

Gonit Sora Notes

Probability Theory

By

Manjil P. Saikia

Rupam Haloi

June 10, 2013

Department of Mathematical Sciences

Tezpur University

Napaam-784028, India

Preface

This is the first of the proposed sets of notes to be published in the website Gonit Sora (<http://gonitsora.com>). The notes will hopefully be able to help the students to learn their subject in an easy and comprehensible way. These notes are aimed at mimicking exactly what would be typically taught in a one-semester course at a college or university. The level of the notes would be roughly at the undergraduate level.

The present sets of notes are not yet complete and this is the second version that is being posted. These notes contain very few proofs and only state the important results in *Probability Theory*. These notes are based on the course taught at Tezpur University, Assam, India by *Dr. Santanu Dutta*. There may be some errors and typos in these notes which we hope the reader would bring to our notice.

Manjil P. Saikia

manjil@gonitsora.com

Rupam Haloi

Contents

1	Introduction	1
2	Preliminaries and Basics of Measure Theory	2
2.1	Preliminary Results	2
2.2	Length of a Borel set	3
2.3	Distribution Functions	6
3	Integration Theory	11
3.1	Riemann Integration	11
3.2	Riemann-Steiltjes Integration	13
3.3	Lebesgue Integration	16
4	Probability Measure and Random Variables	22
4.1	Probability Measure	22
4.2	Random Variable and Mathematical Expectation	23
4.3	Expectations of Some Distributions	24
5	Inequalities involving probabilities	26
	Bibliography	32

Unit 1

Introduction

The objective of this course is to give an introduction to probability theory via the theory of measure and integration. Some amount of inequalities related to probability are also discussed.

This report is divided into five units. The second unit deals with some preliminary concepts related to measure theory, the third unit deals with integration theory where focus is given on Riemann-Steiltjes integrals and Lebesgue integrals. The fourth unit deals with probability spaces and the concept of a random variable as an integral. Finally the fifth unit gives some basic inequalities related to probability.

Almost all the proofs are omitted in this report. Many proofs can be found in either [1] or [2].

Unit 2

Preliminaries and Basics of Measure Theory

In this Unit we shall look into some preliminary concepts of measure theory, with a few proofs and some examples.

2.1 Preliminary Results

We begin with the following important results which we shall be coming back from time to time.

Property 2.1.1 (Archimedean Property). *For any pair of positive real numbers a and b , there is a natural number n for which $na > b$.*

Theorem 2.1.2 (Heine-Borel). *Let F be an open covering of a closed and bounded set A in \mathbb{R}^n , then a finite subcollection of F also covers A .*

Now we have the following

Theorem 2.1.3. *An open set can be expressed as union of atmost countable disjoint intervals.*

Proof. We take $x \in O$, then there exists an $\epsilon > 0$ such that $(x - \epsilon, x + \epsilon) \subset O$.

Now we consider the sets $y_1 = \sup\{z > x : (x - \epsilon, z) \subset O\}$ and $y_2 = \inf\{z < x : (z, x + \epsilon) \subset O\}$. It is clear that $x \in (y_1, y_2) \subset O$.

For if $x \neq y$, let O_x and O_y be the largest open intervals containing x and y such that $O_x, O_y \subset O$, and either $O_x = O_y$ or $O_x \cap O_y = \phi$.

Now by Property 2.1.1, there exists an $r_1 \in O_x, r_2 \in O_y, r_1 \neq r_2, r_1, r_2 \in Q$, where $O_x \cap O_y = \phi$. Since Q is countable, so the no of intervals is countable.

□

From the above we immediately have the following

Corollary 2.1.4. (a) *Open sets are Borel sets.*

(b) *Closed sets are Borel sets.*

It is now clear that $(a, b]$ is a Borel set because

$$(a, b] = \bigcap_{n \in \mathbb{N}} (a, b + \frac{1}{n}).$$

Again the set $[a, b)$ is also a Borel set because

$$[a, b) = \bigcap_{n \in \mathbb{N}} (a - \frac{1}{n}, b).$$

2.2 Length of a Borel set

It is actually possible to define a notion of length for every Borel set.

Definition 2.2.1. Let \mathcal{F} be the class of sets which can be expressed as union of finite number of disjoint semi-open intervals.

As a convention the sets $(a, \infty), (-\infty, b]$ are considered as semi-open intervals.

Clearly the set \mathcal{F} so defined above is a field. We have a natural definition of length of sets in \mathcal{F} . If $B \in \mathcal{F}$, then

$$\begin{aligned} l(B) &= \text{length of } B \\ &= \sum_{i=1}^n l(a_i, b_i], \end{aligned}$$

where $B = \bigcup_{i=1}^n (a_i, b_i]$ with $(a_i, b_i] \cap (a_j, b_j] = \phi$.

Definition 2.2.2. If $\{(a_i, b_i)\}_{i=1,2,\dots}$ is a sequence of non-overlapping or disjoint semi-open intervals, then

$$l(\cup_{i=1}^{\infty} (a_i, b_i]) = \sum_{i=1}^{\infty} l(a_i, b_i]$$

where l denotes the length.

If $(a, b] = \cup_{i=1}^{\infty} (a_i, b_i]$, where $\{(a_i, b_i)\}_{i=1,2,\dots}$ is a sequence of disjoint semi-open intervals, then

$$\begin{aligned} b - a &= \sum_{i=1}^{\infty} l((a_i, b_i]) \\ &= \lim_{n \rightarrow \infty} \sum_{i=1}^n l((a_i, b_i]). \end{aligned} \tag{2.2.1}$$

Given that $\cup_{i=1}^{\infty} (a_i, b_i] = (a, b]$, where $(a_i, b_i] \cap (a_j, b_j] = \phi$.

So for any k , with $a < a_k, b_k < b$, we have

$$b - a > \sum_{i=1}^k (b_i - a_i) = \sum_{i=1}^k l((a_i, b_i]).$$

Taking $k \rightarrow \infty$, we have

$$\sum_{i=1}^{\infty} l((a_i, b_i]) \leq b - a.$$

Now suppose

$$(a, b] \subset \cup_{i=1}^{\infty} (a_i, b_i] = \cup_{i=1}^{\infty} I_i.$$

We now show that $b - a \leq \sum_{i=1}^{\infty} l(I_i)$.

For any small $\epsilon > 0$,

$$[a + \epsilon/2, b] \subset (a, b] \subset \cup_{i=1}^{\infty} (a_i, b_i] \subset \cup_{i=1}^{\infty} (a_i, b_i + \epsilon/2^{i+1}).$$

Then by the Heine-Borel Theorem, there exists a k such that

$$\begin{aligned}
[a + \epsilon/2, b] &\subset \cup_{i=1}^k (a_i, b_i + \epsilon/2^{i+1}) \\
i.e., b - a - \epsilon/2 &\leq \sum_{i=1}^k (b_i - a_i) + \epsilon \sum_{i=1}^k \frac{1}{2^{i+1}} < \sum_{i=1}^k (b_i - a_i) + \epsilon/2 \\
i.e., b - a &< \sum_{i=1}^{\infty} (b_i - a_i) + \epsilon.
\end{aligned}$$

Since ϵ is arbitrary, so Equation 2.0.1 is correct.

