# LECTURE 2

## AGENDA:

1. Basic Set Theory
2. Sample spaces and events
3. Probability

## BASIC SET THEORY

Def: A set is a collection of ~~[illegible]~~ distinct objects, which are called elements or points of the set.

- Sets are denoted by capital letters $A, B, C, \ldots$
- Notation: "$A \subset B$": A is a subset of B, i.e., all points in A are also in B

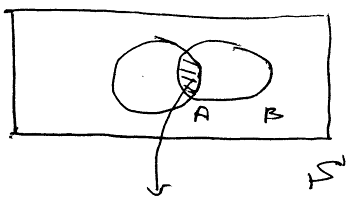
  "$\phi$" denotes the null or empty set, contains no points
- Union "$A \cup B$": Collection of points which are in A or B
- Intersection: "$A \cap B$": Collection of points which are in A and B.

- $S$ denotes the universal set, i.e., collection of all points of interest in the current situation

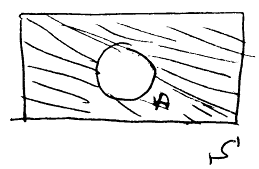
- <u>Complement</u>: "$\overline{A}$": Collection of all points in $S$ that is not in $A$

- $A$ and $B$ are called <u>mutually exclusive</u> if $A \cap B = \phi$ (empty set)

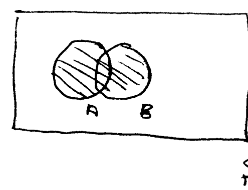
Venn Diagrams are graphical ways of representing sets.




Shaded region: $A \cap B$




Shaded region: $\overline{A}$




Shaded region: $A \cup B$

<u>Distributive laws:</u>

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

<u>De Morgan's Laws:</u>

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

These laws or rules will be useful in probability computations.

Example: $S = \{1,2,3,4\}$

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

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

$A \cap B = \{2,3\}$

$A \cap C = \phi$

$C \subset B \Rightarrow C \cup B = B,\ C \cap B = C$

$\bar{A} = \{4\}$

Homework: Verify the distributive laws and De Morgan's laws for A, B, C above.

---

SAMPLE SPACES AND EVENTS

- Perform a random experiment / observe a random phenomenon. For example, consider the tossing of a coin 4 times or percentage of a population affected by an epidemic.

- Def: The SAMPLE SPACE $S$ of a random experiment is the set of all possible outcomes of the experiment listed in a mutually exclusive and exhaustive way

**Examples:** • Toss a coin 4 times

$S = \{HHHH, HHHT, HHTT, \text{---------}\}$
In total $2^4 = 16$ possible outcomes.
This is an example of a discrete or countable ~~space~~ sample space.

• Percentage of population affected by an epidemic

$S = [0, 100] \rightarrow$ All real numbers from 0 to 100.
This is an example of a continuous or uncountable sample space.

**Def:** An EVENT is any collection of sample points. In other words, any subset of the sample space $S$ is called an event.

**Example:** Toss a coin 3 times

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

A = Event that there is atleast one heads
$\Rightarrow A = \{HHH, HHT, HTH, THH, TTH, THT, HTT\}$
B = Event that there is atmost one heads
$B = \{HTT, THT, TTH, TTT\}$

# FORMAL DEFINITION OF PROBABILITY

Intuitively, "probability" of an event is number between 0 and 1 expressing our belief in the occurrence of the event in a single performance of an experiment

$S$ = Sample space of a random experiment
$\mathcal{A}$ = Collection of all possible events

Def: A PROBABILITY ASSIGNMENT $P$ for a random experiment is a numerically valued function that assigns a value $P(A)$ to every event $A$ so that the following axioms are satisfied:

1) $P(A) \geq 0$ for every event $A$
2) $P(S) = 1$
3) If $A_1, A_2, A_3, \ldots$ is a sequence of mutually exclusive events (i.e., $A_i \cap A_j = \phi$ for every $i \neq j$), then

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

---

- $\boxed{P(\phi) = 0}$. Choose $A_1 = S$ and all others to be $\phi$ in 3).
- If $\boxed{A \cap B = \phi}$, then $\boxed{P(A \cup B) = P(A) + P(B)}$. Choose $A_1 = A$, $A_2 = B$ and all others to be $\phi$ in 3).
- If $\boxed{A \subset B}$, then $\boxed{P(A) \leq P(B)}$.

DEFINING AND CALCULATING THE PROBABILITY OF AN EVENT BY THE SAMPLE POINT METHOD (DISCRETE SAMPLE SPACE)

1) DEFINE THE EXPERIMENT.
2) CONSTRUCT THE SAMPLE SPACE.
3) ASSIGN PROBABILITIES TO EACH OF THE SAMPLE POINTS, MAKING SURE THEY ADD UP TO 1.
4) EXPRESS EVENT OF INTEREST AS A COLLECTION OF SAMPLE POINTS.
5) FIND P(A) BY SUMMING THE PROBABILITIES OF SAMPLE POINTS IN A.

$\boxed{\text{Proof of } P(\phi)=0}$ For axiom 3), choose

$A_1 = S$, $A_i = \phi$ for all $i \geq 2$. It is easy to verify that these events are mutually exclusive. Hence, we get that,

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

$$\Rightarrow P(S) = P(S) + \sum_{i=2}^{\infty} P(A_i), \text{ since } \bigcup_{i=1}^{\infty} A_i = S.$$

$$\Rightarrow \sum_{i=2}^{\infty} P(\phi) = 0.$$

Since $P(\phi) \geq 0$, by axiom 1), it follows that $P(\phi) = 0$.

$\boxed{\text{Proof of } P(A \cup B) = P(A) + P(B) \text{ if } A \cap B = \phi}$

For axiom 3), choose $A_1 = A$, $A_2 = B$, $A_i = \phi$ for all $i \geq 3$. It is easy to verify that these events are mutually exclusive. Hence, we get that,

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

$$\Rightarrow P(A \cup B) = P(A) + P(B) + \sum_{i=3}^{\infty} P(\phi)$$

$$\Rightarrow P(A \cup B) = P(A) + P(B), \text{ since } P(\phi) = 0.$$

# LECTURE 3

Agenda:

1. Some examples to understand the formal definition of probability
2. Fundamental principle of counting
3. Evaluating probabilities using permutations.

## EXAMPLES

Random Experiment: Choose a person from 4 persons at random, with no preference to any person.

- What is the sample space?
  Since there are 4 possible outcomes, the sample space is $S = \{1, 2, 3, 4\}$.

- What is the set of all possible events?
  Recall that $\mathcal{A}$, which denotes the set of all possible events, is essentially the collection of all possible subsets of $S$. Hence,

$$\mathcal{A} = \left\{ \begin{array}{l} \phi, \{1\}, \{2\}, \{3\}, \{4\}, \{1,2\}, \{1,3\}, \{1,4\}, \{2,3\}, \\ \{2,4\}, \{3,4\}, \{1,2,3\}, \{1,2,4\}, \{1,3,4\}, \{2,3,4\}, \\ \{1,2,3,4\} \end{array} \right\}$$

- How to provide a probability assignment P, which reflects our belief in how the random experiment is conducted, and also satisfies the 3 axioms?

<u>SOLUTION:</u> Based on your belief of how the random experiment is conducted, assign a chance to EACH POINT IN THE SAMPLE SPACE $S$, such that these numbers add up to 1. For this experiment, since there is no preference for any person,

$$P(\{1\}) = \frac{1}{4}, \quad P(\{2\}) = \frac{1}{4}, \quad P(\{3\}) = \frac{1}{4}, \quad P(\{4\}) = \frac{1}{4}$$

DEFINE THE PROBABILITY OF ANY EVENT AS THE SUM OF PROBABILITIES OF THE SAMPLE POINTS IN THE GIVEN EVENT.

Hence, for $A$ = Event that Person 1 or Person 2 is chosen,

$$P(A) = P(\{1,2\}) = P(\{1\}) + P(\{2\}) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$$

This procedure guarantees that the probability assignment satisfies the 3 axioms, for discrete sample spaces.

<u>Random Experiment</u>: There are 3 mailboxes. Three people come and put letters at random (with no preference to any of the three mailboxes).

<u>Task</u>: Compute the probability of the event that each mailbox is chosen once.

- Each outcome is a sequence of three numbers, each number denoting the mailbox chosen by the respective person. Hence,

$$S = \{123, 122, 131, \text{----------}\}$$

$$\boxed{27 \text{ points in total}}$$

- Since the mailboxes have no special preference by any person, all outcomes are equally likely.

Hence we assign $P(\{s\}) = \frac{1}{27}$ for all $s \in S$.

- $A$ = Event that each mailbox is chosen once.

$\Rightarrow A = \{123, 132, 213, 231, 312, 321\}$

Hence, $P(A) = P(\{123\}) + P(\{132\}) + P(\{213\}) + P(\{231\}) + P(\{312\}) + P(\{321\})$

$= \frac{1}{27} + \frac{1}{27} + \frac{1}{27} + \frac{1}{27} + \frac{1}{27} + \frac{1}{27}$

$= \frac{6}{27}$

Do we always need to go through this procedure for calculating probabilities of events? No. We can often use counting rules to get around the situation.

**COUNTING RULE # 1** Suppose we are performing an experiment where all outcomes are equally likely, hence we assign the same probability to every sample point (say there are N sample points).

Suppose the event of interest, say A, consists of $n_A$ sample points. Then,

$$P(A) = \frac{n_A}{N} = \frac{\#\text{ sample points in } A}{\#\text{ total sample points}}$$

## THE FUNDAMENTAL PRINCIPLE OF COUNTING

Suppose that an experiment consists of two successive tasks. The first task can result in $n_1$ outcomes and for each such outcome, the second task can result in $n_2$ outcomes. Then the total number of outcomes of the experiment is $n_1 n_2$.

Example: Birthday problem

Random experiment: Choose 25 people.

Task: Evaluate the probability of the event that there is atleast 1 match in the 25 birthdays.

- Sample space is the collection of all possible 25-sequences of birthdays. A typical sample point looks like

$$\underbrace{\_\ \_\ \_\ \_\ \_\ \cdots\cdots\ \_\ \_\ \_}_{25 \text{ of these}}$$

Each an integer between 1 and 365.

Hence, number of sample points $N = (365)^{25}$ (ignoring leap year)

- Assuming all birthdays are equally likely, each sample point has probability $\frac{1}{(365)^{25}}$

- How many different ways are there for 25 people to have no common birthdays?

The answer is $365 \cdot 364 \cdot 363 \cdot \ldots \cdot 341 = \frac{365!}{340!}$

Hence, $n_A = (365)^{25} - \frac{365!}{340!}$

Hence, $P(A) = \frac{n_A}{N} = \frac{(365)^{25} - \frac{365!}{340!}}{(365)^{25}} = 0.5687$

## PERMUTATIONS

Here are some identities which we have been using informally uptil now, and which are quite helpful in counting principles for determining probabilities of events.

- What is the number of ways of choosing $r$ objects with replacement from $n$ objects (order is important)?

  There are $n$ different ways to choose each object. Hence the answer is $n \times n \times - - - \times n = n^r$

- What is the number of ways to choosing $r$ objects without replacement from $n$ objects (order is important)?

  The first slot can be filled in $n$ ways, the second slot in $(n-1)$ ways, the third slot in $(n-2)$ ways and so on. Hence, the answer is.

  $n(n-1)(n-2) - - - (n-r+2)(n-r+1)$ ▲.

# LECTURE 4

Agenda:

① Counting rules.
② Examples

## COUNTING RULES

Recall from Lecture 3 that

- # of ways of choosing $r$ objects from $n$ distinct objects with replacement (order is important) is $n^r$
- # of ways of choosing $r$ objects from $n$ distinct objects without replacement (order is important) is $P^n_r$, where
$$P^n_r = \frac{n!}{(n-r)!} . \quad \text{(PERMUTATIONS)}$$

RESULT: # of ways of choosing $r$ objects from $n$ distinct objects without replacement (order is not important) is
$$C^n_r = \binom{n}{r} = \frac{n!}{r!(n-r)!} \quad \text{((COMBINATIONS)}$$

Proof: Note that now we are only interested in which objects were chosen and not the order in which they are chosen. Since every collection of $r$ objects can be ordered in $r!$ ways, it follows that

1 combination corresponds to $r!$ permutations. Hence, the number of ways of choosing $r$ ~~distinct~~ objects without replacement from $n$ distinct objects (order is not important) is $\frac{1}{r!} P^n_r = \frac{n!}{(n-r)!\,r!}$.

---

**RESULT:** # of ways of partitioning $n$ distinct objects into $k$ groups containing $n_1, n_2, \ldots, n_k$ objects, where $\sum_{i=1}^{k} n_i = n$, is

$$\frac{n!}{n_1!\, n_2! \cdots n_k!}$$

**PROOF:** The $n_1$ objects for the first group can be chosen in $\binom{n}{n_1}$ ways. The $n_2$ objects for the second group can be chosen in $\binom{n-n_1}{n_2}$ ways, ------,

the $n_k$ objects for the $k^{th}$ group can be chosen in $\binom{n-n_1-n_2-\cdots-n_{k-2}-n_{k-1}}{n_k}$ ways. Hence, the

number of ways of partitioning $n$ distinct objects into $k$ groups containing $n_1, n_2, \ldots, n_k$ objects is

$$\binom{n}{n_1}\binom{n-n_1}{n_2}\cdots\binom{n-n_1-n_2-\cdots-n_{k-2}-n_{k-1}}{n_k}$$

$$= \frac{n!}{n_1!\, n_2!\, n_3! \cdots n_{k-1}!\, n_k!}$$

**RESULT:** # of ways of choosing $r$ objects from $n$ distinct objects with replacement (order is not important) is $\binom{n+r-1}{r}$.

**Proof:** Since order is not important, every way of choosing $r$ objects uniquely corresponds to a vector of the form $(l_1, l_2, \ldots, l_n)$ where $l_i$ represents the number of times the $i^{th}$ object is chosen in the $r$ draws. Let us represent the $n$-tuple $(l_1, l_2, \ldots, l_n)$ as follows

$$\underbrace{00\cdots0}_{l_1 \text{ times}} \Big| \underbrace{00\cdots0}_{l_2 \text{ times}} \Big| \cdots\cdots \Big| \underbrace{00\cdots0}_{l_n \text{ times}}$$

Note that $\sum_{i=1}^{n} l_i = r$. It follows that the number of vectors $(l_1, l_2, \ldots, l_n)$ with $l_i \geq 0$ and $\sum_{i=1}^{n} l_i = r$ is same as the number of ways of arranging $n-1$ "|" symbols and $r$ "0" symbols. But the number of such ways is $\frac{(n-1+r)!}{(n-1)!\, r!}$ (we divide by $(n-1)!$ and $r!$ because the $(n-1)$ "|" symbols are indistinguishable and the $r$ "0" symbols are indistinguishable. Hence the number of ways of choosing $r$ objects with replacement from $n$ distinct objects (order is not important) is $\binom{n+r-1}{r}$.

# EXAMPLES

**Example 1:** Suppose a company manufactures 50 different machines out of which 4 are defective. A customer buys 3 machines. Find the probability that all three machines are defective, assuming no preference for any machine

**Solution:** Let us number the machines from 1 to 50. A typical sample point looks like

$$\underbrace{\_\_\ \ \_\_\ \ \_\_}_{\text{3 different integers between 1 to 50}} .$$

Hence, # of sample points $= 50 \times 49 \times 48 \ (= N)$.

$A$ = Event that all machines are defective.

Hence, # of sample points in $A = 4 \times 3 \times 2 \ (= n_A)$.

Hence, $P(A) = \dfrac{n_A}{N} = \dfrac{4 \times 3 \times 2}{50 \times 49 \times 48} = 0.0002$.

(Assuming no preference for any machine in the buy/sell process.)

Example 2: A company makes $n$ orders. There are $n$ distributors. There is no preference for a specific distributor. Find the probability that distributor 1 gets exactly $k$ orders.

Solution: # of sample points $= M^n$.

(There are $M$ distributors to choose from for each of $n$ orders.)

$A$ = Event that distributor 1 gets exactly $k$ orders.

$$\# \text{ of sample points in } A = \underbrace{\binom{n}{k}}_{\substack{\#\text{ of ways of choosing } k \\ \text{orders for distributor 1}}} \underbrace{(M-1)^{n-k}}_{\substack{\#\text{ of ways of assigning the} \\ \text{remaining orders to the} \\ \text{other distributors}}}$$

Hence, $$P(A) = \frac{\binom{n}{k}(M-1)^{n-k}}{M^n}.$$

Example 3: A company has 20 new jobs for which it has recruited 20 employees. There are 6 jobs in City 1, 4 jobs in City 2, 5 jobs in City 3 and 5 jobs in City 4. Out of the 20 employees, 4 are friends. Assuming that the company gives no preference to any person in assigning jobs, find the probability that all 4 friends land in the same city.

<u>Solution:</u>

# of sample points = # of ways of partitioning 20 people into 4 groups of 6, 4, 5, 5 respectively

$$= \frac{20!}{6!\,4!\,5!\,5!}$$

# of sample points with all 4 friends in City 1

$$= \frac{16!}{2!\,4!\,5!\,5!}$$

# of sample points with all 4 friends in City 2

$$= \frac{16!}{6!\,5!\,5!}$$

# of sample points with all 4 friends in City 3

$$= \frac{16!}{6!\,4!\,1!\,5!}$$

# of sample points with all 4 friends in City 4

$$= \frac{16!}{6!\,4!\,5!\,1!}$$

Hence, P( All friends land in the same city)

$$= \frac{\dfrac{16!}{2!\,4!\,5!\,5!} + \dfrac{16!}{6!\,5!\,5!} + \dfrac{16!}{6!\,4!\,1!\,5!} + \dfrac{16!}{6!\,4!\,5!\,1!}}{\dfrac{20!}{6!\,4!\,5!\,5!}}$$

# LECTURE 5

**Agenda:**

1. Conditional Probability
2. Independence

## CONDITIONAL PROBABILITY

Many times, some partial information about the outcome of a random experiment is available and we want to revise our estimates of the chance of an event accordingly.

**Definition:** Let A and B be two events in a random experiment with sample space S. Then, the probability of the event A given that the event B has occurred is defined as

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

provided that $P(B) > 0$.

The symbol $P(A|B)$ is read "THE PROBABILITY OF A GIVEN B".

Example: Toss a fair die. Let A denote the event that the outcome is 2, 4 or 6. If somebody asks you to play the following game:

"If event A occurs, you pay me \$10, otherwise I will pay you \$10"

will you play the game?

Ans: $S' = \{1, 2, 3, 4, 5, 6\}$
All outcomes are equally likely, as it is a fair die with no preference for any outcome.

$A = \{2, 4, 6\}$.

Hence, $P(A) = \frac{1}{2}$.

It seems like a fair game and you would bet.

Suppose the die is cast in a secret chamber and you have a helpful source who tells you that the result was 4, 5 or 6. You have a chance to withdraw or remain in the game. What would you do?

Ans: Let $B = \{4, 5, 6\}$. We are told that $B$ has occured. Let us evaluate $P(A \mid B)$.

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{P(\{4,6\})}{P(\{4,5,6\})} = \frac{2}{3}.$$

Hence, the chance that you will have to pay the other person \$10 is $\frac{2}{3}$ given this additional information. You would withdraw and not play the game.

---

Conditional probabilities satisfy the 3 axioms of probability.

**Result:** Let $B$ be an event with $P(B) > 0$. Then,

1. $0 \le P(A \mid B) \le 1$ for every event $A$.
2. $P(S \mid B) = 1$.
3. If $A_1, A_2, A_3, \ldots$ are mutually exclusive events, then

$$P\left(\bigcup_{i=1}^{\infty} A_i \,\middle|\, B\right) = \sum_{i=1}^{\infty} P(A_i \mid B)$$

Proof: (1) Note that $0 \leq P(A \cap B) \leq P(B)$.
Dividing everything by $P(B)$, we get that,

$$0 \leq \frac{P(A \cap B)}{P(B)} \leq 1$$

$$\Rightarrow \quad 0 \leq P(A|B) \leq 1.$$

(2) $$P(S|B) = \frac{P(S \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1.$$

(3) $$P\left(\bigcup_{i=1}^{\infty} A_i \,\middle|\, B\right) = \frac{P\left(\left(\bigcup_{i=1}^{\infty} A_i\right) \cap B\right)}{P(B)}$$

$$= \frac{P\left(\bigcup_{i=1}^{\infty} (A_i \cap B)\right)}{P(B)}$$

($\because$ By distributive laws)

$$= \frac{\sum_{i=1}^{\infty} P(A_i \cap B)}{P(B)}$$

($\because$ $A_1 \cap B$, $A_2 \cap B$, $A_3 \cap B$, ------- are mutually exclusive)

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

## INDEPENDENCE

If the extra information provided by knowing that an event B has occured does not change the probability of A, i.e, if $P(A|B) = P(A)$, then the events A and B are said to be independent.

**Definition:** Two events A and B are said to be independent if

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

- $P(A \cap B) = P(A)P(B)$ is equivalent to stating that $P(A|B) = P(A)$ or $P(B|A) = P(B)$, if the conditional probabilities $P(A|B)$ or $P(B|A)$ exist.

- Multiplicative rule: If $A_1, A_2, ---, A_n$ are $n$ events, then

$$P(A_1 \cap A_2 \cap --- \cap A_n) = P(A_1) \times P(A_2|A_1) \times P(A_3|A_1 \cap A_2) \times --- \times P(A_{n-1}|A_1 \cap A_2 \cap -- \cap A_{n-2}) \times P(A_n|A_1 \cap A_2 \cap --- A_n)$$

---

Example: Draw a card from a shuffled 52-card deck with no preference to any card. Let A denote the event that a king is drawn, and let B denote the event that a diamond $\diamondsuit$ is drawn.

$$P(A) = \frac{4}{52}, \qquad P(B) = \frac{13}{52}, \qquad P(A \cap B) = \frac{1}{52}.$$

Hence, $P(A \cap B) = P(A)\,P(B)$. This gives independence of $A$ and $B$.

# LECTURE 6

**Agenda:**

1. Inclusion - Exclusion principle
2. Theorem of total probability
3. Bayes rule

## INCLUSION - EXCLUSION PRINCIPLE

The inclusion - exclusion principle provides an identity for computing the probability of union of a set of events in terms of intersections of various orders of these events

**2 events:** If $A, B$ are two events,

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

**3 events:** If $A, B, C$ are three events, then

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

**k events:** If $A_1, A_2, ---, A_k$ are k events,

$$P(A_1 \cup A_2 \cup --- \cup A_k) = \sum_{i=1}^{k} P(A_i) - \sum_{\substack{\text{All unordered} \\ \text{pairs } (i_1, i_2)}} P(A_{i_1} \cap A_{i_2})$$

$$+ \sum_{\substack{\text{All unordered} \\ \text{triplets } (i_1, i_2, i_3)}} P(A_{i_1} \cap A_{i_2} \cap A_{i_3}) -$$

$$--- + (-1)^{k-1} P(A_{i_1} \cap A_{i_2} \cap --- \cap A_{i_k})$$

**Example:** An absent-minded secretary prepared five letters and envelopes to send to five different people. Then he randomly placed letters in the envelopes. A match occurs if the letter and its envelope are addressed to the same person. What is the probability that atleast one of the five letters and envelopes match?

Let, $A$ = Event that atleast one match occurs.

$A_i$ = Event that Letter $i$ is placed in envelope $i$, for $i = 1, 2, 3, 4, 5$.

Then, $A = A_1 \cup A_2 \cup A_3 \cup A_4 \cup A_5$

Note that all arrangements of the 5 letters are equally likely.

Note that the total number of arrangements possible is $5!$

The number of arrangements in the event $A_i$ is $4!$ for $i = 1, 2, 3, 4, 5$ (Why?)

The number of arrangements in the event $A_{i_1} \cap A_{i_2}$ is $3!$ for all unordered pairs $(i_1, i_2)$ (Why?)

The number of arrangements in the event $A_{i_1} \cap A_{i_2} \cap A_{i_3}$ is $2!$ for all unordered triplets $(i_1, i_2, i_3)$ (Why?)

The number of arrangements in the event $A_{i_1} \cap A_{i_2} \cap A_{i_3} \cap$ is $1!$ for all unordered quadruplets $(i_1, i_2, i_3, i_4)$ (Why

The number of arrangements in the event $A_1 \cap A_2 \cap A_3 \cap A_4 \cap A$ is $1$.

Hence, by the inclusion-exclusion principle,

$$P(A) = \sum_{i=1}^{5} \frac{4!}{5!} - \sum_{\substack{\text{All unordered pairs} \\ (i_1, i_2)}} \frac{3!}{5!} + \sum_{\substack{\text{All unordered} \\ \text{triplets } (i_1, i_2, i_3)}} \frac{2!}{5!}$$

$$- \sum_{\substack{\text{All unordered quadruplets} \\ (i_1, i_2, i_3, i_4)}} \frac{1!}{5!} + \frac{1}{5!}$$

$$= 5 \cdot \frac{4!}{5!} - \binom{5}{2} \frac{3!}{5!} + \binom{5}{3} \frac{2!}{5!} - \binom{5}{4} \frac{1!}{5!} + \frac{1}{5!}$$

Hence, $P(A) = 1 - \frac{1}{2!} + \frac{1}{3!} - \frac{1}{4!} + \frac{1}{5!} = 0.6417$

---

## THEOREM OF TOTAL PROBABILITY

If $B_1, B_2, \ldots, B_k$ is a collection of mutually exclusive and exhaustive events, then for any event $A$,

$$P(A) = \sum_{i=1}^{k} P(B_i)\, P(A/B_i)$$

**Example:** A company buys microchips from three suppliers. Supplier I microchips have 10% chance of being defective, Supplier II microchips have 5% chance and Supplier III microchips have 2% chance of being defective. Suppose 20%, 35% and 45% of the current supply came from Suppliers I, II, III respectively. If a microchip is selected randomly from this supply, what is the probability that it is defective?

