1 Basics of Probability Theory

1.1 Infinite Probability Spaces

• Used to model a random experiment with infinitely many possible outcomes.

• Two examples to keep in mind:

i) Choosing a number from the interval $[0, 1]$

ii) Tossing a coin infinitely many times.

For i) the sample space is $[0, 1]$. For ii) the sample space is the set of all infinite sequences of heads and tails. One important issue with infinite sample spaces is that the classical definition fails. For example, what is the probability that you choose a number less than or equal to 1/2. In this case both the total number of outcomes as well as the number of favorable outcomes are infinite. This leads us to what is called the axiomatic definition of probability. Before that we need one more definition.

Definition 1.1. *Let Ω be a non-empty set and let $\mathcal{F}$ be a collection of subsets of Ω. We say that $\mathcal{F}$ is a $\sigma-$algebra provided:*

i) the empty set ϕ belongs to $\mathcal{F}$

ii) if $A \in \mathcal{F}$ then $A^c \in \mathcal{F}$ (closed under complementation)

iii) whenever a sequence of sets $A_1, A_2, \ldots$ belongs to $\mathcal{F}$ then $\cup_{n=1}^{\infty} A_n$ also belongs to $\mathcal{F}$ (closed under countable union).

Two trivial examples $\mathcal{F}_1 = \mathcal{P}(\Omega)$, $\mathcal{F}_2 = \{\phi, \Omega\}$.

Result: Arbitrary intersection of $\sigma-$algebras is again a $\sigma-$algebra.

Proof: Exercise.

Exercise: Is arbitrary union of $\sigma-$algebras again a $\sigma-$algebra ?

Definition 1.2. *Let $\mathcal{A}$ be a collection of subsets of Ω. The $\sigma-$algebra generated by $\mathcal{A}$ is given by*

$$\sigma(\mathcal{A}) \doteq \bigcap \mathcal{F} \,,$$

where the intersection is over all $\sigma-$algebras containing $\mathcal{A}$. Thus this is the smallest $\sigma-$algebra containing $\mathcal{A}$.

Example 1.3. *Let A be a non-empty subset of Ω. Then $\sigma(\{A\}) = \{\phi, A, A^c, \Omega\}$.*

Example 1.4. *Let Θ be the collection of open subsets of $\mathbb{R}$ then $\sigma(\Theta) = \mathcal{B}(\mathbb{R})$ is called Borel $\sigma-$algebra. The sets in this $\sigma-$algebra are called Borel sets.*

Exercise: Let A be a non-empty subset of Ω. Let B be another non-empty subset of Ω such that $A \cap B \neq \phi$ and $A \cup B \neq \Omega$. Find $\sigma(\{A, B\})$.

Exercise: Suppose $\Omega = [0, 1]$. Find the $\sigma-$algebra generated by the collection of singletons.

Definition 1.5 (Axiomatic Definition)**.** *Let Ω be a non-empty set and $\mathcal{F}$ be a $\sigma-$algebra of subsets of Ω. A probability measure $\mathbb{P}$ is a function that to every set $A \in \mathcal{F}$ assigns a number in $[0,1]$, called the probability of the event A and denoted by $\mathbb{P}(A)$. We require:*
i) $\mathbb{P}(\Omega) = 1$
ii) whenever $A_1, A_2, \ldots$ is a sequence of disjoint sets in $\mathcal{F}$, then $\mathbb{P}(\cup_{n=1}^{\infty} A_n) = \sum \mathbb{P}(A_n)$.

• The triple $(\Omega, \mathcal{F}, \mathbb{P})$ is called the probability space.
• The first requirement is just a normalizing property.
• The second requirement is called the countable additivity property.
The following are easy consequences of the definition.

- $\mathbb{P}(\phi) = 0$.
- Finite additivity holds.
- $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.
- If $A \subset B$ then $\mathbb{P}(A) \leq \mathbb{P}(B)$. This is called the monotonicity property.
- If $\{A_i\}$ is a sequence of events not necessarily disjoint then $\mathbb{P}(\cup_{i=1}^{\infty})A_i \leq \sum_{i=1}^{\infty} \mathbb{P}(A_i)$ (countable subadditivity).

Result: Let $A_1 \subset A_2 \subset A_3 \ldots$ be an increasing sequence of events. Then $\mathbb{P}(\cup_{n=1}^{\infty} A_n) = \lim_{n\to\infty} \mathbb{P}(A_n)$. This is called the continuity from below property.
Proof: Exercise.
Result: Let $A_1 \supset A_2 \supset A_3 \ldots$ be a decreasing sequence of events. Then $\mathbb{P}(\cap_{n=1}^{\infty} A_n) = \lim_{n\to\infty} \mathbb{P}(A_n)$. This is called the continuity from above property.
Proof: Exercise.

Definition 1.6. *Let Ω be a non-empty set and let $\mathcal{F}$ be a collection of subsets of Ω. We say that $\mathcal{F}$ is a algebra provided:*
i) the empty set ϕ belongs to $\mathcal{F}$
ii) if $A \in \mathcal{F}$ then $A^c \in \mathcal{F}$ (closed under complementation)
iii) If A_1, A_2 belongs to $\mathcal{F}$ then $A_1 \cup A_2$ also belongs to $\mathcal{F}$ (closed under finite union).

Example 1.7. *Let $\mathcal{L}$ be the collection of all finite disjoint unions of all intervals of the form $(-\infty, a], (a, b], (b, \infty], \phi, \Omega$. Then $\mathcal{L}$ is an algebra not a $\sigma-$algebra.*

Example 1.8. *Let Ω be an infinite set and $\mathcal{L}$ be the collection of all subsets of Ω which are finite or have finite complement. Then $\mathcal{L}$ is an algebra not a $\sigma-$algebra.*

Theorem 1.9. *(Caratheodary extension theorem)*
Let Ω be a non-empty set and $\mathcal{F}_0$ be an algebra on it. Let $\mathbb{P}_0$ be a probability measure on $\mathcal{F}_0$. Then there is a unique probability measure $\mathbb{P}$ on $\sigma(\mathcal{F}_0)$ such that

$$\mathbb{P}_0(A) = \mathbb{P}(A) \; \forall \; A \in \mathcal{F}_0.$$

A Probabilistic Model for Tossing a Coin Infinitely Many Times:

• $\Omega = \{\omega = \{\omega_n\} : \omega_n = HorT\}$.

• Let $\mathcal{F}_\infty$ be the $\sigma-$algebra generated by sets which can be described in terms of finitely many coin tosses. Let $A = \{\{\omega_n\} : \omega_1 = H\}, B = \{\{\omega_n\} : \omega_1 = H, \omega_2 = H\}$. Clearly $A, B \in \mathcal{F}_\infty$. Let $C = \{\{\omega_n\} : \lim_{n\to\infty} \frac{H_n(\omega)}{n} = 1/2\}$, where $H_n(\omega)$ is the number of heads in first n coin tosses. Clearly C is not determined by finitely many coin tosses. For fixed n and m, define

$$C_{n,m} = \{\omega : |\frac{H_n(\omega)}{n} - 1/2| \le 1/m\} .$$

Clearly $C_{n,m} \in \mathcal{F}_\infty$. From the definition of limit

$$C = \cap_{m=1}^{\infty} \cup_{N=1}^{\infty} \cap_{n=N}^{\infty} C_{n,m} .$$

Thus $C \in \mathcal{F}_\infty$.

1.2 Random Variable

Let $\mathcal{B}(\mathbb{R})$ denote the Borel $\sigma-$algebra of $\mathbb{R}$, the $\sigma-$algebra generated by closed intervals.