From the above we get the following as a consequence

Proposition 2.2.1. *If $[a, b] \subset \cup_{i=1}^k (a_i, b_i)$, then*

$$b - a \leq \sum_{i=1}^k (b_i - a_i), k \geq 1.$$

Let \mathcal{F} be the class of all *Borel* sets such that each set can be expressed as union of atmost countable disjoint semi-open intervals (say I_i) *i.e.*, it is the collection

$\{B : B \text{ is the Borel sets and } B \text{ is the union of atmost countable semi - open intervals.}\}$

We have the following basic properties of l

1. l is non-negative.
2. l is finite additive on \mathcal{F} , *i.e.*, if $A, B \in \mathcal{F}$, then $l(A \cup B) = l(A) + l(B)$, where $A \cap B = \phi$.
3. l is countably addtive on \mathcal{F} *i.e.*, if $\{B_1, B_2, \dots\}$ is a sequence of sets in \mathcal{F} such that $B_i \cap B_j = \phi$, then $l(\cup_{i=1}^{\infty} B_i) = \sum_{i=1}^{\infty} l(B_i)$.

Definition 2.2.3 (Borel σ -field). The smallest σ -field containing all open intervals and semi open intervals is called the Borel σ -field.

Theorem 2.2.2 (Caratheodory Extension Theorem). *Let \mathcal{F} be a class (say a field) containing all semi-open intervals such that a function μ on \mathcal{F} satisfies the non-negativity and countable subadditivity properties. Then μ can be extended to a measure on the Borel σ -field. This extension is unique.*

The above theorem ensures that l extends to a measure λ on the Borel σ -field, where λ is called the Lebesgue measure which has the following properties.

1. Lebesgue measure of a finite set is zero.
2. $\lambda(x + cA) = c\lambda(A)$.

It is important to note the following

1. Probability is a finite measure.
2. Lebesgue measure is not finite, but it is σ -finite as $\mathbb{R} = \cup_{n \in \mathbb{I}} (n, n + 1]$.

Definition 2.2.4. Let \mathcal{C} be a class of subsets of \mathbb{R} . Then $\sigma(\mathcal{C})$ is the smallest σ -field containing the sets in \mathcal{C} , or $\sigma(\mathcal{C})$ is generated by \mathcal{C} .

Theorem 2.2.3. For any class of sets \mathcal{C} , $\mathcal{C} \cap B = \{A \cap B : A \in \mathcal{C}\}$. Let B be any Borel set. Then $\sigma(\mathcal{C}) \cap B = \sigma(\mathcal{C} \cap B)$.

Definition 2.2.5 (Counting Measure). Let S be a finite or countable set and \mathcal{F} be a field or σ -finite of subsets on S . Then μ on (S, \mathcal{F}) to be

$$\mu(B) = \begin{cases} \text{number of elements in } B, & \text{for any finite set } B \in \mathcal{F} \\ \infty, & \text{otherwise} \end{cases}$$

Remark 2.2.1. Counting measure and Lebesgue measure are mutually singular and length of an unbounded set is infinite.

2.3 Distribution Functions

We now generalize the notion of length in the following manner

Definition 2.3.1. Let $F : \mathbb{R} \rightarrow (0, \infty)$ be a right continuous monotonically non-decreasing

function F , we define a notion of length as follows:

$$l(a, b] = F(b) - F(a)$$

$$l(a, b) = F(b-)F(a)$$

$$l[a, b) = F(b-) - F(a-)$$

$$l[a, b] = F(b) - F(a-)$$

where $F(x-) = \lim_{\substack{y \rightarrow x \\ y < x}} F(y)$

Theorem 2.3.1. *If $(a, b] \subset \cup_{i=1}^{\infty} I_i$, where $I_i = (a_i, b_i]$ and $I_i \cap I_j = \phi$, then*

$$F(b) - F(a) \leq \sum_{i=1}^{\infty} \{F(b_i) - F(a_i)\}.$$

Proof. Let $\epsilon > 0$ be given. Then we have

$$[a + \delta, b] \subset (a, b] \subset \cup_{i=1}^{\infty} (a_i, b_i] \subset \cup_{i=1}^{\infty} (a_i, b_i + \delta_i)$$

where $F(a + \delta) - F(a) \leq \epsilon/2$ and $F(b + \delta_i) - F(b) \leq \epsilon/2^{i+1}$.

By the Heine-Borel Theorem, there exists an N such that

$$\begin{aligned} [a + \delta, b] &\subset \cup_{i=1}^N (a_i, b_i + \delta_i) \\ \Rightarrow F(b) - F(a + \delta) &\leq \sum_{i=1}^N (F(b_i + \delta_i) - F(a_i)) \\ \Rightarrow F(b) - F(a + \delta) &\leq \sum_{i=1}^N (F(b_i + \delta_i) - F(b_i)) + \sum_{i=1}^N (F(b_i) - F(a_i)) \\ &\leq \sum_{i=1}^N \epsilon/2^{i+1} + \sum_{i=1}^N (F(b_i) - F(a_i)) \\ &\leq \epsilon/2 + \sum_{i=1}^N (F(b_i) - F(a_i)). \end{aligned}$$

Now

$$\begin{aligned} F(b) - F(a) &= F(b) - F(a + \delta) + F(a + \delta) - F(a) \\ &\leq F(b) - F(a + \delta) + \epsilon/2 \\ &\leq \sum_{i=1}^{\infty} (F(b_i) - F(a_i)) + \epsilon \quad \forall \epsilon > 0. \end{aligned}$$

Since ϵ is arbitrary, hence it completes the rest of the proof. \square

Lemma 2.3.2. *If $[x, y] \subset \cup_{i=1}^N (a_i, b_i)$, then $F(y) - F(x) \leq \sum_{i=1}^N (F(b_i) - F(a_i))$, where F is a non-decreasing function.*

Definition 2.3.2 (Probability Distribution). A probability distribution is a (probability) measure on $(\mathbb{R}, \mathcal{B})$.

Note 2.3.1. F_0 = the field of all possible finite disjoint union of semi-open (right closed) intervals.

$$B \in F_0 \Rightarrow B = \cup_{i=1}^k (a_i, b_i], k \geq 1, \text{ with } (a_i, b_i] \cap (a_j, b_j] = \phi.$$

Definition 2.3.3. The definition of length on F_0 to be

1. $l(B) = \sum_{i=1}^k (F(b_i) - F(a_i))$, where $B = \cup_{i=1}^k (a_i, b_i]$,
2. $l(B) = \sum_{i=1}^{\infty} (F(b_i) - F(a_i))$, where $B = \cup_{i=1}^{\infty} (a_i, b_i]$.

