MATH 323
Notes

Emily Wang

Contents

Chapter 1

Basic probability

1.1 Foundation

1.1.1 Set identities

Let A, B, C be sets, all subsets of U.

1. $A \cup U = A$ and $A \cup \emptyset = A$: Identity laws
2. $A \cup U = U$ and $A \cap \emptyset = \emptyset$: Domination laws
3. $A \cup A$ and $A \cap A = A$: Idempotent laws
4. $\overline{\overline{A}} = A$: Double complement
5. $A \cup B = B \cup A$, $A \cap B = B \cap A$: Commutative Laws
6. $A \cup (B \cup C) = (A \cup B) \cup C$: Associative laws
7. $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

   $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$: Distributive laws
8. $\overline{A \cup B} = \overline{A} \cap \overline{B}$

   $\overline{A \cap B} = \overline{A} \cup \overline{B}$: De Morgan's laws

   An intuitive explanation: We can say that A is all the days that it rains and B is all the days that it snows. Then $\overline{A \cup B}$ means all days where it does not rain or snow. This is the same as saying that we are looking at all days where it does not rain *and* all the days it does not snow ($\overline{A} \cap \overline{B}$).

   Similarly, we can say that $\overline{A \cap B}$ means all days where it does not snow and rain simultaneously. Then this is the same as saying all days where it either does not snow, or all days where it does not rain. If it never snows, then we'll never get a day where it rains or snows simultaneously. If it never rains, then we'll never get a day where it rains or snows simultaneously. Hence this is the same as saying that $\overline{A} \cup \overline{B}$.
9. $A \cup (A \cap B) = A$

   $A \cap (A \cup B) = A$: Absorption laws
10. $A \cup \overline{A} = U$

    $A \cap \overline{A} = \emptyset$: Complement laws

Definition 1.1.1: Random experiment

A process for making an observation of observations whose outocme cannot be predicted with certainty.

Definition 1.1.2: Event

Each repetition of a random experiment is known as a **trial**. The outcome is known as an **event**. Events can be represented as a set of outcomes.

A **simple event** consists of the outcome of one trial.

A **compound event** consists of two or more composed simple events.

Definition 1.1.3: Sample space

The set of all possible outcomes of an experiment, denoted S.

Example 1.1.1

The **sample space** of a die toss is $\{1, 2, 3, 4, 5, 6\}$.

The **sample space** of picking 2 workers in any order from 3 (where W_i denotes the worker) is: $\{(W_1, W_2), (W_2, W_3), (W_1, W_3)\}$.

The **sample space** of picking 2 workers in a specific order from 3 (where W_i denotes the worker) is: $\{(W_1, W_2), (W_1, W_3), (W_2, W_1), (W_2, W_3), (W_3, W_1)(W_3, W_3)\}$.

1.2 Kolmogorov's Axioms

Definition 1.2.1: Kolmogorov's Axioms

Consider an experiment with sample space S. To every event $A \subseteq S$, we associate a number $P(A)$ called the probabiltiy of A, such that following axioms hold:

1. $P(A) \geq 0$
2. $P(S) = 1$
3. If $E_1, E_2, ...$ are events in S such that $E_1 \cap E_j = \emptyset$ for $i \not= j$ (disjoint, mutually exclusive events), then:

$$P(E_1 \cup E_2 \cup E_3 ...) = \sum_{i=1}^{\infty} P(E_i)$$

Note:-

Axiom 3 can be stated in terms of a finite union of events, as

$$P(E_1 \cup E_2 \cup E_3 ... \cup E_n = \sum_{i=1}^{n} P(E_i)$$

.
Axiom 3 is necessary to prove the five theorems - we can express the probabilities of disjoint events as a sum of their individual probabilities.

Theorem 1.2.1 Five Theorems

1. For any event A, $P(A^c) = 1 - P(A)$.
2. $P(\emptyset) = 0$.

3. $P(A \cap B^c) = P(A) - P(A \cap B)$.
4. If $A \subseteq B$, then $P(A) \leq P(B)$.
5. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Note:-

For proofs (of not just the five theorems), it is useful to rewrite events as disjoint events so we can use axiom 3.

Theorem 4:

$$B = A \cup (A^c \cap B)$$
$$P(B) = P(A) + P(A^c \cap B)$$
$$P(A^c \cap B) \geq 0$$
$$\implies P(B) \geq P(A)$$

☺

Theorem 5:

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

☺

Theorem 1.2.2 Probability of an event

Let S be a finite sample space with N equally likely events. Let E be an event in S. Then:

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

Where n is the number of outcomes in E, or the number of possible ways E can occurs, and N is the number of outcomes in S.

Proof: Write E as union of simple events:

$$E = \bigcup_{i=1}^{n} E_i$$

Since the simply events are disjoint:

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

Since all events E_i are equally likely:
$\sum_{i=1}^{n} P(E_i) = \sum_{i=1}^{n} (\frac{1}{N}) = \frac{n}{N}$ ☺

1.3 Counting rules

1. **Multiplication rule:** If two sets A and B have n_1 distinct objects and n_2 distinct objects respectively, then the number of ways to choose an object from A and an object from B is $n_1 \times n_2$.

   Extension to k sets: The number of ways to form a set by choosing an object from each of the sets is $n_1 \times n_2 \times ... \times n_k$.

2. **Arranging n distinct objects:** The number of ways to arrange n distinct objects (by selecting one without replacement) is $n!$.

3. **Permutations:** Arranging n distinct objects chosen r at a time without replacement, where order is important, is known as permutation.

   Definition 1.3.1: Permutation

   Number of ways to arrange n distinct objects chosen r at a time without replacement, where order matters.

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

4. **Combinations:** Arranging n distinct objects chosen r at a time without replacement, where order is unimportant, is known as combination.

   Definition 1.3.2: Combination

   Number of ways to arrange n distinct objects chosen r at a time without replacement, where order does not matter.

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

   Definition 1.3.3: Multinomial coefficient

   The number of ways of partitioning n distinct objects into k distinct subsets of sizes:

   $$n_1, n_2, ...n_k$$

   where

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

   is

   $$N = \binom{n}{n_1, n_2, \ldots, n_k} = \frac{n!}{k_1!k_2!\cdots k_r!}$$

Question 1: The Birthday Problem

Suppose you have n friends. You wish to find the probability that at least two of your friends have the same birthday.

Solution: Sample space is given by:

$$S = \{(Jan1, Jan1, ...Jan1), (Jan2, Jan2, ...Jan3), ...(Dec31, Dec31, ...Dec31)\}$$

We create a tuple representing a possible combination of everyone's birthdays.

Represent the event that at least two friends have the same birthday as E. Then E^c is the event that no two friends have the same birthday.

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

where n' = number of outcomes in E^c.

Number of simple events is given by 365^n, since each person's birthday is distinct and unaffected by another person's birthday. So $N = 365^n$.

Number of outcomes favorable to E^c includes counting the number of ways to arrange 365 distinct birthdays, chosen n at a time. The order does matter - if person A is born on $Jan1$, then the second person can only be born on $Jan2, Jan3, ...Dec31$. So the next person has 1 less possibilities of birthdays from the previous person.

$$n' = \frac{365!}{(365-n)!}$$
$$n' = 365 \times 364 \times 363 \times (365 - n + 1)$$

$$P(E) = 1 - \frac{n'}{N}$$
$$P(E) = 1 - \frac{365 \times 364 \times 363 \times (365 - n + 1}{365^n}$$
$$P(E) = 1 - \frac{365!}{(365^n)(365-n)!}$$

Question 2: Splitting sample space

A student is instructed to choose m questions from a test with M questions. What is the probability that the student chooses x from the first a and $m - x$ from the last $M - a$?

Solution: First find the number of outcomes in the sample space (N). Order does not matter, the student can choose m questions in any order they wish. Therefore:

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

Choosing from section a will yield a different subset of events than choosing from section $M - a$. In both sections, we do not care about the order. Since these are always disjoint, we can apply the multiplication rule, such that if n_1 is the number of ways of choosing x from the first a, and n_2 is the number of ways of choosing $m - x$ from $M - a$, then the number of ways of picking a certain subset from both is $n_1 \times n_2$.

Choosing x from the first a:

$$n_1 = \binom{a}{x}$$

Choosing $M - x$ from the last $M - a$:

$$n_2 = \binom{M-a}{m-x}$$

Therefore:

$$P(E) = \frac{\binom{a}{x}\binom{M-a}{m-x}}{\binom{M}{m}}$$

1.4 Conditional probability

Definition 1.4.1: Conditional probability

Let A and B be two events, such that $P(A) \neq 0$, then:

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

That is, the probability of A occurring given that B has occurred. In the presence of updated knowledge, probability can change.

Corollary 1.4.1

$$P(B|A) \geq 0$$

Since $P(A \cap B) \geq 0$, whereas $P(A) > 0$.

Corollary 1.4.2

$$P(S|A) = 1$$

Since $P(S|A) = \frac{P(A \cap S)}{P(A)} = \frac{P(A)}{P(A)}$

Corollary 1.4.3

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

where $B_i \cap B_j = \emptyset$ for $i \neq j$

With the definition of conditional probability, we can determine the intersection of several events as multiple chained conditional probabilities. Given:

$$P(A_1 \cap A_2 \cap A_3)$$

We can read this as: the probability that A_3 occurs given that A_1 and A_2 have occurred, times the probability that A_1 and A_2 have occurred. Then the probability of A_1 and A_2 both occurring is the probability that A_2 occurred given A_1 occurred, times the probability that A_1 occurred:

$$P(A_1 \cap A_3 \cap A_3) = P(A_3|A_1 \cap A_2) \times P(A_1 \cap A_2)$$
$$P(A_1 \cap A_3 \cap A_3) = P(A_3|A_1 \cap A_2) \times P(A_2|A_1)P(A_1)$$

Theorem 1.4.1 Multiplication rule for conditional probability

$$P(\cap_{i=1}^{n} A_i) = P(A_n|A_1 \cap ... \cap A_{n-1}) \times P(A_1 \cap ... \cap A_{n-1}) \times ... \times P(A_2|A_1)P(A_1)$$

1.5 The law of total probability

Definition 1.5.1: Partition of S

For some positive integer k, let the sets $B_1, B_2, ...B_k$ be such that:

1. $S = B_1 \cup B_2 \cup ... \cup B_k$.
2. $B_i \cap B_j = \emptyset$ for $i \neq j$

Then the collection of sets $B_1, B_2, ...B_k$ is said to be a **partition** of S.

