High-Dimensional Probability

Lectured by Hao Ge   LaTeXed by Chengxin Gong

October 10, 2023

Contents

0 Appetizer

- Convex combination: For $z_1, z_2, \cdots, z_m \in \mathbb{R}^n$, the form of $\sum_{i=1}^m \lambda_i z_i$ with $\lambda_i \geq 0, \sum_{i=1}^m \lambda_i = 1$. Convex hull of $T \subset \mathbb{R}^n$: $\mathrm{conv}(T) = \{$convex combinations of $z_1, \cdots, z_m \in T, m \in \mathbb{N}\}$.

- Caratheodory's theorem: Every point in the convex hull of a set $T \subset \mathbb{R}^n$ can be expressed as a convex combination of at most $n+1$ points from T.

- Approximate Caratheodory's theorem: Consider $T \subset \mathbb{R}^n$, $\mathrm{diam}(T) = \sup\{||s-t||_2, s, t \in T\} < 1$. Then for any $x \in \mathrm{conv}(T)$ and any k, one can find points $x_1, x_2, \cdots, x_k \in T$ such that $||x - \frac{1}{k}\sum_{i=1}^k x_i||_2 \leq \frac{1}{\sqrt{k}}$ (repetition is allowed).

  Proof WLOG assume $||t||_2 \leq 1, \forall t \in T$. Fix $x \in \mathrm{conv}(T), x = \sum_{i=1}^m \lambda_i z_i, z_i \in T$. Define $Z, \mathbb{P}(Z = z_i) = \lambda_i, \mathbb{E}Z = x$. Consider i.i.d. $Z_1, Z_2, \cdots$ of Z, $\frac{1}{n}\sum_{j=1}^n Z_j \to x$ a.s. $n \to +\infty$. $\mathbb{E}||x - \frac{1}{k}\sum_{j=1}^k Z_j||_2^2 = \frac{1}{k^2}\sum_{j=1}^k \mathbb{E}||Z_j - x||_2^2 = \frac{1}{k^2}\sum_{j=1}^k (\mathbb{E}||Z_j||^2 - ||\mathbb{E}Z_j||_2^2) \leq \frac{1}{k} \Rightarrow \exists$ a realization of $Z_1, \cdots, Z_k$ such that $||x - \frac{1}{k}\sum_{j=1}^k Z_j||_2 \leq \frac{1}{\sqrt{k}}$. □

- Corollary (Covering polytopes by balls): P is a polytope in $\mathbb{R}^n$ with N vertices, $\mathrm{diam}(P) \leq 1$. Then P can be covered by at most $N^{\lfloor 1/\epsilon^2 \rfloor}$ Euclidean balls of radii $\epsilon > 0$.

1 Preliminaries on random variables

- Jensen's inequality: convex ϕ, $\phi(\mathbb{E}X) \leq \mathbb{E}\phi(X)$. $\Rightarrow ||X||_{L^p} \leq ||X||_{L^q}$ for $p \leq q$.

- Minkowski inequality: $p \geq 1, ||X+Y||_{L^p} \leq ||X||_{L^p} + ||Y||_{L^p}$.

- Cauchy-Schwarz inequality: $\mathbb{E}|XY| \leq ||X||_{L^2}||Y||_{L^2}$.

- Holder inequality: $p, q \in (1, +\infty), \frac{1}{p} + \frac{1}{q} = 1$ or $p = 1, q = \infty$, $\mathbb{E}||XY|| \leq ||X||_{L^p}||Y||_{L^q}$.

- $X \geq 0$, then $\mathbb{E}X = \int_0^{+\infty} \mathbb{P}(X > t)\mathrm{d}t$.

- Markov inequality: $X \geq 0, t > 0, \mathbb{P}(X \geq t) \leq \frac{\mathbb{E}X}{t}$.

- LLN: $X_1, \cdots, X_n, \cdots$ i.i.d., $\mathbb{E}X_i = \mu, \mathrm{Var}(X_i) = \sigma^2, S_N = X_1 + X_2 + \cdots + X_N$. Then: (WLLN) $\mathbb{P}(|\frac{S_N}{N} - \mu| > \epsilon) \to 0, \forall \epsilon > 0$; (SLLN) $\mathbb{P}(\frac{S_N}{N} \to \mu, N \to +\infty) = 1$.

