Points for High-Dimensional Probability

Xiang Li

Peking University — July 20, 2020

Important lemmas

1. Prove the inequalities for all integers $m \in [1, n]$,
$$\left(\frac{n}{m}\right)^m \leq \binom{n}{m} \leq \sum_{k=0}^{m} \binom{n}{k} \leq \left(\frac{en}{m}\right)^m$$

2. Assume $g \sim N(0, 1)$, then prove that
   - For all $t > 0$, we have
$$\max\left\{\left(\frac{1}{t} - \frac{1}{t^3}\right)\frac{1}{\sqrt{2\pi}}e^{-t^2/2}, \frac{1}{2}\left(1 - \sqrt{1 - e^{-t^2}}\right)\right\} \leq \mathbb{P}(g \geq t) \leq \frac{1}{t}\frac{1}{\sqrt{2\pi}}e^{-t^2/2}$$
   - $\mathbb{E}g^2 1_{g>t} = t \cdot \frac{1}{\sqrt{2\pi}}e^{-t^2/2} + \mathbb{P}(g > t) \leq (t + \frac{1}{t}) \cdot \frac{1}{\sqrt{2\pi}}e^{-t^2/2}$;
   - $\|g\|_{L^p} = \sqrt{2}\left(\frac{\Gamma((1+p)/2)}{\Gamma(1/2)}\right)^{1/p} = O(\sqrt{p})$ as $p \to \infty$.

3. Here are some numerical inequalities. Prove them.
   - By Taylor expansion, $\cosh(x) \leq \exp(x^2/2)$ for all $x \in \mathbb{R}$.
   - By Stirling's approximation, $(p/e)^p \leq p! \leq p^p$ for $p \in Z_+$.
   - $|x|^p \leq p^p(e^x + e^{-x})$ holds for all $x \in \mathbb{R}$ and $p > 0$.
   - $\frac{1}{1-x} \leq e^{2x}$ holds for all $x \in [0, 1/2]$.

4. (Hoeffding's lemma) Let X be a random variable with $\mathbb{E}[X] = 0$ and $a \leq X \leq b$ with $b > a$. Then, for any $t > 0$, the following inequality holds:
$$\mathbb{E}e^{tX} \leq e^{\frac{t^2(b-a)^2}{8}}$$

5. For all number $z \geq 0$, if $|z - 1| \geq \delta$, then $|z^2 - 1| \geq \max(\delta, \delta^2)$. Conversely, if $|z^2 - 1| \geq \epsilon^2$, we have $|z - 1| \geq \min(\epsilon, \epsilon^2)$. And we have here $\epsilon = \max(\delta, \delta^2), \delta^2 = \min(\epsilon, \epsilon^2)$.

6. (Blow-up) Let A be a subset of the sphere $\sqrt{n}S^{n-1}$, $A_t = A + tB_2^n$ and let σ denote the normalized area on that sphere. If $\sigma(A) \geq 1/2$, then, for every $t \geq 0$,
$$\sigma(A_t) \geq 1 - 2exp(-ct^2)$$

7. (MGF of Gaussian chaos). Let $X, X^{'} \sim N(0, I_n)$ be independent and let $A = (a_{ij})$ be an $n \times n$ matrix. Then
$$\mathbb{E}\exp(\lambda X^T AX^{'}) \leq \exp(C\lambda^2\|A\|_F^2)$$
for all λ satisfying $|\lambda| \leq c\|A\|$.

8. (Comparison). Consider independent, mean zero, sub-gaussian random vectors $X, X^{'}$ in $\mathbb{R}^n$ with $\|X\|_{\Psi_2} \leq K$ and $\|X^{'}\|_{\Psi_2} \leq K$. Consider also independent random vectors $g, g^{'} \sim N(0, I_n)$. Let A be an $n \times n$ matrix. Then for all $\lambda \in \mathbb{R}$, we have
$$\mathbb{E}\exp(\lambda X^T AX^{'}) \leq \mathbb{E}\exp(CK^2\lambda g^T Ag^{'})$$

9. (Multivariate Gaussian integration by parts). Let $X \sim N(0, \Sigma)$. Then for any differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ we have
$$\mathbb{E}Xf(X) = \Sigma \cdot \mathbb{E}\nabla f(X)$$

Chapter 2

1. First, write down the five equivalent properties a sub-gaussian random variable satisfies. Then, write down the five equivalent properties a sub-exponential random variable satisfies.

