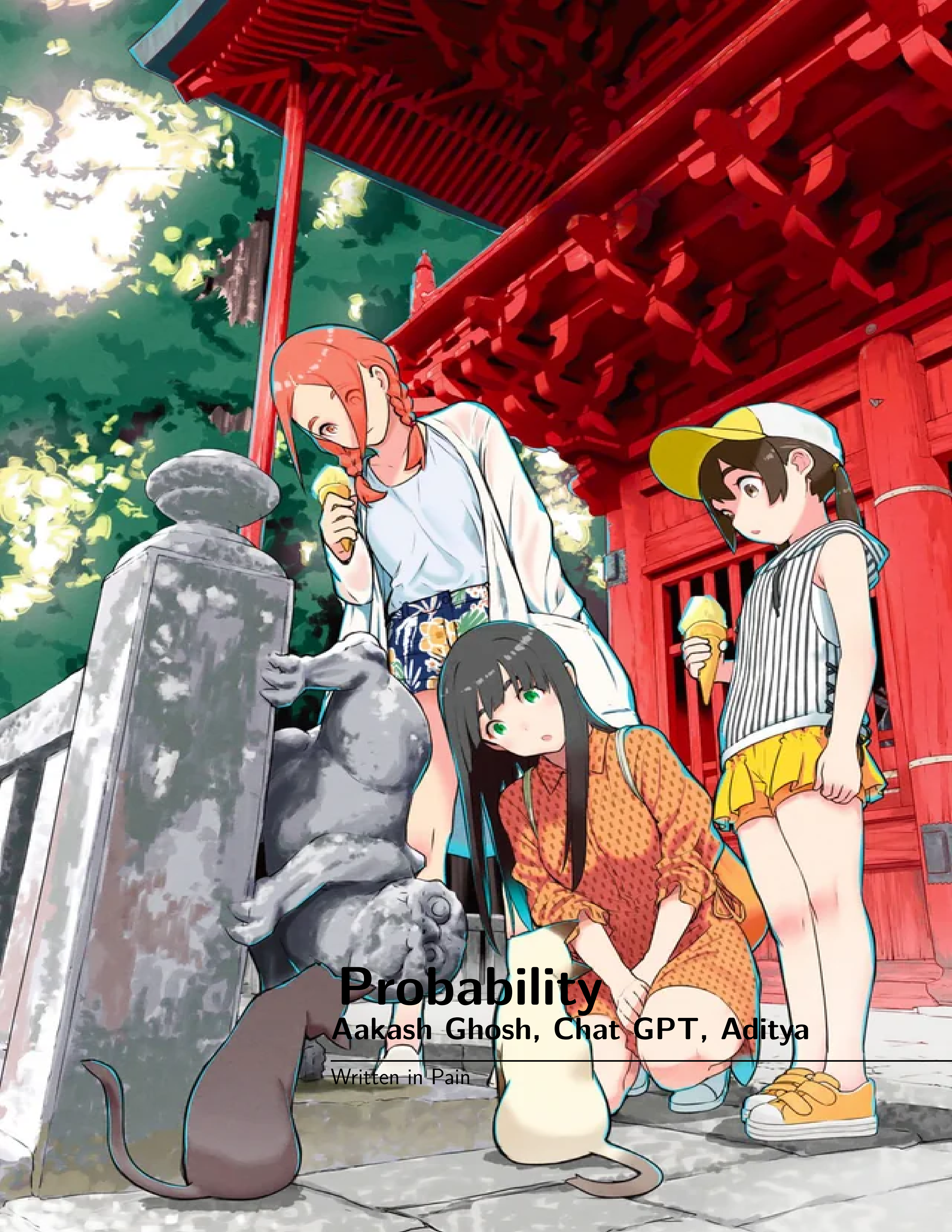


Probability

Aakash Ghosh, Chat GPT, Aditya

Written in Pain

Contents

1 Basic Stuff 2
- 1.1 Probability measures and Random variable . . . 2
- 1.2 Generalized Inverse . . . 3
- 1.3 Independence of events . . . 3
- 1.4 Kolmogorov's 0-1 law . . . 4
- 1.5 Borel-Cantelli lemma . . . 5

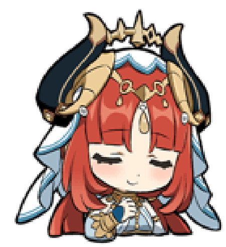
2 Convergence Shenanigans 6
- 2.1 Almost sure convergence . . . 6
- 2.2 Other modes of convergence . . . 7
- 2.3 Uniform Integrability . . . 8
- 2.4 Metrics of convergence . . . 9
- 2.5 Continuous mapping theorem . . . 10
- 2.6 Slutsky's theorem . . . 10
- 2.7 Law of Large number . . . 11
- 2.8 Applications of LLN . . . 13