The above definition holds due to the following theorem.

Theorem 2.3.3. *If $(a, b] = \cup_{i=1}^{\infty} (a_i, b_i]$ such that $(a_i, b_i] \cap (a_j, b_j] = \phi$, then $F(b) - F(a) = \sum_{i=1}^{\infty} (F(b_i) - F(a_i))$.*

Definition 2.3.4 (Distribution Function). A distribution function F is non-negative monotonically non-decreasing right continuous function such that $F(-\infty) = 0, F(\infty) = 1$, where $F : \mathbb{R} \rightarrow [0, 1]$.

Given a distribution function F we have

$$\begin{aligned}
l(x) &= l(\cap_{n=1}^{\infty} (x - \frac{1}{n}, x]) \\
&= \lim_{n \rightarrow \infty} l(x - \frac{1}{n}, x] \\
&= \lim_{n \rightarrow \infty} F(x) - F(x - \frac{1}{n}, x) \\
&= F(x) - F(x-) \\
&= \text{Jump or probability mass at } x.
\end{aligned}$$

Definition 2.3.5 (Continuous distribution). If the distribution function is continuous, then the probability distribution is called the continuous probability distribution.

Note 2.3.2. For any continuous distribution, probability or measure of any interval (a, b) or $(a, b]$ or $[a, b]$ or $[a, b)$ is always $F(b) - F(a)$.

Theorem 2.3.4. *The number of discontinuity points of a monotonic function is at most countable.*

Note 2.3.3. For any distribution function

$$0 \leq F(x) - F(x-) \leq F(x) \leq F(\infty) = 1.$$

Definition 2.3.6 (Purely discrete distribution). Let S be the set of discontinuity points of a distribution function F with S is at most countable. If $\sum_{x \in S} F(x) - F(x-) = 1$, the probability distribution with distribution F is said to be **purely discrete distribution** or **atomic distribution** where S is called the support of the distribution and its elements are called the **mass points**.

Remark 2.3.1. We denote a probability measure by P and uniquely determined by the distribution function F . Also we have $P\{(a, b]\} = F(b) - F(a)$.

Lemma 2.3.5. 1. *If S is the support of a probability distribution P , then $P(\{y\}) = 0$, for $y \notin S$, i.e., $P(S^c) = 0$.*

2. If $P(\{y\}) > 0$, then $P(S \cup \{y\}) > 1$.

3. If P is a purely discrete distribution with support S , then

$$P(B) = P(B \cap S) = \sum_{x \in S \cap B} P(\{x\}) = \sum_{x \in S \cap B} \{F(x) - F(x-)\}.$$

Proposition 2.3.6 (Regularity Property). *Let P be probability distribution on $(\mathbb{R}, \mathcal{B})$. Then for any Borel set B and $\epsilon > 0$, \exists a closed set C and an open set O such that $C \subset B \subset O$ and $P(O - C) < \epsilon$.*

If P is a probability distribution such that $P(B) = 0$ or $P(B) = 1$ for every Borel set B , then $P(\{x_0\}) = 1$. For every probability distribution P on $(\mathbb{R}, \mathcal{B})$, there exists a smallest closed set S such that $P(S) = 1$.

Remark 2.3.2. Such an S is called the support of the distribution.

Unit 3

Integration Theory

Integration theory is a very important aspect of all of analysis specially in probability theory as we shall see in this unit. The normal school level integration that we do is what is called *Riemann integration*, in this section we shall get acquainted with other theories of integration which shall be very useful in our study of probability theory.

3.1 Riemann Integration

In our usual Riemann integration the necessary condition to define $\int_a^b f(x)dx$, is that the function f should be bounded while a sufficient condition to define $\int_a^b f(x)dx$, is that the function f should be continuous.

Remark 3.1.1. Boundedness is not sufficient condition to define Riemann Integration. Here is the counter example of a function which is bounded but not Riemann Integrable,

$$f(x) = \begin{cases} 1, & x \text{ is rational} \\ 0, & x \text{ is irrational.} \end{cases}$$

Note 3.1.1. Let $P = \{x_0, x_1, \dots, x_n\}$, where $x_0 = a, x_n = b$ and $x_0 < x_1 < \dots < x_n$. Let P_1, P_2 be two partition of $[a, b]$. We say P_2 is a refinement of P_1 if $P_2 \supset P_1$.

Definition 3.1.1 (Upper Sums, Lower Sums). Suppose f is bounded on $[a, b]$, that is there exists an m, M such that $m < f(x) < M, \forall x \in [a, b]$. Given a partition $P = \{x_0 =$

$a, \dots, x_i, \dots, x_n = b\}$. Let

$$m_i = \inf_{x \in [x_{i-1}, x_i]} f(x) \quad (3.1.1)$$

$$M_i = \sup_{x \in [x_{i-1}, x_i]} f(x). \quad (3.1.2)$$

Then

$$L(P) = \sum_i (x_i - x_{i-1})m_i,$$

$$L(P) = \sum_i (x_i - x_{i-1})M_i.$$

Also, for each P , we have

$$L(P) \leq U(P).$$

The following are some of the important properties of the upper sums and lower sums that we have just defined.

Lemma 3.1.1. *Let P_1, P_2 be any two partition on $[a, b]$. Then $L(P) \leq U(P)$.*

Lemma 3.1.2. *If $P_1 \subset P_2$, then $L(P_1) \leq L(P_2)$ and $U(P_1) \leq U(P_2)$.*

Proposition 3.1.1. *If a function is continuous, then that function is Riemann Integrable.*

Theorem 3.1.3. *If f is continuous on $[a, b]$, then f is uniformly continuous on $[a, b]$.*

Proposition 3.1.4. *If f is bounded on P , then*

$$\sup_{x \in P} f(x) - \inf_{x \in P} f(x) \geq \sup_{x, y \in P} |f(x) - f(y)|$$

Note 3.1.2. We denote the norm of a partition as the length of the greatest subinterval.

Theorem 3.1.5. *A Riemann Integrable function can have at most countably many discontinuity points.*

Example 3.1.1. $f(x) = \begin{cases} 0, & 0 < x < 1/2 \\ 1, & 1/2 \leq x < 1 \end{cases}$

We have $\int_0^1 f(x)dx = \int_{1/2}^1 dx = \frac{1}{2}$. Therefore the function is Riemann Integrable.

We shall now state some important properties of integration.