Theorem 1.5.1 Law of total probability

If $A \subseteq S$ and $B_1, B_2, ...B_k$ is a partition of S, A can be expressed as:

$$A = (A \cap B_1) \cup (A \cap B_2) \cup ... \cup (A \cap B_k)$$

where each $(A \cap B_i)$ and $(A \cap B_j)$ is disjoint (since each $B_i \cap B_j = \emptyset$).

Let $\{B_1, B_2, ...B_k\}$ be a partition of S, such that $P(B) > 0$, for $i = 1, 2, ...k$. Then for any event A in S:

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

Proof:

$$P(A) = P((A \cap B_1) \cup (A \cap B_2) \cup ... \cup (A \cap B_k))$$
$$P(A) = P(A \cap B_1) + P(A \cap B_2) + ... + P(A \cap B_k))$$
$$P(A) = P(A|B_1)P(B_1) + ... + P(A|B_k)P(B_k)$$
$$P(A) = \sum_{i=1}^{k} P(A|B_i)P(B_i)$$

😄

Question 3: Identical balls in a box

Assume there are three identical boxes with 10 balls each. Box 1 has 7 red and 3 green balls, box 2 has 6 red and 4 green balls, and box 3 has 5 red and 5 green balls. A box is chosen at random and a ball is picked. What is the probability that the ball is red?

Solution: Define the following events: R : The chosen ball is red.
B_1 : The ball is chosen from box 1.
B_2 : The ball is chosen from box 2.
B_3 : The ball is chosen from box 3.

We have the following information:
$P(B_i) = \frac{1}{3}$ for $i = 1, 2, 3$
$P(R|B_1) = 0.7$
$P(R|B_2) = 0.6$
$P(R|B_3) = 0.5$

Then:

$$P(R) = \sum_{i=1}^{3} P(R|B_i)P(B_i)$$
$$P(R) = P(R|B_1)P(B_1) + P(R|B_2)P(B_2) + P(R|B_3)P(B_3)$$

$$P(R) = \frac{7}{10} \times \frac{1}{3} + \frac{6}{10} \times \frac{1}{3} + \frac{5}{10} \times \frac{1}{3}$$

Key takeaway: it is possible to express the probability of an event as multiple conditional probabilities, as well as the probabilities that the conditional probabilities will occur.

Theorem 1.5.2 Bayes' Theorem

Let $\{B_1, B_2, ..., B_k\}$ be a partition of S such that $P(B_i) > 0$, for $i = 1, 2, ...k$. Then for any event A in S:

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

Note:-

This theorem allows us to find the reverse of one or more conditional probabilities. That is, given $P(A|B_i)$, find $P(B_i|A)$. Finding $P(A)$ from multiple conditional and unconditional probabilities requires the law of total probability.

Question 4: Diagnostic test

A diagnostic test for a disease is such that it (correctly) detects the disease in 90% of individuals who actually have the disease. If a person does not have the disease, the test will report that they do not have it with accuracy of 90%.

Only 1% of the population has the disease. If a person is chosen at random and the diagnostic test indicates that they have the disease, what is the probability that the person actually has the disease?

Solution:
Define the following events:
D : Person has the disease
Pos : Test is positive
D^c : Person does not have the disease
Pos^c: Person is negative.
$P(Pos|D) = 0.9$
$P(Pos^c|D^c) = 0.9$
$P(D) = 0.01 \implies P(D^c) = 0.99$

We are looking for $P(D|Pos)$.

$$P(Pos|D^c) = 1 - P(Pos^c|D^c) = 0.1$$

$$P(D|Pos) = \frac{P(Pos|D) \times P(D)}{P(Pos|D) \times P(D) + P(Pos|D^c)P(D^c)}$$

$$P(D|Pos) = \frac{1}{12}$$

Definition 1.5.2: Sensitivity

The probability that a test detects the disease given that the patient has the disease.

$$P(Pos|D)$$

Definition 1.5.3: Specificity

The probability that a test indicates no disease given that the patient is disease free.

$$P(Pos^c|D^c)$$

Definition 1.5.4: Positive predictive value

The probability that a patient has the disease given that the test positive.

$$P(D|Pos)$$

1.6 Independence

Definition 1.6.1: Independence of two events

Two events A and B are independent $\iff P(B|A) = P(B)$.

We write that $A \perp\!\!\!\perp B$, or that $B \perp\!\!\!\perp A$.

Theorem 1.6.1

Two events A and B are independent $\iff P(A \cap B) = P(A)P(B)$.

Proof: Suppose A and B are independent. Then by definition, $P(B|A) = P(B)$. Therefore:

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

Conversely, assume $P(A \cap B) = P(A)P(B)$. Then:

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

😄

Definition 1.6.2: Mutual independence

Events $A_1, A_2, ..., A_n$ are said to be mutually independent if:

$$P(A_1 \cap A_2 \cap ... \cap A_k) = P(A_1) \times P(A_2) \times ... \times P(A_k)$$

for any collection of events selected from $A_1, A_2, ...A_n$.

Definition 1.6.3: Pairwise independence

The events $A_1, A_2, ..., A_n$ are said to be pairwise independent if for any two events A_i and A_j:

$$P(A_i \cap A_j) = P(A_i)P(A_j)$$

for any $i \neq j$.

Note:-

Pairwise independence does not imply mutual independence.

Corollary 1.6.1

If $B \perp\!\!\!\perp A$, then:

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

Proof: Proof that $A \perp\!\!\!\perp B^c$. Show that $P(A \cap B^c) = P(A)P(B^c)$.

$$P(A \cap B^c) = P(A)P(B^c|A)$$
$$P(A \cap B^c) = P(A)(1 - P(B|A))$$
$$P(A \cap B^c) = P(A)(1 - P(B))$$

From the right side:

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

😊

Theorem 1.6.2 Independence between two sets of events

Let $U = \{A_1, ...A_k$ and $W = \{B_1, ...B_l\}$. Then $U \perp\!\!\!\perp W$ if the probability of the intersection of every $A_i's$ with intersection of every set of $B_i's$ is equal to the product of the probabilities of their intersections.

Question 5

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

Solution: We are looking for the event $A = T \cap T \cap ... \cap H$.

Then $P(A) = P(T \cap T \cap ... \cap H)$. Since all coin tosses are independent, we can write:

$$P(A) = (P(T))^5 \times P(H)$$
$$P(A) = (\frac{1}{2})^6$$
$$P(A) = \frac{1}{64}$$

Theorem 1.6.3 Independence and mutual exclusivity

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

Theorem 1.6.4 Independence and mutual exclusivity when empty set

Two mutually exlusive events A and B are independent only if either $P(A) = 0$ or $P(B) = 0$.

Proof: Both of these theorems can be proved by checking $P(A)P(B)$ and comparing to $P(A \cap B)$. 😊

1.6.1 Union of independent events

Identify one of the following:

1. $P(A \cap B) = 0$ (mutual exclusivity)
2. $P(A \cap B) = P(A)P(B)$ (independent events)
3. $P(A \cap B) = P(A)P(B|A) = P(B)P(A|B)$ (non-independent events)

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

Example 1.6.1 (Union of several independent events)

$$P(A_1 \cup A_2 \cup ... \cup A_n) = 1 - P(A_1 \cup A_2 \cup ... \cup A_n)^c$$
$$P(A_1 \cup A_2 \cup ... \cup A_n) = 1 - P(A_1^c \cap A_2^c \cap ... \cap A_n^c)$$
$$P(A_1 \cup A_2 \cup ... \cup A_n) = 1 - P(A_1^c)P(A_2^c)...P(A_n^c)A$$
$$P(A_1 \cup A_2 \cup ... \cup A_n) = 1 - (1 - P(A_1))(1 - P(A_2))...(1 - P(A_n))$$

Chapter 2

Random variables and distributions

2.1 Random variables and sampling

2.1.1 Types of sampling

1. **Sampling with replacement**: Outcomes are independent.
2. **Sampling without replacement**: Outcomes are dependent.
3. **Sampling without replacement with small sample size compared to population**: Can assume that outcomes are pretty much independent.

Question 6

Suppose that in a very large city 20% of people have a genetic mutation. If 10 people are examined, what is:

1. The probability that exactly two have the mutation?
2. The probability that at least two will have the mutation?

Solution: We are sampling without replacement. Let the outcome on trial i be:
M_i: the i-th individual has the mutation.
$M^c i$: the i-th individual does not haev the mutation..

for $i = 1, 2, ...10$.

Let $X = \#$ of mutations in these 10 trials.

To find $P(X = 2)$, we first find probability of a specific configuration of mutations and non-mutations in 10 trials that results in two mutations out of 10. We will sum up the probabilities of all such configurations.

One such configuration is:

$$M_1 \cap M_2 \cap M_3^c \cap M_4^c \cap ... \cap M_{10}^c$$
$$= P(M_1 \cap M_2 \cap M_3^c \cap M_4^c \cap ... \cap M_{10}^c)$$
$$= P(M_1)P(M_2)P(M_3^c)...P(M_{10}^c)$$

(by independence - one person having a mutation does not influence another person having the mutation)

$$= (0.2)^2(0.8)^2$$

We can read this as there being a 20% chance that one person having the mutation. In reality if we are sampling without replacement, there would be a slightly higher chance that the next person picked has the mutation. With a very large population this effect is negligible. So we assume independence of events.

Every configuration includes 2 people with the mutation ($(0.2)^2$ chance) and 8 people without the mutation ($(0.8)^2$ chance).

The number of possible configurations of 2 people with mutations is given by $\binom{10}{2}$ (number of ways to pick 2 mutations from 10, where order does not matter). Each configuration with 2 mutations has a probability $(0.2)^2(0.8)^2$ of occurring. Each configuration is disjoint.

Therefore, the total probability of any of these configurations occurring is $\binom{10}{2}(0.2)^2(0.8)^8$.

General formula:

$$P(X \geq 2) = \sum_{k=2}^{10}(P(X = k))$$

$$P(X \geq 2) = \sum_{k=2}^{10}(0.2)^k(0.8)^{10-k}\binom{10}{k}$$

Or, by using the complement:

$$P(X \geq 2) = 1 - P(X < 2)$$

$$P(X \geq 2) = 1 - \{P(X = 0) + P(X = 1)\}$$

$$P(X \geq 2) = 1 - (0.2)^2(0.8)^{1}0\binom{10}{0} - (0.2)^2(0.8)^9\binom{10}{1}$$

Definition 2.1.1: Random variable

A random value is a real-valued function defined over a sample space.

Example 2.1.1

Toss 3 fair coins. Let $X = \#$ of tails observed. Then $X = 0, 1, 2, 3$. Each has an associated probability: $P(X = x)$.