2. Define
$$\|X\|_{\Psi_2} = \inf\{t > 0 : \mathbb{E}\exp(X^2/t^2) \le 2\}$$
check that $\|\cdot\|_{\Psi_2}$ is indeed a norm on the space of sub-gaussian random variables. Do the same thing for $\|\cdot\|_{\Psi_1}$ on the space of sub-exponential random variables, where
$$\|X\|_{\Psi_1} = \inf\{t > 0 : \mathbb{E}\exp(|X|/t) \le 2\}$$

3. A random variable X is sub-gaussian if and only if X^2 is sub-exponential, what's more,
$$\|X^2\|_{\Psi_1} = \|X\|_{\Psi_2}^2$$

4. Let $X_1, \cdots, X_n$ be independent, mean zero, sub-gaussian random variables. Then $Y = \sum_{i=1}^n X_i$ is also a sub-gaussian random variable, and
$$\|Y\|_{\Psi_2} \precsim \sqrt{\sum_{i=1}^n \|X_i\|_{\Psi_2}^2}$$

5. Let $X_1, \cdots, X_n$ be a sequence of sub-gaussian random variables (which may not necessarily independent), show that
$$\mathbb{E}\max_i \frac{|X_i|}{\sqrt{1 + log(i)}} \precsim K$$
where $K = \max_i \|X_i\|_{\Psi_2}$. Then deduce that for every $n \ge 2$, we have
$$\mathbb{E}\max_i |X_i| \precsim K\sqrt{\log(n)}$$
further prove that this bound is tight when $X_1, \cdots, X_n$ are independent $N(0,1)$.

6. (Hoeffding's inequality, two-sided) Let $X_1, \cdot, X_n$ be independent, mean zero, sub-gaussian random variables, and $a = (a_1, \cdots, a_n)$ in $\mathbb{R}^n$. Then for every $t \ge 0$, we have,
$$\mathbb{P}\left(|\sum_{i=1}^n a_i X_i| \ge t\right) \le 2\exp\left(-\frac{ct^2}{K^2\|a\|_2^2}\right)$$
where $K = \max_i \|X_i\|_{\Psi_2}$.

7. (Chernoff's inequality). Let X_i be independent Bernoulli random variables with parameters p_i. Consider their sum $S_n = \sum_{i=1}^n X_i$ and denote its mean by$\mu = \mathbb{E}Sn$. Then
   - For any $t > \mu$, we have
   $$\mathbb{P}(S_n \ge t) \le e^{-\mu}\left(\frac{e\mu}{t}\right)^t$$
   - For any $t < \mu$, we have
   $$\mathbb{P}(S_n \le t) \le e^{-\mu}\left(\frac{e\mu}{t}\right)^t$$
   - For $\delta \in (0,1]$, we have
   $$\mathbb{P}(|S_n - \mu| \le \delta\mu) \le 2e^{-c\mu\delta^2}$$

8. (Bernsteins's inequality) Let $X_1, \cdot, X_n$ be independent, mean zero, sub-exponential random variables, and $a = (a_1, \cdots, a_n)$ in $\mathbb{R}^n$. Then for every $t \ge 0$, we have,
$$\mathbb{P}\left(|\sum_{i=1}^n a_i X_i| \ge t\right) \le 2\exp\left(-c\min\left(\frac{t^2}{K^2\|a\|_2^2}, \frac{t}{K\|a\|_\infty}\right)\right)$$
where $K = \max_i \|X_i\|_{\Psi_1}$.

9. (McDiarmid's inequality). Let $X_1, \cdots, X_n$ be independent random variables Let $f : \mathbb{R}^n ßR$ be a measurable function. Assume that the value of $f(x)$ can change by at most $c_i > 0$ under an arbitrary change of a single coordinate of $x - \mathbb{R}^n$, i.e. for any index i and any $x_1, x_2, \cdots, x_n, x_i^{'}$, we have

$$|f(x_1, x_2, \cdot, x_{i-1}, x_i, x_{i+1} \cdots, x_n) - f(x_1, x_2, \cdot, x_{i-1}, x_i^{'}, x_{i+1} \cdots, x_n)| \leq c_i$$

Then, for any t > 0, we have

$$\mathbb{P}\left(f(X) - \mathbb{E}f(X) \geq t\right) \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^n c_i^2}\right)$$

where $X = (X_1, \cdots, X_n)$.

10. (Bennett's inequality) Let $X_1, \cdots, X_n$ be independent random variables. Assume that $|X_i - \mathbb{E}X_i| \leq K$ a.s for every i. Then for every $t \geq 0$, we have,

$$\mathbb{P}\left(\sum_{i=1}(X_i - \mathbb{E}X_i) \geq t\right) \leq \exp\left(-\frac{\sigma^2}{K^2} h\left(\frac{Kt}{\sigma^2}\right)\right)$$

where $\sigma^2 = \sum_{i=1}^n Var(X_i)$ and $h(u) = (1+u)\log(1+u) - u$.

11. What's Orlicz spaces? Does L^p space belong to it?

Chapter 3

1. White down the definition of sub-gaussian random vectors, and give three different examples of sub-gaussian random vectors.