1. $\int_a^b f(x)dx$ is linear, that is

$$\int_a^b \{cf(x) + dg(x)\}dx = c \int_a^b f(x)dx + d \int_a^b g(x)dx.$$

2. If $f(x) \leq g(x) \forall x \in [a, b]$, then

$$\int_a^b f(x)dx \leq \int_a^b g(x)dx.$$

3.2 Riemann-Steiltjes Integration

After the preliminaries on Riemann integration, we now move on to another theory of integration called the Riemann-Stieltjes integration theory. We begin with the following definitions of upper and lower sums in the context of Riemann-Steiltjes integration.

Definition 3.2.1. To define $\int_a^b f(t)dF(t)$, we assume (i) f is bounded, (ii) F is monotonically non-decreasing. For a partition, $P = \{x_0 = a, x_1, \dots, x_n = b\}$, we have $F(x_i) - F(x_{i-1}) \geq 0$. And we define Upper sum and Lower sum as

$$U(P, f, F) = \sum_{i=1}^n M_i \{F(x_i) - F(x_{i-1})\},$$

$$L(P, f, F) = \sum_{i=1}^n m_i \{F(x_i) - F(x_{i-1})\}.$$

where m_i and M_i are defined by (3.1.1) and (3.1.2) respectively. For any partition P_1, P_2 , we have

$$L(P, f, F) \leq U(P, f, F).$$

Moreover, if $P_1 \subseteq P_2$, then

$$U(P, f, F) \geq U(P_1, f, F) \text{ and}$$

$$L(P, f, F) \leq L(P_1, f, F).$$

We are now in a position to define what is a Riemann-Steiltjes integral.

Definition 3.2.2 (Riemann-Steiltjes Integral). For every $\epsilon > 0$, there exists a partition P_ϵ of $[a, b]$ such that for every refinement P of P_ϵ , $|U(P, f, F) - L(P, f, F)| < \epsilon$. In that case $f \in R(F)[a, b]$ that is f is Reimann-Stieltjes Integrable with respect to f on $[a, b]$.

We also have the following alternate definiton of the above

Definition 3.2.3. Suppose $f \in R(F)[a, b]$, then

$$\lim_{\delta \rightarrow 0} S(P_\delta, f, F) = \int_a^b f(x) dF(x),$$

where

$$S(P_\delta, f, F) = \sum_{i=1}^k f(t_i) \{F(x_i) - F(x_{i-1})\}$$

where t_i is a point in the i^{th} subinterval of the partition P of Definition 3.2.1.

Note 3.2.1. For any partition P , we have $L(P, f, F) \leq S(P, f, F) \leq U(P, f, F)$.

We now state the following properties of Riemann-Steiltjes integration.

1. If $f, g \in R(F)[a, b]$, then $cf + g \in R(F)[a, b]$.
2. If $f \leq g$ on $[a, b]$, then

$$\int_a^b f(x) dF(x) \leq \int_a^b g(x) dF(x).$$

3. Suppose F, G are two monotonic functions, if $f \in R(F)[a, b]$ and also $f \in R(G)[a, b]$ then

$$\int_a^b f(x) d(F + G)(x) = \int_a^b f(x) dF(x) + \int_a^b f(x) dG(x).$$

4. $\int_a^b f(x) dF(x) = - \int_a^b f(x) dF(x)$.

The following theorem is one of the most important results in integration theory.

Theorem 3.2.1 (Integration by parts). *If $f \in R(F)$ on $[a, b]$, then $F \in R(f)$ on $[a, b]$ and*

$$\int_a^b f(x)dF(x) + \int_a^b F(x)df(x) = F(b)f(b) - F(a)f(a).$$

Theorem 3.2.2. *If f is continuous on $[a, b]$ and F is monotonically non-decreasing, then $f \in R(F)$ on $[a, b]$.*

Exercise 3.2.1. If f is a continuous function on $[a, b]$ and $F(x) = \begin{cases} 0, & 0 \leq x < \frac{1}{2}, \\ 1, & \frac{1}{2} \leq x \leq 1. \end{cases}$

Then find the value of $\int_0^1 f(x)dF(x) = \lim_{\delta \rightarrow 0} S(P, f, F)$.

Lemma 3.2.3. *Let g be a continuous strictly monotonic function on $[a, b]$, then $g\{[a, b]\} = \{g(y) : a \leq y \leq b\}$ is an interval.*

Lemma 3.2.4. *A strictly monotonic continuous function g on $[a, b]$ is invertible.*

Lemma 3.2.5. *The function g^{-1} is also a continuous function.*

The above three lemmas give us the following important result.

Theorem 3.2.6 (Change of variables). *Let $f \in R(F)$ on $[a, b]$ and let g be a strictly monotonic continuous function defined on an interval S having endpoints c and d . Assume that $a = g(c)$, $b = g(d)$. Let h and G be the composite function defined as follows*

$$h(x) = f[g(x)], \quad G(x) = F[g(x)], \quad \text{if } x \in S.$$

Then $h \in R(G)$ on S and we have $\int_a^b f dF = \int_a^b h dG$. That is,

$$\int_{g(c)}^{g(d)} f(t)dF(t) = \int_c^d f[g(x)]d\{F[g(x)]\}.$$

Theorem 3.2.7. *Given $a < c < b$. Define F on $[a, b]$ as $F(x) = \begin{cases} F(a), & a \leq x < c \\ F(b), & c < x \leq b \end{cases}$.*

Suppose f is a function on $[a, b]$ such that atleast one of the function on F is a right continuous at c and the other is left continuous at c , then $f \in R(F)$ on $[a, b]$ and

$$\int_a^b f(x)dF(x) = f(b)\{F(c+) - F(c-)\}.$$

Theorem 3.2.8. Any finite sum $\sum_{k=1}^n a_k$ can be written as

$$\int_0^n f(x)d[x]$$

where $f(x) = a_k$, $k - 1 < x \leq k$, $k = 1, 2, \dots, n$; $f(0) = 0$ and $[x]$ is the greatest integer less than or equal to x .

Theorem 3.2.9 (Euler Summation Formula). If f has a continuous derivative f' on $[a, b]$, then we have

$$\sum_{a < n \leq b} f(n) = \int_a^b f(x)dx + \int_a^b f'(x)(x - [x])dx + f(a)(a - [a]) - f(b)(b - [b]).$$

We can prove the above result using the change of variable theorem and the two other theorems that follow it.

Theorem 3.2.10. Assume $f \in R(F)$ on $[a, b]$ and assume that F has a continuous derivative F' on $[a, b]$. Then the Riemann integral $\int_a^b f(x)F'(x)dx$ exists and we have

$$\int_a^b f(x)dF(x) = \int_a^b f(x)F'(x)dx.$$

3.3 Lebesgue Integration

Having learnt Riemann and Riemann-Stieltjes integration we shall now go into another very important theory of integration called Lebesgue integration, which was devised by Henri Lebesgue, a French mathematician in the twentieth century.