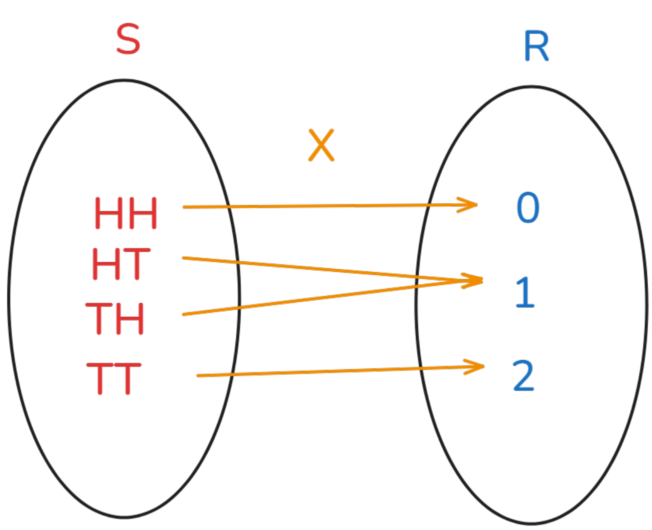

Figure 2.1: Random variable function

2.1.2 Discrete random variable distribution

Definition 2.1.2: Discrete random variable

A random variable is **discrete** if its range is finite or countably infinite. Then the set of values that X can take can be described as: $R_x = \{x_1, x_2, ...\}$.

Definition 2.1.3: Probability distribution

Probability distribution can be described as $\rho(x) = P(X = x)\forall x$. Then $\sum_{x=0}^{n} P(X = x) = 1$

Definition 2.1.4: Cumulative distribution function

Defined as:

$$F_X(x) = P(X \leq x), \forall x \in R$$

The domain of $F_X(x)$ is the real line. The range is $[0, 1]$.

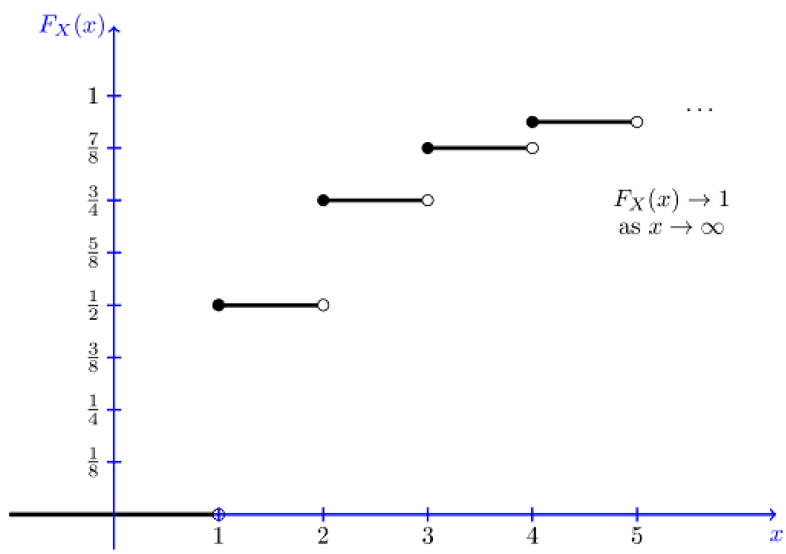

Figure 2.2: Cumulative distribution function

Example 2.1.2

Recall the 3 coin throws.

$$P(X \leq 1) = P(X = 0) + P(X = 1)$$

$$F_X(1) = P(X \leq 1)$$

Note:-

Possible to write individual probabilities as difference of cumulative probabilities. For example:

$$F_X(2) - F_X(1) = P(X = 2)$$

More generally:

$$P(X = x_k) = F_X(x_k) - F_X(x_k - \epsilon)$$

Properties of CDF

1. $x \leq y$, then $F_X(x) \leq F_X(y)$
2. $lim_{x \to + \inf} F_X(x) = 1$
3. $lim_{x \to - \inf} F_X(x) = 0$
4. Right continuous

Note:-

Mathematical expression for right continuity is:

$$lim_{x \to x_k^+} F_X(x) = F_X(x_k)$$

for $x_k \leq x < x_{k+1}$

Definition 2.1.5: Probability mass function

The probability mass function of a discrete random variable is give by:

$$P_X(x_k) = P(X = x_k)$$

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

Definition 2.1.6: Discrete uniform distribution

A random variable X is said to have a discrete uniform distribution if for $x_1, x_2, ...x_n \in R$ where R is the set of real numbers:

$$\rho_X(x) = P(X = x) = \frac{1}{N}$$

for $X = x_1, x_2, ...x_N$

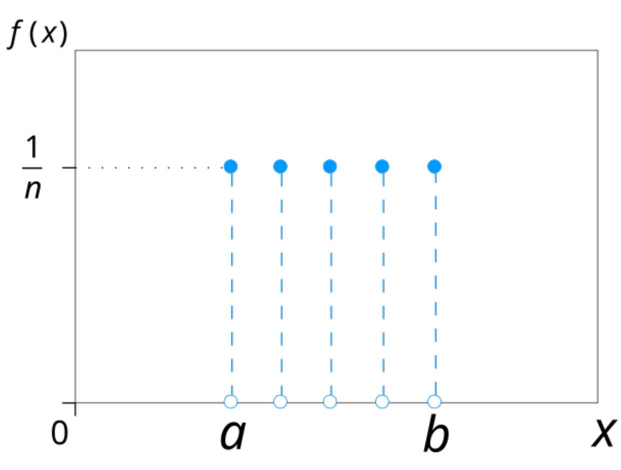

Figure 2.3: Discrete uniform distribution

Example 2.1.3

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

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

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

Definition 2.1.7: Bernoulli Distribution

A random X is said to have a Bernoulli Distirbution if X can take only two possible values, usually 0 and 1. Is used to model situation where random experiment that has two possible outcomes, usually referred to as "success" or "failure".

Probability of success is denoted p, $0 < p < 1$.

Then probabilty distribution is given by:

$$P(X = x) \begin{cases} p^x(1-p)^{1-x} \text{ if } x \in \{0, 1\} \\ 0 \text{ otherwise} \end{cases} \tag{2.1}$$

2.2 Binomial distribution

Definition 2.2.1: Binomial experiment

An experiment is called a binomial experiment if it consists of n independent Bernoulli trials (trial with two outomes: success and failure). The probability of success p remains constant from trial to trial. We are interested in x successes out of n trials.

Definition 2.2.2: Binomial distribution

Let X be the random variable that counts the number of successes in n **independent** Bernoulli trials, where the probability of success remains **constant** from trial to trial. The probability distribution is denoted:

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

Example 2.2.1

According to tables by the National Centre for Health Sciences in Vital Statistics for United States, a person at the age of 20 years has a probability of 0.8 of being alive at the age of 65 years. Suppose three people of age 20 years are chosen at random (all people in the population of 20 years old the US are equally likely to be chosen).
Let A_i be the event that the i^{th} person chosen will be alive at the age of 65. Then A_i^c is the event that the i^{th} person chosen will not be alive at the age of 65.
We are given that a person at the age of 20 years has the probability of 0.8 at the age of 65 years.
Two possible outcomes: alive or dead, so these are Bernoulli trails. Then the random experiment of observing whether a person is alive at the age of 65 can be modelled by using Bernoulli random variable.
Let X be the random variable that takes values 0 or 1 where 0 corresponds to not alive and 1 corresponds to being alive at the age of 65.
Number of possible ways you can observe a given event is the number of ways it occurs $\binom{n}{x}$ times the probability tha it occurs. For example, in a sample of 3, to observe if 2 are alive at age of 65, you would find this by $\binom{3}{2}(0.8)^2(0.2)^1 = 0.384$.

Theorem 2.2.1 Binomial distribution

For a binomial random variable X

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

where $x = 0, 1, ...n$

Question 7

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

Solution: There are two outcomes: defective or not defective. Can represent X (number of successes) in n trials as a binomial random variable.

$$n = 5$$

$$\mathrm{X} \sim \mathrm{Binom}(5, 0.5$$

Then to find at least one defective:

$$P(X \geq 1) = P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4) + P(X = 5)$$

$$= \sum_{x=1}^{5} \binom{5}{x} (0.05^x)(0.95)^{5-x}$$

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

$$= 1 - \binom{5}{0} (0.05)^0 (0.95)^5$$

2.3 Geometric distribution

Definition 2.3.1: Geometric distribution

A random variable X is said to have geometric distribution if the sample space countains a **countably infinite** number of sample points, where we are looking for the probability that an event first occurs on the x^{th} trial.

Theorem 2.3.1

Let X be the trial at which the first success occurs. Then

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

$$0 < p < 1, x = 1, 2, 3, ...$$

Proof: The event $X = x$ can be expressed as :

$$\{X = x\} = \{FFFF...FS\}$$

where we observe $x - 1$ failures. Let p be the probability of success.
$P(X = x) = P(F \cap F \cap F ... \cap F \cap S)$

$$= P(F)^{x-1} P(S)$$

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

😃

Question 8

You ask people outside a polling station who they voted for until you find someone that voted for the Liberal candidate in a recent election. The probability that a randomly chosen person votes a liberal candidate is 0.4 What is the probability that the first person who voted for the Liberals is the fifth person you interviewed?

Solution: Using geometric distribution. We are finding $X = 5$ and $p = 0.4$:

$$P(X = 5) = (1 - 0.4)^4(0.4)$$
$$P(X = 5) = 0.05184$$

2.4 Negative binomial distribution

Definition 2.4.1: Negative binomial distribution

Let X be a negative binomial random variable. Then:

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

where $x = r, r+1, \ldots$. We write this as $X \sim NB(r, p)$ where r is the number of successes. Then this is the probability of getting the r-th success on the x-th trial.

Proof: The event $\{X = x\}$ can be written as:

$$\{X = x\} = \{r - 1 \text{ successes in x - 1 trials }\} \cap \{\text{r-th success on xth trial}\}$$

$$P(X = x) = P(E_1 \cap E_2)$$
$$= P(E_1) \cdot P(E_2)$$
$$= \binom{x-1}{r-1} p^{r-1} (1-p)^{(x-1)-(r-1)} \cdot p$$
$$= \binom{x-1}{r-1} p^r (1-p)^{x-r}$$

☺

Question 9

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

Solution:

$$x = 7$$
$$x \sim NB(3, 0.2)$$

$$P(x = 7) = \binom{6}{2} (0.3)^3 (0.8)^4$$

2.5 Poisson distribution

Definition 2.5.1: Poisson distribution

Let x be the number of successes in a unit time interval or unit space. let λ be the average number of events that occur within a given interval of time or space.

