

Advanced Theory of Probability

Lectured by Xinyi Li

L^AT_EXed by Chengxin Gong

2022 年 11 月 3 日

目录

1	Measure Theory	2
2	Laws of Large Numbers	2
2.1	Independence	2
2.2	Weak Laws of Large Numbers	3
2.3	Borel-Cantelli Lemmas	3
2.4	Strong Law of Large Numbers	4
2.5	Convergence of Random Series	4
2.6	Large Deviations	5
2.7	Percolation	7
3	Central Limit Theorems	7
3.1	The De Moivre-Laplace Theorem	7
3.2	Weak Convergence	8
3.3	Characteristic Functions	8
3.4	Central Limit Theorems	9

1 Measure Theory

- Fatou's lemma: If $f_n \geq 0$ then $\liminf_{n \rightarrow \infty} \int f_n d\mu \geq \int \liminf_{n \rightarrow \infty} f_n d\mu$.
- Monotone convergence theorem: If $f_n \geq 0$ and $f_n \uparrow f$ then $\int f_n d\mu \uparrow \int f d\mu$.
- Dominated convergence theorem: If $f_n \rightarrow f$ a.e., $|f_n| \leq g$ for all n , and g is integrable, then $\int f_n d\mu \rightarrow \int f d\mu$.
- Suppose $X_n \rightarrow X$ a.s. Let g, h be continuous functions with (i) $g \geq 0$ and $g(x) \rightarrow \infty$ as $|x| \rightarrow \infty$; (ii) $|h(x)|/g(x) \rightarrow 0$ as $|x| \rightarrow \infty$; (iii) $\mathbb{E}g(X_n) \leq K < \infty$ for all n . Then $\mathbb{E}h(X_n) \rightarrow \mathbb{E}h(X)$.
- Fubini's theorem: If $f \geq 0$ or $\int |f| d\mu < \infty$, then $\int_X \int_Y f(x, y) \mu_2(dy) \mu_1(dx) = \int_{X \times Y} f d\mu = \int_Y \int_X f(x, y) \mu_1(dx) \mu_2(dy)$.

2 Laws of Large Numbers

2.1 Independence

- Two events A and B are independent if $P(A \cap B) = P(A)P(B)$. Two random variables X and Y are independent if for all $C, D \in \mathbb{R}$, $P(X \in C, Y \in D) = P(X \in C)P(Y \in D)$. Two σ -fields \mathcal{F} and \mathcal{G} are independent if for all $A \in \mathcal{F}$ and $B \in \mathcal{G}$ the events A and B are independent.
- σ -fields $\mathcal{F}_1, \dots, \mathcal{F}_n$ are independent if whenever $A_i \in \mathcal{F}_i$ for $i = 1, \dots, n$, we have $P(\cap_{i=1}^n A_i) = \prod_{i=1}^n P(A_i)$. Random variables X_1, \dots, X_n are independent if whenever $B_i \in \mathbb{R}$ for $i = 1, \dots, n$ we have $P(\cap_{i=1}^n \{X_i \in B_i\}) = \prod_{i=1}^n P(X_i \in B_i)$. Sets A_1, \dots, A_n are independent if whenever $I \subset \{1, \dots, n\}$ we have $P(\cap_{i \in I} A_i) = \prod_{i \in I} P(A_i)$.
- A sequence of events A_1, \dots, A_n with $P(A_i \cap A_j) = P(A_i)P(A_j)$ for all $i \neq j$ is called pairwise independent.
- π - λ theorem: If \mathcal{P} is a π -system and \mathcal{L} is a λ -system that contains \mathcal{P} then $\sigma(\mathcal{P}) \subset \mathcal{L}$.
- Suppose $\mathcal{A}_1, \dots, \mathcal{A}_n$ are independent and each \mathcal{A}_i is a π -system. Then $\sigma(\mathcal{A}_1), \dots, \sigma(\mathcal{A}_n)$ are independent.
- Suppose $\mathcal{F}_{i,j}, 1 \leq i \leq n, 1 \leq j \leq m(i)$ are independent and let $\mathcal{G}_i = \sigma(\cup_j \mathcal{F}_{i,j})$. Then $\mathcal{G}_1, \dots, \mathcal{G}_n$ are independent.
- If for $1 \leq i \leq n, 1 \leq j \leq m(i)$, $X_{i,j}$ are independent and $f_i : \mathbb{R}^{m(i)} \rightarrow \mathbb{R}$ are measurable then $f_i(X_{i,1}, \dots, X_{i,m(i)})$ are independent.
- If X_1, \dots, X_n are independent and have (a) $X_i \geq 0$ for all i , or (b) $\mathbb{E}|X_i| < \infty$ for all i then $\mathbb{E}(\prod_{i=1}^n X_i) = \prod_{i=1}^n \mathbb{E}X_i$.
- If X and Y are independent, $F(x) = P(X \leq x)$, and $G(y) = P(Y \leq y)$, then $P(X + Y \geq z) = \int F(z - y) dG(y)$.