- CLT: $Z_N = \frac{S_N - \mathbb{E}S_N}{\sqrt{\mathrm{Var}(S_N)}} \xrightarrow{d} \mathcal{N}(0,1)$.

- $X_{N,i}, 1 \leq i \leq N$ independent $\mathrm{Ber}(p_{N,i}), \max_{i \leq N} p_{N,i} \to 0, S_N = \sum_{i=1}^N X_{N,i}, \mathbb{E}S_N \to \lambda < +\infty$. Then $S_N \xrightarrow{d} \mathrm{Poisson}(\lambda)$.

2 Concentration of sums of independent random variables

- Question: N times, $\mathbb{P}(\text{head} \geq \frac{3}{4}N) = ?$ Let S_N be the number of heads, $\mathbb{E}S_N = \frac{N}{2}, \mathrm{Var}(S_N) = \frac{N}{4}$. (1) Chebyshev's inequality: $\mathbb{P}(S_N \geq \frac{3}{4}N) \leq \mathbb{P}(|S_N - \frac{N}{2}| \geq \frac{N}{4}) \leq \frac{4}{N}$; (2) $Z_N = \frac{S_N - \frac{N}{2}}{\sqrt{N/4}}$, expect: $\mathbb{P}(Z_N \geq \sqrt{N/4}) \approx \mathbb{P}(g \geq \sqrt{N/4}) \leq \frac{1}{\sqrt{2\pi}}e^{-N/8}$ where $g \sim \mathcal{N}(0,1)$.

- For all $t > 0$, $(\frac{1}{t} - \frac{1}{t^3})\frac{1}{\sqrt{2\pi}}e^{-t^2/2} \leq \mathbb{P}(g \sim \mathcal{N}(0,1) \geq t) \leq \frac{1}{t}\frac{1}{\sqrt{2\pi}}e^{-t^2/2}$.

- Berry-Esseen bound: $|\mathbb{P}(Z_N \geq t) - \mathbb{P}(g \geq t)| \leq \frac{\rho}{\sqrt{N}}$ where $\rho = \mathbb{E}|X_1 - \mu|^3/\sigma^3$. And in general, no improvement since $\mathbb{P}(S_N = \frac{N}{2}) \sim \frac{1}{\sqrt{N}} \Rightarrow \mathbb{P}(Z_N = 0) \sim \frac{1}{\sqrt{N}}$ but $\mathbb{P}(g = 0) = 0$.

- Hoeffding's inequality: $X_1, \cdots, X_N$ i.i.d. symmetric Bernoulli ($\mathbb{P}(X = 1) = \mathbb{P}(X = -1) = \frac{1}{2}$), $a = (a_1, \cdots, a_N) \in \mathbb{R}^N$. Then $\forall t \geq 0, \mathbb{P}(\sum_{i=1}^N a_i X_i \geq t) \leq e^{-t^2/2||a||_2^2}, \mathbb{P}(|\sum_{i=1}^N a_i X_i \geq t|) \leq 2e^{-t^2/2||a||_2^2}$.

  Proof WLOG, $||a||_2^2 = 1$. For $\lambda > 0, \mathbb{P}(\sum_{i=1}^N a_i X_i \geq t) = \mathbb{P}(e^{\lambda \sum a_i X_i} \geq e^{\lambda t}) \leq e^{-\lambda t}\mathbb{E}e^{\lambda \sum_{i=1}^N a_i X_i} = e^{-\lambda t}\prod_{i=1}^N \mathbb{E}e^{\lambda a_i X_i} = e^{-\lambda t}\prod_{i=1}^N \cosh(\lambda a_i) \leq e^{-\lambda t}\prod_{i=1}^N e^{\lambda^2 a_i^2/2} = e^{-\lambda t + \frac{\lambda^2}{2}} \Rightarrow \mathbb{P}(\sum_{i=1}^N a_i X_i \geq t) \leq \inf_{\lambda \geq 0} e^{-\lambda t + \frac{\lambda^2}{2}} = e^{-\frac{t^2}{2}} (\lambda = t)$. □