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

Example 2.5.1

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

Solution:

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

We recognize this as a Maclaurin series for e^{λ}.

$$= e^{-\lambda}(\frac{\lambda^0}{0!} + \frac{\lambda^1}{1!} + ...)$$
$$= e^{-\lambda} \cdot e^{\lambda}$$
$$= 1$$

Theorem 2.5.1

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

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

In other words, the bimomial distribution is approximated by the Poisson approximation when $n \longrightarrow \infty$ and $p \longrightarrow 0$.

Proof: Let $X \sim Binom(n, p)$, where $n \longrightarrow \infty$ and $p \longrightarrow 0$ and $np = \lambda$. Then:

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

$$= \lim_{n\to\infty} \binom{n}{x} (\frac{\lambda}{n})^x (1 - \frac{\lambda}{n})^{n-x}$$

$$= \lim_{n\to\infty} \frac{n!}{(n-x)!x!} \frac{\lambda^2}{n^2} (1 - \frac{\lambda}{n})^n \cdot (1 - \frac{\lambda}{n})^{-x}$$

$$= \frac{\lambda^x}{x!} (\lim_{n\to\infty} \frac{1}{n^x} \frac{n!}{(n-x)!} (1 - \frac{\lambda}{n})^n)(1 - \frac{\lambda}{n})^{-x})$$

We consider each of the factors separately. Considering only

$$\lim_{n\to\infty} \frac{1}{n^x} \frac{n!}{(n-x)!}$$

$$= \lim_{n\to\infty} \frac{n(n-1)(n-2)...(N-(x-1))}{n^x}$$

$$\lim_{n\to\infty} 1 \cdot (1-\frac{1}{n})(1-\frac{2}{n})...(1-\frac{x-1}{n})$$

Recall that:

$$lim_{n\to\infty}(1+\frac{1}{n})^n = e$$

$$lim_{n\to\infty}(1-\frac{\lambda}{n})^n = e^{-\lambda}$$

This means that

$$lim_{n\to\infty}(1-\frac{\lambda}{n})^n = e^{-\lambda}$$

Finally, consider

$$lim_{n\to\infty}(1-\frac{\lambda}{n})^{-x}$$

$$= \lim_{n\to\infty} \frac{1}{(1-\frac{\lambda}{n})^x}$$

Therefore, multiply all these limits (taken separately) together:

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

😁

Note:-

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

Question 10

Suppose that sections of textile length 1cm have a flaw in them with probability 0.01. If 1000 such sections are examined, what is the probability that:
a. exactly 10 will have a flaw? b. at least 50 will have a flaw?

Solution: a. Looking for $P(X=10)$. We can model this using a binomial distribution: $x \sim Binom(1000, 0.01)$. Then:

$$P(X=10) = \binom{n}{x} p^x (1-p)^{n-x}$$

$$= \binom{1000}{10}(0.01)^{10}(0.9)^{990}$$

But because n is large and np is small, we can approximate the binomial distribution by the Poisson distribution.

$$P(X=10) = \frac{\lambda^x e^{-\lambda}}{x!}$$

$$= \frac{10^{10}e^{-10}}{10!}$$

Question 11

Industrial accidents occur according to a Poisson process with an average of 3 accidents pers month. What is the probbaility of 10 accidents during the last 2 months?

Solution: The number of accidents in two months, X has Poisson probability distribution with mean $\lambda^t = 2(3) = 6$.

2.6 Hypergeometric distribution

Definition 2.6.1

Let N be the sample size. Let a be the number of favorable items in total sample space. Let x be the number of favorable items in a sample of size n. Then:

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

Write as $X \sim hypergeometric(N, n, a)$ where $x = 0, 1, 2, ...min(a, n)$.

2.7 Expectation and variance

The expected value of a random variable X is the measure of centrality of X. It is the weighted mean or average of all possible values of the random variable X.

Definition 2.7.1

Let X be a discrete random variable with the possible values $\{x_1, x_2, x_3, ...\}$ (finite or countably infinite). The expected value of X, denoted $E(X)$, is defined as:

$$E(X) = \sum x_k P(X = x_k)$$

Example 2.7.1

Rolling a fair die. The expected value is given by:

$$E(X) = \frac{1}{6}(1 + 2 + 3 + 4 + 5 + 6)$$
$$E(X) = 3.5$$

Properties of expected variable

1. The expected value of X can be thought of as a long run average of X.
2. The expected value of X exists if E|X| converges.

Question 12

Consider the probability mass function $P(X = x) = \frac{c}{x^2}$ of the random variable X, where $x = 1, 2, ...$ Find the expected value of X.

Solution:

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

$$E(x) = \sum_{x=1}^{\infty} x \cdot \frac{x}{x^2}$$

$$E(x) = \sum_{i=1}^{\infty} \frac{c}{x}$$

$$E(x) = c \sum_{i=1}^{\infty} \frac{1}{x}$$

This is a divergent series, so the expected value of x does not exist.

Definition 2.7.2: Mathematical expectation

When it exists, the mathematical expectation satisfies the following properties. if a is a constant, then:

1. $E(a) = a$
2. If a is a constant, then $E(aX) = aE(X)$.
3. The mathematical expectation is linear. That is:

$$E(X_1 + X_2 + ... + X_n) = E(X_1) + E(X_2) + ...E(X_n)$$

for any set of random variables $X_1, x_2, ...X_n$.

Mathematical expectation follows properties of linear transformations. Let a_1, a_2 be two constants that when the mathematical expectation exists, satisfies the following property:

$$E(a_1X_1 + a_2X_2) = a_1E(X_1) + a_2E(X_2)$$

Question 13

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

Solution: Let C be premium for policy. Company's profit is $C-15$ if event A does not occur and $C-15-1000$ if A does occur. Let X be a random variable that represents the company's profit:

$$E(X) = 50 = (C - 15)(0.98) + (C - 15 - 1000)(0.02)$$

$$C = 85$$

Definition 2.7.3: Variance

The variance of a random variable X with mean $E(X) = \mu_X$ is defined as:

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

Properties of variance

1. The variance of the random variable X is denoted by

$$Var(X) = \sigma_X^2$$

2. The variance of a random variable X is the expected value of $(X - \mu_X)^2$.
3. $Var(cX) = c^2Var(X)$

Theorem 2.7.1

$$Var(X) = E[X^2] - [E(X)]^2$$

Proof:

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

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

😄

Question 14

What is the variance of a fair dice throw?

Solution: Know that $E(X) = 3.5$. Then:

$$\sigma_2 = Var(X) = E[X^2] - [E(X)]^2$$
$$= \frac{91}{6} - [\frac{7}{2}]^2$$
$$= 2.92$$

$$\sigma_X = \sqrt{2.921} = 1.71$$

2.7.1 Moments of a distribution

Definition 2.7.4

Let X be a discrete random variable. Then the moments of X around the origin are $E(X), E(X^2), ...E(X^k)$, where $E(X^r) < \infty$ for $r = 1, 2, ...k$.

The mean of the random variable $E(X)$ is the first moment of X. The variance of the random variable $E(X - \mu)^2$ is the second moment of X about the mean.

Solving system of linear equations, can obtain the mean and variance from the classical and factorial moments.

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

Definition 2.7.5

Let X be a random variable. Then the r^{th} factorial moment of X is:

$$E(X_r) = E[X(X-1)...(X-(r-1))]$$

$E(X)$: First factorial moment of X.
$E[X(X-1)]$: Second factorial moment of X.

2.7.2 Mean and variance of uniform distribution

Definition 2.7.6

A random variable X is said to have a discrete uniform distirbution if $\rho_X(x) = P(X = x) = \frac{1}{N}$ for $X = x_1, x_2, ...x_N$.

Mean:

$$E(X) = \sum_{k=1}^{N} x_k P(X = x_k)$$
$$= \sum_{k=1}^{N} x_k \rho_X(x)$$

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

Variance:

$$E(X^2) = \sum_{k=1}^{N} x_k^2 P(X = x_k)$$
$$= \sum_{k=1}^{N} x_k^2 \rho_X(x)$$
$$= \frac{1}{N}\sum_{k=1}^{N} x_k^2$$
$$Var(X) = E[X^2] - [E(X)]^2$$
$$= \frac{1}{N}\sum_{k=1}^{N} x_k^2 - \overline{x}^2$$

2.7.3 Mean and variance of Bernoulli distribution

Recall the probability distribution of the Bernoulli random variable X:

Table 2.1: Probability distribution of Bernoulli distribution

x	P(X=x)
0	1-p
1	p

Mean:

$$E(X) = 0(1-p) + 1(p) = p$$
$$E(X^2) = 0^2(1-p) + 1^2(p) = p$$
$$Var(X) = E[X^2] - [E(X)]^2$$
$$= p - p^2 = p(1-p)$$

2.7.4 Summary

Table 2.2: Mean and Variance of Probability Distributions

Distribution	Support	Parameters	$P(X = x)$	Mean	Variance
Uniform (Discrete)	$\{a, a+1, \ldots, b\}$	a, b	$\frac{1}{b-a+1}$	$\frac{a+b}{2}$	$\frac{(b-a+1)^2-1}{12}$
Bernoulli	$\{0, 1\}$	p	$p^x(1-p)^{1-x}$	p	$p(1-p)$
Binomial	$\{0, 1, \ldots, n\}$	n, p	$\binom{n}{x}p^x(1-p)^{n-x}$	np	$np(1-p)$
Poisson	$\{0, 1, 2, \ldots\}$	λ	$\frac{e^{-\lambda}\lambda^x}{x!}$	λ	λ
Geometric	$\{1, 2, 3, \ldots\}$	p	$(1-p)^{x-1}p$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
Negative Binomial	$\{r, r+1, \ldots\}$	r, p	$\binom{x-1}{r-1}(1-p)^{x-r}p^r$	$\frac{r}{p}$	$\frac{r(1-p)}{p^2}$

Chapter 3

Functions of random variables

3.1 Finding functions of random variables

3.1.1 Using the CDF

Let X be a continuous random variable and $Y = f(X)$ be a function of X. Then Y is a continuous random variable too. This method consists of two steps.

1. Find the CDF of Y.
2. Differentiate the CDF to find the PDF of Y.

Example 3.1.1
Let X be a random variable with PDF:

$$f_X(x) = \begin{cases} 2e^{-2x} & \text{for } x > 0 \\ 0 & \text{otherwise} \end{cases} \tag{3.1}$$

Recall that:

$$f(x) = \begin{cases} \frac{1}{\beta}e^{-x/\beta} & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases} \tag{3.2}$$

$X \sim exp(\beta), E(X) = \beta$ and $Var(X) = \beta^2$. Determine the PDF of $Y = \sqrt{X}$.

