CS 58500 – Theoretical Computer Science Toolkit

Lecture 5 (02/03)

Concentration Inequality IV

https://ruizhezhang.com/course_spring_2026.html

Today's Lecture

- Subgaussian and Subgamma Random Variable
- The Entropy Method
- Talagrand's Inequality
- Applications: LIS and TSP

Subgaussian Random Variable

A random variable X is called σ^2-subgaussian if its log-MGF satisfies

$$\psi(\theta) \coloneqq \log \mathbb{E}\left[\exp\left(\theta(X - \mathbb{E}[X])\right)\right] \leq \theta^2\sigma^2/2 \qquad \forall \theta \in \mathbb{R}$$

We call σ^2 the variance proxy.

Equivalently,

- $\Pr[|X - \mathbb{E}[X]| \geq t] \leq 2e^{-t^2/2\sigma^2}$
- $\mathbb{E}[|X - \mathbb{E}[X]|^k] \leq \sigma^k k^{k/2}$ for any $k \in \mathbb{Z}_+$

Examples:

- True Gaussian random variable $\mathcal{N}(0, \sigma^2)$
- Bounded random variable: if $a \leq X \leq b$ a.s., then X is $(b-a)^2/4$-subgaussian

Subgaussian Random Variable

A random variable X is called σ^2-subgaussian if its log-MGF satisfies

$$\psi(\theta) \coloneqq \log \mathbb{E}\left[\exp\left(\theta(X - \mathbb{E}[X])\right)\right] \leq \theta^2\sigma^2/2 \qquad \forall\theta \in \mathbb{R}$$

We call σ^2 the variance proxy.

Equivalently,

- $\Pr[|X - \mathbb{E}[X]| \geq t] \leq 2e^{-t^2/2\sigma^2}$
- $\mathbb{E}[|X - \mathbb{E}[X]|^k] \leq \sigma^k k^{k/2}$ for any $k \in \mathbb{Z}_+$

Lemma. If X_1, X_2 are independent subgaussian random variables with variance proxy σ_1^2 and σ_2^2, then $X_1 + X_2$ is $(\sigma_1^2 + \sigma_2^2)$-subgaussian

➢ It immediately recovers the Hoeffding's inequality

Subgamma Random Variable

A random variable X is called (σ^2, c)-subgamma if

$$\psi(\theta) \leq \frac{\theta^2\sigma^2}{2(1-|\theta|c)} \leq \frac{\theta^2\sigma^2}{2} \qquad \forall|\theta| < 1/c$$

It holds that

$$\Pr[|X - \mathbb{E}[X]| \geq t] \leq 2\max\left\{e^{-\frac{t^2}{2\sigma^2}}, e^{-\frac{t}{2c}}\right\}$$

Examples:

- If $X \sim \mathcal{N}(0,1)$, then X^2 is $(4,3)$-subgamma
- If X is σ^2-subgaussian, then X is $(\sigma^2, 0)$-subgamma
- Bounded random variable: if $|X - \mathbb{E}[X]| \leq b$ a.s., then X is $(\mathrm{Var}[X], b/3)$-subgamma
- If X is (σ^2, c)-subgamma, then αX is $(\alpha^2\sigma^2, \alpha c)$-subgamma

Subgamma Random Variable

A random variable X is called (σ^2, c)-subgamma if

$$\psi(\theta) \le \frac{\theta^2\sigma^2}{2(1-|\theta|c)} \le \frac{\theta^2\sigma^2}{2} \qquad \forall|\theta| < 1/c$$

It holds that

$$\Pr[|X - \mathbb{E}[X]| \ge t] \le 2\max\left\{e^{-\frac{t^2}{2\sigma^2}}, e^{-\frac{t}{2c}}\right\}$$

Lemma. If X_1, X_2 are independent subgamma random variables with parameters (σ_1^2, c_1) and (σ_2^2, c_2), then $X_1 + X_2$ is $(\sigma_1^2 + \sigma_2^2, \max\{c_1, c_2\})$-subgamma

- It immediately recovers the Bernstein inequality

Today's Lecture

- Subgaussian and Subgamma Random Variable
- The Entropy Method
- Talagrand's Inequality
- Applications: LIS and TSP

The Entropy Method

The entropy of a random variable X is defined as

$$\mathrm{Ent}[X] \coloneqq \mathbb{E}[X \log X] - \mathbb{E}[X] \log \mathbb{E}[X]$$

Lemma (Herbst). Suppose that

$$\mathrm{Ent}\left[e^{\theta X}\right] \leq \frac{\theta^2 \sigma^2}{2} \mathbb{E}\left[e^{\theta X}\right] \quad \forall \theta \geq 0$$

Then, X is σ^2-subgaussian.