2.2 Weak Laws of Large Numbers

- L^2 weak law: Let X_1, X_2, \dots be uncorrelated random variables with $\mathbb{E}X_i = \mu$ and $\text{var}(X_i) \leq C < \infty$. If $S_n = X_1 + \dots + X_n$, then as $n \rightarrow \infty$, $S_n/n \rightarrow \mu$ in L^2 and in probability.
- Let $\mu_n = \mathbb{E}[S_n]$, $\sigma_n^2 = \text{var}(S_n)$. If $\sigma_n^2/b_n^2 \rightarrow 0$ then $\frac{S_n - \mu_n}{b_n} \rightarrow 0$ in probability.
- Truncation: To truncate a random variable X at level M means to consider $\bar{X}_M = X1_{\{|X| \leq M\}}$.
- For each n , let $X_{n,k}$, $1 \leq k \leq n$ be independent. Let $0 < b_n \rightarrow \infty$ and $\bar{X}_{n,k} = X_{n,k}1_{\{|X_{n,k}| \leq b_n\}}$. Suppose that as $n \rightarrow \infty$ (1) $\sum_{k=1}^n P(|X_{n,k}| > b_n) \rightarrow 0$; (2) $b_n^{-2} \sum_{k=1}^n \text{var}(\bar{X}_{n,k}) \rightarrow 0$. If we let $S_n = \sum_{k=1}^n X_{n,k}$ and $a_n = \sum_{k=1}^n \mathbb{E}[\bar{X}_{n,k}]$, then $\frac{S_n - a_n}{b_n} \rightarrow 0$ in probability.
- Let X_1, X_2, \dots be i.i.d. with $xP(|X_1| > x) \rightarrow 0$ as $x \rightarrow \infty$. Let $S_n = X_1 + \dots + X_n$ and let $\mu_n = \mathbb{E}[X_1 1_{\{|X_1| \leq n\}}]$. Then $S_n/n - \mu_n \rightarrow 0$ in probability.
- If $Y \geq 0$ and $p > 0$ then $\mathbb{E}[Y^p] = \int_0^\infty py^{p-1}P(Y > y)dy$.
- Let $\{X_i\}_{i=1}^\infty$ be i.i.d. with $\mathbb{E}[|X_i|] < \infty$. Let $S_n = X_1 + \dots + X_n$ and let $\mu = \mathbb{E}[X_1]$. Then $S_n/n \rightarrow \mu$ in probability.
- The distribution of X is infinitely divisible iff for any $n \in \mathbb{N}$, there exists i.i.d. Y_i 's such that $X = \sum_{i=1}^n Y_i$.
- The distribution of X is stable if for all $a, b > 0$, and X_1, X_2 i.i.d. copies of X , $aX_1 + bX_2 \stackrel{d}{=} cX + d$ for some $c > 0$.

2.3 Borel-Cantelli Lemmas

- If A_n is a sequence of subsets of Ω , then we write

$$\limsup A_n = \cap_{n=1}^\infty \cup_{m=n}^\infty A_m = \{\omega : \omega \text{ in infinitely many } A_i \text{'s}\}$$

$$\liminf A_n = \cup_{n=1}^\infty \cap_{m=n}^\infty A_m = \{\omega : \omega \text{ in all but finitely many } A_i \text{'s}\}$$

- $P(\limsup A_n) \geq \limsup P(A_n)$, $P(\liminf A_n) \leq \liminf P(A_n)$.
- Borel-Cantelli lemma: If $\sum_i P(A_i) < \infty$, then $P(A_n \text{ i.o.}) = 0$.
- Let y_n be a sequence of elements of a topological space. If every subsequence $y_{n(m)}$ has a further subsubsequence $y_{n(m_k)}$ that converges to y , then $y_n \rightarrow y$.
- $X_n \rightarrow X$ in probability iff for every subsequence $X_{n(m)}$ there is a further subsubsequence $X_{n(m_k)}$ that converges a.s. to X .
- If f is continuous and $X_n \rightarrow X$ in probability then $f(X_n) \rightarrow f(X)$ in probability. If in addition f is bounded then $\mathbb{E}[f(X_n)] \rightarrow \mathbb{E}[f(X)]$.
- Let X_1, X_2, \dots be i.i.d. with $\mathbb{E}[X_i] = \mu$ and $\mathbb{E}[X_i^4] < \infty$. Then $S_n/n \rightarrow \mu$ a.s.
- For events A_n , $n = 1, 2, \dots$, independent such that $\sum_{n=1}^\infty P(A_n) = \infty$, then $P(A_n \text{ i.o.}) = 1$.

LAWS OF LARGE NUMBERS

- If X_1, X_2, \dots are i.i.d. random variables with $\mathbb{E}[X_i] = \infty$, then $P(|X_n| \geq n \text{ i.o.}) = 1$. Let $C = \{\lim S_n/n \text{ exists \& is finite}\}$. Then $P(C) = 0$.
- If A_1, A_2, \dots are pairwise independent and $\sum_{n=1}^{\infty} P(A_n) = \infty$ then $\sum_{i=1}^n 1_{A_i} / \sum_{i=1}^n P(A_i) \rightarrow 1$ a.s. as $n \rightarrow \infty$.
- For a sequence of increasing events A_n , $P(A_n \text{ i.o.}) = 1$ iff $\sum_n P(A_n | A_{n-1}^c) = \infty$.