$B_1$ = Event that microchip comes from Supplier I
$B_2$ = Event that microchip comes from Supplier II
$B_3$ = Event that microchip comes from Supplier III

$A$ = Event that microchip is defective

It is given that

$$P(A|B_1) = 0.1, \quad P(A|B_2) = 0.05, \quad P(A|B_3) = 0.02$$

Note that $B_1, B_2, B_3$ are mutually exclusive and exhaustive, since $B_1 \cup B_2 \cup B_3 = S$, and $B_1 \cap B_2 = \phi$, $B_2 \cap B_3 = \phi$ and $B_1 \cap B_3 = \phi$. Also,

$$P(B_1) = 0.2, \quad P(B_2) = 0.35, \quad P(B_3) = 0.45.$$

Hence, by the Theorem of total probability,

$$P(A) = 0.1 \times 0.2 + 0.05 \times 0.35 + 0.02 \times 0.45 = 0.046$$

---

## BAYES RULE

If the events $B_1, B_2, ---, B_k$ are mutually exclusive and exhaustive, then for any event $A$,

$$P(B_i|A) = \frac{P(A|B_i)\, P(B_i)}{\sum_{j=1}^{k} P(A|B_j)\, P(B_j)}.$$

Example: In the previous example, if a randomly selected microchip is defective, what is the probability that it came from supplier II ?

By Bayes rule,

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

$$= \frac{0.05 \times 0.35}{0.0465}$$

$$= 0.376.$$

# LECTURE - 7

## Agenda:

- The Monty Hall problem
- Other problems

## THE MONTY HALL PROBLEM

Consider a game where there are 3 doors. 2 goats and 1 car are placed randomly behind the 3 doors. The player, who cannot see through the doors, is asked to choose a door. The door is kept closed for the time being, and the host (who knows which door has the car) then opens one of the two remaining doors which has a goat (if both remaining doors have goats, he randomly picks one). The player is now given a choice: Open the door which he/she originally chose or switch to the door which still remains closed. He/she wins whatever is behind the door he/she chooses to open. What is the right strategy for the player?

Suppose the player originally picked door 1 and the host opens door 3. Let,

$C_i$ = Event that car is behind door $i$, $i = 1, 2, 3$

$H_3$ = Event that host opens door 3

We need to evaluate

$$P(C_2 \mid H_3) = P(\text{Car is behind door } 2 \mid \text{Host picks door } 3)$$

Note that $C_1 \cup C_2 \cup C_3 = S$ and $C_1, C_2, C_3$ are mutually exclusive. Hence, by the law of total probability,

$$P(H_3) = \sum_{i=1}^{3} P(H_3 \mid C_i)\, P(C_i),$$

and by Bayes rule

$$P(C_2 \mid H_3) = \frac{P(H_3 \mid C_2)\, P(C_2)}{\sum_{i=1}^{3} P(H_3 \mid C_i)\, P(C_i)}$$

$$= \frac{1 \cdot \frac{1}{3}}{\frac{1}{2} \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3}}$$

$$= \frac{2}{3}.$$

Hence the right strategy is to always switch doors.

OTHER PROBLEMS

(i) Prove that $\binom{n}{r} = \binom{n-1}{r} + \binom{n-1}{r-1}$.

$\binom{n}{r}$ = # of ways of choosing $r$ objects from $n$ objects without replacement, order is not important

Here is another way of counting these ways. Fix one of the objects, call it "J". Every subset of size $r$ chosen either contains "J" or does not contain "J".

$\binom{n-1}{r}$ = # of ways of choosing $r$ objects from $n$ objects WITHOUT choosing "J".

$\binom{n-1}{r-1}$ = # of ways of choosing $r$ objects from $n$ objects WITH "J" being one of them.

(ii) Prove that $\sum_{i=0}^{n} \binom{n}{i} = 2^n$.

$2^n$ = # of possible outcomes of $n$ successive tosses.

$\binom{n}{i}$ = # of outcomes with $i$ heads, $i = 0, 1, 2, ---, n-1, n$

Since the number of possible heads is $0, 1, ---, n$, it follows that $\sum_{i=0}^{n}\binom{n}{i} = 2^n$

(iii) $$k^r = \sum_{\substack{(r_1, r_2, \ldots, r_k) \\ \sum_{i=1}^{k} r_i = r,\ r_i \geq 0}} \frac{r!}{r_1!\, r_2! \cdots r_k!}$$

$k^r$ = # of ways of choosing $r$ objects from $k$ objects with replacement and order is important

Note that if $r_1$ is the number of times 1 is chosen, $r_2$ is the number of times 2 is chosen and so on, then $r_1 + r_2 + \cdots + r_k = r$, $r_i \geq 0$. However, there are $\frac{r!}{r_1!\, r_2! \cdots r_k!}$ such arrangements for any given $k$-type $(r_1, r_2, \ldots, r_k)$ of counts.

It follows that

$$k^r = \sum_{\substack{(r_1, r_2, \ldots, r_k) \\ \sum_{i=1}^{k} r_i = r}} \frac{r!}{r_1!\, r_2! \cdots r_k!}$$

(iv) For $N \geq n$, $\binom{N}{n} = \sum_{i=0}^{n} \binom{M}{i}\binom{N-M}{n-i}$

where $0 < M < N$.

Suppose you have N objects and we want to choose n objects without replacement (order is not important) from them. The # of ways is $\binom{N}{n}$.

Divide the N objects into two groups of size M and N − M. The number of ways of choosing n objects from N objects such that i come from the first group and n − i come from the second group is $\binom{M}{i}\binom{N-M}{n-i}$. Since the values that i can take vary from 0 to n, it follows that

$$\binom{N}{n} = \sum_{i=0}^{n} \binom{M}{i}\binom{N-M}{n-i}.$$

# LECTURE 8

**Agenda:**

1. Random variables
2. Probability mass function
3. Probability distribution function

## RANDOM VARIABLES

Many times when we observe a random experiment, there is a numerical quantity associated with the experiment that we are interested in. For example,

(i) Experiment : Sample $n$ people (WOR) from a population of $N$ people
Quantity of interest : The average height

(ii) Experiment : A pandemic spreading through town
Quantity of interest : Percentage of people affected

(iii) Experiment : Toss a coin 3 times
Quantity of interest : # of heads

These quantities of interest associated with random experiments are called RANDOM VARIABLES, generally denoted by $X, Y, Z, \text{-----------}$ .

**DEFINITION:** A **random variable** is a real-valued function whose domain is the sample space.

Notationally, $X : \mathcal{S} \longrightarrow \mathbb{R}$.

The set of possible values that a random variable $X$ takes, or equivalently, the range of $X$ is generally denoted by $\mathcal{X}$.

**DEFINITION:** A random variable is called **discrete** if the set of possible values that it takes is discrete.

PROBABILITY MASS FUNCTION

We focus on discrete random variables for a while. It is quite important to know the various chances with which the random variables takes various values. Mathematically, we need to know $P(X = x)$ for every $x \in \mathcal{X}$.

**DEFINITION:** The **probability mass function** of a random variable $X$ (discrete), is given by

$$p_X(x) \triangleq P(X = x) \text{ for every } x \in \mathcal{X}.$$

Note that $p_X(x) \geq 0$ and $\sum_{x \in \mathcal{X}} p_X(x) = 1$.

__Example 1:__ Sample $n$ people (WOR) from a population of $N$ people, with no preference to any specific person.

Quantity of interest $(X)$ = The average height

$S$ = All subsets of size $n$ of the $N$ people

Typical sample point $\underline{x} = \underbrace{(x_1, x_2, ---, x_n)}_{\text{unordered}}$.

Mathematically $X : S \longrightarrow \mathbb{R}$ with

$$X(\underline{x}) = \frac{\sum_{i=1}^{n} x_i}{n}.$$

Suppose $N = 3$, $n = 2$. Hence, we have 3 people with heights say 50, 60, 70 inches. Hence,

$$S = \{(M1, M2), (M2, M3), (M3, M1)\}$$

$$P((M1, M2)) = P((M2, M3)) = P((M3, M1)) = \frac{1}{3}.$$

$$\left.\begin{aligned} X((M1, M2)) &= 55 \\ X((M2, M3)) &= 65 \\ X((M3, M1)) &= 60 \end{aligned}\right\} \Rightarrow \mathcal{X} = \{55, 60, 65\}$$

Also, $P(X = 55) = P(X = 60) = P(X = 65) = \frac{1}{3}$.

Hence, the probability mass function is given by

$$p_X(55) = p_X(60) = p_X(65) = \frac{1}{3}.$$

<u>EXAMPLE 2:</u> Toss a fair coin 3 times.
$X =$ # of heads (Quantity of Interest)

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

$$X : S \to \mathbb{R}$$

$$X(HHH) = 3$$

$$X(HHT) = X(HTH) = X(THH) = 2$$

$$X(TTH) = X(THT) = X(HTT) = 1$$

$$X(TTT) = 0$$

This gives $\mathcal{X} = \{0, 1, 2, 3\}$.

Also, $P(X=0) = \frac{1}{8}$, $P(X=1) = \frac{3}{8}$, $P(X=2) = \frac{3}{8}$,

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

Hence, the probability mass function is given by

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

Sometimes it is important to study a random variable by looking at its cumulative properties, i.e., probabilities of the type $P(X \le b)$ for any real number $b$.

> DEFINITION: The distribution function $F$ for a random variable $X$ is defined by
>
> $$F_X(b) \triangleq P(X \le b) \quad \text{for all } b \text{ in } \mathbb{R} \text{ (real numbers)}$$

If $X$ is a discrete random variable,

$$F_X(b) = P(X \le b) = \sum_{x \in \mathcal{X} : x \le b} p_X(x) .$$

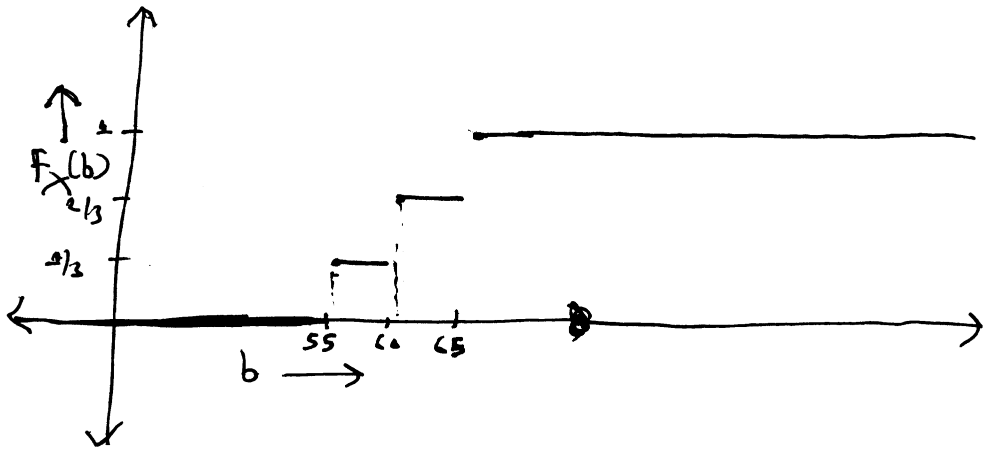
Consider the random variable $X$ in Example 1. Note that

$$p_X(55) = p_X(60) = p_X(65) = \frac{1}{3} .$$

Hence,

$$F_X(b) = \begin{cases} 0 & \text{if } b < 55, \\ \frac{1}{3} & \text{if } 55 \le b < 60, \\ \frac{2}{3} & \text{if } 60 \le b < 65, \\ 1 & \text{otherwise.} \end{cases}$$

Let us look at the plot of $F_X$



# LECTURE – 9

<u>Agenda:</u>

(1) Properties of distribution function
~~(2) Expected values~~
~~(3) Variance / standard deviation~~

## PROPERTIES OF DISTRIBUTION FUNCTION

Let us recollect that if $X$ is a random variable, then its distribution function $F_X : \mathbb{R} \to [0, 1]$ is defined by

$$F_X(b) = P(X \le b) \quad \text{for all } b \in \mathbb{R}.$$

$\boxed{\text{(i)}}$ $\lim_{b \to -\infty} F_X(b) = 0$

<u>Proof:</u> $\lim_{b \to -\infty} F_X(b) = \lim_{b \to -\infty} P(X \le b)$

$$= P(X \le -\infty)$$

$$= 0.$$

$\boxed{\text{(ii)}}$ $\lim_{b \to \infty} F_X(b) = \lim_{b \to \infty} P(X \le b)$

$$= P(X \le \infty)$$

$$= 1.$$

$\boxed{\text{(iii)}}$ $F_X$ is a non-decreasing function.

Proof: Let $a < b$. Then,

$$F_X(a) = P(X \leq a) \quad , \quad F_X(b) = P(X \leq b).$$

Since the event $\{X \leq a\}$ is contained in the event $\{X \leq b\}$,

$$P(X \leq a) \leq P(X \leq b).$$

Hence $F_X(a) \leq F_X(b)$.

$\boxed{\text{(iv)}}$ $F_X$ is right-hand continuous, i.e.,

$$\lim_{h \to 0^+} F_X(b+h) = F_X(b).$$

Proof: Note that,

$$\{X \leq b\} = \bigcap_{m \geq 1} \{X \leq b + \tfrac{1}{m}\}$$

Hence, $$\lim_{n \to \infty} P\left(X \leq b + \frac{1}{n}\right) = \lim_{n \to \infty} P\left(\bigcap_{m \geq 1}^{n} \{X \leq b + \tfrac{1}{m}\}\right)$$

$$= P\left(\bigcap_{m \geq 1} \{X \leq b + \tfrac{1}{m}\}\right)$$

$$= P(X \leq b)$$

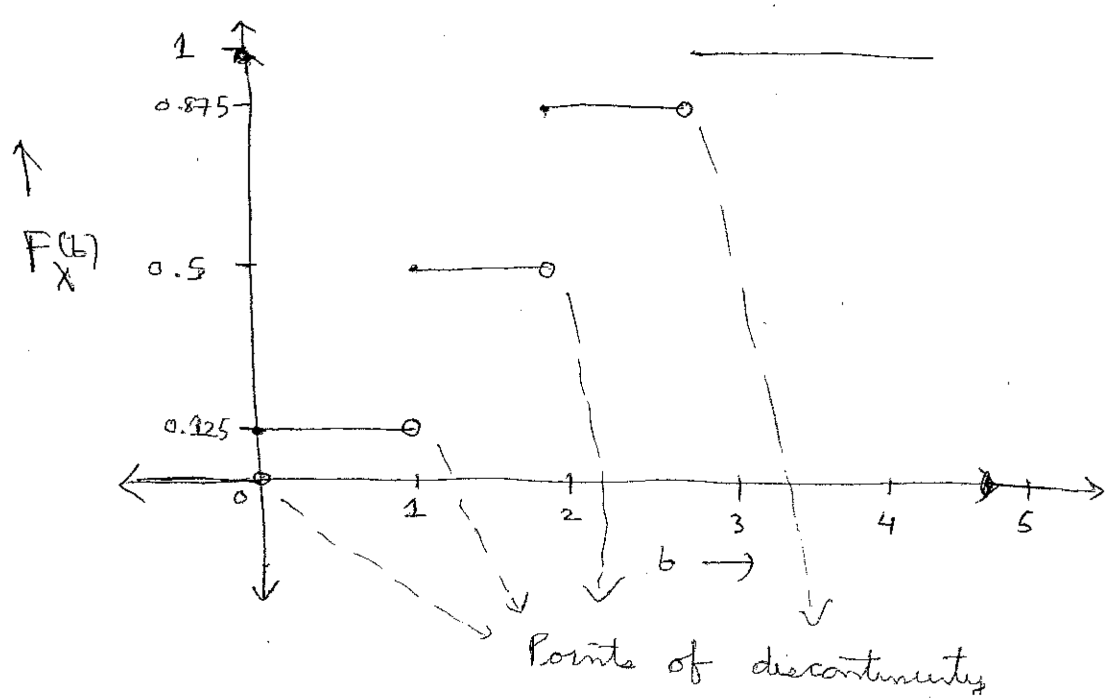
A similar argument holds for any sequence

However, it is not true that $F_X$ is left-hand continuous, especially for discrete random variables.

<u>Example:</u> Let $X$ be the number of heads for 3 tosses of a fair coin. Then $P(X=0) = \frac{1}{8}$, $P(X=1) = \frac{3}{8}$, $P(X=2) = \frac{3}{8}$, $P(X=3) = \frac{1}{8}$.

Then, $F_X(b) = \sum_{x \in \{0,1,2,3\}:\, x \le b} P(X=x)$. Hence,

$$F_X(b) = \begin{cases} 0 & \text{if } b < 0, \\ 0.125 & \text{if } 0 \le b < 1, \\ 0.5 & \text{if } 1 \le b < 2, \\ 0.875 & \text{if } 2 \le b < 3, \\ 1 & \text{if } 3 \le b. \end{cases}$$



$\boxed{\text{FACT:}}$ Any function satisfying properties (i), (ii), (iii) and (iv) is a distribution function of some random variable

# LECTURE - 10

Agenda:

1. Expected values
2. Variance
3. Properties of expected values and variance

## EXPECTED VALUES

Suppose we are interested in a random variable $X$ arising out of a random experiment. Based on our understanding of the random experiment, we have a probability model. Often, we want to summarize our understanding of the random variable in one number, "The expected value" of the random variable.

DEFINITION: The "expected value" of a discrete random variable $X$ with probability mass function $p_X$ is given by

$$E(X) = \sum_{x \in \mathcal{X}} x \, p_X(x) = \sum_{x \in \mathcal{X}} x \, P(X=x).$$

It is also understood as our estimate of the "average" value that the random variable will take.

NOTE: The expected value of a discrete random variable is defined only if $\sum_{x \in \mathcal{X}} |x| \, P(X=x) < \infty$.

Example: Consider the following game. We toss a six-faced die 2 times. If the sum of the two values is 3 or lower, we have to pay 10 dollars. If the sum of the two values is 4, 5 or 6 we pay 4 dollars. If the sum of the two values is 7, 8 or 9 we earn 4 dollars. If the sum of the two values is 10, 11 or 12 we earn 10 dollars. What are the expected winnings?

Let X = Sum of two values on the die

$$P(X = 2 \text{ or } 3) = \frac{3}{36} = \frac{1}{12}$$

$$P(X = 4, 5 \text{ or } 6) = \frac{12}{36} = \frac{1}{3}$$

$$P(X = 7, 8 \text{ or } 9) = \frac{15}{36} = \frac{5}{12}$$

$$P(X = 10, 11 \text{ or } 12) = \frac{6}{36} = \frac{1}{6}$$

Let W = Winnings in the game

$$P(W = -10) = \frac{1}{12}, \quad P(W = -4) = \frac{1}{3}, \quad P(W = 4) = \frac{5}{12},$$

$$P(W = 10) = \frac{1}{6}$$

$$E[W] = \left(-10 \times \frac{1}{12}\right) + \left(-4 \times \frac{1}{3}\right) + \left(4 \times \frac{5}{12}\right) + \left(10 \times \frac{1}{6}\right)$$

$$= \frac{-10 - 16 + 20 + 20}{12}$$

$$= \frac{7}{6}$$

RESULT: If $X$ is a discrete random variable with probability distribution $p_X$ and if $g : \mathbb{R} \to \mathbb{R}$ is a real-valued function, then

$$E(g(X)) = \sum_{x \in \mathcal{X}} g(x)\, p_X(x),$$

If $g(x) = x^2$, then $E[g(W)]$ is given by

$$E[W^2] = 100 \times \frac{1}{12} + 16 \times \frac{1}{3} + 16 \times \frac{5}{12} + 100 \times \frac{1}{6}$$

$$= 25 + 12$$

$$= 37.$$

---

VARIANCE

DEFINITION: The variance of a random variable $X$ with expected value $\mu$ is given by

$$V(X) = E(X-\mu)^2 = \sum_{x \in \mathcal{X}} (x-\mu)^2 p_X(x).$$

## PROPERTIES OF EXPECTATION AND VARIANCE

Note that the variance $V(X)$ of a random variable is the average squared distance between the values of $X$ and the expected value $\mu$.

DEFINITION: The standard deviation of a random variable $X$ is the square root of the variance and is given by

$$SD(X) = \sqrt{E(X-\mu)^2}$$

The standard deviation is also a measure of the variability of a random variable, but it maintains the original units of measure. It can be thought of as the size of a typical deviation between an observed outcome and the expected value.

Example: If $W$ is the winnings in the game discussed in the previous lecture, remember

$P(W=-20) = \frac{1}{12}$, $P(W=-4) = \frac{1}{3}$, $P(W=4) = \frac{5}{12}$,

$P(W=10) = \frac{1}{6}$.

$\boxed{(ii)}$ $V(aX+b) = a^2 V(X)$

Proof:

$$V(aX+b) = E\left[(aX+b) - E(aX+b)\right]^2$$

$$= E\left[(aX+b) - aE(X) - b\right]^2$$

$$= E\left[a(X - E(X))\right]^2$$

$$= a^2 E\left[(X - E(X))^2\right]$$

$$= a^2 V(X)$$

$\boxed{(iii)}$ $V(X) = E(X^2) - (E(X))^2$

$$V(X) = E(X-\mu)^2$$

$$= E(X^2 - 2X\mu + \mu^2)$$

$$= E(X^2) - 2\mu E(X) + E(\mu^2)$$

$$= E(X^2) - 2\mu^2 + \mu^2$$

$$= E(X^2) - \mu^2$$

---

RESULT: Let $X$ be a random variable with mean $\mu$ and variance $\sigma^2$. Then for any positive $k$,

$$P(|X-\mu| < k\sigma) \geq 1 - \frac{1}{k^2}$$

Proof: Let $V(X) = \sigma^2$ and $E(X) = \mu$

$$V(X) = \sum_{x \in \mathcal{X}} (x-\mu)^2 p_X(x)$$

$$= \sum_{x \in \mathcal{X} : |x-\mu| \geq k\sigma} (x-\mu)^2 p_X(x) + \sum_{x \in \mathcal{X} : |x-\mu| < k\sigma} (x-\mu)^2 p_X(x)$$

$$\geq \sum_{x \in \mathcal{X} : |x-\mu| \geq k\sigma} (x-\mu)^2 p_X(x)$$

$$\geq k^2 \sigma^2 \sum_{x \in \mathcal{X} : |x-\mu| \geq k\sigma} p_X(x)$$

$$= k^2 \sigma^2 P(|X-\mu| \geq k\sigma)$$

Hence, $\sigma^2 \geq k^2 \sigma^2 P(|X-\mu| \geq k\sigma)$

Hence, $\frac{1}{k^2} \geq P(|X-\mu| \geq k\sigma)$

This gives us the required identity as

$$\frac{1}{k^2} \geq 1 - P(|X-\mu| < k\sigma)$$

$$\Rightarrow P(|X-\mu| < k\sigma) \geq 1 - \frac{1}{k^2}.$$

LECTURE 11

Agenda:

(1) Tchebysheff's theorem (from Lecture 10)
(2) Bernoulli random variables

TCHEBYSHEFF'S THEOREM (CONTINUED)

Tchebysheff's theorem is typically quite conservative, but thats the price to pay for a general result.

<u>Example:</u> Let $X$ = # of heads in 3 tosses of a fair coin (independently)

Then, $P(X=0) = \frac{1}{8}$, $P(X=1) = \frac{3}{8}$, $P(X=2) = \frac{3}{8}$, $P(X=3) = \frac{1}{8}$.

$E(X) = \frac{3}{2}$, $V(X) = \frac{3}{4}$.

Applying Tchebysheff's theorem with $k = 2$, we get that,

$$P\left( \left| X - \tfrac{3}{2} \right| < \sqrt{3} \right) \geq 1 - \frac{1}{2^2}$$

i.e $P(-0.232 < X < 3.232) \geq 0.75$

But this is very conservative.

WARNING: For a discrete random variable $X$,

$$E(X) = \sum_{x \in \mathcal{X}} x\, p_X(x) = \sum_{x \in \mathcal{X}} x\, P(X = x) \text{ is}$$

well-defined only if $\sum_{x \in \mathcal{X}} |x|\, p_X(x) < \infty$.

If $\mathcal{X}$ is a finite space, this condition will always hold. However, when $\mathcal{X}$ is not a finite space, one should always make sure that the expectations are finite.

<u>Example</u>: Suppose $X$ is a random variable with range $\mathcal{X} = \{1, 2, 3, \ldots\}$, and the probability mass function of $X$ is given by

$$P(X = x) = \frac{6}{\pi^2 x^2} \quad \text{for every } x \in \mathcal{X}.$$

(Note that $\sum_{x=1}^{\infty} \frac{1}{x^2} = \frac{\pi^2}{6}$, hence the above assignment of probabilities is valid.)