2. (Concentration of the norm). Let$X = (X_1, \cdots, X_n) \in \mathbb{R}^n$ be a random vector with independent, sub-gaussian coordinates Xi that satisfy $\mathbb{E}X_i^2 = 1$. Then

$$\left\| \|X\|_2 - \sqrt{n} \right\|_{\Psi_2} \precsim K^2$$

where $K = \max_i \|X_i\|_{\Psi_2}$.

3. (Sub-gaussian distributions with independent coordinates) Let $X = (X_1, \cdots, X_n) \in \mathbb{R}^n$ be a random vector with independent, mean zero, sub-gaussian coordinate X_i. Then X is a sub-gaussian random vector and

$$\|X\|_{\Psi_2} \precsim \max_i \|X_i\|_{\Psi_2}$$

However, if we remove the condition of independence, find an example of a random vector X with $\|X\|_{\Psi_2} \gg \max_i \|X_i\|_{\Psi_2}$.

4. Let the random vector $X \in \mathbb{R}^n$ is the uniform distribution in the set $\{\sqrt{n}e_i, i = 1, \cdots, n\}$, show that

$$\|X\|_{\Psi_2} \asymp \sqrt{\frac{n}{\log(n)}}$$

5. Let X be an isotropic random vector supported in a finite set $T \subset \mathbb{R}^n$. Show that in order for X to be sub-gaussian with $\|X\|_{\Psi_2} = O(1)$, the cardinality of the set must be exponentially large in n:

$$|T| \geq e^{cn}$$

We can see that discrete distributions do not make nice sub-gaussian distributions, unless they are supported on exponentially large sets.

6. Consider a ball of the l_1 norm in $\mathbb{R}^n$:

$$K := \{x \in \mathbb{R}^n : \|x\|_1 \leq r\}$$

   - Show that the uniform distribution on K is isotropic for $r \sim n$.

- When $r \sim n$, show that this distribution is not sub-gaussian.

7. (Grothendieck's inequality). Consider an $m \times n$ matrix (a_{ij}) of real numbers. Assume that, for any numbers $x_i, y_j \in \{-1, 1\}$, we have
$$|\sum_{i,j} a_{ij} x_i y_j| \leq 1$$
Then, for any Hilbert space H and any vectors $u_i, v_j \in H$ satisfying $\|u_i\| = \|vj\| = 1$, we have
$$|\sum_{i,j} a_{ij} \langle u_i, v_j \rangle| \leq K$$
where $K \leq 1.783$ is an absolute constant.

Chapter 4

1. Let (T, d) be a metric space. Consider a subset $K \subset T$ and let $\epsilon > 0$, give the definition of the packing number of K denoted as $\mathcal{P}(K, d, \epsilon)$ and the covering number of K denoted as $\mathcal{N}(K, d, \epsilon)$. Then prove
$$\mathcal{P}(K, d, 2\epsilon) \leq \mathcal{N}(K, d, \epsilon) \leq \mathcal{P}(K, d, \epsilon)$$

2. Let K is a subset of $\mathbb{R}^n$, $\epsilon > 0$, then
$$\frac{|K|}{|\epsilon B_2^n|} \leq \mathcal{N}(K, \epsilon) \leq \mathcal{P}(K, \epsilon) \leq \frac{|K + (\epsilon/2) B_2^n|}{|(\epsilon/2) B_2^n|}$$
where $|\cdot|$ denotes the volume in $\mathbb{R}^n$, B_2^n denotes the unit Euclidean ball in $\mathbb{R}^n$, ϵB_2^n is a Euclidean ball with radius ϵ.

3. (Norm of matrices with sub-gaussian entries). Let A be an $m \times n$ random matrix whose entries A_{ij} are independent, mean zero, sub-gaussian random variables. Then, for any $t > 0$ we have
$$\|A\| \precsim K(\sqrt{m} + \sqrt{n} + t)$$
with probability at least $1 - 2exp(-t^2)$. Here $K = \max_{ij} \|A_{ij}\|_{\Psi_2}$.

4. Let A be an $m \times n$ random matrix whose entries A_i are independent, mean zero, sub-gaussian isotropic random vectors. Then, for any $t > 0$ we have,
$$\sqrt{m} - CK^2(\sqrt{n} + t) \leq s_n(A) \leq s_1(A) \leq \sqrt{m} + CK^2(\sqrt{n} + t)$$
with probability at least $1 - 2exp(-t^2)$. Here $K = \max_i \|A_i\|_{\Psi_2}$.