2.4 Strong Law of Large Numbers

- Strong law of large numbers: Let X_1, X_2, \dots be pairwise independent identically distributed random variables with $\mathbb{E}[X_i] < \infty$. Let $\mathbb{E}X_i = \mu$ and $S_n = X_1 + \dots + X_n$. Then $S_n/n \rightarrow \mu$ a.s. as $n \rightarrow \infty$.
- Let X_1, X_2, \dots be i.i.d. with $\mathbb{E}[X^+] = \infty$ and $\mathbb{E}[X^-] < \infty$, then $S_n/n \rightarrow \infty$ a.s.
- Let X_1, X_2, \dots be i.i.d. with $0 < X_i < \infty$, write $T_n = X_1 + \dots + X_n$ and let $N_t = \sup\{n : T_n \leq t\}$. If $\mathbb{E}[X_1] = \mu \leq \infty$, then as $t \rightarrow \infty$, $N_t/t \rightarrow 1/\mu$, a.s.
- If $X_n \rightarrow X_\infty$ a.s. and $N(n) \rightarrow \infty$ a.s. then $X_{N(n)} \rightarrow X_\infty$ a.s. But the analogous result for convergence in probability is false!
- Empirical distribution functions: Let X_1, X_2, \dots be i.i.d. with distribution F and let $F_n(x) = \frac{\sum_{i=1}^n 1_{X_i \leq x}}{n}$. As $n \rightarrow \infty$, $\sup_x |F_n(x) - F(x)| \rightarrow 0$ a.s.
- Uniform law of large numbers: Suppose $f(x, \theta)$ is continuous in $\theta \in \Theta$ for some compact Θ . Let X_1, X_2, \dots be a sequence of i.i.d. random variables. If f is continuous at θ for a.s. all $x \in \mathbb{R}$ and measurable of x at each θ and there exists some function $d(x)$ such that $\mathbb{E}[d(X_i)] < \infty$ and for all $\theta \in \Theta$, $|f(x, \theta)| \leq d(x)$. Then $\sup_{\theta \in \Theta} |\frac{1}{n} \sum_{i=1}^n f(X_i, \theta) - \mathbb{E}[f(X_1, \theta)]| \xrightarrow{\text{a.s.}} 0$.

2.5 Convergence of Random Series

- Let $X_1, X_2, \dots, X_n, \dots$ be a sequence of random variables. Define $\mathcal{F}'_n = \sigma(X_n, X_{n+1}, \dots)$ as the information of the future after time n . Let $\mathcal{I} = \cap_{n=1}^{\infty} \mathcal{F}'_n$ be the tail σ -field, i.e., the information in the remote future. Intuitively, $A \in \mathcal{T}$ if and only if changing a finite number of values does not affect the occurrence of the event.
- Kolmogorov's 0-1 law: If $X_1, X_2, \dots, X_n, \dots$ are independent and $A \in \mathcal{I}$, then $P(A) = 0$ or 1 .
- A finite permutation of \mathbb{N} is a map from \mathbb{N} onto \mathbb{N} such that there is a finite I with $\pi(i) = i$ for all $i \geq I$. For $S^{\mathbb{N}}$, associated with its natural product sigma field $\mathcal{F}^{\mathbb{N}}$, and any $\omega = (\omega_1, \omega_2, \dots)$, let $\pi(\omega) = (\omega_{\pi(1)}, \omega_{\pi(2)}, \dots)$. An event $A \in \mathcal{F}^{\mathbb{N}}$ is permutable if $\pi^{-1}(A) = A$ for any finite permutation π . All permutable events form the exchangeable σ -field, denoted by \mathcal{E} . All events in the tail σ -field \mathcal{T} are permutable.
- Hewitt-Savage 0-1 law: If X_1, X_2, \dots , are i.i.d. and $B \in \mathcal{E}(\mathbb{R}^{\mathbb{N}})$. Denote $X = (X_1, X_2, \dots)$. Then $P(X \in B) = 0$ or 1 .

