Probability theory I Facts

Taeyoung Chang

Textbook : R. Durrett 「 Probability : Theory and Examples 」 5th edition

Last Update : April 29, 2022

Contents

1 Probability Space

* Sigma field $\mathcal{F}$ and event $A \in \mathcal{F}$
  - A family of subsets of Ω, named $\mathcal{F}$, is said to be a σ-field if
    (a) $\mathcal{F}$ contains Ω (b) $\mathcal{F}$ is closed under taking complement
    (c) $\mathcal{F}$ is closed under taking countable union.
    (d) $\mathcal{F}$ is closed under taking countable intersection. : by (a),(b),(c)
  - A σ-field $\mathcal{F}$ is usally called as an event space and
    an element $A \in \mathcal{F}$ is said to be an event.
* Sigma field $\sigma(\mathcal{A})$ generated by a collection $\mathcal{A}$
  - Given a collection $\mathcal{A}$ of subsets of Ω, the smallest σ-field containing $\mathcal{A}$ is said to be a σ-field generated by $\mathcal{A}$ and denoted as $\sigma(\mathcal{A})$
* Borel field $\mathcal{B}(\mathbb{R})$
  - Borel field $\mathcal{B}(\mathbb{R})$ is a σ-field on $\mathbb{R}$ generated by the family of all open subsets of $\mathbb{R}$.

- Various collections generating $\mathcal{B}(\mathbb{R})$
  i. Collections of open sets
  ii. Collections of bounded open intervals
  iii. Collections of bounded closed intervals
  iv. Collections of bounded half open intervals
  v. Collections of open rays
  vi. Collections of closed rays

* Measure μ and Probability Measure P
  - A set function $\mu : (\Omega, \mathcal{F}) \to \mathbb{R}$ is said to be a measure if
    (a) μ is nonnegative (b) μ is countably additive
    (c) $\mu(\phi) = 0$: by convention
  - Additionally, if $\mu(\Omega) = 1$, then it is called as probability measure and denoted as P instead of μ.

- Elementary properties of measure
  i. Monotonicity
  ii. Subadditivity
  iii. Continuity from above
  iv. Continuity from below

* π system $\mathcal{P}$ and λ system $\mathcal{L}$
  - A collection $\mathcal{P}$ of subsets of Ω is a π-system if $\mathcal{P}$ is closed under taking intersection.
  - A collection $\mathcal{L}$ of subsets of Ω is a λ-system if
    (a) $\mathcal{L}$ contains Ω (b) $\mathcal{L}$ is closed under taking complement
    (c) $\mathcal{L}$ is closed under taking countable disjoint union.

- A $\pi - \lambda$ system is a sigma field.

- $\pi - \lambda$ system thm
  - let $\mathcal{P}, \mathcal{L}$ be a π-system and λ-system on Ω respectively.
    If $\mathcal{P} \subset \mathcal{L}$ then $\sigma(\mathcal{P}) \subset \mathcal{L}$

- Checking two probability measures are the same
  - If P_1 and P_2 are two probability measures on the same event space and $P_1 = P_2$ on a π-system $\mathcal{P}$ then $P_1 = P_2$ on $\sigma(\mathcal{P})$
  - ☐ If two probability measures are the same on π-system generating a given event space then two probability measures are the same

2 Random Variable

* Random Variable X

- A function $X : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is said to be a random variable if
  $(X \in B) = X^{-1}(B) = \{\omega \in \Omega : X(\omega \in B)\} \in \mathcal{F} \quad \forall B \in \mathcal{B}(\mathbb{R})$

- Checking whether a mapping is a random variable is nearly same with checking whether a function is measurable.

- Elementary properties of random variable

  i. If X is a r.v. then $c + X$ and cX are r.v.'s for any real num. c

  ii. If X and Y are r.v.'s then $X + Y$ and XY are r.v.'s

  iii. If $\{X_n\}$ is a random seq. then $\inf X_n$, $\sup X_n$, $\liminf X_n$, $\limsup X_n$ are all r.v.'s

  iv. If X is a r.v. and f is Borel measurable then $f(X)$ is a r.v.

* Simple random variable

- A random variable X is called simple if X takes a finite number of values.

- If X is a nonnegative random variable then $\exists$ a seq of nonnegative simple random variables $\{X_n\}$ s.t. $X_n \nearrow X$

  - For each $n \in \mathbb{N}$, we can define X_n by

$$X_n = \sum_{k=1}^{n \cdot 2^n} \frac{k-1}{2^n} I\Big(\frac{k-1}{2^n} \le X < \frac{k}{2^n}\Big)$$

□ If X is a random varaible then $\exists$ a seq of simple random variables $\{X_n\}$ s.t. $X_n \to X$

* σ-field $\sigma(X)$ generated by random variable X

- $\sigma(X) = \{(X \in B) : B \in \mathcal{B}(\mathbb{R})\}$

* $\mathcal{G}$-measurable random variable

- For a sub σ-field $\mathcal{G} \subset \mathcal{F}$, a random variable X is said to be $\mathcal{G}$-measurable
  if $(X \in B) \in \mathcal{G} \quad \forall B \in \mathcal{B}(\mathbb{R})$. Denote it as $X \in \mathcal{G}$

- If X and Y are random variables and Y is $\sigma(X)$-measurable
  then $\exists$ a Borel function f s.t. $Y = f(X)$

3 Distributions

* Distribution function F

  – A function $F : \mathbb{R} \to \mathbb{R}$ is said to be a distribution function if
  (a) F is monotone increasing
  (b) F is right continuous (c) F has left limits.
  (d) $F(x) \to 1$ as $x \to \infty$ & $F(x) \to 0$ as $x \to -\infty$

* The inverse of distribution function $F^{-1}(u)$

  – let F be a distribution function. For each $u \in [0,1]$,
  $F^{-1}(u)$ is defined as $F^{-1}(u) = \inf\{x \in \mathbb{R} : F(x) \geq u\}$