Definition 1.10. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A function $X : \Omega \to \mathbb{R}$ is said to be a real-valued random variable if $X^{-1}(A) = \{\omega : X(\omega) \in A\} \in \mathcal{F}$ for all $A \in \mathcal{B}(\mathbb{R})$.*

Proposition 1.11. *Let $\mathcal{G}$ be a $\sigma-$algebra generated by a collection $\mathcal{S}$. If $X^{-1}(A) \in \mathcal{F}$ for all $A \in \mathcal{S}$, then $X^{-1}(A) \in \mathcal{F}$ for all $A \in \mathcal{G}$.*

Proof: Let

$$\mathcal{A} = \{A : X^{-1}(A) \in \mathcal{F}\} .$$

Then $\mathcal{S} \subset \mathcal{A}$. We will now show that $\mathcal{A}$ is a $\sigma-$algebra. $X^{-1}(A^c) = (X^{-1}(A))^c$ Thus if $A \in \mathcal{A}$, then $A^c \in \mathcal{A}$. Again, $X^{-1}(\cup A_i) = \cup X^{-1}(A_i)$. Thus if $A_i \in \mathcal{A}$ then $\cup A_i \in \mathcal{A}$. Hence we have shown that $\mathcal{A}$ is a $\sigma-$algebra, as a result $\sigma(\mathcal{S}) = \mathcal{G} \subset \mathcal{A}$. □

Fact: $\sigma-$algebra generated by open rays is $\mathcal{B}(\mathbb{R})$. Thus if $\mathcal{S} = \{(r, \infty) : r \in \mathbb{R}\}$, then $\sigma(\mathcal{S}) = \mathcal{B}(\mathbb{R})$.
Thus using the above proposition X is a random variable iff $X^{-1}(r, \infty) \in \mathcal{F}$ for all $r \in \mathbb{R}$.

Example 1.12. *Recall the independent, infinite coin-toss probability space $(\Omega, \mathcal{F}_\infty, \mathbb{P})$. Let $S : \Omega \to \mathbb{R}$ be defined by*

$$S(w) = \begin{cases} 8 & \text{if } \mathrm{w}_1 = \mathrm{H} \\ 2 & \text{if } \mathrm{w}_1 = \mathrm{T}. \end{cases}$$

Then

$$S^{-1}(A) = \begin{cases} \phi & \text{if } 8, 2 \notin \mathrm{A} \\ \Omega & \text{if } 8, 2 \in \mathrm{A} \\ A_H & \text{if } 8 \in \mathrm{A} \ \& \ 2 \notin \mathrm{A} \\ A_T & \text{if } 8 \notin \mathrm{A} \ \& \ 2 \in \mathrm{A}. \end{cases}$$

Also note that

$$S^{-1}(a, \infty) = \begin{cases} \phi & \text{if } \mathrm{a} > 8 \\ A_H & \text{if } 2 < \mathrm{a} \le 8 \\ \Omega & \text{if } \mathrm{a} \le 2. \end{cases}$$

Therefore S is a random variable.

Lemma 1.13. *If X and Y are random variables then $X + Y$ is also a random variable.*

Proof:

$$X + Y > r$$
$$\Leftrightarrow X > r - Y$$
$$\Leftrightarrow \text{there exists a rational no. q such that } X > q > r - Y\,.$$

Thus $(X + Y)^{-1}(r, \infty) = \cup_{q \in \mathbb{Q}} X^{-1}(q, \infty) \cap Y^{-1}(r - q, \infty)$. $\square$

Exercise: Let A be a subset of Ω. Then show that 1_A is a random variable if and only if $A \in \mathcal{F}$.

Exercise: Prove that if X is a random variable then X^2 is also a random variable. Prove that if X and Y are random variables then XY is also a random variable.

Exercise Let $f : \mathbb{R} \to \mathbb{R}$ be a function such that $f^{-1}(B) \in \mathcal{B}(\mathbb{R})$ for all $B \in \mathcal{B}(\mathbb{R})$ (such a function is called Borel measurable function). Show that if X is a random variable then $f(X)$ is also a random variable.

Definition 1.14. *Let X be a random variable on $(\Omega, \mathcal{F}, \mathbb{P})$. The distribution measure of X is the probability measure μ_X on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ that assigns to each Borel set B of $\mathbb{R}$ the measure $\mu_X(B) = \mathbb{P}(X \in B)$.*

Two examples:

Example 1.15. *Let $\Omega = [0, 1]$, $\mathcal{F} = \sigma-$algebra generated by closed intervals. Let $\mathbb{P}$ be the probability measure which assigns to each interval its length. Define $X(\omega) = \omega$ and $Y(\omega) = 1 - \omega$. Then*

$$\mu_X[a, b] = \mathbb{P}(X \in [a, b]) = \mathbb{P}[a, b] = b - a\,.$$

Now

$$\mu_Y[a, b] = \mathbb{P}(Y \in [a, b]) = \mathbb{P}(a \le 1 - \omega \le b) = \mathbb{P}[1 - b, 1 - a] = b - a\,.$$

Thus $\mu_X = \mu_Y$ (X and Y are said to be uniformly distributed.)

Definition 1.16. *The distribution function of a random variable X defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is the function $F_X : \mathbb{R} \to [0, 1]$ given by*

$$F_X(x) = \mathbb{P}(X \le x)\,.$$

Proposition 1.17. *The distribution function of a random variable has the following properties:*

(1) $F_X(\cdot)$ *is non-decreasing and hence has only jump discontinuities.*

(2) $\lim_{x \uparrow \infty} F_X(x) = 1, \lim_{x \downarrow -\infty} F_X(x) = 0.$

(3) $\lim_{h \downarrow 0} F_X(x + h) = F_X(x), \forall x \in \mathbb{R}$*, thus CDF is right continuous.*

(4) $\lim_{h \downarrow 0} F_X(x - h) = F_X(x) - P(X = x), \forall x \in \mathbb{R}.$

Theorem 1.18. *Let F be a function from $\mathbb{R}$ to $[0, 1]$ satisfying the properties of the above proposition, then there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a random variable X defined on it whose distribution function is F.*

Two Special Cases

• There exists a non-negative function f on $\mathbb{R}$ such that

$$\mu_X[a,b] = \mathbb{P}(a \leq X \leq b) = \int_a^b f(x)dx\,.$$

Thus

$$1 = \mathbb{P}(X \in \mathbb{R}) = \lim_{n\to\infty} \mathbb{P}(-n \leq X \leq n) = \lim_{n\to\infty} \int_{-n}^{n} f(x)dx = \int_{-\infty}^{\infty} f(x)dx\,.$$

• X takes only countably many values x_i. Define $p_i = \mathbb{P}(X = x_i)$. Then

$$\mu_X(B) = \sum_{\{i: x_i \in B\}} p_i\,.$$

In the first case X is said to have an absolutely continuous distribution with probability density function f and in the second case X is said to have a discrete distribution with probability mass function $\{p_i\}$.

Example: Consider the functions:

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}\,,\, N(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{\frac{-y^2}{2}} dy\,.$$

Let X be uniformly distributed on $[0,1]$. Notice that N is a strictly increasing function. So it has an inverse N^{-1}. Define the random variable $Z = N^{-1}(X)$. Then

$$\begin{aligned}
\mu_Z[a,b] &= \mathbb{P}(\omega \in \Omega : a \leq Z(\omega) \leq b) \\
&= \mathbb{P}(\omega \in \Omega : a \leq N^{-1}(X)(\omega) \leq b) \\
&= \mathbb{P}(\omega \in \Omega : N(a) \leq NN^{-1}(X)(\omega) \leq N(b)) \\
&= \mathbb{P}(\omega \in \Omega : N(a) \leq X(\omega) \leq N(b)) \\
&= N(b) - N(a) = \int_a^b \varphi(x)dx\,.
\end{aligned}$$

The measure μ_X on $\mathbb{R}$ given by this formula is called the standard normal distribution. Any random variable that has this distribution, regardless of the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ on which it is defined, is called a standard normal random variable.

1.3 Expectation

Let X be a random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We want to compute an "average value" of X, where we take the probabilities into account while computing the average.

If Ω is countable then we can simply define

$$\text{average value" of X} := \mathbb{E}(X) := \sum_{k=0}^{\infty} X(w_k)\mathbb{P}(w_k)\,,$$

where $\Omega = \{w_1, w_2, \ldots\}$

But if Ω is uncountable then we must think in terms of integrals.