To motivate the need for this new theory of integration we look at the following function called Dirichlet's function

$$f(x) = \begin{cases} 1, & x \text{ is rational,} \\ 0, & x \text{ is irrational.} \end{cases}$$

We know that for any monotonic function F , $\int_0^1 f(x)dF(x)$ is not defined in Riemann-Stieltjes sense.

Also we have $f(x) = 1 \times I_Q(x) + 0 \times I_{Q^c}(x)$ where I stands for indication or characteristic function, defined by $I_Q(x) = \begin{cases} 1, & x \in Q, \\ 0, & x \notin Q. \end{cases}$

Definition 3.3.1 (Simple function). A function f is said to be simple if it is measurable and its range is a finite set.

Definition 3.3.2. Let $(\mathbb{R}, \mathcal{B})$ be a measurable space, where \mathcal{B} is a Borel σ -field. Let λ be a measure on $(\mathbb{R}, \mathcal{B})$. Consider the map $f : \mathbb{R} \rightarrow E$, where E is a finite subset of \mathbb{R} and $f^{-1}(B) = \{x \in \mathbb{R} : f(x) \in B\} \in \mathcal{B}$. Then we have

$$\begin{aligned} \int_B f d\lambda &= \int_B (1 \times I_Q(x) + 0 \times I_{Q^c}(x)) d\lambda(x) \\ &= 1 \times \lambda(B \cap Q) + 0 \times \lambda(B \cap Q^c). \end{aligned}$$

Definition 3.3.3. If f is simple function with range $\{a_1, \dots, a_k\}$ on a measurable space $(\mathbb{R}, \mathcal{B}, \lambda)$, then

$$\int_B f d\lambda = \sum_{i=1}^k a_i \lambda(B \cap A_i)$$

where $A_i = \{x \in \mathbb{R} : f(x) = a_i\}$.

Definition 3.3.4. We have

$$\begin{aligned} f(x) &= f^+(x) - f^-(x) \\ f^+(x) &= \max\{f(x), 0\} \\ f^-(x) &= \max\{-f(x), 0\} \\ \int_B f d\lambda &= \int_B f^+ d\lambda - \int_B f^- d\lambda. \end{aligned}$$

Remark 3.3.1. In general $\lim_{n \rightarrow \infty} \int_0^1 f_n(x) dx = \int_0^1 f(x) dx$ is not true. Following example shows that this condition is violated.

Example 3.3.1. $f_n(x) = \begin{cases} n, & 0 \leq x < \frac{1}{n} \\ 1, & \frac{1}{n} < x \leq 1 \end{cases}$. In this example $\lim_{n \rightarrow \infty} \int_0^1 f_n(x) dx \neq \int_0^1 f(x) dx$.

Remark 3.3.2. If

$$|f_n(x)| \leq g(x) \quad (\text{free of } n)$$

and $\int_0^1 g(x)dx$ is finite, then

$$\lim_{n \rightarrow \infty} \int_0^1 f_n(x)dx = \int_0^1 f(x)dx.$$

We now state a very important result in the theory of Lebesgue integration.

Theorem 3.3.1 (Dominated Convergence Theorem). *Suppose $\{f_n\}$ is a sequence of integrable functions on $(\mathbb{R}, \mathcal{B}, \lambda)$ such that $f_n(x) \rightarrow f(x)$ for almost all x and $|f_n(x)| \leq g(x)$ for almost all x such that g is an integrable function. Then f is integrable and $\int f d\lambda = \lim_{n \rightarrow \infty} \int f_n d\lambda$.*

The following example illustrates the need for the boundedness condition in the Dominated Convergence Theorem.

Example 3.3.2. Define $f : [0, \infty] \rightarrow [0, \infty]$ by

$$f(x) = \begin{cases} 0, & 0 \leq x \leq 1 \\ 1, & 1 < x \leq 1 + \frac{1}{2^3} \\ 0, & 1 + \frac{1}{2^3} < x \leq 2 \\ 2, & 2 < x \leq 2 + \frac{1}{2^3} \\ 0, & 2 + \frac{1}{2^3} < x \leq 3 \\ \vdots & \\ n, & n < x \leq n + \frac{1}{(n+1)^3} \\ \vdots & \end{cases}$$

Here we have

$$\int_0^\infty f(x)dx = \sum_{n=1}^{\infty} n \cdot \frac{1}{(n+1)^3} \leq \sum_{n=1}^{\infty} \frac{1}{(n+1)^2} < \infty.$$

The above series is convergent, but the function is unbounded.

Below we state some important results without proof. Almost all the results below follow from definitions or from the results stated earlier to them.

Theorem 3.3.2. Suppose f is a non-negative finite valued measurable function on a measure space $(\Omega, \mathcal{F}, \mu)$. Then there exists a sequence of non-negative simple functions $\{f_n\}_{n=1,2,\dots}$ such that $f_n \leq f_{n+1}$ (pointwise) and if $f(x) < \infty \forall x \in \Omega$, then $f_n(x) \rightarrow f(x) \forall x$.

Lemma 3.3.3. If f and g are two simple functions on $(\Omega, \mathcal{F}, \mu)$ such that $f(x) \leq g(x)$. Then $\int f d\mu \leq \int g d\mu$.

Remark 3.3.3. For any non-negative measurable function f on $(\Omega, \mathcal{F}, \mu)$ there exists a sequence $\{f_n\}_{n=1,2,\dots}$ of non-negative simple functions such that $f_n \leq f_{n+1}$ and $f_n \rightarrow f$ a.e. Then $\int f_n d\mu \leq \int f_{n+1} d\mu$.

Definition 3.3.5. For any non-negative measurable function f on $(\Omega, \mathcal{F}, \mu)$ satisfying $\mu(\{x : f(x) = \infty\}) = 0$, define $\int_B f d\mu = \lim_{n \rightarrow \infty} \int_B f_n d\mu$ for $B \in \mathcal{F}$

Definition 3.3.6. For any measurable function f on $(\Omega, \mathcal{F}, \mu)$ satisfying $\mu(\{x : f(x) = \pm\infty\}) = 0$, define $\int_B f d\mu = \int_B f^+ d\mu - \int_B f^- d\mu$ provided $\int_\Omega f^+ d\mu$ or $\int_\Omega f^- d\mu < \infty$ where f^+ and f^- are integrable functions.

Below we list some of the important properties of Lebesgue integrals.

If f, g are integrable functions on $(\Omega, \mathcal{F}, \mu)$, then

1. $\int (af + g) d\mu = a \int f d\mu + \int g d\mu$.
2. $|\int f d\mu| \leq \int |f| d\mu$.
3. If $f \leq g$, then $\int f d\mu \leq \int g d\mu$.
4. If $\int_B f d\mu \leq \int_B g d\mu \forall B \in \mathcal{F}$, then $f \leq g$ almost everywhere.
5. If $\int_B f d\mu = \int_B g d\mu \forall B \in \mathcal{F}$, then $f = g$ almost everywhere.