5. (Covariance estimation). Let X be a sub-gaussian random vector in $\mathbb{R}^n$. More precisely, assume that there exists $K \geq 1$ such that $\|\langle X, x\rangle\|_{\Psi_2} \leq K \|\langle X, x\rangle\|_{L^2}$ for any x in $\mathbb{R}$. Then, for every positive integer m, we have
$$\mathbb{E}\,\|\Sigma_m - \Sigma\| \precsim K^2 \left(\sqrt{\frac{n}{m}} + \frac{n}{m}\right) \|\Sigma\|$$

Chapter 5

1. (Concentration of Lipschitz functions) Consider a Lipschitz function $f : \mathcal{X} \to \mathbb{R}$ and a random vector $X \in \mathcal{X}$, prove following results:
   - If $\mathcal{X} = \sqrt{n} S^{n-1}$ and $X \sim Unif(\sqrt{n} S^{n-1})$, then $\|f(X) - \mathbb{E} f(X)\|_{\Psi_2} \precsim \|f\|_{Lip}$;
   - If $\mathcal{X} = S^{n-1}$ and $X \sim Unif(S^{n-1})$, then $\|f(X) - \mathbb{E} f(X)\|_{\Psi_2} \precsim \|f\|_{Lip}/\sqrt{n}$;
   - If $\mathcal{X} = [0, 1]^n$ and $X \sim Unif([0, 1]^n)$, then $\|f(X) - \mathbb{E} f(X)\|_{\Psi_2} \precsim \|f\|_{Lip}$;
   - If $\mathcal{X} = \sqrt{n} B_2^n$ and $X \sim Unif(\sqrt{n} B_2^n)$, then $\|f(X) - \mathbb{E} f(X)\|_{\Psi_2} \precsim \|f\|_{Lip}$;
   - If $\mathcal{X} = \mathbb{R}^n$ and $X \sim N(0, I_n)$, then $\|f(X) - \mathbb{E} f(X)\|_{\Psi_2} \precsim \|f\|_{Lip}$.

2. State and prove Johnson-Lindenstrauss Lemma.

3. (Matrix Bernstein's inequality) Let $X_1, \cdots, X_n$ be independent, mean zero, $n \times n$ symmetric random matrices, such that $\|X_i\| \le K$ almost surely for all i. Then, for every t $\ge$ 0, we have,

$$\mathbb{P}\left(\left\|\sum_{i=1} X_i\right\| \ge t\right) \le 2n \cdot \exp\left(-\frac{t^2/2}{\sigma^2 + Kt/3}\right)$$

Here $\sigma^2 = \left\|\sum_{i=1}^n \mathbb{E}X_i^2\right\|$ is the norm of the matrix variance of the sum.

4. (Matrix Bernstein's inequality: expectation). Let $X_1, \cdots, X_n$ be independent, mean zero, $n \times n$ symmetric random matrices, such that $\|X_i\| \le K$ almost surely for all i. Deduce from Bernstein's inequality that

   - With probability at least $1 - 2exp(-u)$, we have
   $$\mathbb{E}\left\|\sum_{i=1}^n X_i\right\| \precsim \sigma\sqrt{\log(n) + u} + K(\log(n) + u)$$
   - Check that
   $$\mathbb{E}\left\|\sum_{i=1}^n X_i\right\| \precsim \sigma\sqrt{\log(n)} + K\log(n)$$

   where $\sigma^2 = \left\|\sum_{i=1}^n \mathbb{E}X_i^2\right\|$ is the norm of the matrix variance of the sum.

5. Let A be an $m \times n$ random matrix whose rows A_i are independent, isotropic random vectors in $\mathbb{R}^n$. Assume that for some $K > 0$, $\|A_i\|_2 \le K\sqrt{n}$ almost surely for every i. Prove that for any $t > 1$ we have,
$$\sqrt{m} - Kt\sqrt{nlog(n)} \le s_n(A) \le s_1(A) \le \sqrt{m} + Kt\sqrt{nlog(n)}$$
with probability at least $1 - 2n^{-ct^2}$. (Hint: use Matrix Bernstein's inequality.)

6. (Matrix Hoeffding's inequality). Let $\epsilon_1, \cdots, \epsilon_n$ be independent symmetric Bernoulli random variables and let $A_1, \cdots, A_n$ be symmetric $n \times n$ matrices (deterministic). Prove that, for any $t \ge 0$, we have
$$\mathbb{P}\left(\left\|\sum_{i=1}^n \epsilon_i A_i\right\| \ge t\right) \le 2n \cdot \exp(-t^2/2\sigma^2)$$
where $\sigma^2 = \left\|\sum_{i=1}^n A_i^2\right\|$. (Hint: use $\mathbb{E}\exp(\lambda\epsilon_i A_i) \le \exp(\lambda^2 A_i^2/2)$.)

