

# MAT 3172 Lecture Notes

Last Updated:

June 2, 2023

# Contents

<b>1</b>	<b>Probability Measures</b>	<b>2</b>
1.1	Review of Set Theory . . . . .	2
1.1.1	Properties of Sets . . . . .	4
1.2	Indicator Function . . . . .	5
1.2.1	Properties of Indicator Functions: . . . . .	6
1.3	Set Theoretic Limits . . . . .	6
1.4	Fields and Algebras . . . . .	11
1.4.1	Properties of Probability Measures . . . . .	12
1.4.2	Expectations . . . . .	13
1.5	Finding Probabilities Using Expectations . . . . .	14
1.5.1	Properties . . . . .	14
1.6	Independence . . . . .	19
1.6.1	Properties of Independence . . . . .	20
<b>2</b>	<b>Important Inequalities and Probability Theorems</b>	<b>24</b>
2.1	Important Inequalities . . . . .	26
2.2	Moment Generating Functions . . . . .	29
2.2.1	Properties of Moment Generating Functions . . . . .	29
2.2.2	Applications of Moment Generating Functions . . . . .	31
2.3	Holder, Lyapunov and Minkowski Inequalities . . . . .	34

# Chapter 1

## Probability Measures

### 1.1 Review of Set Theory

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

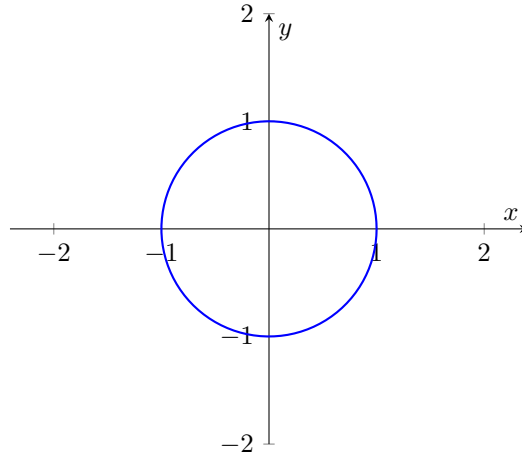
$$\Omega = \{1H, 2H, \dots, 6H, 1T, \dots 6T\}$$

If we flip a coin until we observe a head,

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

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:**

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

$$A = \{1, 2, 3\} \quad B = \{1, 2, 3\}$$

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

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

$$A^c = \{1, 3, 5\}$$

$$A \setminus B = \{x : x \in A, x \notin B\} = \{4, 6\}$$

$$A \Delta B = (A \cup B) \setminus (A \cap B)$$

**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 arbitrarily 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 arbitrarily 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 sets with  $\left(0, \frac{1}{n}\right)$  which become arbitrarily small, we have the empty set.

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

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

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

□

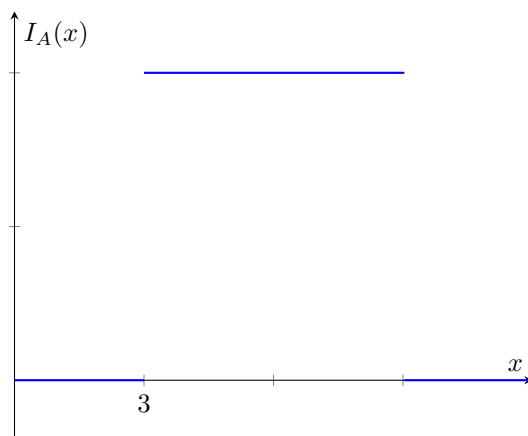
## 1.2 Indicator Function

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  $A_1, A_2, A_3, \dots$ , we can define

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

So its sequence members are

$$B_n = A_n \cap A_{n+1} \cap \dots$$

$$B_{n+1} = A_{n+1} \cap A_{n+2} \cap \dots$$

$$B_{n+2} = A_{n+2} \cap A_{n+3} \cap \dots$$

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

$$B_n = A_n \cup A_{n+1} \cup \dots$$

$$B_{n+1} = A_{n+1} \cup A_{n+2} \cup \dots$$

$$B_{n+2} = A_{n+2} \cup A_{n+3} \cup \dots$$

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

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

**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

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

and

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



*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

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

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  $\omega$  is in infinitely many of the  $A_i$ 's. For the limit inferior, it means that  $\omega$  is in all but finitely many of the  $A_i$ 's.