Solution: First observe that if $x \geq 0$, $y \geq 0$, and when $x < 0$, y is not defined. So support is the same.

$$\begin{aligned} P(Y \leq y) &= P(\sqrt{X} \leq y) \\ &= P(X \leq y^2) \end{aligned}$$

See that $F_Y(y) = F_X(y^2)$. Now we differentiate the CDF of Y to find PDF of Y.

$$F'_Y(y) = \frac{d}{dx}(F_X(y^2))$$

$$\begin{aligned} P(X \leq y^2) &= \int_0^{y^2} 2e^{-2x} dx \\ &= \left| \frac{2e^{-2x}}{-2} \right|_0^{y^2} \\ &= \left| -e^{-2x} \right|_0^{y^2} \\ &= -e^{-2y^2} - (-1) \end{aligned}$$

$$= 1 - e^{-2y^2}$$

Now differentiate with respect to y:

$$F'(y) = \frac{d}{dy}(1 + -e^{-2y^2})$$

$$= -e^{-2y^2} \cdot (4y)$$

$$= 4ye^{-2y^2}$$

$$f_Y(y) = \begin{cases} 4ye^{-2y^2} & y > 0 \\ 0 & \text{otherwise} \end{cases} \tag{3.3}$$

Example 3.1.2

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

$$f_Y(y) = \begin{cases} 2y & 0 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases} \tag{3.4}$$

Let U be the profit given by

$$U = 3Y - 1$$

Find the PDF of U.

Solution: Find the CDF of U in terms of the CDF of Y.

$$P(U \leq u) = P(3Y - 1 \leq u)$$

Looking for $P(Y \leq \frac{u+1}{3})$.

The support for y is $0 \leq y \leq 1$. So $0 \leq \frac{u+1}{3} \leq 1$. The support for y is therefore:

$$0 \leq u + 1 \leq 3$$

$$F_Y(\tfrac{u+1}{3}) = \int_0^{\frac{u+1}{3}} 2y dy$$

$$= \Big| \frac{2y^2}{2} \Big|_0^{\frac{u+1}{3}}$$

$$= (\frac{u+1}{3})^2$$

$$F_U(u) = F_Y(\frac{u+1}{3}) = (\frac{u+1}{3})^2$$

$$F'_U(u) = \frac{d}{du}(\frac{u+1}{3})^2 = \frac{2}{9}(u+1)$$

$$f_U(u) = \begin{cases} \frac{2}{9}(u+1) & -1 \leq u \leq 2 \\ 0 & \text{otherwise} \end{cases} \tag{3.5}$$

3.1.2 Change of variable technique

Let X be a continuous random variable with CDF $F_X(x)$ and PDF $f_X(x)$. Let $Y = g(X)$ for some function g. If g is strictly increasing or strictly decreasing, then we proceed as follows to find the PDF of Y.

1. Write the CDF of Y in terms of the CDF of X. So we have $P_Y(y) = P(Y \leq y) = P(g(x) \leq y)$. As $x_2 > x_1$, $g(x_2) > g(x_1)$. So we have that:

$$F_Y(y) = P(x \leq g^{-1}(y))$$

$$= F_X(g^{-1}(y))$$

2. Differentiate the above expression with respect to y.

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

$$= f_X(g^{-1}(y)) \cdot |\frac{d}{dy}(g^{-1}(y))|$$

$$= f_X(g^{-1}(y)) \cdot |\frac{d}{dy}(x)|$$

Basically, find an expression for X in terms of Y to get $g^{-1}(y)$. Then find the derivative of $g^{-1}(y)$. Plug into this formula:

$$f_Y(y) = f_X(g^{-1}(y)) \cdot |\frac{d}{dy}g^{-1}(y)|$$

Example 3.1.3
Let $Y = \sqrt{X}$ where the PDF of X is:

$$f_X(x) = \begin{cases} 2e^{-2x} & x \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{3.6}$$

This is an exponential distribution with $\beta = \frac{1}{2}$.

Solution: Set $F_Y(y) = P(Y \leq y) = P(\sqrt{X} \leq y)$

$$= P(X \leq y^2)$$

$$= F_X(y^2)$$

Now we differentiate:

$$F_Y'(y) = f_Y(y)$$

$$= F_X'(y^2)\frac{d}{dy}(y^2)$$

$$= f_X(y^2) \cdot (2y)$$

$$= 2e^{-2y^2} \cdot 2y$$

$$= 4e^{-2y^2}$$

when $y \geq 0$, 0 otherwise.

Example 3.1.4
Use the transformation method to find the PDF of $Z = \frac{x-\mu}{\sigma}$.

Solution: Know the PDF of X:

$$f_X(x) = \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sigma}e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} \tag{3.7}$$

Now we write $F_Z(z)$ in terms of x:

$$F_Z(z) = P(\frac{x-\mu}{\sigma} \leq z)$$

$$= P(x \leq \mu + \sigma z)$$

$$= F_X(\mu + \sigma z)$$

Differentiate our CDF to find our PDF:

$$f_Z(z) = F_Z'(z)$$

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

For $-\infty < x < \infty$, $-\infty < z < \infty$. See that $Z \sim N(0, 1)$.

Example 3.1.5

Let $Z \sim N(0, 1)$, then show that $Z^2 \sim \chi^2(1)$. Recall that $\chi^2(U)$ is a special case of $Gamma(\alpha, \beta)$ where $\alpha = U/2$ and $\beta = 2$.

Solution: Let $Y = Z^2$.

$$F_Y(y) = P(Y \leq y) = P(Z^2 \leq y)$$

Have that

$$
\begin{aligned}
&Z^2 \leq y \\
&\implies \pm z \leq \sqrt{y} \\
&\implies z \leq \sqrt{y} \\
&-z \leq \sqrt{y} \\
&-\sqrt{y} \leq z \leq \sqrt{y}
\end{aligned}
$$

Therefore:

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

Now we can differentiate:

$$
\begin{aligned}
f_Y(y) = F'_Y(y) &= F'_Z(\sqrt{y}) \cdot \frac{d}{dz}(\sqrt{y}) - F'_Z(-\sqrt{y}) \cdot \frac{d}{dz}(-\sqrt{y}) \\
f_Y(y) &= f_Z(\sqrt{y}) \cdot \frac{1}{2} y^{-1/2} - f_z(-\sqrt{y}) \cdot (-\frac{1}{2} y^{-1/2}) \\
&= \frac{1}{2} y^{-1/2}(f_Z(\sqrt{y}) + f_Z(-\sqrt{y}))
\end{aligned}
$$

Due to symmetry, $f_Z(\sqrt{y}) = f_Z(\sqrt{-\sqrt{y}}$.

$$
\begin{aligned}
&= \frac{1}{2} y^{-1/2}(2 f_Z(\sqrt{y})) \\
&= y^{-1/2} f_Z(\sqrt{y})
\end{aligned}
$$

Important that $y \geq 0$.

$$= y^{-1/2} \cdot \frac{1}{2\pi} e^{-1/2y}$$

Let $x \sim Gamma(\alpha, \beta)$.

$$f_X(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}$$

Rewrite expression in this form:

$$= \frac{1}{\sqrt{2}} \cdot \frac{1}{\sqrt{\pi}} y^{-1/2} e^{-y/2}$$

Recall that $\Gamma(1/2) = \sqrt{\pi}$.

$$= \frac{1}{2^{1/2}} \frac{1}{\Gamma 1/2} y^{1/2-1} \cdot e^{-y/2}$$

This is exactly the $\chi^2(1)$ distribution, where $\alpha = 1/2$ and $\beta = 2$.

3.1.3 Probability integral transformation

Theorem 3.1.1

Let X be a continuous random variable with CDF $F_X()$. Then the function $F_X(X)$ (notice that this is a different random variable) is uniformly distributed, i.e.

$$F_X(x) \sim U[0,1]$$

Note:-

Note that $F_X(X)$ is not a probability. Whereas $F_X(x)$ is a probability. Therefore:

$$F_X(X) \neq P(X \leq X)$$

So $F_X(x)$ gives cumulative probabilities, but $F_X(X)$ is a transformation of X and $F_X(X) \sim U[0,1]$.

Example 3.1.6

Let X be a continuous random variable with PDF:

$$f_X(x) = \begin{cases} 3x^2 & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases} \tag{3.8}$$

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

Solution: Let us find $F_X(X) = \int_0^X 3t^2 dt$.

$$= \left| \frac{3t^3}{3} \right|_0^X$$

$$= X^3$$

Now we have obtained the transformation:

$$F_X(X) = X^3$$

This is a new function of X. Let $U = X^3$. We have to show that $U = X^3 \sim Unif(0,1)$.

First, express $F_U(u) = P(U \leq u) = P(X^3 < u) = P(X \leq U^{1/3})$

$$= F_X(u^{1/3})$$

Now differentiate:

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

Therefore $u \sim Unif(0,1)$.

Example 3.1.7

Let $X \sim exp(\beta)$, where

$$f_X(x) = \begin{cases} \frac{1}{\beta} e^{-x/\beta} & x \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{3.9}$$

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

Solution: First, find $F_X(x)$.

$$\begin{aligned} F_X(x) &= \int_0^X \frac{1}{\beta} e^{-x/\beta} \\ &= \frac{1}{\beta}\int_0^x e^{\frac{-x}{\beta}} \\ &= \frac{1}{\beta}(-\beta)e^{-x/\beta}\Big|_0^x \\ &= 1 - e^{-x/\beta} \end{aligned}$$

Now we have

$$F_X(x) = -e^{-X/\beta}$$

Let $U = 1 - e^{-x/\beta}$. Then $x = -\beta ln(1-U)$. So:

$$F_X(-\beta ln(1-U)))$$

Differentiating:

$$\begin{aligned} f_U(u) &= F'_X(-\beta ln(1-u)) \cdot \frac{d}{du}(-\beta ln(1-u)) \\ &= f_X(-\beta ln(1-u)) \cdot \frac{\beta}{1-u} \\ &= \frac{1}{\beta} e^{\frac{-(-\beta ln(1-u))}{\beta}} \cdot \frac{\beta}{1-u} \\ &= \frac{1}{\beta} e^{ln(1-u)} \cdot \frac{\beta}{1-u} \\ &= \frac{1}{\beta} ln(1-u)\frac{\beta}{1-u} \\ &= 1 \end{aligned}$$

We see that $0 \le u \le 1$ in order for x to be defined. AKA $u \ge 0$. So $u \sim Unif(0,1)$.

Proof: Let X be a continuous random variable. Let $P(X \le x) = F_X(x)$. We get a new function of x from the CDF by replacing 'x' by X. Let $U = F_X(X)$. Want to find the pdf of U.

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

This is the PDF of the uniform random U over the interval $[0,1]$. Therefore $F_X(X) \sim Unif(0,1)$

😃

Note:-

Suppose we want to generate random variables from the exponential distribution $exp(\beta)$.

