Foundations of Probability

Last Updated:

April 30, 2024

Contents

Chapter 1

Probability Measures

1.1 Review of Set Theory

Let Ω be an abstract set representing the sample space of a random experiment. The power set of Ω by $\mathcal{P}(\Omega)$ is defined to be the set of all subsets of Ω. Elements of Ω are outcomes and its subsets are events. Therefore,

$$\mathcal{P}(\Omega) = \{A : A \subseteq \Omega\}$$

For $A, B \in \mathcal{P}(\Omega)$, we define

$$A \cup B = \{x : x \in A \text{ or } x \in B\}$$
$$A \cap B = \{x : x \in A \text{ and } x \in B\}$$
$$\bar{A} = A^c = \{x : x \notin A\}$$
$$A \Delta B = (A \cup B) \setminus (A \cap B)$$

In terms of events $A \cup B$ occurs if and only if at least one of the two events A and B occurs. Also, $A \cap B$ occurs if both A and B occurs. The empty set is denoted by $\emptyset$.

Examples of Sample Spaces: When flipping a coin, we have two outcomes, so

$$\Omega = \{H, T\}$$

If we flip a coin and role a dice,

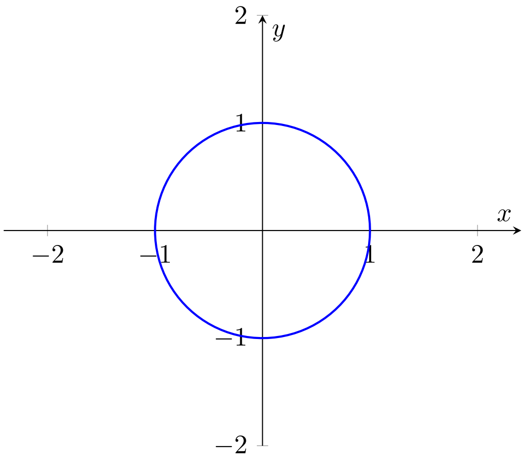
$$\Omega = \{1H, 2H, \ldots, 6H, 1T, \ldots 6T\}$$

If we flip a coin until we observe a head,

$$\Omega = \{H, TH, TTH, TTTH, \ldots\}$$

Here, there are infinite outcomes so the sample space is infinite, but it is countable since we can list all the possibilities.

If we pick a choose are sample space to be the points with distance one from the origin, we have the points in the unit circle,



The sample space is defined by

$$\Omega = \{(x, y) : d((x, y), (0, 0)) \leq 1\} = \{(x, y) : x^2 + y^2 \leq 1\}$$

In this example, the sample space omega is infinite as well, but it is uncountable.

Examples of Events: An event is a subset of the sample space. For example, in the case of rolling a dice and flipping a coin, we have

$$\Omega = \{1H, 2H, \dots, 6H, 1T, \dots 6T\}$$

And we can define an event E as

$$E = \{\text{Coin is heads and the dice is even}\} = \{2H, 4H, 6H\} \subset \Omega$$

If we flip a coin until the first head appears,

$$\Omega = \{H, TH, TTH, \dots\}$$

And we can define an event E as

$$E = \{\text{First head appears before the 5th trial}\} = \{H, TH, TTH, TTTH, TTTTH\} \subset \Omega$$

Examples of Power Sets: Consider the sample space obtained by rolling a dice,

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

And let $E = \{2, 4, 6\}$ be the event we roll an even number. Then, the power set of E is

$$\mathcal{P}(E) = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}\}$$

The cardinality of the power set is

$$|\mathcal{P}(E)| = 2^{|E|} = 8$$

Examples of Set Operations:

$$\Omega = \{1, 2, \dots, 6\}$$
$$A = \{1, 2, 3\}\ B = \{1, 2, 3\}$$
$$A \cup B = \{1, 2, 3, 4, 6\}$$
$$A \cap B = \{2\}$$
$$A^c = \{1, 3, 5\}$$
$$A \setminus B = \{x : x \in A, x \notin B\} = \{4, 6\}$$
$$A \Delta B = (A \cup B) \setminus (A \cap B)$$

Example of Empty Set: Is $\{\{\}\} = \{\}$? **No**. $\{\{\}\}$ is a set with one element, which is the empty set, so $\{\{\}\} = \{\emptyset\}$.

1.1.1 Properties of Sets

- $A \subset A$, $\emptyset \subset A$
- $A \subset B$ and $B \subset A$ implies $A = B$
- $A \subset C$ and $B \subset C$ implies $A \cup B \subset C$ and $A \cap B \subset C$.
- $A \subset B$ if and only if $B^c \subset A^c$
- $(A^c)^c = A$, ${}^c = \Omega$, $\Omega^c = \emptyset$
- $A \cup B = B \cup A$, $A \cap B = B \cap A$
- $A \cup A = A$, $A \cap \Omega = A$, $A \cup A^c = \Omega$, $A \cap A^c = \emptyset$.
- $A \cap (B \cup C) = (A \cap B) \cup (A \cap C) \cup (A \cap C)$
- $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

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

Example. We have

$$\bigcup_{n=1}^{\infty} \left[0, \frac{n}{n+1}\right) = [0, 1)$$

To show that these sets are equal, consider the limit of the sequence $\left(\frac{n}{n+1}\right)_{n=1}^{\infty}$. As n becomes large, the limit approaches 1. So, the union of all these sets will contain elements that become arbitrarly close to 1 but do not reach 1, so we have $[0, 1)$.

$$\bigcap_{n=1}^{\infty} \left(0, \frac{1}{n}\right) = \emptyset$$

To show that these sets are equal, consider the sequence $\left(\frac{1}{n}\right)_{n=1}^{\infty}$, As this sequence approaches 0, the intersection of all these sets will contain elements that become arbitrarly close to 0 but do not reach 0, so we have the set $(0, 0)$ which is empty. Therefore, when taking the intersection of all these sents with $\left(0, \frac{1}{n}\right)$ which become arbitrarly small, we have the empty set.

Example. Prove that $A\Delta B = A^c \Delta B^c$.

Proof. Note that $A \setminus B = A \cap B^c$.

$$\begin{aligned}
A^c \Delta B^c &= (A^c \cup B^c) \setminus (A^c \cap B^c) \\
&= (A \cap B)^c \cap ((A \cup B)^c)^c \\
&= (A \cap B)^c \cap (A \cup B) = (A \cup B) \setminus (A \cap B) \\
&= A\Delta B
\end{aligned}$$

□

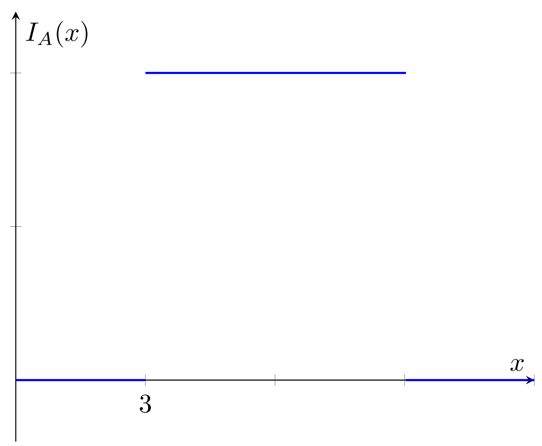
1.2 Indicator Function

Let $A \subset \Omega$. The indicator function of A is defined as

$$I(x \in A) = I_A(x) = \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases}$$

Example. $A = [1, 3]$

$$I_A(x) = I_{[1,3]}(x)$$



1.2.1 Properties of Indicator Functions:

- $I_{A \cup B} = \max(I_A, I_B)$
- $I_{A \cap B} = I_A \cdot I_B$
- $I_{A \Delta B} = I_A + I_B \pmod 2$
- $A \subset B$ if and only if $I_A \leq I_B$
- $I_{\cup_i A_i} \leq \sum_i I_{A_i}$

1.3 Set Theoretic Limits

Exercise: Prove that

$$I_{\cup_{i=1}^{\infty} A_i} = 1 - \prod_{i=1}^{\infty}(1 - I_{A_i})$$

and

$$I_{A \Delta B} = (I_A - I_B)^2$$

Definition 1.3.1. *Let $\{A_n\}$ be a sequence of events. Then*

$$\liminf A_n = \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} A_m$$

and

$$\limsup A_n = \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m$$

Example. Suppose we have a sequence of events $A_1, A_2, A_3, \ldots$, we can define

$$B_n = \bigcap_{m=n}^{\infty} A_m$$

So its sequence members are

$$B_n = A_n \cap A_{n+1} \cap \cdots$$
$$B_{n+1} = A_{n+1} \cap A_{n+2} \cap \cdots$$
$$B_{n+2} = A_{n+2} \cap A_{n+3} \cap \cdots$$

These sets are getting smaller because surely the intersection of more sets will be smaller since $|A \cap B| \leq \min(|A|, |B|)$. So as we take more intersections, the sets become smaller and smaller. Now we can look at the union of these sets,

$$\bigcup_{n=1}^{\infty} B_n = \liminf A_n$$

If instead we take B_n to be the union of all the sets,

$$B_n = \bigcup_{m=n}^{\infty} A_m$$

So the sequence members are

$$B_n = A_n \cup A_{n+1} \cup \cdots$$
$$B_{n+1} = A_{n+1} \cup A_{n+2} \cup \cdots$$
$$B_{n+2} = A_{n+2} \cup A_{n+3} \cup \cdots$$

These sets are getting larger since we are taking the union of more sets, then we can look at the intersection of these sets,

$$\bigcap_{m=n}^{\infty} = \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_n = \limsup A_n$$

Lemma 1.3.1. *We have*

$$\limsup A_n = \left\{ \omega : \sum_{i=1}^{\infty} I_{A_i}(\omega) = \infty \right\}$$

and

$$\liminf A_n = \left\{ \omega : \sum_{i=1}^{\infty} I_{A_i^c}(\omega) < \infty \right\}$$

We can also express this as

$$\limsup_{n\to\infty} A_n = \left\{ x \in X : \limsup_{n\to\infty} I_{A_n}(x) = 1 \right\}$$

and

$$\liminf_{n\to\infty} A_n = \left\{ x \in X : \liminf_{n\to\infty} I_{A_n}(x) = 1 \right\}$$

Proof. If $\omega \in \limsup A_n$, then $\omega \in \bigcup_{m=n}^{\infty} A_m$ for all integers n. Therefore, for any integer n there exists an integer k_n such that $\omega \in A_{k_n}$, since

$$\sum_{i=1}^{\infty} A_{A_i}(\omega) \geq \sum_{i=1}^{\infty} I_{A_{k_i}}(\omega) = \infty$$

Conversely, for any integer n, by definition of the limit superior,

$$\sum_{i=n}^{\infty} I_{A_i}(\omega) = \infty$$

This implies that $\omega \in \bigcup_{j=n}^{\infty} A_j$ for all integers n. Therefore,

$$\omega \in \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m = \limsup A_n$$

Then, we can notice that

$$\omega \in \liminf A_n = \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} A_m$$

implies that there exists an integer n_0 such that

$$\omega \in \bigcap_{k=n_0}^{\infty} A_k$$

Therefore,

$$\sum_{n=1}^{\infty} I_{A_n^c}(\omega) = \sum_{n=1}^{n_0-1} I_{A_n^c}(\omega) \leq n_0 < \infty$$

□

Note: For this reason, sometimes we write $\limsup A_n = A_n$ infinitely often. If $\liminf A_n = \limsup A_n$, then

$$\lim A_n = \liminf A_n = \limsup A_n$$

Remark: The proof of the lemma above can be simplified by noticing the fact that

$$(\limsup A_n)^c = \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} A_m^c = \liminf A_n^c$$

$$(\liminf A_n)^c = \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m^c = \limsup A_n^c$$

To summarize, with the limit superior we have infinitely many cases where $I_{A_i}(\omega) = 1$. So what it means for $\omega \in \limsup A_n$ is that ω is in infinitely many of the A_i's. For the limit inferior, it means that ω is in all but finitely many of the A_i's.

Example. Consider $\omega \in A_i$ when i is odd, so

$$\omega \in A_1, \omega \notin A_2, \omega \in A_3, \omega \notin A_4, \omega \in A_5, \ldots$$

ω is in infinitely many (but countable) number of the A_i's. So $\omega \in \limsup A_n$. Now consider $\omega \in A_i$ when $i \geq 10$, so

$$\omega \notin A_1, \omega \notin A_2, \omega \notin A_3, \ldots, \omega \notin A_9, \omega \in A_{10}, \omega \in A_{11}, \omega \in A_{12}, \ldots$$

So, ω is not in finitely many of the A_i's, therefore $\omega \in \liminf A_n$.

Example. Consider sample space of flipping a coin and infinite number of times, and the event $E = \{HTTHT\}$, so the event that we get $HTTHT$ in that order. Because this outcome is possible, it will occur an infinite number of times in the sequence of events, so E is in the limit superior, and the probability that E occurs infinitely often is 1.

Lemma 1.3.2. *Let $\{A_n\}$ be a sequence of events, then*

1. *If $A_n \subset A_{n+1}$ for any integer n, then*

$$\lim A_n = \bigcup_{n=1}^{\infty} A_n$$

2. *If $A_{n+1} \subset A_n$ for any integer n, then*

$$\lim A_n = \bigcap_{n=1}^{\infty} A_n$$

Proof. We can prove (1) and (2) similarly as follows, not that in this case,

$$\bigcap_{m=n}^{\infty} A_m = \bigcap_{m=1}^{\infty} A_m$$

for all integers n. If $A_n \subset A_{n+1}$ for any integer n, then we have that

$$A_1 \subset A_2 \subset A_3 \subset \cdots$$

So the set is getting bigger, now consider the limit superior

$$\limsup_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m = \left(\bigcup_{m=1}^{\infty} A_m\right) \cap \left(\bigcup_{m=2}^{\infty} A_m\right) \cap \left(\bigcup_{m=3}^{\infty} A_m\right) \cap \cdots$$

These sets are equal in size since if $A_1 \subset A_2$, then $A_1 \cup A_2 = A_2$. Therefore, we get that the intersection of these sets is $\bigcup_{m=1}^{\infty} A_m$, and thus

$$\limsup_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m = \bigcup_{n=1}^{\infty} A_n$$

Furthermore,

$$\begin{aligned}
\liminf_{n\to\infty} A_n &= \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m \\
&= \left(\bigcap_{m=1}^{\infty} A_m\right) \cup \left(\bigcap_{m=2}^{\infty} A_m\right) \cup \left(\bigcap_{m=3}^{\infty} A_m\right) \cup \cdots \\
&= A_1 \cup A_2 \cup A_3 \cup \cdots \\
&= \bigcup_{n=1}^{\infty} A_n
\end{aligned}$$

Therefore

$$\limsup A_n = \liminf A_n \implies \limsup A_n = \liminf A_n = \lim A_n$$

The proof for (2) follows the same. ☐

Example.

$$\lim_{n\to\infty} \left[0, 1 - \frac{1}{n}\right] = \lim_{n\to\infty} \left[0, 1 - \frac{1}{n}\right) = [0, 1)$$

To see this, we have that

$$A_1 = \{0\}, A_2 = \left[0, \frac{1}{2}\right], A_3 = \left[0, \frac{2}{3}\right], A_4 = \left[0, \frac{3}{4}\right], \ldots$$

So the set A_n is increasing, therefore

$$\limsup A_n = \liminf A_n = \lim A_n = \bigcup_{n=1}^{\infty} A_n = [0, 1)$$

Example.

$$\lim_{n\to\infty} \left[0, 1 + \frac{1}{n}\right] = \lim_{n\to\infty} \left[0, 1 + \frac{1}{n}\right) = [0, 1]$$

Similarly,

$$A_1 = [0,2], A_2 = \left[0, 1+\frac{1}{2}\right], A_3 = \left[0, 1+\frac{1}{3}\right], A_4 = \left[0, 1+\frac{1}{4}\right], \dots$$

The set A_n is decreasing, therefore

$$\limsup A_n = \liminf A_n = \bigcap_{n=1}^{\infty} A_n = \bigcap_{n=1}^{\infty} \left[0, 1+\frac{1}{n}\right] = [0,1]$$

Example. Let $B, C \subset \Omega$ and define the sequence

$$A_n = \begin{cases} B & \text{if } n \text{ is odd} \\ C & \text{if } n \text{ is even} \end{cases}$$

Then we have

$$\bigcup_{m=n}^{\infty} = B \cup C \implies \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m = B \cup C = \limsup A_n$$

Similarly for the limit inferior,

$$\bigcap_{m=n}^{\infty} A_m = B \cap C \implies \bigcup_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m = B \cup C = \liminf A_n$$

Therefore, we have

$$\bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m = B \cup C \text{ and } \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} A_m = B \cap C$$

If $B \cap C \neq B \cup C$, then $B \cap C = \liminf A_n \neq \limsup A_n = B \cup C$.

1.4 Fields and Algebras

Definition 1.4.1 (Fields (Algebras))**.** *A* field *(or* algebra*) is a class of subsets of Ω (called events) that contain Ω and are closed under finite union, finite intersection, and complementation. In otherwords, a family of subsets of Ω (say $\mathcal{A}$) is a field if*

- $\Omega \in \mathcal{A}$
- *If $A \in \mathcal{A}$, then $A^c \in \mathcal{A}$*
- *If $A, B \in \mathcal{A}$, then $A \cup B \in \mathcal{A}$*

Remarks. If $A, B \in \mathcal{A}$ then $A \cap B \in \mathcal{A}$. This is true because

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

Definition 1.4.2 (σ-field)**.** *A σ-field (or σ-algebra) is a field that is closed under countable union (which implies that it is closed under countable intersection).*

Example. Let Ω be a set and $A, B \subset \Omega$. Then,

$$\mathcal{A} = \{\Omega, \emptyset, A, A^c, B, B^c, A \cup B, A \cap B, A \cap B^c, A^c \cap B, A^c \cup B^c, A^c \cap B^c\}$$

Examples of σ-fields.

- The power set $\mathcal{P}(\Omega)$
- $\mathcal{F} = \{\Omega, \emptyset\}$
- The family of subsets or $\mathbb{R}$ which are either countable or their complements are countable.
- Let $\mathcal{B}$ be the smallest σ-field containing all open sets. Then $\mathcal{B}$ is called the Borel σ-field.

Definition 1.4.3 (Probability Measure)**.** *Let Ω be a sample space and $\mathcal{F}$ be a σ-field on Ω. A probability measure P is defined on $\mathcal{F}$ such that*

(i) $P(\Omega) = 1$

(ii) If $A_1, A_2, \dots \in \mathcal{F}$ are disjoint, then

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

1.4.1 Properties of Probability Measures

(i) Since $P(\Omega) = 1 = P(\Omega \cup \emptyset) = P(\Omega) + P(\emptyset)$, we have $P(\emptyset) = 0$

(ii) Since $(A \setminus B) \cup (A \cap B) = A$ and $(A \setminus B) \cap (A \cap B) = \emptyset$, we have

$$P(A \setminus B) = P(A) - P(A \cap B)$$

(iii) Similarly, $(A \setminus B) \cup B = A \cup B$ and $(A \setminus B) \cap B = \emptyset$, which implies

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

(iv) If $A \subset B$ then $A \cup (B \setminus A) = B$. Therefore,

$$P(A) + P(B \setminus A) = P(B)$$

and furthermore, $P(A) \leq P(B)$

1.4.2 Expectations

Definition 1.4.4. *Let $X : \Omega \to \mathbb{R}$, an expecation E be an operator with the following properties,*

(i) If $X \geq 0$ then $E(X) \geq 0$

(ii) If $c \in \mathbb{R}$ is a constant, then $E(cX) = cE(X)$

(iii) $E(X_1 + X_2) = E(X_1) + E(X_2)$

(iv) $E(1) = 1$

(v) If $X_n(\omega)$ is monotonically increasing and $X_n(\omega) \to X(\omega)$, then

$$\lim_{n\to\infty} E(X_n) = E(X)$$

Example. Flip 2 coins,

$$\Omega = \{HH, HT, TH, TT\}$$

Define $X : \Omega \to \mathbb{R}$, with

$$X(HH) = 2, X(HT) = X(TH) = 1, X(TT) = 0$$

So X is a random variable which represents the number of heads.

Example. Flip a coin until a head appears,

$$\Omega = \{H, TH, TTH, TTTH, \ldots\}$$

With

$$X(H) = 1, X(TH) = 2, X(TTH) = 3, X(TTTH) = 4, \ldots$$

So X is a random variable which represents the number of trials until a head appears.

Example. Define

$$\Omega = \{(x, y) : x^2 + y^2 \leq 1\}$$

and

$$X(x, y) = \sqrt{x^2 + y^2} = \text{ Distance of } (x, y) \text{ from } (0, 0)$$

Example. Let X be a random variable, and

$$E(X) = \lim_{D\to\infty} \frac{\int_{-D}^{D} X(\omega) d\omega}{2D}$$

Check if E satisfies the definition for an expectation.

