cases} 2y_1, & 0 \le y_1 \le 1 \quad 0 \le y_2 \le 1 \\ 0, & \text{otherwise} \end{cases}$$

with its marginal distribution as $f_{Y_1}(y_1) = 2y_1$ and $f_{Y_2}(y_2) = 1$. Then, suppose that $Y_2 = 0.5$, the conditional probability of Y_1 given Y_2 happened is

$$f_{Y_1|Y_2=0.5}(y_1 \mid y_2) = \frac{f(y_1, y_2)}{f_{Y_2}(y_2)} = \frac{2y_1}{1} = 2y_1$$

If $Y_1 = 0.5$ then the conditional probability of Y_2 given Y_1 happened is

$$f_{Y_2|Y_1=0.5}(y_2 \mid y_2) = \frac{f(y_1, y_2)}{f_{Y_1}(y_1)} = \frac{2(0.5)}{2(0.5)} = 1$$

Example 4.3.2. A soft-drink machine has a random amount Y_2 (in gallons) in supply at the beginning of a given day and dispenses a random amount Y_1 during the day, with the condition $Y_1 \le Y_2$. It is given that Y_1 and Y_2 have the joint density function:

$$f(Y_1, Y_2) = \begin{cases} \frac{1}{2}, & 0 \le y_1 \le y_2 \le 2 \\ 0, & \text{otherwise} \end{cases}$$

1. Find the conditional probability of Y_1 given $Y_2 = y_2$.
2. Evaluate the probability that less that $\frac{1}{2}$ gallon is sold, given that the machine contains 1.5 gallons at the start of the day.

End of Lecture

Solution: We begin by finding the marginal PDF of Y_2 as we will need it to determine the conditional probability of Y_1. The marginal PDF of Y_2 is given as

$$f_{Y_2}(y_2) = \int_{-\infty}^{\infty} f(y_1, y_2)dy_1 = \int_0^{y_2} f(y_1, y_2)dy_1 = \int_0^{y_2} \frac{1}{2}dy_1 = \frac{1}{2}y_2$$

Next, we will determine the conditional probability of Y_1 given that $Y_2 = y_2$ which is given as

$$f_{Y_1|Y_2=y_2}(y_1 \mid y_2) = \frac{f(y_1, y_2)}{f_{Y_2}(y_2)} = \frac{1/2}{y_2/2} = \boxed{\frac{1}{y_2}}$$

where $0 < y_2 \leq 2$.
We will now determine the probability as outlined in question 2 which is as follows:

$$\begin{aligned} P\left(Y_1 < \frac{1}{2} \middle| Y_2 = 1.5\right) &= \int_0^{1/2} f_{Y_1|Y_2=1.5}(y_1 \mid y_2)dy_1 \\ &= \int_0^{1/2} \frac{1}{1.5}dy_1 \\ &= \frac{1}{1.5}\int_0^{1/2} dy_1 = \frac{1}{1.5}\cdot\frac{1}{2} = \boxed{\frac{1}{3}} \end{aligned}$$

Remark 4.2. *You may see questions similar like $P(0.5 \leq Y_1 \leq 1.5 \mid Y_2 = 1.5)$. In such case, simply integrate over the given range with respect to y_1.*

4.4 Expectation and Variance

Definition 4.8. Let $g(Y_1, Y_2)$ be some function of Y_1 and Y_2. Then, the expectation of this function is given as

$$E[g(Y_1, Y_2)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(Y_1, Y_2)f_{Y_1,Y_2}(y_1, y_2)dy_1dy_2 \tag{4.9}$$

where Y_1, Y_2 are continuous random variables. If they're discrete, then the expectation is given as

$$E[g(Y_1, Y_2)] = \sum_{y_1}\sum_{y_2} g(Y_1, Y_2)P(Y_1 = y_1, Y_2 = y_2) \tag{4.10}$$

Note: for any of the continuous random variable Y_i (in the bivariate system), its expectation is given by:

$$E[Y_i] = \int_{-\infty}^{\infty} y_i f_{Y_i}(y_i) dy_i \tag{4.11}$$

and similarly its moment is

$$E[Y_i^k] = \int_{-\infty}^{\infty} y_i^k f_{Y_i}(y_i) dy_i \tag{4.12}$$

where $i = \{1,2\}$.