• Properties of the inverse of distribution function

  i. $u \mapsto F^{-1}(u)$ is monotone increasing
  ii. $-\infty < F^{-1}(u) < \infty$ whenever $0 < u < 1$
  iii. $F^{-1}(0) = -\infty$ and $F^{-1}(1) = \infty$ or M for some $M < \infty$.
  If $F^{-1}(1) = M$ for some $M < \infty$ then X is bounded above by M a.s. where $X \sim F$
  iv. $F^{-1}(u) \leq x \Leftrightarrow u \leq F(x) \qquad \forall x \in \mathbb{R}, u \in [0,1]$
  $F(x) < u \Leftrightarrow x < F^{-1}(u) \qquad \forall x \in \mathbb{R}, u \in [0,1]$
  v. $u \leq F(F^{-1}(u))$ and $F(F^{-1}(u)-) \leq u \qquad \forall u \in [0,1]$
  vi. If F is continuous then $F(F^{-1}(u)) = u \qquad \forall u \in [0,1]$

☐ If a r.v. $X \sim F$ and $\mathcal{U} \sim unif[0,1]$ then $F^{-1}(\mathcal{U}) \overset{D}{=} X$

☐ If X is a continuous r.v. with $X \sim F$ where F is strictly increasing, then $F(X) \sim unif[0,1]$
(X is said to be continuous r.v. provided there is no point mass i.e. $P(X = x) = 0 \; \forall\, x \in \mathbb{R}$)

* Probability Borel measure μ

  – Any probability measure μ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is called as a probability Borel measure.

• 1-1 correspondence of distribution function and prob. Borel measure

  – For any distribution function F, $\exists$ a unique prob. Borel measure μ s.t.
  $\mu((-\infty, x]) = F(x) \qquad \forall x \in \mathbb{R}$

  – For any prob. Borel measure μ, $F : \mathbb{R} \to \mathbb{R}$ defined by
  $F(x) = \mu((-\infty, x])$ is a distribution function

• If a function F is monotone increasing and right-continuous satisfying $F(-\infty) = 0$ and $F(\infty) = 1$ then $\exists$ a probability space $(\Omega, \mathcal{F}, P)$ and a random varaible X s.t.
$P(X \leq x) = F(x) \; \forall x \in \mathbb{R}$ i.e. Given F is a distribution function for X

4 Expected Value and Independence

* Expected Value $E[X]$ / Integrability of a random variable X
  - Given prob. space $(\Omega, \mathcal{F}, P)$, $E[I_A] = \int I_A \, dP = P(A) \ \forall A \in \mathcal{F}$
  - For simple nonnegative random variable $X = \sum_{i=1}^{k} \alpha_i I_{A_i}$
    $E[X] = \int X \, dP = \sum_{i=1}^{k} \alpha_i P(A_i)$
  - For nonnegative random variable X,

$$\begin{aligned} E[X] &= \sup\{E[Z] : 0 \leq Z \leq X \quad simple\} \\ &= \lim_{n\to\infty} E[X_n] \quad \forall simple \ X_n \quad s.t. \ 0 \leq X_n \nearrow X \end{aligned}$$

  - For a random variable X,
    i. $E[X] = E[X^+] - E[X^-]$
    ii. X is called integrable if $E|X| < \infty$ or if $E[X^+], E[X^-] < \infty$

* Independence of events $\{A_n\}$ & collections of events $\{\mathcal{G}_n\}$
  - $A \perp\!\!\!\perp B \in \mathcal{F}$ if $P(A \cap B) = P(A)P(B)$
  - $A_1, \cdots, A_n$ independent if $P(A_{i_1} \cap \cdots \cap A_{i_k}) = P(A_{i_1}) \cdots P(A_{i_k})$
    $\forall 1 \leq i_1 < i_2 < \cdots < i_k \leq n$
  - $\{A_n\}$ independent if $A_1, \cdots, A_m$ are independent $\forall m \in \mathbb{N}$
  - For subcollections $\{\mathcal{G}_n\} \subset \mathcal{F}$, $\{\mathcal{G}_n\}$ are independent if $\{A_n\}$ are independent $\forall A_i \in \mathcal{G}_i$

• If $\mathcal{G}_1, \cdots, \mathcal{G}_n \subset \mathcal{F}$ are independent and each $\mathcal{G}_i$ is a π-system
  then $\sigma(\mathcal{G}_1), \cdots, \sigma(\mathcal{G}_n)$ are independent

* Independence of Random variables
  - R.V.'s $\{X_n\}$ are independent if $\{\sigma(X_n)\}$ are independent

• For a collection $\mathcal{C}$ of subsets of $\mathbb{R}$ and a r.v. X,
  $X^{-1}(\sigma(\mathcal{C})) = \sigma(X^{-1}(\mathcal{C}))$ where $X^{-1}(\mathcal{A}) = \{(X \in A) : A \in \mathcal{A}\}$ # 1.3.1

□ If $P(X_1 \leq x_1, \cdots, X_n \leq x_n) = P(X_1 \leq x_1) \cdots P(X_n \leq x_n) \quad \forall x_i \in \mathbb{R}$
  then $X_1, \cdots, X_n$ are independent

□ If $(X_1, \cdots, X_n)$ has a joint density $f(x_1, \cdots, x_n)$ and f can be written as $f(x) = g_1(x_1) \cdots g_n(x_n)$
  where g_k's are nonnegative and measurable, then $X_1, \cdots, X_n$ are independent # 2.1.1

□ If $X_1, \cdots, X_n$ are r.v.'s taking values in countable sets $C_1, \cdots, C_n$.
  Then $P(X_1 = x_1, \cdots X_n = x_n) = P(X_1 = x_1) \cdots P(X_n = x_n)$ whenever $\forall x_i \in C_i$
  implies that $X_1, \cdots, X_n$ are independent. # 2.1.2

• If X and Y are independent and f, g are Borel measurable functions,
  then $f(X)$ and $g(Y)$ are independent # 2.1.6

* limsup and liminf of seq of events. $\limsup A_n, \liminf A_n$
  - $\limsup A_n = \bigcap_{n=1}^{\infty} \bigcup_{k\geq n} A_k \quad \liminf A_n = \bigcup_{n=1}^{\infty} \bigcap_{k\geq n} A_k$

□ $\limsup A_n = A_n$ infinitely often $\quad \liminf A_n = A_n$ all but finitely many n's

☐ $(\limsup A_n)^C = \liminf A_n^C$

- Borel Cantelli lemma
  - For a seq. of events $\{A_n\}$, if $\sum_{n=1}^{\infty} P(A_n) < \infty$ then $P(A_n \ \ i.o.) = 0$
  - For a seq. of independent events $\{A_n\}$, if $\sum_{n=1}^{\infty} P(A_n) = \infty$ then $P(A_n \ \ i.o.) = 1$

