Fundamental Theorems and Rules - Probability Theory

Md Jamal Hossain

Sigma-Field

A family A of subsets A of Ω,A $\in \mathcal{P}(\Omega)$ with

- $\emptyset \in \mathcal{A}$
- $A \in \mathcal{A}$, `then` $A^c \in \mathcal{A}$
- $A_1, A_2, ... \in \mathcal{A}$, `then` $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$

is called σ-field

Definition of Measure

The set function μ is called a measure on $(\Omega, \mathcal{A})$ if

- $\mu : \mathcal{A} \to [0, \infty]$ `Positive Mapping`
- $\mu(\emptyset) = 0$
- $A_1, A_2, ... \in \mathcal{A}$ and pairwise disjoint

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} u(A_i) \quad \sigma - additivity$$

Probability Measure

The set function P is called a probability measure on $(\Omega, \mathcal{A})$ if

- $P : \mathcal{A} \to [0, \infty]$ `Positive Mapping`
- $A_1, A_2, ... \in \mathcal{A}$ and pairwise disjoint

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

- $P(\Omega) = 1$

The triplet $(\Omega, \mathcal{A}, P)$ is called a probability space.

Note

For a countable sample space $\Omega = \{\omega_i\} i \in I$ (with countable I) a probability measure P on Ω is specified by

$$p_i = P\left(\{\omega_i\}\right)$$

Then for every $A \subseteq \Omega$

$$P(A) = \sum_{\omega \in A} P(\{\omega\})$$

Then for $A_1, A_2, ... \in \mathcal{A}$ and being pairwise disjoint, it holds that

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

Definition of Field

A family $\mathcal{A}^*$ of subsets of Ω is called field if

- $\emptyset \in \mathcal{A}^*$
- $A \in \mathcal{A}^*, \texttt{ then } A^c \in \mathcal{A}^*$
- $A_1, A_2 \in \mathcal{A}^*, \texttt{ then } A_1 \cup A_2 \in \mathcal{A}^*$

Definition of Pre-Measure

If $\mathcal{A}^*$ is a field and the set function $P^* : \mathcal{A}^* \to [0, \infty]$ and pairwise disjoint $A_1, A_2, ...$ such that $A_1, A_2, ... \in \mathcal{A}^*$ and $\bigcup_{i=1} A_i \in \mathcal{A}^*$, then it holds that

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

P^* is called a **Pre-measure**. If $P^*(\Omega) = 1$, then P^* is called the **Probability pre-measure**.

Simple Function

Given $(\Omega, \mathcal{A})$ and $\alpha_1, \alpha_2, ..., \alpha_n \in (0, \infty)$ and $A_1, ..., A_n \in \mathcal{A}$ are disjoint, then

$$s = \sum_{i=1}^{n} \alpha_i 1_{A_i}$$

is a simple function.

Measureability

Given $(\Omega, \mathcal{A})$ measureable space, the numeric function $f : \Omega \to [0, \infty]$ is measureable if

$$f^{-1}(B) \in \mathcal{A} \quad \texttt{for} \quad B \in \mathbb{B}$$

where $\mathbb{B}$ is a Borel-σ algebra on $[0, \infty]$

NOTE

More precisely measureability is

- $f : (\omega, \mathcal{A}) \to ([0, \infty], \mathbb{B})$
- f is $\mathcal{A} - \mathbb{B}$ measureable

Cumulative Distribution Function

For a probability measure P on $(\mathbb{R}, \mathbb{B})$ the function $F : \mathbb{R} \to [0, 1]$ given by

$$F(b) = P((-\infty, b])$$

is called the cumulative distribution function.

Defining Measure Integral via simple function

$$\int_E s d\lambda = \int_E \sum_{i=1}^{n} \alpha_i 1_{A_i} d\lambda = \sum_{i=1}^{n} \alpha_i \lambda \left(A_i \cap E \right)$$

Defining Measure Integral via measureable function f

If $f : \Omega \to [0, \infty]$ is an arbitrary measureable function and $E \in \mathcal{A}$, then

$$\int_E f d\lambda =: sup \left\{ \int_E s d\lambda : 0 \leq s \leq f, \quad \texttt{s simple function} \right\}$$

Condition of Integrability

The function f is Lebesgue integrable if

$$\int_\Omega |f| d\lambda < \infty$$

Random Variable

A $\mathbb{R}^k$ valued random variable is a function $X : \Omega \to \mathbb{R}$ where $(\Omega, \mathcal{A})$ is a measureable space and X fulfills

$$X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{A} \quad B \in \mathbb{B}^k$$