Tensorization of entropy

- For a function $f(x_1, \dots, x_n)$, and for each $i \in [n]$, define
$$\mathrm{Ent}_i[f(x)] \coloneqq \mathrm{Ent}[f(x_1, \dots, x_{i-1}, X_i, x_{i+1}, \dots, x_n)]$$
- For independent random variables $X_1, \dots, X_n$, we have
$$\mathrm{Ent}[f(X_1, \dots, X_n)] \leq \mathbb{E}\left[\sum_{i=1}^{n} \mathrm{Ent}_i[f(X_1, \dots, X_n)]\right]$$

Proof of Herbst Lemma

Proof.

- We'll verify that $\psi(\theta) \coloneqq \log \mathbb{E}\left[\exp\left(\theta(X - \mathbb{E}[X])\right)\right] \le \theta^2\sigma^2/2$
- $\psi(\theta) = \log \mathbb{E}\left[e^{\theta X}\right] - \theta\mathbb{E}[X]$

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\left(\frac{\psi(\theta)}{\theta}\right) = \frac{\mathbb{E}\left[Xe^{\theta X}\right]}{\theta\mathbb{E}[e^{\theta X}]} - \frac{\mathbb{E}[X]}{\theta} - \frac{\log \mathbb{E}\left[e^{\theta X}\right]}{\theta^2} + \frac{\mathbb{E}[X]}{\theta} = \frac{\mathbb{E}\left[Xe^{\theta X}\right]}{\theta\mathbb{E}[e^{\theta X}]} - \frac{\log \mathbb{E}\left[e^{\theta X}\right]}{\theta^2}$$

- $\mathrm{Ent}\left[e^{\theta X}\right] = \theta\mathbb{E}\left[Xe^{\theta X}\right] - \mathbb{E}\left[e^{\theta X}\right]\log \mathbb{E}\left[e^{\theta X}\right]$
- Thus, $\frac{\mathrm{d}}{\mathrm{d}\theta}\left(\frac{\psi(\theta)}{\theta}\right) = \frac{\mathrm{Ent}\left[e^{\theta X}\right]}{\theta^2\mathbb{E}\left[e^{\theta X}\right]} \le \frac{\sigma^2}{2}$ by assumption
- Then, we have

$$\frac{\psi(\theta)}{\theta} = \int_0^\theta \frac{\mathrm{Ent}[e^{\tau X}]}{\tau^2\mathbb{E}[e^{\tau X}]}\mathrm{d}\tau \le \int_0^\theta \frac{\sigma^2}{2}\mathrm{d}\tau = \frac{\theta\sigma^2}{2}$$

■

The Entropy Method

Lemma (Discrete Modified log-Sobolev (MLS) Inequality). Let $f: \mathbb{R} \to \mathbb{R}$ and let

$$D^- f(x) \coloneqq f(x) - \inf_y f(y)$$

Then for any random variable X,

$$\mathrm{Ent}\left[e^{f(X)}\right] \le \mathrm{Cov}\left[f(X), e^{f(X)}\right] \le \mathbb{E}\left[|D^- f(X)|^2 e^{f(X)}\right],$$

where $\mathrm{Cov}[X, Y] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$

Proof.