Then,

$$\sum_{x \in \mathcal{X}} |x|\, p_X(x) = \sum_{x=1}^{\infty} \frac{1}{x} = \infty.$$

Hence $E(X)$ is not well-defined in this case.

BERNOULLI RANDOM VARIABLES

An experiment with two possible outcomes is called a "Bernoulli experiment." or a "Bernoulli trial". Suppose one outcome of a Bernoulli trial is identified as success and the other outcome is identified as failure. Define the random variable $X$ such that

$$X = \begin{cases} 1 & \text{if the outcome of the trial is a success,} \\ 0 & \text{otherwise.} \end{cases}$$

Let $p$ denote the probability of success in the experiment, i.e.,

$$P(X = 0) = 1 - p, \quad P(X = 1) = p.$$

Such a random variable is said to be a "Bernoulli random variable". The only parameter needed to describe the probability distribution of this random variable is $p$.

$$E(X) = 0 \cdot (1-p) + 1 \cdot p = p.$$

$$\begin{aligned} V(X) &= E(X^2) - (E(X))^2 \\ &= 0 \cdot (1-p) + 1 \cdot p - p^2 \\ &= p(1-p). \end{aligned}$$

# LECTURE 12

Agenda :

(1) Binomial random variable
(2) Geometric random variable

## BINOMIAL RANDOM VARIABLE

Recall that a Bernoulli trial or a Bernoulli experiment has only 2 outcomes. The associated Bernoulli random variable $X$ takes the value 0 if the outcome is a success, and 1 if the outcome is a failure. Let $P(X=1) = p$ and $P(X=0) = 1-p$, i.e., $p$ is the probability of success for the experiment.

$$E(X) = p \quad \text{and} \quad V(X) = p(1-p).$$

Suppose an experiment consists of $n$ independent Bernoulli trials. For eg., toss a coin 1000 times ($n = 1000$) or inspect 1000 items for being defective ($n = 1000$).

Let $X$ = # of successes in the $n$ trials.

If $Y_i = \begin{cases} 0 & \text{if } i^{th} \text{ trial is success,} \\ 1 & \text{if } i^{th} \text{ trial is failure.} \end{cases}$

$$X = \sum_{i=1}^{n} Y_i .$$

DEFINITION: A random variable $X$ is said to be a BINOMIAL RANDOM VARIABLE with parameters $n$ and $p$ if

(i) $X$ is the number of successes in $n$ independent Bernoulli trials

(ii) The probability of success in each trial is $p$.

Let us calculate the probability mass function of a Binomial random variable.

Note that $\mathcal{X} = \{0, 1, 2, ---, n\}$.

$P(X=x) = P($Exactly $x$ trials out of $n$ trials result in a success$)$

$$= \underbrace{\binom{n}{x}}_{\text{\# of ways of choosing } x \text{ positions or trials}} \times \underbrace{p^x}_{\text{Probability of success in the chosen } x \text{ trials}} \times \underbrace{(1-p)^{n-x}}_{\text{Probability of failure in the remaining } n-x \text{ trials}}$$

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

IDENTITY: $\sum_{x=0}^{n} \binom{n}{x} p^x (1-p)^{n-x} = 1$.

<u>Proof:</u> $1 = \sum_{x=0}^{n} p_X(x) = \sum_{x=0}^{n} \binom{n}{x} p^x (1-p)^{n-x}$

Recall that

$$X = \sum_{i=1}^{n} Y_i ,$$

where $Y_i$ is the outcome of the $i^{th}$ trial.

- $E(X) = \sum_{i=1}^{n} E(Y_i) = \sum_{i=1}^{n} p$ [redacted] $= np$ [redacted].

- However, from the basic definition of $E(X)$,

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

$$\boxed{\text{IDENTITY : } \quad \sum_{x=0}^{n} x \binom{n}{x} p^x (1-p)^{n-x} = np}$$

FACT : If $Y_1, Y_2, ---, Y_n$ are random variables arising out of independent experiments, then

$$V\text{[redacted]}\left( \sum_{i=1}^{n} Y_i \right) = \sum_{i=1}^{n} V\text{[redacted]}(Y_i).$$

Hence,

$$V\text{[redacted]}(X) = \sum_{i=1}^{n} V(Y_i) = \sum_{i=1}^{n} p(1-p) = np(1-p).$$

Example: A hospital has to install $n$ generators, with the requirement that with probability 0.99 atleast one generator should be working at any given time. Suppose that the probability that any generator is operating correctly is 0.95. What is the minimum value of $n$?

Let $X$ denote the number of generators which are operating correctly. Note that the probability that each generator operates correctly is 0.95. Assuming that each generator operates independently,

$X$ is a Binomial random variable with parameters $n$ and 0.95.

REQUIREMENT : $P(X \geq 1) = 0.99$

$$\Rightarrow \quad 1 - P(X=0) = 0.99$$

$$\Rightarrow \quad 1 - (1-p)^n = 0.99$$

$$\Rightarrow \quad (0.05)^n = 0.01$$

$$\Rightarrow \quad n = \frac{\log 0.01}{\log 0.05}$$

GEOMETRIC RANDOM VARIABLE

Consider an experiment which consists of repeating independent Bernoulli trials until a success is obtained. Assume that the probability of success in each independent trial is $p$.

Let $X$ = # of failures before the first success.

$$\mathcal{X} = \text{Range}(X) = \{0, 1, 2, \text{---------}\}$$

$P(X = x) = P($ First $x$ trials are failure, $(x+1)^{th}$ trial is success$)$

$$= p(1-p)^x$$

Hence,

$$p_X(x) = p(1-p)^x \quad x = 0, 1, 2, \text{-------}$$

Note that

$$P(X = x) = p(1-p)^x = P(X = x-1)(1-p).$$

Hence, $p_X(x)$ is a decreasing function of $x$.

$$\boxed{\text{IDENTITY: } \sum_{x=0}^{\infty} p(1-p)^x = 1.}$$

$$1 = \sum_{x=0}^{\infty} P(X = x) = \sum_{x=0}^{\infty} p(1-p)^x.$$

Let us calculate the distribution function of X

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

If $x$ is a positive integer,

$$P(X \leq x) = \sum_{t=0}^{x} p(1-p)^{x}$$

$$= p \sum_{t=0}^{x} (1-p)^{x}$$

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

$$= 1 - (1-p)^{x+1}$$

Hence for any positive integer $x$,

$$F_X(x) = 1 - (1-p)^{x+1}.$$

LECTURE 13

Agenda:

1. Geometric distribution continued
2. Negative Binomial distribution

GEOMETRIC RANDOM VARIABLE

Let us recollect that a geometric random variable arises out of an experiment which consists of repeating a Bernoulli trial until the first success. Let $p$ denote the probability of success of the Bernoulli trial.
Let $X$ = # of failures before the first success.

We saw that,

$$\mathcal{X} = \text{Range}(X) = \{0, 1, 2, \text{-----------}\}$$

$$P(X = x) = p(1-p)^x \quad \text{for } x = 0, 1, 2, \text{------}$$

$$P(X \leq x) = \left(1 - (1-p)^{x+1}\right) \text{ for } x = 0, 1, 2, \text{------}$$

$\boxed{\text{RESULT:}}$ $\quad E(X) = \dfrac{1-p}{p}$

Proof: $E(X) = \displaystyle\sum_{x=0}^{\infty} x p (1-p)^x$

$$= p(1-p)\left(1 + 2(1-p) + 3(1-p)^2 + \text{----------}\right)$$

$$= p(1-p)\left(\begin{array}{l} 1 + (1-p) + (1-p)^2 + \cdots\cdots\cdots \\ \quad\quad + (1-p) + (1-p)^2 + \cdots\cdots\cdots \\ \quad\quad\quad\quad + (1-p)^2 + \cdots\cdots\cdots \\ \quad\quad\quad\quad\quad\quad + \cdots\cdots\cdots \end{array}\right)$$

Recall that if $|x| < 1$, then

$$a + ax + ax^2 + \cdots\cdots\cdots = \frac{a}{1-x}.$$

Hence,

$$E(X) = p(1-p)\left(\frac{1}{1-(1-p)} + \frac{q}{1-(1-p)} + \frac{q^2}{1-(1-p)} + \cdots\cdots\right)$$

($q$ denotes $1-p$)

$$= (1-p)\left(1 + q + q^2 + \cdots\cdots\cdots\right)$$

$$= \frac{(1-p)}{1-(1-p)}$$

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

Along identical lines, it can be proved that

$\boxed{\text{RESULT:}}$ $$V(X) = \frac{1-p}{p^2}$$

**Example:** A firm has a new position which needs fluency in both English and Spanish. Applicants are selected randomly from (a very large) pool, and interviewed until the first applicant who is fluent in both English and Spanish is found. If there are roughly 20% applicants in the pool who are fluent in both English and Spanish, what is the expected number of unqualified applicants who will be interviewed before a qualified applicant is found?

Since each applicant is either qualified or unqualified, each interview is a Bernoulli trial.

HOWEVER, EACH BERNOULLI TRIAL IS NOT STRICTLY INDEPENDENT AND IDENTICAL. BUT IF THE POPULATION IS VERY LARGE, WE CAN APPROXIMATELY ASSUME THAT.

Let $X$ = # of unqualified applicants interviewed before the first qualified applicant

Then $X$ is a geometric random variable, where the success probability of the Bernoulli experiment is 0.2. Hence, $E(X) = \dfrac{1-0.2}{0.2} = 4.$

MEMORYLESS PROPERTY OF THE GEOMETRIC DISTRIBUTION

Let $j, k$ be positive integers.

$\boxed{\text{RESULT:}}$ $P(X \geq j+k \mid X \geq j) = P(X \geq k).$

**Proof:** $P(X \geq j+k \mid X \geq j)$

$$= \frac{P(\{X \geq j+k\} \cap \{X \geq j\})}{P(\{X \geq j\})}$$

$$= \frac{P(\{X \geq j+k\})}{P(\{X \geq j\})}$$

$$= \frac{\sum_{x=j+k}^{\infty} p(1-p)^x}{\sum_{x=j}^{\infty} p(1-p)^x}$$

$$= \frac{p(1-p)^{j+k} \sum_{x=0}^{\infty} (1-p)^x}{p(1-p)^j \sum_{x=0}^{\infty} (1-p)^x}$$

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

$$P(X \geq k) = \sum_{x=k}^{\infty} p(1-p)^x$$

$$= p(1-p)^k \sum_{x=0}^{\infty} (1-p)^x$$

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

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

In words, this property means that given that there have been $j$ failures before the first success, the chance of atleast $k$ more failures is exactly the same as if we are just beginning the experiment and want to know the probability of having atleast k failures before the first success.

## THE NEGATIVE BINOMIAL RANDOM VARIABLE

The geometric random variable corresponds to the number of failures before the first success in a sequence of independent Bernoulli trials. But what if we are interested in the number of failures before the $r^{th}$ success for some positive integer $r$?

Let $X$ = # of failures observed before the $r^{th}$ success in the Bernoulli trials

$X$ is called as the NEGATIVE BINOMIAL RANDOM VARIABLE.

Clearly, the random variable $X$ can take any non-negative integer as a value, i.e,

$$\mathcal{X} = \text{Range}(X) = \{0, 1, 2, \text{----------}\}.$$

$$P(X = x) = P\left(\begin{array}{l}\text{The first } x+r-1 \text{ trials contain} \\ x \text{ failures and } r-1 \text{ successes,} \\ (x+r)^{th} \text{ trial is a success}\end{array}\right)$$

$$= P(\text{The first } x+r-1 \text{ trials contain } x \text{ failures and } r-1 \text{ successes})$$
$$P((x+r)^{th} \text{ trial is a success})$$

$$= \Big\{ \underbrace{\binom{x+r-1}{x}}_{\text{\# of ways of choosing the } x \text{ trials with success}} p^{r-1}(1-p)^{x} \Big\} p$$

$$= \binom{x+r-1}{x} p^{r} (1-p)^{x}.$$

Hence, the probability mass function of a negative binomial random variable is given by

$$p_X(x) = \binom{x+r-1}{x} p^{r} (1-p)^{x}, \quad x = 0, 1, 2, \ldots\ldots$$

LECTURE - (14)

Agenda:
(1) Example
(2) Poisson distribution

**Example:** A large lot of tires contains 5% defectives. Four tires are to be chosen from the lot and placed on a car.

(a) Find the probability that 2 defectives are found before four good ones

(b) Find the expected value and the variance of the number of defective tires chosen before finding 4 good tires.

(Assume that the number of tires is large enough so that choosing tires successively can be treated as independent and identical Bernoulli experiments).

As given to us, choosing each tire can be thought of as a Bernoulli experiment with a success if the tire is good (with probability 95%). Hence,

$X$ = # of bad tires chosen before finding 4 good tires is a negative binomial random variable with $r = 4$, $p = 0.95$.

$$P(X=2) = \binom{4+2-1}{2} (0.95)^4 (0.05)^2$$

$$= 10 (0.95)^4 (0.05)^2 = 0.02036$$

$$E(X) = \frac{4 \times 0.05}{0.95} = \frac{4}{19}$$

$$V(X) = 4 \times \frac{0.05}{(0.95)^2} = \frac{400}{19 \times 95} = \frac{80}{361}$$

---

POISSON RANDOM VARIABLE

The poisson distribution was historically derived as a limit of the binomial distribution when the number of trials $n \to \infty$, the probability of success $p \to 0$, but $np \to \lambda > 0$.

It is a very useful model for rare events. For example, the number of accidents on a highway intersection in 1 month, the number of repairs a high quality machine requires within a year, the number of cavalrymen killed by a freak accident in a year, etc.

Let us consider a binomial random variable $X$ with parameters $n$ and $p$.

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

for $x = 0, 1, 2, ---, n-1, n$.

Let us assume $x$ to be fixed. Let $n \to \infty$, $p \to 0$ and $np \to \lambda$.

$$\lim_{n\to\infty, p\to 0} P(X = x) = \lim_{n\to\infty, p\to 0} \binom{n}{x} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x}$$

$$= \lim_{n\to\infty, p\to 0} \frac{\lambda^x}{x!} \frac{n!}{(n-x)!} \left(1 - \frac{\lambda}{n}\right)^{n-x} \frac{1}{n^x}$$

Note that, $\displaystyle \lim_{n\to\infty} \frac{n(n-1) --- (n-x+1)}{n^x} = 1$

Hence $\displaystyle \lim_{n\to\infty, p\to\infty} \frac{n!}{(n-x)!\, n^x} = 1.$

Also, $\displaystyle \lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^n = e^{-\lambda}.$

Hence,

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

Hence, the Poisson random variable is a random variable which takes values in $\{0, 1, 2, \ldots\ldots\}$ and

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

(The earlier calculation is just meant to be an intuitive explanation of why a Poisson random variable can be thought of as a limit of Binomial random variables under appropriate assumptions.)

$$E(X) = \sum_{x=0}^{\infty} x P(X=x) = \sum_{x=0}^{\infty} \frac{x e^{-\lambda}\lambda^x}{x!}$$

$$= \lambda. \quad \text{(Why?)}$$

$$V(X) = \sum_{x=0}^{\infty} (x-\lambda)^2 \frac{e^{-\lambda}\lambda^x}{x!} = \lambda. \quad \text{(Why?)}$$

$$SD(X) = \sqrt{V(X)} = \sqrt{\lambda}.$$

# LECTURE 15

<u>Agenda:</u>

1. Poisson distribution
2. Hypergeometric distribution

## POISSON DISTRIBUTION

A random variable $X$ is said to be a Poisson$(\lambda)$ random variable if

(i) $\mathcal{X} = \text{Range}(X) = \{0, 1, 2, \text{-----------}\}$

(ii) $P(X = x) = \dfrac{e^{-\lambda} \lambda^{x}}{x!}$ for $x = 0, 1, 2, \text{-----}$

We saw in the last lecture that the poisson probabilities can be obtained as limits of binomial probabilities, when $n \to \infty$, $p \to 0$, $np \to \lambda$.

This distribution is generally used to model the number of times an event occurs in a given time frame. For example, the number of accidents at a highway intersection in one month, the number of calls passing through a cellular relay in a five minute period.

VERIFY THAT $P(X=x)$, $x = 0, 1, 2, \ldots$ add up to 1.

$$\sum_{x \in \mathcal{X}} P(X=x) = \sum_{x=0}^{\infty} P(X=x)$$

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

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

$$= e^{-\lambda} e^{\lambda}$$

$$= 1.$$

Hence the probability mass function is valid.

$$E(X) = \sum_{x \in \mathcal{X}} x P(X=x)$$

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

$$= \sum_{x=1}^{\infty} \lambda \frac{e^{-\lambda} \lambda^{x-1}}{(x-1)!}$$

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

$$= \lambda.$$

Similarly,

$$V(X) = \lambda.$$

✲✲ and the number of repairs Y that machine B requires is Poisson random variable with mean 1.12.

<u>Example:</u> The manger of an industrial plant is planning to buy a new machine of either type A or type B. For each day's operation, the number of repairs X that machine A requires is a Poisson random variable with mean 0.96, ✲✲ The daily cost of operating A is $C_A = 160 + 40X^2$; for B, the daily cost of operating is $C_B = 128 + 40Y^2$. Assume that the repairs take negligible time and that each night the machines are cleaned so that they operate like new machines at the start of each day. Which machine minimizes the expected daily cost for the following times of daily operation?

The expected cost for machine A is

$$E[C_A(t)] = 160 + 40E(X^2)$$

$$= 160 + 40\left(V(X) + (E(X))^2\right)$$

$$= 160 + 40\left(0.96 + (0.96)^2\right)$$

$$= 235.264.$$

The expected cost for machine B is

$$E[C_B(t)] = 128 + 40E(Y^2)$$

$$= 128 + 40\left(1.12 + (1.12)^2\right)$$

$$= 222.976$$

Example: The number of calls coming into a hotel's reservation center is a Poisson random variable with mean 3. Find the probability that no calls arrive in a given 1 minute period.

$X$ = # of calls in the given one-minute period

Given that $X$ is Poisson(3).

$$P(X = 0) = \frac{e^{-3} 3^0}{0!} = e^{-3}.$$

Assuming that the number of calls in each minute behaves independently, find the probability that atleast two calls will arrive in a given two-minute period.

Let $X_1$ = # of calls in the first minute
$X_2$ = # of calls in the second minute

We are required to find $P(X_1 + X_2 \geq 2)$

$$P(X_1 + X_2 \geq 2) = 1 - P(X_1 + X_2 < 2)$$

$$= 1 - P(X_1 + X_2 = 0) - P(X_1 + X_2 = 1)$$

$$= 1 - P(X_1 = 0, X_2 = 0) - P(X_1 = 0, X_2 = 1) - P(X_1 = 1, X_2 = 0)$$

$$= 1 - e^{-3} e^{-3} - e^{-3} \frac{e^{-3} 3}{1!} - \frac{e^{-3} 3}{1!} e^{-3}$$

$$= 1 - e^{-6} - \frac{e^{-6}\,6}{1!}$$

$$= 1 - 7e^{-6}.$$

$$= 0.983$$

## HYPERGEOMETRIC DISTRIBUTION

All distributions that we have discussed are in some or the other way related to the Bernoulli distribution. For these random variables, we consider repeated Bernoulli trials which for all practical purposes, are independent and identical. Here is a situation when this is not the case.

Suppose we have a small lot consisting of N items, of which k are defective. Suppose that n items are sampled randomly and sequentially from the lot WITHOUT REPLACEMENT. Let X denote the number of defective items in the n items that are chosen.

$$P(X = x) = \frac{\overbrace{\binom{k}{x}\binom{N-k}{n-x}}^{\text{\# of outcomes with } x \text{ defectives}}}{\underbrace{\binom{N}{n}}_{\text{Total number of outcomes}}}$$

# LECTURE 16

Agenda:

① Hypergeometric distribution
② Summary

## HYPERGEOMETRIC DISTRIBUTION

Suppose we have N dichotomous objects, i.e, N objects each of which is Type I or Type II. Suppose we know that k of the objects are Type I.

<u>Experiment:</u> Draw n objects (WOR) from N objects.

$X$ = # of objects drawn of Type I.

Note that the set of possible values that X can take is $\{0, 1, 2, ---, k-1, k\}$. Hence,

$$\mathcal{X} = \text{Range}(X) = \{0, 1, ---, k-1, k\}.$$

$$P(X=x) = \frac{\underbrace{\binom{k}{x}\binom{N-k}{n-x}}_{\text{\# of ways of choosing n objects out of N objects (WOR) such that x are of type I}}}{\underbrace{\binom{N}{n}}_{\text{\# of ways of choosing n objects out of N objects.}}}$$

$x = 0, 1, ---, k.$

Since $\sum_{x=0}^{k} P(X=x) = 1$, it follows that

$$\sum_{x=0}^{k} \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}} = 1. \qquad \text{——} \quad (*)$$

Note that the identity holds for any fixed integers $N, k, n$.

$$E(X) = \sum_{x=0}^{k} x \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}$$

$$= \sum_{x=1}^{k} x \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}$$

$$= \sum_{x=1}^{k} \frac{k\binom{k-1}{x-1}\binom{N-k}{n-x}}{\frac{N}{n}\binom{N-1}{n-1}} \qquad \left( \begin{array}{l} \because x\binom{k}{x} = k\binom{k-1}{x-1} \\ \binom{N}{n} = \frac{N}{n}\binom{N-1}{n-1} \end{array} \right)$$

$$= \frac{nk}{N} \sum_{x=1}^{k} \frac{\binom{k-1}{x-1}\binom{(N-1)-(k-1)}{(n-1)-(x-1)}}{\binom{N-1}{n-1}}$$

$$= \frac{nk}{N} \sum_{x'=0}^{k-1} \frac{\binom{k-1}{x'}\binom{(N-1)-(k-1)}{(n-1)-x'}}{\binom{N-1}{n-1}},$$

where $x' = x-1$

But using (*) with $N-1, k-1, n-1$ in place of $N, k, n$, we get

$$\sum_{x'=0}^{k-1} \frac{\binom{k-1}{x'}\binom{(N-1)-(k-1)}{(n-1)-x'}}{\binom{N-1}{n-1}} = 1.$$

Hence, $E(X) = \dfrac{nk}{N}$.

Similarly,

$$V(X) = n\left(\frac{k}{N}\right)\left(1-\frac{k}{N}\right)\left(\frac{N-n}{N-1}\right).$$

__Example:__ Two positions are open in a company. Ten men and five women have applied for a job, and all are equally qualified. The manager randomly picks two people to fill the position. What is the probability that one man and one woman are chosen?

This is a hypergeometric experiment with Type I = man, Type II = woman, $N = 15$, $k = 10$, $n = 2$.

$X$ = # of men chosen for the two positions

$$P(\text{One man, one woman}) = P(X=1) = \frac{\binom{10}{1}\binom{5}{1}}{\binom{15}{2}}$$

$$= \frac{10 \times 5}{\frac{15 \times 14}{2!}}$$

$$= \frac{10}{21}.$$

BINOMIAL APPROXIMATION TO HYPERGEOMETRIC

It can be proved that as $N \to \infty$ and $\frac{k}{N} \to p$

$$P(X = x) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}} \longrightarrow \binom{n}{x} p^{x} (1-p)^{n-x}.$$

Hence if the number of objects $N$ is very large, the hypergeometric random variable can be approximated by a binomial random variable with parameters $n$ and $p = \frac{k}{N}$.

But unless the question says that $N$ is large and you can approximate the experiment by a binomial experiment, always assume that the experiment is a hypergeometric experiment.

SUMMARY

In the past 3 weeks, we have studied 6 random variables. Here is a summary of their basic properties.

## BERNOULLI RANDOM VARIABLE

Experiment: Any experiment with two outcomes, success and failure.

$$X = \begin{cases} 0 & \text{if Outcome is failure} \\ 1 & \text{if Outcome is success} \end{cases}$$

Parameters: $p$ = Probability of success

$\mathcal{X}$ = Range $(X) = \{0, 1\}$.

P.M.F. : $P(X = x) = p^{x}(1-p)^{1-x}, \quad x = 0, 1.$

$E(X) = p$

$V(X) = p(1-p)$

## BINOMIAL RANDOM VARIABLE

Experiment: Repeat $n$ Bernoulli trials (independently)

$X$ = # of successes

Parameters : $n$ = Repetitions, $p$ = Probability of success in a single Bernoulli experiment.

$\mathcal{X}$ = Range $(X) = \{0, 1, 2, ---, n-1, n\}$

P.M.F. : $P(X = x) = \binom{n}{x} p^{x}(1-p)^{n-x},$

$x = 0, 1, ---, n$

$E(X) = np$

$V(X) = np(1-p)$

GEOMETRIC RANDOM VARIABLE

<u>Experiment:</u> Repeat Bernoulli trials (independently) until first success.

$X$ = # of failures before first successes

Parameters: $p$ = Probability of success in a single Bernoulli experiment

$\mathcal{X} = \text{Range}(X) = \{0, 1, 2, \text{----------}\}$

P.M.F.: $P(X = x) = (1-p)^x p, \quad x = 0, 1, 2, \ldots$

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

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

NEGATIVE BINOMIAL RANDOM VARIABLE

<u>Experiment:</u> Repeat Bernoulli trials (independently) until $r^{th}$ success.