Monotone Convergence Theorem

Let $X_1, .., X_n$ real valued random variables and X_n converges point wise $\lim n \to \infty X_n(\omega) = X(\omega) \quad \forall \omega \in \Omega$ and $0 \leq X_n \leq X_{n+1}$ for $n \geq 1$, then it holds that

$$\lim_{n \to \infty} E[X_n] = E\left[\lim_{n \to \infty} X_n \right]$$

Fatou's Lemma

Let $X_1, ..., X_n$ real valued random variables such that $X_n \geq 0$ and

$$X(\omega) = \lim_{n\to\infty} \texttt{inf} X_n(\omega)$$

Then it holds that

$$E(X) \leq \lim_{n\to\infty} \texttt{inf} E(X_n)$$

Dominated Convergence Theorem

Let $X_1, .., X_n$ real valued random variables and $X_{_n}$ converges point wise $\lim n \to \infty X_n(\omega) = X(\omega) \quad \forall\omega \in \Omega$ and $|X| \leq Y$ for random variable Y such that $E(Y) < \infty$, then it holds that

$$\lim_{n\to\infty} E[X_n] = E\left[\lim_{n\to\infty} X_n\right]$$

Note

$\lim_{n\to\infty} E[X_n] = E\left[\lim_{n\to\infty} X_n\right]$ holds if $X_n \to X$ and X_n is uniformly bounded $|X_n| \leq C$ for $C \in \mathbb{R}$

Convergence in Probability

Let $X_1, ..., X_n$ and X random variables in the probability space $(\Omega, \mathcal{A}, P)$ with values in $(\mathbb{R}, \mathcal{B})$. Then X_n converges in probability to X for $\epsilon > 0$ if

$$\lim_{n\to\infty} P(|X_n - X| > \epsilon) \to 0$$

Almost Sure Convergence X_n converges almost surely to X if

$$P(\lim_{n\to\infty} X_n \to X) = 1$$

Convergence in p$^{th}Mean$

X_n converges in p^{th} mean to X for $p \geq 1$ if

$$\lim_{n\to\infty} E\left(|X_n - X|^p\right) \to 0$$

Convergence in Distribution

Given $f : \mathbb{R} \to \mathbb{R}$ continuous and bounded function, X_n converges to X in distribution if

$$\lim_{n\to\infty} E\left(f(X_n)\right) = E\left(f(x)\right)$$

Jensen inequality

$$(E(|Y|)^p \leq E\left(|Y|^p\right) \quad \texttt{for} \quad p \geq 1$$

Markov Inequality

For random variable X_n and a monotone increasing function $g : [0, \infty) \to [0, \infty)$ with $g(x) > 0$ for $x > 0$, it holds that for every $\epsilon > 0$

$$P\left(|X_n| > \epsilon\right) \leq \frac{E[g(|x|)]}{g(\epsilon)}$$

Note

Markov inequality says that convergence in p^{th} mean implies convergence in probability

Cheybechev Inequality

For real valued random variable X with finite second moment or variance and $\epsilon > 0$, it holds that

$$P\left(|X - E(X)| \geq \epsilon\right) \leq \frac{Var(X)}{\epsilon^2}$$

Strong Law of Large Number

Let $X_1, ..., X_n$ for $i = 1, 2, 3...$ real valued i, i, d random variables on the probability space $(\Omega, \mathcal{A}, P)$ with finite mean $E(X_i) = \mu < \infty$, then it holds that

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{a.s} \mu$$

Weak Law of Large Number: Version - I

Let $X_1, ..., X_n$ for $i = 1, 2, 3...$ real valued and uncorrelated random variables with $E(X_1) = E(X_2) = ... = E(X_n) = \mu$ for $\mu \in \mathbb{R}$ and $Var(X_i) \leq c$ for $c \in \mathbb{R}$, then it holds that

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{p} \mu$$

Weak Law of Large Number: Version - II

Let $X_1, ..., X_n$ for $i = 1, 2, 3...$ real valued i, i, d random variables with finite mean $E(X_i) = \mu < \infty$, it holds that

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{p} \mu$$

Characteristic Function

Let X be a random variable in $\mathbb{R}$ and the function $\phi_X : \mathbb{R} \to \mathbb{C}$ defined by

$$\phi_X(t) = E\left(e^{itX}\right) = E[cos(tX)] + iE[sin(tX)]$$

is called the characteristic function of X where $i^2 = -1$

Levy's Continuity Theorem

For random variable X and X_n in $\mathbb{R}$, $X_n \xrightarrow{d} X$ if and only if for $t \in \mathbb{R}$ $\phi_{X_n}(t) \to \phi_X(t)$

Lindeberg-Levy Central Limit Theorem

Let $X_1, ..., X_n$ for $i = 1, 2, 3...$ real valued i, i, d random variables with $E(X_i) = \mu$ and $Var(X_i) = \sigma^2 \in (0, \infty)$. Then

$$\frac{X_1 + X_2 + ... + X_n - n\mu}{\sqrt{n}\sigma} \xrightarrow{d} Z \sim N(0, 1)$$

Delta Method

For i,i,d random variables $X_1, X_2, ..., X_n$ with $E(X_i) = \mu$ and $Var(X_i) = \sigma^2 < \infty$, Lindeberg-Levy CLT implies

$$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} N(0, \sigma^2)$$

and for $g(\bar{X}_n)$ where g is continously differentiable with $g'(\mu) \neq 0$, it holds that

$$\sqrt{n}\left(g(\bar{X}_n) - g(\mu)\right) \xrightarrow{d} N\left(0, \left(g'(\mu)\right)^2 \sigma^2\right)$$