LAWS OF LARGE NUMBERS

- Kolmogorov's maximal inequality: Suppose X_1, X_2, \dots, X_n are independent with $\mathbb{E}[X_i] = 0$, $\text{var}(X_i) < \infty$. Let $S_n = X_1 + \dots + X_n$, then $P(\max_{k \leq n} |S_k| \geq x) \leq \frac{\text{var}(S_n)}{x^2}$.
- We call a sequence of r.v's S_1, S_2, \dots a martingale if (i) there is a sequence of σ -algebras $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \dots$ and $S_i \in \mathcal{F}_i$ for all i ; (ii) S_i 's are integrable; (iii) For each k , $\mathbb{E}[S_{k+1} | \mathcal{F}_k] = S_k$. If the "=" in (iii) is replaced by \geq (resp. \leq), then we say that this sequence is a submartingale (resp. supermartingale).
- Second-moment criterion: Suppose X_1, X_2, \dots are independent and centered (i.e., for all i , $\mathbb{E}[X_i] = 0$). If $\sum_{n=1}^{\infty} \text{var}(X_n) < \infty$, then $P(\sum_{n=1}^{\infty} X_n(\omega) \text{ converges}) = 1$.
- Kronecker's lemma: If $a_n \uparrow \infty$ and $\sum_{n=1}^{\infty} x_n/a_n$ converges, then $a_n^{-1} \sum_{m=1}^n x_m \rightarrow 0$.
- Let X_1, X_2, \dots be i.i.d. random variables with $\mathbb{E}[X_i] = 0$ and $\mathbb{E}[X_i^2] = \sigma^2 < \infty$. Let $S_n = X_1 + \dots + X_n$. If $\epsilon > 0$, then $S_n/n^{1/2}(\log n)^{1/2+\epsilon} \rightarrow 0$ a.s.
- Let X_1, X_2, \dots be i.i.d. with $\mathbb{E}[X_i] = 0$ and $\mathbb{E}[|X_i|^p] < \infty$ where $1 < p < 2$. Write $S_n = X_1 + \dots + X_n$. Then $S_n/n^{1/p} \rightarrow 0$ a.s.
- Let X_1, X_2, \dots be i.i.d. with $\mathbb{E}[X_1] = \infty$ and let $S_n = X_1 + \dots + X_n$. Let a_n be a sequence of positive numbers with a_n/n increasing. Then $\limsup_{n \rightarrow \infty} |S_n|/a_n = 0$ or ∞ according as $\sum_n P(|X_1| \geq a_n) < \infty$ or $= \infty$.
- Kolmogorov's three-series theorem: Let $X_1, X_2, \dots, X_n, \dots$ be independent random variables. Let $A > 0$ and $Y_i = X_i 1_{|X_i| \leq A}$. In order to show that $\sum X_i$ converges a.s., it is necessary and sufficient that (i) $\sum_{n=1}^{\infty} P(|X_n| > A) < \infty$; (ii) $\sum_{n=1}^{\infty} \mathbb{E}[Y_n]$ converges; (iii) $\sum_{n=1}^{\infty} \text{var}(Y_n) < \infty$.

2.6 Large Deviations

- Let X_1, X_2, \dots be i.i.d. with $\mathbb{E}[X_1] = \mu$ and let $S_n = X_1 + X_2 + \dots + X_n$. According to CLT, the typical value of $S_n - n\mu$ is $O(\sqrt{n})$. What about atypical deviations of $S_n - n\mu$? According to WLLN, we know that for any $a > \mu$, $P(S_n > na) \rightarrow 0$. We want to discuss the existence and value of the limit: $\lim_{n \rightarrow \infty} \frac{1}{n} \log P(S_n > na)$.
- Let $\pi_n = P(S_n \geq na)$. Then $\pi_{n+m} \geq P(S_n \geq na, S_{n+m} - S_n \geq ma) = \pi_n \pi_m$. Let $\gamma_n = \log \pi_n$, $\gamma_{n+m} \geq \gamma_n + \gamma_m$. As $n \rightarrow \infty$ the limit of γ_n exists and $\lim_{n \rightarrow \infty} \frac{\gamma_n}{n} = \sup_n \frac{\gamma_n}{n}$. We define $\gamma(a) = \lim_{n \rightarrow \infty} \gamma_n/n \leq 0$. Then for any distribution and any n and a , $P(S_n \geq na) \leq e^{n\gamma(a)}$. We want to show $\gamma(a) < 0$ if $a > \mu$.
- If the moment generating function $\psi(\theta) = \mathbb{E}[\exp(\theta X_1)] < \infty$ for some $\theta > 0$, then $P(S_n \geq na) \leq \exp[n(\log \psi(\theta) - \theta a)]$. Let $\kappa(\theta) = \log \psi(\theta)$. If $a > \mu$, then $a\theta - \kappa(\theta) > 0$ for all sufficiently small θ .
- We will further strengthen our upper bounds by finding the maximum of $\lambda(\theta) = a\theta - \kappa(\theta)$. Let $\theta_+ = \sup\{\theta : \psi(\theta) < \infty\}$ and $\theta_- = \inf\{\theta : \psi(\theta) < \infty\}$. Now since that $\psi(\theta) \in C^\infty$ within (θ_-, θ_+) , we have $\lambda'(\theta) = a - \frac{\psi'(\theta)}{\psi(\theta)}$. So the maximal point of λ must satisfy $\psi'(\theta)/\psi(\theta) = a$. For

LAWS OF LARGE NUMBERS

the existence and uniqueness of such point(s), we introduce a new distribution, and use a trick named “tilting”.