Solution. We have to check the 5 axioms,

1. Its clear that if $X \geq 0$ then $E(X) \geq 0$ since the integral of a non-negative function is non-negative.

2.
$$E(cX) = \lim_{D\to\infty} \frac{\int_{-D}^{D} cX(\omega)d\omega}{2D} = c\lim_{D\to\infty} \frac{\int_{-D}^{D} X(\omega)d\omega}{2D} = cE(x)$$

3.
$$\begin{aligned}
E(X_1+X_2) &= \lim_{D\to\infty} \frac{\int_{-D}^{D}(X_1(\omega)+X_2(\omega))d\omega}{2D} \\
&= \lim_{D\to\infty}\left(\frac{\int_{-D}^{D} X_1(\omega)d\omega}{2D} + \frac{\int_{-D}^{D} X_2(\omega)d\omega}{2D}\right) \\
&= \lim_{D\to\infty}\frac{\int_{-D}^{D} X_1(\omega)d\omega}{2D} + \lim_{D\to\infty}\frac{\int_{-D}^{D} X_2(\omega)d\omega}{2D} \\
&= E(X_1)+E(X_2)
\end{aligned}$$

4.
$$E(1) = \lim_{D\to\infty}\frac{\int_{-D}^{D} 1d\omega}{2D} = \lim_{D\to\infty}\frac{2D}{2D} = 1$$

5. The 5th axiom fails however. Take $\Omega = \mathbb{R}$ and $X_n(\omega) = I_{[-n,-n]}(\omega)$, then
$$\lim_{D\to\infty}\frac{\int_{-D}^{D} X(\omega)d\omega}{2D} = 0$$
But, $x_n(\omega) \to 1$. So the operator in this example is not a proper form of expectation.

1.5 Finding Probabilities Using Expectations

Definition 1.5.1. *For any event A, define*
$$P(A) = E(I_A(\omega))$$
For simplicity, we somtimes drop ω and write
$$P(A) = E(I_A)$$

1.5.1 Properties

1. $E\left(\sum_{i=1}^{n} c_i X_i\right) = \sum_{i=1}^{n} c_i E(X_i)$
2. If $X \le Y \le Z$, then $E(X) \le E(Y) \le E(Z)$
3. If $\{A_i\}$ is a sequence of events, then
$$P\left(\bigcup_{i=1}^{\infty} A_i\right) \le \sum_{i=1}^{\infty} P(A_i)$$

The third property (known as Borel's inequality) is an important result and can be proved as follows.

Proof. Using the fact that
$$I_{\cup A_i} \le \sum_{i=1}^{\infty} I_{A_i}$$
and the second property where if $X \le Y$, then $E(X) \le E(Y)$, then
$$E\left(I_{\cup A_i}\right) \le \sum_{i=1}^{\infty} E(I_{A_i})$$
Then this is exactly
$$P\left(\bigcup_{i=1}^{\infty} A_i\right) \le \sum_{i=1}^{\infty} P(A_i)$$
□

We can use these properties to prove the axioms of probability measures.

1. $P(\Omega) = E(I_\Omega)$, we have $I_\Omega(\omega) = 1\ \forall \omega \in \Omega$, so $P(\Omega) = E(I_\Omega) = 1$
2. We want to show that the probability of disjoint events is the sum of their probablities, so we have that

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

   We know from indicator functions that

$$I_{\cup A_i} = \sum_{i=1}^{\infty} I_{A_i}$$

   Therefore,

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

   as required.

Another useful result we can prove is that

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

and furthermore

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

Proof. We know that

$$I_{A \cup B \cup C} = 1 - (1 - I_A)(1 - I_B)(1 - I_C)$$

This can be easily shown by considering the cases when we have $\omega \in A \cup B \cup C$, then at least one of the terms $(1 - I_A)$, $(1 - I_B)$, or $(1 - I_c)$ will be zero, then

$$1 - (1 - I_A)(1 - I_B)(1 - I_C) = 1 - 0 = 1 = I_{A \cup B \cup C}$$

And if $\omega \notin A \cup B \cup C$, then $\omega \notin A$ and $\omega \notin B$ and $\omega \notin C$, so all their indicator functions will be zero, then we get

$$1 - (1 - 0)(1 - 0)(1 - 0) = 0 = I_{A \cup B \cup C}$$

So, we have the result that

$$I_{\cup A_i} = 1 - \prod_{i=1}^{\infty} (1 - A_i)$$

Now if we expand the equality, we get

$$\begin{aligned}
I_{A \cup B \cup C} &= 1 - (1 - I_A)(1 - I_B)(1 - I_C) \\
&= 1 - (1 - I_A - I_B + I_A I_B)(1 - I_C) \\
&= 1 - 1 + I_A + I_B + I_C - I_A I_B - I_A I_C - I_B I_C + I_A I_B I_C \\
&= I_A + I_B + I_C - I_A I_B - I_A I_C - I_B I_C + I_A I_B I_C
\end{aligned}$$

Then, we can take the expected value of both sides,

$$\begin{aligned}
E(I_{A \cup B \cup C}) &= E(I_A + I_B + I_C - I_A I_B - I_A I_C - I_B I_C + I_A I_B I_C) \\
&= E(I_A) + E(I_B) + E(I_C) - E(I_A I_B) - E(I_A I_C) - E(I_B I_C) + E(I_A I_B I_C) \\
&= P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C) \\
&= P(A \cup B \cup C)
\end{aligned}$$

Therefore,

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

Showing that $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ follows the same. $\square$

Then we get that the general case for the union of n events as

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) - \sum_{1\leq i<j\leq n} P(A_i \cap A_j) + \sum_{i\leq i<j<k\leq n} P(A_i \cap A_j \cap A_k) + (-1)^{n+1} P(A_1 \cap A_2 \cap \cdots \cap A_n)$$

And this is again proved using the fact that

$$I_{\cup_{i=1}^{n} A_i} = 1 - \prod_{i=1}^{n}(1 - A_i)$$

Example. (Confused Secretary Problem) Suppose we have 100 distinct letters to be sent to 100 different people. The Secretary confuses the addresses, and sends 100 letters at random to these 100 people. What is the probability that at least one letter is sent to the correct address?

Solution. Let A_i denote the event that the ith letter goes to the right person. We want to find the probability that one of these events occurs, so $P\left(\bigcup_{i=1}^{100} A_i\right)$. From the formulas previously discussed,

$$\begin{aligned}
P\left(\bigcup_{i=1}^{100} A_i\right) &= P(A_1) + \cdots P(A_100) - P(A_1 \cap A_2) - \cdots - P(A_{99} \cap A_{100}) \\
&+ P(A_1 \cap A_2 \cap A_3) + \cdots + P(A_{98} \cap A_{99} \cap A_{100}) \\
&+ P(A_1 \cap A_2 \cap A_3 \cap A_4) - \cdots - P(A_{97} \cap A_{98} \cap A_{99} \cap A_{100}) \\
&+ \cdots - P(A_1 \cap A_2 \cap \cdots \cap A_{100})
\end{aligned}$$

Then we can evaluate each probability,

$$P(A_1) = \frac{1}{100},\ P(A_2) = \frac{1}{100}, \ldots, P(A_{100}) = \frac{1}{100}$$
$$P(A_1 \cap A_2) = \frac{1}{100} \cdot \frac{1}{99} = \cdots = P(A_{99} \cap A_{100})$$
$$P(A_1 \cap A_2 \cap A_3) = \frac{1}{100} \cdot \frac{1}{99} \cdot \frac{1}{98} = \cdots$$

Then continuing this sequence we have

$$\begin{aligned}
P\left(\bigcup_{i=1}^{100}\right) &= 100\left(\frac{1}{100}\right) - \binom{100}{2}\left(\frac{1}{100 \cdot 99}\right) + \cdots - \cdots - \binom{100}{100}\left(\frac{1}{100 \cdot 99 \cdots 1}\right) \\
&= 1 - \frac{100 \cdot 99}{2!} \cdot \frac{1}{100 \cdot 99} + \frac{100 \cdot 99 \cdot 98}{3!} \cdot \frac{1}{100 \cdot 99 \cdot 98} - \cdots - \frac{1}{100!} \\
&= 1 - \frac{1}{2!} + \frac{1}{3!} - \frac{1}{4!} + \frac{1}{5!} - \cdots - \frac{1}{100!}
\end{aligned}$$

Now recall that e^x can be written as the infinite series

$$e^x = \sum_{n=1}^{\infty} \frac{x^n}{n!}$$

So,

$$e^{-1} = 1 - 1 + \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} - \cdots$$
$$1 - e^{-1} = 1 - \frac{1}{2!} + \frac{1}{3!} - \frac{1}{4!} + \cdots$$

Now since 100! is incredibly large, it is sufficiently large for this infinite sum. So we get that

$$P(\text{At least one letter goes to the right address}) = 1 - \frac{1}{e}$$

Lemma 1.5.1 (Fatou's Lemma)**.** *If $\{A_n\}$ is a family of events, then*

1. $P(\liminf A_n) \leq \liminf P(A_n) \leq \limsup P(A_n) \leq P(\limsup A_n)$

2. *If* $\lim A_n = A$, *then* $\lim P(A_n) = P(A)$

To prove this lemma, we first need to prove the following lemma.

Lemma 1.5.2. *If* $A_n \subset A_{n+1}$ *for any* $n \in \mathbb{N}$, *then* $\lim P(A_n) = P(A)$. *Similarly, if* $A_{n+1} \subset A_n$ *for any* $n \in \mathbb{N}$, *then* $\lim P(A_n) = P(A)$ *where in both cases* $\lim A_n = A$.

Proof. If A_n is increasing (i.e $A_n \subset A_{n+1}$), start by defining $B_n = A_n \setminus A_{n-1}$ with $B_1 = A_1$. Then,

$$\bigcup_{i=1}^{n} B_i = \bigcup_{i=1}^{n} A_i = A_n$$

The point of this is now we have that the B_i's are disjoint. Now if we take the limit we get

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

Taking the probability now,

$$P(A) = P\left(\bigcup_{i=1}^{\infty} B_i\right) = \sum_{i=1}^{\infty} P(B_i) = \lim_{n\to\infty} \sum_{i=1}^{n} P(B_i) = \lim_{n\to\infty} P(A_n)$$

If A_n is decreasing ($A_{n+1} \subset A_n$), then A_n^c is increasing. So,

$$\lim_{n\to\infty} P(A_n^c) = \lim(1 - P(A_n)) = P(A^c) = 1 - P(A)$$

Therefore,

$$\lim_{n\to\infty} P(A_n) = A$$

□

To summarize, when A_n is increasing, we have

$$\lim P(A_n) = P\left(\bigcup_{i=1}^{\infty} A_i\right) = P\left(\lim_{n\to\infty} A_n\right)$$

when A_n is decreasing,

$$\lim P(A_n) = P\left(\bigcap_{i=1}^{\infty} A_i\right) = P\left(\lim_{n\to\infty} A_n\right)$$

Now we can prove Fatou's Lemma.

Proof. To prove (1), notice that from the first part of lemma 1.5.2, we can write

$$P(\liminf A_n) = P\left(\lim_{n\to\infty} \cap_{i=n}^{\infty} A_i\right) = \lim_{n\to\infty} P\left(\cap_{i=n}^{\infty} A_i\right) \leq \liminf P(A_n)$$

since $\bigcap_{i=n}^{\infty} A_i \subset A_n$. Likewise,

$$P(\limsup A_n) = P\left(\lim_{n\to\infty} \bigcup_{i=n}^{\infty} A_i\right) = \lim_{n\to\infty} P\left(\bigcup_{i=1}^{\infty} A_i\right) \geq \limsup P(A_n)$$

since $A_n \subset \bigcup_{i=n}^{\infty} A_i$. For part (2), notice that if

$$A = \limsup A_n = \liminf A_n$$

we have

$$P(A) = P(\liminf A_n) \leq \liminf P(A_n) \leq \limsup P(A_n) \leq P(\limsup A_n) = P(A)$$

Then this implies that

$$\lim_{n\to\infty} P(A_n) = P(A)$$

□

1.6 Independence

Definition 1.6.1. *Let A and B be events. We say that A and B are independent if*

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

Example. Flip a coin and roll a die,

$$\Omega = \{1H, 2H, \ldots, 6H, 1T, 2T, \ldots, 6T\}$$

Let $A = \{1H, 1T\}$ and $B = \{1H, 2H, 3H, 4H, 5H, 6H\}$. Are A, B independent?

$$A \cap B = \{1H\} \neq \emptyset$$

$$P(A \cap B) = \frac{1}{12}$$

$$P(A) = \frac{1}{6},\ P(B) = \frac{1}{2}$$

$$P(A)P(B) = \frac{1}{6} \cdot \frac{1}{2} = \frac{1}{12} = P(A \cap B)$$

Therefore, A and B are independent.

1.6.1 Properties of Independence

1. If A and B are independent, then A and B^c are independent, A^c and B are independent, and A^c and B^c are independent.
2. If A, B, C are indepdent then A and $B \cup C$ are independent. similarly A and $B \cap C$ are independent.
3. An event A is independent of itself if and only if $A = \emptyset$ or $A = \Omega$.
4. Any event A is independent of Ω.

Proof. 1. To prove that A^c and B are independent, recall that

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

So,

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

The proof for A and B^c follows the same. To prove A^c and B^c are independent,

$$\begin{aligned} P(A^c \cap B^c) &= P((A \cup B)^c) \\ &= 1 - P(A \cup B) \\ &= 1 - [P(A) + P(B) - P(A \cap B)] \\ &= 1 - P(A) - P(B) + P(A)P(B) \\ &= 1 - P(A) - P(B)(1 - P(A)) \\ &= (1 - P(A))(1 - P(B)) \\ &= P(A^c)P(B^c) \end{aligned}$$

2. Given 3 independent events A, B, C, then any operations between the sets is independent. So we can show A and $B \cup C$ is independent since

$$\begin{aligned} P(A \cap (B \cup C)) &= P((A \cap B) \cup (A \cap C)) \\ &= P(A \cap B) + P(A \cap C) - P(A \cap B \cap A \cap C) \\ &= P(A)P(B) + P(A)P(B) - P(A)P(B)P(C) \\ &= P(A)(P(B) + P(C) - P(B)P(C)) \\ &= P(A)P(B \cup C) \end{aligned}$$

Similarly for A and $B \cap C$,

$$\begin{aligned} P(A \cap (B \cap C)) &= P((A \cap B) \cap (A \cap C)) \\ &= P(A \cap B \cap A \cap C) \\ &= P(A)P(B)P(C) \\ &= P(A)P(B \cap C) \end{aligned}$$

3. ($\Longrightarrow$) If A is independent of itself, then

$$P(A \cap A) = P(A)P(A) \implies P(A) = P(A)^2 \implies P(A) = 0, 1 \implies A = \emptyset, A = \Omega$$

($\Longleftarrow$) If $A = \emptyset$, then

$$P(A \cap A) = P(A) = 0 = P(\emptyset)P(\emptyset)$$

If $A = \Omega$, then

$$P(A \cap A) = P(A) = 1 = P(\Omega)P(\Omega)$$

4. Every event is independent of Ω since

$$P(A \cap \Omega) = P(A)P(\Omega) = P(A) \cdot 1$$

□

Lemma 1.6.1 (Borel Cantelli Lemma)**.** *Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $\{E_i\}$ be a sequence of events. Then,*

(i) If $\sum\limits_{i=1}^{\infty} P(E_i) < \infty$, then $P(\limsup E_n) = 0$

(ii) If $\{E_i\}$ is a sequence of independent events, then $P(\limsup E_n) = 0$ or 1 according to to whether the series $\sum\limits_{i=1}^{\infty} P(E_i)$ diverges or converges respectively.

Proof. Set $E = \limsup E_n$. We have $E = \bigcap\limits_{n=1}^{\infty} F_n$ where $F_n = \bigcup\limits_{m=n}^{\infty} E_m$. For every positive integer n,

$$0 \leq P(F_n) \leq \sum_{m=n}^{\infty} P(E_m)$$

Since $\sum\limits_{i=1}^{\infty} P(E_i) < \infty$, then $\lim\limits_{n\to\infty} P(E_n) = 0$. Since $F_n \downarrow E$, from Fatou's lemma we can write

$$0 = \lim_{n\to\infty} P(F_n) = P\left(\lim_{n\to\infty} F_n\right) = P(\limsup E_n)$$

Thus (i) is proved. For (ii), suppose $E_1, E_2, \ldots$ are independent. From (ii), we know that $P(\limsup E_n) = 0$ if the sum $\sum_{n=1}^{\infty} E_n$ is finite. It remains to show that if the sum is infinite, then E_n occurs infinitely often. Let $E = \limsup E_n$. Then

$$E^c = \liminf E_n^c$$

The sequence of events $\{E_n^c\}$ are also independent, so we have

$$P\left(\bigcap_{m=n}^{\infty} E_m^c\right) \leq P\left(\bigcap_{m=n}^{N} E_m^c\right) = \prod_{m=n}^{N} (1 - P(E_m)) \leq \exp\left(-\sum_{m=n}^{N} P(E_m)\right)$$

This inequality comes from the Talor Series expansion of e^x, so

$$e^{-x} = 1 - x + \frac{x^2}{2!} - \frac{x^3}{3!} + \cdots$$

$$e^{-x} \geq 1 - x$$

As $N \to \infty$, we get $\bigcap\limits_{m=n}^{N} E_m^C \downarrow E^c$. Thus,

$$P\left(\bigcap_{m=n}^{N} E_m^c\right) \leq \exp\left(-\sum_{m=n}^{N} P(E_m)\right) \leq \exp\left(-\sum_{m=n}^{N} P(E_m)\right) \to 0$$

This implies that $P(E^c) = 0$, thus $P(E) = 1$.

□

Corollary 1.6.1. *If $\{E_i\}$ is a sequence of independent events then*

$$P(\limsup E_n) = 0 \iff \sum_{i=1}^{\infty} P(E_i) < \infty$$

Proof. ($\Longleftarrow$) We know from the Borel Cantelli lemma that if $\sum_{i=1}^{\infty} P(E_i) < \infty$, then $P(\limsup E_n) = 0$.

($\Longrightarrow$) Suppose $P(\limsup E_n) = 0$. Then from part (2), $P(\limsup E_n) = 1$ when $\sum_{i=1}^{\infty} P(E_i) = \infty$. So it must be that $\sum_{i=1}^{\infty} P(E_i) < \infty$ as required. $\square$

Remark. Notice that indepdence is required by Corollary 1.6.1. To see this, let $(\Omega = [0,1], B, P)$ be a probability space with

$$P(A) = \int_A dx$$

for a Borel set A. It's easy to show that P is a probability measure on $[0,1]$. Now define $E = \left(0, \frac{1}{n}\right)$ and notice that $E_n \downarrow \emptyset$. Therefore,

$$P(\limsup E_n) = P(\lim E_n) = 0$$

Since

$$P(E_n) = \int_0^{\frac{1}{n}} dx = \frac{1}{n}$$

we have

$$\sum_{n=1}^{\infty} P(E_n) = \sum_{n=1}^{\infty} \frac{1}{n} = \infty$$

This does not violate Corollary 1.6.1 since the events E_n are not independent. For example. consider $E_2 = (0, 1/2)$, and $E_3 = (0, 1/3)$. Then

$$P(E_2) = \frac{1}{2}, \quad P(E_3) = \frac{1}{3}$$

$$P(E_2 \cap E_3) = P(E_3) = \frac{1}{3} \neq P(E_2)P(E_3) = \frac{1}{6}$$

Example. Let E_n denote the event that the result of a fair coin flip is heads on both the nth and the $(n+1)$st toss. Then, $\limsup E_n$ is the event that in repeated tossing of a fair coin two successive head appears infinitely often times. Since E_{2n} is an independent sequence of events and

$$\sum_{n=1}^{\infty} P(E_{2n}) = \sum_{n=1}^{\infty} \frac{1}{4} = \infty$$

we have $P(\limsup E_{2n}) = 1$. This implies that $P(\limsup E_n) = 1$ since

$$\limsup E_{2n} \subset \limsup E_n$$

Chapter 2

Important Inequalities and Probability Theorems

In this chapter we review some important inequalities in probability. We start with some definitions and review of probability topics.

Definition 2.0.1 (Probability Mass Density Function)**.** *A function $f : \mathbb{R} \to \mathbb{R}$ is called a probability density function (p.d.f) if*

(i) $f \geq 0$

(ii) $\int f(x)dx = 1$

similarly f is a probability mass function defined on discrete random variables with

1. $f \geq 0$

2. $\sum f(x) = 1$

Definition 2.0.2 (Cumulative Density Function)**.** *The cumulative density function (c.d.f) of a random variable X is defined as*

$$F(X) = P(X \leq x)$$

In otherwords, its the integral (or sum if X is discrete) of the p.d.f.

Definition 2.0.3. *A random variable (r.v.) is of continuous type if its cumulative distribution function (c.d.f) is continuous.*

Definition 2.0.4. *We define the expected value of a random variable X as*

$$E(X) = \mu = \int_{-\infty}^{\infty} x dF(x)$$

where $dF(x) = f dx$ is the probability density function (p.d.f) of X. The variance is given as

$$\mathrm{Var}(X) = \sigma^2 = E[(X-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2 dF(X) = E(X^2) - E(X)^2$$

Example. Let X be ae random variable with p.d.f

$$f(x) = \begin{cases} \frac{1}{3} & \text{if } 0 \leq x < 0.5 \\ \frac{5}{3} & \text{if } 0.5 \leq x \leq 1 \end{cases}$$

(i) Calculate $F(x)$. Is X a continuous random variable?

(ii) Find $E(X)$ and $\mathrm{Var}(X)$.

(iii) Find $P(0.5 < X < 0.75)$.

Solution.

(i) To find the c.d.f, we integrate the p.d.f to get

$$F(X) = \begin{cases} 0 & \text{if } x < 0 \\ \int_0^x \frac{1}{3} dt = \frac{x}{3} & \text{if } 0 \leq x < 0.5 \\ \int_0^{0.5} \frac{1}{6} dt + \int_{0.5}^x \frac{5}{3} dt = \frac{5x}{3} - \frac{2}{3} & \text{if } 0.5 \leq x \leq 1 \\ 1 & \text{if } x > 1 \end{cases}$$

X is a continuous random variable since its c.d.f is continuous.