2 Riemann integration

Partition: Let $[a,b]$ be a closed and bounded interval. A partition of $[a,b]$ is a finite sequence $P = (x_0, x_1, \cdots, x_n)$ of points of $[a,b]$ such that $a = x_0 < x_1 < \cdots < x_n = b$. The family of all partitions of $[a,b]$ is denoted by $\mathcal{P}[a,b]$ and the partition $P = (x_0, x_1, \cdots, x_n)$ is a member of $\mathcal{P}[a,b]$.
For example, $P = (0, 1/4, 1/3, 1/2, 2/3, 3/4, 1)$ is a partition of $[0,1]$, $Q = (0, 1/4, 3/8, 1/2, 3/4, 7/8, 1)$ is another partition of of $[0,1]$.
Riemann sums:- Let $f : [a,b] \to \mathbb{R}$ be a bounded function on $[a,b]$. Let $P \in \mathcal{P}[a,b]$ (i.e, $P = (x_0, x_1, \cdots, x_n)$, where $a = x_0 < x_1 < \cdots < x_n = b$). Since f is bounded on $[a,b]$, f is bounded on $[x_{r-1}, x_r]$, for $r = 1, 2, \cdots, n$. Let $M = \sup_{x\in[a,b]} f(x)$, $m = \inf_{x\in[a,b]} f(x)$; $M_r = \sup_{x\in[x_{r-1},x_r]} f(x)$, $m_r = \inf_{x\in[x_{r-1},x_r]} f(x)$; for $r = 1, 2, \cdots, n$. Then $m \leq m_r \leq M_r \leq M$, for $r = 1, 2, \cdots, n$. The sum $U(P,f) := \sum_{i=1}^{n} M_r(x_r - x_{r-1})$ is said to be the upper Riemann sum and the sum $L(P,f) := \sum_{i=1}^{n} m_r(x_r - x_{r-1})$ is said to be lower Riemann sum.

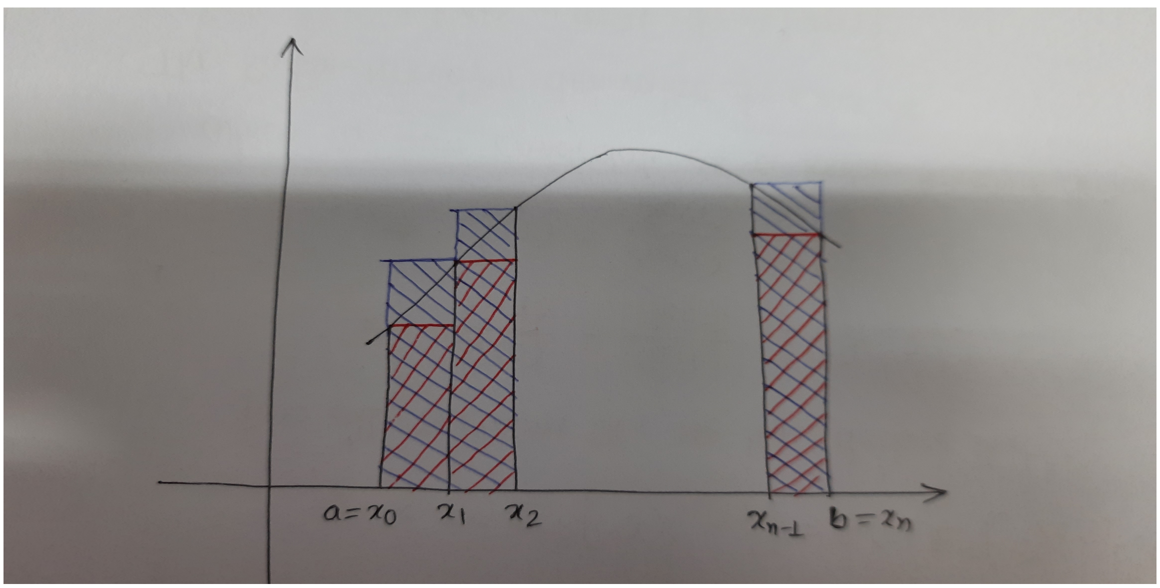

Figure 1:

Here $U(P,f)$ is the blue shaded area (region) and $L(P,f)$ is the red shaded area (region) of Figure 1. Note that $m(x_r - x_{r-1}) \leq m_r(x_r - x_{r-1}) \leq M_r(x_r - x_{r-1}) \leq M(x_r - x_{r-1})$, for $r = 1, 2, \cdots, n$. Therefore,

$$m\sum_{r=1}^{n}(x_r - x_{r-1}) \leq \sum_{r=1}^{n} m_r(x_r - x_{r-1}) \leq \sum_{r=1}^{n} M_r(x_r - x_{r-1}) \leq M\sum_{r=1}^{n}(x_r - x_{r-1}),$$

or, $m(b-a) \leq L(P,f) \leq U(P,f) \leq M(b-a)$. We have two sets of real numbers $\{U(P,f) : P \in \mathcal{P}[a,b]\}$ and $\{L(P,f) : P \in \mathcal{P}[a,b]\}$ both sets are bounded. The supremum of the set $\{L(P,f) : P \in P \in \mathcal{P}[a,b]\}$ exists and it is called the lower integral of f on $[a,b]$ and is denoted by $\underline{\int_a^b} f(x)dx$. The infimum of the set $\{U(P,f); P \in P \in \mathcal{P}[a,b]\}$ exists and it is called the upper integral of f on $[a,b]$ and is denoted by $\overline{\int_a^b} f(x)dx$. f is said to be Riemann integral on

$[a, b]$ if

$$\underline{\int_a^b} f(x)dx = \overline{\int_a^b} f(x)dx.$$

The common value is called the Riemann integral of f on $[a, b]$ and it is denoted by $\int_a^b f(x)dx$.

Exercise:-

(1) Let $f(x) = c$, $x \in [a, b]$. Prove that f is Riemann integral on $[a, b]$.

(2) A function f is defined on $[0, 1]$ by

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

Show that f is not Riemann integral on $[0, 1]$.

(3) Prove that the function f is defined on $[a, b]$ by $f(x) = x$, $x \in [a, b]$ is Riemann integral on $[a, b]$. Evaluate $\int_a^b f(x)dx$.

(4) $f(x) = x^2$.

(5) $f(x) = e^x$.

Refinement of a partition:- Let $P = (x_0, x_1, x_2, \cdots, x_n)$ be a partition of $[a, b]$. A partition Q of $[a, b]$ is said to be a refinement of P if P is a proper subset of Q. That is Q is obtained by adjoining a finite number of additional points to P.

For example, let $P = (0, 1/4, 1/2, 3/4, 1)$ be a parttion of $[0, 1]$ and $Q = (0, 1/8, 1/4, 1/2, 3/4, 7/8, 1)$, then Q is a refinement of P. If $R = (0, 1/8, 1/4, 3/8, 1/2, 3/4, 1)$, then R is a refinement of P but not a refinement of Q.

Lemma 2.1. *Let $f : [a, b] \to \mathbb{R}$ be a bounded function and P be a partition of $[a, b]$. If Q is a refinement of P, then $U(P, f) \geq U(Q, f)$ and $L(P, f) \leq L(Q, f)$.*

Norm of partition:- Let $P = (x_0, x_1, \cdots, x_n)$ be a partition of $[a, b]$. Then norm of a partition denoted by $\|P\|$, is defined by

$$\|P\| = \max_{r \in \{1,2,\cdots,n\}} |x_r - x_{r-1}|.$$

If Q is a refinement of P, then $\|Q\| \leq \|P\|$.

Lemma 2.2. *Let $f : [a, b] \to \mathbb{R}$ be bounded. If $\{P_n\}$ is a sequence of partition of $[a, b]$ such that $\|P_n\| \to 0$, then*

(i) $\lim_{n\to\infty} U(P_n, f) = \overline{\int_a^b} f$

(ii) $\lim_{n\to\infty} L(P_n, f) = \underline{\int_a^b} f.$

Condition for integrability:- Let $f : [a, b] \to \mathbb{R}$ be a bounded function. Then f is integrable on $[a, b]$ if and only if for each $\varepsilon > 0$, there exists a partition of P of $[a, b]$ such that

$$U(P, f) - L(P, f) < \varepsilon.$$

Let $f : [a, b] \to \mathbb{R}$ be a bounded function. Then f is integrable on $[a, b]$ iff for each $\varepsilon > 0$ there exists a positive δ such that

$$U(P, f) - L(P, f) < \varepsilon$$

for every partition P of $[a, b]$ satifying $\|P\| \leq \delta$.

Properties:-

(1) Let $f : [a, b] \to \mathbb{R}$, $g : [a, b] \to \mathbb{R}$ be both Riemann integrable on $[a, b]$. Then $f + g$ is Riemann integrable on $[a, b]$ and $\int_a^b (f + g) = \int_a^b f + \int_a^b g$.

(2) Let $f : [a, b] \to \mathbb{R}$ be Riemann integrable on $[a, b]$ and $c \in \mathbb{R}$. Then cf is integrable on $[a, b]$ and $\int_a^b cf = c \int_a^b f$.

(3) $|f|$, f^2, $f \cdot g$ are Riemann integrable. If $g \geq k > 0$ then $1/g$ is also Riemann integrable.