☐ Given a seq. of independent events $\{A_n\}$,
$\sum_{n=1}^{\infty} P(A_n) < \infty \Leftrightarrow P(A_n \ \ i.o.) = 0$ and $\sum_{n=1}^{\infty} P(A_n) = \infty \Leftrightarrow P(A_n \ \ i.o.) = 1$
This is called as Borel Cantelli 0-1 law

☐ Note that $P(A_n \ \ i.o.) = 0$ implies $P(A_n^C$ all but finitely many n's$) = 1$

* Almost sure convergence $X_n \to X \ \ a.s.$
  - $\{X_n\}$ converges $a.s.$ if $P(C) = 1$ where $C = \{\omega \in \Omega : X_n(\omega) \ converges\}$
  - $X_n \to X \ \ a.s.$ if $P(X_n \to X) = 1$

- Classic results about interchanging limits and integrals(expectations)
  i. [Fatou's lemma] If $X_n \geq 0$ then $E[liminf X_n] \leq liminf E[X_n]$
  If $X_n \geq 0$ and $X_n \to X \ \ a.s.$ then $E[X] \leq liminf E[X_n]$
  ii. [MCT] If $0 \leq X_n \nearrow X \ \ a.s.$ then $E[X_n] \nearrow E[X]$
  If $X_n \nearrow X \ \ a.s.$ and $\exists$ a r.v. Y s.t. $Y \leq X_n \ \forall n \in \mathbb{N}$ and $E|Y| < \infty$
  then MCT $E[X_n] \nearrow E[X]$ also holds
  iii. [DCT] If $|X_n| \leq Y \ \ a.s. \ \ \forall n \in \mathbb{N}$ for some r.v. Y s.t. $E|Y| < \infty$ then $X_n \to X \ \ a.s.$ implies that $E[X_n] \to E[X]$
  iv. [BCT] If $|X_n| \leq B \ \ a.s. \ \ \forall n \in \mathbb{N}$ for some constant $B > 0$ then $X_n \to X \ \ a.s.$ implies that $E[X_n] \to E[X]$

- Almost sure convergence and convergence of expectation
  - $X_n \to X \ \ a.s.$ implies $E[h(X_n)] \to E[h(X)]$ if the following conditions for continuous g and h are satisfied.
    i. $g > 0$ (or $g \geq 0$ and $g(x) > 0$ unless $|x| \leq M$ for some $M > 0$)
    ii. $|h(x)|/g(x) \to 0$ as $|x| \to \infty$ (kind of "h is dominated by g")
    iii. $\exists M > 0$ s.t. $E[g(X_n)] \leq M \ \ \forall n \in \mathbb{N}$

☐ If $p > 1$ and $E|X_n|^p \leq M \ \ \forall n \in \mathbb{N}$ for some $M > 0$
then $X_n \to X \ \ a.s.$ implies $E[X_n] \to E[X]$

- Almost sure convergence is closed under continuous map # 1.3.3
  - If Borel measurable $f : \mathbb{R} \to \mathbb{R}$ is continuous then
  $X_n \to X \ \ a.s.$ implies $f(X_n) \to f(X) \ \ a.s.$

- Change of measure
  - For any Borel measurable function f and a r.v. X,
  if $f \geq 0$ or $E|f(X)| < \infty$ then $E[f(X)]$ is calculated by

$$E[f(X)] = \int_{\Omega} f(X) dP = \int_{\mathbb{R}} f d(PX^{-1})$$

☐ Change of measure and calculating probability

i. $P(X \in B) = E[I_B(X)] \;\; \forall B \in \mathcal{B}(\mathbb{R})$
ii. $P(f(X) \in A) = P(X \in f^{-1}(A)) \;\; \forall A \in \mathcal{B}(\mathbb{R})$
given f is nonnegative or integrable

* $\mathcal{L}_p$ space
  - $\mathcal{L}_0 =$ The class of all random variables on $(\Omega, \mathcal{F}, P)$
    $\mathcal{L}_p = \{X \in \mathcal{L}_0 : E|X|^p < \infty\}$ $(0 < p < \infty)$: normed vector space with $\|X\|_p = (E|X|^p)^{1/p}$

- Essential inequalities

  i. [Markov] $P(|X| \geq c) \leq \frac{1}{c}E|X| \;\; \forall c > 0$
  ii. [Hölder] Given $X \in \mathcal{L}_p$ and $Y \in \mathcal{L}_q$ with $p, q > 1$ and $\frac{1}{p} + \frac{1}{q} = 1$,
  $E|XY| = \|XY\|_1 \leq \|X\|_p \|Y\|_q$
  iii. [Cauchy-Schwarz] $(E|XY|)^2 \leq E[X^2]E[Y^2]$
  iv. [Jensen] If $\phi : \mathbb{R} \to \mathbb{R}$ convex then $\phi(E[X]) \leq E[\phi(X)]$ provided both expectations exist.
  If ϕ strictly convex then $\phi(E[X]) < E[\phi(X)]$ unless $X = E[X]$ $a.s.$ # 1.6.1

- If X is a nonnegative r.v. then $E[X] = \int_0^\infty P(X > t)dt$

- Product measures of independent random variables
  - If $X_1, \cdots, X_n$ are independent with distributions $X_i \sim \mu_i \; \forall i$, then a random vector $(X_1, \cdots X_n)$ has a distribution $\mu = \mu_1 \times \cdots \times \mu_n$

- Fubini theorem
  - Suppose X, Y are independent r.v.'s with distributions $X \sim \mu$, $Y \sim \nu$.
    If $f : \mathbb{R}^2 \to \mathbb{R}$ is Borel measurable with $f \geq 0$ or $E|f(X,Y)| < \infty$
    then $E[f(X,Y)] = \iint f(x,y)\, d\mu(x)d\nu(y) = \iint f(x,y)\, d\nu(y)d\mu(x)$

  ☐ Suppose X, Y are independent r.v's and $f,\, g : \mathbb{R} \to \mathbb{R}$ are Borel measurable functions.
  If $f, g \geq 0$ or $E|f(X)|,\, E|g(Y)| < \infty$ then $E[f(X)g(Y)] = E[f(X)]E[g(Y)]$

* Tail σ-field $\mathcal{T}$
  - The tail σ-field of events $\{A_n\}$ is $\mathcal{T} = \cap_{n=1}^\infty \sigma(A_n, A_{n+1}, \cdots)$

- Kolmogorov's 0-1 law
  - Suppose $\{A_n\}$ is a seq of independent events and $\mathcal{T}$ is the tail σ-field of $\{A_n\}$.
    If $A \in \mathcal{T}$ then $P(A) = 0$ or 1.