3 Convergence in distribution 14
- 3.1 Method of moments . . . 16

4 Characteristic functions and CLT 17
- 4.1 Characteristic function . . . 17
- 4.2 Multivariate CF . . . 19
- 4.3 CF and distributional convergence . . . 19
- 4.4 CF and weak convergence . . . 20
- 4.5 Central limit theorem . . . 20
- 4.6 CLT for independent but not identical sequence of R.V . . . 21

1 | Basic Stuff

1.1 | Probability measures and Random variable

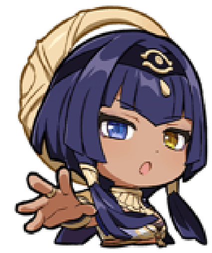


1. A probability measure is a finite measure space with a total measure of 1. The corresponding σ algebra is called the event space.
2. Definition: Random Variable
   A Random Variable is a measurable function.
3. Definition: CDF
   We define the CFD of a Random Variable(RV) as:

   $$F(x) = P(X \leq x) = P \circ X^{-1}(-\infty, x]$$

   Note that this defines a push forward measure on $\mathbb{R}$. We use μ and F interchangeably.
4. Properties of CDF:
   (a) It is monotone increasing
   (b) It is right continuous.
   (c) It has both left and right limits
   (d) If $P(X = x) = 0 \forall x \in \mathbb{R}$, it is said to be continuous RV. It is equivalent to F being continuous.
   (e) For a discrete RV, there are countable many x such that $P(X = x) \neq 0$.
   (f) $\lim_{x\to\infty} F(x) = 1, \lim_{x\to-\infty} F(x) = 0$

5. Definition: *Support*
   We define the support of a discrete distribution as:
   $$S = \{x|P(X = x) \neq 0\}$$
   We define the support of a continuous distribution with PDF f as the set:
   $$S = \{x|f(x) \geq 0\}$$



1.2 | Generalized Inverse

6. Definition: *Generalized Inverse*
   We define the Generalized Inverse of a CDF F as:
   $$F^{-}(u) = \inf\{x|F(x) \geq u\}$$
7. We can generate a RV with CDF F and an uniform distribution U. If $X = F^{-}U$ then $X \sim F$.
8. Properties of Quantile Functions(F^{-})
   (a) $F^{-}(u) \leq x \Leftrightarrow u \leq F(x)$
   (b) $P(X \leq x) = F(x)$
   (c) F^{-} is monotone increasing
   (d) It is left continuous
   (e) It has both left and right limits

1.3 | Independence of events

9. Definition: *Independence of Events*: The set of events $\{A_i\}$ is said to be independent if for all finite subcollection we have:
   $$P\left(\cap A_{i_k}\right) = \prod P(A_{i_k})$$

10. We can generalise this to sigma algebra where for each selection $A_i \in \mathcal{M}_i$ we have Independence.

11. Definition: *Pairwise Independence of Events*: The set of events $\{A_i\}$ is said to be pairwise independent if for all finite subcollection we have:

$$P(A_i \cap A_j) = P(A_i)P(A_j)$$

Independence implies pairwise Independence, but the converse is not true.

1.4 | Kolmogorov's 0-1 law

12. Definition: *Tail of a sequence* Let $\{A_i\}_{i\in\mathbb{N}}$ be a sequence of events. Define:

$$\mathcal{A}_n = \sigma\{A_1, A_2 \ldots A_n\}$$

$$\mathcal{A}'_n = \sigma\{A_n, A_{n+1}, A_{n+2} \ldots\}$$

13. The Kolmogorov 0-1 law states that for any collection X_n of independent random variables with a common distribution function F, and for any event E determined by the tail σ-algebra $\mathcal{T}_n = \bigcap_{n=1}^{\infty} \sigma(X_n, X_{n+1}, \ldots)$, the probability of E is either 0 or 1, i.e.,

$$\mathbb{P}(E) = 0 \quad \text{or} \quad \mathbb{P}(E) = 1.$$

14. Definition: *Limsup of sequence Events*: This sigma algebra consisting of events occurring infinitely often. The limit superior (lim sup) of a sequence of events A_n is defined as:

$$\limsup_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k = \{\omega \in \Omega : \omega \in A_n \text{ for infinitely many } n\}.$$

15. Definition: *Liminf of sequence Events*: This sigma algebra consisting of events occurring always eventually. The limit inferior (lim inf) of a sequence of events A_n is defined as:

$$\liminf_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k = \{\omega \in \Omega : \omega \in A_n \text{ for all but finitely many } n\}.$$

1.5 | Borel-Cantelli lemma

16. Definition: *The first Borel-Cantelli lemma*:The first Borel-Cantelli lemma states that if A_n is a sequence of events such that $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$, then

$$P(\limsup_{n\to\infty} A_n) = 0,$$

where $\limsup_{n\to\infty} A_n$ is the lim sup of the sequence of events A_n. In other words, almost surely, only finitely many of the events A_n occur.