Remark 4.3. *From the expectation and also the second moment, one can directly realize the variance.*

Example 4.4.1. Let Y_1 and Y_2 have joint PDF given by

$$f(y_1,y_2) = \begin{cases} 2y_1 & 0 \le y_1 \le 1 \quad 0 \le y_2 \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Find $E[Y_1]$, $E[Y_2]$, $\text{Var}[Y_1]$ and $\text{Var}[Y_2]$.
Solution: We can find these directly from the definition. First, we will find the marginal PDFs, this can be obtained via example 4.2.1 from previous lecture.

$$\begin{aligned} E[Y_1] &= \int_{-\infty}^{\infty} y_1 f_{Y_1}(y_1) dy_1 \\ &= \int_0^1 y_1 2y_1 dy_1 \\ &= \int_0^1 2y_1^2 dy_1 = \frac{2y_1^3}{3}\Big|_0^1 = \boxed{\frac{2}{3}} \end{aligned} \qquad \begin{aligned} E[Y_2] &= \int_{-\infty}^{\infty} y_2 f_{Y_2}(y_2) dy_2 \\ &= \int_0^1 y_2 dy_2 = \frac{y_2^2}{2}\Big|_0^1 = \boxed{\frac{1}{2}} \end{aligned}$$

To determine the variance, we first determine $E[Y_i^2]$ (Try it yourself!). Then, we will get that $\text{Var}[Y_1] = \frac{1}{18}$ and $\text{Var}[Y_2] = \frac{1}{12}$

4.4.1 Conditional Expectation

Definition 4.9. Let Y_1 and Y_2 be two random variables. The conditional expectation $g(Y_1)$, given that $Y_1 = Y_2$ and is joint continuously, is defined as:

$$E[g(Y_1) \mid Y_2 = y_2] = \int_{-\infty}^{\infty} g(Y_1) f(y_1 \mid y_2) dy_1 \tag{4.13}$$

where $f(y_1 \mid y_2)$ is the conditional PDF. On the other hand, if Y_1 and Y_2 are joint discretely, it's defined as

$$E[g(Y_1) \mid Y_2 = y_2] = \sum_{y_1} g(y_1) P(y_1 \mid y_2) \tag{4.14}$$

where $P(y_1 \mid y_2)$ is the conditional probability function.

Example 4.4.2. Using the same joint PDF as example 4.3.2. Find the conditional expectation of the amount of the liquid dispensed given that $Y_2 = 1.5$

Solution: Using the definition and solution of example 4.3.2 directly, we get:

$$\begin{aligned} E[Y_1 \mid Y_2 = 1.5] &= \int_0^{y_2} y_1 f(y_1 \mid y_2) dy_1 \\ &= \int_0^{y_2} y_1 \cdot \frac{1}{y_2} dy_1 \\ &= \frac{1}{y_2} \int_0^{y_2} y_1 dy_1 = \frac{y_2}{2} = \frac{1.5}{2} = \boxed{0.25} \end{aligned}$$

4.4.2 Independence

Theorem 4.2. *If Y_1 and Y_2 are discrete random variables with joint probability function $p(y_1, y_2)$ and marginal pmfs as $p_1(y_1)$ and $p_2(y_2)$ respectively. Then, Y_1 and Y_2 are independent if and only if*

$$p(y_1, y_2) = p_1(y_1) p_2(y_2) \tag{4.15}$$

$\forall (y_1, y_2) \in \mathbb{R}^2$

Theorem 4.3. *If Y_1 and Y_2 are continuous random variables with joint PDF $f(y_1, y_2)$ and marginal PDFs as $f_1(y_1)$ and $f_2(y_2)$ respectively. Then, Y_1 and Y_2 are independent if and only if*

$$f(y_1, y_2) = f_1(y_1) f_2(y_2) \tag{4.16}$$

$\forall (y_1, y_2) \in \mathbb{R}^2$

Example 4.4.3. Let X and Y be joint continuous random variables with joint PDF:

$$f_{X,Y}(x, y) = \begin{cases} 8xy & 0 < x < y < 1 \\ 0 & \text{otherwise} \end{cases}$$

Determine whether $X \perp Y$ or not.