* Tail σ-field of random seq $\{X_n\}$
  - The tail σ-field of random seq. $\{X_n\}$ is $\mathcal{T} = \cap_{n=1}^\infty \sigma(X_n, X_{n+1}, \cdots)$
    $\surd$ $\sigma(X_n, X_{n+1}, \cdots) = \sigma(\{(X_i \in B) : B \in \mathcal{B}(\mathbb{R}), i = n, n+1, \cdots\})$

☐ If $\{X_n\}$ is a seq of independent r.v.'s and $\mathcal{C} = \{\omega \in \Omega : X_n(\omega) \;\; \textit{converges}\}$ then $\mathcal{C}$ lies in a tail σ-field of $\{X_n\}$ and $P(\mathcal{C}) = 0$ or 1

- Constructing independent random variables
  - Given a finite number of distribution functions, $F_1, \cdots, F_n$, it is possible to construct independent random variables $X_1, \cdots, X_n$ with $X_i \sim F_i$ for each $i = 1, \cdots, n$

5 Convergence in Probability

* Convergence in probability $X_n \xrightarrow{P} X$
  - $X_n \xrightarrow{P} X$ if $P(|X_n - X| \geq \epsilon) \to 0$ as $n \to \infty \quad \forall \epsilon > 0$
- Equivalent condition with almost sure convergence
  - $X_n \to X \;\; a.s. \Leftrightarrow \;\; \forall \epsilon > 0, \; P(\bigcup_{k \geq n}(|X_k - X| \geq \epsilon)) \to 0 \; as \;\; n \to \infty$
    $\Leftrightarrow \;\; \forall \epsilon > 0, \; P(|X_n - X| > \epsilon \;\; i.o.) = 0$
- Almost sure convergence is stronger than convergence in probability
  - $X_n \to X \;\; a.s.$ implies $X_n \xrightarrow{P} X$
  - Converse does not hold.
    (Counterexample) $X_n \sim \text{Bern}(\frac{1}{n}) \quad \forall \, n \in \mathbb{N} \;\Rightarrow X_n \xrightarrow{P} 0$ but $X_n \to 0 \;\; a.s.$ does not hold.
- The limit is unique both for in almost sure sense or in probability sense
  - $X_n \to X \;\; a.s.$ and $X_n \to Y \;\; a.s.$ then $X = Y \;\; a.s.$
  - $X_n \xrightarrow{P} X$ and $X_n \xrightarrow{P} Y$ then $X = Y \;\; a.s.$
- Convergence in probability is closed under continuous map
  - If Borel measurable $f : \mathbb{R} \to \mathbb{R}$ is continuous then $X_n \xrightarrow{P} X$ implies $f(X_n) \xrightarrow{P} f(X)$

☐ If $X_n \xrightarrow{P} X$ and $Y_n \xrightarrow{P} Y$ then $X_n + Y_n \xrightarrow{P} X + Y$ and $X_n Y_n \xrightarrow{P} XY$

- Convergence in $\mathcal{L}_p$ is stronger than convergence in probability
  - If $E|X_n - X|^p \to 0$ for some $p \geq 1$ then $X_n \xrightarrow{P} X$
- If $X_n \xrightarrow{P} X$ then $\exists$ a subseq $\{X_{n_k}\}$ of $\{X_n\}$ s.t. $\{X_{n_k}\} \to X \;\; a.s.$
- Lemma about convergence of real seq
  - For real seq. $\{x_n\}$, if every subseq. of $\{x_n\}$ has a further subseq. converging to x, then $\{x_n\}$ converges to x
- $X_n \xrightarrow{P} X$ if for every subseq. $\{X_{n_m}\}$ of $\{X_n\}, \exists$ a further subseq. $\{X_{n_{m_k}}\}$ s.t. $X_{n_{m_k}} \to X \;\; a.s.$
- Revisiting classic convergence thm
  1. [Fatou's lemma] If $X_n \geq 0$ and $X_n \xrightarrow{P} X$ then $E[X] \leq liminf E[X_n]$
  2. [MCT] If $0 \leq X_n$ increasing and $X_n \xrightarrow{P} X$ then $E[X_n] \nearrow E[X]$
  3. [DCT] If $|X_n| \leq Y \;\; a.s. \;\; \forall n \in \mathbb{N}$ for some r.v. Y s.t. $E|Y| < \infty$ then $X_n \xrightarrow{P} X$ implies that $E[X_n] \to E[X]$
- If $X_n \xrightarrow{P} X$ and f is continuous and bounded
  then not only $f(X_n) \xrightarrow{P} f(X)$ but also $E[f(X_n)] \to E[f(X)]$

6 Convergence in Distribution

* Sub probability Borel measure
  - a prob. Borel measure μ s.t. $\mu(\mathbb{R}) \leq 1$
* Weak convergence of sub prob. Borel measures $\mu_n \xrightarrow{w} \mu$
  - For sub prob. Borel measures $\{\mu_n\}$ and μ, $\quad \mu_n \xrightarrow{w} \mu$ if
    $\exists$ a dense $D \subset \mathbb{R}$ s.t. $\mu_n(a,b] \to \mu(a,b]$ as $n \to \infty$ $\forall a,b \in D$
- Lemma about countable set and dense set
  - If $D \subset \mathbb{R}$ and D^C is countable then D is dense in $\mathbb{R}$

□ Point mass set of a finite measure is at most countable.
  - If μ is a measure on measurable space $(\mathcal{S}, \mathcal{A})$ with $\mu(\mathcal{S}) < \infty$ then
    $E = \{x \in S : \mu(\{x\}) > 0\}$ is at most countable.