- For the first inequality,

$$\mathrm{Ent}\left[e^f\right] = \mathbb{E}\left[f e^f\right] - \mathbb{E}\left[e^f\right] \log \mathbb{E}\left[e^f\right] \le \mathbb{E}\left[f e^f\right] - \mathbb{E}\left[e^f\right]\mathbb{E}[f] = \mathrm{Cov}\left[f, e^f\right]$$

- For the second inequality,

$$\mathrm{Cov}\left[f, e^f\right] = \mathbb{E}\left[(f - \inf f)\left(e^f - \mathbb{E}\left[e^f\right]\right)\right] \le \mathbb{E}\left[(f - \inf f)\left(e^f - e^{\inf f}\right)\right]$$

- The convexity of e^x implies that $e^f - e^{\inf f} \le e^f (f - \inf f)$ ∎

The Entropy Method: Sharper Bounded Differences

Define one-sided differences for multivariate function:

$$D_i^- f(x) \coloneqq f(x_1, \dots, x_n) - \inf_z f(x_1, \dots, x_{i-1}, z, x_{i+1}, \dots, x_n)$$
$$D_i^+ f(x) \coloneqq \sup_z f(x_1, \dots, x_{i-1}, z, x_{i+1}, \dots, x_n) - f(x_1, \dots, x_n)$$

Theorem (Bounded differences inequality).

Let $X_1, \dots, X_n$ be independent random variables. Then, $f(X_1, \dots, X_n)$ is subgaussian with variance proxy $2\|\sum_{i=1}^n |D_i f|^2\|_\infty$. Moreover,

$$\Pr[f - \mathbb{E}[f] \geq t] \leq \exp\left(-\frac{t^2}{4\left\|\sum_{i=1}^n \left|D_i^- f\right|^2\right\|_\infty}\right)$$

$$\Pr[f - \mathbb{E}[f] \leq -t] \leq \exp\left(-\frac{t^2}{4\left\|\sum_{i=1}^n \left|D_i^+ f\right|^2\right\|_\infty}\right)$$

The Entropy Method: Sharper Bounded Difference

Theorem (Bounded differences inequality)**.**

Let $X_1, \ldots, X_n$ be independent random variables. Then, $f(X_1, \ldots, X_n)$ is subgaussian with variance proxy $2\|\sum_{i=1}^{n} |D_i f|^2\|_\infty$

Theorem (McDiarmid inequality)**.**

Let $X_1, \ldots, X_n$ be independent random variables. Then, $f(X_1, \ldots, X_n)$ is subgaussian with variance proxy $\frac{1}{4}\sum_{i=1}^{n} \|D_i f\|_\infty^2$

In many cases, $\|\sum_{i=1}^{n} |D_i f|^2\|_\infty$ can be much smaller than $\sum_{i=1}^{n} \|D_i f\|_\infty^2$

The Entropy Method: Sharper Bounded Difference

Theorem (Bounded difference inequality)**.**

Let $X_1, \dots, X_n$ be independent random variables. Then, $f(X_1, \dots, X_n)$ is subgaussian with variance proxy $2\|\sum_{i=1}^n |D_i f|^2\|_\infty$

Proof.

- By the discrete MLS lemma,

$$\mathrm{Ent}_i[e^f] \le \mathbb{E}[|D_i^- f|^2 e^f \,|\, X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n]$$

- By tensorization, for any $\theta \ge 0$,

$$\mathrm{Ent}[e^{\theta f}] \le \mathbb{E}\left[\sum_{i=1}^n \mathrm{Ent}_i[e^{\theta f}]\right] \le \mathbb{E}\left[\left(\sum_{i=1}^n |D_i^-(\theta f)|^2\right) e^{\theta f}\right] \le \theta^2 \left\|\sum_{i=1}^n |D_i^- f|^2\right\|_\infty \mathbb{E}[e^{\theta f}]$$

- We finish the proof by Herbst lemma.

■

Today’s Lecture

- Subgaussian and Subgamma Random Variable
- The Entropy Method
- Talagrand’s Inequality
- Applications: LIS and TSP

Talagrand's Inequality: Motivating Question

Let V be a fixed d-dimensional subspace. Let $\boldsymbol{x} \sim \mathrm{Unif}\{-1,1\}^n$. How well is $\mathrm{dist}(\boldsymbol{x}, V)$ concentrated?