Ex. A function f is defined by $f(x) = x^2$, $x \in [a, b]$, where $a > 0$. Find $\overline{\int_a^b} f$ and $\underline{\int_a^b} f$. Deduce that f is integrable on $[a, b]$.

Ans:- f is bounded on $[a, b]$. Let $P_n = (a, a + h, a + 2h, \cdots, a + nh)$ where $h = \frac{b-a}{n}$. Then P_n is partition of $[a, b]$ with $\|P_n\| = \frac{b-a}{n}$. Since f is increasing function on $[a, b]$,
$M_r = (a + rh)^2$, $m_r = [a + (r - 1)h]^2$ for $r = 1, 2, \cdots, n$.

$$\begin{aligned}
U(P_n, f) &= h\Big[(a + h)^2 + (a + 2h)^2 + \cdots + (a + nh)^2\Big] \\
&= h\Big[(a^2 + a^2 + \cdots + a^2) + 2ah(1 + 2 + 3 + \cdots + n) + h^2(1^2 + 2^2 + 3^3 + \cdots + n^2)\Big] \\
&= h\Big[na^2 + 2ah\frac{n(n + 1)}{2} + h^2\frac{n(n + 1)(2n + 1)}{6}\Big] \\
&= nha^2 + anh(nh + h) + \frac{nh(nh + h)(2nh + h)}{6} \\
&= (b - a)a^2 + a(b - a)^2(1 + \frac{1}{n}) + \frac{1}{6}(b - a)^3(1 + \frac{1}{n})(2 + \frac{1}{n})
\end{aligned}$$

and

$$\begin{aligned}
L(P_n, f) &= h\Big[a^2 + (a + h)^2 + (a + 2h)^2 + \cdots + (a + (n - 1)h)^2\Big] \\
&= h\Big[na^2 + 2ah\frac{n(n - 1)}{2} + h^2\frac{n(n - 1)(2n - 1)}{6}\Big] \\
&= nha^2 + anh(nh - h) + \frac{nh(nh - h)(2nh - h)}{6} \\
&= (b - a)a^2 + a(b - a)^2(1 - \frac{1}{n}) + \frac{1}{6}(b - a)^3(1 - \frac{1}{n})(2 - \frac{1}{n}).
\end{aligned}$$

Consider the sequence of partitions $\{P_n\}$ of $[a, b]$ with $\lim_{n\to\infty} \|P_n\| = \lim_{n\to\infty} \dfrac{b - a}{n} = 0$. Then $\overline{\int_a^b} f(x)dx = \lim_{n\to\infty} U(P_n, f) = (b - a)a^2 + a(b - a)^2 + \dfrac{(b - a)^3}{3} = \dfrac{b^3 - a^3}{3}$ and

$$\begin{aligned}
\underline{\int_a^b} f(x)dx &= \lim_{n\to\infty} L(P_n, f) \\
&= (b - a)a^2 + a(b - a)^2 + \frac{(b - a)^3}{3} = \frac{b^3 - a^3}{3}.
\end{aligned}$$

As $\overline{\int_a^b} f(x)dx = \underline{\int_a^b} f(x)dx$, f is integrable on $[a, b]$ and $\int_a^b f(x)dx = \frac{b^3 - a^3}{3}$.

Ex. A function f is defined on $[0, 1]$ by

$$f(x) = \begin{cases} x \text{ if } x \in [0, 1] \cap \mathbb{Q} \\ 0 \text{ if } x \in [0, 1] \backslash \mathbb{Q}.. \end{cases}$$

Find $\underline{\int_0^1} f(x)dx$ and $\overline{\int_0^1} f(x)dx$. Deduce that f is not integrable on $[0,1]$.
Ans:- f is bounded on $[0,1]$. Let us take the partition P_n of $[0,1]$ defined by $P_n = (0, 1/n, 2/n, \cdots, n/n)$. Let $M_r = \sup_{x\in[\frac{r-1}{n},\frac{r}{n}]} f(x)$, $m_r = \inf_{x\in[\frac{r-1}{n},\frac{r}{n}]} f(x)$, for $r = 1, 2, \cdots, n$. Then $M_r = r/n$ and $m_r = 0$ for $r = 1, 2, \cdots n$.

$$\begin{aligned} U(P_n, f) &= M_1(\frac{1}{n} - 0) + M_2(\frac{2}{n} - \frac{1}{n}) + \cdots M_n(\frac{n}{n} - \frac{n-1}{n}) \\ &= \frac{1}{n}[\frac{1}{n} + \frac{2}{n} + \cdots \frac{n}{n}] \\ &= \frac{n(n+1)}{2n^2} = \frac{n+1}{2n} \end{aligned}$$

and

$$L(P_n, f) = m_1(\frac{1}{n} - 0) + m_2(\frac{2}{n} - \frac{1}{n}) + \cdots m_n(\frac{n}{n} - \frac{n-1}{n}) = 0.$$

Let us consider the sequence of partitions $\{P_n\}$ of $[0,1]$ with $\|P_n\| = \frac{1}{n}$ and $\lim_{n\to\infty} \|P_n\| = 0$. Then $\lim_{n\to\infty} U(P_n, f) = \overline{\int_0^1} f(x) = \lim_{n\to\infty} \frac{n+1}{2n} = \frac{1}{2}$ and $\lim_{n\to\infty} L(P_n, f) = \underline{\int_0^1} f(x)dx = 0$. Since $\overline{\int_0^1} f(x)dx \neq \underline{\int_0^1} f(x)dx$, f is not Reimann integrable on $[0,1]$.

3 Lebesgue Integral

Definition:- A random variable $s : \Omega \to [0, \infty)$ is defined by $s(\omega) = \sum_{i=1}^{n} a_i \mathcal{X}_{A_i}(\omega)$, $\omega \in \Omega$, where n is some positive integer, $a_1, a_2, \cdots, a_n$ are non-negative real-numbers, $A_i \in \mathcal{F}$ for every i; $A_i \cap A_j = \emptyset$ for $i \neq j$ and $\cup_{i=1}^{n} A_i = \Omega$. Such a function s is called a non-negative simple random variable. We say that $\sum_{i=1}^{n} a_i \mathcal{X}_{A_i}(\omega)$ is the standard representation of s if $a_1, a_2, \cdots, a_n$ are all distinct. We denote by $\mathbb{L}_0^+$ the class of all non-negative simple random variables on $(\Omega, \mathcal{F})$.
Examples:

- If $s(\omega) \equiv c$ for some $c \in [0, \infty)$, then $s \in \mathbb{L}_0^+$.
- For $A \subset \Omega$, consider $\mathcal{X}_A : \Omega \to [0, \infty)$, the indicator function of the set A, i.e.,

$$\mathcal{X}(\omega) = \begin{cases} 0 \text{ if } \omega \notin A \\ 1 \text{ if } \omega \in A. \end{cases}$$

  Then $\mathcal{X}_A \in \mathbb{L}_0^+$ iff $A \in \mathcal{F}$.
- Let $A, B \in \mathcal{F}$. then $s = \mathcal{X}_A \mathcal{X}_B \in \mathbb{L}_0^+$ since $s = \mathcal{X}_{A\cap B}$.
- Let $A, B \in \mathcal{F}$. If $A \cap B = \emptyset$, then clearly, $\mathcal{X}_A + \mathcal{X}_B = \mathcal{X}_{A\cup B} \in \mathbb{L}_0^+$.

Definition:- For $s \in \mathbb{L}_0^+$ with a representation $s = \sum_{i=1}^{n} a_i \mathcal{X}_{A_i}$, we define $\int_\Omega s(\omega)d\mathbb{P}(\omega)$, the integral of s with respect to $\mathbb{P}$, by $\int_\Omega s(\omega)d\mathbb{P}(\omega) := \sum_{i=1}^{n} a_i \mathbb{P}(A_i)$.