(ii) To find $E(X)$, we use the definition of expected value,

$$E(X) = \int_0^1 x f(x) dx = \int_0^{0.5} \frac{x}{6} dx + \int_{0.5}^1 \frac{5x}{3} dx = \frac{2}{3}$$

$$E(X) = \int_0^1 x^2 f(x) dx = \int_0^{0.5} \frac{x^2}{6} dx + \int_{0.5}^1 \frac{5x}{3} dx = \frac{1}{2}$$

Therefore $E(X) = \frac{2}{3}$ and $\text{Var}(X) = \frac{2}{3} - \frac{4}{9} = \frac{1}{16}$

(iii)

$$\begin{aligned} P(0.5 < X < 0.75) &= P(X < 0.75) - P(X < 0.5) \\ &= F(0.75) - F(0.5) \\ &= \int_{0.5}^{0.75} f(x) dx \\ &= \int_{0.5}^{0.75} \frac{5}{3} dx \\ &= \frac{5}{12} \end{aligned}$$

All the above ways of calculating $P(0.5 < X < 0.75)$ are equivalent.

2.1 Important Inequalities

Theorem 2.1.1 (Markov's inequality)**.** *Let $X \geq 0$ be a random variable and a be a positive constant. Then*

$$P(X \geq a) \leq \frac{E(X)}{a}$$

Proof. To prove this, notice that

$$I(X \geq a) \leq \frac{X}{a}$$

This is because for a finite a, we have when $X < a$,

$$\frac{X}{a} \geq 0 = I(X \geq a)$$

Then when $X \geq a$,

$$\frac{X}{a} \geq 1 = I(X \geq a)$$

so in either case $\frac{X}{a} \geq I(X \geq a)$. Then we can take expected value of each side of the inequality and get

$$E(I(X \geq a)) = P(X \geq a) \leq E\left(\frac{X}{a}\right) = \frac{E(X)}{a}$$

□

Similarly, we can write

$$I(|X - b| \geq a) \leq \frac{(X - b)^2}{a^2}$$

for all $a > 0$ and $b \in \mathbb{R}$. Then we can take the expected value of each side to get

$$P(|X - b| \geq a) \leq \frac{E((X - b)^2)}{a^2}$$

Taking $b = \mu$ and $a = k\sigma$ where $k > 0$, $\mu = E(X)$ and $\sigma^2 = \text{Var}(X)$, we get Chebyshev's inequality.

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

We can rewrite this as

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

This means that in the set $(\mu - k\sigma, \mu + k\sigma)$, there is at least $100\left(\frac{k^2-1}{k^2}\right)$ % of the population.

Example. Let X be a random variable with

$$P(X = -1) = P(X = 1) = \frac{1}{8} \quad P(X = 0) = \frac{6}{8}$$

We can see that $E(X) = 0$ and $\sigma^2 = \text{Var}(X) = 1/4$. Let $k = 2$, then we can calculate

$$P(|X - \mu| \geq k\sigma) = P(|X| \geq 1) = \frac{1}{4} = \frac{1}{k^2}$$

This example shows Chebyshev's inequality cannot be improved since in this case it becomes an equality.

Definition 2.1.1. *An $n \times n$ matrix U is non-negative definite if for any $n \times 1$ vector c, we have*

$$c'Uc \geq 0$$

Furthermore, all of U eigenvalues are non-negative.

Example. Let

$$U = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$$

Then

$$\begin{aligned} c'Uc &= \begin{bmatrix} c_1 & c_2 \end{bmatrix} \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} \\ &= \begin{bmatrix} c_1 + 0.5c_2 & 0.5c_1 + c_2 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} \\ &= c_1^2 + c_2^2 + c_1c_2 \end{aligned}$$

Now it remains to prove that $c_1^2 + c_2^2 + c_1c_2 \geq 0$. This task is difficult so alternatively, it is much simpler to check that all eigenvalues are non-negative.

$$\det \begin{bmatrix} 1 - \lambda & 0.5 \\ 0.5 & 1 - \lambda \end{bmatrix} = (1 - \lambda)^{-}0.25 = 0 \implies \lambda = 1 \pm \frac{1}{2}$$

So our eigenvalues are $\frac{1}{2}$, and $\frac{3}{2}$ which are both positive so U is non-negative definite.

Note. With the vector $p \times 1$ vector $X = [X_1, \ldots, X_p]$, we have

$$XX' = \begin{bmatrix} X_1 & \cdots & X_p \end{bmatrix} \begin{bmatrix} X_1 \\ \vdots \\ X_p \end{bmatrix} = \begin{bmatrix} X_1^2 & X_1X_2 & \cdots & X_1X_p \\ X_1X_2 & X_2^2 & \cdots & X_2X_p \\ \vdots & \vdots & \ddots & \vdots \\ X_1X_p & X_2X_p & \cdots & X_p^2 \end{bmatrix}$$

Then,

$$E(XX') = \begin{bmatrix} E(X_1^2) & E(X_1X_2) & \cdots & E(X_1X_p) \\ E(X_1X_2) & E(X_2^2) & \cdots & E(X_2X_p) \\ \vdots & \vdots & \ddots & \vdots \\ E(X_1X_p) & E(X_2X_p) & \cdots & E(X_p^2) \end{bmatrix}$$

We can see that XX' is non-negative definition since $c'XX'c = (X'c)'(X'c)$. Denote $(X'c)' = V$, then we have

$$(X'c)'(X'c) = Y'Y = \begin{bmatrix} y_1 & \cdots & y_p \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_p \end{bmatrix} = y_1^2 + \cdots y_2^2$$

In conclusion, XX' is non-negative definite and $E(XX')$ is non-negative definite since, for example, take $X = [X_1, X_2]^T$. Then

$$XX' = \begin{bmatrix} X_1^2 & X_1X_2 \\ X_1X_2 & X_2^2 \end{bmatrix}$$

and $c'E(XX')c = E(c'XX'c) \geq 0$, so

$$E(XX') = \begin{bmatrix} E(X_1^2) & E(X_1X_2) \\ E(X_1X_2) & E(X_2^2) \end{bmatrix}$$

is also non-negative definite.

Theorem 2.1.2 (Cauchy-Shwarz Inequality). *Let $X' = [X_1, \ldots, X_p]$ be a random vector. Define the matrix $U = E(XX')$. Notice that for any $p \times 1$ vector c we have*

$$c'Uc = E((c'X)^2) \geq 0$$

Therefore the matrix U is non-negative definite. Take $p = 2$ to see that

$$0 \leq \det(U) = E(X_1^2)E(X_2^2) - E((X_1X_2))^2$$

Equality holds when $c'X = 0$, $c_1x_1 + c_2x_2 + \cdots + c_px_p = 0$. In other words,

$$E(XY)^2 \leq E(X^2)E(Y^2)$$

and equality holds if there is a linear relationship between X and Y.

$$aX + bY = 0 \quad a, b \in \mathbb{R}$$

Remark. If we replace $X \to (X - \mu_X)$ and $Y \to (Y - \mu_Y)$, then we get

$$E[(X - \mu_X)(Y - \mu_Y)]^2 \leq E(X - \mu_X)E(Y - \mu_Y)$$

Notice that $E(X - \mu_X) = \mathrm{Var}(X)$ and $E[(X - \mu_X)(Y - \mu_Y)] = \mathrm{Cov}(X, Y)$, so can conclude

$$\mathrm{Cov}(X, Y)^2 \leq \mathrm{Var}(X)\,\mathrm{Var}(Y)$$

Furthermore,

$$\rho^2 = \frac{\mathrm{Cov}(X, Y)^2}{\mathrm{Var}(X)\,\mathrm{Var}(Y)} \leq 1$$

Where ρ^2 is the correlation between X and Y. Therefore, when X and Y are linear and the Cauchy-Shwarz inequality becomes an equality, we have

$$Y - \mu_Y = \beta(X - \mu_X)$$

for some constant β. Then, $\rho^2 = 1$.

2.2 Moment Generating Functions

Definition 2.2.1. *Let X be a random variable wtih cdf F, then the* moment generating function *(m.g.f) of X is defined as*

$$M(t) = E(e^{tX}) = \int_{-\infty}^{\infty} e^{tX} dF(X) < \infty$$

Note. Moment generating functions are unique, so with a (m.g.f) we can uniquely determine F and vice-versa.

2.2.1 Properties of Moment Generating Functions

(i) $M(0) = E(e^{0X}) = E(1) = 1$

(ii) $\frac{d}{dt}M(t) = \int_{-\infty}^{\infty} xe^{tX} dF(X) = E(Xe^{tX}) \implies M'(0) = E(X)$

(iii) $\frac{d^2}{dt^2}M(t) = \int_{-\infty}^{\infty} x^2e^{tX} dF(X) = E(X^2e^{tX}) \implies M''(0) = E(X^2)$. $E(X^2)$ is the 2nd moment of X.

(iv) In general,

$$\frac{d^k}{dt^k}M(t) = \int_{-\infty}^{\infty} x^k e^{tx} dF(x)$$

$$M^{(k)}(0) = E(X^k)$$

(v) We can use the Taylor series expansion of the moment generating function to get

$$M^{(k)}(t) = \sum_{k=0}^{\infty} \frac{M^{(k)}(0)}{k!} t^k = \sum_{k=0}^{\infty} \frac{E(X^k)}{k!} t^k$$

Then if we multiply by $k!$ to get

$$E(X^k) = (\text{Multiple of } t^k) \times k! = \frac{M^{(k)}(0)}{k!} \cdot k! = M^{(k)}(0)$$

Example. Recall the gamma distribution

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} I(x > 0)$$

With

$$\begin{aligned}
\Gamma(\alpha) &= \int_0^\infty e^{-x} x^{\alpha-1} dx \\
&= \left[x^{\alpha-1} e^{-x}\right]_0^\infty + \int_0^\infty e^{-x}(\alpha-1)x^{\alpha-2} dx \\
&= (\alpha-1) \int_0^\infty e^{-x}(\alpha-1)x^{\alpha-2} dx \\
&= (\alpha-1)\Gamma(\alpha-1)
\end{aligned}$$

Then, we have

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

$$\Gamma(2) = \int_0^\infty x e^{-x} dx = \left[-x e^{-x}\right]_0^\infty + \int_0^\infty e^{-x} dx = 1$$

So,

$$\Gamma(3) = 2\Gamma(2) = 2,\ \Gamma(4) = 3\Gamma(3) = 6 = 3!,\ \Gamma(5) = 4!,\ \ldots$$

Therefore,

$$\alpha \in \mathbb{N} \implies \Gamma(\alpha) = (\alpha - 1)!$$

If we set $x = \beta u$, we get

$$\Gamma(\alpha) = \int_0^\infty e^{-\beta u} \beta^{\alpha-1} u^{\alpha-1} \beta du = \beta^\alpha \int_0^\infty e^{-\beta u} u^{\alpha-1} du$$

This gives us the equation for the Laplace transform for $u^{\alpha-1}$.

$$\frac{\Gamma(\alpha)}{\beta^\alpha} = \int_0^\infty e^{\beta u} u^{\alpha-1} du = \mathcal{L}(u^{\alpha-1})$$

This can be useful for calculating integrals, such as

$$\int_0^\infty e^{-3x} x^5 dx = \frac{5!}{3^6}$$

If we multiply the Laplace transform by $\frac{\beta^\alpha}{\Gamma(\alpha)}$,

$$\frac{\Gamma(\alpha)}{\beta^\alpha} \cdot \frac{\beta^\alpha}{\Gamma(\alpha)} = \int_0^\infty \frac{\beta^\alpha}{\Gamma(\alpha)} e^{-\beta u} u^{\alpha-1} du = 1$$

Since this function is positive and integrates to 1, it is a p.d.f.

$$f(u) = \frac{\beta^\alpha}{\Gamma(\alpha)} e^{-\beta u} u^{\alpha-1}$$

Thus, this is the Gamma distribution. If we take $\alpha = r/2$, $\beta = 1/2$. We get a special case of the Gamma distribution

$$f(u) = \frac{1}{\Gamma\left(\frac{r}{2}\right) 2^{r/2}} e^{-u/2} u^{\frac{r}{2}-1}$$

This is the chi-squared distribution with r degrees of freedom. Now we can find the moment generating function for the Gamma distribution

$$\begin{aligned}
M(t) = E(e^{tx}) &= \int_0^\infty \frac{\beta^\alpha}{\Gamma(\alpha)} e^{-\beta x} x^{\alpha-1} e^{tx} dx \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty e^{-(\beta-t)x} x^{\alpha-1} e^{tx} dx \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)} \cdot \frac{\Gamma(t)}{(\beta-t)^\alpha} \\
&= \beta^\alpha (\beta - t)^{-\alpha} = \left(1 - \frac{t}{\beta}\right)^{-\alpha}
\end{aligned}$$

We can also calculate the expected value and variance

$$M'(t) = \frac{\alpha}{\beta}\left(1 - \frac{t}{\beta}\right)^{-\alpha-1} \implies M'(0) = \frac{\alpha}{\beta} = 1$$

$$M''(t) = \frac{\alpha}{\beta} \cdot -\frac{1}{\beta}(-\alpha - 1)\left(1 - \frac{t}{\beta}\right)^{-\alpha-2} \implies M''(0) = \frac{\alpha(\alpha+1)}{\beta^2}$$

$$\text{Var}(X) = E(X^2) - E(X)^2 = \frac{\alpha(\alpha+1)}{\beta^2} - \frac{\alpha^2}{\beta^2} = \frac{\alpha}{\beta^2}$$

Similarly, for the chi-squared distribution, we have $\alpha = r/2$ and $\beta = 1/2$,

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

$$E(X \sim \chi^2(r)) = \frac{\alpha}{\beta} = \frac{r/2}{1/2} = r$$

$$\text{Var}(X \sim \chi^2(r)) = \frac{\alpha}{\beta^2} = \frac{r/2}{1/4} = 2r$$

2.2.2 Applications of Moment Generating Functions

1. Suppose we have k random variables $X_1 \sim \chi^2(r_1), \ldots, X_k \sim \chi^2(r_k)$. If they're independent, then we can define $Y = \sum_{i=1}^{k} X_i$, and calculate the moment generating function

$$\begin{aligned}
E(e^{tY}) &= E\left(\exp\left(t \sum_{i=1}^{k} X_i\right)\right) \\
&= E(e^{tX_1} \cdot e^{tX_2} \cdots e^{tX_k}) \\
&= E(e^{tX_1}) \cdot E(e^{tX_2}) \cdots E(e^{tX_k}) \\
&= (1-2t)^{-r_1/2} \cdot (1-2t)^{-r_2/2} \cdots (1-2t)^{-r_k/2} \\
&= (1-2t)^{-(r_1+r_2+\cdots+r_k)/2}
\end{aligned}$$

   Therefore, $Y \sim \chi^2(r_1 + r_2 + \cdots + r_k)$

2. Let $Z \sim N(0,1)$ with pdf

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

   We know that since $f(z)$ is a p.d.f,

$$\int_{-\infty}^{\infty} \frac{1}{2\sqrt{\pi}} e^{-z^2/2} dz = 1$$

We can calculate the m.g.f

$$\begin{aligned}
M(t) = E(e^{tZ}) &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} e^{tz} dz \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{\frac{-(z-t)^2}{2}} e^{t^2/2} dz \\
&= e^{t^2/2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{\frac{-(z-t)^2}{2}} dz \\
&= e^{t^2/2} \cdot 1
\end{aligned}$$

The integral is 1 since it is the integral of the p.d.f of the normal distrubtion with $N(t,1)$. Recall that the p.d.f for a general normal distribution $N(\mu, \sigma^2)$ is

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

Using the Taylor expansion for e^x we can rewrite the m.g.f as

$$E(e^{tZ}) = e^{t^2/2} = \sum_{k=0}^{\infty} \frac{(t^2/2)^k}{k!} = \sum_{k=0}^{\infty} \frac{t^{2k}}{k!2^k}$$

We say earlier that $E(X^k) =$ Some multiple of $t^k \cdot k!$. So this means that $E(X^{2k+1}) = 0$ since $2k+1$ is odd but we have only even multiples of t^k. Therefore, all odd moments have 0 expecation. All even moments are

$$E(X^{2k}) = \frac{(2k)!}{k!2^k}$$

Example. Let $Z \sim N(0,1)$, find the m.g.f for Z^2.

Solution.

$$\begin{aligned}
M(t) &= E(e^{tZ^2}) \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} e^{tz^2} dz \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-z^2(-t+1/2)} dz \\
&= (1-2t)^{-1/2} = \frac{1}{\sqrt{1-2t}}
\end{aligned}$$

Note that in general if $Z \sim N(0, 2\alpha)$, we have

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\alpha z^2} dz = \frac{1}{\sqrt{2\alpha}}$$

To summarize,

$$Z \sim N(0,1) \implies M(t) = e^{t^2/2}$$

$$Y = Z^2 \implies M_Y(t) = (1-2t)^{-1/2} \to \chi^2(1)$$

$$M_{\chi^2}(t) = (1-2t)^{-r/2}$$

$$Z \sim N(0,1) \implies Z^2 \sim \chi^2(1)$$

Example. Let $X_1, \ldots, X_n$ be independent random variables with $X_i \sim N(\mu_i, \sigma_i^2)$. What is the distribution for $Y = \sum_{i=1}^{n} X_i$?

Solution. We can calculate the m.g.f for X_i

$$M_{X_i}(t) = E(e^{t(\sigma_i z + \mu_i)}) = e^{t\mu_i} E(e^{(t\sigma_i)z}) = e^{t\mu_i + t^2\sigma_i^2/2}$$

Then, the m.g.f for Y is

$$\begin{aligned}
M_Y(t) &= E(e^{tY}) \\
&= E\left(\exp\left(t \sum a_i X_i\right)\right) \\
&= E(e^{ta_i X_1} \cdot e^{ta_i X_2} \cdots e^{ta_i X_n}) && \text{(Independence)} \\
&= \exp\left(ta_1\mu_1 + \frac{t^2 a_1^2 \sigma_n^2}{2}\right) \cdots \exp\left(ta_n\mu_n + \frac{t^2 a_n^2 \sigma_n^2}{2}\right) \\
&= \exp\left(t(a_1\mu_1 + \cdots + a_n\mu_n) + \frac{t^2}{2}(a_1^2\sigma_1^2 + \cdots + a_n^2\sigma_n^2)\right)
\end{aligned}$$

Notice that the m.g.f for the general normal distribution is $M(t) = \exp(t\mu + t^2\sigma^2/2)$. So in this case we have $\mu = a_1\mu_1 + \cdots + a_n\mu_n$, and $\sigma^2 = a_1^2\sigma_1^2 + \cdots + a_n^2\sigma_n^2$. Therefore,

$$Y = \sum_{i=1}^{n} a_i X_i \sim N\left(\sum_{i=1}^{n} a_i\mu_i, \sum_{i=1}^{n} a_i^2\sigma_i^2\right)$$

Binomial Distribution

Recall a Bernoulli random variable X is a random variable with $P(X = 1) = p$, and $P(X = 0) = 1 - p = q$, then $X \sim \text{Bernoulli}(p)$, $E(X) = 0q + 1p = p$, $E(X^2) = 0^2q + 1^2p = p$, $\text{Var}(X) = p - p^2 = p(1 - p) = pq$. Let $X_1, \dots X_n \overset{\text{iid}}{\sim} \text{Bernoulli}(p)$, then $Y = \sum_{i=1}^{n} X_i$ is the number of 1's observed, or in other words the number of successes. Then, we have

$$P(Y = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

We can prove this by finding the m.g.f for Y,

$$\begin{aligned}
E(e^{tY}) &= E(e^{tX_1}) \cdots E(e^{tX_n}) \\
&= (pe^t + q)^n
\end{aligned}$$

$$\begin{aligned}
E(e^{tY}) &= \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} e^{tk} \\
&= \sum_{k=0}^{n} \binom{n}{k} (pe^t)^k (1-p)^{n-k} && = (pe^t + q)^n
\end{aligned}$$

Notice that from the binomial expansion we have

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

and replace a with pe^t and b with $q = 1 - p$. In conclusion, the sum of Bernoulli random variables $Y = \sum_{i=1}^{n} X_i$ is a binomial random variable $Y \sim \text{Bin}(n, p)$.