- Let P be the orthogonal projection onto $V^\perp$. Then, $\mathrm{tr}[P] = \dim(V^\perp) = n - d$
- $\mathrm{dist}(\boldsymbol{x}, V)^2 = |P\boldsymbol{x} \cdot P\boldsymbol{x}| = |\boldsymbol{x}^\top P\boldsymbol{x}| = \sum_{i,j\in[n]} x_i x_j P_{ij}$
- Thus, $\mathbb{E}[\mathrm{dist}(x, V)^2] = \sum_{i\in[n]} P_{ii} = n - d$

How well is $\mathrm{dist}(\boldsymbol{x}, V)$ concentrated around $\sqrt{n-d}$?

- Consider $f(\boldsymbol{x}) \coloneqq \mathrm{dist}(\boldsymbol{x}, V)$ for $\boldsymbol{x} \in \{-1,1\}^n$
- For any $i \in [n]$, by triangle inequality,
$$|D_i f(\boldsymbol{x})| = |\mathrm{dist}(\boldsymbol{x}_{-i}, V) - \mathrm{dist}(\boldsymbol{x}, V)| \le \|\boldsymbol{x} - \boldsymbol{x}_{-i}\|_2 = 2$$
- By the bounded differences inequality, $\Pr\left[\left|\mathrm{dist}(\boldsymbol{x}, V) - \sqrt{n-d}\right| \ge t\right] \le 2e^{-2t^2/n}$
- Useless since $\mathrm{dist}(\boldsymbol{x}, V) \le \mathrm{dist}(\boldsymbol{x}, \boldsymbol{0}) = \sqrt{n}$

Talagrand’s Inequality: Motivating Question

Let V be a fixed d-dimensional subspace. Let $\boldsymbol{x} \sim \mathrm{Unif}\{-1,1\}^n$. How well is $\mathrm{dist}(\boldsymbol{x}, V)$ concentrated?

- Let P be the orthogonal projection onto $V^\perp$. Then, $\mathrm{tr}[P] = \dim(V^\perp) = n - d$
- $\mathrm{dist}(\boldsymbol{x}, V)^2 = |P\boldsymbol{x} \cdot P\boldsymbol{x}| = |\boldsymbol{x}^\top P \boldsymbol{x}| = \sum_{i,j \in [n]} x_i x_j P_{ij}$
- Thus, $\mathbb{E}[\mathrm{dist}(x, V)^2] = \sum_{i \in [n]} P_{ii} = n - d$

How well is $\mathrm{dist}(\boldsymbol{x}, V)$ concentrated around $\sqrt{n-d}$?

Corollary (Talagrand’s inequality). For $\boldsymbol{x} \sim \mathrm{Unif}\{-1,1\}^n$, we have

$$\Pr\left[\left|\mathrm{dist}(\boldsymbol{x}, V) - \sqrt{n-d}\right| \geq t\right] \leq Ce^{-ct^2}$$

where C, c are universal constants

Talagrand's Inequality: Convex Lipschitz Functions

Theorem (Talagrand)**.**

Let $A \subseteq \mathbb{R}^n$ be a convex set. Let $\boldsymbol{x} \sim \mathrm{Unif}\{0,1\}^n$. Then

$$\Pr[\boldsymbol{x} \in A] \Pr[\mathrm{dist}(\boldsymbol{x}, A) \geq t] \leq e^{-t^2/4} \qquad \forall t \geq 0$$

Equivalently, for a convex 1-Lipschitz function $f: \mathbb{R}^n \to \mathbb{R}$ (i.e., $|f(x) - f(y)| \leq \|x - y\|_2$ for any $x, y \in \mathbb{R}^n$) and $\boldsymbol{x} \sim \mathrm{Unif}\{0,1\}^n$,

$$\Pr[f(x) \leq r] \Pr[f(x) \geq r + t] \leq e^{-t^2/4} \qquad \forall r \in \mathbb{R}, t \geq 0$$

Proof of the equivalence.

- "$\Longrightarrow$": let $A := \{x \in \mathbb{R}^n : f(x) \leq r\}$. Then f is convex implies that A is convex. We also have $\mathrm{dist}(x, A) \leq t \Longrightarrow f(x) \leq r + t$ by the 1-Lipschitzness. Thus, $\Pr[f(\boldsymbol{x}) \leq r] = \Pr[\boldsymbol{x} \in A]$ and $\Pr[f(\boldsymbol{x}) \geq r + t] \leq \Pr[\mathrm{dist}(\boldsymbol{x}, A) \geq t]$
- "$\Longleftarrow$": let $r = 0$ and $f(x) = \mathrm{dist}(x, A)$

Talagrand's Inequality: Convex Lipschitz Functions

Theorem (Talagrand).