Solution: We first determine the marginal PDFs for each variables:

$$\begin{aligned} f_X(x) &= \int_x^1 8xy\,dy & f_Y(y) &= \int_0^y 8xy\,dx \\ &= 8x\left|\frac{y^2}{2}\right|_x^1 & &= 8y\left|\frac{x^2}{2}\right|_0^y \\ &= 4y^3 & &= 4x(1-x^2) \end{aligned}$$

Then, $f_X(x)f_Y(y) = 4x(1-x^2)\cdot 4y^3 \neq f_{X,Y}(x,y) = 8xy$. Thus, $X \not\perp Y$.

4.4.3 Covariance

Definition 4.10. Let Y_1 and Y_2 be 2 joint random variables with expectation $E[Y_1]$ and $E[Y_2]$, respectively. Then, the **covariance** between Y_1 and Y_2 is given as

$$\text{Cov}[Y_1, Y_2] = E[(Y_1 - E[Y_1])(Y_2 - E[Y_2])] = E[Y_1Y_2] - E[Y_1]E[Y_2] \tag{4.17}$$

Theorem 4.4. *If $Y_1 \perp Y_2$. Then, Cov$[Y_1, Y_2] = 0$.*

End of Lecture

4.5 Topics: Distributions of Sums of Random Variables

Theorem 4.5. *If $X \perp Y$ be 2 independent random variables. Then, the mgf of $X+Y$ is given by*

$$M_{X+Y}(t) = M_X(t)\cdot M_Y(t) \tag{4.18}$$

Proof. Let $X \perp Y$. Then,

$$M_{X+Y}[t] = E[e^{t(X+Y)}] = E[e^{tX}e^{tY}] = E[e^{tX}]E[e^{tY}] = M_X(t)\cdot M_Y(t)$$

Thus, $M_{X+Y}(t) = M_X(t)\cdot M_Y(t)$. □

Remark 4.4. *Mgf of a random variable id unique which means obtianing enable us to find its probability distribution*

Example 4.5.1. Suppose that $X_1 \sim \text{Bernoulli}(p)$, $X_2 \sim \text{Bernoulli}(p)$ such that $X_1 \perp X_2$. Then, we can determine the mgf of its $X_1 + X_2$ as

$$\begin{aligned} M_{X_1+X_2}(t) &= M_{X_1}(t)M_{X_2}(y) \\ &= (pe^t + 1 - p)(pe^t + 1 - p) = \boxed{(pe^t + 1 - p)^2} \end{aligned}$$

We can thus see that $X_1 + X_2 \sim \text{Binom}(2, p)$.

Theorem 4.6. *If $X_1, X_2, \ldots, X_n$ are independent random variables such that $X_i \sim$ Bernoulli(p) where $i \in \{1, \ldots, n\}$. Then,*

$$X_1 + X_2 + \cdots + X_n = \sum X_i \sim \text{Binom}(n, p) \tag{4.19}$$

Proof. Consider the mgf of the sum of these random variables. Then,

$$\begin{aligned} M_{X_1+X_2+\cdots+X_n}(t) &= M_{X_1}(t) \cdot M_{X_2}(t) \cdot \ldots \cdot M_{X_n}(t) \\ &= \underbrace{(pe^t + 1 - p) \cdot \ldots \cdot (pe^t + 1 - p)}_{n \text{ times}} \\ &= (pe^t + 1 - p)^n \end{aligned}$$

Thus, $\sum X_i \sim \text{Binom}(n, p)$. □

Theorem 4.7. *Let $X_1 \sim \text{Binom}(n, p)$ and $X_2 \sim \text{Binom}(m, p)$. Then, $X_1 + X_2 \sim \text{Binom}(n + m, p)$*

Theorem 4.8. *Let $X \perp Y$ be random variables where $X \sim \text{Poiss}(\lambda_1)$ and $Y \sim \text{Poiss}(\lambda_2)$. Then, $X + Y \sim \text{Poiss}(\lambda_1 + \lambda_2)$*

Theorem 4.9. *Let $X \perp Y$ be random variables where $X \sim \chi^2(v_1)$ and $Y \sim \chi^2(v_2)$. Then, $X + Y \sim \chi^2(v_1 + v_2)$*

Theorem 4.10. *Let $X \perp Y$ be random variables where $X \sim N(\mu_1, \sigma_1^2)$ and $Y \sim N(\mu_2, \sigma_2^2)$. Then, $X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$*

Remark 4.5. *The proofs of these theorem can be directly realized using mgf of the sum of the 2 variables.*

Now, the above theorem is especially important as we can extend it to something much more elegant and that is the central limit theorem.