Example. Let $X_1 \sim \text{Bin}(m, p)$, $X_2 \sim \text{Bin}(n, p)$ be independent random variables. Then $X_1 + X_2 \sim \text{Bin}(m + n, p)$. We can show this using the moment generating function.

$$M_{X_1+X_2}(t) = E(e^{t(X_1+X_2)}) = E(e^{tX_1})E(e^{tX_2}) = (pe^t + q)^{m+n}$$

2.3 Holder, Lyapunov and Minkowski Inequalities

Corollary 2.3.1. *If $P(X \geq 0) = 1$ and $E(X) = \mu$, then*

$$P(X \geq 2\mu) \leq 0.5$$

Proof. Using Markov's Inequality,

$$P(X \geq 2\mu) \leq \frac{E(X)}{2\mu} = \frac{1}{2}$$

□

Lemma 2.3.1. *If $\alpha \geq 0$, $\beta \geq 0$, and*

$$\frac{1}{p} + \frac{1}{p} = 1,\ p > 1,\ q > 1$$

then

$$0 \leq \alpha\beta \leq \frac{\alpha^p}{p} + \frac{\beta^p}{q}$$

Proof. If $\alpha\beta = 0$, then the inequality holds trivially. Therefore, let $\alpha > 0$, $\beta > 0$. Then, define for $t > 0$

$$\phi(t) = \frac{t^p}{p} + \frac{t^{-q}}{q}$$

Differenting this function we get

$$\phi'(t) = t^{p-1} - t^{-q-1}$$

We can see that $\phi'(1) = 0$, $\psi'(t) < 0$ when $t \in (0, 1)$, and $\psi'(t) > 0$ when $t > 1$. Thus, t minimizes ϕ on $(0, \infty)$. Set $t = \frac{a^{1/q}}{\beta^{1/p}}$ to get

$$\frac{\alpha^{p/q}}{p\beta} + \frac{\alpha^{-1}}{q\beta^{-q/p}} \geq 1$$

Multyipling both sides by $\alpha\beta$ and using

$$p/q + 1 = p \ \text{ and } \ q/p + 1 = q$$

we get

$$\alpha\beta \leq \frac{\alpha^{p/q+1}}{p} + \frac{\beta}{q\beta^{-q/p}} = \frac{\alpha^p}{p} + \frac{\beta^q}{q}$$

□

Theorem 2.3.1 (Holder's Inequality)**.** *Let X and Y be two random variables and*

$$\frac{1}{p} + \frac{1}{q} = 1,\ p > 1,\ q > 1$$

We have

$$E(|XY|) \leq (E(|X|^p))^{1/p}(E(|Y|^q))^{1/q}$$

Proof. In the case that $E(|X|^p)E(|Y|^q) = 0$, the result follows. Otherwise, from Lemma 2.3.1, take

$$\alpha = \frac{|X|}{(E(|X|^p))^{1/p}},\ \beta = \frac{|Y|}{(E(|Y|^q))^{1/q}}$$

Then using

$$\alpha\beta \leq \frac{\alpha^p}{p} + \frac{\beta^q}{q}$$

we get

$$\frac{|XY|}{(E(|X|^p))^{1/p}(E(|Y|^q))^{1/q}} \leq \frac{|X|^p}{pE(|X|^p)} + \frac{|Y|^q}{qE(|Y|^q)}$$

Now taking the expected value, we have

$$\frac{E(|X|^p)}{pE(|X|^p)} + \frac{E(|Y|^q)}{qE(|Y|^q)} = \frac{1}{p} + \frac{1}{q} = 1$$

Therefore,

$$\frac{E(|XY|)}{(E(|X|^p))^{1/p}(E(|Y|^q))^{1/q}} \leq 1 \implies E(|XY|) \leq (E(|X|^p))^{1/p}(E(|Y|^q))^{1/q}$$

□

Theorem 2.3.2 (Minkowski's Inequality)**.** *For $p \geq 1$, we have*

$$E(|XY|) \leq (E(|X|^p))^{1/p}(E(|Y|^q))^{1/q}$$

Proof. Since $|X+Y| \leq |X| + |Y|$, the case that $p = 1$ is obvious. Let $p > 1$, choose q such that

$$\frac{1}{p} + \frac{1}{q} = 1$$

Then, use Holder's inequality to write

$$\begin{aligned} E(|X+Y|^p) &= E(|X+Y||X+Y|^{p-1}) \\ &\leq E(|X||X+Y|^{p-1}) + E(|Y||X+Y|^{p-1}) \\ &\leq (E(|X|^p))^{1/p}(E(|X+Y|^{(p-1)q}))^{1/q} + (E(|Y|^p))^{1/p}(E(|X+Y|^{(p-1)q}))^{1/q} \\ &= (E(|X|^p))^{1/p}(E(|X+Y|^p))^{1/q} + (E(|Y|^p))^{1/p}(E(|X+Y|^p))^{1/q} \\ &= (E|X+Y|^p)^{1/q}((E(|X|^p))^{1/p} + (E(|Y|^p))^{1/p}) \end{aligned}$$

Now we can divide both sides by $(E(|X+Y|^p))^{1/q}$ to get

$$\frac{E(|X+Y|^p)}{(E|X+Y|^p)^{1/q}} = (E|X+Y|^p)^{1-1/q} = (E|X+Y|^p)^p$$

Thus,

$$(E|X+Y|^p)^p \leq (E(|X|^p))^{1/p} + (E(|Y|^p))^{1/p}$$

□

We can define

$$||X||_p - E(|X|^p)^{1/p}$$

And we can define a space of random variables whose pth moment exists as

$$\mathcal{X} = \{X : E|X|^p < \infty\}$$

Then we can calculate a metric on $\mathcal{X}$ as

$$d(X,Y) = (E|X-Y|^p)^{1/p}$$

We can confirm that $d(X,Y)$ is a metric by confirming the axioms

(i) $d(X,X) = 0$

(ii) $d(X,Y) = d(Y,X)$

(iii) $d(X,Z) \leq d(X,Y) + d(Y,Z)$

The triangular inequality is Minkowski's inequality which we proved previously. Thus we have the Hilbert Space $(\mathcal{X}, d)$.

$$\langle X, Y \rangle = \int X(\omega)Y(\omega)d\omega$$

this is an inner product space. A special case of Mikowski's inequality is when $p = 1$ and $p = 2$,

$$(E|X+Y|^2)^{1/2} \leq (E(X^2))^{1/2} + (E(Y^2))^{1/2}$$

$$E|X+Y| \leq E(X) + E(Y)$$

Theorem 2.3.3 (Jensen's Inequality)**.** *Let $\phi(x)$ be a convex function, then*

$$\phi(E(X)) \leq E(\phi(X))$$

The proof for this is simple since if the function is convex, its derivatives are increasing, thus the secant lines from any points x_1, x_2 on the function are above the curve, therefore the average of the function and the average of the secant lines is

$$\phi\left(\frac{x_1+x_2}{2}\right) \leq \frac{f(x_1)+f(x_2)}{2}$$

Theorem 2.3.4 (Lyapunov's Inequality)**.** *If $r > s > 0$, then*

$$||X|_r = (E|X|^r)^{1/r} \geq (E|X|^s)^{1/s} = ||X||_s$$

Proof. Define $g(x) = |x|^u$ with $u > 1$ so that g is convex. Then, we know that $r/s > 1$ since $r > s > 0$. From Jensen's inequality, we can take $u = r/s > 1$ to get

$$E(g(x)) \geq g(E(X)) \iff E(|X|^{r/s}) \geq (E|X|)^{r/s}$$

Then we can rewrite this as

$$E(|X|^r) = E\left[(|X|^s)^{r/s}\right] \geq (E|X|^s)^{r/s}$$

Replacing $|X|$ with $|X|^s$,

$$E(|X|^r) \geq (E|X|^s)^{r/s}$$

We can take the rth root of each side and this inequality will still hold,

$$E(|X|^r)^{1/r} \geq (E|X|^s)^{1/s}$$

as required. $\square$

Chapter 3

Induced Probability Measures

Recall that a random variable X is a map from the sample space Ω to the real number line, so $X : \Omega \to \mathbb{R}$.

Example. Consider the flip of a coin with $\Omega = \{H, T\}$. Let X be the number of heads observed, then $X(H) = 1$, and $X(T) = 0$.

Another Example. Consider the result of rolling 2 die, so

$$\Omega = \{(1,1), (1,2), (1,3), \ldots, (6,5), (6,6)\}$$

So $|\Omega| = 36$. Then define an $\omega = (x, y) \in \Omega$ as the result of rolling the first die x and the second die y, and the random variable X to be $X(\omega) = x + y$. So $X(2,3) = 5$, $X(6,6) = 12$ etc. We can define a different random variable $Y(x, y) = |x - y|$. We can ask questions about probabilities on Y, such as what is the probability $Y = 1$?

$$P(Y = 1) = P(\{\omega : Y(\omega) = 1\}) = P(\{(1,2), (2,1), (2,3), (3,2), \ldots\}) = \frac{10}{36}$$

So $\{\omega : Y(\omega) = 1\} = Y^{-1}(\{1\})$. In otherwords the probability of an event E occuring can be written as

$$P_Y(E) = P(Y^{-1}(E))$$

Where P_Y is called the *induced probability measure* by Y. We can check that P_Y satisfies all conditions of a probability measure.

$$P_Y(\mathbb{R}) = 1$$

If E_i's are disjoint, then

$$P_Y\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P_Y(E_i)$$

since

$$Y^{-1}\left(\bigcup_{i=1}^{\infty} E_i\right) = \bigcup_{i=1}^{\infty} Y^{-1}(E_i)$$

Example. Flip 2 coins and let X be the number of heads observed.

$$\Omega = \{HH, HT, TH, TT\}$$

and

$$X(HH) = 2, X(HT) = X(TH) = 1, X(TT) = 0$$

So $X : \Omega \to \{0, 1, 2\} \subset \mathbb{R}$. We can define the induced probability measure P_X by

$$P_X(\{0\}) = P(X = 0) = P(\{TT\}) = \frac{1}{4}$$

$$P_X(\{1\}) = P(X = 1) = P(\{TH, HT\}) = \frac{1}{2}$$

$$P_X(\{1, 2\}) = P(X = 0 \text{ or } X = 1) = P(\{TH, HT, HH\}) = \frac{3}{4}$$

$$P_X(\{0, 2\}) = P(X = 0) + P(X = 2) = P(\{TT, HH\}) = \frac{1}{2}$$

$$P_X(\{0, 1, 2\}) = 1$$

3.1 Cumulative Distribution Functions

We can know look at a more exact definition of c.d.f's.

Definition 3.1.1 (Cumulative Distribution Functions)**.** *If X is a random variable $X : \Omega : \mathbb{R}$, then the c.d.f of X is*

$$F_X(x) = P(X \leq x) = P(\{\omega : X(\omega) \leq x\}) = P_X((-\infty, x])$$

3.1.1 Properties of Cumulative Distribution Functions

Theorem 3.1.1. *Let F be a c.d.f, then*

(i) $F(x) \geq 0$ *and* $F(x)$ *is non-decreasing.*

(ii) $\lim_{x \to t^+} F(x) = F(t)$

(iii) $F(-\infty) = 0$ *and* $F(-\infty) = 1$.

Proof. (i) To prove F is non-decreasing, we need to show

$$F(x+h) - F(x) \geq 0, \ \forall h \geq 0$$

$$F(x+h) - F(x) = P(x < X \leq x+h) \geq 0$$

Since probability measures are positive, therefore F is non-decreasing.

(ii) Consider

$$E_n = (-\infty, x + h_n)$$

With $h_n \downarrow 0$, for example $h_n = \frac{1}{n} \downarrow 0$. Then, E_n is decreasing so the limit is

$$\lim_{n \to \infty} E_n = \bigcap_{n=1}^{\infty} E_n = (-\infty, x]$$

Thus, from Lemma 1.5.2, we have

$$\lim_{n \to \infty} P(E_n) = P\left(\lim_{n \to \infty} E_n\right) = P\left((-\infty, x]\right) = F(x)$$

Therefore we showed that

$$\lim_{n \to \infty} F(x + h_n) = F(x)$$

So F is right continuous.

(iii) We know $F(N) - F(-N) = P(-N < X \leq N) = P((-N, N])$ for any, N, so

$$\lim_{N \to \infty} = (-\infty, \infty) \implies P((-N, N]) \uparrow 1$$

Therefore,

$$\lim_{N \to \infty} [F(N) - F(-N)] = 1$$

Thus $F(\infty) = 1$ and $F(-\infty) = 0$.

□

Remark. A distrbiution function F is continuous at $x \in \mathbb{R}$ if and only if $P(X = x) = 0$, since

$$P(X = x) = F(x) - \lim_{x \to x^-} F(x) = \text{The size of the jump at } x$$

Definition 3.1.2. *A random variable X is of continuous type if $F(x) = P(X \leq x)$ is a continuous function.*

Example. Let

$$F(x) = \frac{1}{2} I(x \in [0,1)) + \frac{2x}{3} I\left(x \in [1, 3/2)\right) + I\left(x \in [3/2, \infty)\right)$$

This c.d.f is not continuous since there is a jump at $x = 0$ and $x = 1$, so we have

$$P(X = 0) = F(0) - \lim_{x \to 0^-} F(x) = \frac{1}{2} - 0$$

$$P(X=1) = \frac{3}{2} - \frac{1}{2} = \frac{1}{6}$$

Recall that a set is called *countable* if we can define a bijection from the natural numbers to the set. For example

$$f : \mathbb{N} \mapsto \{2,4,6,8,\ldots\} \implies f(n) := 2n$$

We say a set is at most countable if it is finite or countable.

Lemma 3.1.1. *If F is a c.d.f, then the number of discontinuities of F is at most countable.*

Proof. Define

$$p(x) := F(x) - F(x^-) = P(X=x)$$

Let D be the set of discontinuities of F, so

$$D = \{x : p(x) > 0\}$$

We need to show that D is at most countable. Let

$$D_n := \left\{x : \frac{1}{n+1} < p(x) \le \frac{1}{n}\right\}$$

Then,

$$\bigcup_{n=1}^{\infty} D_n = D$$

We must prove that D_n is finite. D_n is bounded below by $\frac{1}{n+1}$, so $|D_n|$ is at most n, since if we have $n+1$ points,

$$P(D_n) > \frac{n+1}{n+1} = 1$$

which is a contradiction. Therefore, D is at most countable. □

Lemma 3.1.2. *Let X be a random variable with c.d.f F and $p(x) = F(x) - F(x^-) = P(X=x)$. Let $D = \{x_1, x_2, \ldots\}$ be the set of discontinuities of F. Define the step function*

$$G(x) = \sum_{n=1}^{\infty} p(x_i) I(x \ge x_i)$$

Then $H(x) = F(x) - G(x)$ is non-decreasing and continuous on $\mathbb{R}$.

Note. We call G a step function since it increases in discrete intervals and is constant in between.

Proof. Obviously H is right continuous since $F(x)$ and $G(x)$ are right continuous. We want to show that H is also left continuous. Not that if $x' < x$, then

$$H(x) - H(x') = F(x) - F(x') - G(x) + G(x')$$

As $x' \uparrow x$, then $F(x) - F(x')$ converges to the size of jummp of F at x, and $G(x) - G(x')$ converges to the size of the jump of G at x. The size of jump in both cases is $p(x)$ which shows that $H(x) - H(x') \to 0$ as $x' \uparrow x$. Therefore H is continuous. Now we want to show that H is non-decreasing. Note that

$$\sum_{x' < x_n \le x} p(x_n) \le P(x' < X \le x) = F(x) - F(x')$$

We want to show that $H(x) \ge H(x')$.

$$\begin{aligned} H(x') &= F(x') - G(x') \\ \implies H(x) - H(x') &= F(x) - F(x') - (G(x) - G(x')) \\ &= F(x) - F(x') - \sum_{x' < x_n \le x} p(x_n) \end{aligned}$$

We know that $\sum_{x' < x_n \le x} p(x_n) \le F(x) - F(x')$, therefore $H(x) - H(x') \ge 0$. Thus H is non-decreasing. □

Theorem 3.1.2. *Let F be a c.d.f, then there exists two c.d.f's F_d, F_c where F_d is a discrete step function and F_c is a continuous function such that*

$$F = \alpha F_d + (1-\alpha)F_c$$

for some $\alpha \in [0,1]$, and

$$F_d = \frac{G(x)}{\alpha}$$

$$F_c = \frac{H(x)}{1-\alpha}$$

Example. Find F_c, F_d, and α for the following c.d.f

$$F(x) = \frac{1}{2}I(0 \leq x < 1) + \frac{2x}{3}I(1 \leq x < 3/2) + I(x \geq 3/2)$$

Solution. Consider the set of discontinuities of F, we have $\{x_1 = 0, x_2 = 1\}$. Then,

$$p(x_1) = p(0) = \frac{1}{2}$$
$$p(x_2) = p(1) = \frac{2}{3} - \frac{1}{2} = \frac{1}{6}$$

Then we can define $G(x)$,

$$G(x) = \sum_{x_i \leq x} p(x_i)$$

If $x < 0$, then $G(x) = 0$, similarly

$$0 \leq x < 1 \implies G(x) = \frac{1}{2}$$
$$1 \leq x \implies G(x) = \frac{1}{2} + \frac{1}{6} = \frac{2}{3}$$

So we can define $G(x)$ with indicator functions

$$G(x) = \frac{1}{2}I(0 \leq x < 1) + \frac{2}{3}I(1 \leq x)$$

We can see that $G(x)$ is not a c.d.f since $G(\infty) \neq 1$, then our α is

$$\alpha = G(\infty) = \frac{2}{3} \implies 1 - \alpha = \frac{1}{3}$$

To find $H(x)$, we can write $G(x)$ and $F(x)$ as

$$F(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{2} & 0 \leq x < 1 \\ \frac{2x}{3} & 1 \leq x < \frac{3}{2} \\ 1 & x \geq \frac{3}{2} \end{cases}$$

$$G(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{2} & 0 \leq x < 1 \\ \frac{2}{3} & 1 \leq x < \frac{3}{2} \\ \frac{2}{3} & x \geq \frac{3}{2} \end{cases}$$

Then,

$$H(x) = F(x) - G(x) = \begin{cases} 0 & x < 0 \\ 0 & 0 \leq x < 1 \\ \frac{2x}{3} - \frac{2}{3} = \frac{2}{3}(x-1) & 1 \leq x < \frac{3}{2} \\ \frac{1}{3} & x \geq \frac{3}{2} \end{cases}$$

Now we can find F_d and F_c,

$$F_d(x) = \frac{G(x)}{\alpha} = \frac{3G(x)}{2} = \begin{cases} 0 & x < 0 \\ \frac{3}{4} & 0 \leq x < 1 \\ 1 & x \geq 1 \end{cases}$$

$$F_c(x) = \frac{H(x)}{1-\alpha} = 3H(x) = \begin{cases} 0 & x < 1 \\ 2(x-1) & 1 \leq x < \frac{3}{2} \\ 1 & x \geq \frac{3}{2} \end{cases}$$

Notice that the p.d.f

$$\frac{dF_d(x)}{dx} = \begin{cases} \frac{3}{4} & x = 0 \\ \frac{1}{4} & x = 1 \end{cases}$$

This is the p.d.f of a Bernoulli random variable with $p = 1/4$, and

$$\frac{dF_c(x)}{dx} = \begin{cases} 2 & 1 \leq x < \frac{3}{2} \\ 0 & \text{otherwise} \end{cases}$$

This is the p.d.f of a uniform random variable on $[1, 3/2]$. Therefore, F is a linear combination of a Bernoulli random variable and a uniform random variable. We can calculate the expected value

$$E(X) = \int x dF(x) = \alpha \int x dF_d(x) + (1-\alpha) \int x dF_c(x)$$

Example. Let X be a random variable with continuous c.d.f F. Find the distrbiution for

(i) $U = F(X)$

(ii) If $U \sim [0,1]$, then $X \stackrel{\text{d}}{=} F^{-1}(U)$

(iii) Use (i) and (ii) to show that $X = -\ln U$ has an exponentional distrbution.

Note. If $U \sim \text{Unif}(0,1)$, then

$$U \stackrel{\text{d}}{=} 1 - U$$

since the c.d.f for $1-U$ is

$$G(t) = P(1 - U \leq t) = P(U \geq 1 - t) = \int_{1-t}^{1} du = t$$

Solution.

(i) We have

$$G(u) = P(F(X) \leq u) = P(F^{-1}(F(X)) \leq F^{-1}(u)) = F(F^{-1}(u)) = u$$

Therefore, $g(u) = G'(u) = 1$ for $u \in [0,1]$.

(ii) Notice that

$$\begin{aligned} P(F^{-1}(U) \leq x) &= P(F(F^{-1}(U)) \leq F(s)) \\ &= P(U \leq F(x)) \\ &= \int_0^{F(x)} du = F(x) \end{aligned}$$

(iii) Now using (i), (ii)

$$\begin{aligned} F(x) &= P(-\ln U \leq x) \\ &= P(U \geq e^{-x}) \\ &= 1 - P(U \leq e^{-x}) \\ &= 1 - e^{-x} \\ \implies f(x) &= F'(x) = e^{-x} I(x \geq 0) \end{aligned}$$

3.2 Conditional Probability Measure

Definition 3.2.1. *Let $P(X \leq x) = F(x)$ be continuous. If there exists a non-negative function f such that*

$$P(X \leq x) = \int_{-\infty}^{x} f(t)dt$$

Then f is called the probability density function (p.d.f).

Definition 3.2.2. *Let $(\Omega, \mathcal{F}, P)$ be a probability space and $B \in \mathcal{F}$ with $P(B) > 0$. The conditional probability measure given B as*

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

Theorem 3.2.1. *Q_B is a probability measure on $(\Omega, \mathcal{F})$.*

Proof. We can check each property of probability measures

(i) $Q_B(\emptyset) = 0$ since

$$Q_B(\emptyset) = \frac{P(\emptyset \cap B)}{P(B)} = \frac{P(\emptyset)}{P(B)} = 0$$

(ii) The sum of disjoint sets is

$$\begin{aligned} Q_B\left(\bigcup_{i=1}^{\infty} A_i\right) &= \frac{P(B \cap (A_1 \cup A_2 \cup \cdots))}{P(B)} \\ &= \frac{P((B \cap A_1) \cup (B \cap A_2) \cup \cdots)}{P(B)} \\ &= \sum_{i=1}^{\infty} \frac{P(B \cap A_i)}{P(B)} \\ &= \sum_{i=1}^{\infty} Q_B(A_i) \end{aligned}$$

(iii) $Q_B(\Omega) = 1$ since

$$Q_B(\Omega) = \frac{P(\Omega \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1$$

(iv) $P(A \cap B) \leq P(B)$ since $A \cap B \subset B$, therefore $Q_B(A) \in [0, 1]$.