- Bounded r.v.s: $X_1, \cdots, X_N$ independent, $X_i \in [m_i, M_i]$. Then $\forall t \geq 0, \mathbb{P}(\sum_{i=1}^N (X_i - \mathbb{E}X_i) \geq t) \leq e^{-\frac{2t^2}{\sum_{i=1}^N (M_i - m_i)^2}}$.

- Chernoff's inequality: $X_i \sim \mathrm{Ber}(p_i)$ independent, $S_N = \sum_{i=1}^N X_i, \mu = \mathbb{E}S_N \Rightarrow \forall t > \mu, \mathbb{P}(S_N \geq t) \leq e^{-\mu}(\frac{e\mu}{t})^t$.

  Proof $\mathbb{P}(S_N \geq t) \leq e^{-\lambda t} \prod_{i=1}^N \mathbb{E}e^{\lambda X_i}$. $\mathbb{E}e^{\lambda X_i} = e^\lambda p_i + (1-p_i) = 1 + (e^\lambda - 1)p_i \leq e^{(e^\lambda - 1)p_i} \Rightarrow \mathbb{P}(S_N \geq t) \leq e^{-\lambda t} e^{(e^\lambda - 1)\mu}$. Take $\lambda^* = \log(t/\mu)$. □

- $d = (n-1)p$ is the expected degree. There is an absolute constant C s.t. for $G(n,p)$, $d \geq C \log n$. Then with high prob (for example 0.9), all vertices of G have degrees between $0.9d$ and $1.1d$.

  Proof Ex 2.3.5 $\Rightarrow \mathbb{P}(|d_i - d| \geq \delta d) \leq 2e^{-c\delta^2 d}$. Union bound: $\mathbb{P}(\exists i, |d_i - d| \geq \delta d) \leq n \cdot 2e^{-c\delta^2 d} \leq n \cdot 2 \cdots n^{-Cc\delta^2} = 2n^{1-Cc\delta^2} \leq 1 - p^*$ (let $Cc\delta^2 > 1$). □

- Sub-gaussian properties: The following are equivalent: (i) $\mathbb{P}(|X| \geq t) \leq 2e^{-t^2/k_1^2}$ for all $t \geq 0$; (ii) $||X||_{L^p} = (\mathbb{E}|X|^p)^{1/p} \leq k_2\sqrt{p}$ for all $p \geq 1$; (iii) $\mathbb{E}e^{\lambda^2 X^2} \leq e^{k_3^2 \lambda^2}$ for all λ s.t. $|\lambda| \leq \frac{1}{k_3}$; (iv) $\mathbb{E}e^{X^2/k_4^2} \leq 2$; (v) $\mathbb{E}e^{\lambda X} \leq e^{k_5^2 \lambda^2}$, for all $\lambda \in \mathbb{R}$ (if $\mathbb{E}X = 0$).

  Proof (i) $\Rightarrow$ (ii): WLOG $k_1 = 1$. $\mathbb{E}|X|^p = \int_0^{+\infty} \mathbb{P}(|X| \geq t) p t^{p-1} \mathrm{d}t \leq \int_0^{+\infty} 2e^{-t^2} p t^{p-1} \mathrm{d}t = p\Gamma(\frac{p}{2}) \overset{\Gamma(x) \leq 3x^x \text{ for } x \geq \frac{1}{2}}{\leq} 3p(\frac{p}{2})^{p/2} \Rightarrow ||X||_p \leq \frac{1}{\sqrt{2}}(3p)^{1/p} p^{1/2} \leq 3\sqrt{p}$.