- We now introduce the distribution F_θ by “reweighting F ”: $F_\theta(x) = \frac{1}{\psi(\theta)} \int_{-\infty}^x e^{y\theta} dF(y)$. By simple calculus, $\int x dF_\theta(x) = \frac{\psi'(\theta)}{\psi(\theta)}$, $\psi''(\theta) = \int x^2 e^{\theta x} dF(x)$, $\frac{d}{d\theta} \frac{\psi'(\theta)}{\psi(\theta)} = \int x^2 dF_\theta(x) - (\int x dF_\theta(x))^2 \geq 0$. If we assume the distribution F is not a point mass at μ , then $\frac{\psi'(\theta)}{\psi(\theta)}$ is strictly increasing and $a\theta - \log \psi(\theta)$ is concave. Since we have $\frac{\psi'(0)}{\psi(0)} = \mu$, this shows that for each $a > \mu$ there is at most one $\theta_a \geq 0$ that solves $a = \frac{\psi'(\theta_a)}{\psi(\theta_a)}$, and this value of θ maximizes $a\theta - \log \psi(\theta)$. Let F^n be the c.d.f. of $S_n = X_1 + \dots + X_n$ and F_λ^n be the c.d.f. of $S_n^\lambda = X_1^\lambda + \dots + X_n^\lambda$ where X_i i.i.d. $\sim F$ and X_i^λ i.i.d. $\sim F_\lambda = \frac{1}{\psi(\lambda)} \int_{-\infty}^x e^{y\lambda} dF(y)$. By induction, $\frac{dF^n}{dF_\lambda^n} = e^{-\lambda x} \psi(\lambda)^n$. Then as $n \rightarrow \infty$, $n^{-1} \log P(S_n \geq na) \rightarrow -a\theta_a + \log \psi(\theta_a)$.
- Some important information: $\kappa(\theta) = \log \psi(\theta)$, $\kappa'(\theta) = \frac{\psi'(\theta)}{\psi(\theta)}$, θ_a solves $\kappa'(\theta_a) = a$, $\gamma(a) = \lim_{n \rightarrow \infty} \frac{1}{n} \log P(S_n \geq na) = -a\theta_a + \kappa(\theta_a)$.
- Suppose $x_o = \sup\{x : F(x) < 1\} = \infty$, $\theta_+ < \infty$, and $\psi'(\theta)/\psi(\theta)$ increases to a finite limit a_0 as $\theta \uparrow \theta_+$. If $a_0 \leq a < \infty$, $n^{-1} \log P(S_n \geq na) \rightarrow -a\theta_+ + \log \psi(\theta_+)$, i.e. $\gamma(a)$ is linear for $a \geq a_0$.
- Suppose $x_o = \sup\{x : F(x) < 1\} < \infty$ and F has no mass at x_o . Then $\psi(\theta) < \infty$ for all $\theta > 0$ and $\psi'(\theta)/\psi(\theta) \rightarrow x_o$ as $\theta \rightarrow \infty$.
- Now, we have shown the decaying asymptotic for all possible situations:

$$\left\{ \begin{array}{l} \text{If } x_o < \infty : \left\{ \begin{array}{l} a < x_o : \text{exponential, rate} = \theta_a \\ a = x_o : \text{exponential if } P(X_1 = x_o) > 0, 0 \text{ otherwise} \\ a > x_o : 0 \end{array} \right. \\ \text{If } x_o = \infty : \left\{ \begin{array}{l} \text{If } \theta_+ = \infty : \text{exponential, rate} = \theta_a \\ \text{If } \theta_+ < \infty : \left\{ \begin{array}{l} \text{If } \psi'(\theta)/\psi(\theta) \rightarrow \infty \text{ as } \theta \rightarrow \theta_+ : \text{exponential, rate} = \theta_a \\ \text{If } \psi'(\theta)/\psi(\theta) \rightarrow a_0 \text{ as } \theta \rightarrow \theta_+ : \left\{ \begin{array}{l} a < a_0 : \text{exponential, rate} = \theta_a \\ a \geq a_0 : \text{exponential, rate} = \theta_+ \end{array} \right. \end{array} \right. \end{array} \right.$$
- Cramér’s theorem: Let $I(a)$ be the Legendre transform of $\log \psi(\cdot)$: $I(a) := \sup_{\theta \in \mathbb{R}} (\theta a - \log \psi(\theta))$. Then for any closed set F , $\limsup_{n \rightarrow \infty} n^{-1} \log P(\frac{S_n}{n} \in F) \leq -\inf_{x \in F} I(x)$; for any open set G , $\liminf_{n \rightarrow \infty} n^{-1} \log P(\frac{S_n}{n} \in G) \geq -\inf_{x \in G} I(x)$.
- Intuition behind the tilting: Why do we want to introduce the measure F_θ ? Intuitively, the new measure is like a “distorting mirror” – it “distorts” our view on how each event is likely to happen. So, when we want to estimate a rare event A under P , suppose (1) we can construct a new measure Q such that $Q[A]$ is easily calculable, e.g., $Q[A] \approx 1$; (2) we have a uniform lower bound of the R-N derivative $dP/dQ \geq c$ on A . Then we can conclude that $P[A] = \int_A \frac{dP}{dQ} dQ \geq cQ[A]$.
- Let $\Sigma = \{a_1, \dots\}$ stand for a finite-size alphabet. Let $M_1(\Sigma)$ be the space of all probability measures on Σ . The entropy of some $\nu \in M_1(\Sigma)$ is $H(\nu) := -\sum_{i=1}^{|\Sigma|} \nu(a_i) \log(\nu(a_i))$. The relative entropy of ν with respect to some other $\mu \in M_1(\Sigma)$ is $H(\nu|\mu) := \sum_{i=1}^{|\Sigma|} \nu(a_i) \log \frac{\nu(a_i)}{\mu(a_i)}$.