17. Definition: *The second Borel-Cantelli lemma*:The second Borel-Cantelli lemma states that if A_n is a sequence of independent events and $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$, then

$$P(\limsup_{n\to\infty} A_n) = 1,$$

where $\limsup_{n\to\infty} A_n$ is the $\limsup$ of the sequence of events A_n. In other words, almost surely, infinitely many of the events A_n occur. This also holds for pairwise independent events.

2 | Convergence Shenanigans

2.1 | Almost sure convergence

18. Definition: *Absolute convergence*: Absolute convergence of a random variable is a property that applies to a sequence of random variables X_n. We say that the sequence is absolutely convergent if

$$\mathbb{P}\left(\omega \text{ such that } X_n(\omega) \to X(\omega)\right) = 1.$$

19. Equivalently if $X_n \to X$ almost surely then:

$$P\left(|X_n - X| \geq 1/k \text{ infinitely often}\right) = 0 \forall k$$

20. Suppose $\sum_{n \in \mathbb{N}} P\left(|X_n - X| \geq 1/k\right) < \infty \forall k$ then $X_n \to X$ a.s. (almost surely). This is sufficient but not necessary for a.s convergence.

21. If X_n is sequence of independent R.V then $X_n \to c$ a.s.where c is a constant iff $\sum_{n \in \mathbb{N}} P\left(|X_n - c| \geq \epsilon\right) < \infty \forall \epsilon > 0$

22. Let $\{X_n\}$ be a sequence of i.i.d random variable. Define $S_n = \sum_{i=1}^{n} X_i$. Then:

    (a) $P\left(\left|\frac{S_n}{n} - \mu\right| > \epsilon \text{ i.o}\right) = 0 \Leftrightarrow E(|X_1|) < \infty, E(|X_1|) = \mu$

    (b) $\sum_{n \in \mathbb{N}} P\left(\left|\frac{S_n}{n} - \mu\right| > \epsilon \text{ i.o}\right) < \infty \Leftrightarrow E(|X_1|^2) < \infty, E(|X_1|) = \mu$

23. Definition: *Markov's inequality*: Markov's inequality for random variables states that for any non-negative random variable X and any $a > 0$.

$$\mathbb{P}(|X| \geq a) \leq \frac{\mathbb{E}\left[|X|^p\right]}{a^p} \quad \forall p \in (0, \infty).$$

24. Suppose $E[e^{tz}] < \infty$. Then:
$$P(z \geq \delta) \geq \frac{E[e^{tz}]}{e^{-t\delta}}$$

2.2 | Other modes of convergence

25. Definition: *Convergence in probability*: A sequence of random variables X_n converges to a random variable X in probability, denoted by $X_n \xrightarrow{p} X$, if for every $\epsilon > 0$,
$$\lim_{n\to\infty} \mathbb{P}\left(|X_n - X| > \epsilon\right) = 0.$$

26. Definition: *L^r convergence*: Let X and X_n be random variables, and let r be a positive real number. We say that the sequence of random variables X_n converges to the random variable X in L^r sense, denoted by $X_n \xrightarrow{L^r} X$, if
$$\lim_{n\to\infty} \mathbb{E}\left[|X_n - X|^r\right] = 0.$$

27. Definition: *Convergence in distribution*: Let X and X_n be random variables with distribution functions F and F_n, respectively. We say that the sequence of random variables X_n converges to the random variable X in distribution, denoted by $X_n \xrightarrow{d} X$, if
$$\lim_{n\to\infty} F_n(x) = F(x),$$
for all continuity points x of the distribution function F.

    In other words, X_n converges to X in distribution if the distribution function F_n of X_n converges pointwise to the distribution function F of X at all continuity points of F.

28. a.s convergence $\Rightarrow$ convergence in probability

29. convergence in probability $\Rightarrow$ convergence in distribution

30. L^r convergence $\Rightarrow$ convergence in probability for $r \geq 1$

31. L^s convergence $\Rightarrow L^r$ convergence, for $s \geq r \geq 1$.

32. If $X_n \to C$ in distribution for $c \in \mathbb{R}$, then $X_n \to C$ in probability.

33. If $X_n \to X$ in probability, there exists a subsequence $\sigma(i)$ such that $X_{\sigma(n)} \to X$ almost surely

34. If X_n is almost surely monotone, and if $X_n \to X$ in probability then $X_n \to X$ almost surely.

35. Ref:

    https://en.wikipedia.org/wiki/Convergence_of_random_variables#

2.3 | Uniform Integrability

36. Definition: *Uniform integrability*: Let $X_1, X_2, \ldots$ be a sequence of random variables defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We say that the sequence is uniformly integrable if

$$\lim_{M\to\infty} \sup_n \mathbb{E}[|X_n|\mathbb{I}_{|X_n|>M}] = 0,$$

where $\mathbb{I}_{|X_n|>M}$ is the indicator function of the event $|X_n| > M$.

In other words, a sequence of random variables is uniformly integrable if the tails of the sequence are small in an integrable sense, meaning that the expectation of the absolute value of each random variable times an indicator function of its tail decreases to zero uniformly as the tail threshold increases.