- Equivalent condition of weak convergence
  - $\mu_n \xrightarrow{w} \mu \Leftrightarrow \quad \mu_n(a,b] \to \mu(a,b]$ whenever $\mu(\{a\}) = \mu(\{b\}) = 0$
- The limit is unique in weak convergence sense
  - If $\mu_n \xrightarrow{w} \mu$ and $\mu_n \xrightarrow{w} \nu$ then $\mu = \nu$ i.e. $\mu(B) = \nu(B)$ $\forall B \in \mathcal{B}(\mathbb{R})$

* Weak convergence of distribution functions $F_n \Rightarrow F$
  - $F_n \Rightarrow F$ if $\mu_n \xrightarrow{w} \mu$ where $\mu_n \sim F_n$ and $\mu \sim F$

- Continuity set C_F of a distribution function F is dense in $\mathbb{R}$
- $F_n \Rightarrow F \Leftrightarrow F_n(x) \to F(x)$ $\forall x \in C_F$

* Convergence in distribution $X_n \xrightarrow{D} X$
  - $X_n \xrightarrow{D} X$ if $\mu_n \xrightarrow{w} \mu$ where $X_n \sim \mu_n$ and $X \sim \mu$

- Convergence in probability is stronger than convergence in distribution
  - $X_n \xrightarrow{P} X$ implies $X_n \xrightarrow{D} X$
- For a constant $c \in \mathbb{R}$, if $X_n \xrightarrow{D} c$ then $X_n \xrightarrow{P} c$
- Slutsky's thm
  i. If $X_n \xrightarrow{D} X$ and $Y_n \xrightarrow{P} c$ where c is a constant, then $X_n + Y_n \xrightarrow{D} X + c$ # 3.2.13
     Especially, if $X_n \xrightarrow{D} X$ and $Z_n - X_n \xrightarrow{P} 0$ then $Z_n \xrightarrow{D} X$
  ii. If $X_n \xrightarrow{D} X$ and $\delta_n \xrightarrow{P} 0$ then $X_n \delta_n \xrightarrow{P} 0$
  iii. If $X_n \xrightarrow{D} X$ and $Y_n \xrightarrow{P} c$ where c is a constant, then $X_n Y_n \xrightarrow{D} cX$ # 3.2.14

- Scheffe's thm
  - Sps $\{X_n\}$ and X have density functions $\{f_n\}$ and f respectively. If $f_n \to f$ $\mu - a.e.$ where μ is Lebesgue measure, then $X_n \xrightarrow{D} X$
- Skorohod's thm
  - Suppose that $\{\mu_n\}$ and μ are prob. Borel measures s.t. $\mu_n \xrightarrow{w} \mu$. Then $\exists$ a prob. space $(\Omega', \mathcal{F}', P')$ and r.v $\{X_n'\}$ and X' s.t. $X_n' \sim \mu_n$, $X' \sim \mu$ and $X_n' \to X'$ $P' - a.s.$
- Continuous mapping thm
  - If Borel measurable $f : \mathbb{R} \to \mathbb{R}$ satisfies $P(X \in D_f) = 0$ where D_f is discontinuity set of f then $X_n \xrightarrow{D} X$ implies $f(X_n) \xrightarrow{D} f(X)$
- If $X_n \xrightarrow{D} X$ then $E[g(X_n)] \to E[g(X)]$ $\forall$ continuous bounded g
- If $E[g(X_n)] \to E[g(X)]$ $\forall$ unif. continuous(or Lipschitz) bounded g then $X_n \xrightarrow{D} X$
- $X_n \xrightarrow{D} X$ if for every subseq. $\{X_{n_m}\}$ of $\{X_n\}$, $\exists$ a further subseq. $\{X_{n_{m_k}}\}$ s.t. $X_{n_{m_k}} \xrightarrow{D} X$
- The Portmanteau thm
  - The followings are equivalent
    1. $X_n \xrightarrow{D} X$
    2. $\liminf P(X_n \in G) \geq P(X \in G)$ $\forall$ open $G \subset \mathbb{R}$
    3. $\limsup P(X_n \in F) \leq P(X \in F)$ $\forall$ closed $F \subset \mathbb{R}$
    4. $P(X_n \in A) \to P(X \in A)$ $\forall A \in \mathcal{B}(\mathbb{R})$ with $P(X \in \partial A) = 0$
- Polya's thm   # 3.2.9
  - If $F_n \Rightarrow F$ and F is continuous then $F_n \to F$ uniformly on $\mathbb{R}$
- $\{X_n\}$ and X are integer valued random variables.
  Then $X_n \xrightarrow{D} X$ iff $P(X_n = m) \to P(X = m)$ $\forall m \in \mathbb{Z}$ # 3.2.12

* Big O_p and small o_p notation
  - $X_n = o_p(1)$ if $X_n \xrightarrow{P} 0$
  - $X_n = O_p(1)$ if $\lim_{M\to\infty} \sup_n P(|X_n| > M) = 0$
    or equivalently $\forall \epsilon > 0$, $\exists M_\epsilon \& N_\epsilon$ s.t. $P(|X_n| > M_\epsilon) < \epsilon$ $\forall n \geq N_\epsilon$
    $O_p(1)$ is also called as 'stochasticallly bounded'

- Elementary properties of Big O_p and small o_p
  1. $X_n = o_p(1), Y_n = o_p(1) \Rightarrow X_n + Y_n = o_p(1), X_nY_n = o_p(1)$
  2. $X_n = O_p(1), Y_n = O_p(1) \Rightarrow X_n + Y_n = O_p(1), X_nY_n = O_p(1)$
  3. $X_n = O_p(1), Y_n = o_p(1) \Rightarrow X_n + Y_n = O_p(1), X_nY_n = o_p(1)$
  4. $X_n \xrightarrow{D} X \Rightarrow X_n = O_p(1)$

- Helly's selection principle
  - For a seq. $\{F_n\}$ of distribution functions, $\exists$ a subseq. $\{F_{n_k}\}$ and a distribution-like func. F s.t. $F_{n_k}(x) \to F(x)$ as $k \to \infty$ $\forall x \in C_F$

* Tightness of seq. of distribution functions
  - A seq. of distribution functions $\{F_n\}$ is called tight if $\forall \epsilon > 0$, $\exists M_\epsilon > 0$ s.t. $limsup_n\, 1 - F_n(M_\epsilon) + F_n(-M_\epsilon) \le \epsilon$