$X$ = # of failures before the $r^{th}$ success

Parameters: $r$ = Number of successes after which we stop the experiment

$p$ = Probability of success in a singe Bernoulli trial.

$\mathcal{X} = \text{Range}(X) = \{0, 1, 2, \text{----------}\}$

P.M.F.: $P(X = x) = \binom{x+r-1}{x} p^r (1-p)^x, \quad x = 0, 1, 2, \ldots$

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

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

## POISSON RANDOM VARIABLE

**Experiment:** This random variable arises from experiments which can be approximated by Binomial experiments with LARGE $n$, SMALL $p$ and $np \to \lambda > 0$. It is generally used to model the number of times a certain event occurs in a given time frame or a given area.

Parameters: $\lambda > 0$.

*Range$(X) = \{0, 1, 2, \text{---------}\}$

$$P(X = x) = \frac{e^{-\lambda}\lambda^{x}}{x!}, \quad x = 0, 1, 2, \text{---------}.$$

$E(X) = \lambda$

$V(X) = \lambda$

## HYPERGEOMETRIC RANDOM VARIABLE

**Experiment:** Draw $n$ objects from $N$ objects of two types, type I and type II. The objects are drawn without replacement.

$X$ = # of objects of type I

Parameters: $N$ = Total number of objects
$n$ = Number of objects drawn
$k$ = Total number of objects of Type I.

$\mathcal{X} = \text{Range}(X) = \{0, 1, ---, k\}$

P.M.F: $P(X = x) = \dfrac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}, \quad x = 0, 1, ---, k$

$E(X) = \dfrac{nk}{N}$

$V(X) = n\left(\dfrac{k}{N}\right)\left(1 - \dfrac{k}{N}\right)\left(\dfrac{N-n}{N-1}\right)$

# LECTURE – 17

Agenda:

1. Continuous random variables
2. Probability density functions
3. Probability distribution function

## CONTINUOUS RANDOM VARIABLES

Till now we were dealing with discrete random variables, i.e., random quantities which can possibly take finite or countably infinite values. But there are many random quantities that we observe, which can possibly take uncountably many values. For example,

(i) Proportion of people infected by a pandemic. The set of possible values is $[0, 1]$, i.e., any real number in the interval $[0, 1]$.

(ii) Height of a randomly chosen individual in a country

(iii) The time it takes to get served in a French restaurant

and lots more of them.

For discrete random variables, we had the probability mass function where we assign $P(X=x)$ for every $x \in \mathcal{X} = \text{Range}(X)$, such that

$$\sum_{x \in \mathcal{X}} P(X=x) = 1.$$

HOWEVER, WE CANNOT DO THIS FOR CONTINUOUS RANDOM VARIABLES, IF WE WANT TO BE CONSISTENT WITH THE THREE AXIOMS OF PROBABILITY.

<u>Mathematical fact:</u> For a continuous random variable, if we insist on assigning a positive probability to each single outcome of the experiment i.e., if we insist that $P(X=x) > 0$ for every $x \in \mathcal{X} = \text{Range}(X)$, then $P(S) > 1$, which is in deviation from the second axiom of probability (and also intuition)

WHAT IS A WAY OUT?

Instead of working with the probability mass function, we work with what is known as a probability density function which is supposed to indicate the relative proportion of each value in the range of the random variable under consideration.

# PROBABILITY DENSITY FUNCTION FOR A CONTINUOUS RANDOM VARIABLE

<u>Definition</u>: A continuous random variable $X$ is said to have a probability density function $f_X : \mathbb{R} \to \mathbb{R}$ if,

(i) $f_X(x) \geq 0$ for every $x \in \mathbb{R}$

$(f_X(x) = 0$ for $x \notin \mathcal{X})$

(ii) $\int_{-\infty}^{\infty} f_X(x) = 1$, or equivalently $\int_{\mathcal{X}} f_X(x)\,dx = 1$.

(iii) $P(a \leq X \leq b) = \int_a^b f_X(x)\,dx$

for all $a < b \in \mathbb{R}$

It turns out that this definition is consistent with the axioms of probability

Recall that for any random variable $X$, the probability distribution function $F_X : \mathbb{R} \to \mathbb{R}$ is defined as

$$F_X(b) = P(X \leq b) \text{ for every } b \in \mathbb{R}$$

Then, by the definition of the probability density function,

$$\boxed{F_X(b) = P(-\infty \leq X \leq b) = \int_{-\infty}^{b} f_X(x)\,dx}$$

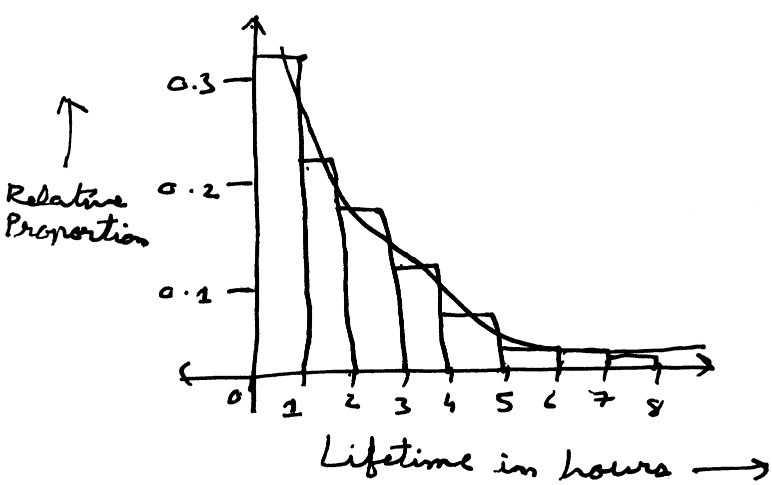
By the fundamental theorem of calculus,

$$\frac{d}{db} F_X(b) = \frac{d}{db}\left(\int_{-\infty}^{b} f_X(x)\,dx\right) = f_X(b).$$

Hence, THE DENSITY FUNCTION $f_X$ IS THE DERIVATIVE OF THE DISTRIBUTION FUNCTION OF A RANDOM VARIABLE

Let us look at a practical example to see how people decide what is an appropriate density function for a random quantity which takes a continuous set of values.

<u>Example:</u> Suppose that we are interested in the battery life of a transistor randomly chosen from a large collection of transistors. Suppose that in the past, somebody chose 50 of these transistors randomly and measured their lifetimes (see Table 5.1 in the textbook). They found that 32% of the 50 observations fall into [0, 1], 22% fall into [1, 2], and so on.



This histogram gives us a good insight into the possible probabilistic model for the random variable $X$, where

$X$ = Lifetime of a randomly chosen transistor.

The histogram ~~[illegible]~~ seems to be very well approximated by a negative exponential curve, in particular the function

$$f_X(x) = \begin{cases} \frac{1}{2} e^{-x/2}, & x > 0, \\ 0, & x \le 0. \end{cases}$$

Note that $f_X(x) \ge 0$ and $\int_0^\infty f_X(x) = 1$.

Hence, after looking at the historical data, it seems reasonable to assume that $X$ is a random variable with probability density function $f_X$.

This is a standard way of coming up with probabilistic models for continuous random quantities. Divide the range into small subintervals. From available historical data, find the proportion of observations in each subinterval. Find a function $f_X$ which approximates the resulting histogram, and satisfies $f_X(x) \ge 0$ for every $x \in \mathbb{R}$, $\int_{-\infty}^{\infty} f_X(x)\,dx = 1$.

# LECTURE 18

Agenda:

1. Properties of probability density functions
2. Examples

## PROPERTIES OF PROBABILITY DENSITY FUNCTIONS

We saw that continuous random variables cannot be described in terms of probability mass functions, as this leads to inconsistencies with the three basic axioms of probability. As a solution, the concept of a probability density function for continuous random variables was introduced.

Definition: If $X$ is a continuous random variable, then there exists a function $f_X$ such that

(i) $f_X(x) \geq 0$ for all $x \in \mathbb{R}$ ( $f_X(x) = 0$ for $x \notin \mathcal{X} = \text{Range}(X)$)

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

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

We start by looking at some properties of the probability density function.

We saw that if $F_X(b) = P(X \leq b)$ is the probability distribution function of a continuous random variable, then

$$f_X(b) = \frac{d}{db} F_X(b).$$

Hence,

$$f_X(b) = \lim_{h \to 0} \frac{F_X(b+h) - F_X(b-h)}{2h} \quad \text{(by definition of derivatives)}$$

$$= \lim_{h \to 0} \frac{P(X \leq b+h) - P(X \leq b-h)}{2h}$$

$$= \lim_{h \to 0} \frac{P(b-h < X \leq b+h)}{2h}$$

Hence, for any two points $b$ and $b'$,

$$\frac{f_X(b)}{f_X(b')} = \lim_{h \to 0} \underbrace{\frac{P(b-h < X \leq b+h)}{P(b'-h < X \leq b'+h)}}$$

$$\boxed{\text{Ratio of the probability the } X \text{ is very close to } b, \text{ to the probability that } X \text{ is very close to } b'.}$$

~~Hence, although the probability density function does not directly relate to~~

Hence, the ratio of the probability density function at two points can be roughly interpreted as the ratio of the probabilities that X is close to the two points respectively.

If X is a continuous random variable with density $f_X$, then

$$P(X = x) = P(x \leq X \leq x) = \int_x^x f_X(y)\,dy = 0.$$

Hence, we inherently assign $P(X = x) = 0$ for every $x$.

CONFUSED ? Remember that we are in a difficult situation and are trying to come up with a reasonable framework for describing continuous random variables probabilistically. The framework we have suggested is indeed the best possible way out.

1. Note that X can take uncountably many values. Hence saying $P(X=x) = 0$ DOES NOT RULE OUT $x$ as a possible value (ATLEAST THAT IS NOT HOW WE SHOULD INTERPRET IT)

2. Note that as long as $x \in \mathcal{X} = \text{Range}(X)$,

$$P(x-h \leq X \leq x+h) = \int_{x-h}^{x+h} f_X(y)\,dy > 0.$$

Hence, for continuous random variables, any interval (in the range of X) is assigned a positive probability, but every point in the range of X is assigned zero probability.

EXAMPLES

Example 1: The distribution function of the random variable X, the time (in months) from the diagnosis age until death for a population of patients with AIDS, is as follows

$$F_X(x) = \begin{cases} 1 - e^{-0.03x^{1.2}}, & x \geq 0, \\ 0 & x < 0. \end{cases}$$

(a) Find the probability density function of X

Note that for $x \geq 0$.

$$f_X(x) = \frac{d}{dx} F_X(x) = \frac{d}{dx}\left(1 - e^{-0.03x^{1.2}}\right)$$

$$= 0.03 \times 1.2 x^{0.2} e^{-0.03x^{1.2}}$$

$$= 0.036 x^{0.2} e^{-0.03x^{1.2}},$$

and for $x < 0$,

$$f_X(x) = \frac{d}{dx} F_X(x) = \frac{d}{dx}(0) = 0.$$

Hence, the density of X is given by

$$f_X(x) = \begin{cases} 0.036 x^{0.2} e^{-0.03x^{1.2}} & x \geq 0, \\ 0 & x < 0. \end{cases}$$

$\boxed{(b)}$ Find the probability that a randomly selected person survives atleast 12 months.

$$P(X \geq 12) = 1 - P(X \leq 12) \quad (\because P(X=12)=0)$$

$$= 1 - F_X(12)$$

$$= 1 - \left(1 - e^{-0.03(12)^{1.2}}\right)$$

$$= 0.55$$

<u>Example 2:</u> Suppose that a random variable X has a probability density function given by

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

$\boxed{(a)}$ Find the probability that $-1 < X < 1$.

$$P(-1 < X < 1) = P(-1 \leq X \leq 1) \quad (\because P(X=-1) = P(X=1) = 0)$$

$$= \int_{-1}^{1} f_X(x)\, dx$$

$$= \int_{-1}^{1} \frac{x^2}{3}\, dx$$

$$= \left[\frac{x^3}{9}\right]_{-1}^{1}$$

$$= \frac{1-(-1)}{9}$$

$$= \frac{2}{9}.$$

$\boxed{b}$ Find the distribution function of $X$.

Note that for $x \leq -1$,

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(y)\,dy = 0$$

$$(\because f_X(y) = 0 \text{ if } y < -1)$$

For $-1 < x < 2$,

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

$$= \int_{-1}^{x} \frac{y^2}{3}\,dy$$

$$= \left[\frac{y^3}{9}\right]_{-1}^{x}$$

$$= \frac{x^3+1}{9}$$

For $x \geq 2$,

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

$$= \int_{-1}^{2} f_X(y)\,dy$$

$$= \frac{2^3+1}{9}$$

$$= 1.$$

Hence,
$$F_X(x) = \begin{cases} 0 & \text{if } x \le -1, \\ \frac{x^3+1}{9} & \text{if } -1 < x < 2, \\ 1 & \text{if } x \ge 2. \end{cases}$$

<u>Example 3:</u> Suppose that the weekly repair cost for a certain machine (in units of \$100) denoted by the random variable $X$, has probability density function given by

$$f_X(x) = \begin{cases} cx(1-x), & 0 \le x \le 1, \\ 0 & \text{otherwise}. \end{cases}$$

Find $c$.

Since $f_X$ is a probability density function,
$$\int_{-\infty}^{\infty} f_X(x) = 1.$$

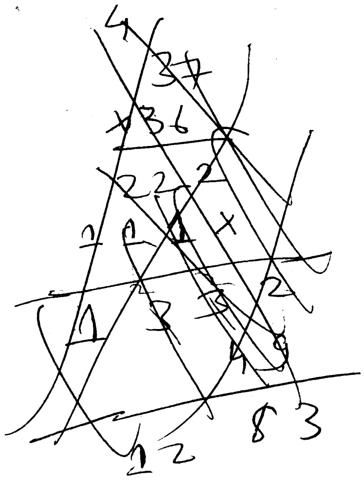
Hence, $\int_0^1 cx(1-x)\,dx = 1$

$\Rightarrow \quad c\left\{ \int_0^1 x\,dx - \int_0^1 x^2 dx \right\} = 1$

$\Rightarrow \quad c\left\{ \left[\frac{x^2}{2}\right]_0^1 - \left[\frac{x^3}{3}\right]_0^1 \right\} = 1$

$\Rightarrow \quad c\left\{ \frac{1}{2} - \frac{1}{3} \right\} = 1$

$\Rightarrow \quad c = 6.$



# LECTURE 19

Agenda:

(1) Expected values of continuous random variables
(2) Tchebysheff's inequality for continuous random variables
(3) ~~[illegible]~~ Examples

## EXPECTED VALUES OF CONTINUOUS RANDOM VARIABLES

**Definition:** If $X$ is a continuous random variable with density $f_X$, then the expected value of $X$ is defined as

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

(Assuming that $\int_{-\infty}^{\infty} |x| f_X(x)\, dx < \infty$).

Notice the parallels with discrete random variables. The sum is replaced by an integral, and the probability mass function $p_X$ is replaced by the probability density function $f_X$.

**RESULT:** If $X$ is a continuous random variable with density $f_X$, then for any function $g: \mathbb{R} \to \mathbb{R}$,

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

(Assuming that $\int_{-\infty}^{\infty} |g(x)| f_X(x)\, dx < \infty$).

The variance of a continuous random variable is defined in a similiar fashion as a discrete random variable.

**Definition:** If $X$ is a continuous random variable with probability density function $f_X$, then the variance of $X$ is given by

$$
\begin{aligned}
V(X) &= E[(X - E(x))^2] \\
&= \int_{-\infty}^{\infty} (x - E(x))^2 f_X(x)\, dx.
\end{aligned}
$$

Let us look at an example to understand these defintions and results.

**Example:** For a given teller in a bank, let $X$ denote the proportion of time, out of a 40-hour work week, that he is directly serving the customers. Suppose that $X$ has a probability density function given by

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

$\boxed{(a)}$ Find the mean proportion of time during a 40-hour work week the teller directly serves customers.

We [redacted] required $E(X)$.

By definition of the expected value,

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

$$= \int_0^1 x . 3x^2 \, dx \quad \left( \because f_X(x) \text{ is zero if } x \text{ is outside } [0,1]. \right)$$

$$= 3 \int_0^1 x^3 \, dx$$

$$= 3 \left[ \frac{x^4}{4} \right]_0^1$$

$$= \frac{3}{4}(1-0)$$

$$= \frac{3}{4}.$$

Hence, the teller on an average spends 75% of his time directly serving customers.

$\boxed{(b)}$ Find the variance of the proportion of time during a 40-hour work week the teller directly serves customers.

Before we go forward, we state some linearity properties of $E(X)$ and $V(X)$, that were true for discrete random variables, and are also true for continuous random variables.

(i) If $a, b$ are constants, $E(aX+b) = aE(X) + b$.

(ii) If $a, b$ are constants, $V(aX+b) = a^2 V(X)$.

(iii) $V(X) = E(X^2) - (E(X))^2$.

By definition of the expected value of $g(X)$ with $g(x) = x^2$, we get that,

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f_X(x)\,dx$$

$$= \int_0^1 x^2 \cdot 3x^2\,dx$$

$$= 3\int_0^1 x^4\,dx$$

$$= 3\left[\frac{x^5}{5}\right]_0^1$$

$$= \frac{3}{5}(1-0)$$

$$= \frac{3}{5}.$$

Hence, $V(X) = \frac{3}{5} - \left(\frac{3}{4}\right)^2 = 0.60 - (0.75)^2 = 0.0375.$

---

TCHEBYSHEFF'S THEOREM FOR CONTINUOUS RANDOM VARIABLES

The Tchebysheff's theorem holds for continuous random variables in the same way as for discrete random variables.

**RESULT:** If $X$ is a continuous random variable with $E(X) = \mu_X$ and $SD(X) = \sigma_X$, then for any $k \geq 0$,

$$P(|X-\mu_X| < k\sigma_X) \geq 1 - \frac{1}{k^2}.$$

There is another way to determine expectations of $\boxed{\text{non-negative}}$ continuous random variables directly from their distribution functions.

**RESULT:** If $X$ is a $\boxed{\text{non-negative}}$ continuous random variable, then

$$E[X] = \int_0^\infty P(X \geq x)\,dx = \int_0^\infty [1 - F_X(x)]\,dx.$$

**Example:** The distribution function of the random variable $X$, the time (in years) from the time a machine is serviced until it breaks down, is as follows:

$$F_X(x) = \begin{cases} 1 - e^{-4x} & \text{if } x > 0, \\ 0 & \text{otherwise.} \end{cases}$$

$\boxed{(9)}$ Find $E(X)$, $V(X)$.

$$E(X) = \int_0^\infty (1 - F_X(x))\,dx \quad \left(\because X \text{ is } \boxed{\text{non-negative}}.\right)$$

$$= \int_0^\infty e^{-4x}\,dx$$

$$= \left[-\frac{e^{-4x}}{4}\right]_0^\infty$$

$$= \left\{(-0) - \left(-\frac{1}{4}\right)\right\}$$

$$= \frac{1}{4}.$$

Since $f_X(x) = \frac{d}{dx}F_X(x)$, $f_X(x) = \begin{cases} 4e^{-4x} & \text{if } x > 0, \\ 0 & \text{otherwise.} \end{cases}$

It can be established from integration by parts that,

$$E[X^2] = \int_0^\infty x^2 f_X(x)\, dx = 4\int_0^\infty x^2 e^{-4x}\, dx = \frac{1}{8}.$$

Hence,

$$V(X) = E[X^2] - (E[X])^2 = \frac{1}{8} - \left(\frac{1}{4}\right)^2 = \frac{1}{16}.$$

$\boxed{(b)}$ Find an interval such that the probability that $X$ lies in the interval is atleast 75%.

By Tchebysheff's inequality, with $k=2$,

$$P\left(|X - \tfrac{1}{4}| < 2\, SD(X)\right) \geq 1 - \frac{1}{4} = 0.75$$

Since $V(X) = \frac{1}{16}$, $SD(X) = \sqrt{\frac{1}{16}} = \frac{1}{4}$

Hence $$P\left(|X - \tfrac{1}{4}| < 2 \cdot \tfrac{1}{4}\right) \geq 0.75$$

$$\Rightarrow P\left(|X - \tfrac{1}{4}| < \tfrac{1}{2}\right) \geq 0.75$$

$$\Rightarrow P\left(-\tfrac{1}{4} \leq X \leq \tfrac{3}{4}\right) \geq 0.75$$

Hence, $X$ lies in the interval $\left[-\frac{1}{4}, \frac{3}{4}\right]$ with probability atleast 75%.

Agenda:

1. Uniform random variable

~~[illegible]~~

Having established a general framework to study continuous random variables, we now move on to the study of special classes of continuous random variables that have been found useful in practical applications

UNIFORM RANDOM VARIABLE

Perhaps the simplest and most natural class of continuous random variables is the class of Uniform random variables.

Consider an experiment which consists of choosing a point from the interval $[a, b]$ such that "all points are equally likely to be chosen.". The random variable of interest is

$$X = \text{Value of the point chosen.}$$

We need to specify a density function for X which satisfies the intuitive notion that "all points are equally likely to be chosen.".

Remember from Lecture 18 that the ratio of the density functions of X at two points, say $x, x'$, has the following interpretation:

$$\frac{f_X(x)}{f_X(x')} \approx \frac{P(X \in (x-h, x+h))}{P(X \in (x'-h, x'+h))} \quad (h \text{ very small}).$$

Our intuition about the experiment immediately tells us that the ratio of probabilities on the right hand side of the above expression should be 1. Hence

$$f_X(x) = f_X(x') \quad \text{for every pair } x, x' \text{ in } [a,b],$$

i.e., $f_X$ is constant in the interval $[a,b]$. In other words,

$$f_X(x) = \begin{cases} c & \text{if } x \in [a,b], \\ 0 & \text{if } x \notin [a,b], \end{cases}$$

where c is an unknown constant. How do we figure out the value of c? Since $f_X$ is a probability density function, it follows that

$$\int_{-\infty}^{\infty} f_X(x) = 1.$$

Hence,

$$\int_a^b c\,dx = 1 \quad \Rightarrow c = \frac{1}{b-a}.$$

In conclusion, the density of $X$ is given by

$$f_X(x) = \begin{cases} \dfrac{1}{b-a} & \text{if} \quad x \in [a,b], \\ 0 & \text{if} \quad x \notin [a,b]. \end{cases}$$

This density is known as the "UNIFORM DENSITY ON [a, b]", and $X$ is known as the "UNIFORM [a, b] RANDOM VARIABLE". Let us now derive some properties of the random variable.

(1) DISTRIBUTION FUNCTION OF $X$.

$$F_X(x) = P(X \leq x) = \begin{cases} 0 & \text{if} \quad x < a, \\ \displaystyle\int_a^x \frac{1}{b-a}\,dt & \text{if} \quad a \leq x \leq b, \\ 1 & \text{if} \quad x > b \end{cases}$$

$$= \begin{cases} 0 & \text{if} \quad x < a, \\ \dfrac{x-a}{b-a} & \text{if} \quad a \leq x \leq b, \\ 1 & \text{if} \quad x > b. \end{cases}$$

(2) Let $[c, d] \subseteq [a, b]$, i.e., $a < c < d < b$.

$$P(X \in [c, d]) = \int_c^d f_X(x)\,dx$$

$$= \int_c^d \frac{1}{b-a}\,dx$$

THIS PROPERTY CHARACTERIZES THE UNIFORM RANDOM VARIABLE

$$= \frac{d-c}{b-a}.$$

Note that the probability that $X$ is in $[c, d]$ depends only on the length of the interval $[c, d]$.

WARNING: The above property is true only if $[c, d] \subseteq [a, b]$. For example, if $X$ is Uniform $[0, 1]$, then

$$P\left(X \in \left[0, \tfrac{1}{2}\right]\right) = \frac{\frac{1}{2} - 0}{1 - 0} = \frac{1}{2},$$

whereas

$$P\left(X \in \left[1, \tfrac{3}{2}\right]\right) = 0.$$

Both $\left[0, \frac{1}{2}\right]$ and $\left[1, \frac{3}{2}\right]$ are intervals of length $\frac{1}{2}$, yet their probabilities are different.

(3) $$\boxed{E[X] = \frac{a+b}{2}}$$

This is quite an intuitive result. All values in the interval $[a, b]$ are "equally likely". Hence, the mean value or the average value should be the mid-point of the interval.

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

$$= \int_a^b x \cdot \frac{1}{b-a}\, dx$$

$$= \frac{1}{b-a} \left[ \frac{x^2}{2} \right]_a^b$$

$$= \frac{b^2 - a^2}{2(b-a)}$$

$$= \frac{a+b}{2} .$$

(4) $$V(X) = \frac{(b-a)^2}{12} .$$

NOTE: This shows that the variance is a function of the length of the interval $[a, b]$ only.