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

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

$\omega$  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

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

So,  $\omega$  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

$$A_1 \subset A_2 \subset A_3 \subset \dots$$

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

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

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

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

Furthermore,

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

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.**

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

To see this, we have that

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

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.**

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

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

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

Similarly for the limit inferior,

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

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  $\Omega$  (called events) that contain  $\Omega$  and are closed under finite union, finite intersection, and complementation. In other words, a family of subsets of  $\Omega$  (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** ( $\sigma$ -field). *A  $\sigma$ -field (or  $\sigma$ -algebra) is a field that is closed under countable union (which implies that it is closed under countable intersection).*

**Example.** Let  $\Omega$  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  $\sigma$ -fields.**

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

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

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

(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  $X : \Omega \rightarrow \mathbb{R}$ , an expectation  $E$  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  $X_n(\omega) \rightarrow X(\omega)$ , then

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

**Example.** Flip 2 coins,

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

Define  $X : \Omega \rightarrow \mathbb{R}$ , 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,

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

With

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

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

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

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.

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

3.

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

4.

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

5. The 5th axiom fails however. Take  $\Omega = \mathbb{R}$  and  $X_n(\omega) = I_{[-n, -n]}(\omega)$ , then

$$\lim_{D \rightarrow \infty} \frac{\int_{-D}^D X(\omega) d\omega}{2D} = 0$$

But,  $x_n(\omega) \rightarrow 1$ . 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 sometimes drop  $\omega$  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  $X \leq Y \leq Z$ , then  $E(X) \leq E(Y) \leq E(Z)$

3. If  $\{A_i\}$  is a sequence of events, then

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

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

*Proof.* Using the fact that

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

and the second property where if  $X \leq Y$ , then  $E(X) \leq E(Y)$ , then

$$E(I_{\bigcup A_i}) \leq \sum_{i=1}^{\infty} E(I_{A_i})$$

Then this is exactly

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

□

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 probabilities, so we have that

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

We know from indicator functions that

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

Therefore,

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

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

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

And this is again proved using the fact that

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

**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  $i$ th 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,

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

Then we can evaluate each probability,

$$P(A_1) = \frac{1}{100}, P(A_2) = \frac{1}{100}, \dots, P(A_{100}) = \frac{1}{100}$$

$$P(A_1 \cap A_2) = \frac{1}{100} \cdot \frac{1}{99} = \dots = P(A_{99} \cap A_{100})$$

$$P(A_1 \cap A_2 \cap A_3) = \frac{1}{100} \cdot \frac{1}{99} \cdot \frac{1}{98} = \dots$$

Then continuing this sequence we have

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

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

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

So,

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

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

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,

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

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

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

Therefore,

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

□

To summarize, when  $A_n$  is increasing, we have

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

when  $A_n$  is decreasing,

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

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

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

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

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

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

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

□

## 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,

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

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 independent 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  $\Omega$ .