7. (General covariance estimation). Let X be a random vector in $\mathbb{R}^n$. Assume that for some $K \ge 1$, $\|X\|_2 \le K(\mathbb{E}\|X\|_2^2)^{1/2}$ almost surely. Then, for every positive integer m, we have
$$\mathbb{E}\left\|\Sigma_m - \Sigma\right\| \precsim \left(\sqrt{\frac{K^2nlog(n)}{m}} + \frac{K^2nlog(n)}{m}\right)\|\Sigma\|$$

Chapter 6

1. (Decoupling in normed space) Let $(u_{ij})_{i,j=1}^n$ be fixed vectors in some normed space. Let $X_1, \cdots, X_n$ be independent, mean zero random variables. Show that, for every convex function F , one has
$$\mathbb{E}F\left(\left\|\sum_{i,j,i\neq j} X_i X_j u_{i,j}\right\|\right) \le \mathbb{E}F\left(4\left\|\sum_{i,j} X_i X_j^{'} u_{i,j}\right\|\right)$$
where $\left(X_i^{'}\right)$ is an independent copy of X_i.

2. (Hanson-Wright inequality) Let $X = (X_1, \cdots, X_n) \in \mathbb{R}^n$ be a random vector with independent, mean zero, sub-gaussian coordinates. Let A be an $n \times n$ matrix. Then, for every $t \geq 0$, we have

$$\mathbb{P}\left(|X^\top AX - \mathbb{E}X^\top AX| \geq t\right) \leq 2\exp\left(-\min\left(\frac{t^2}{K^4\|A\|_F^2}, \frac{t}{K^2\|A\|}\right)\right)$$

where $K = \max_i \|X_i\|_{\Psi_2}$.

3. (Symmetrization) Let $X_1, \cdots, X_n$ be independent, mean zero random vectors in a normed space and ϵ_i are independent symmetric Bernoulli. Then

$$\frac{1}{2}\mathbb{E}\left\|\sum_{i=1}^n \epsilon_i X_i\right\| \leq \mathbb{E}\left\|\sum_{i=1}^n X_i\right\| \leq 2\mathbb{E}\left\|\sum_{i=1}^n \epsilon_i X_i\right\|$$

4. Let A be an $m \times n$ random matrix whose entries are independent, mean zero random variables. Show that

$$\mathbb{E}\|A\| \lesssim \sqrt{\log(m+n)}\left(\mathbb{E}\max_i \|A_i\|_2 + \mathbb{E}\max_j \|A^j\|_2\right)$$

where A_i and A^j denote the rows and columns of A, respectively.

5. (Contraction principle) Let $x_1, \cdots, x_n$ be (deterministic) vectors in some normed space, ϵ_i are independent symmetric Bernoulli, and let $a = (a_1, \cdots, a_n) \in \mathbb{R}^n$. Then

$$\mathbb{E}\left\|\sum_{i=1}^n a_i\epsilon_i x_i\right\| \leq \|a\|_\infty \mathbb{E}\left\|\sum_{i=1}^n \epsilon_i x_i\right\|$$

6. (Symmetrization with gaussian) Let $X_1, \cdots, X_n$ be independent, mean zero random vectors in a normed space. Let $g_1, \cdots, g_n \sim N(0,1)$ be independent Gaussian random variables, which are also independent of X_i. Then

$$\frac{c}{\sqrt{\log(n)}}\mathbb{E}\left\|\sum_{i=1}^n g_i X_i\right\| \leq \mathbb{E}\left\|\sum_{i=1}^n X_i\right\| \leq 3\mathbb{E}\left\|\sum_{i=1}^n g_i X_i\right\|$$

Further prove the factor $\sqrt{\log(n)}$ in the lower bound is optimal.

7. (Rademecher complexity) Let G be a family of functions mapping from Z to $[a, b]$ and $S = (z_1, \cdots, z_m)$ a fixed sample of size m with elements in Z. Then, the empirical Rademacher complexity of G with respect to the sample S is defined as:

$$\widehat{\mathcal{R}}_S(G) = \mathbb{E}_\sigma\left[\frac{1}{m}\sup_{g\in G}\sum_{i=1}^m \sigma_i g(z_i)\right]$$

where σ_i are independent symmetric Bernoulli. Let D denote the distribution according to which samples are drawn. Now for any sample $S = (z_1, \cdots, z_m)$ and any $g \in G$, we denote by $\widehat{\mathbb{E}}_S[g]$ the empirical average of g over S i.e. $\widehat{\mathbb{E}}_S[g] = \frac{1}{m}\sum_{i=1}^n g(z_i)$. Define the function Φ over S as $\Phi(S) = \sup_{g\in G}\mathbb{E}[g] - \widehat{\mathbb{E}}_S[g]$, prove that

$$\mathbb{E}_{S\sim D^m}[\Phi(S)] \leq 2\mathbb{E}_{S\sim D^m}\widehat{\mathcal{R}}_S(G)$$

(Hint: apply decoupling and symmetrization to upper bound $\mathbb{E}_{S\sim D^m}[\Phi(S)]$.)