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

$$= \int_{-\infty}^{\infty} x^2 f_X(x)\, dx - \left(\frac{a+b}{2}\right)^2$$

$$= \int_a^b \frac{x^2}{b-a}\, dx - \left(\frac{a+b}{2}\right)^2$$

$$= \frac{b^3 - a^3}{3(b-a)} - \left(\frac{a+b}{2}\right)^2$$

$$= \frac{1}{12} \left[ 4(b^2 + ab + a^2) - 3(a+b)^2 \right]$$

$$= \frac{1}{12} \left[ a^2 + b^2 - 2ab \right]$$

$$= \frac{1}{12} (b-a)^2 .$$

EXAMPLE: A researcher has been observing a certain volcano for a long time. He knows that an eruption is imminent and is equally likely to occur at any time in the next 24 hours.

(a) What is the probability that the volcano will not erupt for atleast 15 hours?

Let $X$ = Time (from now) that eruption of volcano occurs

Then, $X$ is Uniform $[0, 24]$ based on the specifications in the problem.

$$P(X > 15) = 1 - P(X < 15)$$
$$= 1 - \frac{15 - 0}{24 - 0}$$
$$= \frac{3}{8}.$$

(b) Find a time such that there is only a 10% chance that the volcano will not have erupted by that time.

We want to find $t$ such that

$$P(X > t) = 10\%$$

$$\Rightarrow \quad 1 - \frac{t - 0}{24 - 0} = \frac{1}{10}$$

$$\Rightarrow \quad t = \frac{24 \times 9}{10}$$

$\Rightarrow$ $t = 21.6$[redacted] hours.

# LECTURE - 21

<u>Agenda:</u>

1. Exponential distribution

There are several random variables X that are observed in real life experiments, which have the property that "the chance that X takes values close to $x$, decreases exponentially with $x$.". Often, Lifetimes of various objects have been found to satisfy this property. How do we convert this intuitive notion into the mathematical form for the density function of X?

Firstly, the assumption essentially says that

$$P(X \in (x-h, x+h)) \propto e^{-\lambda x}$$ for some $\lambda > 0$ ($\lambda$ can be thought of as the rate of decay).

$$\Rightarrow \quad P(X \in (x-h, x+h)) = c(h)\, e^{-\lambda x},$$

where $c(h)$ is a constant depending on $h$.

$$\Rightarrow f_X(x) = \lim_{h \downarrow 0} \frac{P(X \in (x-h, x+h))}{h} = e^{-\lambda x} \left( \lim_{h \downarrow 0} \frac{c(h)}{h} \right)$$

Let us assume that $\lim_{h \downarrow 0} \frac{c(h)}{h} = K$.

Since lifetimes are always non-negative, we have proved until this point that

$$f_X(x) = \begin{cases} Ke^{-\lambda x} & \text{if } x \geq 0, \\ 0 & \text{otherwise} \end{cases}$$

Note that $K$ is unknown. But $f_X$ is a density function. Hence,

$$\int_{-\infty}^{\infty} f_X(x)\,dx = K\int_0^{\infty} e^{-\lambda x}\,dx = 1 \Rightarrow K = \lambda.$$

Hence,

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

For reasons which will become clear shortly, we will define the density in terms of $\theta = \frac{1}{\lambda}$.

DEFINITION: A random variable $X$ is said to have an Exponential distribution with parameter $\theta$, if

$$f_X(x) = \begin{cases} \frac{1}{\theta} e^{-\frac{x}{\theta}} & \text{if } x \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

- $F_X(x) = P(X \leq x) = \int_0^x \frac{1}{\theta} e^{-t/\theta}\, dt$ if $x \geq 0$

$$= \left[-e^{-t/\theta}\right]_0^x \quad \text{if } x \geq 0$$

$$= 1 - e^{-x/\theta} \quad \text{if } x \geq 0$$

Hence, $F_X(x) = \begin{cases} 1 - e^{-x/\theta} & \text{if } x \geq 0, \\ 0 & \text{if } x < 0. \end{cases}$

- $E[X] = \int_0^\infty \frac{x}{\theta} e^{-x/\theta} dx$ (Substitute $y = \frac{x}{\theta}$)

$$= \theta \int_0^\infty y\, e^{-y} dy$$

$$= 1 \quad \text{(Integration by parts).}$$

However, there is an easier way of evaluating integrals of the form $\int_0^\infty x^n e^{-x} dx$.

In mathematics, the "Gamma function" is defined as

$$\boxed{\Gamma(\alpha) \triangleq \int_0^\infty x^{\alpha-1} e^{-x} dx \quad \text{for } \alpha > 0}$$

FACT: $\Gamma(\alpha+1) = \alpha \Gamma(\alpha)$.

FACT: $\Gamma(n) = (n-1)!$ for every positive integer $n$.

Hence, $\Gamma(1) = \int_0^\infty e^{-x}\,dx = 0! = 1$

$$\Gamma(2) = \int_0^\infty x\,e^{-x}\,dx = 1! = 1$$

$$\Gamma(3) = \int_0^\infty x^2 e^{-x}\,dx = 2! = 2$$

$$\Gamma(4) = \int_0^\infty x^3 e^{-x}\,dx = 3! = 6$$

and so on.

- $V(X) = \int_0^\infty \frac{x^2}{\theta} e^{-x/\theta}\,dx - (E(X))^2$

Substituting $y = \frac{x}{\theta}$, we get

$$V(X) = \theta^2 \int_0^\infty y^2 e^{-y}\,dy - \theta^2$$

$$= \theta^2\,\Gamma(3) - \theta^2$$
$$= 2\theta^2 - \theta^2$$
$$= \theta^2.$$

MEMORYLESS PROPERTY OF THE EXPONENTIAL DISTRIBUTION

Note that we had seen the memoryless property of the geometric distribution. It is the only discrete distribution with this property. The exponential distribution also shares this property. It is the only continuous distribution with this property. To see this,

$$P(X \geq a+b \mid X \geq a) = \frac{P(X \geq a+b)}{P(X \geq a)}$$

$$= \frac{1-(1-e^{-\frac{a+b}{\theta}})}{1-(1-e^{-\frac{a}{\theta}})}$$

$$= e^{-\frac{b}{\theta}}$$

$$= P(X \geq b).$$

Hence, given that an exponential random variable is greater than a, the conditional probability that it is greater than a+b is the same as the unconditional probability that it is greater than b.

<u>Example:</u> The magnitudes of earthquakes recorded in a region of North America can be modeled by an exponential distribution with a mean of 2.4, as measured on the Richter scale.

(a) Find the probability that the next earthquake will be no more than 2.5 on the Richter scale.

Note that if $X$ = Magnitude of next earthquake, then $X$ is an exponential random variable with parameter 2.4. Hence,

$$P(X \le 2.5) = 1 - e^{-\frac{2.5}{2.4}} =$$

(b) Given that the next earthquake will be more than 2 on the Richter scale, what is the probability that it will be more than 3 on the Richter scale.

By the memoryless property of the exponential distribution,

$$P(X \ge 3 \mid X \ge 2) = P(X \ge 1)$$
$$= e^{-\frac{1}{2.4}}$$
$$=$$

~~THE GAMMA DISTRIBUTION~~

~~[illegible]~~

Agenda:

(1) Gamma random variable

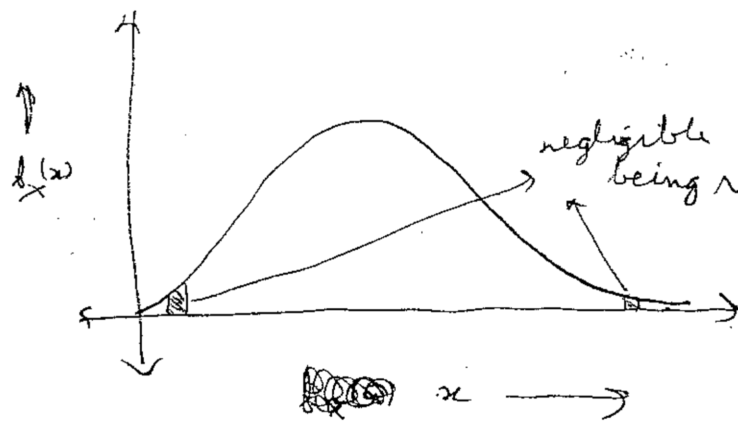
Some random variables have the property that $P(X \in (x-h, x+h))$ increases as we move $x$ away from zero, and then decreases again as we get close to large values of $x$. In other words, the chance of observing very small or very large values is negligible, and most of the values observed are close to the average associated with the random variable $X$, i.e., $E(X)$.

A random variable $X$ is said to be a Gamma random variable if

$$\mathcal{X} = \text{Range}(X) = (0, \infty)$$

$$f_X(x) = \begin{cases} \dfrac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-x/\beta}, & \text{if } x \geq 0, \\ 0 & \text{if } x < 0. \end{cases}$$

Here $\alpha, \beta > 0$ are fixed parameters for the distribution




NOTE THAT

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

= AREA UNDER $f_X$ FROM a TO b.

Note that when $\alpha = 1$, the gamma density reduces to the exponential density with parameter $\beta$. Hence the exponential random variable is a special case of the gamma random variable with $\alpha = 1$.

The first thing that should be verified is, whether the density integrates to 1.

$$\int_{-\infty}^{\infty} f_X(x)\, dx = \int_0^{\infty} \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}\, dx$$

(Substitute $y = \frac{x}{\beta}$)

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

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

$$= 1.$$

$$\boxed{\text{FACT: If } \alpha \geq 1, \quad \Gamma(\alpha+1) = \alpha\Gamma(\alpha)}$$

We use this fact to derive $E(X)$.

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

$$= \int_0^\infty x \cdot \frac{x^{\alpha-1} e^{-\frac{x}{\beta}}}{\Gamma(\alpha)\beta^\alpha} dx$$

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

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

($\because$ Substitute $y = \frac{x}{\beta}$)

$$= \frac{\beta}{\Gamma(\alpha)} \Gamma(\alpha+1)$$

$$= \beta\alpha .$$

($\because$ $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$)

One can also establish that $V(X) = \beta^2\alpha$.

NORMAL DISTRIBUTION

The most widely used continuous random variab in probability and its applications is the "normal random variable". A random variab X is said to be a normal random variable with parameters $\mu \in \mathbb{R}$ and $\sigma > 0$, if

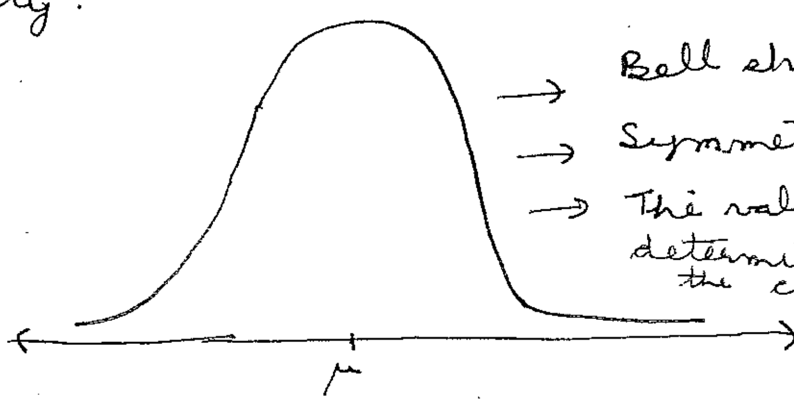
$\mathcal{X} = \text{Range}(X) = \mathbb{R}$, and

$$f_X(x) = \frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}}, \quad x \in \mathbb{R}.$$

What do the parameters $\mu$ and $\sigma$ mean in terms of the random variable X?

$$\boxed{\text{RESULT:} \quad E(X) = \mu \text{ and } V(X) = \sigma^2}$$

Let us postpone the proof of this result for later and look at the shape of the normal density.




→ Bell shaped curve

→ Symmetric around $\mu$

→ The value of $\sigma$ determines the spread of the curve.

The normal distribution works as a good model for a lot of measurements that are observed in real experiments. There is a valid reason for this and we will see that a few weeks later, but for now let us contend ourselves with the knowledge that if the random quantity under consideration is an average of independent random quantities, then the normal distribution is quite likely an appropriate model for that random quantity.

## THE STANDARD NORMAL DISTRIBUTION

<u>Definition</u>: If $Z$ is a normal random variable with parameters $\mu = 0$ and $\sigma = 1$, then $Z$ is said to be a "standard normal random variable".

Note that, by definition, the density of a standard normal random variable is given by

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \quad \text{for } z \in \mathbb{R}.$$

Let us calculate $E[Z]$ and $V[Z]$.

$$E[Z] = \int_{-\infty}^{\infty} z f_Z(z)\, dz$$

$$= \int_{-\infty}^{\infty} \frac{z e^{-\frac{z^2}{2}}}{\sqrt{2\pi}}\, dz$$

Here is a nice trick to evaluate this integral. Substitute $y = -z$.

$$E[Z] = -\int_{-\infty}^{\infty} \frac{y e^{-\frac{y^2}{2}}}{\sqrt{2\pi}}\, dy = -E[Z].$$

Hence, $E[Z] = 0$.

The main reason which this trick worked is that

$g(z) = \frac{z e^{-z^2/2}}{\sqrt{2\pi}}$ is an "odd function", i.e.,

$$g(z) = \frac{z e^{-z^2/2}}{\sqrt{2\pi}} = -\left\{\frac{(-z)\, e^{-(-z)^2/2}}{\sqrt{2\pi}}\right\} = -g(-z) \text{ for } z \in \mathbb{R}$$

$$\begin{aligned} V(Z) &= E(Z^2) - (E(Z))^2 \\ &= E(Z^2) \\ &= \int_{-\infty}^{\infty} z^2 \frac{e^{-\frac{z^2}{2}}}{\sqrt{2\pi}}\, dz \\ &= \frac{1}{\sqrt{2\pi}} \int_{0}^{\infty} u^{1/2} e^{-\frac{u}{2}}\, du \end{aligned}$$

$(\because$ Substitute $u = z^2)$

$$= \frac{1}{\sqrt{2\pi}} \Gamma\left(\frac{3}{2}\right) (2)^{3/2}$$

$(\because$ By the definition of the gamma function$)$

$$= 1$$

$(\because$ By the property of the gamma function$)$

RESULT: If $Z$ is a standard normal random variable, then for any $\mu \in \mathbb{R}$ and $\sigma \geq 0$,

$X = \mu + \sigma Z$ is a normal random variable with parameters $\mu$ and $\sigma$, i.e., the probability density function of $X$ is given by

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \text{ for } x \in \mathbb{R}.$$

Proof of $E(X) = \mu$ and $V(X) = \sigma^2$

Remember the result that we had stated earlier about $E(X)$ and $V(X)$. Let us prove the result using the property described above.

$$E(X) = E(\mu + \sigma Z) = \mu + \sigma E(Z) = \mu + 0 = \mu$$

$$V(X) = V(\mu + \sigma Z) = V(\sigma Z) = \sigma^2 V(Z) = \sigma^2$$

# LECTURE – 23

Agenda:

1. The standard normal distribution function and its properties.
2. Examples

## THE STANDARD NORMAL DISTRIBUTION FUNCTION

Let us recollect that $Z$ is said to be a standard normal random variable, if the probability density of $Z$ is given by

$$f_Z(z) = \frac{e^{-\frac{z^2}{2}}}{\sqrt{2\pi}} \quad \text{for } z \in \mathbb{R}.$$

The distribution function of a standard normal random variable has a special name and place in the theory of probability.

**Definition**: The "$\Phi$ - function" from $\mathbb{R}$ to $[0,1]$ is defined as

$$\Phi(z) = \int_{-\infty}^{z} \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}\, dx = P(Z \leq z)$$

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

Hence $\Phi$ is essentially the distribution function of a standard normal random variable. It cannot be evaluated in closed form (except at special values), but with modern computing technology there are programs which will give you the value of $\Phi$ at any given point. Let us study some of the important properties of this function.

PROPERTY 1: $0 \le \Phi(z) \le 1$ for $z \in \mathbb{R}$

This follows as $\Phi(z) = P(Z \le z)$ where $Z$ is $\underbrace{N(0,1)}_{\text{notation for standard normal}}$

PROPERTY 2: $\Phi(z) + \Phi(-z) = 1$

$$\Phi(z) + \Phi(-z) = \int_{-\infty}^{z} \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}\,dx + \int_{-\infty}^{-z} \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}\,dx$$

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

(Substitute $y = x$ in the first integral) (Substitute $y = -x$ in the second integral)

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

$$= \int_{-\infty}^{\infty} \frac{e^{-y^2/2}}{\sqrt{2\pi}}$$

$$= 1.$$

PROPERTY 3: If $X$ is a normal random variable with mean $\mu$ and variance $\sigma^2$, then

$$F_X(x) = P(X \le x) = \Phi\left(\frac{x-\mu}{\sigma}\right).$$

Note that if $Z$ is $N(0,1)$, then $\mu + \sigma Z$ is $N(\mu, \sigma^2)$. Hence,

$$P(X \le x) = P(\mu + \sigma Z \le x) = P\left(Z \le \frac{x-\mu}{\sigma}\right) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

Example: Sick leave time used by employees of a firm in the course of 1 month has approximately a normal distribution with a mean of 180 hours and a variance of 350 hours. Find the probability that the total sick leave for a given month is less than 150 hours. Your answer may be expressed in terms of the $\Phi$-function.

Let $X$ = Total sick leave time.

Then $X$ is $N(150, 350)$. $(\mu = 150, \sigma^2 = 350)$.

$$P(X < 150) = P(X \le 150) \quad (\because X \text{ is continuous r.v. hence } P(X=150)=0)$$

$$= \Phi\left(\frac{150-180}{\sqrt{350}}\right)$$

$$= \Phi\left(\frac{-30}{\sqrt{350}}\right)$$

$$= 0.05440$$

Note that $\Phi(0) + \Phi(-0) = 1$. Hence $\Phi(0) = P(Z \le 0) = \frac{1}{2}$.

Hence, if we want to find the probability that $X$ is less than its mean 180, then

$$P(X < 180) = P(X \le 180)$$

($\because X$ is a continuous random variable, hence $P(X=180)=0$)

$$= \Phi\left(\frac{180-180}{\sqrt{350}}\right)$$

$$= \Phi(0)$$

$$= \frac{1}{2}$$

Here are some other facts about the $\Phi$-function.

$$\Phi(1) - \Phi(-1) = P(-1 \le Z \le 1) \approx 68\%$$
$$\Phi(2) - \Phi(-2) = P(-2 \le Z \le 2) \approx 95\%$$
$$\Phi(3) - \Phi(-3) = P(-3 \le Z \le 3) = 99.7\%$$

This means that if $X$ is $N(\mu, \sigma^2)$, then

$$P(\mu - \sigma \leq X \leq \mu + \sigma) = P\left(-1 \leq \frac{X-\mu}{\sigma} \leq 1\right)$$

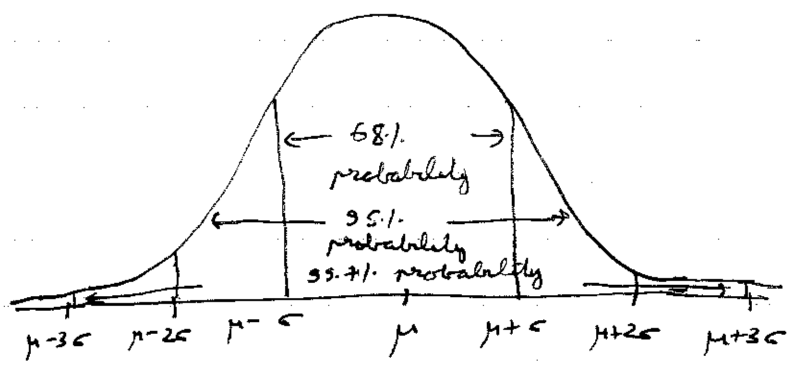
$$= P(-1 \leq Z \leq 1)$$

$$\approx 68\%$$

Similarly,

$$P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) \approx 95\%$$

$$P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) = 99.7\%$$




Hence if $X$ is $N(\mu, \sigma^2)$, there is 99.7% probability that it takes a value in the interval $[\mu-3\sigma, \mu+3\sigma]$. Since the length of the interval $[\mu-3\sigma, \mu+3\sigma]$ is $6\sigma$, it is used as the "$6\sigma$" rule often used in quality control divisions.

Example: Suppose that men's neck sizes are approximately normally distributed with a mean of 16.2 inches and variance of 0.81 inches. Find the probability that the neck size of a randomly chosen man lies between 13.5 inches and 18.9 inches.

$X$ = Man's neck size.

$X$ is $N(16.2,\ 0.81)$, $(\mu = 16.2,\ \sigma^2 = 0.81)$

$$P(13.5 \le X \le 18.9) = P\left(-3 \le \frac{X - 16.2}{0.9} \le 3\right)$$

$$= P(-3 \le Z \le 3)$$

$$\approx 99.7\%$$

LECTURE - 24

<u>Agenda:</u>

1. Beta distribution
2. Moment generating function

BETA DISTRIBUTION

Every continuous distribution that we have encountered, except the uniform, takes values over an infinite interval ( $(0, \infty)$ or $\mathbb{R}$). The beta distribution is an alternative model for random variables which are constrained to lie in the interval $(0, 1)$.

A random variable $X$ is said to have a Beta$(\alpha, \beta)$ distribution (where $\alpha$ and $\beta$ are fixed constants) if

$$\mathcal{X} = \text{Range}(X) = (0, 1)$$

$$f_X(x) = \begin{cases} \dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} & \text{if } 0 \le x \le 1, \\ 0 & \text{if } x \notin [0, 1]. \end{cases}$$

The first thing that we need to check is
$$\int_{-\infty}^{\infty} f_X(x)\,dx = 1.$$

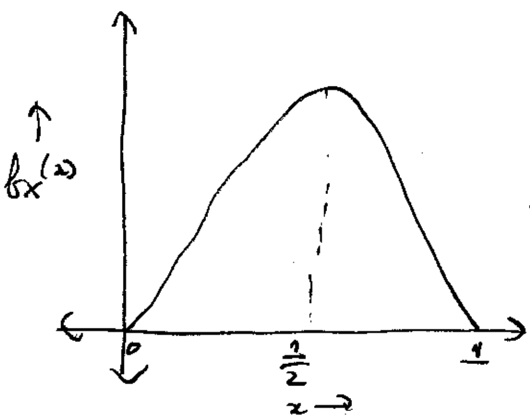
To prove this, we use the following identity.

**RESULT:** If $\alpha, \beta$ are positive constants, then
$$\int_0^1 x^{\alpha-1}(1-x)^{\beta-1}\,dx = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

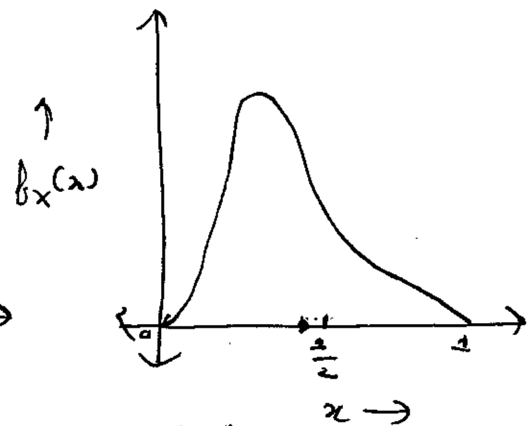
The proof of this identity is omitted. Hence,
$$\int_{-\infty}^{\infty} f_X(x)\,dx = \int_0^1 \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \cdot x^{\alpha-1}(1-x)^{\beta-1}\,dx$$
$$= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \, \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$
$$= 1.$$

The distribution function $F_X$ is not available in closed form, but can be evaluated using various software packages. The Beta density can take a varie of shapes which points to the flexibility o this collection of random variables.

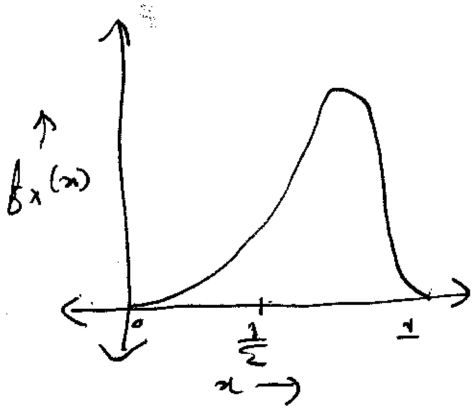
**RESULT:** The Uniform $[0,1]$ random variable is a special case of the Beta random variable with $\alpha = 1, \beta = 1$.



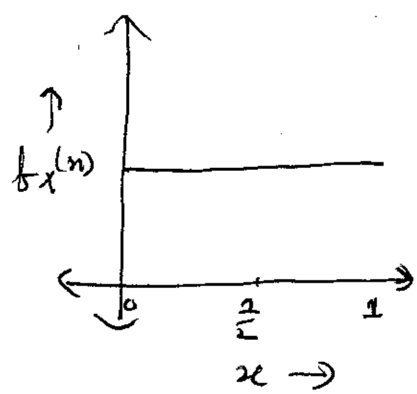

$\alpha = \beta > 1$




$1 < \alpha < \beta$




$\alpha > \beta > 1$




$\alpha = \beta = 1$

The graph of the beta density for four possible situations is illustrated.

Let us calculate $E[X]$.

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

$$= \int_0^1 x \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1}\, dx$$

$$= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 x^{\alpha} (1-x)^{\beta-1}\, dx$$

$$= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha+1)\Gamma(\beta)}{\Gamma(\alpha+\beta+1)}$$

$$= \frac{\Gamma(\alpha+1)}{\Gamma(\alpha)} \Big/ \frac{\Gamma(\alpha+\beta+1)}{\Gamma(\alpha+\beta)}$$

$$= \alpha / \alpha+\beta .$$

($\because$ By the properties of the Gamma function).