CENTRAL LIMIT THEOREMS

- Let Y_i be i.i.d. r.v.s, $\mu \in M_1(\Sigma)$. For $n \geq 1$, write $Y = (Y_1, \dots, Y_n)$ and call $L_n^Y \in M_1(\Sigma)$ be the empirical frequency of Y . Let $T_n(\nu)$ be the set of y a sequence of n letters whose empirical measure is ν .
- If $y \in T_n(\nu)$, then $P_\mu(Y = y) = e^{-n(H(\nu) + H(\nu|\mu))}$. In particular, if $y \in T_n(\mu)$, then $P_\mu(Y = y) = e^{-nH(\mu)}$.
- For every possible empirical measure ν of n letters, $(n+1)^{-|\Sigma|} e^{nH(\nu)} \leq |T_n(\nu)| \leq e^{nH(\nu)}$.
- For every possible empirical measure ν of n letters, $(n+1)^{-|\Sigma|} e^{nH(\nu|\mu)} \leq P_\mu(L_n^T = \nu) \leq e^{nH(\nu|\mu)}$.
- Sanov's theorem: For every set $\Gamma \subset M_1(\Sigma)$, $-\inf_{\nu \in \Gamma^\circ} H(\nu|\mu) \leq \liminf \frac{1}{n} \log P_\mu(L_n^Y \in \Gamma) \leq \limsup \frac{1}{n} \log P_\mu(L_n^Y \in \Gamma) \leq -\inf_{\nu \in \Gamma} H(\nu|\mu)$.

2.7 Percolation

- Fix $p \in [0, 1]$ and consider the d -dimensional lattice \mathbb{Z}^d . Assign to each edge $e \in \mathbb{E}$ an independent Bernoulli r.v. $I(e)$ with parameter p . If $I(e) = 1$, we say that this edge is open, otherwise closed. Consider the connected components of open edges, then for any $p \in [0, 1]$, $P_p(A) = 0$ or 1 where $A = \{\exists \text{ infinite open clusters}\}$.
- If A is translation-invariant, then $P(A) = 0$ or 1.
- Actually we can go further and show that for any $N = 0, 1, \dots, \infty$, $P_p[A(N)] = 0$ or 1, where $A(N) = \{\exists N \text{ infinite open clusters}\}$. Or even further: for $N = 2, 3, \dots$ and $N = \infty$, $P_p[A(N)] = 0$.
- Let $p_c = p_c(d) = \sup\{p : P_p(A) = 0\}$. Then one can show that $1/3 \leq p_c(2) \leq 2/3$. More generally, $p_c(1) = 1$ and for $d \geq 2$, $1/(2d-1) \leq p_c(d) \leq p_c(2)(= 1/2)$.
- By knowledge of Galton-Watson tree and the analogy between \mathbb{Z}^d and $2d$ -regular tree in high dimensions, we can take an educated guess that $p_c(d) \sim \frac{1}{2d}$ as $d \rightarrow \infty$.

3 Central Limit Theorems

3.1 The De Moivre-Laplace Theorem

- Central Limit Theorem: Let X_1, X_2, \dots be i.i.d. with mean μ and variance $\sigma^2 \in (0, \infty)$. Write $S_n = X_1 + \dots + X_n$, then $\frac{S_n - \mu n}{\sqrt{n}\sigma} \Rightarrow \mathcal{N}(0, 1)$.
- Before discussing the central limit theorem in full generality, we first see a special example for Bernoulli random variables. Let X_1, X_2, \dots be i.i.d. random variables such that $P(X_1 = 1) = P(X_1 = -1) = 1/2$ and write $S_n = X_1 + \dots + X_n$. For integers $|k| \leq n$, $P(S_{2n} = 2k) = C_{2n}^{n+k} 2^{-2n}$ since $(S_{2n} + 2n)/2 \sim \text{Binomial}(2n, 1/2)$.
- Local central limit theorem: If $2k/\sqrt{2n} \rightarrow x$, then $\lim_{n \rightarrow \infty} (\pi n)^{1/2} e^{x^2/2} P(S_{2n} = 2k) = 1$.
- The De Moivre-Laplace Theorem: For $a < b$, $P(a \leq S_n/\sqrt{n} \leq b) \rightarrow \int_a^b (2\pi)^{-1/2} e^{-x^2/2} dx$.