4.5.1 Central Limit Theorem

The central limit theorem is perhaps the most widely used theorem in statistics.

Theorem 4.11. ***(Central Limit Theorem [CLT]).*** *Sums of large number of independently and identically distributed random variables, X_i converge to normal distribution, no matter what the distribution of X_i is.*

Note that the distribution of X_i should have a defined first and second moments. As a consequence, the sum centered at its mean and scaled by its standard deviation converges to the standard normal distribution. This convergence is the reason why the central limit theorem is so powerful. Also note that there are several versions Central Limit Theorem based on different assumptions and constraints

Theorem 4.12. *We will extend the theorem on independence a little more. The random variables $X_1, \ldots, X_n$ are independent iff*

- $f_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = f_{X_1}(x_1) \cdot \ldots \cdot f_{X_n}(x_n)$.
- $Var[\sum X_i] = \sum Var[X_i]$
- $M_{X_1+X_2+\cdots+X_n}(t) = M_{X_1}(t) \cdot M_{X_2}(t) \cdot \ldots \cdot M_{X_n}(t)$

Theorem 4.13. ***(CLT)***[1]**.** *Let $X_1, \ldots, X_n$ be independent random variables where $X_i \sim N(\mu, \sigma^2)$. Then,*

$$\sum X_i \sim N(n\mu, n\sigma^2) \tag{4.20}$$

and

$$\frac{\sum X_i - E[\sum X_i]}{\sqrt{Var[\sum X_i]}} \sim N(0,1) \tag{4.21}$$

Example 4.5.2. Let $S = \sum X_i$. Furthermore, let's define

$$S' = \frac{S - E[S]}{\sqrt{\text{Var}[S]}}$$

Basically, what happened here is that we standardize S. Then, according to CLT and its consequences, $S \sim N(n\mu, n\sigma^2)$ and $S' \sim N(0,1)$ as $n \to \infty$.

Now, let's consider the sample mean of the every independent random variable defined as

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

[1] This is more mathematically rigorous as compared to theorem 4.11 which is only wordings

Then, its expectation is given as $E[\overline{X}] = \mu$ and its variance as $\text{Var}[\overline{X}] = \sigma^2/n$ (simple calculations). Then, by the CLT, $\overline{X} \sim N(\mu, \sigma^2/n)$. Now, define $\overline{X}'$ as

$$\overline{X}' = \frac{\overline{X} - E[\overline{X}]}{\sqrt{\text{Var}[X]}}$$

Then, by the CLT also, $\overline{X}' \sim N(0,1)$ as $n \to \infty$

Question: How big is should n be in order for CLT to work?
Answer: It should be at least 30 i.e. $n \geq 30$.

Remark 4.6. *Remember that $Var[X] = \sigma^2$. Then, $\sqrt{Var[nS]} = \sqrt{n\,Var[S]} = \sigma \cdot \sqrt{n}$. This is the typical form that the equation will have as most words problem will have standard deviation instead of variance.*

Thus, if $S = \sum X_i$ which are independent and identically distributed, then

$$\frac{S - E[S]}{\sqrt{\text{Var}[S]}} = \frac{S - E[S]}{\sigma \cdot \sqrt{n}} \sim N(0,1) \tag{4.22}$$

Example 4.5.3. The customers enter at a local convenience store randomly. The service time X_i for customer i has mean $E[X_i] = 4$ (minutes) and $\text{Var}[X_i] = 4$. Assume that service times for different customers are independent. Let Y be the total time the bank store keeper spends serving 64 customers. Find $P(Y > 280)$.