1. Simulate $U \sim Unif(0,1)$.
2. $U = F_X(X)$ equal to transformation obtained from the CDF.
3. Find $X = F_X^{-1}(U)$.

3.2 Moment generating functions

Definition 3.2.1: Moment generating function

The moment generating function (mgf) of the distribution of the random variable X is the function of a real parameter t defined by:

$$M_X(t) = E[e^{tX}]$$

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

From the MGF, we have:

1. For the discrete random variables, MGF is:

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

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

2. For continuous random variables,

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

Note that the PDF (or PMF) of the random variable X can be obtained from its moment generating function and vice-versa.

Let X be a continuous random variable. Then

$$M_X(t) = E[e^{tx}]$$
$$= \int_{-\infty}^{\infty} e^{tx} f_X(x) dx$$
$$M'_X(t) = \frac{d}{dt} \int_{-\infty}^{\infty} e^{tx} f(x) dx$$
$$= \int_{-\infty}^{\infty} \frac{d}{dt} e^{tx} f_X(x) dx$$
$$= \int_{-\infty}^{\infty} e^{tx} x f_X(x) dx$$
$$M'_X(t)\Big|_{t=0}$$
$$= \int_{-\infty}^{\infty} e^{0} x f_X(x) dx$$
$$= \int_{-\infty}^{\infty} x f_X(x) dx$$
$$M_X(t)\Big|_{t=0} = E(X)$$

The first derivative of $M_X(t)$ evaluated at $t = 0$ gives $E(X)$. This is the first moment about the origin.
Similarly, $M''_X(t)\Big|_{t=0} = E(x^2)$
Generalizing

$$M_X^{(k)}(t)\Big|_{t=0} = E(X^k)$$

Example 3.2.1

Let $X \sim Binom(n, p)$.

1. Find the moment generating function of the distribution of X.
2. Show that $M'_X(t)\big|_{t=0} = E(X) = np$.
3. Show that $M''_X(t) = E(x^2) = n^2p^2 - np^2 + np$
4. Find $Var(X)$.

Solution: (1)

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

Binomial form:

$$(a+b)^n = \sum_{x=0}^{n}\binom{n}{x}a^x b^{n-x}$$

Using this, we get:

$$M_X(t) = (pe^t + 1 - p)^n$$

(2)

$$\begin{aligned} M'_X(t) &= \frac{d}{dt}(pe^t + 1 - p)^n \\ &= n(pe^t + 1 - p)^{n-1} \cdot pe^t \\ M'_X(t)\big|_{t=0} & n(pe^0 + 1 - p)^{n-1} \cdot pe^0 \\ &= n(p + 1 - p)^{n-1} \cdot p \\ &= np = E(X) \end{aligned}$$

(3)

$$M''_X(t)\big|_{t=0} = E(x^2)$$

Since we have $M'_X(t)$, differentiate again:

$$M''_X(t) = \frac{d}{dt}[n(pe^t + 1 - p)^{n-1} \cdot pe^t]$$

$$= n(n-1)(pe^t + 1 - p)^{n-2} \cdot (pe^t)^2 + n(1 - p + pe^t)^{n-1} \cdot pe^t$$

Substitute $t = 0$.

Question 15

Let $X \sim Poiss(x, \lambda)$. Show that $E(X) = M'_X(t)\big|_{t=0}$ and that $E(X^2) = M''_X(t)\big|_{t=0}$. Note that $P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}$. Use the Maclaurin's series.

Question 16

Let $X \sim Gamm(X, \lambda)$. Show that $M_X(t) = \frac{1}{(1-BtP^{\alpha}}$. Hint: set up your expression as the gamma function instead of evaluating the integral.

Example 3.2.2

The normal distribution. Start with $Z \sim N(0,1)$. Recall the normal distribution:

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$$

$$M_Z(t) = E(e^{tz}) = \int_{-\infty}^{\infty} e^{tz} f_Z(z) dz$$
$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tz-\frac{1}{2}z^2} dz$$

Completing the square:

$$tz - \frac{1}{2}z^2$$
$$= \frac{-1}{2}(z^2 - 2tz + t^2) + \frac{1}{2}t^2$$
$$= -\frac{1}{2}(z-t)^2 + \frac{1}{2}^2$$

So:

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(z-t)^2 + \frac{1}{2}t^2} dz$$
$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(z-t)^2} \cdot e^{\frac{1}{2}t^2} dz$$
$$= e^{\frac{1}{2}t^2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z-t)^2} dz$$
$$= e^{\frac{1}{2}t^2}$$

This is the PDF of the normal distribution with $\sigma = 1$ and $\mu = t$.

Result: Let X be a random variable with the moment generating function $M_X(t)$. Let $Y = ax + b$. Then

$$M_Y(t) = e^{bt} M_X(at)$$

To establish this result, consider:

$$M_Y(t) = E[e^{yt}]$$
$$= E[e^{t(ax+bt)}]$$
$$= E[e^{axt+bt}]$$
$$= E[e^{axt} \cdot e^{bt}]$$
$$= e^{bt} E[xat]$$
$$= e^{bt} M_X(at)$$

This is a scaling.

Example 3.2.3

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

$$X = \mu + \sigma Z$$

where $Z \sigma N(0,1)$. In order to find $M_X(t)$:

$$M_X(t) = M_{\mu+\sigma z}(t)$$

$$= e^{\mu t} M_Z(\sigma t)$$

Chapter 4

Multivariate distribution

Definition 4.0.1: Multivariate distribution

A multivariate distribution describes the joint behaviour of two or more random variables. That is, the simultaneous occurrence of two distinct events.

Example 4.0.1

Consider tossing two fair coins. Define the first variable y_1 to represent the number of heads for coin 1, with support $\{0, 1\}$. Define the second variable y_2 to represent the number of heads for coin 2, with support $\{0, 1\}$.
The values that both of them take, jointly, can be represented as:

$$\{(0, 0), (1, 0), (0, 1), (1, 1)\}$$

We are looking for $P(Y_1 = y_1 \cap Y_2 = y_2)$, where y_1 and y_2 represent events.
Notice that these can be represented as points on a 2D plane.

Definition 4.0.2: Discrete multivariable probability distribution

Let Y_1 and Y_2 be discrete random variables. The joint (or bivariate) probability function for Y_1 and Y_2 is given by:

$$\begin{aligned} p(y_1, y_2) &= P(Y_1 = y_1 \cap Y_2 = y_2) \\ &= P(Y_1 = y_1, Y_2, = y_2) \\ -\infty < y_1 &< \infty, -\infty < y_1 < \infty \end{aligned}$$

If Y_1 and Y_2 are discrete random variables with joint probability mass function $p(y_1, y_2)$, then:

1. $p(y_1, y_2) \geq 0$ for all y_1, y_2.
2. $\sum_{y_1, y_2} p(y_1, y_2) = 1$

Example 4.0.2

Roll two fair dice and define:

1. Y_1: The number appearing on die 1.
2. Y_2: The number appearing on die 2.

Then $p(y_1, y_2) = \frac{1}{36}$.

Definition 4.0.3: CDF of discrete multivariable probability distribution

$$F_{Y_1,Y_2}(y_1,y_2) = P(Y_1 \leq y_1, Y_2 \leq y_2)$$
$$\sum_{t_1 \leq y_1} \sum_{t_2 \leq y_2} P_{Y_1,Y_2}(t_1,t_2)$$

Definition 4.0.4: Jointly continuous multivariate probability distribution

The random variables Y_1 and Y_2 are jointly continuous if their joint CDF $F_{Y_1,Y_2}(y_1,y_2)$ is continuous in both y_1 and y_2. Then the joint CDF of Y_1, Y_2 is the function such that

$$F_{Y_1,Y_2}(y_1,y_2) = P(Y_1 \leq y_1, Y_2 \leq y_2)$$
$$= \int_{-\infty}^{y_1} \int_{-\infty}^{y_2} f_{Y_1,Y_2}(t_1,t_2) dt_2 dt_1$$

Generally,

$$P(Y_1,Y_2) \in A = \int \int_A f_{Y_1,Y_2}(y_1,y_2) dy_1 dy_2$$

$$P(a_1, \leq Y_1 \leq a_2, b_1 \leq Y_2 \leq b_2)$$
$$= \int_{a_1}^{a_2} \int_{b_1}^{b_2} f_{Y_1,Y_2}(y_1,y_2) dy_2 dy_1$$

Example 4.0.3

Consider the PDF:

$$f(y_1,y_2) = \begin{cases} 1 & 0 \leq y_1 \leq 1, 0 \leq y_2 \leq 1 \\ 0 & \text{otherwise} \end{cases} \tag{4.1}$$

To find $F(0.2, 0.4)$:

$$F_{Y_1,Y_2}(y_1,y_2) = \int_{-\infty}^{y_1} \int_{-\infty}^{y_2} f(y_1,y_2) dy_2 dy_1$$
$$F_{Y_1,Y_2}(0.2, 0.4) = \int_{-\infty}^{0.2} \int_{-\infty}^{0.4} 1 dy_2 dy_1$$
$$= \int_0^{0.2} |y_2|_0^{0.4} dy_1$$
$$= \int_0^{0.2} 0.4 dy_1$$
$$= 0.08$$

Definition 4.0.5: Marginal probability distribution

Let Y_1 and Y_2 be jointly discrete random variables with PMF $p(y_1,y_2)$. Then the marginal probability function of Y_1 and Y_2 are:

$$P_{Y_1}(y_1) = \sum_{\text{all } y_2} P(y_1,y_2)$$

Example 4.0.4

The marginal probability for a dice roll is $P(y_1, 1) + P(y_1, 2)... + P(y_1, 6) = \frac{6}{36}$.

Definition 4.0.6: PMF of marginal probability distributions

Let Y_1, Y_2 be jointly discrete random variables with pmf $p(y_1, y_2)$. Then the marginal pmfs of Y_1 and Y_2 are:

$$P(Y_1 = y_1) = \sum_{y_2} p(y_1, y_2)$$

$$P(Y_2 = y_2) = \sum_{y_1} p(y_1, y_2)$$

Let Y_1, Y_2 be jointly continuous random variables with pdf $f_{Y_1,Y_2}(y_1, y_2)$. Then the marginal pdfs of Y_1 and Y_2 denoted by $f_{Y_1}(y_1)$ and $f_{Y_2}(y_2)$ are given by:

$$f_{Y_1}(y_1) = \int_{-\infty}^{\infty} f_{Y_1,Y_2}(y_1, y_2) dy_2$$

$$f_{Y_2}(y_2) = \int_{-\infty}^{\infty} f_{Y_1,Y_2}(y_1, y_2) dy_1$$

Example 4.0.5

Consider

$$f(y_1, y_2) = \begin{cases} 2y_1 & 0 \le y_1 \le 1, 0 \le y_2 \le 1 \\ 0 & \text{otherwise} \end{cases} \tag{4.2}$$

Find the marginal PDFs.