Similarly,

$$V(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

<u>Example:</u> A gasoline wholesale distributor uses bulk storage tanks to hold a fixed supply. The tanks are

filled every Monday. Of interest to the wholesaler is the proportion of the supply sold during the week. Over many weeks, this proportion has been observed to match fairly well a beta distribution with $\alpha = 4$ and $\beta = 2$.

(a) Find the expected value of this proportion

$$E[X] = \frac{\alpha}{\alpha+\beta} = \frac{4}{4+2} = \frac{2}{3}.$$

(b) Find the probability that the wholesaler will sell atleast 90% of the stock in a given week.

$$P(X \geq 0.9) = \int_{0.9}^{\infty} f_X(x)\,dx$$

$$= \int_{0.9}^{1} \frac{\Gamma(6)}{\Gamma(4)\Gamma(2)} x^{4-1}(1-x)^{2-1}\,dx$$

$$= \frac{5!}{3!\,1!} \int_{0.9}^{1} x^3(1-x)\,dx$$

$(\because \Gamma(\alpha) = (\alpha-1)!$ if $\alpha$ is a positive integer$)$

$$= 20 \int_{0.9}^{1} (x^4 - x^3)\,dx$$

$$= 20\,(0.004)$$

$= \quad 0.08.$

Hence there is an 8% chance that the wholesaler will sell atleast 90% of the stock in a given week.

MOMENT GENERATING FUNCTION

For a random variable $X$, (discrete or continuous), the moment generating function is a special function associated to it. The moment generating function is a powerful theoretical tool. The name "moment generating function" comes from the fact that it can be used to obtain the "moments" of the random variable. We clarify this shortly.

<u>Definition:</u> If $X$ is a discrete or continuous random variable, then the moment generating function of $X$ is denoted by $M_X : \mathbb{R} \to (0, \infty)$, and is defined by

$$M_X(t) = E[e^{tX}] = \begin{cases} \sum_{x \in \mathcal{X}} e^{tx} P(X = x), & \text{if } X \text{ is discrete}, \\ \int_{-\infty}^{\infty} e^{tx} f_X(x)\, dx, & \text{if } X \text{ is continuous}. \end{cases}$$

IMPORTANT: The moment generating function is defined for only those $t$, where the sums or integrals considered in the definition exist.

For a random variable $X$, the values

$E(X), E(X^2), E(X^3), \ldots\ldots$ are known

as the moments of the random variable. They provide useful information about the random variable.

**RESULT:**

$$\frac{d}{dt} M_X(t)\Big|_{t=0} = E(X)$$

$$\frac{d^2}{dt^2} M_X(t)\Big|_{t=0} = E(X^2)$$

Infact, for any positive integer $k$,

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

Hence the moment generating function can be used to compute the moments of the random variable (provided that the moment generating function is well defined in an interval around zero).

# LECTURE 25

**Agenda**

① Moment generating functions (MGFs)

② Properties

In the last lecture, we defined the notion of a moment generating functions. Today, we will derive the moment generating functions for some standard random variables, and learn a very useful property of moment generating functions

BINOMIAL MGF

If $X$ is Binomial $(n, p)$, then

$$M_X(t) = E[e^{tX}]$$

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

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

$$= (pe^t + 1-p)^n$$

(By the binomial theorem,
$$\sum_{x=0}^{n} \binom{n}{x} a^x b^{n-x} = (a+b)^n \Big)$$

Note that

$$\frac{d}{dt} M_X(t) = \frac{d}{dt}\left[(pe^t + 1-p)^n\right]$$

$$= npe^t (pe^t + 1-p)^{n-1}$$

Hence,

$$\left.\frac{d}{dt} M_X(t)\right|_{t=0} = npe^0 (pe^0 + 1-p)^{n-1}$$

$$= np.$$

$$= E[X]$$

We just verified that the first derivative of the moment generating function evaluated at 0, gives us the expected value

STANDARD NORMAL MGF

If $Z$ IS NORMAL$(0,1)$, then

$$M_Z(t) = E[e^{tZ}]$$

$$= \int_{-\infty}^{\infty} e^{tz} \frac{e^{-\frac{z^2}{2}}}{\sqrt{2\pi}} dz$$

$$= e^{\frac{t^2}{2}} \int_{-\infty}^{\infty} \underbrace{\frac{e^{-\frac{(t-z)^2}{2}}}{\sqrt{2\pi}}}_{\text{Density of Normal}(\mu=t,\ \sigma^2=1)} dz$$

$$= e^{\frac{t^2}{2}}.$$

GAMMA MGF

If $X$ IS GAMMA$(\alpha, \beta)$, then

$$M_X(t) = E[e^{tX}]$$

$$= \int_0^{\infty} e^{tx} \frac{x^{\alpha-1} e^{-\frac{x}{\beta}}}{\Gamma(\alpha)\beta^{\alpha}} dx$$

$$= \int_0^{\infty} \frac{x^{\alpha-1} e^{-x\left(\frac{1}{\beta}-t\right)}}{\Gamma(\alpha)\beta^{\alpha}} dx$$

Note that if $t \geq \frac{1}{\beta}$, then the integral is infinite. If $t < \frac{1}{\beta}$, then

$$M_X(t) = \int_0^\infty \frac{x^{\alpha-1} e^{-x\left(\frac{1}{\beta}-t\right)}}{\Gamma(\alpha)\beta^\alpha} \, dx$$

Let $y = x\left(\frac{1}{\beta} - t\right)$. Then,

$$M_X(t) = \frac{\left(\frac{1}{\beta}-t\right)^{-\alpha}}{\Gamma(\alpha)\beta^\alpha} \int_0^\infty y^{\alpha-1} e^{-y} \, dy$$

$$= \frac{\Gamma(\alpha)}{\Gamma(\alpha)(1-\beta t)^\alpha}$$

$$= \frac{1}{(1-\beta t)^\alpha}.$$

PROPERTY OF MGF: If $X$ and $Y$ are two random variables, such that

$$M_X(t) = M_Y(t) \text{ for every } t \in \mathbb{R},$$

(assuming that $M_X(t) < \infty$ for an interval around 0), then $X$ and $Y$ have the same distribution.

APPLICATION:

Let $Z$ be STANDARD NORMAL. We are interested in finding the distribution of $Z^2$. Note that

$$M_{Z^2}(t) = E[e^{tZ^2}]$$

$$= \int_{-\infty}^{\infty} e^{tz^2} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \, dz$$

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

This is infinity if $t \geq \frac{1}{2}$.

If $t < \frac{1}{2}$

$$= \frac{1}{\sqrt{1-2t}} \left\{ \int_{-\infty}^{\infty} \underbrace{\sqrt{\frac{1-2t}{2\pi}} \, e^{-\frac{z^2}{2}(1-2t)}}_{\text{Density of Normal}\left(\mu=0,\, \sigma^2=\frac{1}{1-2t}\right)} dz \right\}$$

$$= \frac{1}{\sqrt{1-2t}}$$

This is exactly the MGF of a Gamma$\left(\alpha=\frac{1}{2}, \beta=2\right)$ random variable.

Hence $Z^2$ has a Gamma$\left(\alpha=\frac{1}{2}, \beta=2\right)$ distribution.

# LECTURE – 26

Agenda:

1. Joint probability distributions for discrete random variable
2. Examples

## JOINT PROBABILITY DISTRIBUTIONS

Until now we have been studying situations, where we are only interested in one random quantity (a random variable) arising out of an experiment. But there are various situations, when there is more than one quantity associated with the experiment, are we are interested in the JOINT BEHAVIOUR of these random quantities. Let us consider a simple motivating example.

On April 15, 1912, the ocean liner Titanic collided with an iceberg and sank. Of the 2201 passengers on board, 1495 perished. The question as to whether passenger class was related to survival has been discussed extensively. Here is a table summarizing the details of number of survivors by class.

| Passenger Status | Survivors | Fatalities | Total |
|---|---|---|---|
| First class | 203 | 122 | 325 |
| Second class | 118 | 167 | 285 |
| Third class | 178 | 528 | 706 |
| Crew | 212 | 673 | 885 |
| Total | 711 | 1490 | 2201 |

There are two discrete random variables in play here,

$$X = \begin{cases} 0 & \text{if passenger survived,} \\ 1 & \text{if passenger did not survive} \end{cases}$$

$$Y = \begin{cases} 1 & \text{if passenger was in first class,} \\ 2 & \text{if passenger was in second class} \\ 3 & \text{if passenger was in third class} \\ 4 & \text{if passenger was a crew member} \end{cases}$$

Experiment: Select a Titanic passenger randomly. What is the chance that he/she is a non-crew member who survived? To answer questions like these, we develop a concept of the joint probability mass function of two discrete random variables.

DEFINITION: Let $X, Y$ be discrete random variable. Let $\mathcal{X} = \text{Range}(X)$, $\mathcal{Y} = \text{Range}(Y)$. The JOINT PROBABILITY MASS FUNCTION of $(X, Y)$ is defined as

$$p_{X,Y}(x, y) = P(X = x, Y = y) \text{ for } x \in \mathcal{X}, y \in \mathcal{Y}.$$

Clearly,

$$\sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p_{X,Y}(x,y) = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} P(X=x, Y=y)$$

$$= P\Big(\bigcup_{\{x \in \mathcal{X},\, y \in \mathcal{Y}\}} \{X=x, Y=y\}\Big)$$

$$= 1 .$$

In the same fashion, we define the concept of a joint probability distribution function of two random variables $X$ and $Y$.

---

<u>DEFINITION:</u> Let $X$ and $Y$ be two random variables (arising out of an experiment). Then, the JOINT PROBABILITY DISTRIBUTION FUNCTION of $X$ and $Y$ is defined as

$$F_{X,Y}(a,b) = P(X \leq a, Y \leq b) \text{ for any } a, b \in \mathbb{R}.$$

If $X, Y$ are discrete, then

$$F_{X,Y}(a,b) = \sum_{x \in \mathcal{X} : x \leq a} \sum_{y \in \mathcal{Y} : y \leq b} P(X=x, Y=y).$$

---

Clearly,

① $\lim_{a \to -\infty} \lim_{b \to -\infty} F_{X,Y}(a,b) = P(X \leq -\infty, Y \leq -\infty) = 0$

② $\lim_{a \to \infty} \lim_{b \to \infty} F_{X,Y}(a,b) = P(X \leq \infty, Y \leq \infty) = 1$

③ $P(a < X \leq b, c < Y \leq d)$

$$= P(X \leq b, Y \leq d) - P(X \leq a, Y \leq d) - P(X \leq b, Y \leq c) + P(X \leq a, Y \leq c)$$

~~$\because$ Use $\{a < X \leq b, c < Y \leq d\} \cup \{X \leq a, Y \leq d\} \cup \{X \leq b, Y \leq c$~~

$$= F_{X,Y}(b,d) - F_{X,Y}(a,d) - F_{X,Y}(b,c) + F_{X,Y}(a,c)$$

④ For every fixed $a$, the function $F_{X,Y}$ is a right continuous in the second co-ordinate. For every fixed $b$, the function $F_{X,Y}$ is right continuous in the first co-ordinate.

It can be proved that these properties characterize a probability distribution function.

Getting back to the example, the joint probability density function of X and Y (based on the information provided in the table) is given by

| | | X | |
|---|---|---|---|
| | | 0 | 1 |
| Y | 1 | 0.09 | 0.06 |
| | 2 | 0.05 | 0.08 |
| | 3 | 0.08 | 0.24 |
| | 4 | 0.10 | 0.30 |

P( Non-crew member, who survived)

$$= P(X=0,\ Y \in \{1,2,3\})$$

$$= P(X=0, Y=1) + P(X=0, Y=2) + P(X=0, Y=3)$$

$$= 0.09 + 0.05 + 0.08$$

$$= 0.22$$

Using the joint probability mass function of X and Y, we can calculate the individual or marginal probability mass functions of X and Y.

**RESULT:** If X and Y are discrete random variables with joint probability mass function $p_{X,Y}$, then the individual or marginal probability mass functions of X and Y are given by

$$p_X(x) = P(X=x) = \sum_{y \in \mathcal{Y}} P(X=x, Y=y) = \sum_{y \in \mathcal{Y}} p_{X,Y}(x,y)$$

$$p_Y(y) = P(Y=y) = \sum_{x \in \mathcal{X}} P(X=x, Y=y) = \sum_{x \in \mathcal{X}} p_{X,Y}(x,y)$$

As an example, suppose we are interested in the probability that a randomly chosen passenger is a survivor.

Then,

$$\begin{aligned} P(\text{Survivor}) &= P(X=0) \\ &= \sum_{y=1}^{4} P(X=0, Y=y) \\ &= 0.32. \end{aligned}$$

However, many times we need to calculate conditional probabilities of events related to the random variables X and Y.

**DEFINITION:** Let $X$ and $Y$ be two discrete random variables. The conditional probability mass function of $X$ given $Y = y$ is defined as

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

Similarly, the conditional probability mass function of $Y$ given $X = x$ is defined as

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

In the Titanic examples what is the probability that a passenger survived given that he/she was a first class passenger?

P( Survived | First class passenger )

$$= P(X = 0 \mid Y = 1)$$

$$= \frac{P(X=0, Y=1)}{P(Y=1)}$$

$$= \frac{0.09}{0.15}$$

$= 0.06$

Getting back to the question of whether the class of a passenger was related to his or her survival, note that

| $y$ | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| $P(Y=y \mid X=0)$ | 0.28 | 0.16 | 0.25 | 0.31 |

| $y$ | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| $P(Y=y)$ | 0.15 | 0.13 | 0.32 | 0.40 |

If we compare the proportion of survivors for any given class, with the proportion of members of that class in the whole population, then we clearly see that there are more survivors in the first and second class and less in the third class and the crew members. Hence, one could make a point that the survival strategy was not equal to everyone. However, one should look into these issues more carefully.

LECTURE (27)

Agenda:

(1) Joint distribution functions of continuous random variables

(2) Examples

JOINT DISTRIBUTION FUNCTIONS OF CONTINUOUS RANDOM VARIABLES

We previously studied the joint probability mass functions for jointly describing the probability behaviour of two discrete random variables. Today, we repeat the same exercise for continuous random variables.

Let us recollect that a random variable $X$ is said to be continuous, if it has a density function $f_X$, such that

(i) $f_X(x) \geq 0$ for every $x \in \mathbb{R}$, $f_X(x) = 0$ for $x \notin \underset{=\text{Range}(}{\mathcal{X}}$

(ii) $P(a \leq X \leq b) = \underbrace{\int_a^b f_X(x)\,dx}_{\text{Area under the curve } f_X \text{ from } a \text{ to } b}$.

Suppose we have two continuous random variables X and Y. The joint probability behaviour of X and Y is described by the "joint probability density function" $f_{X,Y}$ which has the following properties.

(i) $f_{X,Y}(x,y) \geq 0$ for every $(x,y) \in \mathbb{R}^2$

(ii) $P(a \leq X \leq b,\ c \leq Y \leq d) = \underbrace{\int_c^d \int_a^b f_{X,Y}(x,y)\,dx\,dy}_{\text{Volume under surface } f_{X,Y} \text{ from } [c,d]\times[a,b].}$

It follows that the "joint probability distribution function" $F_{X,Y}$ is given by

$$F_{X,Y}(a,b) = P(X \leq a,\ Y \leq b) = \underbrace{\int_{-\infty}^{b} \int_{-\infty}^{a} f_{X,Y}(x,y)\,dx\,dy}_{\text{Volume under surface } f_{X,Y} \text{ from } [-\infty,b]\times(-\infty,a]}.$$

Example: A certain process for producing an industrial chemical yields a product that contains two main types of impurities. Let X denote the proportion of impurities of Type I and Y denotes the proportion of impurities of Type II. Suppose that the joint density of X and Y can be adequately modeled by the following function.

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

Compute $P(0 \leq X \leq 0.5, 0.4 \leq Y \leq 0.7)$

$$P(0 \leq X \leq 0.5,\ 0.4 \leq Y \leq 0.7)$$

$$= \int_{0.4}^{0.7} \int_{0}^{0.5} 2(1-x)\,dx\,dy$$

$$= \int_{0.4}^{0.7} \left[-(1-x)^2\right]_0^{0.5} dy$$

$$= \int_{0.4}^{0.7} 0.75\,dy$$

$$= (0.75) \int_{0.4}^{0.7} dy$$

$$= (0.75) \times (0.3)$$

$$= 0.225.$$

<u>RESULT:</u> If X and Y are continuous random variables with joint probability density function $f_{X,Y}$, then the individual or marginal density functions $f_X$ and $f_Y$ are given by the following:

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)\,dy \quad \text{for every } x \in \mathbb{R},$$

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y)\,dx \quad \text{for every } y \in \mathbb{R}.$$

In the industrial production example considered previously, compute $f_X$ and $f_Y$.

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

$$= \begin{cases} \int_0^1 2(1-x)\,dy & 0 \le x \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

$$= \begin{cases} 2(1-x)\,[y]_0^1 & 0 \le x \le 1, \\ 0 & \text{otherwise} \end{cases}$$

$$= \begin{cases} 2(1-x) & 0 \le \cancel{y} \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx$$

$$= \begin{cases} \int_0^1 2(1-x)\,dx & 0 \le y \le 1, \\ 0 & \text{otherwise} \end{cases}$$

$$= \begin{cases} 1 & 0 \le y \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

Suppose we are interested in the behaviour of the random variable X given that $Y = y$. To develop a framework for expressing this behaviour, we need the notion of "conditional probability density function".

**DEFINITION:** Let X and Y be continuous random variables with joint probability density function $f_{X,Y}$ and marginal densities $f_X$ and $f_Y$. Then the conditional probability density function of X given $Y = y$ is defined by

$$f_{X|Y=y}(x) = \begin{cases} \dfrac{f_{X,Y}(x,y)}{f_Y(y)} & y \in \mathcal{Y} = \text{Range}(Y), \\ 0 & \text{otherwise.} \end{cases}$$

Similarly, the conditional probability density function of Y given $X = x$, is defined by

$$f_{Y|X=x} = \begin{cases} \dfrac{f_{X,Y}(x,y)}{f_X(x)} & x \in \mathcal{X} = \text{Range}(X), \\ 0 & \text{otherwise.} \end{cases}$$

Compute the conditional probability density functions $f_{X|Y=y}$ and $f_{Y|X=x}$ for the industrial impurities example.

$$f_{X|Y=y} = \begin{cases} \frac{2(1-x)}{1} & 0 \le x \le 1, \ (0 \le y \le 1) \\ 0 & \text{otherwise.} \end{cases}$$

$$= \begin{cases} 2(1-x) & 0 \le x \le 1, \ (0 \le y \le 1), \\ 0 & \text{otherwise.} \end{cases}$$

$$f_{Y|X=x} = \begin{cases} \frac{2(1-x)}{2(1-x)} & 0 \le y \le 1, \ (0 \le x \le 1), \\ 0 & \text{otherwise.} \end{cases}$$

$$= \begin{cases} 1 & 0 \le y \le 1 \quad (0 \le x \le 1), \\ 0 & \text{otherwise.} \end{cases}$$

We see that $f_{X|Y=y}$ is the same as $f_X$, and $f_{Y|X=x}$ is the same as $f_Y$. This means that the random variables $X$ and $Y$ behave independently since fixing one to any value (in the appropriate range) does not affect the probability behaviour of the other. We consider this more formally in the next lecture.

LECTURE - (28)

<u>Agenda:</u>

(1) Independent Random Variables
(2) Expected values of functions of random variables

INDEPENDENT RANDOM VARIABLES

Let us recollect the notion of "independent events". We say that events A and B are independent if

$$P(A \cap B) = P(A)P(B) \quad \text{or} \quad P(A \mid B) = P(A).$$

The way we understand it intuitively is that even if we are given the information that B has occured, that does not change the probability of A. The same notion can be generalized to random variables. Let us consider the case of discrete random variables.

<u>Definition:</u> Let $X, Y$ be discrete random variables. Then, $X$ and $Y$ are said to be independent if

$$P(X=x, Y=y) = P(X=x)P(Y=y)$$

<u>for every $x \in \mathcal{X}$, $y \in \mathcal{Y}$.</u>

Note that

$$P(X = x, Y = y) = P(X = x)P(Y = y)$$

is the same as

$$P(X = x \mid Y = y) = P(X = x) \quad \text{or} \quad P(Y = y \mid X = x) = P(Y = y)$$

Hence, saying that $X$ and $Y$ are independent, means that even if we are given the information that $Y = y$, that does not change the probability behaviour of $X$. Similarly, even if we are given the information that $X = x$, that does not change the probability behaviour of $Y$.

Let us now turn our attention to continuous random variables.

Definition: Let $X, Y$ be continuous random variables. Then, $X$ and $Y$ are said to be independent if

$$f_{X,Y}(x, y) = f_X(x) f_X(y) \quad \text{for every } x \in \mathbb{R}, y \in \mathbb{R}.$$

Note that

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

is the same as

$$f_{X|Y=y}(x) = f_X(x) \quad \text{or} \quad f_{Y|X=x}(y) = f_Y(y).$$

Hence, saying that X and Y are independent, implies that the probability behaviour of X is unaffected by information about Y, and the probability behaviour of Y is unaffected by information about X.

<u>Example:</u> A bus arrives at a bus stop at a randomly selected time within a 1- hour period. A passenger arrives at the bus stop at a randomly selected time within the same hour. The passenger is willing to wait for the bus for up to $\frac{1}{4}$ of an hour. What is the probability that the passenger will catch the bus?

Let $X$ = Bus arrival time and $Y$ = Passenger arrival time

$X$ is uniform on $[0,1]$ and $Y$ is uniform on $[0,1]$. Hence,

$$f_X(x) = \begin{cases} 1 & \text{if } 0 \le x \le 1, \\ 0 & \text{otherwise.} \end{cases} \quad \text{and } f_Y(y) = \begin{cases} 1 & \text{if } 0 \le y \le \\ 0 & \text{otherwis} \end{cases}$$

Since X and Y are independent, the joint probability density of $(X, Y)$ is the product of the individual or marginal probability densities of X and Y.

Hence, the joint probability density of $X$ and $Y$ is given by

$$f_{X,Y}(x,y) = \begin{cases} 1 & \text{if } 0 \le x \le 1, 0 \le y \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

In terms of the random variables,

P( Passenger catches the bus)

$= P(Y \le X \le Y + \frac{1}{4})$

$= P((X,Y) \in A)$

where $A = \{(x,y) \in \mathbb{R}^2 : y \le x \le y + \frac{1}{4}\}$

**RESULT:** If $X$ and $Y$ are continuous random variables with joint density $f_{X,Y}$, then

$$P((X,Y) \in A) = \iint_A f_{X,Y}(x,y)\,dx\,dy$$ for any subset $A$ of $\mathbb{R}^2$.

Hence, in our particular example,

$$P((X,Y) \in A) = \iint_{\{(x,y)\in\mathbb{R}^2 :\; y \le x \le y+\frac{1}{4}\}} f_{X,Y}(x,y)\,dx\,dy$$

$$= \int_{-\infty}^{\infty}\left(\int_{y}^{y+\frac{1}{4}} f_{X,Y}(x,y)\,dx\right)dy$$

$$= \int_{0}^{1}\left(\int_{y}^{y+\frac{1}{4}} f_{X,Y}(x,y)\,dx\right)dy$$

(Since $f_{X,Y}(x,y) = 0$ if $y<0$ or $y>1$)

$$= \int_{0}^{3/4}\left(\int_{y}^{y+\frac{1}{4}} 1\,dx\right)dy + \int_{3/4}^{1}\left(\int_{y}^{1} 1\,dx\right)dy$$

(Since $f_{X,Y}(x,y) = 0$ if $x<0$ or $x>1$, and if $\frac{3}{4}<y$, then the interval $(y, y+\frac{1}{4})$ crosses the point $1$)

$$= \int_{0}^{3/4} \frac{1}{4}\,dy + \int_{3/4}^{1} (1-y)\,dy$$

$$= \frac{1}{4}\times\frac{3}{4} + \left[-\frac{(1-y)^2}{2}\right]_{3/4}^{1}$$

$$= \frac{3}{16} + \frac{1}{32}$$

$$= \frac{7}{32}.$$

Hence the probability that the passenger catches the bus is $\frac{7}{32}$.

## EXPECTED VALUES OF FUNCTIONS OF RANDOM VARIBLES

If $X, Y$ are discrete random variables with joint probability mass function $p_{X,Y}(x,y)$, then for any function $g: \mathcal{X} \times \mathcal{Y} \longrightarrow \mathbb{R}$ (where $\mathcal{X} = \text{Range}(X)$ and $\mathcal{Y} = \text{Range}(Y)$),

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

If $X, Y$ are continuous random variables with joint probability density function $f_{X,Y}(x,y)$, then for any function $g: \mathbb{R}^2 \longrightarrow \mathbb{R}$,

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

# LECTURE – 29

1. Probabilities involving two random variables
2. Expectations involving two random variables

We learnt in the last lecture that if $(X, Y)$ have joint density $f_{X,Y}$, then for any $A \subseteq \mathbb{R}^2$,

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

Also, typically in this course, the joint density of $(X,Y)$ will have the form

$$f_{X,Y}(x,y) = \begin{cases} h(x,y) > 0 & \text{if } (x,y) \in R \\ 0 & \text{otherwise.} \end{cases}$$

We call $R$ as the range of the joint density $f_{X,Y}$.
We now provide a general procedure for evaluating $P((X,Y) \in A)$.