Multivariate CLT

Let $X_1, X_2, ..., X_n$ are i, i, d $\mathbb{R}^k$-valued random variables with mean vector μ and finite positive definite covariance matrix Σ, then

$$\frac{X_1 + X_2 + ... + X_n - n\mu}{\sqrt{n}} \xrightarrow{d} Z \sim N(0_k, \Sigma)$$

Variance of Conditional Expectation

$$Var(Y|X) = E\left[\left(Y - E(Y|X)\right)^2 |X\right]$$

Iterated Expectation

$$E[E(Y|X)] = E(Y)$$

$E[E(Y|X, Z)|Z] = E(Y|Z)$ `Expectation with smaller set prevails`

$E[E(Y|X)|X,Z] = E(Y|X)$ `Expectation with smaller set prevails`

Variance of Y

$$Var(Y) = E\left(Var(Y|X)\right) + Var(E(Y|X))$$

Proof

Now using $Var(Y|X) = E\left[(Y - E(Y|X))^2 |X\right]$

$$Var(Y) = E\left(E\left[(Y - E(Y|X))^2 |X\right]\right) + E\left(E\left((Y|X) - E(E(Y|X))\right)^2\right)$$
$$= E(E(Y^2|X)) - E((E(Y|X))^2) + E((E(Y|X))^2) - (E(Y))^2$$
$$= E(Y^2) - (E(Y))^2 = E[(Y - E(Y))^2] = Var(Y)$$

One inequality

$$P(|Y| \geq \epsilon|X) \leq \frac{E(Y^2|X)}{\epsilon^2}$$

This is the conditional version of Markov inequality.

Inequality relating to expectation of Conditional Variance

$$E[Var(Y|X)] \geq E[Var(Y|X,Z)]$$

Conditional Expectation Formulae - I

$$E(E(Y|X)) = E(Y)$$

Proof

We know that marginal density for x can be found by

$$f(y) = \int f(x,y)dx$$

where $f(x,y)$ is the joint density.

$$E(Y) = \int yf(y)dy$$
$$= \int\int yf(x,y)dxdy$$
$$= \int\int yf_{Y|X}(y|x)f(x)dxdy \quad \texttt{since} \quad f_{Y|X} = \frac{f(X,Y)}{f(X)}$$

$$= \int \int y f_{Y|X}(y|x) dy f(x) dx$$
$$= \int E(Y|X = x) f(x) dx$$
$$= E[E(Y|X = x)]$$

Conditional Expectation Formulae - II

$$E[E(Y|X,Z)|X] = E(Y|X)$$

Proof

$$E[E(Y|X,Z)|X] = \int \int y f_{Y|X,Z}(y|x,z) dy f_{Z|X}(z|x) dz$$

Then we use the following

$$f_{y|x,z} = \frac{f(x,y,z)}{f(x,z)} \quad \texttt{and} \quad f_{z|x} = \frac{f(x,z)}{f(x)}$$

$$E[E(Y|X,Z)|X] = \int y \frac{1}{f(x)} \left(\int f(x,y,z) dz \right) dy$$

Since

$$\int f(x,y,z) dz = f(x,y)$$

$$= \int y \frac{1}{f(x)} f(x,y) dy$$

Now we can write

$$f(x,y) = f(x) f(y|x)$$

Then

$$E[E(Y|X,Z)|X] = \int y \frac{f_{Y|X} f(x)}{f(x)} dy = \int y f_{Y|X} dy = E(Y|X)$$

Expectation Property - I

$$Var(X) = 0 \implies X = E(X) \implies P(X = E(X)) = 1$$

That means $Var(X) = 0$ means X is constant.

Expectation Property - II

$$E[g(X)] = \int g(x)f(x)dx$$

Expectation Property - III

$$X_1 \leq X_2 \implies E(X_)\leq E(X_2) \quad \texttt{Monotoncity Property}$$

Expectation Property - IV

$$E(1_A) = P(A)$$

Proof

$$E(1_A) = \int_A 1_A dP = \int_A dP = P(A)$$

One can alternatively prove this using Dirac delta function as well.

Lemma

If X and Y are independent and $g : \mathbb{R}^k \to \mathbb{R}^l$ and $h : \mathbb{R}^m \to \mathbb{R}^n$ and $\mathbb{B}^k - \mathbb{B}^l$ and $\mathbb{B}^m - \mathbb{B}^n$ measureable, then

$$g(X) \quad \texttt{and} \quad h(Y) \quad \texttt{are independent}$$

Note

$$\sqrt{n}\left(\bar{X}_n - \mu\right) = O_p(1) \iff \bar{X}_n = \mu + O_p\left(\frac{1}{\sqrt{n}}\right)$$

$$\sqrt{n}\left(\bar{X}_n - \mu\right) = o_p(1) \iff \bar{X}_n = \mu + o_p\left(\frac{1}{\sqrt{n}}\right)$$