We should check that $\displaystyle\int_\Omega s(\omega)d\mathbb{P}(\omega)$ is well defined i.e., if $s = \sum_{i=1}^n a_i \mathcal{X}_{A_i} = \sum_{j=1}^m b_i \mathcal{X}_{B_j}$ where $\{A_1, A_2, \cdots, A_n\}$ and $\{B_1, B_2, \cdots, B_m\}$ are partitions of Ω by elements of $\mathcal{F}$, then

$$\sum_{i=1}^n a_i \mathbb{P}(A_i) = \sum_{b=1}^m b_j \mathbb{P}(B_j).$$

For this, we note that we can write

$$s = \sum_{i=1}^n a_i \sum_{j=1}^m \mathcal{X}_{A_i \cap B_j} = \sum_{j=1}^m b_j \sum_{i=1}^n \mathcal{X}_{A_i \cap B_j}.$$

Thus if $A_i \cap B_j \neq \emptyset$ then $a_i = b_j$. Hence using finite additivity of $\mathbb{P}$,

$$\begin{aligned}\sum_{i=1}^n a_i \mathbb{P}(A_i) &= \sum_{i=1}^n a_i \sum_{j=1}^m \mathbb{P}(A_i \cap B_j) \\ &= \sum_{j=1}^m b_j \sum_{i=1}^n \mathbb{P}(A_i \cap B_j) = \sum_{j=1}^m b_j \mathbb{P}(B_j).\end{aligned}$$

Thus $\int_\Omega s(\omega)d\mathbb{P}(\omega)$ is independent of the representation of $s = \sum_{i=1}^n a_i \mathcal{X}_{A_i}$.

Proposition 3.1. *For $s, s_1, s_2 \in \mathbb{L}_0^+$ and $\alpha \in \mathbb{R}$ with $\alpha \geq 0$, the following hold:*

(i) $0 \leq \int_\Omega s \, d\mathbb{P} < +\infty$

(ii) $\alpha s \in \mathbb{L}_0^+$ *and* $\int_\Omega (\alpha s) \, d\mathbb{P} = \alpha \int_\Omega s \, d\mathbb{P}$

(iii) $s_1 + s_2 \in \mathbb{L}_0^+$ *and* $\int_\Omega (s_1 + s_2) \, d\mathbb{P} = \int_\Omega s_1 \, d\mathbb{P} + \int_\Omega s_2 \, d\mathbb{P}$.

Proof. Statements (i) and (ii) are obvious.

For (iii), let $s_1 = \displaystyle\sum_{i=1}^n a_i \mathcal{X}_{A_i}$ and $s_2 = \sum_{j=1}^m b_j \mathcal{X}_{B_j}$. Then we can write $s_1 = \displaystyle\sum_{i=1}^n \sum_{j=1}^m a_i \mathcal{X}_{A_i \cap B_j}$ and $s_2 = \displaystyle\sum_{i=1}^n \sum_{j=1}^m b_j \mathcal{X}_{A_i \cap B_j}$. Thus $s_1 + s_2 = \displaystyle\sum_{i=1}^n \sum_{j=1}^m (a_i + b_j) \mathcal{X}_{A_i \cap B_j}$. Hence $s_1 + s_2 \in \mathbb{L}_0^+$ and using these representations, it is clear that $\int_\Omega (s_1 + s_2) d\mathbb{P} = \int_\Omega s_1 d\mathbb{P} + \int_\Omega s_2 d\mathbb{P}$. $\square$

Excercise:- Let $s_1, s_2 \in \mathbb{L}_0^+$. Then prove the followings

1. Let $s_1 \geq s_2$. Set $\phi = s_1 - s_2$. Show that $\phi \in \mathbb{L}_0^+$.
2. If $s_1 \geq s_2$, then $\int s_1 d\mathbb{P} \geq \int s_2 d\mathbb{P}$.

Proposition 3.2. *Let $X : \Omega \to \mathbb{R}$ a non-negative bounded random variable, then there exists a sequence $\{s_n\}_{n\geq 1}$ of random variables in $\mathbb{L}_0^+$ such that $\mathbb{P}\{\omega \in \Omega : \lim_{n\to\infty} s_n(\omega) = X(\omega)\} = 1$.*

Proof. Let X be bounded by M. Then the sets $A_k^n = \{\omega : \dfrac{(k-1)M}{2^n} \leq X(\omega) < \dfrac{kM}{2^n}\}$, $1 \leq k \leq 2^n$. Then $\{A_k^n\}$ are disjoint, $A_k^n \in \mathcal{F}$ and have union $\cup_{k=1}^{2^n} A_k^n = \Omega$. We define function s_n on Ω by

$$s_n(\omega) = \sum_{k=1}^{2^n} \frac{M(k-1)}{2^n} \mathcal{X}_{A_k^n}(\omega).$$

Clearly, $s_n \in \mathbb{L}_0^+$ and it is easy to check that for every n,

$$s_n(\omega) \leq s_{n+1}(\omega), \ \forall \omega \in \Omega.$$

If $\omega \in A_k^n$ for some k, $1 \leq k \leq 2^n$. Then

$$s_n(x) = \frac{(k-1)M}{2^n}$$

and $X(\omega) \in \left[\frac{(k-1)M}{2^n}, \frac{kM}{2^n}\right)$. Thus we have $s_n(\omega) \leq X(\omega)$ and $X(\omega) - s_n(\omega) \leq \frac{M}{2^n}$. In other words, $\lim_{n\to\infty} s_n(\omega) = X(\omega)$. □

Consider the case n=1

Then $A_1^1 = [a, a_1) \cup (a_2, b]$

$A_2^1 = [a_1, a_2]$

$s_1 = 0.\mathfrak{X}_{A_1^1} + \frac{M}{2}\mathfrak{X}_{A_2^1}$.

Consider the case n=2

$A_1^2 = [a, a_2^{'}) \cup (a_2^{''}, b]$, $A_2^2 = [a_2^{'}, a_1) \cup (a_2, a_2^{''}]$, $A_3^2 = [a_1, c_2) \cup (c_2^{'}, a_2]$, $A_4^2 = [c_2, c_2^{'}]$.

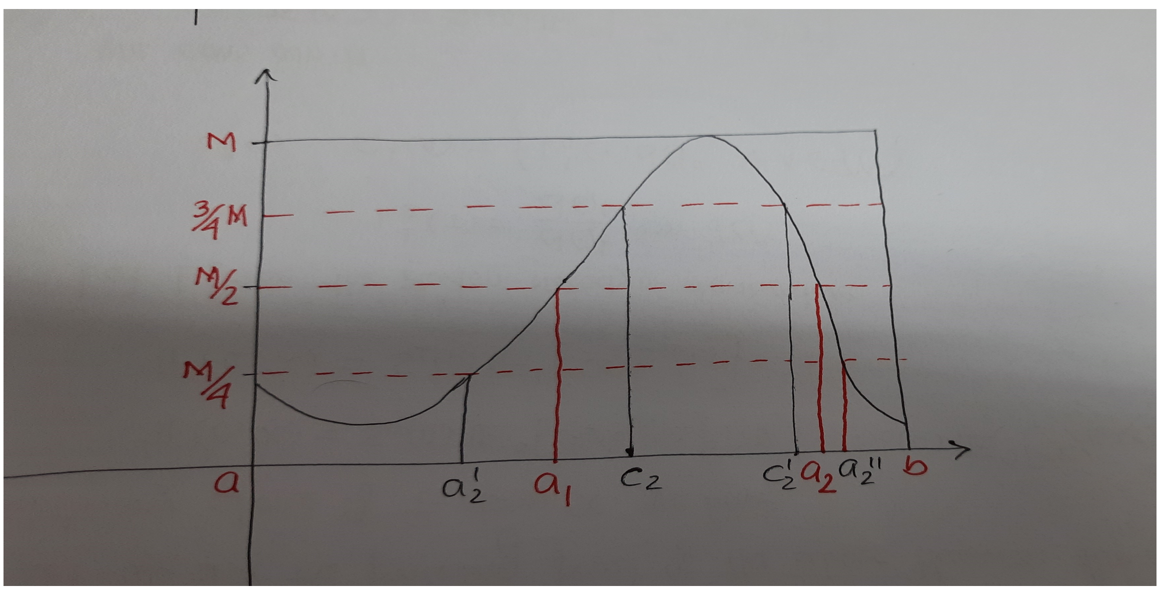

Figure 2:

$s_2 = 0.\mathfrak{X}_{A_1^2} + \frac{M}{4}\mathfrak{X}_{A_2^2} + \frac{M}{2}\mathfrak{X}_{A_3^2} + \frac{3M}{4}\mathfrak{X}_{A_4^2}$.

$$\begin{aligned}\int_\Omega s_1 \, d\mathbb{P}(\omega) &= 0.\mathbb{P}(A_1^{'}) + \frac{M}{2}\mathbb{P}(A_2^{'}) \\ &= \frac{M}{2}[a_2 - a_1].\end{aligned}$$

$$\int_\Omega s_2 \, d\mathbb{P}(\omega) = \frac{M}{4}[(a_1 - a_2^{'}) + (a_2^{''} - a_2)] + \frac{M}{2}[(c_2 - a_1) + (a_2 - c_2^{'})] + \frac{3M}{4}(c_2^{'} - c_2).$$