The following are some easy consequences of the above properties.

Lemma 3.3.4. If f be a simple function on $(\Omega, \mathcal{F}, \mu)$ such that $\int_B f d\mu \geq 0 \forall B \in \mathcal{F}$, then $f \geq 0$.

Corollary 3.3.5. If f is simple function on $(\Omega, \mathcal{F}, \mu)$ such that $\int_B f d\mu = 0 \quad \forall B \in \mathcal{F}$, then $f = 0$.

Lemma 3.3.6. If f is an integrable function on $(\Omega, \mathcal{F}, \mu)$, then \exists a sequence of simple functions $\{f_n\}_{n=1,2,\dots}$ such that $\int_B f d\mu = \lim_{n \rightarrow \infty} \int_B f_n d\mu$.

Definition 3.3.7. A measurable function f on $(\Omega, \mathcal{F}, \mu)$ is said to be integrable if

$$\int f_+ d\mu < \infty \quad \text{and} \quad \int f_- d\mu < \infty$$

$$\text{or,} \quad \int |f| d\mu < \infty \Leftrightarrow f \in L_1(\Omega, \mathcal{F}, \mu)$$

$$\text{or,} \quad \int f d\mu \text{ exists} \Leftrightarrow \int f_+ d\mu < \infty \quad \text{or} \quad \int f_- d\mu < \infty.$$

Proposition 3.3.1. If f is integrable, then f is finite almost everywhere, i. e. $\mu\{\{x\} : |f(x)| = \infty\} = 0$.

Proposition 3.3.2. If $f \geq 0$, then $\int f d\mu = 0$ implies $f = 0$ almost everywhere.

Theorem 3.3.7. Suppose that f and g are two integrable functions on $(\Omega, \mathcal{F}, \mu)$ such that $\int_B f d\mu \leq \int_B g d\mu, \quad \forall B \in \mathcal{F} \Rightarrow f \leq g$ almost everywhere.

Theorem 3.3.8. Suppose that f and g are two integrable functions on $(\Omega, \mathcal{F}, \mu)$ such that $\int_B f d\mu = \int_B g d\mu, \quad \forall B \in \mathcal{F} \Rightarrow f = g$ almost everywhere.

Theorem 3.3.9. Let f be a nonnegative integrable function and $\int_{\Omega} f d\mu < \infty$. Define $\lambda(B) = \int_B f d\mu$. Then λ is a measure on (Ω, \mathcal{F}) .

Theorem 3.3.10. Let f be a nonnegative measurable function on $(\Omega, \mathcal{F}, \mu)$. Let $\{f_n\}_{n=1}^{\infty}$ be a sequence of non-negative simple function on $(\Omega, \mathcal{F}, \mu)$ such that $f_n(x) \uparrow f(x)$. Then $\lim_{n \rightarrow \infty} \int_{\Omega} f_n d\mu = \sup\{\int_{\Omega} s d\mu : 0 \leq s \leq f\}$, where s is a non-negative simple valued functions on $(\Omega, \mathcal{F}, \mu)$.

Definition 3.3.8. If f is a non-negative measurable function, then we define

$$\int_{\Omega} f d\mu = \sup \left\{ \int_{\Omega} s d\mu : 0 \leq s \leq f \right\},$$

where s is a non-negative simple valued functions on $(\Omega, \mathcal{F}, \mu)$.

Theorem 3.3.11. If $f \leq g$, then

$$\int f d\mu \leq \int g d\mu \tag{3.3.1}$$

in the sense that if $\int g d\mu < \infty$, then $\int f d\mu$ exists and (3.3.1) holds. Similarly, if $\int f d\mu > -\infty$, then $\int g d\mu$ exists and (3.3.1) holds.

We end this discussion with the following famous theorem.

Theorem 3.3.12 (Monotone Convergence Theorem). Let $\{f_n\}_{n=1}^{\infty}$ be a sequence of non-negative measurable functions $(\Omega, \mathcal{F}, \mu)$ such that $f_n(x) \uparrow f(x) \forall x$. Then $\int f d\mu = \lim_{n \rightarrow \infty} \int f_n d\mu$.

Unit 4

Probability Measure and Random Variables

In this unit we shall study what are probability measures and random variables. The objective here is to demonstrate them as integrals.

4.1 Probability Measure

Let f be a measurable function on $(\Omega, \mathcal{F}, \mu)$, where $f : \Omega \rightarrow \mathcal{R}$. The objective is to define a measure on $(\mathcal{R}, \mathcal{B})$. If f is a measurable function on $(\mathcal{R}, \mathcal{B})$, then $B \in \mathcal{B} \Rightarrow f^{-1}(B) \in \mathcal{F}$. Thus for $B \in \mathcal{B}$ we define $\mu_0(B) = \mu(f^{-1}(B))$. Then μ_0 is a measure on $(\mathcal{R}, \mathcal{B})$ is called probability measure.

Definition 4.1.1. If μ is a probability measure and f is a random variable, then μ_0 is called the probability distribution of f , where μ_0 is defined as $\mu_0(B) = \mu(f^{-1}(B))$.

With the above two definitions and the results from the previous unit we have the following.

Theorem 4.1.1. Let f be a measurable function on $(\Omega, \mathcal{F}, \mu)$ and g be an integrable function on $(\mathcal{R}, \mathcal{B}, \mu_0)$. Then $g(f)$ is integrable on $(\Omega, \mathcal{F}, \mu)$ and

$$\int_{f^{-1}(B)} g(f) d\mu = \int_B g d\mu,$$

where $\mu_0(B) = \mu(f^{-1}(B))$.

4.2 Random Variable and Mathematical Expectation

Definition 4.2.1 (Random Variable, Mathematical Expectation). Let (Ω, \mathcal{F}, P) be a probability space. A random variable on X on (Ω, \mathcal{F}, P) is a real valued measurable mapping on (Ω, \mathcal{F}, P) . Then the mathematical expectation $E(X)$ is defined as

$$\begin{aligned} E(X) &= \int_{\Omega} X dP \\ &= \int_{-\infty}^{\infty} x dF(x) \\ &= \int_{\mathbb{R}} x d\mu, \end{aligned}$$

where μ is a measure on $(\mathbb{R}, \mathcal{B})$ defined by the distribution function.

Remark 4.2.1. Given a random variable X , we have the distribution function F defined as $F(x) = P(X \leq x)$. Here $g(x) = x$ and $f = X$. So,

$$\int_{\Omega} X dP = \int_{\Omega} g(f) dP = \int_{\mathbb{R}} x d\mu.$$

We now state some important properties of mathematical expectation which can be proved very easily by applying the definition.