- For a seq. of distribution functions $\{F_n\}$, every subsequential limit is a distribution function iff $\{F_n\}$ is tight

7 Random Series

- Komogorov's inequality
  - $\{X_n\}$: a seq. of indep. r.v.'s with mean zero and finite variance.

$$\forall \epsilon > 0,\ \ P(\max_{1\le k\le n} |S_k| \ge \epsilon) \le \frac{1}{\epsilon^2}\sum_{k=1}^{n}\sigma_k^2$$

- Convergence of random series
  - $\{X_n\}$: a seq. of indep. r.v.'s with mean zero and finite variance.

$$\sum_{n=1}^{\infty}\sigma_n^2 < \infty \Rightarrow \sum_{n=1}^{\infty} X_n\ \ converges\ \ a.s.$$

- Etamadi's inequality
  - $\{X_n\}$: a seq. of independent r.v.'s.

$$\forall \epsilon > 0,\ \ P(\max_{1\le k\le n} |S_k| \ge 3\epsilon) \le 3\max_{1\le k\le n} P(|S_k| \ge \epsilon)$$

- Levy's thm
  - $\{X_n\}$: a seq. of independent r.v.'s. If $S_n \xrightarrow{P} S$ then $S_n \to S$ a.s.

- Lemma for Kolmogorovs's three series thm
  - $\{X_n\}$: a seq. of independent r.v.'s. If $|X_n - E(X_n)| \le A\ \ a.s.$ for some $A > 0$ $\forall n \in \mathbb{N}$, then

$$\forall \epsilon > 0,\ \ P(\max_{1\le k\le n} |S_k| \le \epsilon) \le \frac{(2A + 4\epsilon)^2}{Var(S_n)}$$

* Eventual equivalence of random sequences $\{X_n\} \sim \{Y_n\}$
  - Random seq. $\{X_n\}$ and $\{Y_n\}$ are said to be (eventually) equivalent if $\sum_{n=1}^{\infty} P(X_n \ne Y_n) < \infty$. Denote it as $\{X_n\} \sim \{Y_n\}$

- ☐ If $\{X_n\} \sim \{Y_n\}$ then $P(X_n = Y_n$ all but finitely many n's$) = 1$
- Kolmogorov's three series thm
  - $\{X_n\}$: a seq. of independent r.v.'s.
    For $A > 0$, $\{Y_n\}$ is defined by $Y_n = X_n I(|X_n| \le A)$ Then $\sum_n X_n$ converges a.s.
    $\Leftrightarrow$ (a)$\sum_n P(|X_n| > A) < \infty$ (b)$\sum_n E(Y_n)$ converges. (c) $\sum_n Var(Y_n) < \infty$

8 Law of Large Numbers

- Lemma about eventually equivalent random sequences
  - If random seq $\{X_n\} \sim \{Y_n\}$ and real seq. $\{a_n\}$ s.t. $0 < a_n \to \infty$
    then for a random variable Z,
    i. $\frac{1}{a_n}\sum_{j=1}^n X_j \to Z \ a.s. \Leftrightarrow \frac{1}{a_n}\sum_{j=1}^n Y_j \to Z \ a.s.$
    ii. $\frac{1}{a_n}\sum_{j=1}^n X_j \xrightarrow{P} Z \Leftrightarrow \frac{1}{a_n}\sum_{j=1}^n Y_j \xrightarrow{P} Z$
- Equivalent condition for integrability
  - $\sum_n P(|X| \ge n) \le E|X| \le 1 + \sum_n P(|X| \ge n)$
  - $E|X| < \infty \Leftrightarrow \sum_n P(|X| \ge n) < \infty$
- Weak Law of Large numbers ; W.L.L.N
  - $\{X_n\}$ i.i.d. random seq. with $E|X_1| < \infty$, $E(X_1) = \mu$ Then $\frac{1}{n}\sum_{j=1}^n X_j \xrightarrow{P} \mu$
- Lemmas about convergence of real series
  - [Cesàro mean] If $x_n \to x$ then $\frac{1}{n}\sum_{j=1}^n x_j \to x$
  - [Kronecker's lemma] $0 < a_n \nearrow \infty$. If $\sum_{n=1}^\infty \frac{1}{a_n} x_n$ converges then $\frac{1}{a_n}\sum_{j=1}^n x_j \to 0$
- Strong Law of Large numbers ; S.L.L.N
  - $\{X_n\}$ i.i.d. random seq. with $E|X_1| < \infty$, $E(X_1) = \mu$. Then $\frac{1}{n}\sum_{j=1}^n X_j \to \mu \ a.s.$
- Strong Law holds even if the mean is infinite
  - $\{X_n\}$ i.i.d. random seq. with $E(X_1^+) = \infty$. $E(X_1^-) < \infty$.
    Then $\frac{1}{n}\sum_{j=1}^n X_j \to \infty \ a.s.$
  - $\{X_n\}$ i.i.d. random seq. with $E(X_1^+) < \infty$. $E(X_1^-) = \infty$.
    Then $\frac{1}{n}\sum_{j=1}^n X_j \to -\infty \ a.s.$
- If $\{X_n\}$ i.i.d. random seq. with $E|X_1| = \infty$ then
  (a)$P(|X_n| \ge n \ i.o) = 1$ (b)$P(\lim_{n\to\infty} \frac{1}{n} S_n \ exists \ and \ finite) = 0$
- Glivenko Cantelli thm
  - $\{X_n\}$ i.i.d. random seq. $X_1 \sim F$. Empirical distribution function is defined as $F_n(x) = \frac{1}{n}\sum_{i=1}^n I(X_i \le x)$ $\forall x \in \mathbb{R}$ for each $n \in \mathbb{N}$
    Then $F_n \rightrightarrows F \ a.s.$ i.e. $sup_{x\in\mathbb{R}}|F_n(x) - F(x)| \to 0 \ a.s.$

- $\{X_n\}$ i.i.d. random seq. with $E|X_1| = \infty$.
  $\{a_n\}$ positive real seq s.t. a_n/n is monotone increasing. Then
  $limsup_n \frac{|S_n|}{a_n} = \begin{cases} 0 \quad a.s. & if \ \sum_n P(|X_1| \geq a_n) < \infty \\ \infty \quad a.s. & if \ \sum_n P(|X_1| \geq a_n) = \infty \end{cases}$