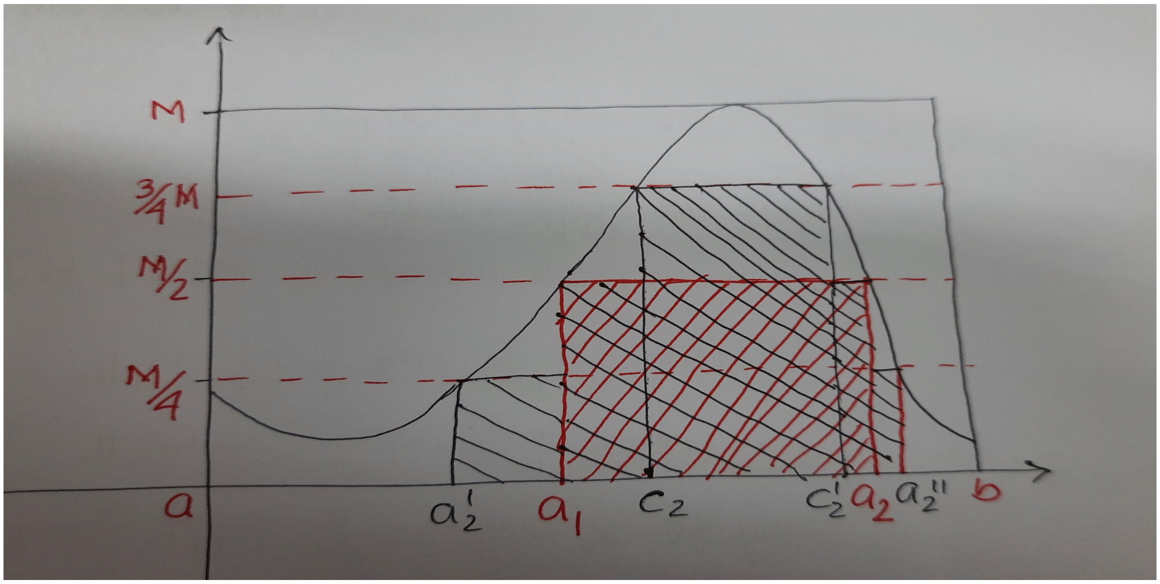

Figure 3:

Set $\mathbb{L}^+ = \{f : \Omega \to [0, \infty) : \exists$ an increasing sequence of random variables $\{s_n\}_{n\geq 1}$ in $\mathbb{L}_0^+$ such that $s_n(\omega)$ converges to $s(\omega)$ almost surely$\}$. For $f \in \mathbb{L}^+$, we define the integral of f w.r.t. $\mathbb{P}$ by

$$\int_\Omega f(\omega) d\mathbb{P}(\omega) := \lim_{n\to\infty} \int_\Omega s_n(\omega) d\mathbb{P}(\omega).$$

Proposition 3.3. *If $f \in \mathbb{L}^+$ and $s \in \mathbb{L}_0^+$ are such that $0 \leq s \leq f$, then $\int_\Omega s(\omega)d\mathbb{P}(\omega) \leq \int_\Omega f(\omega)d\mathbb{P}(\omega)$ and $\int_\Omega f(\omega)d\mathbb{P}(\omega) = \sup\{\int_\Omega s d\mathbb{P} | 0 \leq s \leq f, s \in \mathbb{L}_0^+\}$.*

Now for any random variable X, define

$$X^+(\omega) = \max\{X(\omega), 0\},\ X^- = \max\{-X(\omega), 0\}.$$

Then

$$X = X^+ - X^-.$$

We can define $\int_\Omega X^+ d\mathbb{P}(\omega)$ and $\int_\Omega X^- d\mathbb{P}(\omega)$ provided both of them are not infinite. Then we define

$$\int_\Omega X(\omega)d\mathbb{P}(\omega) = \int_\Omega X^+(\omega)d\mathbb{P}(\omega) - \int_\Omega X^-(\omega)d\mathbb{P}(\omega).$$

We say that X is integrable if both $\int_\Omega X^+(\omega)d\mathbb{P}(\omega)$ and $\int_\Omega X^-(\omega)d\mathbb{P}(\omega)$ are finite. If both are infinite, then $\int_\Omega X(\omega)d\mathbb{P}(\omega)$ is not defined. If $\int_\Omega X^+(\omega)d\mathbb{P}(\omega)$ is finite and $\int_\Omega X^-(\omega)d\mathbb{P}(\omega) = \infty$ then $\int_\Omega X(\omega)d\mathbb{P}(\omega) = -\infty$. If $\int_\Omega X^-(\omega)d\mathbb{P}(\omega)$ is finite and $\int_\Omega X^+(\omega)d\mathbb{P}(\omega) = \infty$ then $\int_\Omega X(\omega)d\mathbb{P}(\omega) = \infty$.

4 Comparison of Riemann and Lebesgue integrals:-

Let f be a bounded function defined on $\mathbb{R}$, and let $a < b$ be numbers.

1. The Riemann integral $\int_a^b f(x)dx$ is defined iff the set of points $x \in [a, b]$ where $f(x)$ is not continuous has Lebesgue measure zero.

2. If the Riemann integral $\int_a^b f(x)dx$ is defined, then f is Borel measurable and so the Lebesgue integral $\int_a^b f(x)dx$ is also defined and the Riemann and Lebesgue integrals agree.

Definition:- Let X be an integrable random variable. Then the expectation of X is defined by $\mathbb{E}[X] := \int_\Omega X(\omega)d\mathbb{P}(\omega)$.

If $X \geq 0$, then $\mathbb{E}[X]$ is always defined [can be $+\infty$ as well].

Examples:-

1. Consider the infinite independent coin-toss space $(\Omega_\infty, \mathcal{F}_\infty, \mathbb{P}_\infty)$ with $p = \frac{1}{2}$. Let

$$Y_n(\omega) = \begin{cases} 1 \text{ if } \omega_n = H \\ 0 \text{ if } \omega_n = T. \end{cases}$$

$$\begin{aligned} \mathbb{E}[Y_n] &= 1 \cdot \mathbb{P}(Y_n = 1) + 0 \cdot \mathbb{P}(Y_n = 0) \\ &= 1 \cdot \frac{1}{2} + 0 \cdot \frac{1}{2} = \frac{1}{2}. \end{aligned}$$

2. Let $\Omega = [0,1], \mathcal{F} = \mathcal{B}[0,1]$ and let $\mathbb{P}$ be the Lebesgue measure on $[0,1]$. Consider the random variable

$$X(\omega) = \begin{cases} 1 \text{ if } \omega \text{ is irrational} \\ 0 \text{ if } \omega \text{ is rational.} \end{cases}$$

$$\begin{aligned} \mathbb{E}[X] =& 1 \cdot \mathbb{P}(X = 1) + 0 \cdot \mathbb{P}(X = 0) \\ &= 1 \cdot \mathbb{P}(\omega \in [0,1] \backslash \mathbb{Q}) + 0 \cdot \mathbb{P}(\omega \in [0,1] \cap \mathbb{Q}) \\ &= 1 \cdot 1 + 0 \cdot 0 = 1. \end{aligned}$$

Properties:-

1. If X takes only finitely many values $x_0, x_1, \cdots, x_n$, then

$$\mathbb{E}[X] = \int_\Omega X(\omega)d\mathbb{P}(\omega) = \sum_{k=0}^{n} x_k \mathbb{P}(X = x_k).$$

2. The random variable X is integrable iff $\int_\Omega |X(\omega)|d\mathbb{P}(\omega) < \infty$.

   Note: $|X| = X^+ + X^-$, $X^+ \leq |X|$, $X^- \leq |X|$.

3. If $X \leq Y$ and X and Y are integrable then

$$\int_\Omega X(\omega)d\mathbb{P}(\omega) \leq \int_\Omega Y(\omega)d\mathbb{P}(\omega).$$

4. If α and β are real constant and X and Y are integrable or if α, β are non-negative constant and X and Y are non-negative. Then

$$\mathbb{E}[\alpha X + \beta Y] = \alpha \mathbb{E}[X] + \beta \mathbb{E}[Y].$$

Two important convergence theorems:-

Definition:- Let $X_1, X_2, \cdots, X_n, \cdots$ be a sequence of random variables, all defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let X be another random variable defined on the same space. We say that $X_1, X_2, \cdots$ converges to X almost surely and write $\lim_{n\to\infty} X_n = X$ a.s. if $\mathbb{P}\{\omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\} = 1$.