37. If $\{X_n\}$ is U.I then $\sup_{n\geq 1} E\left(|X_n|\right) < \infty$

38. If $\{X_n\}$ is sequence of random variable, such that $\sup_{n\geq 1} E\left(|X_n|^k\right) < \infty$ for some $k > 1$. Then $\{X_n\}$ is U.I. and $\{|X_n|^l\}$ is U.I. for $1 < l < k$.

39. Let X and $\{X_n\}$ be a sequence of R.V. and suppose that $X_n \to X$ in probability. Let $r \in (0, \infty)$ and suppose $E\left(|X_n|^r\right) < \infty \forall n$. Then the following are equivalent:

    $\{|X_n|^r\}$ is U.I.

    $X_n \to X$ in L^r.

    $E\left(|X_n|^r\right) \to E\left(|X|^r\right)$ as $n \to \infty$.

2.4 | Metrics of convergence

40. L^r convergence follows just the L^r metric.
41. Convergence in probability is given by:

$$d_P(X,Y) = E\left[\frac{|X_n - X|}{1+|X_n - X|}\right]$$

42. Definition: *Ky Fan metric*: Let X and Y be two random variables with distributions P_X and P_Y on a metric space $(\mathcal{X}, d)$. The Ky Fan metric between X and Y, denoted by $d_{KF}(P_X, P_Y)$, is defined as

$$d_{KF}(P_X, P_Y) = \inf \epsilon > 0 : \text{there exists a coupling } \Pi \text{ such that } \mathbb{P}(d(X,Y) > \epsilon) \leq \epsilon,$$

where $\epsilon > 0$ and $d(X,Y)$ denotes the distance between X and Y in the metric space $(\mathcal{X}, d)$.

In other words, the Ky Fan metric measures the minimum amount of coupling required to ensure that the probability of the distance between X and Y exceeding a given threshold ϵ is no larger than ϵ itself.

For a sequence of random variables X_n converging in distribution to X, the Ky Fan metric can be used to measure the rate of convergence, as $d_{KF}(P_{X_n}, P_X)$ measures the distance between the distribution of X_n and X in the metric space $(\mathcal{X}, d)$ with respect to the Ky Fan metric.

43. Almost sure convergence is not metrizable. A simple counterexample to this fact is as follows:

Consider:

$$P(X_n = 0) = 1 - \frac{1}{n} \qquad P(X_n = 1) = \frac{1}{n}$$

Doesn't converge absolutely as:

$$\sum P\left(|X_n| > \epsilon\right) < \infty$$

But $X_n \to X$ in probability. Therefore, for all subsequence there is some sub subsequence which converge absolutely. So the whole sequence converges almost surely. But this is clearly not true.

Definition: *Supremum norm metric*: Let X and Y be two random variables with cumulative distribution functions F_X and F_Y. Then, the supremum norm metric between X and Y, denoted by $d_\infty(X,Y)$, is defined as

$$d_\infty(X,Y) = \sup_{x\in\mathbb{R}} |F_X(x) - F_Y(x)|.$$

In other words, the supremum norm metric measures the maximum difference between the cumulative distribution functions of X and Y. The supremum norm metric is a metric on the space of all random variables equipped with the convergence in distribution topology.

2.5 | Continuous mapping theorem

44. Theorem: *Continuous Mapping Theorem*: Let X_n be a sequence of random variables that converges in probability to a random variable X, and let $g : \mathbb{R} \to \mathbb{R}$ be a continuous function. Then, the sequence of random variables $Y_n = g(X_n)$ converges in probability to $Y = g(X)$.

    In other words, if $X_n \xrightarrow{P} X$, and g is a continuous function, then $g(X_n) \xrightarrow{P} g(X)$. It extends similarly to a.s and convergence in distribution.

2.6 | Slutsky's theorem

45. Theorem: *Slutsky's Theorem*: Let X_n and Y_n be sequences of random variables, and let c be a constant. Assume that X_n converges in distribution to a random variable X, and Y_n converges in probability to a constant c. Then, the following statements hold:

    (a) $X_n + Y_n$ converges in distribution to $X + c$.

    (b) $X_n Y_n$ converges in distribution to cX.

    (c) If $c \neq 0$, then $\frac{X_n}{Y_n}$ converges in distribution to $\frac{X}{c}$, provided that Y_n does not converge to zero in probability.

46. Definition: *Convergence in distribution of (X_n, Y_n)*: Let (X_n, Y_n) be a sequence of bivariate random variables, and let (X, Y) be a bivariate random variable. We say that (X_n, Y_n) converges in distribution to (X, Y), denoted as $(X_n, Y_n) \xrightarrow{d} (X, Y)$, if for any continuous bounded function $f : \mathbb{R}^2 \to \mathbb{R}$, we have:

$$\lim_{n\to\infty} \mathbb{E}[f(X_n, Y_n)] = \mathbb{E}[f(X, Y)].$$

In other words, (X_n, Y_n) converges in distribution to (X, Y) if and only if the expected value of any continuous bounded function of (X_n, Y_n) converges to the expected value of the same function of (X, Y).