□

Remark. We have

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

but we this does not apply for B^c,

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

Theorem 3.2.2 (Law of Total Probability)**.** *Let $\{E_i\}$ be a sequence of disjoint events such that*

$$\Omega = \bigcup_{i=1}^{\infty} E_i$$

Then

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

Proof.

$$
\begin{aligned}
P(A) &= P(A \cap \Omega) \\
&= P\left(A \cap \bigcup_{i=1}^{\infty} E_i\right) \\
&= P\left(\bigcup_{i=1}^{\infty} A \cap E_i\right) \\
&= \sum_{i=1}^{\infty} P(A \cap E_i) \\
&= \sum_{i=1}^{\infty} P(A|E_i)P(E_i)
\end{aligned}
$$

□

Theorem 3.2.3 (Bayes' Theorem). *Let $\{E_i\}$ be a sequence of disjoint events such that*

$$\Omega = \bigcup_{i=1}^{\infty} E_i$$

Then

$$P(E_i|A) = \frac{P(A|E_i)P(E_i)}{P(A)} = \frac{P(A|E_i)P(E_i)}{\sum_{i=1}^{\infty} P(A|E_i)P(E_i)}$$

Proof. □

Example. Roll a die and flip a coin the number of times that appears on the die. Let X denote the random variable of the number of heads observed. What is the distrbiution of X? If we observed 3 heads, what is the probability that the die is 4?

Solution. Let Y denote the number on the die, then

$$
\begin{aligned}
P(X = k) &= \sum_{i=1}^{6} P(X = k|Y = i)P(Y = i) \\
&= \sum_{i=1} \binom{i}{k} \left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{i-k} \frac{1}{6} \\
&= \frac{1}{6} \sum_{i=1}^{6} \binom{i}{k} \left(\frac{1}{2}\right)^i
\end{aligned}
$$

Then, we want to find $P(Y = 4|X = 3)$, using Bayes' theorem

$$
\begin{aligned}
P(Y = 4|X = 3) &= \frac{P(X = 3|Y = 4)P(Y = 4)}{P(X = 3)} \\
&= \frac{\binom{4}{3}\left(\frac{1}{2}\right)^k \frac{1}{2}\frac{1}{6}}{\frac{1}{6}\sum_{i=3}^{6}\binom{i}{3}\left(\frac{1}{2}\right)^i} \\
&= \frac{4}{16}
\end{aligned}
$$

Example. An urn contains $m + n$ chips of which M are white and the rest are black. A chip is drawn at random without observing its color, then another chip is drawn. What is the probability that the second chip is white?

Solution. Let E_1 be the event that the first chip is white and E_2 be the event that the first chip is black. Also let A be the event the second chip is black. We have

$$P(E_1) = \frac{m}{m+n}$$

and

$$P(A|E_1) = \frac{m-1}{m+n-1},\ P(A|E_2) = \frac{m}{m+n-1}$$

Then, from the law of total probability,

$$\begin{aligned}P(A) &= P(A|E_1)P(E_1) + P(A|E_2)P(E_2)\\ &= \left(\frac{m-1}{m+n-1}\right)\left(\frac{m}{m+n}\right) + \left(\frac{m}{m+n-1}\right)\left(\frac{n}{m+n}\right)\\ &= \frac{m}{m+n}\end{aligned}$$

Chapter 4

Expecations, Moments, Characteristic Functions and Functions of Random Variables.

Let X be a random variable with c.d.f F such that

$$\int_{-\infty}^{\infty} |U(x)| dF(x) < \infty$$

Then we can define

$$E(U(X)) = \int_{-\infty}^{\infty} U(x) dF(x)$$

We denote the mean by $\mu = E(X)$, the kth moment by $\mu_k = E(X^k)$, and the variance by

$$\sigma^2 = E[(X - E(X)^2)] = E(X^2) - E(X)^2$$

with moment generating function

$$M(t) = E(e^{tX})$$

Definition 4.0.1 (Characteristic Functions). *Let X be a random variable. The characteristic function of X is defined by*

$$\phi(t) = E(\exp(itX))$$

where $i = \sqrt{-1}$. If X is discrete then

$$g(s) = E(s^X) = \sum_k s^k P(X = k)$$

is called the generating function.

Example. Show that

(i) $f(x) = \frac{1}{\pi(1+x^2)} I(x \in \mathbb{R})$ is a p.d.f.

(ii) $E(X^k)$ does not exist if $k \geq 1$.

Proof. It's clear that

$$\int_{-\infty}^{\infty} \frac{1}{\pi(1+x^2)} dx = \left[\frac{\tan^{-1}(x)}{\pi}\right]_{-\infty}^{\infty} = 1$$

Then,

$$\begin{aligned}
E(|X|^k) &= \int_{-\infty}^{\infty} \frac{|x|^k}{\pi(1+x^2)} dx \\
&= \int_0^{\infty} \frac{x^k}{\pi(1+x^2)} dx + \int_0^{\infty} \frac{x^k}{\pi(1+x^2)} dx \\
&= 2\int_0^{\infty} \frac{x^k}{\pi(1+x^2)} dx \\
&\geq \frac{2}{\pi} \int_1^{\infty} \frac{x^k}{1+x^2} dx \\
&\geq \frac{2}{\pi} \int_1^{\infty} \frac{x^k}{x^2} dx
\end{aligned}$$

This integral diverges to infinity when $k > 1$, therefore $E(|X|^k)$ does not exist. □

Example. Let Z be a random variable with p.d.f

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

Find the c.d.f fpr $X = \sigma Z + \mu$ for $\sigma > 0$ and $\mu \in \mathbb{R}$.

Solution. Note that $f(x) \geq 0$. Let

$$I = \int_{-\infty}^{\infty} e^{-x^2/2} dx$$

Then,

$$I^2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp(-(x^2+y^2)/2) dx dy$$

Using polar coordinates, set $x = r\cos\theta$, and $y = r\sin\theta$,

$$I^2 = \int_0^{\infty} \int_0^{2\pi} \exp(-r^2/2) r dr d\theta = 2\pi$$

Thus $I = \sqrt{2\pi}$. Now we can calculate

$$F(x) = P(\sigma Z + \mu \leq X) = P\left(Z \leq \frac{x-\mu}{\sigma}\right) = \int_{-\infty}^{\frac{x-\mu}{\sigma}} f(z) dz$$

Then taking the derivative, we have

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

4.0.1 Properties of Characteristic Functions

The moment generating functions of a random variable may not exist. For example, it can be shown that the moment generating function for the Cauchy distribution may not exist

$$\int_{-\infty}^{\infty} \frac{\exp(\theta x)}{\pi(1+x^2)} dx = \infty$$

However characteristic functions always exist.

Theorem 4.0.1. *If X is a random variable, then*

(i) $\phi(\theta) = E(\exp(i\theta X))$ *always exists, with* $\phi(0) = 1$ *and* $|\phi(\theta)| \leq 1$.

(ii) $\overline{\phi(\theta)} = \phi(\theta)$ *where* $\overline{\phi(\theta)}$ *is the complex conjugate.*

(iii) If X is symmetric, then $\phi(X) \in \mathbb{R}$.

(iv) $\phi(\theta)$ *has linearity, i.e*

$$\phi_{aX+B}(\theta) = E(\exp(i\theta(aX+b))) = \phi_X(i\theta a)\exp(i\theta b)$$

(v) *The characteristic function for any random variable X is uniformly continuous.*

(vi) *There is a 1 to 1 correspondence between a c.d.f of a random variable and its characteristic function, (Uniquness Theorem, without proof).*

(vii) *X_1 and X_2 are independent if and only if*

$$\phi(\theta_1, \theta_2) = E(\exp(i\theta_1 X_1 + i\theta_2 X_2)) = \phi_{X_1}(\theta_1)\phi_{X_2}(\theta_2)$$

Proof. (i) Set $U = \cos(\theta)$, and $V = \sin(\theta)$, then using

$$e^{i\theta X} = \cos(\theta X) + i\sin(\theta X)$$

We have

$$|E(\exp(i\theta X))|^2 = |E(U + iV)|^2 = E(U)^2 + E(V)^2 \leq E(U^2) + E(V^2) = 1$$

(ii) Using the linearity of expecations,

$$\overline{E(\cos(\theta X)) + iE(\sin(\theta X))} = E(\cos(\theta X)) - iE(\sin(\theta X)) = \phi(-\theta)$$

(iii) This follows from $X \stackrel{d}{=} -X$.

(iv) This follows from linearity of expecations,

$$E(\exp(i\theta(aX + b))) = \exp(i\theta b)E(\exp(i\theta aX))$$

□

Theorem 4.0.2. *Let X be a random variable with c.d.f F such that $E(X)$ exists. Then*

$$E(X) = \int_0^\infty (1 - F(x))dx - \int_0^\infty F(-x)dx$$

Proof. Assume $P(X \geq 0) = 1$, then

$$\begin{aligned} E(X) &= \int_0^\infty x dF(x) \\ &= \int_0^\infty \int_0^x dy dF(x) \\ &= \int_0^\infty \int_y^\infty dF(x) dy \\ &= \int_0^\infty (1 - F(y))dy \end{aligned}$$

In general, $X = X^+ - X^-$ where

$$\begin{aligned} X^+ &= \max(0, X) = \frac{X + |X|}{2} \\ X^- &= \max(0, -X) = \frac{-X + |X|}{2} \end{aligned}$$

Threfore, $E(X) = E(X^+) - E(X^-)$. So we can write

$$E(X) = \int_0^\infty P(\max(0, X) > x)dx - \int_0^\infty P(\max(0, -X) > x)dx$$

With

$$\begin{aligned} P(\max(0, X) > x) &= P\left(\frac{X + |X|}{2} > x\right) \\ &= P(|X| > 2x - X) \\ &= P(X > 2x - X \text{ or } X < -2x + X) = P(X > x) \\ P(X^- > x) &= P(|X| > 2x + X) \\ &= P(X < -x) \end{aligned}$$

□

Example. Let $r \geq 0$, and $X \geq 0$, then

$$\begin{aligned} E(X^r) &= \int_0^\infty P(X^r > x)dx \\ &= \int_0^\infty P(X > x^{1/r})dx \\ &= \int_0^\infty P(X > u)ru^{r-1}du && (u = x^{1/r}) \end{aligned}$$

Therefore

$$E(X^r) = \int_0^\infty P(X^r > u)ru^{r-1}$$

Example.

$$\begin{aligned} E(|X|) &= \int_0^\infty P(|X| > x)dx \\ &= \int_0^\infty (1 - P(|X| \leq x))dx \\ &= \int_0^\infty (1 - P(-x \leq X \leq x))dx \\ &= \int_0^\infty 1 - (F(x) - F(-x))dx \end{aligned}$$

4.1 Distribution of Functions of Random Variables

Our goal is to find a c.d.f or p.d.f for a function $Y = U(X)$ for a random variable X.

Example. Let X be a random variable with $f(x) = 2xI(0 \leq x \leq 1)$. Let $Y = X^2$. Find the p.d.f for Y.

Solution. We can compute the c.d.f directly

$$\begin{aligned} G(y) = P(Y \leq y) &= P(X^2 \leq y) \\ &= P(-\sqrt{y} \leq X \leq \sqrt{y}) \\ &= P(0 \leq X \leq \sqrt{y}) && (\text{Since } x \in [0,1]) \\ &= \int_0^{\sqrt{y}} 2xdx \\ &= \left[x^2\right]_0^{\sqrt{y}} = y \end{aligned}$$

Therefore $G(y) = y$, and $g(y) = G'(y) = 1I(0 \leq y \leq 1)$. This means $X^2 \sim \text{Unif}(0,1)$.

Example. Let $X \sim f(x) = e^{-x}I(x > 0)$. Find the p.d.f for $Y = (\ln X)^2$.

Solution. Similarly to the previous example,

$$\begin{aligned} G(y) = P((\ln X)^2 \leq y) &= P(|\ln X| \leq \sqrt{y}) \\ &= P(-\sqrt{y} \leq \ln X \leq \sqrt{y}) \\ &= P(e^{-\sqrt{y}} \leq X \leq e^{\sqrt{y}}) \\ &= P(e^{-\sqrt{y}} < X \leq e^{\sqrt{y}}) \\ &= \int_{-\sqrt{y}}^{e^{\sqrt{y}}} e^{-x}dx = e^{\sqrt{y}} - e^{-\sqrt{y}} \end{aligned}$$

Then

$$g(y) = G'(y) = \frac{1}{\sqrt{y}}e^{\sqrt{y}}I(0 < y < \infty)$$

Note. The following formula will be useful for the next examples

$$\frac{\Gamma(\alpha)}{\beta^\alpha} = \int_0^\infty x^{\alpha-1}\exp(-\beta x)dx$$

Example. Let $X_1, X_2 \overset{\text{iid}}{\sim} f(x) = \frac{1}{\Gamma(\alpha_i)} e^{-x} x^{\alpha_i - 1} I(x > 0)$. . Define $U = \frac{X_1}{X_1 + X_2}$. Find p.d.f for U.

Solution. Note that $0 \leq U \leq 1$ since $X_1 + X_2 \geq X_1$, so $U : [0, \infty) \times [0, \infty) \mapsto [0, 1]$. We start with the c.d.f of U. We need to have the joint distribution for X_1 and X_2, so

$$f(x_1, x_2) = \frac{1}{\Gamma(\alpha_1)\Gamma(\alpha_2)} e^{-x_1} e^{-x_2} x_1^{\alpha_1 - 1} x_2^{\alpha_2 - 1} I(x_1, x_2 > 0)$$

Now

$$\begin{aligned} G(u) &= P\left(\frac{X_1}{X_1 + X_2} \leq u\right) \\ &= P(X_1 \leq u(X_1 + X_2)) \\ &= P(X_1 - uX_1 \leq uX_2) \\ &= P\left(X_1 \leq \frac{uX^2}{1 - u}\right) \\ &= \int_0^\infty \int_0^{\frac{x_2 u}{1-u}} \frac{e^{-x_1 - x_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} x_1^{\alpha_1 - 1} x_2^{\alpha_2 - 1} dx_1 dx_2 \end{aligned}$$

Differentiating this function we get

$$g(u) = G'(u) = \frac{\Gamma(\alpha_1 + \alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} u^{\alpha_1 - 1}(1 - u)^{\alpha_2 - 1}$$

Note. From the beta distribution with parameters α_1, α_2 we get the useful formula

$$\int_0^1 \frac{\Gamma(\alpha_1 + \alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} u^{\alpha_1 - 1}(1 - u)^{\alpha_2 - 1} du = 1$$

$$\implies \int_0^1 u^{\alpha_1 - 1}(1 - u)^{\alpha_2 - 1} du = \frac{\Gamma(\alpha_1 + \alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)}$$

Another way to solve the previous example is to introduce another variable $V = X + Y$. The joint distribution for X and Y

$$f(x, y) = \frac{1}{\Gamma(\alpha_1)\Gamma(\alpha_2)} e^{-(x+y)} x^{\alpha_1 - 1} y^{\alpha_2 - 1}$$

Then we can solve for X and Y in terms of U and V and we get $X = UV$, $Y = V(1 - U)$. Then the Jacobian for this transformation is

$$J = \begin{vmatrix} \frac{\partial X}{\partial u} & \frac{\partial X}{\partial v} \\ \frac{\partial Y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} \begin{vmatrix} v & u \\ -v & 1 - u \end{vmatrix} = v$$

Now our joint p.d.f for U, V is

$$\begin{aligned} g(u, v) &= f(x(u, v), y(u, v))|J| = f(uv, v(1 - u))|J| \\ &= \frac{1}{\Gamma(\alpha_1)\Gamma(\alpha_2)} e^{-v}(uv)^{\alpha_1 - 1}(v(1 - u))^{\alpha_2 - 1} v I(0 < v < 1, u > 0) \end{aligned}$$

We don't want the joint p.d.f for U and V, we want it for U so we can integrate over V

$$\begin{aligned} g(u) = \int_0^\infty g(u, v) dv &= \int_0^\infty \frac{1}{\Gamma(\alpha_1)\Gamma(\alpha_2)} e^{-v}(uv)^{\alpha_1 - 1}(v(1 - u))^{\alpha_2 - 1} v dv \\ &= \frac{1}{\Gamma(\alpha_1)\Gamma(\alpha_2)} u^{\alpha_1 - 1}(1 - u)^{\alpha_2 - 1} \int_0^\infty e^{-v} v^{\alpha_1 + \alpha_2} dv \end{aligned}$$

Notice that the integral is $\Gamma(\alpha_1 + \alpha_2)$ so we get

$$g(u) = \frac{\Gamma(\alpha_1 + \alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} u^{\alpha_1 - 1}(1 - u)^{\alpha_2 - 1}$$

Example. Let $X \sim \text{Unif}(0, 1)$. Find the p.d.f for

(i) $Y = a + (b - a)X$ where $a < b$

(ii) $W = \tan\left(\frac{\pi(2X-1)}{2}\right)$

Solution.

(i) Let G denote the c.d.f for Y, then

$$\begin{aligned} G(y) = P(Y \le y) &= P(a + (b-a)X \le y) \\ &= P((b-a)X \le y - a) \\ &= P\left(X \le \frac{y-a}{b-a}\right) \\ &= \frac{y-a}{b-a} \end{aligned}$$

Then the p.d.f is

$$g(y) = G'(y) = \frac{1}{b-a} I(a \le y \le b)$$

Therefore, $Y \sim \text{Unif}(a,b)$.

(ii) Let H denote the c.d.f for W, then

$$\begin{aligned} H(w) = P(W \le w) &= P\left(\tan\frac{\pi(2X-1)}{2} \le 2\right) \\ &= P\left(\frac{\pi(2X-1)}{2} \le \arctan(w)\right) \\ &= P\left(X \le \frac{2\arctan(w)+1}{2\pi}\right) \\ &= \frac{2\arctan(w)+1}{2\pi} \end{aligned}$$

Then the p.d.f is

$$h(w) = H'(w) = \frac{1}{\pi(1+w^2)} I(w \in \mathbb{R})$$

Thus W has a Cauchy distribution.

Example. Let $W \sim N(0,1)$, $V \sim \chi^2(r)$ and V is independent of W. Define

$$T = \frac{W}{\sqrt{V/r}} \sim t(r)$$

Find the p.d.f for $t(r)$.