Monotone convergence theorem:- Let $X_1, X_2, \cdots$ be a sequence of random variables converging almost surely to another random variable X. If $0 \le X_1 \le X_2 \le X_3 \cdots$ almost surely, then

$$\lim_{n\to\infty} \mathbb{E}[X_n] = \mathbb{E}[X].$$

Corollary 4.1. *Suppose the non-negative random variable X takes countably many values $x_0, x_1 \cdots$, then*

$$\mathbb{E}[X] = \sum_{k=0}^{\infty} x_k \mathbb{P}(X = x_k).$$

Proof. Let $A_k = \{X = x_k\}$. Then X can be written as

$$X = \sum_{k=0}^{\infty} x_k \mathcal{X}_{A_k}.$$

Define

$$X_n = \sum_{k=0}^{n} x_k \mathcal{X}_{A_k}.$$

Then $0 \le X_1 \le X_2 \le \cdots$ and $\lim_{n\to\infty} X_n = X$. Note that

$$\mathbb{E}[X_n] = \sum_{k=0}^{n} x_k \mathbb{P}(X = x_k).$$

Using Monotone convergence theorem, we obtain

$$\mathbb{E}[X] = \lim_{n\to\infty} \mathbb{E}[X_n] = \lim_{n\to\infty} \sum_{k=0}^{n} x_k \mathbb{P}(X = x_k) = \sum_{k=0}^{\infty} x_k \mathbb{P}(X = x_k).$$

□

Dominated Convergence Theorem:- *let $X_1, X_2, \cdots$ be a sequence of random variables converging almost surely to a random variable X. If there is another random variable Y such that $\mathbb{E}[Y] < \infty$ and $|X_n| \le Y$ almost surely for every n, then*

$$\lim_{n\to\infty} \mathbb{E}[X_n] = \mathbb{E}[X].$$

1. Consider the space $(\Omega, \mathcal{B}[0,1], \mathcal{L})$, where $\mathcal{L}$ is the Lebesgue measure. Define

$$X_n(\omega) = \begin{cases} n \text{ if } \omega \in [0, \frac{1}{n}) \\ 0 \text{ otherwise.} \end{cases}$$

Then $\lim_{n\to\infty} X_n(\omega) = 0$ a.s.

$$\lim_{n\to\infty} \int_0^1 X_n(\omega) d\mathbb{P}(\omega) = \lim_{n\to\infty} n \cdot \frac{1}{n} = 1 \neq 0 = \int_0^1 \lim_{n\to\infty} X_n(\omega) d\mathbb{P}(\omega)$$

2. Consider a sequence of normal densities, each with mean zero and and the n^{th} having variance $\frac{1}{n}$.

$$f_n(x) = \sqrt{\frac{n}{2\pi}} e^{\frac{-nx^2}{2}}.$$

If $x \neq 0$, then $\lim_{n\to\infty} f_n(x) = 0$ but

$$\lim_{n\to\infty} f_n(0) = \lim_{n\to\infty} \sqrt{\frac{n}{2\pi}} = \infty.$$

Therefore

$$f_n(x) \to 0 \text{ a.s.}$$

and

$$\int_{-\infty}^{\infty} f_n(x)dx = 1 \neq 0 = \int_{-\infty}^{\infty} \lim_{n\to\infty} f_n(x)dx.$$

Definition 4.2. *If X is an integrable random variable then the expectation of X is defined to be*

$$\mathbb{E}(X) = \int_\Omega X d\mathbb{P} = \mathbb{E}(X^+) - \mathbb{E}(X^-).$$

Definition 4.3. *Let X be an integrable random variable or a non-negative random variable and $A \in \mathcal{F}$. Then*

$$\int_A X d\mathbb{P} = \int_\Omega X \mathbb{I}_A d\mathbb{P}.$$

Proposition 4.4. *Suppose $X \geq 0$ and $\mathbb{E}(X) = 0$. Then $\mathbb{P}(X = 0) = 1$.*

Proof: Let $E = \{\omega \in \Omega : X(\omega) > 0\}$ and let $E_n = \{\omega \in \Omega : X(\omega) \geq 1/n\}$. Then by definition $0 = \mathbb{E}(X) \geq \int \frac{1}{n} 1_{E_n} d\mathbb{P} = \frac{1}{n}\mathbb{P}(E_n)$. Thus $\mathbb{P}(E_n) = 0$. Hence

$$\mathbb{P}(E) = \mathbb{P}(\cup_{n=1}^\infty E_n) \leq \sum_{n=1}^{\infty} \mathbb{P}(E_n) = 0.$$

Thus the conclusion. □

Exercise: Suppose $X > 0$ on A and $\int_A X d\mathbb{P} = 0$, then show that $\mathbb{P}(A) = 0$.

Theorem 4.5. *Let X be a random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and let g be a Borel-measurable function on $\mathbb{R}$. Then*

$$\mathbb{E}[|g(X)|] = \int_{\mathbb{R}} |g(x)| d\mu_X(x)$$

and if this quantity is finite, then

$$\mathbb{E}[g(X)] = \int_{\mathbb{R}} g(x) d\mu_X(x).$$

Definition:- Let $f(x)$ be a real-valued function defined on $\mathbb{R}$. The function $f(x)$ is said to be Borel measurable if for every Borel subset B of $\mathbb{R}$, the set $\{x : f(x) \in B\}$ is also a Borel subset of $\mathbb{R}$.

Theorem 4.6. *Let X be a random variable with density f. Then for any Borel-measurable function g on $\mathbb{R}$, we have,*

$$\mathbb{E}[|g(X)|] = \int_{-\infty}^{\infty} |g(x)| f(x) dx.$$

If this quantity is finite, then

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

Proof. **Step 1:-** If $g(x) = \mathbb{I}_B(x)$. Then L.H.S= $\mathbb{E}[\mathbb{I}_B(X)] = 1 \cdot \mathbb{P}(X \in B) = \mu_x(B)$. Since X has density so, $\mu_x(B) = \int_B f(x)dx = \int_{-\infty}^{+\infty} g(x)f(x)dx$=R.H.S.

Step 2:-If $g(x) = \sum_{k=1}^{n} \alpha_k \mathbb{I}_{B_k}$, then

$$\begin{aligned}\text{L.H.S.} &= \mathbb{E}(g(X)) = \mathbb{E}(\sum_{k=1}^{n} \alpha_k \mathbb{I}_{B_k}(X)) = \sum_{k=1}^{n} \alpha_k \mathbb{E}(\mathbb{I}_{B_k}(X)) \\ &= \sum_{k=1}^{n} \alpha_k \int_{-\infty}^{+\infty} \mathbb{I}_{B_k}(x)f(x)dx = \int_{-\infty}^{+\infty} \sum_{k=1}^{n} \alpha_k \mathbb{I}_{B_k}(x)f(x)dx \\ &= \int_{-\infty}^{+\infty} g(x)f(x)dx.\end{aligned}$$

Step 3:- Let $g(x)$ be a given non-negative Borel-measurable function. Then $\exists$ a sequence of simple functions $0 \leq g_1 \leq g_2 \leq \cdots$ such that

$$\lim_{n\to\infty} g_n(x) = g(x).$$

Now by Monotone Convergence Theorem and previous step we have

$$\begin{aligned}\mathbb{E}(g(X)) &= \lim_{n\to\infty} \mathbb{E}(g_n(X)) \\ &= \lim_{n\to\infty} \int_{-\infty}^{+\infty} g_n(x)f(x)dx = \int_{-\infty}^{+\infty} g(x)f(x)dx.\end{aligned}$$

Step 4:- For any Borel measurable function g, we have $\mathbb{E}[g^+(X)] = \int_{-\infty}^{+\infty} g^+(x)f(x)dx$ and $\mathbb{E}[g^-(X)] = \int_{-\infty}^{+\infty} g^-(x)f(x)dx$. Thus $\mathbb{E}|g(X)| = \int_{-\infty}^{+\infty} |g(x)|f(x)dx$ and $\mathbb{E}(g(X)) = \int_{-\infty}^{+\infty} g(x)f(x)dx$ provided $\mathbb{E}[|g(X)|] < +\infty$. $\square$

5 Change of Measure:-

Theorem 5.1. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let Z be an almost surely non-negative random variable with $\mathbb{E}[Z] = 1$, for $A \in \mathcal{F}$, define*

$$\tilde{\mathbb{P}}(A) = \int_A Z(\omega)d\mathbb{P}(\omega).$$

Then $\tilde{\mathbb{P}}$ is a probability measure. Furthermore, if X is non-negative random variable, then

$$\tilde{\mathbb{E}}[X] = \int_\Omega X(\omega)d\tilde{\mathbb{P}} = \mathbb{E}[XZ]. \tag{1}$$

If Z is almost surely strictly positive, we have

$$\mathbb{E}[Y] = \tilde{\mathbb{E}}\left[\frac{Y}{Z}\right], \tag{2}$$

for every non-negative random variable Y.