- Convergence rate of random series

  - $\{X_n\}$ i.i.d. random seq. with mean zero and finite variance. Then

$$\forall \epsilon > 0, \ \frac{S_n}{\sqrt{n}(\log n)^{\frac{1}{2}+\epsilon}} \to 0 \quad a.s.$$

- The law of iterated logarithm(L.I.L)

  - $\{X_n\}$ i.i.d. random seq. with mean zero and finite variance σ^2. Then

$$\begin{aligned} &\limsup_n \frac{S_n}{\sqrt{2n \log \log n \sigma}} = 1 \ a.s. \\ &\liminf_n \frac{S_n}{\sqrt{2n \log \log n \sigma}} = -1 \ a.s. \\ &\limsup_n \frac{|S_n|}{\sqrt{2n \log \log n \sigma}} = 1 \ a.s. \\ \forall \, \varepsilon > 0, \quad &P(S_n \geq (1+\varepsilon)\sigma\sqrt{2n \log \log n} \ i.o.) = 0 \\ &P(S_n \leq -(1+\varepsilon)\sigma\sqrt{2n \log \log n} \ i.o.) = 0 \end{aligned}$$

9 Characteristic function

* Characteristic function
  - A char. func. ψ corresponding to a prob. Borel measure μ is

$$\begin{aligned}\psi(t) &= \int e^{itx}\,d\mu(x)\\ &= \int \cos tx\,d\mu(x) + i\int \sin tx\,d\mu(x)\end{aligned}$$

  - A char. func. of a r.v. $X \sim \mu$ is $\psi_X(t) = E[e^{itX}] = \int e^{itx}\,d\mu(x)$

- Elementary properties of characteristic functions
  i. $\psi(0) = 1$
  ii. $|\psi(t)| \leq 1 \;\; \forall t \in \mathbb{R}$
  iii. $\sup_{t\in\mathbb{R}} |\psi(t+h) - \psi(t)| \to 0$ as $h \to 0$. ψ is uniformly continuous
  iv. $\psi_{aX+b}(t) = e^{itb}\psi_X(at) \;\; \forall a, b, \in \mathbb{R}$
  v. $\psi(-t) = \overline{\psi}(t)$. If $X \sim \psi$ then $-X \sim \overline{\psi}$

- Additional properties of characteristic functions
  i. If $X_1 \perp\!\!\!\perp X_2$ and $X_1 \sim \psi_1, X_2 \sim \psi_2$ then $X_1 + X_2 \sim \psi_1\psi_2$
  If $X_1, \cdots, X_n$ independent with $X_i \sim \psi_i$ then $\sum_{i=1}^n X_i \sim \prod_{i=1}^n \psi_i$
  ii. If $\psi_1, \cdots \psi_n$ are char. func.s then $\prod_{i=1}^n \psi_i$ is also a char. function.
  iii. If $\psi_1, \cdots \psi_n$ are char. func.s with $\psi_i \sim \mu_i$ then $\sum_{i=1}^n \lambda_i\psi_i$ is also a char. function given $\lambda_i > 0$ and $\sum_{i=1}^n \lambda_i = 1$
  iv. If ψ is a char. func then $Re(\psi)$ and $|\psi|^2$ are also char. functions.

☐ Characteristic function of $X \sim N(0,1)$ is $exp(-\frac{1}{2}t^2)$ $X \sim N(\mu, \sigma)$ is $exp(i\mu t - \frac{1}{2}\sigma^2 t^2)$
Characteristic function of $X \sim Poi(\lambda)$ is $exp(\lambda(e^{it} - 1))$

- The Inversion formula
  - If ψ is a char. func. corresponding to a distribution μ, then whenever $a < b$, the equation below holds true.

$$\frac{1}{2}\{\mu(a,b] + \mu[a,b)\} = \lim_{T\to\infty} \frac{1}{2\pi}\int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it}\psi(t)\,dt$$

- If two distributions μ and ν corresponds to same characteristic function, then μ and ν are indeed the same distributions. In this sense, characteristic functions uniquely determine the distribution.

- Suppose ψ is a char. func. corresponding to $\mu \sim F$.
  If $\int_{-\infty}^{\infty} |\psi(t)|dt < \infty$ then F is diff.able and the density f is derived as

$$f(x) = F'(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\psi(t)\,dt$$

- Suppose ψ is a char. func. corresponding to μ. For each $x_0 \in \mathbb{R}$,

$$\mu(\{x_0\}) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} e^{-itx_0}\psi(t)\,dt$$

- Suppose ψ is a char. func. corresponding to μ.

$$\sum_x [\mu(\{x\})]^2 = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} |\psi(t)|^2\,dt$$

- Suppose ψ is a char. func. corresponding to $\mu \sim F$. If F is diff.able with density f then $\lim_{t\to\infty}\psi(t) = 0$

☐ If a distribution μ has a density f where $\mu \sim F$ and $f = F'$ then μ has no point mass. i.e. $\mu(\{x\}) = 0\ \forall x \in \mathbb{R}$ # 3.3.3

- Lemma for continuity thm
  - μ : a sub prob. Borel measure. $\psi(t) = \int e^{itx} d\mu(x)$. Then $\forall \delta > 0$,

$$\mu[-\delta.\delta] \geq \frac{\delta}{2}\left|\int_{-2/\delta}^{2/\delta} \psi(t)\,dt\right| - 1$$

- Levy's Continuity thm
  - let $\{\mu_n\}$ be a seq. of distributions and $\mu_n \sim \psi_n\ \forall n \in \mathbb{N}$.
    If $\psi_n \to \psi$ pointwise and ψ is continuous at zero,
    then ψ is a char. func. correspoding to μ and $\mu_n \xrightarrow{w} \mu$
  - let $\{\mu_n\}$ be a seq. of distributions and $\mu_n \sim \psi_n\ \forall n \in \mathbb{N}$.
    If $\mu_n \xrightarrow{w} \mu$ and $\mu \sim \psi$ then $\psi_n \to \psi$ pointwise.

- Suppose X, Y i.i.d. with mean zero, variance one.
  If $X + Y \perp\!\!\!\perp X - Y$ then X, Y are normal r.v.'s.