Solution.

$$\begin{aligned} G(t) = P(T \le t) &= P\left(\frac{W}{\sqrt{v/r}} \le t\right) \\ &= P\left(W \le t\sqrt{v/r}\right) \end{aligned}$$

W and V are independent, so their joint distribution is

$$f(x,y) = \frac{1}{\sqrt{2\pi}} e^{-w^2/2} \frac{1}{\Gamma(r/2)2^{r/2}} e^{-v/2} v^{r/2-1} I(v > 0, w \in \mathbb{R})$$

Note. Recall this very important formula

$$\int_0^\infty e^{\beta u} u^{\alpha-1} du = \frac{\Gamma(\alpha)}{\beta^\alpha}$$

Now we can compute $G(t)$

$$
\begin{aligned}
G(t) = P(W \le t\sqrt{v/r}) &= \frac{1}{\sqrt{2\pi}\Gamma(r/2)2^{r/2}} \int_0^\infty \int_{-\infty}^{t\sqrt{v/r}} e^{-w^2/2} e^{-v/2} v^{r/2-1} dw dv \\
G'(t) = g(t) &= \frac{1}{\sqrt{2\pi}\Gamma(r/2)2^{r/2}} \int_0^\infty \left(\frac{d}{dt} \int_0^{t\sqrt{v/r}} e^{-w^2/2} dw \right) e^{-v/2} v^{r/2-1} dv \\
&= \frac{1}{\sqrt{2\pi}\Gamma(r/2)2^{r/2}} \int_0^\infty \sqrt{v/r} e^{-t^2 v/2r} e^{-v/2} v^{r/2-1} dv \\
&= \frac{1}{\sqrt{2\pi r}\Gamma(r/2)2^{r/2}} \int_0^\infty e^{-v/2(1+t^2/r)} v^{r/2-1+1/2} dv \\
&= \frac{1}{\sqrt{2\pi r}\Gamma(r/2)2^{r/2}} \frac{\Gamma\left(\frac{r+1}{2}\right)}{\left(\frac{1}{2}\left(1+\frac{t^2}{r}\right)\right)^{(r+1)/2}} \\
&= \frac{2^{r/2}\Gamma\left(\frac{r+1}{2}\right)}{\Gamma(r/2)\sqrt{\pi r}} \left(\frac{1}{1+\frac{t^2}{r}} \right)^{(r+1)/2}
\end{aligned}
$$

Example. Let $X, Y \overset{\text{iid}}{\sim} \text{Unif}[0,1]$. Let $U = X+Y$, $V = X-Y$. Find the joint and marginal p.d.f's for U and V

Solution. We start by calculating X and Y in terms of U and V, solving the 2 equations we get

$$X = \frac{U+V}{2}, Y = \frac{U-V}{2}$$

Now we can compute the Jacobian

$$J = \begin{vmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{vmatrix} = -\frac{1}{2}$$

Then the joint p.d.f of U and V is

$$g(u,v) = f(x(u,v), y(u,v))|J| = \frac{1}{2} I((0 \le u \le 1, |v| \le u) \text{ or } (1 \le u \le v, |v| \le 2))$$

For marginal p.d.f's,

$$
\begin{aligned}
g(u) = \int g(u,v)dv &= \begin{cases} \displaystyle\int_{-u}^{u} \frac{1}{2} dv = u & 0 \le u \le 1 \\ \displaystyle\int_{-(2-u)}^{2-u} \frac{1}{2} dv = 2-u & 1 \le u \le 2 \end{cases} \\
&= uI(0 \le u \le 1) + (2-u)I(1 \le u \le 2)
\end{aligned}
$$

Then similarly for V,

$$g(v) = \int g(u,v)du = \int_{|v|}^{2-|v|} \frac{1}{2} du = \frac{1-|v|}{2} I(-1 \le v \le 1)$$

Example. Let $U_1, U_2 \overset{\text{iid}}{\sim} \text{Unif}[0,1]$. Then define

$$X_1 = \cos(2\pi U_1)\sqrt{-2\ln U_2} \; X_2 = \cos(2\pi U_1)\sqrt{-2\ln U_2}$$

Solution. We start by finding U_1, U_2 in terms of X_1 and X_2. We can do that as follows

$$
\begin{aligned}
X_1^2 + X_2^2 &= \cos^2(2\pi U_1)(-\ln U_2) + \sin^2(2\pi U_1)(-\ln U_2) \\
&= (-\ln U_2)(\cos^2(2\pi U_1) + \sin^2(2\pi U_1)) \\
&= -\ln U_2 \cdot 1 \implies U_2 = \exp(-(X_1^2 + X_2^2))
\end{aligned}
$$

Then for U_1,

$$\frac{X_2}{X_1} = \frac{\sin(2\pi U_1)}{\cos(2\pi U_1)} = \tan(2\pi U_1) \implies U_1 = \frac{1}{2\pi} \arctan\left(\frac{X_2}{X_1}\right)$$

Now we can calculate the Jacobian

$$J = \begin{vmatrix} \frac{\partial U_1}{\partial x_1} & \frac{\partial U_1}{\partial x_2} \\ \frac{\partial U_2}{\partial x_1} & \frac{\partial U_2}{\partial x_2} \end{vmatrix} = \frac{1}{2\pi} \exp\left(-\frac{x_1^2 + x_2^2}{2}\right)$$

Unif$[0,1]$ has p.d.f 1, so our joint p.d.f for X_1, X_2 is

$$g(x_1,x_2) = f(u_1(x_1,x_2), u_2(x_1,x_2))|J| = \frac{1}{2\pi}\exp\left(-\frac{x_1^2+x_2^2}{2}\right)$$

We can split this up as

$$g(x_1,x_2) = \left(\frac{1}{\sqrt{2\pi}}e^{-x_1^2/2}\right)\left(\frac{1}{\sqrt{2\pi}}e^{-x_2^2/2}\right)$$

These are 2 standard normal distributions, so this tells us how to simulate normal distribution from Unif$(0,1)$.

4.2 Order Statistics

Definition 4.2.1. *Let $X_1, \dots, X_n \overset{\text{iid}}{\sim} f(x)$ be a random sample. Then the order statistics for this sample are*

$$Y_1 = \min(X_1,\dots,X_n), Y_2 = \min\left(\{X_1,\dots X_n\}\setminus Y_1\right), \dots, Y_n = \max(X_1,\dots,X_n)$$

So

$$Y_1 < Y_2 < \cdots < Y_{n-1} < Y_n$$

In otherwords, $Y_1, \dots, Y_n$ are $X_1, \dots, X_n$ sorted with Y_1 being the minimum and Y_n being the maximum.

Example. Consider the sample

$$X_1 = 3.1, X_2 = 4.5, X_3, 3.4, X_4 = 4.1, X_5 = 2$$

Then its order statistics are

$$Y_1 = 2 < Y_2 = 3.1 < Y_3 = 3.4 < Y_4 = 4.1 < Y_5 = 4.5$$

It's clear that $Y_3 = 3.4$ is the median. In general the median is given as

$$\begin{cases} Y_{(n+1)/2} & \text{If } n \text{ is odd} \\ \frac{1}{2}\left(Y_{n/2} + Y_{(n/2)+1}\right) & \text{If } n \text{ is even} \end{cases}$$

It is important in many statistical problems to find the p.d.f for functions of order statistics, to find the p.d.f for Y_i, we can simply write

$$g_i(y) = \frac{n!}{(i-1)!(n-1)!}(F(y))^{i-1}f(y)(1-F(y))^{n-i}$$

Then the joint p.d.f for Y_i, Y_j is

$$g_{ij}(u,v) = \frac{n!}{(i-1)!(j-i-1)!(n-j)!}F(u)^{i-1}(F(v)-F(u))^{j-i-1}(1-F(v))^{n-j}f(u)f(v)$$

with $-\infty < u < v < \infty$.

Example. Let $X_1, X_2, X_3 \overset{\text{iid}}{\sim} f(x)$ for a continuous p.d.f f. Let $Y_1 < Y_2 < Y_3$ be the order statistics. Find the joint p.d.f for Y_1, Y_2, Y_3.

Solution. Notice that there are 6 possible outcomes in this case, order from smallest to largest we have

$$Y_1 = X_1, Y_2 = X_2, Y_3 = X_3; Y_1 = X_2, Y_2 = X_1, Y_3 = X_3$$

$$Y_1 = X_1, Y_2 = X_3, Y_3 = X_2; Y_1 = X_3, Y_2 = X_1, Y_3 = X_2$$

$$Y_1 = X_2, Y_2 = X_3, Y_3 = X_1; Y_1 = X_3, Y_2 = X_2, Y_3 = X_1$$

Consider the last case where $Y_1 = X_3$, $Y_2 = X_2$, $Y_3 = X_1$, then the Jacobian for this transformation is

$$|J| = \begin{vmatrix} \frac{\partial X_1}{y_1} & \frac{\partial X_1}{y_2} & \frac{\partial X_1}{y_3} \\ \frac{\partial X_2}{y_1} & \frac{\partial X_3}{y_2} & \frac{\partial X_2}{y_3} \\ \frac{\partial X_3}{y_1} & \frac{\partial X_3}{y_2} & \frac{\partial X_3}{y_3} \end{vmatrix} = \begin{matrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{matrix} = 1$$

Notice that for any combination, $J = \pm 1 \implies |J| = 1$. Thus the joint p.d.f is the some of each of these cases

$$g(y_1,y_2,y_3) = f(y_1,y_2,y_3)\cdot 1 + \cdots + f(y_3,y_2,y_1)\cdot 1$$

This gives us

$$g(y_1, y_2, y_3) = 3!f(y_1)f(y_2)f(y_3)I(y_1 < y_2 < y_n)$$

This holds with our general formula for n order statistics

$$g(y_1, \dots y_n) = n!f(y_1)\cdots f(y_n)I(y_1 < \cdots < y_n)$$

Example. Let $X_1, X_2 \overset{\text{iid}}{\sim} f(x)$. Then

$$Y_1 = \min(X_1, X_2), Y_2 = \min(X_1, X_2)$$

Here again we can break this up into cases where $Y_1 = X_1$, $Y_2 = X_2$, or $Y_1 = X_2$, $Y_2 = X_1$. Then our Jacobians are

$$J_1 = \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} = 1$$

$$J_2 = \begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix} = -1$$

So, the joint p.d.f is

$$g(y_1, y_2) = f(x_1, x_2)|J_1| + f(x_2, x_1)|J_2| = 2f(y_1)f(y_2)I(y_1 < y_2)$$

Example. Let $X_1, \dots, \overset{\text{iid}}{\sim} f$ with continuous c.d.f F. Let $Y_1 < Y_2 < \cdots < Y_i < \cdots Y_n$ be the order statistics. Find the p.d.f for Y_i.

Solution. We want to derive the general equation for the p.d.f of any Y_i. Notice that

$$P(y < Y_i < y + dy) = g(y)dy$$

Intuitivively, dy is *very* small, so the area of this region below the curve will be the p.d.f of Y_i multiplied by the infinitesimal width dy. To the left of Y_i, we'll have $i-1$ order statistics, and $n-i$ to the right. For any of the X's,

$$P(X \leq y) = F(y),\ P(X > y) = 1 - F(y)$$

In otherwords, the probability that an observation falls below y is $F(y)$ and above y is $1 - F(y)$. Now, we have n total observations, with $i-1$ below y, and $n-i$ above, so the total number of ways to arrange these is

$$\frac{n!}{(i-1)!(n-i)!}$$

We have $i-1$ (i.i.d) falling below y with probability $F(y)^{i-1}$, one at y with probability $f(y)dy$, and $n-i$ (i.i.d) above y with probability $(1 - F(y))^{n-1}$, therefore we have the equation

$$P(y \leq Y_i \leq y + dy) = g(y)dy = \frac{n!}{(i-1)!(n-i)!}(F(y))^{i-1}f(y)dy(1 - F(y))^{n-i}$$

Dividing by dy on both sides gives us

$$g(y) = \frac{n!}{(i-1)!(n-i)!}(F(y))^{i-1}f(y)(1 - F(y))^{n-i}$$

Example. Let $X_1, \dots, X_n \sim f(x) = 2xI(0 \leq x \leq 1)$. Let $Y_1 < Y_2 < Y_3 < Y_4 < Y_5$ be the orderered statistics.

(i) Find the p.d.f for the median Y_3.

(ii) Find the joint p.d.f for Y_1,Y_5.

(iii) Find the joint p.d.f for Y_1, Y_3, Y_5.

Solution.

(i) To use the equation we derived in the previous example, we must first find the c.d.f

$$F(y) = P(X \leq y) = \int_0^y 2x dx = y^2$$

Then, let $g_3(y)$ denote the p.d.f for Y_3,

$$\begin{aligned} g_3(y) &= \frac{5!}{2!2!}(F(y))^2 f(y)(1 - F(y))^2 I(0 \leq y \leq 1) \\ &= 30y^4(2y)(1 - y^2)^2 I(0 \leq y \leq 1) \\ &= 60y^5(1 - y^2)^2 I(0 \leq y \leq 1) \end{aligned}$$

(ii) We want to find the p.d.f for $Y_1 = \min(X_1, \dots, X_n)$, $Y_5 = \max(X_1, \dots, X_n)$. Let $g_{15}(y_1, y_5)$ denote the p.d.f for Y_1,Y_5. We the c.d.f

$$P(y_1 \le X \le y_5) = \int_{y_1}^{y_5} 2x dx = y_5^2 - y_1^2$$

Then, using the forumla we have

$$\begin{aligned} g_{15}(y_1, y_5) &= \frac{5!}{3!} f(y_1) f(y_5)(F_5 - F)(y_1) I(0 < y_1 < y_5 < 1) \\ &= 80 y_1 y_5 (y_5^2 - y_1^2)^3 I(0 < y_1 < y_5 < 1) \end{aligned}$$

(iii) Similarly, for the joint p.d.f for Y_1, Y_3, Y_5, we have

$$\begin{aligned} g(y_1, y_3, y_5) &= \frac{5!}{1!1!1!1!1!} f(y_1) f(y_3) f(y_5) (F(y_3) - F(y_1))^1 (F(y_5) - F(y_1))^1 \\ &= 120(2y_1)(2y_3)(2y_5)(y_3^2 - y_1^2)(y_5^2 - y_3^2) I(0 < y_1 < y_3 < y_5) \end{aligned}$$

Example. Let $X_1, \dots, X_n \overset{\text{iid}}{\sim} f(x) = e^{-x} I(x > 0)$. Find the joint distribution for Y_1, Y_n.

Solution. First we find the c.d.f

$$F(x) = \int_0^x e^{-t} dt = 1 - e^{-x} I(x \ge 0)$$

Then,

$$\begin{aligned} g(y_1, y_n) &= \frac{n!}{(n-2)!} f(y_1) f(y_n) (F(y_n) - F(y_1))^{n-2} I(y_n > y_1 > 0) \\ &= n(n-1) e^{-y_1} e^{-y_n} (1 - e^{-y_n} - 1 + e^{-y_1})^{n-2} I(y_n > y_1 > 0) \\ &= n(n-1) e^{-(y_1 + y_n)} (e^{-y_1} - e^{-y_n})^{n-2} I(y_n > y_1 > 0) \end{aligned}$$

Example. Let $X_1, \dots X_{2n+1}$ Unif$[\theta_1, \theta_2]$. Find the p.d.f for the median Y_{n+1}.

Solution. The p.d.f for this distribution is

$$f(x) = \frac{1}{\theta_2 - \theta_1} I(\theta_1 \le x \le \theta_2)$$

and the c.d.f is

$$F(x) = \int_{\theta_1}^x \frac{1}{\theta_2 - \theta_1} du = \frac{x - \theta_1}{\theta_2 - \theta_1}$$

Since Y_{n+1} is the median, we have n observations to the left and n to the right of Y_{n+1}, so the p.d.f is

$$\begin{aligned} g(y) &= \frac{(2n+1)}{n!n!} f(y) (F(y))^n (1 - F(y))^n \\ &= \frac{(2n+1)!}{(n!)^2} \frac{1}{\theta_2 - \theta_1} \left(\frac{y - \theta_1}{\theta_2 - \theta_1} \right)^n \left(1 - \frac{y - \theta_1}{\theta_2 - \theta_1} \right)^n I(y \in [\theta_1, \theta_2]) \end{aligned}$$

Example. Let Y_1, Y_2, Y_3 be the order statistics of a random sample from Unif$(0, 1)$. We would liek to find the p.d.f for the range

$$Z = Y_3 - Y_1$$

Solution. Since these are samples for uniform distribution, $f(x) = 1$ and $F(x) = x$. Then the joint p.d.f for Y_1, Y_3 is

$$g(y_1, y_3) = 3! f(y_1) f(y_3) (F(y_3) - F(y_1))^1 = 6(y_3 - y_1) I(0 \le y_1 < y_3 \le 1)$$

Then, we can define another random variable $Z_2 = Y_3$. So

$$Z_1 = Y_3 - Y_1, Z_3 = Y_3 \implies Y_1 = Z_1 - Z_2$$

Now we can compute the joint distribution for these random variables. It's easy to see that $|J| = 1$, therefore

$$g(z_1, z_2) = f(y_1, y_2)|J| = 6(z_2 - z_2 + z_1) = 6z_1 I(0 \le z_1 < z_2 \le 1)$$

Then we want the distribution for Z_2, so we can integrate over Z_2,

$$g(z_1) = \int_{z_1}^1 6 z_1 dz_2 = 6 z_1 (1 - z_1) I(0 < z_1 < 1)$$

4.3 Independence of Random Variables

Recall that if $(X, Y) \sim f(x, y)$, with mariginal p.d.f's $f_1(x)$ and $f_2(y)$, we define the conditional p.d.f for $Y|X = x$ and $X|Y = y$ as

$$f(y|x) = \frac{f(x,y)}{f_1(x)}, f(x|y) = \frac{f(x,y)}{f_2(y)}$$

If X and Y are independent, then

$$f(x|y) = f_1(x), f(y|x) = f_2(y) \implies f(x,y) = f_1(x)f_2(x)$$

To check for Independence, finding the marginal distributions may not be necessary. In general, if

$$f(x,y) = u(x)v(y)$$

Where u and v are two functions, not necessarily the marginals of X and Y,then we can still say X and Y are independent. To see this, notice that for independence we need to check that

$$f(x,y) = f_1(x)f_2(y)$$

Suppose $f(x, y) = u(x)v(y)$, then we can calculate

$$f_1(x) = \int_{-\infty}^{\infty} u(x)v(y)dy = u(x)\int_{-\infty}^{\infty} v(y)dy = c_1u(x)$$

Where c_1 is the constant obtained from evaluating the integral of $v(y)$. Then similarly,

$$f_1(x) = \int_{-\infty}^{\infty} u(x)v(y)dx = v(y)\int_{-\infty}^{\infty} u(x)dx = c_2v(y)$$

Then, $f(x, y)$ is a p.d.f so its integral is 1, thus

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} u(x)v(y)dxdy = c_1c_2 = 1$$

Therefore,

$$f(x,y) = u(x)v(y) = c_1c_2u(x)v(y) = f_1(x)f_2(y)$$

and X is independent of Y.

Definition 4.3.1. *If (X, Y) is a random vector with p.d.f $f(x, y)$, then the conditional expecation is defined as*

$$E(Y|X = x) = \int_{-\infty}^{\infty} yf(y|x)dy, E(X|Y = y) = \int_{-\infty}^{\infty} xf(x|y)dx$$

Similarly the variance is defined as

$$\text{Var}(Y|X = x) = E(Y^2|X = x) - E(Y|X = x)^2$$

$$\text{Var}(X|Y = y) = E(X^2|Y = y) - E(X|Y = y)^2$$

We usually call $E(Y|X = x)$ the regression of Y on $X = x$.

Example. Let $Y_1 < \cdots Y_5$ be the order statistics of a random sample of size $n = 5$ from an exponentional distribution with mean 1. Show that $Z_1 = Y_2$ and $Z_2 = Y_4 - Y_2$ are independent random variables.

Solution. $Y_1, \ldots, Y_5$ are samples from exponential with mean 1, so

$$f(x) = e^{-x}I(x \geq 0), F(x) = 1 - e^{-x}$$

We want to find the joint p.d.f of $Z_1 = Y_2$, and $Z_2 = Y_4 - Y_2$ and show that it can be written as a product of 2 functions $u(z_1)$,$v(z_2)$. First, we find the joint p.d.f for Y_2, Y_4.

$$\begin{aligned}
g_{24}(y_2, y_4) &= 5!f(y_2)f(y_4)F(y_2)(1 - F(y_4))(F(y_4) - F(y_2))I(y_4 > y_2 > 0) \\
&= 120e^{-y_2}e^{-y_4}(1 - e^{-y_2})(e^{-y_4})(1 - e^{-y_4} - 1 + e^{-y_2})I(y_4 > y_2 > 0) \\
&= 120e^{-y_2}e^{-y_4}(1 - e^{-y_2})(e^{-y_4})(e^{-y_2} - e^{-y_4})I(y_4 > y_2 > 0)
\end{aligned}$$

Now we want to find the joint p.d.f for $Z_1 = Y_2$, and $Z_2 = Y_4 - Y_2$. We rearrange these to find Y_2 and Y_4 in terms of Z_1, Z_2 to get

$$Y_2 = Z_1, Y_4 = Z_2 + Z_1$$

It's easy to calculate the Jacobian $J = 1$, so

$$\begin{aligned} h(z_1, z_2) &= 120e^{-z_1}e^{-z_2-z_1}(1 - e^{-z_1})(e^{-z_2-z_1})(e^{-z_1} - e^{-z_2-z_1}) \\ &= 120e^{-z_1}e^{-z_2}e^{-z_1}(1 - e^{-z_1})(e^{-z_2}e^{-z_1})(e^{-z_1} - e^{-z_1}e^{-z_2}) \\ &= 120e^{-z_1}e^{-z_2}e^{-z_1}(1 - e^{-z_1})(e^{-z_2}e^{-z_1})e^{-z_1}(1 - e^{-z_2}) \\ &= 120e^{-4z_1}(1 - e^{-z_1})e^{-2z_2}(1 - e^{-z_2})I(z_1 > 0)I(z_2 > 0) \end{aligned}$$

Now we can set $u(z_1) = 120e^{-4z_1}(1 - e^{-z_1})$, and $v(z_2) = e^{-2z_2}(1 - e^{-z_2})$, and we have $h(z_1, z_2) = u(z_1)v(z_2)$.

Example. Let $Y_1 < \cdots Y_5$ be the order statistics of a random sample of size 5 coming from the distribution

$$f(x) = 3x^2 I(0 < x < 1)$$

Show that $Z_1 = Y_2/Y_4$ is independent from $Z_2 = Y_4$.

Solution. First we need the c.d.f,

$$F(x) = \int_0^x 3t^2 dt = x^3$$

Now we can find the joint distribution for Y_2, Y_4,

$$\begin{aligned} g(y_2, y_4) &= 5! f(y_2)f(y_4)F(y_2)(1 - F(y_4))(F(y_4) - F(y_2))I(0 < y_2 < y_4 < 1) \\ &= 120(3y_2^2)(3y_4^2)(y_2^3)(1 - y_4^3)(y_4^3 - y_2^3)I(0 < y_2 < y_4 < 1) \\ &= 1080y_2^5 y_4^2(1 - y_4^3)(y_4^3 - y_2^3) \end{aligned}$$

Then we write Y_2 and Y_4 in terms of Z_1, Z_2,

$$Z_1 = \frac{Y_2}{Y_4}, Z_2 = Y_4 \implies Y_2 = Z_1 Z_2$$

The Jacobian is easy to calculate $|J| = z_2$. Now the joint p.d.f for Z_1, Z_2 is

$$\begin{aligned} h(z_1, z_2) &= g(y_2, y_4)|J| \\ &= 1080(z_1 z_2)^5 z_2^2(1 - z_2^3)(z_2^3 - z_1^3 z_2^3)z_2 I(0 \le z_1)I(0 \le z_1 \le 1, 0 \le z_2 \le 1) \\ &= 1080z_1^5 z_2^7(1 - z_2^3)z_2^3(1 - z_1^3)z_2 I(0 \le z_1 \le 1, 0 \le z_2 \le 1) \\ &= 1080z_1^5(1 - z_1^3)z_2^{11}(1 - z_2^3)^2 I(0 \le z_1 \le 1, 0 \le z_2 \le 1) \\ &= u(z_1)v(z_2) \end{aligned}$$

Chapter 5

Convergence of Random Variables

Definition 5.0.1. *Let $\{X_n\}$ be a sequence of random variables. We define convergence as*

(i) We say X_n converges to X in probability if

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

We denote this by $X_n \xrightarrow{P} X$.