1. $E(aX + b) = aE(X) + b.$

2.

$$X \geq 0, \quad E(X) = 0 \Rightarrow X = 0 \quad a. e.$$

$$\Leftrightarrow P(X > 0) = 0$$

$$\Leftrightarrow P(X = 0) = 1.$$

3. $E(XI_A) = 0, \quad \forall A \in \mathcal{F} \Rightarrow P(X = 0) = 1.$

4. $X \leq Y \Rightarrow E(X) \leq E(Y).$

It is important to note that X is a random variable on a probability space (Ω, \mathcal{F}, P) if and only if X is a measurable function on the measure space (Ω, \mathcal{F}, P) .

We also have the following definition.

Definition 4.2.2. A random variable X is said to be integrable if $E(|X|) = \int_{-\infty}^{\infty} |x|dP < \infty$

4.3 Expectations of Some Distributions

We now mention below the mathematical expectations of some well known distributions.

1. If $X \sim$ Bernouli distribution (P), then the distribution function F is defined as

$$F(x) = \begin{cases} 0, & x < 0, \\ p, & 0 \leq x \leq 1, \\ 1, & x \geq 1. \end{cases}$$

Then $E(X) = \int_{-\infty}^{\infty} xdF(x) = 1 - p$.

2. If X is a random variable of a purely discrete distribution with support $\{x_1, x_2, \dots, x_n\}$, then $E(X) = \int_{-\infty}^{\infty} xdF(x) = \sum_{n=1}^k x_i(F(x_i) - F(x_{i-1})) = \sum_{n=1}^k x_iP(X = x_i)$.
3. If $X \sim$ Binomial distribution (n, P), that means $P(X = k) = \binom{n}{k}P^k(1 - P)^{n-k}$, $k = 0, 1, 2, \dots, n$, then $E(X) = \sum_{k=0}^n \binom{n}{k}P^k(1 - P)^{n-k} = nP$.
4. If $X \sim$ Poisson distribution (λ), then $E(X) = \sum_{k=0}^{\infty} ke^{-\lambda} \frac{\lambda^k}{k!}$.
5. If $X \sim$ Normal distribution ($0, 1$) and $P(X = k) = 0$, then

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} xdF(x) \\ &= \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx. \end{aligned}$$

6. If $X \sim$ Exponential distribution (1), and $F(x) = 1 - e^{-x}$, $x \geq 0 \Rightarrow dF(x) = e^{-x}dx$, then $E(X) = \int_{-\infty}^{\infty} xdF(x) = \int_0^{\infty} xe^{-x}dx$.

7. If X is a continuous random variable with density f , that is $F(x) = \int_{-\infty}^x f(t)dt$, so $dF(x) = f(x)dx$, then $E(X) = \int_{-\infty}^{\infty} x dF(x) = \int_{-\infty}^{\infty} xf(x)dx$.

Unit 5

Inequalities involving probabilities

The study of inequalities in probability theory is very essential as they sometimes form the crux of the problem at hand that is to be solved. Inequalities inherently are quite beautiful and they never lose that beauty even if we put them in a probabilistic setting. This unit shall examine some basic and famous inequalities involving probabilities.

We begin with the following well known inequality and some of its corollaries.

Theorem 5.0.1 (Markov). *If X is a non-negative valued random variable on a probability space (Ω, \mathcal{F}, P) , then*

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

for any positive number a .

We have the following corollary of the above.

Corollary 5.0.2. *If X is a random variable on a probability space (Ω, \mathcal{F}, P) , then*

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

for any number $a > 0$.

We can generalize Markov's inequality in the following way.

Theorem 5.0.3. *If X is a non-negative valued random variable on a probability space (Ω, \mathcal{F}, P) and g is a non-negative measurable function on $(\mathbb{R}, \mathcal{B})$, then*

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

for any positive number a .

We now come to a very easy application of Markov's inequality given below.

Theorem 5.0.4 (Chebyshev). *If X is a non-negative random variable on a probability space (Ω, \mathcal{F}, P) then we have*

$$P(|X - E(X)| > \epsilon) \leq \frac{\text{Var}(X)}{\epsilon^2}$$

for any positive ϵ .

Before proceeding further we define the following.

Definition 5.0.1 (Convex function). A function f is said to be convex if for $0 < t < 1$ we have $f(tx + (1-t)y) \leq tf(x) + (1-t)f(y)$, where x and y are real numbers.

We have now the following very famous result.

Theorem 5.0.5 (Jensen). *If X is a non-negative valued random variable on a probability space (Ω, \mathcal{F}, P) , then*

$$\psi(E(X)) \leq E(\psi(X))$$

where ψ is a convex function.

After the above inequalities we now turn our discussion to inequalities which involve probabilities of sums.

Theorem 5.0.6 (Hoeffding). *If X_1, X_2, \dots, X_n are independent random variables and $a_i \leq X_i \leq b_i$ then for any $\epsilon > 0$ we have*

$$P\left(\sum_{i=1}^n (X_i - E(X_i)) > n\epsilon\right) \leq e^{\frac{-n^2\epsilon^2}{2\sum_{i=1}^n (b_i - a_i)^2}}.$$

From the above we get the following corollary very easily.

Corollary 5.0.7. *Let X_1, X_2, \dots, X_n be i.i.d random variables with $a_i \leq X_i \leq b_i$ and probability one, then for any $\epsilon > 0$ we have*

$$P(|\bar{X} - E(X_1)| > n\epsilon) \leq e^{\frac{-n\epsilon^2}{2(b-a)^2}}$$

where $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$.

Theorem 5.0.8 (Cauchy-Schwarz). *If X is a random variable then we have $E(X^2) \leq (E(X))^2$.*

Theorem 5.0.9 (Normal Tail inequality). *If X follows $N(0, 1)$ and for $\epsilon > 0$ where X is a random variable then we have*

$$P(X > \epsilon) \leq \frac{1}{\epsilon} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\epsilon^2}.$$

Theorem 5.0.10. *If X follows $N(0, 1)$ and for $\epsilon > 0$ where X is a random variable then we have*

$$P(|X| > \epsilon) \leq \frac{2}{\epsilon\sqrt{2\pi}} e^{-\frac{1}{2}\epsilon^2}.$$

We now have the following corollary from the above two theorems.

Corollary 5.0.11. *If X_1, X_2, \dots, X_n are i.i.d random variables and X_i follows $N(0, 1)$ then \bar{X} also follows $N(0, 1)$ and hence $\sqrt{n}\bar{X}$ follows $N(0, 1)$ and we have*

$$P(|\sqrt{n}\bar{X}| > \epsilon) \leq \frac{2}{\epsilon\sqrt{2n\pi}} e^{-\frac{1}{2}n\epsilon^2}$$

where $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$.

We have the following simple result from analysis.

Proposition 5.0.12. *If $g(u) = -\theta u + \log(1 - \theta + \theta e^u)$ where $0 < \theta < 1$ then $g(0) = 0, g'(0) = 0$ and $g''(u) \leq \frac{1}{4}$.*

Using the above we get the following.