Example 4.0.6

$f_{Y_1}(y_1) = \int_{-\infty}^{\infty} f(y_1, y_2) dy_2$

$$= \int_0^1 2y_1 dy_2$$

$$= 2y_1 \int_0^1 dy_2$$

$$= 2y_1 |y_2|_0^1$$

$$f_{Y_1}(y_1) = \begin{cases} 2y_1 & 0 \le y_1 \le 1 \\ 0 & \text{otherwise} \end{cases} \tag{4.3}$$

$f_{Y_2}(y_2) = \int_{-\infty}^{\infty} f(y_1, y_2) dy_1$

$$= \int_0^1 2y_1 dy_1$$

$$= |y_1^2|_0^1$$

$$= 1$$

$$f_{Y_2}(y_2) = \begin{cases} 1 & 0 \le y_2 \le 1 \\ 0 & \text{otherwise} \end{cases} \tag{4.4}$$

Definition 4.0.7: Conditional probability distribution

Discrete case: The conditional probability distribution is derived from the joint probability and the marginal probability.

$$P_{Y_1|Y_2=y_2}(y_1|y_2) = \frac{p(y_1, y_2)}{p_{y_2}(y_2)}$$

Continuous case:

$$f_{Y_1|Y_2=y_2}(y_1|y_2) = \frac{f(y_1, y_2)}{f_{y_2}(y_2)}$$

$$f_{Y_2|Y_1=y_1}(y_2|y_1) = \frac{f(y_1, y_2)}{f_{y_1}(y_1)}$$

Example 4.0.7

Consider

$$f(y_1, y_2) = \begin{cases} 2y_1 & 0 \le y_1 \le 1, 0 \le y_2 \le 1 \\ 0 & \text{otherwise} \end{cases} \tag{4.5}$$

Where the maginal probabilities are:

$$f_{Y_1}(y_1) = 2y_1$$
$$0 \le y_1 \le 1$$
$$f_{Y_2}(y_2) = 1$$
$$0 \le y_2 \le 1$$
$$f_{Y_1|Y_2=0.5} = \frac{f(y_2, y_1)}{f_{Y_2}(y_1)}$$
$$= \frac{2y_1}{1}$$
$$= 2y_1$$
$$f_{Y_2|Y_1=0.5} = \frac{f(y_2, y_1)}{f_{Y_1}(y_1)}$$
$$-\frac{2y_1}{2y_1}$$
$$= 1$$

This implies that

$$f(Y_2|Y_1 = y_1) = f(y_2)$$
$$f(Y_1|Y_2 = y_2) = f(y_1)$$

Or the probabilities are independent of each other.

Example 4.0.8

A soft-drink machine has a random amount Y_2 (in gallons) in supply at the beginning of a given day and dispenses a random amount Y_1 during the day, with the condition $Y_1 \le Y_2$. Joint density function:

$$f(y_1, y_2) = \begin{cases} \frac{1}{2} & 0 \le y_1 \le y_2 \le 2 \\ 0 & \text{elsewhere} \end{cases} \tag{4.6}$$

1. Find the conditional density of Y_1 given $Y_2 = y_2$.

2. Evaluate the probability that less than $\frac{1}{2}$ gallons is sold, given that the machine contains 1.5 gallons at the start of the day.

Solution: (1) Have that:

$$f_{Y_1|Y_2=y_2}(y_1|y_2) = \frac{f(y_1, y_2)}{f(y_2)}$$

$$\begin{aligned} f_{Y_2}(y_1) &= \int_0^{y_2} f(y_1, y_2) dy_1 \\ &= \int_0^{y_2} \frac{1}{2} dy_1 \\ &= \frac{1}{2} \int_0^{y_2} dy_1 \\ &= \frac{1}{2} |y_1|_0^{y_2} \\ &= \frac{y_2}{2} \end{aligned}$$

$$\begin{aligned} f_{Y_1|Y_2=y_2}(y_1|y_2) &= \frac{f(y_1, y_2)}{f_{Y_2}(y_2)} \\ &= \frac{1/2}{y_2/2} \\ &= \frac{1}{y_2} \end{aligned}$$

where $0 < y_2 \leq 2$.

(2)

$$\begin{aligned} P(Y_1 < 1/2 | Y_2 = 1.5) &= \int_0^{1/2} f_{Y_1|Y_2=y_2}(y_1|y_2) dy_1 \\ &= \int_0^{1/2} \frac{1}{y_2} dy_1 \\ &= \frac{1}{y_2} |y_1|_0^{1/2} \\ &= \frac{1}{2} \cdot \frac{1}{y_2} \\ &= \frac{1}{2} \cdot \frac{1}{1.5} \\ &= \frac{1}{3} \end{aligned}$$

4.1 Expectation and variance

Definition 4.1.1: Expected value

Let $g(Y_1, Y_2)$ be some function of Y_1 and Y_2, then

$$E(g(Y_1,Y_2)) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(g(Y_1,Y_2))f_{Y_1,Y_2}(y_1,y_2)dy_1dy_2$$

$$E(g(Y_1,Y_2)) = \sum_{\text{all } y_1}\sum_{\text{all } y_2}(g(Y_1,Y_2)P(Y_1 = y_1, Y_2 = y_2))$$

The expected value of $E(Y_1)$ is given by:

$$E(Y_1) = \int_{-\infty}^{\infty} y_i f_{Y_i}(y_i)dy_i$$

The expected value of any power of the random variable by:

$$E(Y_i^k) = \int_{-\infty}^{\infty} y_i^k f_i(y_i)dy_i$$

Example 4.1.1

Let Y_1, Y_2 be joint density function given by

$$f_{Y_1}(y_1) = \begin{cases} 2y_1 & 0 \leq y_1 \leq 1 \\ 0 & \text{otherwise} \end{cases} \tag{4.7}$$

Find $E(Y_1)$ and $E(Y_2)$.

Solution:

$$E(Y_1) = \int_{-\infty}^{\infty} y_1 f(y_1)dy_1$$

We know that $f_{Y_1}(y_1) = 2y_1$, and also $f_{Y_2}(y_2) = 1$. Therefore:

$$E(Y_1) = \int_0^1 y_1 \cdot 2y_1 dy_1$$
$$= 2 \cdot \frac{y_1^3}{3}|_0^1$$
$$= \frac{2}{3}$$

$$E(Y_2) = \int_0^1 y_2 \cdot 1 dy_2$$
$$= |\frac{y_2^2}{2}|_0^1$$
$$= \frac{1}{2}$$

Now to find $Var(Y_1)$ and $Var(Y_2)$.

$$Var(Y_1) = E(Y_1^2) - (E(Y_1))^2$$

$$E(Y_1^2) = \int_{-\infty}^{\infty} y_1^2(2y_1)dy_1$$

$$= |\frac{2y_1^4}{4}|_0^1$$
$$= \frac{1}{2}$$

$$Var(Y_1) = E(Y_1^2) - [E(Y_1)]^2$$
$$= \frac{1}{2} - (\frac{2}{3})^2$$
$$= \frac{1}{2} - \frac{4}{9} = \frac{1}{18}$$

Similarly for $Var(Y_2)$, to get $Var(Y_2) = \frac{1}{2}$.

4.2 Conditional expectation

Definition 4.2.1: Conditional expectations

If Y_1 and Y_2 are any two random variables, the conditional expectation of $g(Y_1)$, given that $Y_2 = y_2$, is defined as follows:

1. If Y_1 and Y_2 are jointly continuous:
$$E(g(Y_1)|Y_2 = y_2) = \int_{-\infty}^{\infty} g(y_1)f(y_1|y_2)dy_1$$
where $f(y_1|y_2)$ is the continuous density function.
2. If Y_1 and Y_2 are jointly discrete:
$$E(g(Y_1)|Y_2 = y_2) = \sum_{y_2} g(y_1)p(y_1|y_2)$$
where $p(y_1|y_2)$ is the probability density function.
3. For $g(Y_1) = Y_1$, the conditional expectation $E(Y_1|Y_2 = y_2)$ is given by:
$$E(Y_1|Y_2 = y_2) = \int_{-\infty}^{\infty} y_1 f(y_1|y_2)dy_1$$
where $f(y_1|y_2)$ is the conditional density function of Y_1 given $Y_2 = y_2$.

Example 4.2.1

Recall the soft-drink example.

$$f(y_1, y_2) = \begin{cases} \frac{1}{2} & 0 \le y_1 \le y_2 \le 2 \\ 0 & \text{elsewhere} \end{cases} \tag{4.8}$$

Find the conditional expectation of the amount of the liquid dispensed given $Y_2 = 1.5$

Solution: Looking for $E(Y_1|Y_2 = 1.5)$:

$$E(Y_1|Y_2 = 1.5) = \int_0^{y_2} y_1 f(y_1|y_2)dy_1$$
$$= \int_0^{y_2} y_1 \cdot \frac{1}{y_2} dy_1$$

$$= \frac{1}{y_2} \int_0^{y_2} y_1 dy_1$$

$$= \frac{1}{y_2} | \frac{y_1^2}{2} |_0^{y_2}$$

$$= \frac{y_2^2}{2y_2}$$

$$= \frac{y_2}{2}$$

$$E(Y_1|Y_2 = 1.5) = \frac{1.5}{2} = 0.75$$

Can also find the variance.

$$Var(Y_1|Y_2 = 1.5) = E(Y_1^2|Y_2 = 1.5) - [E(Y_1|Y_2 = 1.5)]^2$$

Definition 4.2.2: Covariance

$$Cov(Y_1, Y_2) = E[(Y_1 = U_{Y_1})(Y_2 - U_{Y_2})$$
$$= E(Y_1, Y_2) - E(Y_1) \cdot E(Y_2)$$

If Y_1 and Y_2 are independent, then $E(Y_1, Y_2) = E(Y_1) \cdot E(Y_2)$. This would mean that $Cov(Y_1, Y_2) = 0$.