3.2 Weak Convergence

- A sequence of distribution function F_n is said to converge weakly to a limit F , denoted by $F_n \Rightarrow F$, if $F_n(y) \rightarrow F(y)$ at every point of continuity of F , i.e. every $y \in \mathbb{R}$ such that $F(\cdot)$ is continuous at y .
- A sequence of random variables X_n is said to converge weakly or converge in distribution / law to a limit X_∞ if their distribution functions F_n converges weakly.
- Skorokhod's representation theorem: If $F_n \Rightarrow F$ then there are random variables $Y_n, 1 \leq n < \infty$ and Y with living in the same probability space such that $Y_n \sim F_n, Y \sim F$ and $Y_n \rightarrow Y$ a.s.
- $X_n \Rightarrow X$ if and only if for every bounded continuous function g we have $\mathbb{E}g(X_n) \rightarrow \mathbb{E}g(X)$.
- Continuous mapping theorem: Let g be a measurable function and $D_g = \{x : g \text{ is discontinuous at } x\}$. If $X_n \Rightarrow X$, and $P(X \in D_g) = 0$, then $g(X_n) \Rightarrow g(X)$.
- Portmanteau theorem: The following statements are equivalent: (1) $X_n \Rightarrow X$; (2) G open, $\liminf_{n \rightarrow \infty} P(X_n \in G) \geq P(X \in G)$; (3) G closed, $\limsup_{n \rightarrow \infty} P(X_n \in G) \leq P(X \in G)$; (4) If $P(X \in \partial A) = 0$, then $\lim_{n \rightarrow \infty} P(X_n \in A) = P(X \in A)$.
- Helly's selection theorem: For every sequence F_n of distribution functions, there is a subsequence $F_{n(k)}$ and a right continuous nondecreasing function F so that at all points of continuity y of F , $\lim_{k \rightarrow \infty} F_{n(k)}(y) = F(y)$.
- Every subsequential limit of the sequence F_n is the distribution function of a probability measure iff the sequence is tight, i.e., for all $\epsilon > 0$, there is an M_ϵ so that $\limsup_{n \rightarrow \infty} [1 - F_n(M_\epsilon) + F_n(-M_\epsilon)] \leq \epsilon$.
- If there is a function $\phi \geq 0$ so that $\phi(x) \rightarrow \infty$ as $|x| \rightarrow \infty$ and $C = \sup_n \int \phi(x) dF_n(x) < \infty$, then F_n is tight.

3.3 Characteristic Functions

- If X is a random variable, we define its Characteristic function (ch.f.) by $\phi(t) := \mathbb{E}[e^{itX}] = \mathbb{E}[\cos(tX)] + i\mathbb{E}[\sin(tX)]$.
- All characteristic functions have the following properties: (i) $\phi(0) = 1$; (ii) $\phi(-t) = \overline{\phi(t)}$; (iii) $|\phi(t)| = |\mathbb{E}e^{itX}| \leq \mathbb{E}|e^{itX}| = 1$; (iv) $|\phi(t+h) - \phi(t)| \leq \mathbb{E}|e^{itX} - 1|$, so $\phi(t)$ is uniformly continuous on \mathbb{R} ; (v) $\mathbb{E}e^{it(aX+b)} = e^{itb}\phi(at)$.
- If X_1 and X_2 are independent and have ch.f.'s ϕ_1 and ϕ_2 . Then $X_1 + X_2$ has ch.f. $\phi_1 \cdot \phi_2$.
- Stein's Lemma: If X, Y are jointly Gaussian, then for differentiable $g : \mathbb{R} \rightarrow \mathbb{R}$, as long as the expectations are well-defined, $\text{cov}(g(X), Y) = \text{cov}(X, Y)\mathbb{E}[g'(X)]$.
- If F_1, \dots, F_n have ch.f. ϕ_1, \dots, ϕ_n and $\lambda_i \geq 0, 1 \leq i \leq n$ have $\lambda_1 + \dots + \lambda_n = 1$. Then $\sum \lambda_i F_i$ has ch.f. $\sum \lambda_i \phi_i$.

CENTRAL LIMIT THEOREMS

- The inversion formula: For a probability measure μ , recall $\phi(t) = \int e^{itx} \mu(dx)$. If $a < b$, then $\frac{1}{2\pi} \lim_{T \rightarrow \infty} \int_{-T}^T \frac{e^{-ita} - e^{-itb}}{it} \phi(t) dt = \mu(a, b) + \frac{1}{2} \mu(\{a, b\})$.
- If $\int |\phi(t)| dt < \infty$, then μ has bounded continuous density $f(y) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ity} \phi(t) dt$.
- Continuity theorem: Let $\mu_n, 1 \leq n \leq \infty$ be probability measures with ch.f. ϕ_n . (i) If $\mu_n \Rightarrow \mu_\infty$ then $\phi_n(t) \rightarrow \phi_\infty(t)$ for all t . (ii) If $\phi_n(t) \rightarrow \phi(t)$ for all t , and $\phi(t)$ is continuous at 0. Then $\{\mu_n\}_{n=1}^\infty$ is tight and has a weak limit with ch.f. ϕ .
- Let μ be a probability measure and ϕ be its ch.f. Then $\mu(\{x : |x| \geq 2u^{-1}\}) \leq u^{-1} \int_{-u}^u [1 - \phi(t)] dt$.
- If $\int |x|^n \mu(dx) < \infty$, then its ch.f. ϕ has a continuous derivative of order n given by $\phi^{(n)}(t) = \int (ix)^n e^{itx} \mu(dx)$. In particular, $\phi^{(n)}(0) = \mathbb{E}[(iX)^n]$.
- However, if a characteristic function ϕ_X has a k -th derivative at zero, then the random variable X has all moments up to k if k is even, but only up to $(k-1)$ if k is odd.
- $|e^{ix} - \sum_{m=0}^n \frac{(ix)^m}{m!}| \leq \min(\frac{|x|^{n+1}}{(n+1)!}, \frac{2|x|^n}{n!})$.
- If $\mathbb{E}|X|^2 < \infty$, then $\phi(t) = 1 + it\mathbb{E}X - t^2\mathbb{E}|X|^2/2 + o(t^2)$.
- If $\limsup_{h \downarrow 0} \frac{\phi(h) - 2\phi(0) + \phi(-h)}{h^2} > -\infty$, then $\mathbb{E}[X^2] < \infty$.
- Given ϕ and $x_1, \dots, x_n \in \mathbb{R}$, we can consider the matrix with (i, j) entry given by $\phi(x_i - x_j)$. Call ϕ positive definite if this matrix is always positive semi-definite Hermitian.
- Bochner's theorem: A function from \mathbb{R} to \mathbb{C} which is continuous at origin with $\phi(0) = 1$ is a ch.f. of some probability measure on \mathbb{R} if and only if it is positive definite.
- Pólya's theorem: If ϕ is real-valued, even and continuous such that (i) $\phi(0) = 1$; (ii) ϕ is convex for $t > 0$; (iii) $\phi(\infty) = 0$; then $\phi(t)$ is the ch.f. of a distribution symmetric about 0.