- Suppose $\{X_n\}$ and $\{Y_n\}$ are independent and $X \perp\!\!\!\perp Y$.
  If $X_n \xrightarrow{D} X$ and $Y_n \xrightarrow{D} Y$ then $X_n + Y_n \xrightarrow{D} X + Y$ # 3.3.8

10 Central Limit Theorem

- Lemma about convergence of sequence and exponential
  i. $(1 + \frac{a_n}{n})^n \to e^a$ if $a_n \to a$ $\quad (1 + c_n)^n \to e^c$ if $nc_n \to c$
  ii. $(1 + \frac{a_n}{\lambda_n})^{\lambda_n} \to e^a$ if $a_n \to a,\ 0 < \lambda_n \nearrow \infty$
  $(1 + c_n)^{\lambda_n} \to e^c$ if $\lambda_n c_n \to c,\ 0 < \lambda_n \nearrow \infty$

- Lemma about taylor expansion error term of e^{ix}

$$\left|e^{ix} - \sum_{k=0}^{n} \frac{(ix)^k}{k!}\right| \leq min\left\{\frac{|x|^{n+1}}{(n+1)!}, \frac{2|x|^n}{n!}\right\} \quad \forall x \in \mathbb{R}\ \forall n \in \mathbb{N}$$

- Lemma about taylor expansion of characteristic function

  i. If $E|X|^n < \infty$ then

  $$\left| E\left[e^{itx} - \sum_{k=0}^{n} \frac{(itx)^k}{k!} \right] \right| \le |t|^n E[min\{|t||X|^{n+1}, 2|X|^n\}]$$

  $$\psi(t) = \sum_{k=0}^{n} \frac{i^k E(X^k)}{k!} t^k + o(|t|^n) \quad as \ t \to 0$$

  ii. Especially if $E[X] = 0$ and $E[X^2] = \sigma^2 < \infty$, then

  $$\psi(t) = 1 - \frac{\sigma^2}{2} t^2 + o(t^2) \quad as \ t \to 0$$

- Central Limit Thm

  – $\{X_n\}$ i.i.d. random seq. $E[X_1] = \mu, Var(X_1) = \sigma^2 < \infty$. Then

  $$\frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{D} N(0,1) \quad i.e. \quad \frac{S_n - E[S_n]}{\sqrt{Var(S_n)}} \xrightarrow{D} N(0,1)$$

* Lindberg's condition

  – $\{X_{nk} : k = 1, \cdots, r_n\}$: row-wise independent double array of r.v.'s with mean zero and finite variance. $\{X_{nk}\}$ is said to be satisfying Lindberg's condition if

  $$\epsilon > 0, \quad \lim_{n\to\infty} \frac{1}{\mathcal{D}_n^2} \sum_{k=1}^{r_n} \int_{(|X_{nk}| \ge \epsilon \mathcal{D}_n)} X_{nk}^2 \, dP = 0$$

  where $S_n = X_{n1} + \cdots + X_{nr_n}$ and $\mathcal{D}_n^2 = Var(S_n) = \sigma_{n1}^2 + \cdots + \sigma_{nr_n}^2$

- Lemma about complex numbers

  i. $|\prod_{i=1}^n z_i - \prod_{i=1}^n w_i| \le \sum_{i=1}^n |z_i - w_i|$ if $|z_i|, |w_i| \le 1$

  ii. $|e^z - (1+z)| \le \frac{1}{2} e^c |z|^2$ whenever $|z| \le c$
  Especially, $|e^z - (1+z)| \le |z|^2$ if $|z| \le 1/2$

  iii. $|e^z| = e^{Re(z)} \le e^{|z|} \quad \forall z \in \mathbb{C}$

  iv. For $\{z_n\} \in \mathbb{C}$, $e^{z_n} \to e^z \Rightarrow e^{Re(z_n)} \to e^{Re(z)}$

- Feller's Thm

  – $\{X_{nk} : k = 1, \cdots, r_n\}$: row-wise independent double array of r.v.'s with mean zero and finite variance. $S_n = X_{n1} + \cdots + X_{nr_n}$ and $\mathcal{D}_n^2 = Var(S_n) = \sigma_{n1}^2 + \cdots + \sigma_{nr_n}^2$
  Lindberg's condition is satisfied if and only if
  (a) $S_n/\mathcal{D}_n \xrightarrow{D} N(0,1)$ (b) $\frac{1}{\mathcal{D}_n^2} \max_{1\le k\le r_n} \sigma_{nk}^2 \to 0$ as $n \to \infty$

* Lyapunov condition

- $\{X_{nk} : k = 1, \cdots, r_n\}$: row-wise independent double array of r.v.'s with mean zero. $\{X_{nk}\}$ is said to be satisfying Lyapunov condition if $\exists\, \delta > 0$ s.t.
  i. $E|X_{nk}|^{2+\delta} < \infty$
  ii. $\lim_{n\to\infty} \frac{1}{\mathcal{D}_n^{2+\delta}} \sum_{k=1}^{r_n} E|X_{nk}|^{2+\delta} = 0$

- Lyapunov condition is stronger than Lindberg's condition

- Poisson approximation of binomial random variable

  i. $\{X_{nk} : k = 1, \cdots, n\}$: row-wise independent double array of Bernoulli r.v.'s.
  $p_{nk} = P(X_{nk} = 1)$. If (a) $\sum_{k=1}^{n} p_{nk} \to \lambda$ where $\lambda \in (0, \infty)$ and (b) $\max_{1\le k\le n} p_{nk} \to 0$
  then $S_n = X_{n1} + \cdots + X_{nn} \xrightarrow{D} Poi(\lambda)$

  ii. $\{X_{nk} : k = 1, \cdots, n\}$: row-wise independent double array of r.v.'s having nonnegative integer values. $p_{nk} = P(X_{nk} = 1)$ and $\epsilon_{nk} = P(X_{nk} \ge 2)$
  If (a) $\sum_{k=1}^{n} p_{nk} \to \lambda$ where $\lambda \in (0, \infty)$ (b) $\max_{1\le k\le n} p_{nk} \to 0$
  and (c) $\sum_{k=1}^{n} \epsilon_{nk} \to 0$ (which means X_{nk} is nearly Bernoulli r.v.)
  then $S_n = X_{n1} + \cdots + X_{nn} \xrightarrow{D} Poi(\lambda)$