(1) Note that $P((X,Y)\in A) = \iint_A f_{X,Y}(x,y)\,dx\,dy = \iint_{A \cap R} h(x,y)\,dx\,dy$

Hence the job is now to evaluate $\iint_{A\cap R} h(x,y)\,dx\,dy$.

(2) Figure out the range of $x$ values that can be taken in $A \cap R$.

- Call this set $\mathcal{X}_A$. For every $x \in \mathcal{X}_A$, find the projection set

$$A_x = \{ y : (x,y) \in A \cap R \}.$$

(3) Basic theory of calculus tells us that

$$\iint_{A \cap R} h(x,y)\, dy\, dx = \int_{\mathcal{X}_A} \int_{A_x} h(x,y)\, dy\, dx$$

In most examples, both $\mathcal{X}_A$ and $A_x$ will be either intervals or union of intervals. Hence, simplifying the integral further involves standard integration techniques.

<u>EXAMPLE:</u> Suppose that the joint density of $(X,Y)$ is given by

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

Find $P(X < 2, Y > 1)$.

Always convert the probability into the form $P((X,Y) \in A)$.
Here,

$$P(X<2, Y<1) = P((X,Y) \in A), \text{ where}$$

$$A = \{(x,y) : x<2, y<1\}.$$

① Note that the range of the density function is

$$R = \{(x,y) : 0 \le y \le x < \infty\}.$$

Hence $A \cap R = \{(x,y) : x<2, y<1, 0 \le y \le x < \infty\}$.

It follows that

$$P((X,Y) \in A) = \iint_{A \cap R} e^{-x}\, dy\, dx$$

② Clearly, the range of $x$ values that can be taken in $A \cap R$ is $\mathcal{X}_A = [0,2]$.

Fix $x \in \mathcal{X}_A$. ~~...~~ Then

$$\begin{aligned} A_x &= \{y : (x,y) \in A \cap R\} \\ &= \{y : y<1, 0 \le y \le x < \infty\} \\ &= \{y : 0 \le y \le \min(1,x)\} \end{aligned}$$

③ It follows that

$$P((X,Y) \in A) = \int_0^2 \int_0^{\min(1,x)} e^{-x}\, dy\, dx$$

$$= \int_0^2 \min(1,x)\, e^{-x}\, dx$$

$$= \int_0^1 x e^{-x}\, dx + \int_1^2 e^{-x} dx$$

$$= \int_0^1 \underset{\mathrm{I}}{x}\, \underset{\mathrm{II}}{e^{-x}}\, dx + \left[-e^{-x}\right]_1^2$$

(Integration by parts)

$$= \left[-xe^{-x}\right]_0^1 + \int_0^1 e^{-x} + (e^{-1} - e^{-2})$$

$$= e^{-0} - e^{-1} - e^{-2}$$

$$= 1 - e^{-1} - e^{-2}.$$

Hence,

$$P(X<2, Y<1) = 1 - e^{-1} - e^{-2}.$$

EXPECTATIONS OF FUNCTIONS OF TWO CONTINUOUS RANDOM VARIABLES

If $(X,Y)$ have joint density $f_{X,Y}$, then

$$E[g(X,Y)] \triangleq \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)\, f_{X,Y}(x,y)\, dx\, dy.$$

Note again that if $f_{X,Y}(x,y) = \begin{cases} h(x,y) > 0 & \text{if } (x,y) \in A, \\ 0 & \text{otherwise.} \end{cases}$

then,

$$E[g(X,Y)] = \iint_{R} g(x,y)\, h(x,y)\, dy\, dx.$$

Hence, the same process as earlier needs to be repeated with $R$ (instead of $A \cap R$) and $g(x,y)\, h(x,y)$ instead of $h(x,y)$. For example, if $g(x,y) = y$, then using the same joint density as in the previous example,

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

$$= \iint_{\{(x,y):\, 0 \le y \le x < \infty\}} y e^{-x}\, dy\, dx$$

$$= \int_{0}^{\infty}\int_{0}^{x} y e^{-x}\, dy\, dx$$

$$= \int_{0}^{\infty} e^{-x} \left[\frac{y^2}{2}\right]_{0}^{x} dx$$

$$= \int_{0}^{\infty} \frac{x^2 e^{-x}}{2}\, dx$$

$$= \frac{\Gamma(3)}{2}$$

$$= \frac{2!}{2}$$

$$= 1.$$

You need not always use the order $\boxed{dy\,dx}$. If it is more convenient, you can also use $\boxed{dx\,dy}$. For example, if $g(x,y)=x$, then

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

$$= \iint_{\{(x,y):\, 0\le y\le x<\infty\}} x e^{-x}\,dx\,dy$$

$$= \int_0^{\infty}\left(\int_y^{\infty} \underbrace{x}_{\mathrm{I}}\,\underbrace{e^{-x}}_{\mathrm{II}}\,dx\right)dy \qquad \text{(Integration by parts)}$$

$$= \int_0^{\infty}\left(\Big[-xe^{-x}\Big]_y^{\infty} + \int_y^{\infty} e^{-x}\,dx\right)dy$$

$$= \int_0^{\infty}\left(y e^{-y} + e^{-y}\right)dy$$

$$= \Gamma(2) + \Gamma(1)$$

$$= 1! + 0!$$

$$= 2.$$

LECTURE - (20)

Agenda:

1. General definition of independence
2. Covariance and correlation
3. Examples

GENERAL DEFINITION OF INDEPENDENCE

Definition: If $X$ and $Y$ are two random variables, then $X$ and $Y$ are said to be independent if for every function $g$ of $X$, and $h$ of $Y$,

$$E[g(X)h(Y)] = E[g(X)]E[h(Y)]$$

In the last lectures, we defined independence for a pair of discrete random variables, and also for a pair of continuous random variables. The above definition generalizes this notion of independence for an arbitrary pair of random variables.

COVARIANCE AND CORRELATION

Definition: The covariance between two random variables $X$ and $Y$ is defined as

$$Cov(X,Y) = E[(X-E(X))(Y-E(Y))].$$

Just as $E(X)$, $V(X)$ help us understand the probabilist behaviour of a single random variable $X$ in specific ways, $Cov(X,Y)$ helps us understand the joint probabilistic behaviour of two random variables $X$ and $Y$ as follows:

(1) If $Y$ tends to be large when $X$ is large, and $Y$ tends to be small when $X$ is small, then $X$ and $Y$ have a positive covariance

(2) If $Y$ tends to be large when $X$ is small, and $Y$ tends to be small when $X$ is large, then $X$ and $Y$ have a negative covariance.

Here are two properties of $Cov(X,Y)$ which are useful for practical purposes.

(i) $Cov(X,Y) = E[XY] - E(X)E(Y)$

(ii) $Cov(X,X) = V(X)$

<u>Definition:</u> The correlation coefficient between two random variables $X$ and $Y$ is defined as

$$\rho_{XY} = \frac{Cov(X,Y)}{\sqrt{V(X)\,V(Y)}}.$$

Properties of $\rho_{XY}$

① $\rho_{XY}$ is a unitless quantity

② $-1 \leq \rho_{XY} \leq 1$

③ If $\rho_{XY} = -1$, $Y = aX + b$ where $a < 0$.

④ If $\rho_{XY} = +1$, $Y = aX + b$ where $a > 0$.

⑤ $\rho_{XY}$ is a measure of "linear association" between $X$ and $Y$. As $\rho_{XY}$ becomes closer to zero, it indicates lesser and lesser linear relationship between $X$ and $Y$.

**Result:** If $X$ and $Y$ are independent,

$$\rho_{XY} = Cov(X, Y) = 0.$$

**Proof:** Choose $g(x) = x$ and $h(y) = y$. Then

$$E[g(X)h(Y)] = E[g(X)]E[h(Y)] \quad \text{(by independence)}$$

Hence,

$$E[XY] - E(X)E(Y) = 0.$$

However, the converse is not true, i.e.,

$$\operatorname{cov}(X,Y) = 0 \not\Longrightarrow X \text{ and } Y \text{ are independent}$$

Example: Consider discrete random variables $X$ and $Y$ with joint p.m.f. given by

$$P(X=x, Y=y) = \begin{cases} \frac{1}{8} & \text{if } (x,y) \neq (0,0), \\ 0 & \text{if } (x,y) = 0 \\ \mathcal{X} = \{-1, 0, 1\}, \\ \mathcal{Y} = \{-1, 0, 1\}. \end{cases}$$

$$E[XY] = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} (xy)\, P(X=x, Y=y)$$

$$= 0$$

$$E[X] = \sum_{x \in \mathcal{X}} x\, P(X=x) = 0$$

$$E(Y) = \sum_{y \in \mathcal{Y}} y\, P(Y=y) = 0$$

Hence, $E(XY) - E(X)E(Y) = 0$.

However,

$$P(X = -1, Y = -1) = \frac{1}{8}$$

$$P(X = -1)\, P(Y = -1) = \frac{3}{8} \cdot \frac{3}{8} = \frac{9}{64}$$

Hence,

$$P(X = -1, Y = -1) \neq P(X = -1) P(Y = -1),$$

which means $X$ and $Y$ are not independent.

Suppose we have a random experiment which gives rise to two random variables $X$ and $Y$. Suppose we are interested in $V(X - Y)$. Here is a general identity which helps us compute this quantity in terms of $V(X)$, $V(Y)$ and $Cov(X,Y)$.

<u>Result:</u> Let $X_1, X_2, ---, X_n$ and $Y_1, Y_2, ---, Y_m$ be two sequences of random variables. Let us define

$$U_1 = \sum_{i=1}^{n} a_i X_i \;,\quad U_2 = \sum_{j=1}^{m} b_j Y_j .$$

Then,

(1) $$V(U_1) = \sum_{i=1}^{n} a_i^2 V(X_i) + \sum_{1 \le i \ne j \le n}\sum a_i a_j Cov(X_i, X_j)$$

(2) $$Cov(U_1, U_2) = \sum_{i=1}^{n} \sum_{j=1}^{m} a_i b_j \, Cov(X_i, Y_j).$$

Example: Suppose that random variables $X$ and $Y$ have the joint density

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{2} & \text{if } 0 \le x \le y, \quad 0 \le y \le 2, \\ 0 & \text{otherwise.} \end{cases}$$

Calculate $V(Y-X)$.

$$\boxed{V(Y-X) = V(Y) + V(X) + 2(1)(-1)\,\mathrm{Cov}(X,Y)}$$

(Use the previous result with $U_1 = 1.Y + (-1)X$, i.e, $a_1 = 1,\ a_2 = -1,\ X_1 = Y,\ X_2 = X,\ n = 2$ )

$$E(X) = \int_0^2\int_0^y x \cdot \frac{1}{2}\,dx\,dy = \int_0^2 \frac{y^2}{4}\,dy = \frac{2}{3}.$$

$$E(Y) = \int_0^2\int_0^y y \cdot \frac{1}{2}\,dx\,dy = \int_0^2 \frac{y^2}{2}\,dy = \frac{4}{3}.$$

$$E(X^2) = \int_0^2\int_0^y x^2 \cdot \frac{1}{2}\,dx\,dy = \frac{2}{3}.$$

$$E(Y^2) = \int_0^2\int_0^y y^2 \cdot \frac{1}{2}\,dx\,dy = 2.$$

$$E(XY) = \int_0^2\int_0^y xy\,dx\,dy = \int_0^2 \frac{y^3}{4}\,dy = 1.$$

Using these values,

$$V(X) = E(X^2) - (E(X))^2 = \frac{2}{9},$$

$$V(Y) = E(Y^2) - (E(Y))^2 = \frac{2}{9}.$$

$$Cov(X, Y) = E(XY) - E(X)E(Y) = \frac{1}{9}.$$

Hence,

$$V(Y - X) = \frac{2}{9} + \frac{2}{9} - 2 \times \frac{1}{9} = \frac{2}{9}.$$

We have been using informally that if $X_1, X_2, \ldots, X_n$ are random variables which arise independently, then $V(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} V(X_i)$. Now, we are in a position to see a formal proof of this fact. Note that,

$$V\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} V(X_i) + \sum_{1 \le i \ne j \le n}\sum Cov(X_i, X_j)$$

$$= \sum_{i=1}^{n} V(X_i) + \sum_{1 \le i \ne j \le n}\sum 0$$

$$= \sum_{i=1}^{n} V(X_i).$$

LECTURE - 31

Agenda:

(1) Covariance and correlation (continued)
(2) Example: Variance of the hyperglometric

COVARIANCE AND CORRELATION

Let us recollect that covariance and correlation are two quantities often used to summarize the linear association between two random variables X and Y.

$$Cov(X, Y) = E(XY) - E(X)E(Y)$$

$$\rho_{XY} = \frac{Cov(X, Y)}{\sqrt{V(X)\,V(Y)}}$$

Here are two useful identities involving covariance.

**RESULT:** Let $Y_1, Y_2, ---, Y_n$ be a sequence of random variables. Then,

$$V\left(\sum_{i=1}^{n} a_i Y_i\right) = \sum_{i=1}^{n} a_i^2 V(Y_i) + 2 \sum_{i<j}\sum a_i a_j Cov(Y_i, Y_j)$$

**RESULT:** Let $Y_1, Y_2, ---, Y_n$ and $X_1, X_2, ---, X_m$ be two sequences of random variables. Then,

$$Cov\left(\sum_{i=1}^{n} a_i Y_i, \sum_{j=1}^{m} b_j X_j\right) = \sum_{i=1}^{n} \sum_{j=1}^{m} a_i b_j \, Cov(Y_i, X_j)$$

VARIANCE OF THE HYPERGEOMETRIC

Let us recollect the hypergeometric experiment. We have a collection of N objects, k are of type I and N-k are of Type II. We choose n objects without replacement from this collection. Let

$$X = \# \text{ of objects of Type I.}$$

Then X has a hypergeometric distribution.

We saw that $V(X) = n \frac{k}{N}\left(1 - \frac{k}{N}\right) \frac{N-n}{N-1}$.

Let us prove this identity.

Define the random variables $Y_1, Y_2, ---, Y_n$ as follows.

$$Y_i = \begin{cases} 1 & \text{if } i^{th} \text{ draw is of Type I,} \\ 0 & \text{otherwise.} \end{cases}$$

Then, clearly

$$X = \sum_{i=1}^{n} Y_i .$$

It follows that,

$$V(X) = \sum_{i=1}^{n} V(Y_i) + 2 \sum\sum_{i<j} Cov(Y_i, Y_j).$$

Let us figure out $V(Y_i)$ and $Cov(Y_i, Y_j)$.

Note that

$$P(X_1 = 1) = \frac{k}{N}. \Rightarrow \begin{array}{l} E(X_1) = 1.P(X_1=1) + 0.P(X_1=0) = \frac{k}{N}. \\ E(X_1^2) = \frac{k}{N}. \end{array}$$

$$P(X_2 = 1) = P(X_1 = 1) P(X_2 = 1 | X_1 = 1) + P(X_1 = 0) P(X_2 = 1 | X_1 = 0)$$

$$= \frac{k}{N} \cdot \underbrace{\frac{(k-1)}{N-1}}_{\text{The collection has N-1 objects with k-1 objects of Type I}} + \left(1 - \frac{k}{N}\right) \cdot \underbrace{\frac{k}{N-1}}_{\text{The collection has N-1 objects with k objects of Type I}}$$

$$= \frac{k(N-1)}{N(N-1)}$$

$$= \frac{k}{N}.$$

$$\Rightarrow E(X_2) = 1 \cdot P(X_2 = 1) + 0 \cdot P(X_2 = 0) = \frac{k}{N}.$$

Similarly, by conditioning on $\sum_{i=1}^{2} X_i$, one can prove that $P(X_3 = 1) = \frac{k}{N}$, $E(X_3) = \frac{k}{N}$, $E(X_3^2) = \frac{k}{N}$.

Continuing like this, we get,

$$E(X_i) = \frac{k}{N}, \quad E(X_i^2) = \frac{k}{N} \text{ for every } i = 1, 2, \ldots, n.$$

Hence,

$$V(X_i) = \frac{k}{N}\left(1 - \frac{k}{N}\right) \text{ for every } i = 1, 2, \ldots, n.$$

Note that, 
$$\begin{aligned} Cov(X_i, X_j) &= E(X_i X_j) - E(X_i)E(X_j) \\ &= \underbrace{P(X_i = 1, X_j = 1)}_{\substack{\text{The only way} \\ X_i X_j \text{ is not zero} \\ \text{is if } X_i = 1, X_j = 1}} - \left(\frac{k}{N}\right)^2 \end{aligned}$$

It is easy to see that 
$$P(X_1 = 1, X_2 = 1) = P(X_1 = 1)\, P(X_2 = 1 \mid X_1 = 1) = \frac{k}{N}\frac{(k-1)}{(N-1)}$$

Using similar techniques as above, one can prove for any $i < j$

$$P(X_i = 1, X_j = 1) = P(X_i = 1)\, P(X_j = 1 \mid X_i = 1) = \frac{k}{N}\frac{(k-1)}{(N-1)}$$

Hence, for every $i < j$

$$\mathrm{Cov}(X_i, X_j) = \frac{k(k-1)}{N(N-1)} - \frac{k^2}{N^2}$$

$$= -\frac{k}{N}\left(1 - \frac{k}{N}\right)\frac{1}{N-1}.$$

Substituting this in the formula for $V(Y)$, we get

$$V(Y) = \sum_{i=1}^{n} V(Y_i) + 2\sum_{i<j}\sum \mathrm{Cov}(Y_i, Y_j)$$

$$= \sum_{i=1}^{n} \frac{k}{N}\left(1 - \frac{k}{N}\right) + 2\sum_{i<j}\sum\left(-\frac{k}{N}\left(1-\frac{k}{N}\right)\frac{1}{N-1}\right)$$

$$= n\frac{k}{N}\left(1 - \frac{k}{N}\right) + 2\underbrace{\binom{n}{2}}_{\substack{\#\text{ of ways of} \\ \text{choosing } i,j \\ \text{such that} \\ 1 \le i < j \le n}} \frac{-k}{N}\left(1 - \frac{k}{N}\right)\frac{1}{N-1}$$

$$= n\frac{k}{N}\left(1 - \frac{k}{N}\right)\left(1 - \frac{n-1}{N-1}\right)$$

$$= n\frac{k}{N}\left(1 - \frac{k}{N}\right)\left(\frac{N-n}{N-1}\right).$$

LECTURE - (32)

Agenda:

(1) Conditional expectation of random variables
(2) Properties involving conditional expectations

CONDITIONAL EXPECTATION OF RANDOM VARIABLES

Let us recollect that for two random variables $X$ and $Y$, the conditional distribution of $X$ given $Y=y$, describes the probability behaviour of the random variable $X$ given the information that $Y=y$. In the same spirit, we defined the notion of the conditional expectation of $X$ given $Y=y$.

**Definition:** If $X$ and $Y$ are two continuous random variables with joint probability density $f_{X,Y}$, then the conditional expectation of $X$ given $Y=y$ is defined as

$$E(X \mid Y=y) = \int_{-\infty}^{\infty} x \underbrace{f_{X \mid Y=y}(x)}_{\text{conditional probability density of } X \text{ given } Y=y.} \, dx .$$

**Definition:** If $X$ and $Y$ are two discrete random variables with joint probability mass function $p_{X,Y}$, then the conditional expectation of $X$ given $\cancel{X} = \cancel{y}$ is defined as

$$E(X|Y=y) = \sum_{x\in\mathcal{X}} x \underbrace{p_{X|Y=y}(x)}_{\substack{\text{Conditional probability mass} \\ \text{function of } X \text{ given } Y=y}}$$

**Example:** A soft-drink machine has a random supply $Y$ at the beginning of a given day and dispenses a random amount $X$ during the day (with measurements in gallons). It has been observed that $X$ and $Y$ have joint density

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{2} & \text{if } 0 \le x \le y,\ 0 \le y \le 2, \\ 0 & \text{otherwise.} \end{cases}$$

If it is known that on a particular day 1 gallon was supplied at the beginning of the day, find the expected value of the amount of soft-drink consumed in that day.

Note that we need to find $E(X|\ Y=1)$.

Let us first find the marginal density of $Y$.

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

$$= \begin{cases} \int_{0}^{y} \frac{1}{2}\,dx & \text{if } 0 \le y \le 2, \\ 0 & \text{otherwise.} \end{cases}$$

$$= \begin{cases} \frac{y}{2} & \text{if } 0 \le y \le 2, \\ 0 & \text{otherwise.} \end{cases}$$

Hence, the conditional density of $X$ given $Y = y$ is given by

$$f_{X|Y=y}(x) = \begin{cases} \frac{f_{X,Y}(x,y)}{f_Y(y)} & \text{if } 0 \le x \le y, \\ 0 & \text{otherwise.} \end{cases}$$

$$= \begin{cases} \frac{1/2}{y/2} & \text{if } 0 \le x \le y, \\ 0 & \text{otherwise.} \end{cases}$$

$$= \begin{cases} \frac{1}{y} & \text{if } 0 \le x \le y. \\ 0 & \text{otherwise.} \end{cases}$$

Hence,

$$E(X|Y=1) = \int_{-\infty}^{\infty} x f_{X|Y=1}(x)\,dx$$

$$= \int_{0}^{1} x \cdot 1 \, dx$$

$$= \frac{1}{2}.$$

Hence, the expected amount of the amount of soft drink consumed in that day is $\frac{1}{2}$ gallon.

PROPERTIES INVOLVING CONDITIONAL EXPECTATIONS

**Result:** Let X and Y be two random variables. Then,

$$E(X) = E(E(X|Y)),$$

where on the right-hand side, the inside expectation is stated with respect to the conditional distribution of X given Y, and the outside expectation is stated with respect to the distribution of Y.

CONFUSED ??

Sometimes, the way the distribution of the random variable $X$ is specified, it is not possible to directly evaluate its expectation, and the previous identity comes in very handy.

HERE IS AN ALTERNATE WAY OF EVALUATING $E(X)$ BASED ON THE PREVIOUS IDENTITY.

① For every $y$, find $E(X|Y=y)$.

② Define the function $h$ as $h(y) \triangleq E(X|Y=y)$.

③ $E(X) = E(h(Y))$.

Hence, the way to interpret the previous relationship is as follows.

$$E(X) = E(h(Y)), \quad \text{where } h(y) = E(X|Y=y).$$

<u>Example:</u> A quality control plan for an assembly line involves sampling $n$ finished items per day and counting $X$, the number of defective items. If $p$ denotes the probability that an item is defective, then given $p$, $X$ has a binomial distribution with parameters $n$ and $p$. However, it is observed that $p$ is a random quantity,

and has a uniform distribution on the interval $[0, \frac{1}{4}]$.

(a) Find $E(X)$.

Note that given $p = p_0$, X has a binomial distribution with parameters $n$ and $p_0$. Hence,

$$E(X \mid p = p_0) = np_0.$$

Hence, $E(X) = E(E(X \mid p))$
$= E(h(p))$, where $h(p_0) = E(X \mid p = p_0)$.

Hence,
$$\begin{aligned} E(X) &= E(np) \\ &= n\,E(p) \\ &= n\,\frac{(0 + \frac{1}{4})}{2} \end{aligned}$$

($\because p$ is uniform on the interval $[0, \frac{1}{4}]$)

$$= \frac{n}{8}.$$

(b) Find $SD(X)$.

**Result:** Let X and Y be two random variables. Then,
$$V(X) = E(V(X \mid Y)) + V(E(X \mid Y)).$$

HERE IS AN ALGORITHM BASED ON THIS IDENTITY.

TO EVALUATE $V(X)$.

1. For every $y$, find $E(X \mid Y=y)$ and $V(X \mid Y=y)$

2. Define the function $h$ as $h(y) \triangleq E(X \mid Y=y)$.
3. Define the function $\tilde{h}$ as $\tilde{h}(y) \triangleq V(X \mid Y=y)$.

4. $V(X) = V(h(Y)) + E(\tilde{h}(Y))$.

(Note that for any function $g$, $E(g(X) \mid Y=y)$ is $\int_{-\infty}^{\infty} g(x) f_{X \mid Y=y}(x)\, dx$ if $X$ is continuous, and

$\sum_{x \in \mathcal{X}} g(x)\, p_{X \mid Y=y}(x)$ if $X$ is discrete.)

Hence, the way to interpret the previous relationship is as follows.

$V(X) = V(h(Y)) + E(\tilde{h}(Y))$, where $h(y) = E(X \mid Y=y)$ and $\tilde{h}(y) = V(X \mid Y=y)$.

Returning to the example,

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

$$= (n^2-n)V(p) + nE(p) - n(E(p))^2$$

$$(\because \quad E(p^2) = V(p) + (E(p))^2)$$

$$= (n^2-n) \cdot \frac{\left(\frac{1}{4}\right)^2}{12} + \frac{n}{8} - \frac{n}{64}$$

$$= \frac{n^2}{192} + \frac{5n}{48}$$

Hence, $SD(x) = \sqrt{V(x)} = \sqrt{\frac{n^2}{192} + \frac{5n}{48}}$ .