(ii) We say that X_n converges to X almost surely if

$$P(|X_n - X| > \epsilon \ i.o) = 0$$

This is denoted by $X_n \xrightarrow{a.s} X$.

(iii) We say X_n converges to X in L^p if

$$\lim_{n\to\infty} E(|X_n - X|^p) = 0$$

This is denoted by $X_n \xrightarrow{L^p} X$.

(iv) We say X_n converges in distribution to X if

$$\lim_{n\to\infty} P(X_n \le x) = P(X \le x) = F(x)$$

For all $x \in C_F$ where C_F is the set of continuity points of F. This is denoted by $X_n \xrightarrow{d} X$.

Example. Let $X_i, \dots, X_n \overset{\text{iid}}{\sim} f(x)$ with $E(X_i) = \mu$. Use sample mean $\bar{X}$ to estimate μ. Assume $\text{Var}(X_i) = \sigma^2 < \infty$. Prove $\bar{X} \xrightarrow{L^p} \mu$ for $p \in [1, 2]$. **Solution.** We have

$$\bar{X} \xrightarrow{L^2} \mu \iff E[(\bar{X} - \mu)^2] \to 0$$

So we want to show that the variance converges to 0. We have that

$$E(\bar{X}) = \frac{E(X_1) + \cdots + E(X_n)}{n} = \frac{n\mu}{n} = \mu$$

The variance is

$$\text{Var}(\bar{X}) = E[(\bar{X} - \mu)^2] = \text{Var}\left(\frac{X_1 + \cdots + X_n}{n}\right) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

Now taking the limit

$$\lim_{n\to\infty} \frac{\sigma^2}{n} == 0$$

To prove $E[(\bar{X} - \mu)^p] \to 0$, for $1 \le p \le 2$, we'll use Lyapunov's inequality

$$(E|u|^p)^{1/p} \le (E|u|^q)^{1/q}$$

for $q \ge p$. If $E(|\bar{X} - \mu|^2) \to 0$, then

$$(E|\bar{X} - \mu|^p)^{1/p} \le (E|\bar{X} - \mu|^2)^{1/2}$$

Since $E(\bar{X} - \mu)^2 \to 0$, then $E|X - \mu|^p \to 0$ for $1 \le p \le 2$. Thus $\bar{X} \xrightarrow{L^2} \mu$.

Theorem 5.0.1. *Let $(X_n)_{n\geq 1}$ be a sequence of random variables such that $E(X_n) = \mu$ and $\mathrm{Var}(X_n) \to 0$ as $n \to \infty$. Then $X_n \xrightarrow{P} \mu$.*

Proof. Suppose $E(X_n) = \mu$ and $\mathrm{Var}(X_n) \to 0$, then we use Markov's inequality,

$$0 \leq P(|X_n - \mu| \geq \epsilon) \leq \frac{\mathrm{Var}(X_n)}{\epsilon^2}$$

The right side converges to 0 as $n \to \infty$. Therefore,

$$\lim_{n\to\infty} P(|X_n - \mu| \geq \epsilon) = 0$$

Thus $X_n \xrightarrow{P} \mu$. $\square$

Example. Let X_n be a random variable with p.d.f

$$f_n(x) = \begin{cases} 1 & x = 2 + \frac{1}{n} \\ 0 & x \neq 2 + \frac{1}{n} \end{cases}$$

Check to see if there i sa limiting distribution for X_n.

Solution. We can see that

$$E(X_n) = 2 + \frac{1}{n} \to 2$$

$$\mathrm{Var}(X_n) = E\left(X_n - 2 - \frac{1}{n}\right)^2 = 0$$

So its clear that $X_n \xrightarrow{P} 2$ and $X_n \xrightarrow{L^p} 2$. We want the limiting distribution,

$$F_n(x) = P(X_n \leq x) \begin{cases} 0 & x < 2 + \frac{1}{n} \\ 1 & x \geq 2 + \frac{1}{n} \end{cases}$$

$$\lim_{n\to\infty} F_n(x) = F(x) = \begin{cases} 0 & x < 2 \\ 1 & x \geq 2 \end{cases}$$

So $F_n(x) \to F(x)$ except for $x = 2$, but there is a jump at 2 so $2 \notin C_F$ so we don't care about $x = 2$.

Theorem 5.0.2. *Let $\{X_i\}$ be a sequence of random variables. Let $c \in$ real, then*

$$X_n \xrightarrow{P} c \implies X_n \xrightarrow{d} c$$

Proof. Suppose that $X_n \xrightarrow{P} c$. Then $\forall \epsilon > 0$,

$$P(|X_n - c| < \epsilon) = 1 - P(|X_n - c| \geq \epsilon) = 0 \implies \lim_{n\to\infty} P(|X_n - c| < \epsilon) = 1$$

$$\begin{aligned} P(|X_n - c| < \epsilon) &= P(-\epsilon < X_n - c < \epsilon) \\ &= P(c - \epsilon < X_n < c + \epsilon) \\ &= F_n((c+\epsilon)^-) - F_n(c - \epsilon) \end{aligned}$$

Then

$$1 = \lim_{n\to\infty} P(|X_n - c| < \epsilon) = \lim_{n\to\infty} \left(F_n((c+\epsilon)^-) - F_n(c-\epsilon)\right)$$

So we must have that $\lim_{n\to\infty} F_n((c+\epsilon)^-) = 1$, and $\lim_{n\to\infty} F_n(c-\epsilon) = 0$. Thus we can conclude

$$\lim_{n\to\infty} F_n(x) = \begin{cases} 0 & x < c \\ 1 & x > c \end{cases} = F(x)$$

Therefore $X_n \xrightarrow{d} c$. $\square$

Example. From our previous example with $\bar{X}$, with $E(X_1) = \mu$ and $\text{Var}(X_i) = \sigma^2 < \infty$, we have $\bar{X} \xrightarrow{L^p} \mu$, $\bar{X} \xrightarrow{P} \mu$. So we also have $\bar{X} \xrightarrow{d} \mu$.

Theorem 5.0.3. *Let X_n be a sequence of random variables, then*

$$X_n \xrightarrow{d} c \implies X_n \xrightarrow{P} c$$

Proof. We know that since $X_n \xrightarrow{d} c$,

$$\lim_{n\to\infty} F_n(x) = \begin{cases} 0 & x < c \\ 1 & x \geq c \end{cases}$$

For all $x \neq c$. Now we want to find

$$\begin{aligned} &\lim_{n\to\infty} P(|X_n - c| < \epsilon) = \lim_{n\to\infty} F_n((c+\epsilon)^-) - \lim_{n\to\infty} F_n(c-\epsilon) \\ \implies &\lim_{n\to\infty} P(|X_n - c| < \epsilon) = 1 \\ \implies &\lim_{n\to\infty} P(|X_n - c| \geq \epsilon) = 0 \end{aligned}$$

Thus $X_n \xrightarrow{P} c$. So we have shown $X_n \xrightarrow{P} c \iff X_n \xrightarrow{d} c$ for a constant $c \in \mathbb{R}$. □

Example. Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{Unif}[0, \theta]$ with density

$$f(x) = \frac{1}{\theta} I(0 \leq x \leq \theta)$$

Let

$$Y_n = \max(X_1, \ldots, X_n)$$

(i) Find c.d.f and p.d.f for Y_n.

(ii) Prove $Y_n \xrightarrow{P} \theta$.

(iii) Find the limiting distribution for $U_n = n(\theta - Y_n)$.

Solution.

(i) We can compute the c.d.f directly

$$\begin{aligned} G_n(y) = P(Y_n \leq y) &= P(\max(X_1, \ldots, X_n) \leq y) \\ &= P(X_1 \leq y, X_2 \leq y, \ldots, X_n \leq y) \\ &= P(X_1 \leq y)^n \\ &= \left(\frac{y}{\theta}\right)^n I(0 \leq y \leq \theta) + I(y > \theta) \end{aligned}$$

Then the p.d.f is

$$g_n(y) = G_n'(y) = \frac{n}{\theta^n} y^{n-1} I(0 \leq y \leq \theta)$$

(ii) *Proof.* From Theorem 5.0.1, it suffices to show $E(Y_n) \to \theta$, and $\text{Var}(Y_n) \to 0$ then $Y_n \xrightarrow{P} \theta$.

$$E(Y_n) = \int_0^\theta \frac{n}{\theta^n} y^n dy = \frac{\theta^{n+1}}{\theta^n} \frac{n}{n+1} = \frac{n\theta}{n+1}$$

So $\lim_{n\to\infty} E(Y_n) = \theta$.

$$E(Y_n^2) = \int_0^\infty \frac{n}{\theta^n} y^{n+1} dy = \frac{n\theta^2}{n+2}$$

Now, the variance is

$$\text{Var}(Y_n) = E(Y_n^2) - E(Y_n)^2 = \frac{n\theta^2}{n+2} - \frac{n^2\theta^2}{(n+1)^2} \to 0$$

Therefore we have $E(Y_n) \to \theta$, and $\text{Var}(Y_n) \to 0$, so $Y_n \xrightarrow{P} \theta$. □

(iii) Let G_n be the c.d.f for U_n,

$$\begin{aligned}
G_n(u) = P(n(\theta - Y_n) \leq u) &= P\left(\theta - Y - n \leq \frac{u}{n}\right) \\
&= P\left(Y_n \geq \theta - \frac{u}{n}\right) \\
&= 1 - P\left(Y_n \leq \theta - \frac{u}{n}\right) \\
&= 1 - \left(\frac{\theta - u/n}{\theta}\right)^n \\
&= 1 - \left(1 - \frac{u}{n\theta}\right)^n
\end{aligned}$$

Now we can take the limit, we use logarithms to solve

$$\begin{aligned}
\lim_{n\to\infty} \ln\left(1 - \frac{u}{n\theta}\right)^n &= \lim_{n\to\infty} n \ln\left(1 - \frac{u}{n\theta}\right) \\
&= \lim_{n\to 0} \frac{\ln\left(1 - \frac{un}{\theta}\right)}{n} && \text{(Change } n \text{ to } 1/n\text{)} \\
&= \frac{0}{0} && \text{(Apply L'hoptial's Rule)} \\
&= \lim_{n\to 0} \frac{-u/\theta}{1 - un/\theta} \\
&= -\frac{u}{\theta}
\end{aligned}$$

Therefore,

$$G(u) = 1 - \lim_{n\to\infty}\left(1 - \frac{u}{n\theta}\right)^n = 1 - e^{-u/\theta}$$

Then the p.d.f is

$$g(u) = G'(u) = \frac{1}{\theta} e^{-u/\theta} I(0 < u < \infty) \sim \exp(\theta)$$

So we have shown $n(\theta - Y_n)$ converges to an exponential distribution with mean θ.

Example. Let $X_1, \ldots, X_n \stackrel{\text{iid}}{\sim} F$ for some unknown continuous c.d.f F with p.d.f $F' = f$. Let

$$Y_n = \max(X_1, \ldots, X_n)$$

Find the limiting distribution for $n(1 - F(Y_n))$.

Solution. Let $Y_1 < Y_2 < \cdots < Y_n$ be the order statistics for $X_1, \ldots, X_n$. Since F is a c.d.f, it is increasing so

$$F(Y_1) < F(Y_2) < \cdots < F(Y_n)$$

We proved previously that $F(x) \sim \text{Unif}(0,1)$. Then, we can treat $F(Y_1), \ldots, F(Y_n)$ as order statistics from n observations $U_1, \ldots, U_n \stackrel{\text{iid}}{\sim} \text{Unif}(0,1)$. Let $Z_1 = F(Y_1), \ldots Z_n = F(Y_n)$ be the order statistics for these uniform observations, then

$$g_{Z_N}(z) = n g(z) G(z)^{n-1} = n z^{n-1}$$

where $G(z)$ is the c.d.f for uniform distribution. Now we can use the previous example where we showed

$$n(\theta - Y_n) \xrightarrow{P} \exp\left(\frac{1}{\theta}\right)$$

Here in this example we have $\theta = 1$ and Y_n is $Z_n = F(Y_n)$, therefore

$$n(1 - F(Y_n)) \xrightarrow{P} \exp(1)$$

Therefore the limiting distribution for $n(1 - F(Y)n)$ is exponential with mean 1.

$$f(x) = e^{-x} I(0 < x < \infty)$$

Example. Let $X_1, \ldots, X_n \stackrel{\text{iid}}{\sim} f(x) = e^{-(x-\theta)} I(x > \theta)$. Find the assymptotic distrbiution for

$$U_n = n(\min(X_1, \ldots, X_n) - \theta)$$

Solution. First we find the c.d.f

$$
\begin{aligned}
G_n(u) = P(U_n \le u) &= P(n(\min(X_1,\dots,X_n) - \theta) \le u) \\
&= P\left(\min(X_1,\dots,X_n) - \theta \le \frac{u}{n}\right) \\
&= P\left(\min(X_1,\dots,X_n) \le \frac{u}{n} + \theta\right) \\
&= 1 - P\left(\min(X_1,\dots,X_n) \ge \frac{u}{n} + \theta\right) \\
&= 1 - P\left(X_1 > \frac{u}{n} + \theta\right)^n \\
&= 1 - \left(\int_{u/n+\theta}^{\infty} e^{-(x-\theta)} dx\right)^n \\
&= 1 - \left(e^{-u/n}\right)^n = 1 - e^{-u}
\end{aligned}
$$

Notice that $g(u) = e^{-u} I(0 < u < \infty)$ is free of n so we do not need to take the limit. Thus we get

$$
n(\min(X_1,\dots,X_n) - \theta) \xrightarrow{d} \exp(1)
$$

Example. Let $X_1,\dots,X_n \overset{\text{iid}}{\sim} \text{Bernoulli}(p)$. Prove

$$
\hat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} p
$$

Solution. We can calculate the expected value and variance

$$
\begin{aligned}
E(X_i) &= 1 \cdot P(X_i = 1) + 0 \cdot P(X_i = 0) = p \\
\text{Var}(X_i) &= 1^2 \cdot P(X_i = 1) + 0^2 P(X_i = 0) - p^2 = p(1-p)
\end{aligned}
$$

Then,

$$
\begin{aligned}
E(\bar{X}) &= E\left(\frac{X_1 + \cdots + X_n}{n}\right) = \frac{np}{n} = p \\
\text{Var}(\bar{X}) &= \frac{\sigma^2}{n} = \frac{p(1-p)}{n}
\end{aligned}
$$

Then, using Theorem 5.0.1, $E(\bar{X}) = p \to p$, and $\text{Var}(\bar{X}) = p(1-p)/n \to 0$, so $\bar{X} \xrightarrow{P} p$.

Example. Define the c.d.f for empirical process

$$
F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I(X_i \le x)
$$

Prove that empirical process is consistent estimator of F, in otherwords show $F_n(x) \xrightarrow{P} F(x)$.

Solution. Again, we useTheorem 5.0.1,

$$
E(F_n(x)) = \frac{1}{n}\sum_{i=1}^{n} E(I(X_i \le x)) = \frac{nP(X_i \le x)}{n} = F(x)
$$

So $E(F_n(x)) = F(x) \to F(x)$.

$$
\text{Var}(F_n(x)) = \frac{1}{n^2}\sum_{i=1}^{n} \text{Var}(I(X_i \le n))
$$

First we find the variance of X_i,

$$
\begin{aligned}
\text{Var}(I(X_i \le x)) &= E(I^2(X_i \le x)) - E(I(X_i \le x))^2 \\
&= E(I(X_i \le x)) - E(I(X_i \le x))^2 \\
&= F(x) - (F(x))^2 = F(x)(1 - F(x)) = \sigma^2
\end{aligned}
$$

Thus

$$
\text{Var}(F_n(x)) = n\frac{F(x)(1-F(x))}{n^2} = \frac{F(x)(1-F(x))}{n} \to 0
$$

Therefore we have $E(F_n(x)) \to F(x)$, and $\text{Var}(F_n(x)) \to 0$, therefore by Theorem 5.0.1, $F_n(x) \xrightarrow{P} F(x)$.

Theorem 5.0.4. *Let X_n be a sequence of random variables. If $X_n \xrightarrow{a.s} X$, then $X_n \xrightarrow{P} X$.*

Proof. Suppose $X_n \xrightarrow{a.s} X$, we know from the definition of a.s convergence that

$$P(|X_n - X| > \epsilon \ i.o) = 0$$

We can write this as

$$X_n \xrightarrow{a.s} X \iff P\left(\bigcup_{m=n}^{\infty} |X_m - X| > \epsilon\right) \to 0$$

$$P\left(\bigcup_{m=n}^{\infty} |X_m - X| > \epsilon\right) \geq P(|X_m - X| > \epsilon) \geq 0$$

If we take limits on both sides, we have

$$0 \leq P(|X_m - X| > \epsilon) \leq 0 \implies P(|X_m - X| > \epsilon) \to 0$$

Therefore, $X_n \xrightarrow{a.s} X \implies X_n \xrightarrow{P} X$. □

Definition 5.0.2. *We say that X_n converges completely to X, which is denoted by $X_n \xrightarrow{c} X$ if*

$$\sum_{n=1}^{\infty} P(|X_n - X| > \epsilon) < \infty$$

In otherwords, the series converges.

Theorem 5.0.5. *Let X_n be a sequence of random variables. If $X_n \xrightarrow{c} X$, then $X_n \xrightarrow{a.s} X$.*

Proof. We need to show that

$$P\left(\bigcup_{m=n}^{\infty} |X_m - X| > \epsilon\right) \to 0$$

Recall Borel's inequality where $P\left(\bigcup A_i\right) \leq \sum P(A_i)$. Then,

$$P\left(\bigcup_{m=n}^{\infty} |X_m - X| > \epsilon\right) \leq \sum_{m=n}^{\infty} P\left(|X_m - X| > \epsilon\right)$$

Then,

$$\lim_{n\to\infty} \sum_{m=n}^{\infty} P\left(|X_m - X| > \epsilon\right) = 0$$

Thus as n approaches infinity,

$$0 \leq P\left(\bigcup_{m=n}^{\infty} |X_m - X| > \epsilon\right) \leq 0 \implies P\left(\bigcup_{m=n}^{\infty} |X_m - X| > \epsilon\right) \to 0$$

□

Remark. If $X_n \xrightarrow{a.s} X$ and g is a continuous function, then $g(X_n) \xrightarrow{a.s} g(x)$.

Theorem 5.0.6. *Let X_n be a sequence of random variables. If $X_n \xrightarrow{P} X$, then $X_n \xrightarrow{d} X$.*

Proof. Let F be the c.d.f for X. We are only interested in values for $x \in C_F$. Take $x' \in \mathbb{R}$. Then using the fact that $P(A) = P(A \cap B) + P(A \cap B^c)$,

$$P(X \leq x') = P(X \leq x', X_n \leq x) + P(X \leq x', X_n > x)$$

Then since $P(A \cap B) \leq P(B)$,

$$P(X \leq x') \leq P(X_n \leq x) + P(X \leq x', X_n > x)$$

If $x' < x$, then

$$\begin{aligned} P(X \leq x') &\leq P(X_n \leq x) + P(X_n - X \geq x - x') \\ &\leq P(X_n \leq x) + P(|X_n - X| \geq x - x') \end{aligned}$$

Then taking the limit on both sides,

$$F(x') \leq \liminf_{n\to\infty} P(X_n \leq x)$$

Now if we have $x'' > x$, repeating what we had before

$$F_n(x) = P(X_n \leq x) \leq F(x'') + P(|X_n - X| \geq x'' - x)$$

Then taking the limit again

$$\limsup_{n\to\infty} F_n(x) \leq F(x'')$$

Then this tells us

$$F(x') \leq \liminf_{n\to\infty} F_n(x) \leq \limsup_{n\to\infty} F_n(x) \leq F(x'')$$

We have $x' < x \leq x''$ if we take the limit as $x'' \to x'$, then since $F(x)$ is a c.d.f and is right continuous,

$$F(x) \leq \liminf_{n\to\infty} F_n(x) \leq \limsup_{n\to\infty} F_n(x) \leq F(x)$$

□

This theorem gives us a sort of ordering for the strength of convergence,

$$X_n \xrightarrow{c} X \implies X_n \xrightarrow{a.s} X \implies X_n \xrightarrow{P} X \implies X_n \xrightarrow{d} X$$

Theorem 5.0.7. *Let X_n and Y_n be sequences of random variables with $X_n \xrightarrow{P} X$, and $Y_n \xrightarrow{P} Y$, then*

(i) $X_n + Y_n \xrightarrow{P} X + Y$

(ii) $X_n Y \xrightarrow{P} XY$

(iii) $X_n Y_n \xrightarrow{P} XY$

Note. We *cannot* conclude from this that $X_n Y_n \xrightarrow{d} XY$.