This definition can be extended to sequences of k-dimensional random vectors similarly. The convergence in distribution of (X_n, Y_n) has many applications in probability theory, statistics, and econometrics, especially in the study of multivariate stochastic processes and time series analysis.

2.7 | Law of Large number

47. Definition: *Kronecker's Lemma*: Let a_n and b_n be sequences of positive real numbers such that $\sum_{n=1}^{\infty} a_n < \infty$. If b_n is a sequence of non-negative real numbers such that b_n increases to infinity, then:

$$\frac{1}{b_n} \sum_{i=1}^{n} a_i b_i \to 0$$

as $n \to 0$. Suppose $\sum_{k=1} \infty a_k/k$ converges then $\frac{1}{n} \sum_{i=1}^{n} a_i$ converges.

48. Theorem (Portmanteau Theorem): Let X_n be a sequence of random variables and let X be another random variable. The following conditions are equivalent:

    (a) $X_n \to X$ almost surely.

    (b) $X_n \to X$ in probability.

    (c) $X_n \to X$ in distribution.

    (d) For all bounded and continuous functions $f : \mathbb{R} \to \mathbb{R}$, we have $\mathbb{E}[f(X_n)] \to \mathbb{E}[f(X)]$.

49. In particular, for a sequence of random variables X_n and a random variable X, the following conditions are equivalent:

(a) $\sum_{n=1}^{\infty} X_n$ converges to X almost surely.
(b) $\sum_{n=1}^{\infty} X_n$ converges to X in probability.
(c) $\sum_{n=1}^{\infty} X_n$ converges to X in distribution.

50. Theorem *Bessel's inequality*: Let X_n be a sequence of independent random variables with $\mathbb{E}[X_n] = 0$ and $\mathrm{Var}(X_n) = \sigma_n^2 < \infty$. Then, the following statements are equivalent:

    (a) The sum $\sum_{n=1}^{\infty} X_n$ converges in L^2.
    (b) The series $\sum_{n=1}^{\infty} \sigma_n^2$ converges.

51. *Kolmogorov's One Series Theorem*: Let $\{X_n\}$ be a sequence of independent random variables. $E(X_n)^2 < \infty$ for all $n \geq 1$. Then, $\sum_{k=1}^{\infty} \frac{Var(X_k)}{k^2} < \infty \implies \frac{1}{n} \sum_{n=1}^{\infty} (X_n - E(X_n)) \xrightarrow{a.s.} 0$.

52. *Kolmogorov's Three Series Theorem*:Suppose $\{X_n\}$ is a sequence of independent random variables. Set $Y_n = X_n 1(|X_n| \leq A)$ (truncations) for any $A > 0$. Then, $\sum_{n=1}^{\infty} X_n$ converges iff

    (a) $\sum_{k=1}^{\infty} P(X_n \neq Y_n) = \sum_{n=1}^{\infty} P(|X_n| > A) < \infty$.
    (b) $\sum_{n=1}^{\infty} E(Y_n)$ converges.
    (c) $\sum_{n=1}^{\infty}$ Var(Y_n) converges.

    For some $A > 0$.

53. *Kolmogorov's strong law of large numbers.*: Let $\{X_n\}$ be a sequence of iid random variables.

    If $E(|X_1|) < \infty, E(X_1) = \mu, \frac{s_n}{n} \xrightarrow{a.s.} \mu$.
    If $\frac{s_n}{n} \xrightarrow{a.s.} c, c \in \mathbb{R}, E(|X_1|) < \infty$ and $E(X_1) = c$.
    If $E(|X_1|) = \infty, P(\frac{|s_n|}{n} = \infty \text{i.o.}) = 1$.

54. *Kolmogorov-Feller weak law of large numbers*: Let $\{X_n\}$ be an iid sequence. Then

$$\frac{s_n - nE\{X_1 1(|X_1| \leq n)\}}{n} \xrightarrow{p} 0 \iff nP(|X_1| > n) \to 0$$

2.8 | Applications of LLN

55. *The empirical cumulative distribution function (CDF)* of a finite sequence of random variables $X_1, X_2, \ldots, X_n$ is defined as:

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} 1_{[X_i \leq n]}$$

where $I_{X_i \leq x}$ is the indicator function that takes the value 1 if $X_i \leq x$ and 0 otherwise.

Geometrically, the empirical CDF is a step function that starts from 0 at $-\infty$ and jumps up by $1/n$ at each observation value X_i. At the end of the sample, the empirical CDF takes the value 1 at $+\infty$.