4.3 Independence

Definition 4.3.1: Independence

If Y_1, Y_2 are discrete random variables with joint probability function $p(y_1, y_2)$ and marginal probability functions $p_1(y_1)$ and $p_2(y_2)$, respectively, then Y_1 and Y_2 are independent if and only if:

$$p(y_1, y_2) = p_1(y_1)p_2(y_2)$$

for all pairs of real numbers (y_1, y_2).
Continuous random variable case:

$$f(y_1, y_2) = f_1(y_1) \cdot f_2(y_2)$$

Example 4.3.1

Let Y_1, Y_2 be joint density function given by

$$f_{Y_1}(y_1) = \begin{cases} 2y_1 & 0 \leq y_1 \leq 1 \\ 0 & \text{otherwise} \end{cases} \tag{4.9}$$

We found that $f_{Y_1}(y_1) = 2y_2$, and that $f_{Y_2}(y_2) = 1$. We can see that:

$$f(y_1, y_2) = 2y_1 = f(y_1) \cdot f(y_2)$$

Therefore, Y_1 and Y_2 are independent.

Example 4.3.2

Determine whether X and Y are independent, where:

$$f_{XY}(2,y) = \begin{cases} 3e^{-x-3y} & x,y > 0 \\ 0 & \text{otherwise} \end{cases} \tag{4.10}$$

Solution:

$$\begin{aligned} f_X(x) &= \int_0^\infty f(x, y_d y \\ &= \int_0^\infty 3e^{-x-3y} dy \\ &= 3\int_0^\infty e^{-x}e^{-3y} dy \\ &= 3e^{-x}|\frac{e^{-3y}}{-3}|_0^\infty \\ &= e^{-x}| - e^{-3y}|_0^\infty \\ &= e^{-x}| - \frac{1}{e^{3y}}|_0^\infty \\ &= e^{-x}[-0-(-1))] \\ &= e^{-x} \end{aligned}$$

$$\begin{aligned} f_Y(y) &= \int_0^\infty f(x,y)dx \\ &= \int_0^\infty 3e^{-x-3y} dx \\ &= \int_0^\infty e^{-x}e^{-3y} dx \\ &= 3e^{-3y}\int_0^\infty e^{-x} dx \\ &= 3e^{-3y}|\frac{e^{-x}}{-1}|_0^\infty \\ &= 3e^{-3y} - x - 3y \end{aligned}$$

$$f(x,y) = f_X(x) \cdot f_Y(y)$$

So X and Y are independent.

Example 4.3.3

Determine whether X and Y are independent.

$$f_X(x,y) = \begin{cases} 8xy & 0 < x < y < 1 \\ 0 & \text{otherwise} \end{cases} \tag{4.11}$$

Solution:

$$\begin{aligned} f_X(x) &= \int_x^1 8xydy \\ &= 8x|\frac{y^2}{2}|_x^1 \end{aligned}$$

$$= 4x(1 - x^2)$$

$$f_Y(y) = \int_0^y 8xydx$$
$$= 8y|\frac{x^2}{2}|_0^y$$
$$= 4y \cdot y^2$$
$$= 4y^3$$

$$f_X(x) \cdot f_Y(y) = 4x(1 - x^2) \cdot 4y^3$$
$$= 16x(1 - x^2)y^3$$

which is not equal to the joint probability density function. So X and Y are not independent.

4.4 Central limit theorem

We can use moment-generating functions to identify distributions.

Theorem 4.4.1

Let $m_X(t)$ and $m_Y(t)$ denote moment-generating functions of random variables X and Y, respectively. If both moment-generating functions exist and $m_X(t) = m_Y(t)$, then for all values of t, X and Y have the same probability distribution.

To find a distribution, if U is a function of n random variables, then:

1. Find $m_U(t) = E[e^{tU}]$
2. Compare the expression with the moment-generating function of a known distribution.

Theorem 4.4.2

Let $Y_1, Y_2, ..., Y_n$ be n independent random variables with moment-generating functions $m_{Y_1}(t), m_{Y_2}(t), ..., m_{Y_n}(t)$. If $U = Y_1 + Y_2 + ... + Y_n$, then:

$$m_U(t) = m_{Y_1}(t) \cdot m_{Y_2}(t) \cdots m_{Y_n}(t)$$

Example 4.4.1

We are given that the number of customer arrivals in a given time period follows an Poisson distribution. If Y_1 denotes the time until the first person's arrival, Y_2 the time between the first and second person's arrival, then Y_n would represent the time between the nth and $n-1$th person's arrival. It can be shown that all random variables Y_i are independent random variables with exponential distribution density function:

$$f_{Y_i}(y_i) = \begin{cases} \frac{1}{\theta}e^{-y_i/\theta} & y_i > 0 \\ 0 & \text{otherwise} \end{cases} \tag{4.12}$$

Find the PDF for the waiting time from the opening of the store until the nth customer.

Solution: We see that $E(Y_i) = \theta$, by the exponential distribution. Therefore, to generate our moment-generating function, we can use the moment-generating function:

$$m_{Y_i}(t) = (1 - \theta t)^{-1}$$

Therefore:

$$m_U(t) = (1 - \theta t)^{-1} \cdot (1 - \theta t)^{-1} ... \cdot (1 - \theta t)^{-1}$$

$$= (1-\theta t)^{-n}$$

We recognize this as the moment-generating function of a gamma-distributed random variable with $\alpha = n$ and $\beta = \theta$. So the density function of U is given by:

$$f_U(u) = \begin{cases} \frac{1}{\Gamma(n)\theta^n}(u^{n-1}e^{-u/\theta}) & u > 0 \\ 0 & \text{otherwise} \end{cases} \tag{4.13}$$

Example 4.4.2

Let Y_1 and Y_2 be 2 random variables, each representing the length of life of a component in a system. We have that both Y_1 and Y_2 follow an exponential distribution with mean 1. That is:

$$Y_1 \sim exp(\theta = 1)$$

$$Y_2 \sim exp(\theta = 1)$$

Find the density function for the average life length of these components, $U = \frac{Y_1+Y_2}{2}$.

Solution: The probability density function of both Y_1 and Y_2 can be expressed as $p(Y_i = y) = e^{-y}$. Then our moment-generating functions are also:

$$m_{Y_i}(t) = \frac{1}{1-t}$$

Let $S = Y_1 + Y_2$. Then:

$$m_S(t) = (\frac{1}{1-t})^2$$

We see that this is a gamma distribution with $\alpha = 2, \beta = 1$. Hence, the probability density function over its support is:

$$f_S(s) = \frac{1}{\Gamma(2)1^2}(u^{2-1}e^{-u/1})$$

Using that $\Gamma(2) = (2-1)! = 1$:

$$f_S(s) = se^{-s}$$

Now we transform this to get $U = \frac{Y_1+Y_2}{2}$:

$$U = \frac{S}{2}$$

$$S = g^{-1}(u) = 2U$$

Differentiating our inverse function:

$$\frac{dg^{-1}(u)}{du} = 2$$

So:

$$f_U(U) = f_S(g^{-1}(u)) \cdot \frac{dg^{-1}(u)}{du}$$

$$= f_S(2U) \cdot 2$$

$$= (2u)e^{-2U} \cdot 2$$

$$= 4ue^{-2u}$$

Therefore:

$$f_U(u) = 4ue^{-2u}$$

Theorem 4.4.3

Let $Y_1, Y_2, ..., Y_n$ be independently normally distributed variables with $E_i = \mu_i$ and $V(Y_i) = \sigma_i^2$ for $i =$

$1, 2, \ldots, n$. Let $a_1, a_2, \ldots, a_n$ be coefficients. If

$$U = \sum_{i=1}^{n} a_i Y_i = a_1 Y_1 + a_2 Y_2 + \ldots + a_n Y_n$$

Then:

$$E(U) = \sum_{i=1}^{n} a_i \mu_i$$

$$V(U) = \sum_{i=1}^{n} a_i^2 \sigma_i^2$$

Theorem 4.4.4

Let $Y_1, Y_2, \ldots, Y_n$ be a random sample of size n from a normal distribution with mean μ and variance σ^2. Then

$$\overline{Y} = \frac{1}{n} \sum_{i=1}^{n} Y_i$$

is also normally distributed, with mean $\mu_Y = \mu$ and variance $\sigma^2_{\overline{Y}} = \sigma^2/n$. This is therefore the **sample mean**. In other words, if $S = Y_1 + Y_2 + \ldots + Y_n$, then:

$$E(S) = n\mu$$

$$V(S) = n\sigma^2$$

Theorem 4.4.5 Central Limit Theorem

Let $Y_1, Y_2, \ldots, Y_n$ be independently and **identically distributed** random variables with $E(Y_i) = \mu$ and $V(Y_i) = \sigma^2 < \infty$. Define:

$$U_n = \frac{\overline{Y} - \mu}{\sigma/\sqrt{n}}$$

where $\overline{Y} = \frac{1}{n} \sum_{i=1}^{n} Y_i$. Then the distribution function of U_n converges to the **standard normal distribution** as $n \longrightarrow \infty$. That is:

$$\lim_{n\to\infty} P(U_n \leq u) = \int_{-\infty}^{u} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$

for all u. That is, $U_n \sim N(0, 1)$.

Example 4.4.3

160 people have booked a flight, but there are only 155 seats. The airline expects that 95% of people will show to the flight. Find the probability that all people who show up to the flight will have a seat.

Solution: Let X_i be the random variable representing whether a person shows up. We have that X_i is modelled by the Bernoulli distribution. Then:

$$E(X_i) = 0.95$$

$$V(X_i) = p(1-p) = 0.95 \cdot 0.05 = 0.0475$$

We have $n = 160$. Let $U = X_1 + X_2 + \ldots + X_{160}$ be the random variable representing the number of people who show up to the flight. We find that $E(U) = 160\mu = 160(0.95) = 152$. Also, $V(Y) = 160\sigma^2 = 160(0.0475) = 7.6$. We are looking for $P(U \leq 155)$. To do this, we can normalize our distribution from the normal distribution, by applying our z-score:

$$z = \frac{155 - E(U)}{\sigma(U)} = \frac{155 - 152}{2.76} \approx 1.09$$

Then

$$P(U \leq 155) = P(Z \leq 1.09)$$
$$\approx 0.862$$

Example 4.4.4

Let a student's mean test score be $\mu = 14$. The variance of this test score is $\sigma^2 = 2^2$. We pick a random sample of 100 students. Find the range of scores between which 95% of students' scores are expected to lie.

Solution: First, we find the z-values between which 95% of values lie. This is $P(-1.96 < z < 1.96) = 0.95$. Transform both of these to the values we want $\overline{Y} = a, b$ such that:

$$-1.96 = \frac{a - \mu}{\sigma/\sqrt{n}}$$

$$1.96 = \frac{b - \mu}{\sigma/\sqrt{n}}$$

We see that $\mu = 14$, $\sigma = 2$, $\sqrt{n} = 10$.

$$-1.96 = \frac{a - 14}{2/10}$$
$$a \approx 13.608$$
$$1.96 = \frac{b - 14}{2/10}$$
$$b \approx 14.392$$