3.4 Central Limit Theorems

- Central Limit Theorem: Let X_1, X_2, \dots be i.i.d. with $\mathbb{E}[X_1] = \mu, \text{var}(X_1) = \sigma^2 \in (0, \infty)$. If $S_n = X_1 + X_2 + \dots + X_n$, then $\frac{S_n - n\mu}{n^{1/2}\sigma} \Rightarrow \mathcal{N}(0, 1)$.
- The Lindeberg-Feller theorem: For each n , let $X_{n,m}, 1 \leq m \leq n$, be independent random variables for each n with $\mathbb{E}[X_{n,m}] = 0$. Suppose (i) $\sum_{m=1}^n \mathbb{E}[X_{n,m}^2] \rightarrow \sigma^2 > 0$; (ii) For all $\epsilon > 0$, $\lim_{n \rightarrow \infty} \sum_{m=1}^n \mathbb{E}[X_{n,m}^2 1_{|X_{n,m}| > \epsilon}] = 0$. Then $S_n = X_{n,1} + \dots + X_{n,n} \Rightarrow \mathcal{N}(0, \sigma^2)$ as $n \rightarrow \infty$.
- Converging together lemma: If $X_n \Rightarrow X$ and $Y_n \Rightarrow c$, $X_n + Y_n \Rightarrow X + c$. A useful consequence of this result is that if $X_n \Rightarrow X$ and $Z_n - X_n \Rightarrow 0$ then $Z_n \Rightarrow X$.
- Lévy's condition for CLT: Let X_1, X_2, \dots be i.i.d. and $S_n = X_1 + \dots + X_n$. In order that there exist constants a_n and $b_n > 0$ so that $(S_n - a_n)/b_n \Rightarrow \mathcal{N}(0, 1)$, it is necessary and sufficient that $\frac{y^2 P(|X_1| > y)}{\mathbb{E}[X_1^2 1_{|X_1| < y}]} \rightarrow 0$.

CENTRAL LIMIT THEOREMS

- Chernoff bound: Let X_i be independent Bernoulli r.v.'s. Write $S_n = X_1 + \cdots + X_n$ and let $\mu = \mathbb{E}[S_n]$. Then for $\delta > 0$, $P(S_n > (1 + \delta)\mu) \leq e^{-\frac{\delta^2 \mu}{2 + \delta}}$, $P(S_n < (1 - \delta)\mu) \leq e^{-\frac{\delta^2 \mu}{2}}$.
- Hoeffding's inequality for bounded r.v. Let X_i be independent r.v.'s such that $X_i \in [a_i, b_i]$ a.s. Write $S_n = X_1 + \cdots + X_n$ and let $\mu = \mathbb{E}[S_n]$. Then for $\delta > 0$, $P(|S_n - \mu| \geq \delta) \leq 2 \exp(-\frac{2n^2 \delta^2}{\sum_{i=1}^n (b_i - a_i)^2})$.
- A random variable is sub-Gaussian, if and only if for some $C < \infty$ and $c > 0$, $P(|X| \geq t) \leq Ce^{-ct^2}$.
- Hoeffding's inequality for sub-Gaussian r.v.'s: Let X_i be independent zero-mean sub-Gaussian r.v.'s. Write $S_n = X_1 + \cdots + X_n$. Then there exists some $c > 0$ such that for any $\delta > 0$, $P(|S_n| \geq \delta) \leq 2 \exp(-c\delta^2 / \sum_{i=1}^n \|X_i\|_{\psi_2})$, where $\|X\|_{\psi_2} = \inf\{c \geq 0 : \mathbb{E}[e^{X^2/c^2}] \leq 2\}$.