56. * Glivenko-Cantelli theorem* For any distribution function F and any sample of n independent and identically distributed random variables $X_1, X_2, \ldots, X_n$ with distribution function F_n, the empirical distribution function F_n converges uniformly to F almost surely, i.e.,

$$\sup_{x\in\mathbb{R}} |F_n(x) - F(x)| \xrightarrow{\text{a.s.}} 0 \quad \text{as } n \to \infty,$$

where the symbol $\xrightarrow{\text{a.s.}}$ denotes almost sure convergence. This means that the probability of the event that F_n converges uniformly to F tends to 1 as n goes to infinity.

57. *Kolmogorov-Feller weak law of large numbers* Let X_n be an independent and identically distributed (iid) sequence of random variables with mean μ and let $s_n = \sum_{i=1}^{n} X_i$ be their partial sum. Then, the following are equivalent:

   (i) $\frac{s_n - nEX_1 1(|X_1| \leq n)}{n} \xrightarrow{p} 0$ as $n \to \infty$;

   (ii) $nP(|X_1| > n) \to 0$ as $n \to \infty$, where $P(|X_1| > n)$ is the tail probability of $|X_1|$.

58. Note that $nP(|X_1|) > n \to 0$ is weakening of $E[|X_1|] < \infty$

59. Let $\{X_n\}$ be a sequence of independent R.V. Let $\{b_n\}$ be a sequence of positive numbers with b_n increasing to ∞. Define $S_n = \sum_{i=1}^{n} X_i$. Set $Y_{i,n} = X_k 1_{|X_k \leq b_n|}$ and $\mu_n = \sum_{k=1}^{n} E[Y_{k,n}]$. If $\sum_{k=1}^{n} P(X_k \neq Y_{k,n}) \to 0$ then $(S_n - \mu_n)/b_n \to 0$ in probability. Further if $\mu_n/b_n \to 0$ then $S_n/b_n \to 0$ in probability.

3 | Convergence in distribution

60. *Convergence in distribution* is a type of convergence for sequences of random variables. Let $X_1, X_2, \ldots$ be a sequence of random variables and let $F_n(x)$ be their cumulative distribution function (CDF). We say that X_n converges in distribution to a random variable X, denoted $X_n \xrightarrow{d} X$ as $n \to \infty$, if $F_n(x)$ converges pointwise to the CDF $F(x)$ of X at all continuity points of $F(x)$, i.e.,

$$\lim_{n\to\infty} F_n(x) = F(x) \quad \text{for all } x \text{ such that } F(x) \text{ is continuous.}$$

61. Let $X_1, X_2, \ldots$ be a sequence of random variables that converges in distribution to a random variable X, denoted $X_n \xrightarrow{d} X$. Let $h : [a, b] \to \mathbb{R}$ be a continuous function where $a, b \in C_c(F)$. Then, we have

$$\lim_{n\to\infty} \mathbb{E}[h(X_n)] = \mathbb{E}[h(X)],$$

62. Let $X_1, X_2, \ldots$ be a sequence of random variables that converges in distribution to a random variable X, denoted $X_n \xrightarrow{d} X$. Let $h : \mathbb{R} \to \mathbb{R}$ be a bounded continuous function such that $\mathbb{E}[|h(X)|] < \infty$. Then, we have

$$\lim_{n\to\infty} \mathbb{E}[h(X_n)] = \mathbb{E}[h(X)],$$

63. Let $X_1, X_2, \ldots$ be a sequence of random variables with distribution functions $F_1, F_2, \ldots$, and let F be another distribution function. If $\mathbb{E}[h(X_n)] \to \mathbb{E}[h(X)]$ for all continuous and bounded functions h, then $F_n(x) \to F(x)$ for all continuity points x of F

64. Pointwise convergence of densities almost everywhere implies convergence in distribution. The converse is not true.

65. *Scheffe's theorem* Let $X, \{X_n\}$ be random variables with density functions f_X, f_{X_n}.

    (a) If they are continuous, $f_{X_n} \xrightarrow{a.e.} f_X \implies X_n \xrightarrow{d} X$.

    (b) If they are integer-valued, $P(X_n = k) \to P(X = k) \forall k \implies X_n \xrightarrow{d} X$.

66. *Tightness* In probability theory, a sequence of random variables X_n is said to be tight if, roughly speaking, the probability mass of the sequence is "concentrated" around a compact set. More precisely, for any $\epsilon > 0$, there exists $M_\epsilon \in (0, \infty)$ such that

$$\sup_{n\in\mathbb{N}} P(|X_n| > M_\epsilon) \leq \epsilon,$$

    or

$$\inf_{n\in\mathbb{N}} P(|X_n| \leq M_\epsilon) \geq 1 - \epsilon,$$

67. Convergence of distribution implies tightness

68.
$$\sup P(|X_n| > M) < \sup \frac{E[|X_n|]}{M}$$

69.
$$\sup P(|X_n| > M) < \sup \frac{E[|X_n|^2]}{M^2}$$

70. Boundedness of all $E[|X_n|]$ or $E[X_n^2]$ implies tightness

71. *Reletively compact* A set $\mathcal{F}$ of probability measures on a metric space (S, d) is said to be relatively compact, or precompact, if for any $\epsilon > 0$, there exists a compact set K such that

$$\sup_{P\in\mathcal{F}} P(S \setminus K) \leq \epsilon.$$

72. *Pohorov's theorem* Pohorov's theorem states that a family of probability measures on a metric space is tight if and only if it is relatively compact, i.e. for any subsequence X_{i_n} there exists a subsubsequence $X_{i_{j_n}}$ which converges to some R.V Y in distribution.