  (ii) $\Rightarrow$ (iii): WLOG $k_2 = 1$. $\mathbb{E}e^{\lambda^2 X^2} = \mathbb{E}[1 + \sum_{p=1}^{+\infty} \frac{(\lambda^2 X^2)^p}{p!}]$. $\mathbb{E}|X|^{2p} \leq (2p)^p, p! \geq (\frac{p}{e})^p \Rightarrow \mathbb{E}e^{\lambda^2 X^2} \leq 1 + \sum_{p=1}^{+\infty} \frac{(2\lambda^2 p)^p}{(\frac{p}{e})^p} = \frac{1}{1 - 2e\lambda^2}$ (if $2e\lambda^2 < 1$) $\overset{\frac{1}{1-x} \leq e^{2x} \text{ for } x \in [0, \frac{1}{2}]}{\leq} e^{4e\lambda^2}$ (if $2e\lambda^2 \leq 1/2 \Leftrightarrow \lambda \leq \frac{1}{2\sqrt{e}}$).

  (iii) $\Rightarrow$ (iv): trivial.

  (iv) $\Rightarrow$ (i): $\mathbb{P}(|X| \geq t) = \mathbb{P}(e^{X^2} \leq e^{t^2}) \leq e^{-t^2}\mathbb{E}e^{X^2} \leq 2e^{-t^2}$.

  (iii) $\Rightarrow$ (v): WLOG $k_3 = 1$. If $|\lambda| \leq 1$, then $\mathbb{E}e^{\lambda X} \leq \mathbb{E}(\lambda X + e^{\lambda^2 X^2}) = \mathbb{E}e^{\lambda^2 X^2} \leq e^{\lambda^2}$. If $|\lambda| \geq 1$, then $\mathbb{E}e^{\lambda X} \leq e^{\frac{\lambda^2}{2}} \mathbb{E}e^{\frac{X^2}{2}} e^{\frac{\lambda^2}{2}} e^{\frac{1}{2}} \leq e^{\lambda^2}$.

  (v) $\Rightarrow$ (i): mimic the proof of (iv) $\Rightarrow$ (i). □

- Sub-gaussian r.v.: satisfy the above sub-gaussian properties. $||X||_{\psi_2} = \inf\{t > 0 : \mathbb{E}e^{\lambda^2/t^2} \leq 2\}$. Thus $\mathbb{P}(|X| \geq t) \leq 2e^{-ct^2/||X||_{\psi_2}^2}$; $||X||_{L^p} \leq C||X||_{\psi_2}\sqrt{p}$; if $\mathbb{E}X = 0$ then $\mathbb{E}e^{\lambda X} \leq e^{C\lambda^2 ||X||_{\psi_2}^2}$.

- Let $X_1, \cdots, X_N$ be i.i.d. and mean zero sub-gaussian, then $||\sum_{i=1}^N X_i||_{\psi_2}^2 \leq C \cdot \sum_{i=1}^N ||X_i||_{\psi_2}^2$.

  Proof $\forall \lambda \in \mathbb{R}, \mathbb{E}e^{\lambda \sum_{i=1}^N X_i} = \prod_{i=1}^N \mathbb{E}e^{\lambda X_i} \leq \prod_{i=1}^N e^{c\lambda^2 ||X_i||_{\psi_2}^2} = e^{c\lambda^2 \sum_{i=1}^n ||X_i||_{\psi_2}^2}$ □

- Centering: X is sub-gaussian $\Rightarrow X - \mathbb{E}X$ is sub-gaussian and $||X - \mathbb{E}X||_{\psi_2} \leq C||X||_{\psi_2}$.

  Proof $||\mathbb{E}X||_{\psi_2} \leq C_1|\mathbb{E}X| \leq C_1\mathbb{E}|X| = C_1||X||_{L^1} \leq C_1C_2||X||_{\psi_2}$. □

- Sub-exponential properties: The following are equivalent: (1) $\mathbb{P}(|X| \geq t) \leq 2e^{-t/k_1}, \forall t \geq 0$; (2) $||X||_{L^p} \leq k_2 p, p \geq 1$; (3) $\mathbb{E}e^{\lambda|X|} \leq e^{k_3\lambda}$ for all $0 \leq \lambda \leq \frac{1}{k_3}$; (4) $\mathbb{E}e^{|X|/k_4} \leq 2$; (5) if $\mathbb{E}X = 0, \mathbb{E}e^{\lambda X} \leq e^{k_5^2\lambda^2}$ for $|\lambda| \leq \frac{1}{k_5}$.