*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

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

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

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

( $\impliedby$ ) 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  $\Omega$  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  $\sum_{i=1}^{\infty} P(E_i) < \infty$ , 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 whether the series  $\sum_{i=1}^{\infty} P(E_i)$  diverges or converges respectively.*

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

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

Since  $\sum_{i=1}^{\infty} P(E_i) < \infty$ , then  $\lim_{n \rightarrow \infty} P(F_n) = 0$ . Since  $F_n \downarrow E$ , from Fatou's lemma we can write

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

Thus (i) is proved. For (ii), suppose  $E_1, E_2, \dots$  are independent. From (ii), we know that  $P(\limsup E_n) = 0$  if the sum  $\sum_{n=1}^{\infty} P(E_n)$  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

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

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

As  $N \rightarrow \infty$ , we get  $\bigcap_{m=n}^N E_m^c \downarrow E^c$ . Thus,

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

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

(  $\Rightarrow$  ) 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. □

**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  $E = (0, \frac{1}{n})$  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  $n$ th 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  $f : \mathbb{R} \rightarrow \mathbb{R}$  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

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

**Example.** Let  $X$  be a 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  $\text{Var}(X)$ .
- (iii) Find  $P(0.5 < X < 0.75)$ .

**Solution.**

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

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

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

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

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

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

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.

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

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  $X = [X_1, \dots, X_p]$ , 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  $(X'c)' = Y$ , then we have

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

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  $X' = [X_1, \dots, X_p]$  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  $X \rightarrow (X - \mu_X)$  and  $Y \rightarrow (Y - \mu_Y)$ , then we get

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

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

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

Furthermore,

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

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

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

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

## 2.2 Moment Generating Functions

**Definition 2.2.1.** Let  $X$  be a random variable with 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)  $\frac{d}{dt} M(t) = \int_{-\infty}^{\infty} x e^{tX} dF(X) = E(X e^{tX}) \implies M'(0) = E(X)$

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

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

Then, we have

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

So,

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

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

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

We can also calculate the expected value and variance

$$\begin{aligned} M'(t) &= \frac{\alpha}{\beta} \left(1 - \frac{t}{\beta}\right)^{-\alpha-1} \implies M'(0) = \frac{\alpha}{\beta} = 1 \\ M''(t) &= \frac{\alpha}{\beta} \cdot -\frac{1}{\beta} (-\alpha-1) \left(1 - \frac{t}{\beta}\right)^{-\alpha-2} \implies M''(0) = \frac{\alpha(\alpha+1)}{\beta^2} \\ \text{Var}(X) &= E(X^2) - E(X)^2 = \frac{\alpha(\alpha+1)}{\beta^2} - \frac{\alpha^2}{\beta^2} = \frac{\alpha}{\beta^2} \end{aligned}$$

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

$$\begin{aligned} M(t) &= (1 - 2t)^{-\frac{r}{2}} \\ E(X \sim \chi^2(r)) &= \frac{\alpha}{\beta} = \frac{r/2}{1/2} = r \\ \text{Var}(X \sim \chi^2(r)) &= \frac{\alpha}{\beta^2} = \frac{r/2}{1/4} = 2r \end{aligned}$$

### 2.2.2 Applications of Moment Generating Functions

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



moment generating function

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

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

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

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

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

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

We can calculate the m.g.f

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

The integral is 1 since it is the integral of the p.d.f of the normal distribution 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  $E(X^k) = \text{Some multiple of } t^k \cdot k!$ . 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 expectation. 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,

$$\begin{aligned} Z \sim N(0, 1) &\implies M(t) = e^{t/2} \\ Y = Z^2 &\implies M_Y(t) = (1-2t)^{-1/2} \rightarrow \chi^2(1) \\ M_{\chi^2}(t) &= (1-2t)^{-r/2} \\ Z \sim N(0, 1) &\implies Z^2 \sim \chi^2(1) \end{aligned}$$

**Example.** Let  $X_1, \dots, X_n$  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

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

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  $\mu = a_1 \mu_1 + \dots + a_n \mu_n$ , and  $\sigma^2 = a_1^2 \sigma_1^2 + \dots + a_n^2 \sigma_n^2$ . 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  $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \text{Bernoulli}(p)$ , 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*

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

*then*

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

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

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

Differentiating 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  $\phi$  on  $(0, \infty)$ . Set  $t = \frac{\alpha^{1/q}}{\beta^{1/p}}$  to get

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

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

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

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*

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

*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

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

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

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

Thus,

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

□