Chapter 7

1. (Sudakov-Fernique's inequality) Let $(X_t)_{t\in T}$ and $(Y_t)_{t\in T}$ be two mean zero Gaussian processes. Assume that for all $t, s \in T$, we have

$$\mathbb{E}(X_t - X_s)^2 \leq \mathbb{E}(Y_t - Y_s)^2$$

Then

$$\mathbb{E}\sup_{t\in T} X_t \leq \mathbb{E}\sup_{t\in T} Y_t$$

2. (Gordon's inequality) Let $(X_{ut})_{u\in U, t\in T}$ and $Y = (Y_{ut})_{u\in U, t\in T}$ be two mean zero Gaussian processes indexed by pairs of points (u, t) in a product set $U \times T$. Assume that we have

$$\begin{aligned} \mathbb{E}X_{ut}^2 &= \mathbb{E}Y_{ut}^2 \ \forall u,\ t \\ \mathbb{E}(X_{ut} - X_{us})^2 &\leq \mathbb{E}(Y_{ut} - Y_{us})^2 \ \forall\ u,\ t,\ s \\ \mathbb{E}(X_{ut} - X_{vs})^2 &\geq \mathbb{E}(Y_{ut} - Y_{vs})^2 \ \forall u \neq v, \forall\ t,\ s \end{aligned}$$

Then for every $\tau \geq 0$ we have

$$\mathbb{P}\left(\inf_{u\in U}\sup_{t\in T} X_{ut} \geq \tau\right) \leq \mathbb{P}\left(\inf_{u\in U}\sup_{t\in T} Y_{ut} \geq \tau\right)$$

Consequently,

$$\mathbb{E}\inf_{u\in U}\sup_{t\in T} X_{ut} \leq \mathbb{E}\inf_{u\in U}\sup_{t\in T} Y_{ut}$$

3. Let A be an $m \times n$ matrix with independent $N(0, 1)$ entries. Then

$$\sqrt{m} - \sqrt{n} \leq \mathbb{E}s_n(A) \leq \mathbb{E}s_1(A) \leq \sqrt{m} + \sqrt{n}$$

(Hint: use results in previous two problems.)

4. (Sudakov's minoration inequality) Let $(X_t)_{t\in T}$ be a mean zero Gaussian process. Then, for any $\epsilon \geq 0$, we have

$$\epsilon\sqrt{\log \mathcal{N}(T, d, \epsilon)} \lesssim \mathbb{E}\sup_{t\in T} X_t$$

5. Given a set $T \in \mathbb{R}^n$, give the definition of $diam(T), rad(T), w(T), w_s(T), h(T), d(T)$ and $\gamma(T)$, and then prove the following inequalities:

   - $\frac{1}{\sqrt{2\pi}} diam(T) \leq w(T) \leq \frac{\sqrt{n}}{2} diam(T)$;
   - $(\sqrt{n} - C)w_s(T) \leq w(T) \leq (\sqrt{n} + C)w_s(T)$;
   - $\gamma(T - T) = 2w(T)$;
   - $\frac{1}{3}\left[w(T) + \|y\|_2\right] \leq \gamma(T) \leq 2\left[w(T) + \|y\|_2\right]$ for arbitrary point $y \in T$;
   - $w(T - T) \leq h(T - T) \leq w(T - T) + C diam(T) \lesssim w(T - T)$;
   - $d(T) \leq dim(T)$.

6. Prove the following propositions:

   - $w(B_\infty^n) = \sqrt{\frac{2}{\pi}}n$;
   - $w(B_1^n) \approx \sqrt{\log(n)}$;
   - $w(S^{n-1}) = w(B_2^n) \approx \sqrt{n}$;
   - Let $p > 1$, check that $w(B_p^n) \lesssim \sqrt{q}n^{1/q}$, where q is the conjugate exponent for p, satisfying $1/p + 1/q = 1$;
   - Let T be a finite set of points in $\mathbb{R}^n$. Check that $w(T) \lesssim \sqrt{\log(|T|)} \cdot diam(T)$;
   - All results above can be extended to $w_s(\cdot)$. State them.

7. (Massrt's lemma) Let $A \subset \mathbb{R}^n$ be a finite set and $\sigma_1, \cdots, \sigma_n$ be independent symmetric Bernoulli, prove that

   - Use the techniques this chapter introduced to prove that

$$\mathbb{E}_\sigma\left[\frac{1}{m}\sup_{g\in G}\sum_{i=1}^{m}\sigma_i x_i\right] \leq C\frac{\sqrt{\log|A|} \cdot rad(A)}{n}$$

   - Further prove that $C \leq \sqrt{2}$ (Hint: you may use other techniques like Hoeffding's lemma to prove the above result again).