LECTURE - 33

Agenda:

(1) The multinomial distribution
(2) Moment generating functions and sums of independen random variables.

THE MULTINOMIAL DISTRIBUTION

Until now we have developed methods to study the joint probability behaviour of two random variables. But often we have experiments where we are interested in the joint probability behaviour of not just two, but several random variables. Let us study one example of such an experiment.

(1) Suppose the experiment consists of $n$ independent trials
(2) Suppose each trial has $k$ possible outcomes, and the probability of outcome $i$ is $p_i$, for every $i=1,2,\ldots$

(If $k=2$, then we have a binomial experiment)

Let $X_i$ = # of outcomes which are $i$, for every $i=1,2,\ldots$

The random variables $X_1, X_2, ---, X_k$ completely describ the randomness in the experiment, and hence it is very natural to ~~[illegible]~~ study the joint probability behaviour of these random variables. As in the case of two random variables, the joint probability behaviour of the discrete random variables $X_1, X_2, ---, X$ is described by their joint probability mass function

$$p_{X_1, X_2, ---, X_k}(x_1, x_2, ---, x_k) \triangleq P(X_1 = x_1, X_2 = x_2, ---, X_k = x$$

for all $x_i \geq 0$ such that $\sum_{i=1}^{k} x_i = n$.

Note that,

$$P(X_1 = x_1, X_2 = x_2, ---, X_k = x_k)$$

$= P(x_1$ outcomes which are $1$, $x_2$ outcomes which are $2$, $---$, $x_k$ outcomes which are $k)$

$$= \underbrace{\frac{n!}{x_1!\, x_2! --- x_k!}}_{\substack{\#\text{ of ways of partitioning} \\ n \text{ outcomes into } k \text{ groups} \\ \text{of size } x_1, x_2, ---, x_k \text{ respectively.}}} p_1^{x_1} p_2^{x_2} --- p_k^{x_k}$$

The vector $(X_1, X_2, ---, X_k)$ is said to follow the multinomial distribution with parameters n and $p_1, p_2, ---, p_k$.

**RESULT:** If $(X_1, X_2, ---, X_k)$ follows a multinomial distribution with parameters n and $p_1, p_2, ---, p_k$, then

$$E(X_i) = np_i, \quad V(X_i) = np_i(1-p_i), \text{ and}$$

$$Cov(X_i, X_j) = -np_i p_j, \text{ for } 1 \le i \ne j \le k.$$

**Example:** The National Fire Incident Reporting Service says that among residential fires, approximately 74% are in 1 or 2 family homes, 20% are in multifamily homes, and 6% in other dwellings. If five fires are reported independently in a day, find the probability that 2 are in 1 or 2 famil homes, 2 are in multifamily homes and one is in other dwellings.

Clearly, this is a multinomial experiment.

Let $X_1$ = # of 1 or 2 family home fires

$X_2$ = # of multifamily home fires

$X_3$ = # of other fires

Then, $(X_1, X_2, X_3)$ is multinomial with parameters 5 and 0.74, 0.20, 0.06.

Hence,

$$P(X_1 = 2,\ X_2 = 2,\ X_3 = 1) = \frac{5!}{2!\,2!\,1!}(0.74)^2(0.2)^2(0.0$$

$$=$$

Suppose we are interested in $V(X_1 + X_2)$.

$$V(X_1 + X_2) = V(X_1) + V(X_2) + 2\,\mathrm{Cov}(X_1, X_2)$$

$$= np_1(1-p_1) + np_2(1-p_2) + 2(-np_1p_2)$$

$$= 5 \times 0.74 \times 0.26 + 5 \times 0.2 \times 0.8 + 2(-5 \times 0.76 \times 0.$$

$$=$$

MOMENT GENERATING FUNCTIONS AND SUMS OF INDEPENDENT RANDOM VARIABLES

By now we are familiar with experiments involving two random variables, and know that it is often of interest to find the distribution of functions of random variables.

Often, moment generating functions are useful for find the distribution of sums of independent random variables. Here are two examples.

<u>Example:</u> Let $X_1$ and $X_2$ be independent exponen random variables, with mean $\theta$. Find the distrib of $Y = X_1 + X_2$.

Let us find the moment generating function of $Y$.

$$
\begin{aligned}
M_Y(t) &= E[e^{tY}] \\
&= E[e^{t(X_1+X_2)}] \\
&= E[e^{tX_1} e^{tX_2}] \\
&= E[e^{tX_1}]\, E[e^{tX_2}]
\end{aligned}
$$

($\because$ By independence of $X_1$ and $X_2$)

$$= \frac{1}{1-\theta t} \cdot \frac{1}{1-\theta t}$$

($\because$ By the formula for the moment generatin function of an exponential random variab

$$= \frac{1}{(1-\theta t)^2}.$$

But note that moment generating functions characteriz the distribution of a random variable, and by observing that $\frac{1}{(1-\theta t)^2}$ is the moment generating function of a Gamma random variable with parameters $\alpha = 2$ and $\beta = \theta$, we conclude that

$Y$ follows a Gamma distribution with parameters $\alpha = 2$ and $\beta = \theta$.

Example 2: Let $X_1$ and $X_2$ be independent Poisson random variables. Let $X_1$ have mean $\lambda_1$ and $X_2$ have mean $\lambda_2$. Find the distribution of $Y = X_1 + X_2$.

Let us get the moment generating function of $Y$.

$$M_Y(t) = E[e^{tY}]$$

$$= E[e^{t(X_1+X_2)}]$$

$$= E[e^{tX_1} e^{tX_2}]$$

$$= E[e^{tX_1}] E[e^{tX_2}]$$

$(\because$ By independence of $X_1$ and $X_2)$

$$= e^{\lambda_1 (e^t - 1)} \cdot e^{\lambda_2 (e^t - 1)}$$

$\left( \because \right.$ By the formula for the moment generating function of a Poisson random variable. $\left. \right)$

$$= e^{(\lambda_1 + \lambda_2)(e^t - 1)}$$

Again, note that moment generating functions characterize the distribution of a random variable. Also, $e^{\lambda(e^t - 1)}$ is the moment generating function of a Poisson random variable with mean $\lambda$. This clearly implies $Y$ follows a Poisson distribut. with mean $\lambda_1 + \lambda_2$.

# LECTURE - 34

## Agenda:

1. Functions of discrete random variables
2. Functions of continuous random variables (Method of distribution functions)

## FUNCTIONS OF DISCRETE RANDOM VARIABLES

We have seen that in many experiments, we have one ~~~~ or more ~~~~ random variables of interest and we often wish to study their joint behaviour. Along the same lines, we are often interested in the probability distribution of a function of these random variables. Let us consider the discrete random variable case first.

Let $X$ be a discrete random variable with p.m.f. $p_X(x) = P(X=x)$ for every $x \in \mathcal{X} = \text{Range}(X)$. Suppose we are interested in the probability distribution of $Y = f(X)$.

General algorithm:

1. Identify the range of $Y$.
2. For every $y \in \mathcal{Y} = \text{Range}(Y)$, find $p_Y(y) = P(Y=y)$

by expressing the event $\{Y = y\}$ in terms of the random variable $X$.

<u>Example:</u> A quality control manager samples from a large lot of items, testing each item until $r$ defectives have been found. Find the distribution of $Y$, the number of items that are tested to obtain $r$ defectives.

Assuming that the lot is large, this experiment is a negative binomial experiment and if

$$X = \# \text{ of non-defectives found before } r \text{ defectives are found,}$$

then

$$P(X = x) = \binom{x+r-1}{r-1} p^r q^x, \quad x = 0, 1, 2, \ldots$$

where $p$ is the probability of finding a defective item.

But we are interested in $Y$ which is the total number of items that are tested to obtain $r$ defectives.

Note that $Y = X + r$.

$\mathcal{Y} = \text{Range}(Y) = \{r, r+1, r+2, \ldots\}$.

For every $y \in \{r, r+1, r+2, \ldots\ldots\}$

$$P(Y = y) = P(X + r = y)$$

$$= P(X = y - r)$$

$$= \binom{y - r + r - 1}{r - 1} p^r q^{y-r}$$

$$= \binom{y-1}{r-1} p^r q^{y-r}.$$

FUNCTIONS OF CONTINUOUS RANDOM VARIABLES (METHOD OF DISTRIBUTION FUNCTIONS)

We have to more careful when dealing with continuous random variables, as they are expressed in terms of density functions.

Let $X$ be a continuous random variable with p.d.f $f_X$. Suppose we are interested in the probability density function of $Y = f(X)$.

General algorithms:

1. Express the event $\{Y \le y\}$ in terms of $X$.
2. Using this express $F_Y(y) = P(Y \le y)$ as

an expression in terms of $F_X$.

(3) Differentiate to get $f_Y(y) = \frac{d}{dy} F_Y(y)$.

<u>Example:</u> Suppose $X$ is a continuous random variable with probability density $f_X$, and we want to find the probability density function for $Y = X^2$.

(1) For every $y \geq 0$,

$$\begin{aligned} P(Y \leq y) &= P(X^2 \leq y) \\ &= P(-\sqrt{y} \leq X \leq \sqrt{y}) \\ &= F_X(\sqrt{y}) - F_X(-\sqrt{y}) \end{aligned}$$

For every $y < 0$,

$$P(Y \leq y) = 0.$$

(2) Hence,

$$F_Y(y) = \begin{cases} F_X(\sqrt{y}) - F_X(-\sqrt{y}) & \text{if } y \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

(3) Differentiating with respect to $y$, we get that,

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \begin{cases} \frac{d}{dy} F_X(\sqrt{y}) - \frac{d}{dy} F_X(-\sqrt{y}) & \text{if } y \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

Hence,

$$f_Y(y) = \begin{cases} \frac{1}{\sqrt{2y}} \left( f_X(\sqrt{y}) + f_X(-\sqrt{y}) \right) & \text{if } y \geq 0, \\ 0 & \text{otherwise}. \end{cases}$$

<u>Example:</u> The proportion of time $X$ that a coffee machine is in use during a typical 40-hour workweek is a random variable whose probability density function is given by

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

The actual number of hours out of a 40-hour week that the coffee machine is not in use is given by

$$Y = 40(1 - X).$$

Find the probability density function of $Y$.

① $\{Y \leq y\} = \{40(1-X) \leq y\}$
$= \left\{ X \geq 1 - \frac{y}{40} \right\}$

(2) Hence,

$$F_Y(y) = P(Y \le y)$$

$$= P\left(X \ge 1 - \frac{y}{40}\right)$$

$$= \int_{1-\frac{y}{40}}^{1} f_X(x)\, dx$$

$$= \begin{cases} 0 & y < 0, \\ 1 - \left(1 - \frac{y}{40}\right)^3 & 0 \le y \le 40, \\ 1 & y > 40. \end{cases}$$

(3) Differentiating with respect to $y$, we get that,

$$f_Y(y) = \frac{d}{dy} F_Y(y)$$

$$= \begin{cases} \frac{d}{dy}\left\{-\left(1 - \frac{y}{40}\right)^3\right\} & \text{if } 0 \le y \le 40, \\ 0 & \text{otherwise.} \end{cases}$$

$$= \begin{cases} \frac{3}{40}\left(1 - \frac{y}{40}\right)^2 & \text{if } 0 \le y \le 40, \\ 0 & \text{otherwise.} \end{cases}$$

# LECTURE – 35

**Agenda:**

(1) Method of transformations
(2) Method of conditioning

## METHOD OF TRANSFORMATIONS

~~Let us recollect the method of transformations from the previous lecture.~~

**Given:** A continuous random variable $X$ with density $f_X$ given by

$$f_X(x) = \begin{cases} > 0 & \text{if } x \in (a,b), \\ 0 & \text{otherwise.} \end{cases}$$

**Task:** Find the probability density function of $Y = g(X)$, where $g$ is a one-to-one differentiable function on the range of $f_X$.

**General algorithm:**

(1) Identify $(\alpha, \beta)$, where $(\alpha,\beta)$ is the image of $(a,b)$ under $g$.
(2) Construct the inverse function $h$, by solving for $X$ in $Y = g(X)$, and expressing it as $X = h(Y)$.

(3) Verify that $h'(y) = \frac{d}{dy} h(y) \neq 0$ for each $y \in (\alpha, \beta)$.

(4) Compute the probability density function of $Y$ as

$$f_Y(y) = \begin{cases} f_X(h(y)) \underbrace{|h'(y)|}_{\frac{d}{dy}h(y)} & \text{if } y \in (\alpha, \beta), \\ 0 & \text{otherwise.} \end{cases}$$

<u>Example:</u> Let $X$ be a random variable with probability density function given by

$$f_X(x) = \begin{cases} (3/2)\, x^2, & \text{if } -1 \leq x \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

(a) Find the density function of $Y = 3X$.

(1) $g(x) = 3x$, $(a, b) = (-1, 1)$.
The image of $(a, b)$ under $g$ is $(\alpha, \beta) = (-3, 3)$.

(2) $Y = 3X \Rightarrow X = \frac{Y}{3}$.

Hence $h(y) = \frac{y}{3}$.

③ $h'(y) = \frac{1}{3} \neq 0$ for every $y \in (-3,3)$.

④ $f_Y(y) = \begin{cases} f_X(h(y))\,|h'(y)| & \text{if } y \in (\alpha, \beta), \\ 0 & \text{otherwise.} \end{cases}$

$= \begin{cases} f_X(\frac{y}{3}) \left|\frac{1}{3}\right| & \text{if } y \in (-3,3), \\ 0 & \text{otherwise.} \end{cases}$

$= \begin{cases} \frac{y^2}{18} & \text{if } y \in (-3,3), \\ 0 & \text{otherwise.} \end{cases}$

(b) Find the probability density function of $Y = 3 - X$

① $g(x) = 3 - x$, $(a,b) = (-1,1)$.
The image of $(-1,1)$ under $g$ is $(\alpha, \beta) = (2,4)$.

② $Y = 3 - X \Rightarrow X = 3 - Y$.

Hence, $h(y) = 3 - y$.

③ $h'(y) = -1 \neq 0$ for every $y \in (2,4)$

④ $f_Y(y) = \begin{cases} f_X(h(y))\,|h'(y)| & \text{if } y \in (\alpha, \beta), \\ 0 & \text{otherwise.} \end{cases}$

$= \begin{cases} f_X(3-y)\,|-1| & \text{if } y \in (2,4), \\ 0 & \text{otherwise} \end{cases}$

$$= \begin{cases} \frac{3}{2}(3-y)^2 & \text{if } y \in (2,4), \\ 0 & \text{otherwise.} \end{cases}$$

(c) Find the probability density function of $Y = X^3$.

① $g(x) = x^3$, $(a, b) = (-1, 1)$.
The image of $(-1, 1)$ under $g$ is $(\alpha, \beta) = (-1, 1)$.

② $Y = X^3 \Rightarrow X = \sqrt[3]{Y}$.

Hence $h(y) = \sqrt[3]{y}$.

③ $h'(y) = \frac{1}{3} \frac{1}{(\sqrt[3]{y})^2} \neq 0$ for every $y \in (-1, 1)$.

④ $$f_Y(y) = \begin{cases} f_X(h(y)) \, |h'(y)| & \text{if } y \in (\alpha, \beta), \\ 0 & \text{otherwise.} \end{cases}$$

$$= \begin{cases} f_X(\sqrt[3]{y}) \left| \frac{1}{3(\sqrt[3]{y})^2} \right| & \text{if } y \in (-1, 1), \\ 0 & \text{otherwise.} \end{cases}$$

$$= \begin{cases} \frac{1}{2} & \text{if } y \in (-1, 1), \\ 0 & \text{otherwise.} \end{cases}$$

METHOD OF CONDITIONING

Conditional probability density functions can also be used to find probability density functions of specific functions of random variables. The basis of this method is the following observation.

Suppose $X_1$ and $X_2$ are continuous random variables with joint probability density function $f_{X_1,X_2}$. Suppose we want to find the density function of $Y = g(X_1, X_2)$. Note that if $f_{Y,X_2}$ is the probability density function of $Y$ and $X_2$,

$$
\begin{aligned}
f_Y(y) &= \int_{-\infty}^{\infty} f_Y(y, x_2)\, dx_2 \\
&= \int_{-\infty}^{\infty} \underbrace{f_{Y|X_2=x_2}(y)}_{\downarrow}\, f_{X_2}(x_2)\, dx_2
\end{aligned}
$$

Find this by the method of transformations.

General algorithm:

① Fix $X_2 = x_2$. Find the conditional probability density function of $X_1$ given $X_2 = x_2$.

② Note that $Y = g(X_1, \underset{\text{fixed}}{x_2})$ is a transformation of $X_1$. Assuming it is one-to-one and differentiable,

Find the conditional probability density function of $Y$ given $X_2 = x_2$.

③ Compute the probability density functions of $Y$ as

$$f_Y(y) = \int_{-\infty}^{\infty} f_{Y|X=x_2}(y)\, f_{X_2}(x_2)\, dx_2 .$$

Example: Let $X_1$ and $X_2$ be independent random variables each following the exponential distribution with mean 1. Find the density function of $Y = \frac{X_1}{X_2}$.

① Fix $X_2 = x_2 > 0$. Since $X_1$ is independent of $X_2$, the conditional probability density function of $X_1$ given $X_2 = x_2$, is the same as the marginal

probability density function of $X_1$, ~~given by~~
Hence, the conditional probability density function is ~~$f_{X_1|X_2=x_2}$~~ given by

$$f_{X_1|X_2=x_2}(x_1) = \begin{cases} e^{-x_1} & \text{if } x_1 > 0, \\ 0 & \text{otherwise.} \end{cases}$$

② Given $X_2 = x_2$, $Y = \dfrac{X_1}{x_2}$. Let us apply the method of transformations.

① $g(x_1) = \dfrac{x_1}{x_2}$, $(a, b) = (0, \infty)$

The image of $(a, b)$ under $g$ is $(\alpha, \beta) = (0, \infty)$.

② $Y = \dfrac{X_1}{x_2} \Rightarrow X_1 = x_2 y$.

Hence, $h(y) = x_2 y$

③ $h'(y) = x_2 \neq 0$ for every $y \in (0, \infty)$.

④ $$f_{Y|X_2=x_2}(y) = \begin{cases} f_{X_1|X_2=x_2}(h(y))\,|h'(y)| & \text{if } y \in (\alpha, \beta), \\ 0 & \text{otherwise.} \end{cases}$$

$$= \begin{cases} f_{X_1|X_2=x_2}(x_2 y)\,|x_2| & \text{if } y > 0, \\ 0 & \text{otherwise.} \end{cases}$$

$$= \begin{cases} x_2 e^{-x_2 y} & \text{if } y > 0, \\ 0 & \text{otherwise.} \end{cases}$$

(3)

$$f_{Y}(y) = \int_{-\infty}^{\infty} f_{Y|X_2=x_2}(y) f_{X_2}(x_2)\, dx_2$$

$$= \begin{cases} \int_0^{\infty} x_2 e^{-x_2 y} e^{-x_2}\, dx_2 & \text{if } y > 0, \\ 0 & \text{otherwise.} \end{cases}$$

$$= \begin{cases} \int_0^{\infty} x_2 e^{-x_2 (y+1)}\, dx_2 & \text{if } y > 0, \\ 0 & \text{otherwise.} \end{cases}$$

$$= \begin{cases} \dfrac{1}{(y+1)^2} & \text{if } y > 0, \\ 0 & \text{otherwise.} \end{cases}$$

LECTURE - 26

Agenda:

1. Method of moment generating functions
2. Examples

METHOD OF MOMENT GENERATING FUNCTIONS

· Moment generating functions can be used to identify the distribution of a sum of independent random variables. The main result that forms the basis of this approach says that if $X$ and $Y$ are two random variables such that their moment generating functions are the same (on an interval around zero), then $X$ and $Y$ have the same probability distribution function. We use this property to establish various identities that are useful in applications. We first recall a simple identity about moment generating functions.

Result: If $X_1, X_2, ---, X_n$ are independent random variables, then

$$M_{X_1+X_2+---+X_n}(t) = M_{X_1}(t) M_{X_2}(t) --- M_{X_n}(t)$$

The proof of this identity is straightforward, but will be omitted due to time constraints.

EXAMPLES

<u>Example 1:</u> Let $X_1, X_2, ---, X_n$ be independent random variables which have a gamma distribution with parameters $\alpha$ and $\beta$. Find the distribution of $X_1 + X_2 + --- + X_n$.

$$M_{X_1+X_2+---+X_n}(t) = E\left[e^{t(X_1+X_2+---+X_n)}\right]$$

$$= E\left[e^{tX_1} e^{tX_2} --- e^{tX_n}\right]$$

$$= E\left[e^{tX_1}\right] E\left[e^{tX_2}\right] --- E\left[e^{tX_n}\right]$$

$$= \left(\frac{1}{1-\beta t}\right)^{\alpha} \left(\frac{1}{1-\beta t}\right)^{\alpha} --- \left(\frac{1}{1-\beta t}\right)^{\alpha}$$

$$= \left(\frac{1}{1-\beta t}\right)^{n\alpha}.$$

But this is precisely the moment generating function of a gamma random variable, with parameters $n\alpha$ and $\beta$. Hence, $X_1 + X_2 + --- + X_n$ has a gamma distribution with parameters $n\alpha$ and $\beta$.

Example 2: Let $X_1, X_2, \ldots, X_n$ be independent normal random variables, where $E(X_i) = \mu_i$ and $V(X_i) = \sigma_i^2$. Find the distribution of $X_1 + X_2 + \ldots + X_n$

Note that if $X \sim N(\mu, \sigma^2)$, then

$$M_X(t) = e^{\mu t + \frac{\sigma^2 t^2}{2}}.$$

Hence,

$$M_{X_1+X_2+\ldots+X_n}(t) = M_{X_1}(t)\, M_{X_2}(t) \ldots M_{X_n}(t)$$

$$= e^{\mu_1 t + \frac{\sigma_1^2 t^2}{2}}\, e^{\mu_2 t + \frac{\sigma_2^2 t^2}{2}} \ldots e^{\mu_n t + \frac{\sigma_n^2 t^2}{2}}$$

$$= e^{\left(\sum_{i=1}^{n} \mu_i\right) t + \frac{\left(\sum_{i=1}^{n} \sigma_i^2\right) t^2}{2}}.$$

But this is precisely the moment generating function of a normal random variable with parameters $\mu = \sum_{i=1}^{n} \mu_i$ and $\sigma^2 = \sum_{i=1}^{n} \sigma_i^2$. Hence, $X_1 + X_2 + \ldots + X_n$ has a normal distribution with parameters $\sum_{i=1}^{n} \mu_i$ and $\sum_{i=1}^{n} \sigma_i^2$.

Example 3: Let $Z$ be a standard normal random variable. Find the distribution of $Z^2$.

$$M_{Z^2}(t) = E[e^{tZ^2}]$$

$$= \int_{-\infty}^{\infty} e^{tz^2} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \, dz$$

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

$$= \frac{1}{\sqrt{1-2t}} \underbrace{\int_{-\infty}^{\infty} \sqrt{\frac{1-2t}{2\pi}} \, e^{-\frac{1}{2}z^2(1-2t)} \, dz}_{}$$

density of a normal random variable with parameters $\mu = 0$, $\sigma^2 = \frac{1}{1-2t}$

(holds only if $t < \frac{1}{2}$)

$$= \frac{1}{\sqrt{1-2t}}.$$

But this is precisely the moment generating function of a gamma random variable with parameters $\alpha = \frac{1}{2}$ and $\beta = 2$.

If $Z_1, Z_2, \ldots, Z_n$ are independent standard normal random variables, then it follows that,

$$M_{\sum_{i=1}^{n} Z_i^2}(t) = M_{Z_1^2}(t) M_{Z_2^2}(t) \cdots M_{Z_n^2}(t),$$

$$= \frac{1}{\sqrt{1-2t}} \cdot \frac{1}{\sqrt{1-2t}} \cdots \frac{1}{\sqrt{1-2t}}$$

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

But this is precisely the moment generating function of a gamma random variable with parameters $\alpha = \frac{n}{2}$ and $\beta = 2$. These random variables have a special name.

**Definition:** A gamma distribution with parameters $\alpha = \frac{n}{2}$ and $\beta = 2$ is known as a chi-square distribution with $n$ degrees of freedom.

**REMARK:** The chi-square distribution appears frequently in applications. An example is provided below.

Let $X_1, X_2, \ldots, X_n$ be a set of independent random variables having a normal distribution with parameters $\mu$ and $\sigma^2$.

Suppose we observe $X_1, X_2, ---, X_n$ and want to estimate $\mu$ and $\sigma^2$.

The sample mean $\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$ is taken to be the estimator of $\mu$, and the sample variance $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2$ is taken to be the estimator of $\sigma^2$. We know that $\overline{X}$ has a normal distribution with $E(\overline{X}) = \mu$ and $V(\overline{X}) = \frac{\sigma^2}{n}$.

Using the method of moment generating functions, it can be proved that $\frac{(n-1)S^2}{\sigma^2}$ has a chi-square distribution with $(n-1)$ degrees of freedom, and $\overline{X}$ and $S^2$ are independent random variables. This example shows the strength of the moment generating functions approach