Proof. (i) We want to show that

$$P(|X_n + Y_n - X - Y| \geq \epsilon) \to 0$$

For any $\epsilon > 0$, we have

$$P(|X_n - X| > \epsilon/2) + P(|Y_n - Y| > \epsilon/2) \geq P(|X_n + Y_n - X - Y| > \epsilon)$$

Since if $|X_n + Y_n - X - Y| > \epsilon$, we must have that $|X_n - X| > \epsilon/2$, or $|Y_n - Y| > \epsilon/2$. As n goes to infinity, $P(|X_n - X| > \epsilon/2) \to 0$ since $X_n \xrightarrow{P} X$, simiarly for Y_n, therefore

$$\lim_{n\to\infty} P(|X_n + Y_n - X - Y| > \epsilon) \to 0$$

(ii) Let k be a constant such that $P(|Y| > k) < \delta$ for some arbitrary $\delta > 0$, then we have

$$P(|X_n Y - XY| > \epsilon) = P(|Y||X_n - X| > \epsilon)$$

Now we have 2 cases where $|Y| > k$, or $|Y| \leq k$, so

$$\begin{aligned}
P(|Y||X_n - X| > \epsilon) &= P(|Y||X_n - X| > \epsilon, |Y| > k) \\
&\quad + P(|Y||X_n - X| > \epsilon, |Y| \leq k) \\
&\leq P(|Y| > k) + P(|Y||X_n - X| > \epsilon, 1/|Y| > 1/k) \\
&\leq \delta + P\left(|Y||X_n - X| \cdot \frac{1}{|Y|} \geq \frac{\epsilon}{k}\right) \\
&= \delta + P(|X_n - X| \geq \epsilon/k)
\end{aligned}$$

Remember we can find a k for any delta to make this statement hold, if we take the limit of this right side $P(|X_n - X| \geq \epsilon/k) \to 0$ as $n \to \infty$ since $X_n \xrightarrow{P} X$, therefore

$$P(|X_n Y - XY| > \epsilon) < \delta \forall \delta > 0$$

As $\delta \downarrow 0^+$, we have

$$P(|X_n Y - XY| > \epsilon) \to 0$$

Note that we could invert $|Y|$ since if $Y = 0$, this case is trivial because $X_n Y = 0$ and $XY = 0$ so of course $X_n Y \to XY$.

(iii) It suffices to show that if $X_n \xrightarrow{P} 0$, and $Y_n \xrightarrow{P} 0$, then $X_nY_n \xrightarrow{P} 0$, because then we have from the previous proofs

$$(X_n - X)(Y_n - Y) = X_nY_n - XY_n - YX_n + XY$$

Notice that $XY_n \xrightarrow{P} XY$, and $YX_n \xrightarrow{P} XY$ from the previous results, so we have

$$X_nY_n - XY_n - YX_n + XY \xrightarrow{P} 0 \implies X_nY_n - XY \xrightarrow{P} 0$$

So we must show that $X_nY_N \xrightarrow{P} 0$. Using a similar strategy as the previous proof, we have that $\forall \delta > 0$, $\exists k$ such that $P(|X_n| \geq k) \leq \delta$, then

$$\begin{aligned} P(|X_nY_n > \epsilon) &= P(|X_nY_n| > \epsilon, |X_n| \geq k) + P(|X_nY_n| > \epsilon, |X_n| < k) \\ &\leq P(|X_n| \geq k) + P(|X_nY_n| > \epsilon, 1/|X_n| > 1/k) \\ &\leq \delta + P\left(|Y_n| > \frac{\epsilon}{k}\right) \end{aligned}$$

$P(|Y_n| > \epsilon/k) \to 0$, so

$$0 \leq \lim_{n\to\infty} P(|X_nY_n| > \epsilon) \leq \delta$$

Then as $\delta \downarrow 0^+$, we have $P(|X_nY_n| > \epsilon) \to 0$ as required.

□

Example. Find what

$$\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^n$$

converges to in probability.

Solution. We can rewrite this as

$$\frac{1}{n}\sum_{i=1}^{n}(X_i^3 - 3Xi^2\bar{X} + 3\bar{X}^2X_i - \bar{X}^3)$$

Then, using weak law of large numbers, we have that these terms converge to

$$\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^n \xrightarrow{P} E[(X-\mu)^3]$$

This is known as skewness in probability.

Theorem 5.0.8. *Let X_n and Y_n be sequences of random variables, then if $X_n \xrightarrow{P} X$, and $X_n \xrightarrow{P} Y$ for random variables X and Y. Then $P(X = Y) = 1$.*

Proof. We want to show that for any $\epsilon > 0$, $P(|X - Y| > \epsilon) = 0$.

$$\begin{aligned} P(|X - Y| > \epsilon) &= P(|X_n - X - X_n + Y| > \epsilon) \\ &= P(|X_n - X| > \epsilon/2) + P(|X_n - Y| > \epsilon/2) \end{aligned}$$

Now we know $X_n \xrightarrow{P} X$ and $X_n \xrightarrow{P} Y$, so the right side converges to 0, therefore

$$P(|X - Y| > \epsilon) \to 0$$

□

Theorem 5.0.9 (Continuity Theorem)**.** *Let X_n be a sequence of random variables with c.d.f F_n and characteristic function $\phi_n(\theta)$. Then*

$$X_n \xrightarrow{d} X \iff \phi_n(\theta) \to \phi(\theta)$$

Where $\phi(\theta) = E(e^{i\theta X})$.

Theorem 5.0.10 (Centeral Limit Theorem). *Let* $X_1, \ldots, X_n \overset{\text{iid}}{\sim} F$*, with*

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Let $E(X_i) = 0$*,* $\text{Var}(X_i) = 1$*, then*

$$\sqrt{n}\frac{\bar{X}-\mu}{\sigma} = \sqrt{n}\bar{X} \xrightarrow{d} N(0,1)$$

Proof. We want to prove $\sqrt{(n)}\bar{X} \xrightarrow{d} N(0,1)$, we'll do this using moment generating functions.

$$\begin{aligned}
M_n(t) &= E(\exp(t\sqrt{n}\bar{X})) \\
&= E\left(\exp\left(t\frac{X_1+\cdots+X_n}{\sqrt{n}}\right)\right) \\
&= E(\exp(tX_1/\sqrt{n}))E(\exp(tX_2/\sqrt{n}))\cdots E(\exp(tX_n/\sqrt{n})) \\
&= M\left(\frac{t}{\sqrt{n}}\right)^n
\end{aligned}$$

where $M(t)$ is the m.g.f of X_i. Now we want to evaluate the limit of this using logarithms again

$$\begin{aligned}
\lim_{n\to\infty} \ln(M(t/\sqrt{n}))^n &= \lim_{n\to\infty} n \ln M(t/\sqrt{n}) \\
&= \lim_{n\to 0} \frac{\ln M(t\sqrt{n})}{n} && \text{(Replace } n \text{ with } 1/n\text{)} \\
&= \frac{\ln M(0)}{0} = \frac{0}{0} && \text{(Use L'Hoptial's Rule)} \\
&= \lim_{n\to 0} \frac{tM'(t\sqrt{n})}{2\sqrt{n}} \\
&= \frac{0}{0} && (M'(0) = \mu = 0) \\
&= \lim_{n\to 0} \frac{t^2}{2} M''(t\sqrt{n}) \\
&= \frac{t^2}{2}E(X^2) = \frac{t^2}{2}
\end{aligned}$$

Therefore,

$$\ln M_n(t) \to \frac{t^2}{2} \implies M_n(t) \to e^{t^2/2}$$

This is the m.g.f for standard normal distribution, therefore

$$\sqrt{n}\bar{X} \xrightarrow{d} N(0,1)$$

□

Note that if $\mu \neq 0$ and $\sigma^2 \neq 0$, then we can standardize the samples with

$$Y_i = \frac{X_i - \mu}{\sigma}$$

Then $E(Y_i) = 0$, $\text{Var}(Y_i) = 1$, then

$$\begin{aligned}
\sqrt{n}\bar{Y} \xrightarrow{d} N(0,1) &\implies \sqrt{n}\frac{1}{n\sigma}\sum_{i=1}^{n}(X_i - \mu) \xrightarrow{d} N(0,1) \\
&\implies \frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma/\sqrt{n}} \xrightarrow{d} N(0,1) \\
&\implies \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \xrightarrow{d} N(0,1)
\end{aligned}$$

Example. Let $X_i \sim \chi^2(1)$, and if $X_n \sim \chi^2(n)$, then

$$\frac{X_n - n}{\sqrt{2n}} \xrightarrow{d} N(0,1)$$

From CLT,

$$\chi^2(n) = \chi^2(1) + \cdots + \chi^2(1) \implies \frac{\sum X_i - n}{\sqrt{2n}} \xrightarrow{d} N(0,1)$$

Example. Let $X_1, \ldots, X_n \sim \text{Bernoulli}(p)$. Recall that

$$\sum_{i=1}^{n} X_i \sim \text{Bin}(n,p)$$

Then

$$\frac{\sum X_i - np}{\sqrt{np(1-p)}} \xrightarrow{d} N(0,1)$$

If $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{Poisson}(\mu)$, then

$$\frac{\sum X_i - n\mu}{\sqrt{n\mu}} \xrightarrow{d} N(0,1)$$

Theorem 5.0.11. *If $E(|X|^m) < \infty$ for a given $m \in \mathbb{N}$, then the characteristic function can be written as*

$$\phi(\theta) = \sum_{j=0}^{m} \frac{(i\theta)^j}{j!} E(X^j) + o(\theta^m)$$

We say a function $g(h)$ is small $o(h)$ if $g(h)/h \to 0$ as $h \to 0$. We say $g(h)$ is $O(h)$ if $|g(h)| \leq M$ for some constant M.

Theorem 5.0.12 (Weak Law of Large Numbers)**.** *Let $\{X_i\}$ be a sequence of i.i.d random variable such that $E(X_i) = \mu$, then*

$$\bar{X} \xrightarrow{P} \mu$$

Theorem 5.0.13. *Let $\{X_n, Y_n\}$ be two sequences of random variables such that $|X_n - Y_n| \xrightarrow{P} 0$, and $Y_n \xrightarrow{d} Y$. Then $X_n \xrightarrow{d} Y$.*

Proof. Let $x \in C_F$ and $\epsilon > 0$ be given. Then

$$\begin{aligned}
P(X_n \leq x) &= P(Y_n \leq x + Y_n - X_n) \\
&= P(Y_n \leq x + Y_n - X_n, Y_n - X_n \leq \epsilon) \\
&+ P(Y_n \leq x + Y_n - X_n, Y_n - X_n > \epsilon) \\
&\leq P(Y_n \leq x + \epsilon) + P(|Y_n - X_n| \geq \epsilon)
\end{aligned}$$

□

Then taking limits we get

$$\limsup_{n\to\infty} P(X_n \leq x) \leq \liminf_{n\to\infty} P(Y_n \leq x + \epsilon)$$

Similarly

$$\begin{aligned}
P(Y_n \leq x - \epsilon) &= P(X_n \leq x + X_n - Y_n - \epsilon) \\
&= P(X_n \leq x + X_n - Y_n - \epsilon, X_n - Y_n \leq \epsilon) \\
&+ P(X_n \leq x + X_n - Y_n - \epsilon, X_n - Y_n > \epsilon) \\
&\leq P(X_n \leq x + X_n - Y_n - \epsilon, X_n - Y_n \leq \epsilon) + P(|X_n - Y_n > \epsilon) \\
&\leq P(X_n \leq x) + P(X_n \leq x) + P(|X_n - Y_n| > \epsilon)
\end{aligned}$$

Then taking limits we get

$$\limsup_{n\to\infty} P(Y_n \leq x - \epsilon) \leq \liminf_{n\to\infty} P(X_n \leq x)$$

Since $\epsilon > 0$ is arbitrary and $x \in C_F$, the result follows when $n \to \infty$.

Remark. If $X_n \xrightarrow{d} X$, $Y_n \xrightarrow{P} c$ for a constant c, then $X_n + Y_n \xrightarrow{d} X + c$.

Proof.

$$(X_n + Y_n) - (X_n + C) = Y_n - c \xrightarrow{d} 0$$

Therefore the limiting distribution for $X_n + Y_n$ and $X_n + c$ is the same since $X_n \xrightarrow{d} X$. Then $X_n + c \xrightarrow{d} X + c$, therefore

$$X_n + Y_n \xrightarrow{d} X + c$$

□

Remark. If $X_n \xrightarrow{d} X$, and $Y_n \xrightarrow{P} c$ for a constant c, then $X_n Y_n \xrightarrow{d} cX$.

Proof. Let's consider the case where $c = 0$, we want to show $X_n Y_n \xrightarrow{P} 0$.

$$\begin{aligned} P(|X_n Y_n| > \epsilon) &= P(|X_n Y_n| > \epsilon, |Y_n| < \epsilon/k) + P(|X_n Y_n| > \epsilon, |Y_n| \geq \epsilon/k) \\ &\leq P(|Y_n > \epsilon/k) + P(|X_n Y_n| > \epsilon, |Y_n| < \epsilon\epsilon/k) \\ &\leq P(|Y_n| > \epsilon/k) + P(|X_n| > k) \end{aligned}$$

Notice that if $|X_n Y_n| > \epsilon$, and $|Y_n| < \epsilon/k$, then $|X_n| > k$. To see this, consider if $|X_n| \leq k$, and $|Y_n| < \epsilon/k$, then $|X_n Y_n| < \epsilon$ if we multiply the inequalites. So it must be that $|X_n| \leq k$. Then as $n \to \infty$, $P(|Y_n| > \epsilon/k) \to \infty$, and

$$0 \leq \limsup_{n\to\infty} P(|X_n Y_n| > \epsilon) \leq \limsup_{n\to\infty} P(|X_n| \geq k)$$

As we mentioned earlier, k can be any value, so if we take $k \to \infty$, then

$$0 \leq \lim_{k\to\infty} P(|X_n Y_n| > \epsilon) \leq 0$$

If $c \neq 0$, then

$$X_n Y_n - cX_n = X_n(Y_n - c)$$

Then $Y_n - c \xrightarrow{P} 0$, so $X_n(Y_n - c) \to 0$, and thus the difference between $X_n Y_n$ and cX_n goes to 0, therefore

$$X_n Y_n - cX_n \xrightarrow{d} 0 \implies X_n Y_n \xrightarrow{d} cX$$

□

To summarize, we proved the two results

$$X_n \xrightarrow{d} X, Y_n \xrightarrow{P} c \implies \begin{cases} X_n + Y_n \xrightarrow{d} X + c \\ X_n Y_n \xrightarrow{d} cX \end{cases}$$

Applications

Example. Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} f(x)$ for some unknown distribution with $E(X_i^2) < \infty$. From the central limit theorem,

$$\frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \xrightarrow{d} N(0, 1)$$

Then

$$\frac{1}{n}\sum_{i}^{n}(X_i - \bar{X})^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2$$

From weak law of large numbers

$$\frac{1}{n}\sum_{i=1}^{n} X_i^2 \xrightarrow{P} E(X_1^2)$$

$$\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \xrightarrow{P} \mu$$

So

$$\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2 \xrightarrow{P} E(X_1^2) - \mu^2 = \sigma^2$$

Then the sample variance

$$\frac{n}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}) \xrightarrow{P} \sigma^2$$

since $\frac{n}{n-1} \to 1$. Using the previous results, we've shown, we also have

$$S \xrightarrow{P} \sigma \implies \frac{S_n}{\sigma} \to 1$$

This gives us the result that

$$\frac{\sqrt{n}(\bar{X} - \mu)}{S} \xrightarrow{d} N(0,1)$$

Example. Let $X_1, \dots X_n \overset{\text{iid}}{\sim} \text{Bernoulli}(p)$. Then from c.l.t

$$\frac{\sum_{i=1}^{n} X_i - n =}{\sqrt{np(1-p)}} \xrightarrow{d} Z \sim N(0,1)$$

On the otherhand,

$$\hat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i = E(X_1) = p \implies \hat{p} \xrightarrow{P} p$$

Similarly from the previous results,

$$\hat{p}(1-\hat{p}) \xrightarrow{P} p(1-p)$$

Therefore,

$$\frac{\hat{p} - p}{\sqrt{\hat{p}(1-\hat{p})/n}} \xrightarrow{d} Z$$

Theorem 5.0.14 (Skorohod Representation)**.** *Let X_n be a sequence of random variables with probability space $(\Omega, \mathcal{F}, P)$ and $X_n \xrightarrow{d} X$. Then there exists a probability space $([0,1], [\prime, \infty], P^*)$ and a sequence of random variables in this space X_n^* with a random variable X^* such that*

$$X_n^* \overset{\text{d}}{=} X_n, \; X^* \overset{\text{d}}{=} X$$

Then

$$X_n^* \xrightarrow{a.s} X^*$$

The proof for this theorem is omitted but we will cover applications of it.

Theorem 5.0.15 (Continuous Mapping Theorem)**.** *Let X_n be a sequence of random variables with $X_n \xrightarrow{d} X$ for a random variable X. Let $g : \mathbb{R} \mapsto \mathbb{R}$ be a continuous function, then*

$$g(X_n) \xrightarrow{d} g(X)$$

Proof. Using Skorohod Representation, $X_d \xrightarrow{d} X$ so there exists a probability space $([0,1], \mathcal{B}([0,1]), P^*)$ and random variables $X_n^* \xrightarrow{d} X_n$, $X^* \xrightarrow{d} X$ with $X_n^* \xrightarrow{a.s} X^*$. Since $X_n^* \xrightarrow{a.s} X^*$ and g is continuous, so

$$g(X_n^*) \xrightarrow{a.s} g(X^*)$$

Then

$$g(X_n^*) \overset{\text{d}}{=} g(X_n), g(X^*) \overset{\text{d}}{=} g(X) \implies g(X_n) \xrightarrow{d} g(x)$$

□

5.1 Delta Method

Theorem 5.1.1. *Let X_n be a sequence of random variables such that*

$$a_n(X_n - \theta) \xrightarrow{d} X$$

Let g be a differential function, then

$$a_n(g(X_n) - g(\theta)) \xrightarrow{d} g'(\theta)X$$

Proof. Using Skorohod Representation, we have a probability space with X_n^*, X^* such that

$$a_n(X_n^* - \theta) \xrightarrow{a.s} X^*$$

Then

$$a_n(g(X_n^*) - g(\theta)) = \frac{g(X_n^*) - g(\theta)}{X_n^* - \theta} a_n(X_n^* - \theta)$$

Then as $n \to \infty$,

$$\frac{g(X_n^*) - g(\theta)}{X_n^* - \theta} \to g'(\theta)$$

Therefore

$$a_n(g(X_n^*) - g(\theta)) \to g'(\theta)X$$

□

What happens if $g'(0) = 0$? Look at the taylor expansion

$$g(X_n) = g(\theta) + g'(\theta)(X_n - \theta) + \frac{g''(\theta)}{2}(X - n - \theta)^2 \cdots$$

Then

$$a_n(g(X_n) - g(\theta)) \xrightarrow{d} \frac{g''(\theta)}{2} X$$

Example. Let $X_1, \ldots, X_n$ be i.i.d random variables. Find the limiting distribution a_n,b_n and σ^2 such that

$$a_n(\bar{X}^2 - b_n) \xrightarrow{d} X \sim N(0, \sigma^2)$$

Solution. Using delta method, define

$$g(x) = x^2$$

Then we have

$$\sqrt{n}(\bar{X} - \mu) \xrightarrow{d} N(0, \sigma_0^2)$$

and

$$\sqrt{n}(\bar{X}^2 - \mu^2) \xrightarrow{d} 2\mu X \sim N(0, 4\mu^2\sigma_0^2)$$

Thus we have $a_n = \sqrt{n}$, $b_n = \mu^2$, and $\sigma^2 = 4\mu^2\sigma_0^2$

Example. Let $X_1, \ldots, X_n$ be i.i.d random variables. With $e(X_1) = 0$, $\mathrm{Var}(X_1) = \sigma_0^2$. Find a_n, b_n, and σ^2 such that

$$a_n(\cos \bar{X} - 1) \xrightarrow{d} N(0, \sigma^2)$$

Solution. Define $g(x) = \cos x$, $g(0) = 1$, $g'(x) = -\sin x$, and $g'(0) = 0$, so we must use $g''(0) = -1$. Then

$$\sqrt{n}(g(\bar{X}) - g(0)) =\xrightarrow{d} \frac{g''(0)}{2} X = -\frac{X}{2} \sim N\left(0, \frac{1}{4}\sigma_0^2\right)$$

Therefore

$$\sqrt{n}(\cos \bar{X} - 1) \xrightarrow{d} N\left(0, \frac{1}{4}\sigma_0^2\right)$$

Example. Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{Bernoulli}(p)$. We know that

$$\sqrt{n}(\hat{p} - p) \xrightarrow{d} N(0, p(1-p))$$

Then we can also find

$$\sqrt{n(\hat{p}^2 - p^2)} \xrightarrow{d} 2pN(0, p(1-p)) = N(0, 4p^3(1-p))$$

Solution. Let $X_1, \ldots, X_n$ be i.i.d random variables with $E(X_i) = 0$, $\mathrm{Var}(X_i) = \sigma_0^2$, find the limiting distribution for

$$\frac{\bar{X}}{1 + \bar{X}}$$

Solution. Define

$$g(u) = \frac{u}{1+u}$$

we know that from the central limit theorem

$$\sqrt{n}(\bar{X} - 0) \xrightarrow{d} N(0, \sigma_0^2)$$

Then

$$\sqrt{n}(g(\bar{X}) - g(0)) \xrightarrow{d} g'(0)N(0, \sigma_0^2) = N(0, \sigma_0^2)$$

Therefore

$$\frac{\bar{X}}{1 + \bar{X}} \xrightarrow{d} N(0, \sigma_0^2)$$