73. By pohorov's theorem, a sequence of R.V converges in distribution iff it is tight and all subsequential limits are same.

74. *Skorod Representation theorem* Let $\{X_n\}$ be a sequence of random variables such that $X_n \xrightarrow{d} X$. Then, there exist random variables $\{Y_n\}, Y$ defined on the Lebesgue measure on $[0,1]$ such that $Y_n = X_n, Y = X$ (up to distribution), and $Y_n \xrightarrow{a.s.} Y$.

75. $X_n \xrightarrow{d} X \implies E(|X|) \leq \text{liminf}_{n\to\infty} E(|X_n|)$

3.1 | Method of moments

76. Let $\{X_n\}$ be a sequence of R.V. such that $m_k = \lim_{n\to\infty} E[X_n^k]$ exits for all k. If $\{m_k\}$ uniquely determine a probability distribution X then $X_n \to X$ in distribution.

77. *Carleman's condition* For the Hamburger moment problem (the moment problem on the whole real line), the theorem states the following: Let μ be a measure on $\mathbb{R}$ such that all the moments

$$m_n = \int_{-\infty}^{+\infty} x^n d\mu(x), \quad n = 0, 1, 2, \cdots$$

are finite. If

$$\sum_{n=1}^{\infty} m_{2n}^{-\frac{1}{2n}} = +\infty$$

then the moment problem for (m_n) is determinate; that is, μ is the only measure on $\mathbb{R}$ with (m_n) as its sequence of moments.

4 | Characteristic functions and CLT

4.1 | Characteristic function

78. **Characteristic function(CF)** The characteristic function of a random variable X is a complex-valued function defined on the real line by

$$\phi_X(t) = \mathbb{E}[e^{itX}],$$

where i is the imaginary unit, t is a real-valued parameter, and $\mathbb{E}$ denotes the expected value operator. The characteristic function of X is essentially the Fourier transform of its probability distribution function.

79. The C.F always exists and is uniformly continuous

80. $|\phi_X(t)| \leq \phi_X(0) = 1$

81. $\overline{\phi_X(t)} = \phi_X(-t) = \phi_{-X}(t)$

82. For independent R.V

$$\phi_{\sum_{i=1}^n X_i}(t) = \prod_{i=1}^n \phi_{X_i}(t)$$

83. $\phi_{(X-a)/b} = e^{-ita/b}\phi_X(t/b)$

84. **Inversion theorem** If X is an R.V with cf ϕ_X and CDF F then

$$F(b) - F(a) + \frac{1}{2}\left(P(X=a) - P(X=b)\right) = \lim_{T\to\infty} \frac{1}{2\pi}\int_{-T}^{T} \frac{e^{-itb} - e^{-ita}}{it} dt$$

85. $\int_0^T \sin t/tdt \leq \pi \forall t$

86. $\lim_{T\to\infty} \int_0^T \sin t/tdt \to \pi/2$

87. *Uniqueness theorem* Let X and Y be two random variables with characteristic functions $\phi_X(t)$ and $\phi_Y(t)$, respectively. If $\phi_X(t) = \phi_Y(t)$ for all t, then X and Y have the same distribution.

88. If ϕ_X is L^1 then X admits a density given by

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\phi_X(t)dt$$

89. If X has a cts CDF F with density f then $\lim_{|t|\to\infty} |\phi_X(t)| = 0$

90. If $P(X=a) > 0$ then $P(X=a) = \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} e^{-ita}\phi_X(t)dt$

91. If ϕ_X has no imaginary part $X \equiv -X$ in distribution i.e distribution is symmetrical about 0.

92. If $E(|X|^n) < \infty$ for all $n \geq 1$ and $\frac{|t|^n}{n!}E(|X|^n) \to 0$ as $n \to \infty$ for all $t \in \mathbb{R}$, then $\phi_X(t) = 1 + \sum_{k=1}^{\infty} \frac{(it)^k}{k!}E(X^k)$.