Chapter 8

We define a random process $(X_t)_{t\in T}$ on a metric space(T,d) with sub-gaussian increments if there exist some constant K and for all $t,s\in T$, it satisfies

$$\|X_t - X_s\| \le Kd(t,s)$$

1. (Dudley's integral inequality). Let $(X_t)_{t\in T}$ be a mean zero random process on a metric space(T,d) with sub-gaussian increments, then

$$\mathbb{E}\sup_{t\in T} X_t \precsim \int_0^{diam(T)} \sqrt{\log \mathcal{N}(T,d,\epsilon)}d\epsilon$$

   Here $diam(T) = \sup_{x,y\in T} d(x,y)$.

2. (Two-sided Sudakov's inequality). Let $T\subset \mathbb{R}^n$ and set

$$s(T) := \sup_{\epsilon\ge 0} \epsilon\sqrt{\log \mathcal{N}(T,\epsilon)}$$

   Then

$$s(T) \precsim w(T) \precsim \log(n)\cdot s(T)$$

3. (Dudley's integral inequality: tail bound). Let $(X_t)_{t\in T}$ be a random process on a metric space(T,d) with sub-gaussian increments. Then for every $u\ge 0$, the event

$$\sup_{t,s\in T} |X_t - X_s| \precsim K\left[\int_0^{\infty} \sqrt{\mathcal{N}(T,d,\epsilon)}d\epsilon + u\cdot diam(T)\right]$$

   holds with probability at least $1-2\exp(-u^2)$.

4. (VC dimension) Give the definition of the VC dimension of a class of subsets of Ω without mentioning any function. Then show that
   - For the class of all intervals on a line, the VC dimension is 2;
   - For the class of all rectangles on the plane (may not be axis-aligned), the VC dimension is 7;
   - For the class of all polygons with k vertices on the plane, the VC dimension is $2k+1$;
   - For the class of half-spaces in $\mathbb{R}^n$, the VC dimension is $n+1$.

5. (Pajor's Lemma) Let $\mathcal{F}$ be a class of Boolean functions on a finite set Ω. Then

$$|\mathcal{F}| \le |\{\Lambda\subset\Omega : \Lambda \ is\ shattered\ by\ \mathcal{F}\}|$$

6. (Sauer-Shelah Lemma) Let $\mathcal{F}$ be a class of Boolean functions on a finite set Ω. Set $d = vc(\mathcal{F})$, then

$$|\mathcal{F}| \le \sum_{k=1}^{d}\binom{n}{k} \le \left(\frac{en}{d}\right)^d$$

7. (Covering numbers via VC dimension). $\mathcal{F}$ be a class of Boolean functions ona probability space (Ω,Σ,μ). Define a metric on $\mathcal{F}$ as

$$d(f,g) = \|f-g\|_{L^2(\mu)} = \left(\int_\Omega |f-g|^2 d\mu\right)^{1/2} \ for\ any\ f,g\ \in\mathcal{F}$$

   Then for every $\epsilon\in(0,1)$, we have

$$\mathcal{N}\left(\mathcal{F}, L^2(\mu),\epsilon\right) \le \left(\frac{2}{\epsilon}\right)^{Cd}$$

   where $d = vc(\mathcal{F})$.

8. (Empirical processes via VC dimension). Let $\mathcal{F}$ be a class of Boolean functions on a probability space (Ω, Σ, μ) with finite VC dimension $vc(\mathcal{F}) \geq 1$. Let $X, X_1, X_2, \cdots, X_n$ be independent random points in Ω distributed according to the law μ. Then

$$\mathbb{E} \sup_{f \in \mathcal{F}} |\frac{1}{n} \sum_{i=1}^{n} f(X_i) - \mathbb{E} f(X)| \precsim \sqrt{\frac{vc(\mathcal{F})}{n}}$$

9. (VC-dimension generalization bounds) Let H be a family of functions taking values in $\{-1, +1\}$ with VC-dimension d. Then, for any $\delta > 0$, with probability at least $1-\delta$, the following holds for all $h \in H$:

$$R(h) \leq \widehat{R}(h) + \sqrt{\frac{2dlog(\frac{em}{d})}{m}} + \sqrt{\frac{\log \frac{1}{\delta}}{2m}}$$

where $R(h) = \mathbb{E}_{(x,y)\sim D}[1_{h(x)\neq y}]$ and $\widehat{R}(h) = \frac{1}{m} \sum_{i=1}^{m} 1_{h(x_i)\neq y_i}$. (Hint: use previous conclusion, Sauer-Shelah Lemma and McDiarmid's inequality.)