Lemma 5.0.13. *If X is a random variable and $E(X) = 0$ and $P(a \leq X \leq b) = 1$ then for any $s > 0$ we have*

$$E(e^{sX}) \leq e^{\frac{s^2(b-a)^2}{8}}.$$

The above lemma gives us another variant of the Hoeffding's inequality as stated below.

Theorem 5.0.14 (Hoeffding). *If X_1, X_2, \dots, X_n are independent random variables and $a_i \leq X_i \leq b_i$ then for any $\epsilon > 0$ we have*

$$P\left(\sum_{i=1}^n (X_i - E(X_i)) > n\epsilon\right) \leq e^{\frac{-2n^2\epsilon^2}{\sum_{i=1}^n (b_i - a_i)^2}}.$$

From the above we get the following corollary very easily.

Corollary 5.0.15. *Let X_1, X_2, \dots, X_n be i.i.d random variables with $a_i \leq X_i \leq b_i$ and probability one, then for any $\epsilon > 0$ we have*

$$P(|\bar{X} - E(X_1)| > n\epsilon) \leq 2e^{\frac{-2n\epsilon^2}{(b-a)^2}}$$

where $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$.

The following is a consequence of the above inequality.

Theorem 5.0.16. *Let $\{X_n\}_{n=1,2,\dots}$ be a sequence of independent Bernoulli (p) random variables then*

$$\frac{1}{n} \sum_{i=1}^n X_i \rightarrow p$$

almost surely.

We now come to an inequality involving medians.

Theorem 5.0.17 (Levy). *If X_1, X_2, \dots, X_n are i.i.d random variables then*

$$P\left\{\max_{1 \leq j \leq n} (S_j - m(S_j - S_n)) \leq \epsilon\right\} \leq 2P(S_n \geq \epsilon)$$

where $\epsilon > 0$ and $S_j = \sum_{i=1}^j X_i$ and $m(Y)$ denotes the median of Y .

In this final part of the report we now mention two very famous inequalities related to expectation.

Definition 5.0.2. We define the norm of the expectation of a random variable as $\|X\|_p = (E(|X|^p))^{\frac{1}{p}}$ where $p > 0$.

Theorem 5.0.18 (Holder Inequality). *For any $1 < p < \infty$ and $1 < q < \infty$ such that $\frac{1}{p} + \frac{1}{q} = 1$ we have*

$$E(|XY|) \leq \|X\|_p \|Y\|_q$$

where X and Y are two random variables.

Proof. Let us define the following,

$$U = \frac{X}{\|X\|_p}, \quad V = \frac{Y}{\|Y\|_q}.$$

Now we have

$$E(|UV|) \leq \frac{E(|U|^p)}{p} + \frac{E(|V|^q)}{q} = \frac{1}{p} + \frac{1}{q} = 1.$$

Also

$$E(|U|^p) = E\left(\frac{|X|^p}{E(|X|^p)}\right) = 1.$$

Similarly we have $E(|V|^q) = 1$.

Again from the above definition we have

$$\begin{aligned} \|X\|_p &= (E(|X|^p))^{\frac{1}{p}} \\ \Rightarrow E\left(\left|\frac{XY}{\|X\|_p \|Y\|_q}\right|\right) &\leq 1 \\ \Rightarrow E(|XY|) &\leq \|X\|_p \|Y\|_q. \end{aligned}$$

□

Theorem 5.0.19 (Minkowski Inequality). *We have for any two random variables X and Y and $1 \leq p < \infty$*

$$\|X + Y\|_p \leq \|X\|_p + \|Y\|_p.$$

Proof. For $p = 1$, the proof is trivial, as $E(|X + Y|) \leq E(|X|) + E(|Y|)$, i.e., $\|X + Y\|_1 \leq \|X\|_1 + \|Y\|_1$.

For $1 < p$, $f(x) = x^p$ is a convex function on $(0, \infty)$ as $f'(x) \uparrow$.

Now we define like earlier

$$U = \frac{X}{\|X\|_p}, \quad V = \frac{Y}{\|Y\|_q}.$$

Let $\lambda = \frac{\|X\|_p}{\|X\|_p + \|Y\|_p}$, then we have $1 - \lambda = \frac{\|Y\|_p}{\|X\|_p + \|Y\|_p}$.

Now

$$\begin{aligned} E(|X + Y|^p) &\leq E(|X| + |Y|)^p, \quad p > 1 \\ &\leq E(\|X\|_p U + \|Y\|_p V)^p \\ &= E\{(\|X\|_p + \|Y\|_p)^p (\lambda U + (1 - \lambda)V)^p\} \\ &= (\|X\|_p + \|Y\|_p)^p E(\lambda U + (1 - \lambda)V)^p \\ &\leq (\|X\|_p + \|Y\|_p)^p \{\lambda E(U^p) + (1 - \lambda)E(V^p)\}. \end{aligned}$$

Note that $E(U^p) = 1$ and $E(V^p) = 1$. So from above we have

$$\begin{aligned} E(|X + Y|^p) &\leq (\|X\|_p + \|Y\|_p)^p \\ \Rightarrow \|X + Y\|_p &\leq \|X\|_p + \|Y\|_p. \end{aligned}$$

□

Bibliography

- [1] *T. M. Apostol*, *Mathematical Analysis*, 2nd ed., Narosa Publishing House, New Delhi, 2002.
- [2] *H. L. Royden, P. M. Fitzpatrick*, *Real Analysis*, 4th ed., Prentice Hall of India, 2010.

Index

- Archimedean Property, 2
- atomic distribution, 9

- Bernoulli distribution, 24
- Binomial distribution, 24
- Borel σ -field, 5

- Caratheodory Extension Theorem, 5
- Cauchy-Schwarz inequality, 28
- change of variables, 15
- characteristic function, 17
- Chebyshev's inequality, 27
- continous distribution, 9
- convex function, 27
- counting measure, 6

- Dirichlet function, 16
- distribution function, 8
- Dominated convergence theorem, 18

- Euler summation formula, 16
- expectation, 23
- exponential distribution, 24

- Heine-Borel Theorem, 2

- Hoeffding's inequality, 27, 29
- Holder's inequality, 30

- indication function, 17
- integrability, 20
- integration by parts, 15

- Jensen's inequality, 27

- Lebesgue integration, 16
- Lebesgue measure, 6
- Levy's inequality, 29
- lower sum, 11, 13

- Markov inequality, 26
- mass points, 9
- Minkowski's inequality, 30
- monotone convergence theorem, 21

- normal distribution, 24
- normal tail inequality, 28

- partition norm, 12
- Poisson distribution, 24
- probability distribution, 8, 22

probability measure, 22
purely discrete distribution, 9, 24

random variable, 23
regularity property, 10
Riemann integration, 11
Riemann-Steiltjes integration, 14

simple function, 17
support, 10

upper sum, 11, 13