  Proof (2) $\Rightarrow$ (5): $k_2 = 1, \mathbb{E}e^{\lambda X} = 1 + \sum_{p=2}^{+\infty} \frac{\lambda^p \mathbb{E}X^p}{p!} \leq 1 + \sum_{p=2}^{+\infty} \frac{\lambda^p p^p}{(p/e)^p} = 1 + \frac{(e\lambda)^2}{1 - e\lambda}(|e\lambda| < 1)$. If $|e\lambda| \leq \frac{1}{2}, 1 + \frac{(e\lambda)^2}{1 - e\lambda} \leq 1 + 2e^2\lambda^2 \leq e^{2e^2\lambda^2} \leq e^{4e^2\lambda^2}$, i.e. $k_5 = 2e$.

  (5) $\Rightarrow$ (1): $k_5 = 1, |x|^p \leq p^p(e^x + e^{-x}) \Rightarrow \mathbb{E}|X|^p \leq p^p(\mathbb{E}e^X + \mathbb{E}e^{-X}) \leq 2ep^p$. □

- $||X||_{\psi_1} = \inf\{t > 0 : \mathbb{E}e^{|X|/t} \leq 2\}$. X is sub-gaussian $\Leftrightarrow X^2$ is sub-exponential. $||X^2||_{\psi_1} = ||X||_{\psi_2}^2$.

- X, Y are sub-gaussian $\Rightarrow XY$ is sub-exponential and $||XY||_{\psi_1} \leq ||X||_{\psi_2}||Y||_{\psi_2}$.

  Proof WLOG $||X||_{\psi_2} = ||Y||_{\psi_2} = 1$. $\mathbb{E}e^{XY} \leq \mathbb{E}e^{\frac{X^2 + Y^2}{2}} = \mathbb{E}[e^{\frac{X^2}{2} + \frac{Y^2}{2}}] \leq \frac{1}{2}(\mathbb{E}e^{X^2} + \mathbb{E}e^{Y^2}) = 2$. □

- Orlicz function/space: $\psi : [0, +\infty) \to [0, +\infty)$, convex, increasing, $\psi(0) = 0$, $\psi(x) \to +\infty, x \to +\infty$. $||X||_\psi := \inf\{t > 0 : \mathbb{E}\psi(|X|/t) \leq 1\}$. $L_\psi := \{X : ||X||_\psi < +\infty\}$ is Banach space. Examples: (1) $L_p : \psi(x) = x^p, p \geq 1$; (2) $L_{\psi_2} : \psi_2(x) = e^{x^2} - 1, L_\infty \subset L_{\psi_2} \subset L_p$.

- Bernstein's inequality: $X_1, \cdots, X_N$ i.i.d., mean zero and sub-exponential. Then for $t \geq 0, \mathbb{P}(|\sum_{i=1}^N X_i| \geq t) \leq 2e^{-c\min(\frac{t^2}{\sum_{i=1}^N ||X_i||_{\psi_1}^2}, \frac{t}{\max_i ||X_i||_{\psi_1}})}$.

Proof $S = \sum_{i=1}^N X_i$. $\mathbb{P}(S \geq t) \leq e^{-\lambda t} \prod_{i=1}^N \mathbb{E}e^{\lambda X_i}$. $\mathbb{E}e^{\lambda X_i} \leq e^{c\lambda^2 ||X_i||_{\psi_1}^2}$ if $|\lambda| \leq \frac{c}{\max ||X_i||_{\psi_1}}$. Then $\mathbb{P}(S \geq t) \leq e^{-\lambda t + c\lambda^2\sigma^2}$ where $\sigma^2 := \sum_{i=1}^N ||X_i||_{\psi_1}^2$. The following is to find the minimum of a quadratic function with the restriction $|\lambda| \leq \frac{c}{\max ||X_i||_{\psi_1}}$. □