93. If $E(|X|^n) < \infty$ for some $n \in \mathbb{N}$,

$$|\phi_X(t) - \sum_{k=0}^{n} \frac{(it)^k}{k!}E(X^k)| \leq E\{\min[\frac{2|tx|^n}{n!}, \frac{|tx|^{n+1}}{(n+1)!}]\}$$

In particular,

$$|\phi_X(t) - 1| \leq E[\min\{2, tX\}]$$

If $E[|X|]$ exists,

$$|\phi_X(t) - 1 - itE[X]| \leq E\left[\min\left(|Xt|, \frac{X^2t^2}{2}\right)\right]$$

94. If $E[|X^n|]$ is finite then k^{th} derivative of ϕ_X exists for $k \leq n$, they are uniformly continuous and

$$\phi_X^{(k)} = \int_{-\infty}^{\infty} (ix)^k e^{itx} dF(x)$$

$$\frac{d^k}{dx^k}E[itX] = \int_{-\infty}^{\infty} \left(\frac{d^k}{dx^k}e^{itx}\right) dF(x)$$

95. $\phi_X^k(0) = i^k E[X^k]$

96. $\phi_X(t) = 1 + \sum_{k=1}^{n} \frac{(it)^k}{k!} E[X^k] + o(|t|^n)$ as $t \to 0$.

97. If X is an R.V such that ϕ_X has a finite derivative of order $2n$ at $t = 0$ then $E(|X|^2 n) < \infty$ (93,94 holds)

4.2 | Multivariate CF

98. *CF of a random vector* The characteristic function of a random vector $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ is defined as $\varphi_{\mathbf{X}}(\mathbf{t}) = \mathbb{E}\left[e^{i\mathbf{t}^\top \mathbf{X}}\right]$, where $\mathbf{t} = (t_1, t_2, \ldots, t_n)^\top \in \mathbb{R}^n$ is a vector of real numbers and $\mathbf{t}^\top \mathbf{X} = t_1 X_1 + t_2 X_2 + \cdots + t_n X_n$.

99. If $t = (0, 0, ...t_j, 0, 0, 0)$ then $\phi_X(t) = \phi_{X_j}(t_j)$

100. If $t = (s, s, s, ..)$ then $\phi_X(t) = E[e^{is \sum X_j}]$

101. Uniqueness theorem holds for multivariate CF

102. If $\phi_X(t) = e^{i(a_1 t_1 + a_2 t_2 - b_1 t_1^2 - b_2 t_2^2 - b_3 t_1 t_2)}$ then $X \sim N(\mu, \sum)$ where $\mu = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}$ and $\sum = \begin{bmatrix} b_1 & b_3/2 \\ b_3/2 & b_2 \end{bmatrix}$

4.3 | CF and distributional convergence

103. For a sequence of R.V. $X_n \to X$ in distribution iff $\phi_{X_n} \to \phi_X$ pointwise.

104. For $h > 0$:

$$P(|X| > 2/h) \leq \frac{1}{h} \int_{|t|<h} (1 - \phi_X(t)) dt$$

4.4 | CF and weak convergence

105. Levy's continuity theorem states that for any sequence of random variables X_n with characteristic functions φ_n. if $\varphi_n \to \varphi$ and if φ is continuous at $t = 0$ then $X_n \to X$ in distribution where $\phi_X = \varphi$.

106. The Cramer-Wold device states that for a sequence of random vectors X_n. $X_n \to X$ in distribution iff $a^T X_n \to a^T X \forall a \in \mathbb{R}^d$

4.5 | Central limit theorem

107. *CLT* For iid $\{X_n\}$. if $0 < E[X_1^2] < \infty$ then

$$\sqrt{n}\frac{\overline{X_n} - \mu}{\sigma} \xrightarrow{d} N(0,1) \quad \overline{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

108. For complex $z_1, z_2 ...$ and $w_1, w_2 ....$ where $|z_i|, |w_i| < 1$. we have

$$\left|\prod_{j=1}^{n} z_j - \prod_{j=1}^{n} w_j\right| \leq \sum_{j=1}^{n} |z_j - w_j|$$

109. *Multivariate CLT* Let $\mathbf{X}nn \geq 1$ be a sequence of independent and identically distributed random vectors with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Define $\mathbf{X}n = \frac{1}{n}\sum i = 1^n \mathbf{X}_i$. Then. as n tends to infinity.

$$\sqrt{n}(\mathbf{X}_n - \boldsymbol{\mu}) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}),$$

where $\xrightarrow{d}$ denotes convergence in distribution and $\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$ is a multivariate normal distribution with mean vector $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}$.

4.6 | CLT for independent but not identical sequence of R.V

110. Define $s_n^2 = \sum_{i=1}^n \sigma_n^2$

111. Define $L_1(n) = \max_{1\leq i\leq n} \sigma_n^2/s_n^2$

112. For $\epsilon > 0$, define $L_2(n) = \frac{1}{s_n^2}\sum_{i=1}^n E[(X_i = \mu_i)^2 1_{|X_i-\mu_i|>\epsilon s_n}]$

113. (a) If $L_2(n) \to 0$ for each $\epsilon > 0$ then $L_1(n) \to 0$ and

$$\frac{1}{s_n}\sum_{i=1}^n (X_i - \mu_i) \xrightarrow{d} N(0,1)$$

(b) If $L_1(n) \to 0$ and $\frac{1}{s_n}\sum_{i=1}^n (X_i - \mu_i) \xrightarrow{d} N(0,1)$ then $L_2(n) \to 0$