Solution: First, in order to use CLT, n must be at least 30, which it is since $n = 60$. So now, let $Y = \sum X_i$ then its expectation and variance is defined as

$$\begin{aligned} E[Y] &= nE[X_i] & \text{Var}[Y] &= n\text{Var}[Y] \\ &= 64(4) = 256 & &= 64(4) = 256 \end{aligned}$$

Then, using what we've known from standardized normal distribution,

$$P(Y < 280) = P\left(\frac{Y - E[Y]}{\sqrt{\text{Var}[Y]}} < \frac{280 - 256}{\sqrt{256}}\right) = P(Z < 1.5) = \boxed{0.0668}$$

Example 4.5.4. The fracture strength of tempered glass averages 14 (measured in thousands of pounds per square inch) and has standard deviation 2.

a. What is the probability that the average fracture strength of 100 randomly selected pieces of this glass exceeds 14.5?

b. Find an interval that includes, with probability 0.95, the average fracture strength of 100 randomly selected pieces of this glass.

Solution: a, first, let X_i be the fracture strength of i pieces of the tempered glass, where $i = \{1,\dots,100\}$. Then, define $\overline{X} = \sum X_i$ to be the average fracture strength of all 100 pieces, and $\overline{X} \sim N(\mu, \sigma^2/n)$. We're asked to find $P(\overline{X} > 14.5)$ where we're given $E[\overline{X}] = 14$ and $\sigma = 2$. Then,

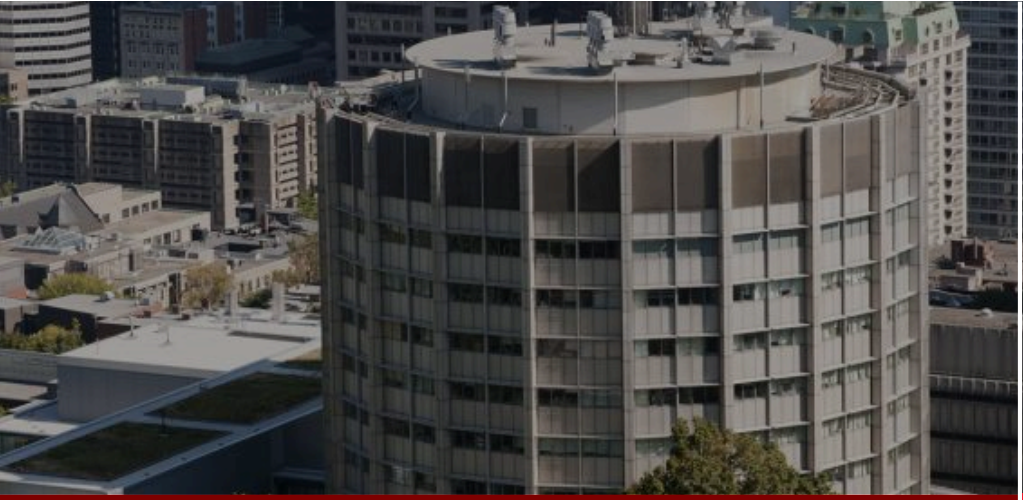
$$\begin{aligned} P(Y > 14.5) &= P\left(\frac{\overline{X} - E[\overline{X}]}{\sqrt{\sigma^2/n}} > \frac{14.5 - E[\overline{X}]}{\sqrt{\sigma^2/n}}\right) \\ &= P\left(Z > \frac{14.5-14}{0.2}\right) \\ &= P(Z > 2.5) = 1 - P(Z < 2.5) = \boxed{0.062} \end{aligned}$$

b, for the interval with probability 0.95, we have

$$\begin{aligned} P(a \le \overline{X} \le b) &= 0.95 \\ P\left(\frac{a-14}{0.2} \le Z \le \frac{b-14}{0.2}\right) &= 0.95 \\ P(-1.96 \le Z \le 1.96) &= 0.95 \end{aligned}$$

which means $\frac{a-14}{0.2} = -1.96 \iff a = 13.608$ and $\frac{b-14}{0.2} = 1.96 \iff b = 14.392$. Then, the interval that include $\overline{X}$ with probability 0.95 is $\boxed{[13.608, 14.392]}$.

End of Lecture



Sample space, events, conditional probability, independence of events, Bayes' Theorem. Basic combinatorial probability, random variables, discrete and continuous univariate and multivariate distributions. Independence of random variables. Inequalities, weak law of large numbers, central limit theorem.

Pigeon's Bindery