Proof. $\tilde{\mathbb{P}}(\Omega) = \int_\Omega Z(\omega)d\mathbb{P}(\omega) = \mathbb{E}[Z] = 1$. Let $A_1, A_2, \cdots$ be a sequence of disjoint sets in $\mathcal{F}$, define $B_n = \cup_{k=1}^n A_k$

and $B_\infty = \cup_{k=1}^{\infty} A_k$, then $\mathbb{I}_{B_n}(w) = \sum_{k=1}^{n} \mathbb{I}_{A_k}(w)$ and $\mathbb{I}_{B_\infty}(w) = \sum_{k=1}^{\infty} \mathbb{I}_{A_k}(w)$ and $\mathbb{I}_{B_n}(w) \uparrow \mathbb{I}_{B_\infty}(w)$. By MCT

$$\begin{aligned}
\tilde{\mathbb{P}}(B_\infty) &= \int_\Omega \mathbb{I}_{B_\infty}(\omega) Z(\omega) d\mathbb{P} \\
&= \lim_{n\to\infty} \int_\Omega \mathbb{I}_{B_n}(\omega) Z(\omega) d\mathbb{P} \\
&= \lim_{n\to\infty} \int_\Omega \sum_{k=1}^{n} \mathbb{I}_{A_k}(\omega) Z(\omega) d\mathbb{P}(\omega) \\
&= \lim_{n\to\infty} \sum_{k=1}^{n} \tilde{\mathbb{P}}(A_k) = \sum_{k=1}^{\infty} \mathbb{P}(A_k).
\end{aligned}$$

Therefore $\tilde{\mathbb{P}}$ is a probability measure. Now suppose X is a non-negative random variable. If $X = \mathbb{I}_A$, then $\tilde{\mathbb{E}}[X] = \tilde{\mathbb{P}}(A) = \int_\Omega \mathbb{I}_A(\omega) Z(\omega) d\mathbb{P}(\omega) = \mathbb{E}[\mathbb{I}_A Z] = \mathbb{E}[XZ]$. Now one can complete the proof using standard machine develoved in previous theorem. When $Z > 0$ a.s., $\frac{Y}{Z}$ is defined and we may replace X in (1) by $\frac{Y}{Z}$ to obtain (2). $\square$

Definition:- Let Ω be a non-empty set and $\mathcal{F}$ a σ-algebra of subsets of Ω. Two probability measures $\mathbb{P}$ and $\tilde{\mathbb{P}}$ on $(\Omega, \mathcal{F})$ are said to be equivalent if they agree on which sets in $\mathcal{F}$ have probability zero.

Under the assumptions of the above theorem and $Z > 0$ a.s., $\mathbb{P}$ and $\tilde{\mathbb{P}}$ are equivalent. Let $A \in \mathcal{F}$ and $\mathbb{P}(A) = 0$. Then the random variable $1_A Z$ is $\mathbb{P}$ a.s. zero $\Rightarrow \tilde{\mathbb{P}}(A) = \int_\Omega 1_A(\omega) Z(\omega) d\mathbb{P}(\omega) = 0$. On the other hand, suppose $B \in \mathcal{F}$ satisfies $\tilde{\mathbb{P}}(B) = 0$. Then $\frac{1}{Z} 1_B = 0$ almost surely under $\tilde{\mathbb{P}}$, so $\tilde{\mathbb{E}}\left[\frac{1}{Z} 1_B\right] = 0 = \mathbb{P}(B)$. Hence $\mathbb{P}$ and $\tilde{\mathbb{P}}$ are equivalent.

Example:- Let $\Omega = [0,1]$, $\mathcal{F} = \mathcal{B}[0,1]$ $\mathbb{P} = \mathcal{L}$, Lebesgue measure, and let $0 \le a \le b \le 1$

$$\begin{aligned}
\tilde{\mathbb{P}}[a,b] &= \int_a^b 2\omega d\omega = b^2 - a^2 \\
&= \int_a^b 2\omega d\mathbb{P}(\omega) \text{ [using the fact that } d\mathbb{P}(\omega) = d\omega].
\end{aligned}$$

So, $\tilde{\mathbb{P}}(B) = \int_B 2\omega d\mathbb{P}(\omega)$ for every Borel set $B \in \mathcal{B}[0,1]$. Set $Z(\omega) = 2\omega > 0$ a.s. in $\mathbb{P}$ and $\mathbb{E}[Z] = \int_0^1 2\omega d\omega = 1$. By (1), for every non-negative random variable X, we have

$$\int_0^1 X(\omega) d\tilde{\mathbb{P}}(\omega) = \int_0^1 X(\omega) 2\omega d\mathbb{P}(\omega).$$

This suggests the notation

$$d\tilde{\mathbb{P}}(\omega) = 2\omega d\omega = 2\omega d\mathbb{P}(\omega).$$

Let X be a standard normal random variable on $(\Omega, \mathcal{F}, \mathbb{P})$. Define $Y = X + \theta$, $\theta > 0$. Then $\mathbb{E}[Y] = \theta$ and $var(Y) = 1$. We want to change to a new probability measure $\tilde{\mathbb{P}}$ on Ω under which Y is a standard normal random variable i.e., $\tilde{\mathbb{E}}[Y] = 0$ and $\tilde{Var}(Y) = \tilde{\mathbb{E}}(Y - \tilde{\mathbb{E}}(Y))^2 = 1$. Define the random variable

$$Z(\omega) = \exp\{-\theta X(\omega) - \frac{1}{2}\theta^2\} \; \forall \omega \in \Omega.$$

We see

$$Z(\omega) > 0 \text{ and } \mathbb{E}[Z] = 1.$$

$$\begin{aligned}
\mathbb{E}[Z] &= \int_{-\infty}^{\infty} \exp\{-\theta x - \frac{1}{2}\theta^2\}\phi(x)dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\{-\theta x - \frac{1}{2}\theta^2 - \frac{1}{2}x^2\}dx \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\{\frac{1}{2}(\theta + x)^2\}dx \text{ (put } \theta + x = y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\{\frac{-y^2}{2}\}dy = 1.
\end{aligned}$$

Define $\tilde{\mathbb{P}}(A) = \int_A Z(\omega)d\mathbb{P}(\omega)\ \forall A \in \mathcal{F}$

$$
\begin{aligned}
\tilde{\mathbb{P}}(Y \leq b) &= \int_\Omega \mathbb{I}_{\{Y(\omega)\leq b\}} Z(\omega)d\mathbb{P}(\omega) = \int_\Omega \mathbb{I}_{\{X(\omega)\leq b-\theta\}} \exp\{-\theta X(\omega) - \frac{1}{2}\theta^2\}d\mathbb{P}(\omega) \\
&= \int_{-\infty}^{\infty} \mathbb{I}_{\{x\leq b-\theta\}} \exp\{-\theta x - \frac{1}{2}\theta^2\}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx \\
&= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{b-\theta} \exp\{-\theta x - \frac{1}{2}\theta^2 - \frac{1}{2}x^2\}dx \\
&= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{b} e^{-\frac{y^2}{2}}dy \text{ (put } y = \theta + x).
\end{aligned}
$$

So, $\tilde{\mathbb{P}}(Y \leq b) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{b} e^{-\frac{y^2}{2}}dy$ which shows that Y is a standard normal random variable under the probability measure $\tilde{\mathbb{P}}$.

Theorem 5.2. ***(Radon-Nikodym Theorem):-*** *Let $\mathbb{P}$ and $\tilde{\mathbb{P}}$ be equivalent probability measures defined on $(\Omega, \mathcal{F})$. Then there exists an almost surely positive random variable Z such that $\mathbb{E}[Z] = 1$ and*

$$\tilde{\mathbb{P}}(A) = \int_A Z(\omega)d\mathbb{P}(\omega),$$

for every $A \in \mathcal{F}$.