10. Give the definition of Talagrand's γ_2 functional, and prove

$$\gamma_2(T, d) \precsim \int_0^{\infty} \sqrt{\log \mathcal{N}(T, d, \epsilon)} d\epsilon$$

11. (Talagrand's comparison inequality) Let $(X_t)_{t\in T}$ be a mean zero random process on a set T and let$(Y_t)_{t\in T}$ be a Gaussian process. Assume that for all $t, s \in T$, we have

$$\|X_t - X_s\|_{\Psi_2} \leq K\|Y_t - Y_s\|_2$$

Then

$$\mathbb{E} \sup_{t\in T} X_t \precsim K \mathbb{E} \sup_{t\in T} Y_t$$

12. Let $(X_x)_{x\in T}$ be a mean zero random process on a subset $T \subset \mathbb{R}^n$. Assume that for all $x, y \in T$, we have

$$\|X_x - X_y\|_{\Psi_2} \leq K\|x - y\|_2$$

Prove

- $\mathbb{E} \sup_{x\in T} X_x \precsim K w(T)$;
- $\mathbb{E} \sup_{x\in T} |X_x| \precsim K \gamma(T)$;
- For every $u \geq 0$, we have the event

$$\sup_{x\in T} |X_x| \precsim K(w(T) + u \cdot rad(T))$$

  holds with probability at least $1 - 2exp(-u^2)$;
- For $p > 1$, $(\mathbb{E} \sup_{x\in T} |X_x|^p)^{1/p} \precsim \sqrt{p} K \gamma(T)$.

13. (Sub-gaussian Chevet's inequality). Let A be an$m \times n$ random matrix whose entries A_{ij} are independent, mean zero, sub-gaussian random variables. Let $T \subset \mathbb{R}^n$ and $S \subset \mathbb{R}^m$ be arbitrary bounded sets and $K = \max_{i,j} \|A_{ij}\|_{\Psi_2}$. Then

$$\mathbb{E} \sup_{x\in T, y\in S} \langle Ax, y\rangle \leq CK\left[w(T)rad(S) + w(S)rad(T)\right]$$

If suppose the entries of A are $N(0, 1)$, then $CK = 1$.

Chapter 9, 11

1. (Matrix deviation inequality). Let A be an $m \times n$ matrix whose rows A_i are independent, isotropic and sub-gaussian random vectors in $\mathbb{R}^n$. Then for any subset $T \subset \mathbb{R}^n$, we have

$$\mathbb{E} \sup_{x \in T} |\|Ax\|_2 - \sqrt{m}\|x\|_2| \lesssim K^2 \gamma(T)$$

Here $\gamma(T)$ is the Gaussian complexity and $K = \max_i \|A_i\|_{\Psi_2}$.

2. (General matrix deviation inequality). Let A be an $m \times n$ Gaussian random matrix with i.i.d. $N(0,1)$ entries. Let f : $\mathbb{R}^m$ $\mathbb{R}$ be a positive-homogeneous and subadditive function, and let $b \in \mathbb{R}$ be such that $f(x) \leq b\|x\|_2$ for all $x \in \mathbb{R}^n$. Then for any subset $T \subset \mathbb{R}^n$, we have

$$\mathbb{E} \sup_{x \in T} |f(Ax) - \mathbb{E}f(Ax)| \lesssim b\gamma(T)$$

3. State and prove M^* bound and Escape theorem.

4. (Dvoretzky-Milman's theorem for Grassmanian) Let P be a random projection onto a random m-dimensional subspace in $\mathbb{R}^n$. $T \subset \mathbb{R}^n$ be a bounded set, and let $\epsilon \in (0,1)$.

   - Suppose $m \lesssim \epsilon^2 d(T)/n$, where $d(T)$ is the stable dimension of T. Then with probability at least 0.99, we have
   $$(1-\epsilon)B \subset conv(PT) \subset (1+\epsilon)B$$
   where B is a Euclidean ball with radius $w_s(T)$. Then $diam(PT) \approx w_s(T)$ when $m \lesssim d(T)$.
   - With probability $1 - 2e^{-m}$, we have
   $$diam(PT) \lesssim \left[w_s(T) + \sqrt{\frac{m}{n}} diam(T) \right]$$
   Thus if $m \gtrsim d(T)$, $diam(PT) \approx \sqrt{\frac{m}{n}} diam(T)$.
   - A random projection of a set T in $\mathbb{R}^n$ onto an m-dimensional subspace approximately preserves the geometry of T if $m \gtrsim d(T)$. For smaller m, the projected set PT becomes approximately a round ball of diameter $\sim w_s(T)$, and its size does not shrink with m.

Chapter 10

1. What's RIP condition? Explain its intuition.

2. Suppose the rows A_i of A are independent, isotropic and sub-gaussian random vectors, and let $K := \max_i \|A_i\|_{\Psi_2}$. Then the following happens with probability at least $1-2$ $exp(-cm/K^4)$. Assume an unknown signal $x \in \mathbb{R}^n$ is s-sparse and the number of measurements m satisfies
$$m \gtrsim K^4 s \log(n)$$
Then a solution $\widehat{x}$ of the following program is exact, i.e. $\widehat{x} = x$.
$$\min \|x\|_1 \; s.t. \; y = Ax$$