- Corollary 1: $\mathbb{P}(|\sum_{i=1}^N a_i X_i| \geq t) \leq 2e^{-c\min(\frac{t^2}{K^2||a||_2^2}, \frac{t}{K||a||_\infty})}$ where $K = \max_i ||X_i||_{\psi_1}$.
- Corollary 2: $|X_i| \leq K$, then $\mathbb{P}(|\sum_{i=1}^N X_i| \geq t) \leq 2\exp(-\frac{t^2/2}{\sigma^2 + Kt/3})$ where $\sigma^2 = \sum_{i=1}^N \mathbb{E}X_i^2$.

3 Random vectors in high dimensions

- $X \in \mathbb{R}^n$, independent sub-gaussian coordinate X_i, $\mathbb{E}X_i^2 = 1$. Then $\|\|X\|_2 - \sqrt{n}\|_{\psi_2} \leq CK^2, K = \max_i ||X_i||_{\psi_2}$.

  Proof $\mathbb{E}X_i^2 = 1 \Rightarrow K \geq 1$. $||X_i^2 - 1||_{\psi_1} \leq C||X_i^2||_{\psi_1} = C||X_i||_{\psi_2}^2 \leq CK^2$. Bernstein's inequality: $\mathbb{P}(|\frac{1}{n}||X||_2^2 - 1| \geq u) = \mathbb{P}(|\frac{1}{n}\sum_{i=1}^n (X_i^2 - 1)| \geq u) \leq 2e^{-cn\min(\frac{u^2}{K^4}, \frac{u}{K^2})} \leq 2e^{-\frac{cn}{K^4}\min(u^2, u)}$. For any $\delta > 0$, $\mathbb{P}(|\frac{1}{\sqrt{n}}||X||_2 - 1| \geq \delta) \leq \mathbb{P}(|\frac{1}{n}||X||_2^2 - 1| \geq \max(\delta, \delta^2)) \leq 2e^{-\frac{cn}{K^4}\delta^2} \Rightarrow \mathbb{P}(|\|X\|_2 - \sqrt{n}| \geq t) \leq 2e^{-ct^2/K^4}$. □

- Isotropy: $\Sigma(X) = \mathbb{E}XX^T = I$. If $\Sigma \neq I_n$, then let $Z = \Sigma^{-1/2}X$. X is isotropic $\Leftrightarrow \mathbb{E}\langle X, x\rangle^2 = ||x||_2^2$ for any $x \in \mathbb{R}^n$.

  Proof $\mathbb{E}\langle X, x\rangle^2 = \mathbb{E}(x^T XX^T x) = x^T(\mathbb{E}XX^T)x$. $||x||_2^2 = x^T I_n x$. $\Rightarrow \mathbb{E}XX^T = I_n$. □

- X is isotropic $\Rightarrow \mathbb{E}||X||_2^2 = n$. If X, Y are independent and isotropic $\Rightarrow \mathbb{E}\langle X, Y\rangle^2 = n$.

  Proof $\mathbb{E}||X||_2^2 = \mathbb{E}(X^T X) = \mathbb{E}(\mathrm{tr}(X^T X)) = \mathrm{tr}(\mathbb{E}XX^T) = n$.

  $\mathbb{E}\langle X, Y\rangle^2 = \mathbb{E}(X^T YY^T X) = \mathbb{E}(\mathrm{tr}(X^T YY^T X)) = \mathbb{E}(\mathrm{tr}(XX^T YY^T)) = \mathrm{tr}((\mathbb{E}XX^T)(\mathbb{E}YY^T)) = n$. □

- Examples: $X \sim U(\sqrt{n}\mathbb{S}^{n-1}), X \sim U(\{-1,1\}^n), X = (X_1, \cdots, X_n)$ i.i.d., $\mathbb{E}X_i = 0, \mathrm{Var}(X_i) = 1$ are all isotropic.
- $g \sim \mathcal{N}(0, I_n)$, then $\mathbb{P}(|\|g\|_2 - \sqrt{n}| \geq t) \leq 2e^{-ct^2}$.