For a convex 1-Lipschitz function $f: \mathbb{R}^n \to \mathbb{R}$ (i.e., $|f(x) - f(y)| \leq \|x - y\|_2$ for any $x, y \in \mathbb{R}^n$) and $\boldsymbol{x} \sim \text{Unif}\{0,1\}^n$,

$$\Pr[f(x) \leq r] \Pr[f(x) \geq r + t] \leq e^{-t^2/4} \qquad \forall r \in \mathbb{R}, t \geq 0$$

Corollary. Let $\text{med}(X)$ be the median of the random variable X. That is, $\Pr[X \geq \text{med}(X)] \geq 1/2$ and $\Pr[X \leq \text{med}(X)] \geq 1/2$. Then, for a convex 1-Lipschitz function $f: \mathbb{R}^n \to \mathbb{R}$ and $\boldsymbol{x} \sim \text{Unif}\{0,1\}^n$,

$$\Pr\left[\left|f(x) - \text{med}\left(f(x)\right)\right| \geq t\right] \leq 4e^{-t^2/4}$$

Proof.

- Let $r \coloneqq \text{med}\left(f(\boldsymbol{x})\right)$. Then $\Pr[f(x) \leq r] \geq 1/2$ and $\Pr[f(x) \geq r + t] \leq 2e^{-t^2/4}$
- Let $r \coloneqq \text{med}\left(f(\boldsymbol{x})\right) - t$. Then $\Pr[f(x) \geq r + t] \geq 1/2$ $\Pr[f(x) \leq r] \leq 2e^{-t^2/4}$ ∎

Talagrand's Inequality: Convex Distance

Let the probability space be $\Omega = \Omega_1 \times \cdots \times \Omega_n$ with product probability measure

Weighted Hamming distance

- Given $\alpha \in \mathbb{R}^n_{\geq 0}$, $x, y \in \Omega$, define

$$d_\alpha(x, y) \coloneqq \sum_{i=1}^{n} \alpha_i \mathbf{1}[x_i \neq y_i]$$

- For a subset $A \subseteq \Omega$, $d_\alpha(x, A) \coloneqq \inf_{y \in A} d_\alpha(x, y)$

Talagrand's convex distance

- For $x \in \Omega$ and $A \subseteq \Omega$,

$$d_T(x, A) \coloneqq \sup_{\substack{\alpha \in \mathbb{R}^n_{\geq 0} \\ \|\alpha\|_2 = 1}} d_\alpha(x, A)$$

Talagrand's Inequality: Convex Distance

Let the probability space be $\Omega = \Omega_1 \times \cdots \times \Omega_n$ with product probability measure

Talagrand's convex distance

$$d_T(x, A) \coloneqq \sup_{\substack{\alpha \in \mathbb{R}^n_{\geq 0} \\ \|\alpha\|_2 = 1}} \inf_{y \in A} \sum_{i=1}^{n} \alpha_i \mathbf{1}[x_i \neq y_i]$$

Properties:

- If $A \subseteq \{0,1\}^n$ and $x \in \{0,1\}^n$, then $d_T(x, A) = \mathrm{dist}\big(x, \mathrm{conv}(A)\big)$
- For any $x \in \Omega$, define $\phi_x(y) \coloneqq (\mathbf{1}[x_1 \neq y_1], \ldots, \mathbf{1}[x_n \neq y_n]) \in \{0,1\}^n$, and $\phi_x(A) \coloneqq \{\phi_x(y) : y \in A\} \subseteq \{0,1\}^n$ for any $A \subseteq \Omega$. Then

$$d_T(x, A) = \mathrm{dist}\Big(\mathbf{0}, \mathrm{conv}\big(\phi_x(A)\big)\Big)$$

Talagrand's Inequality: Convex Distance

Theorem (Talagrand's inequality, general form)**.**

Let $A \subseteq \Omega = \Omega_1 \times \cdots \times \Omega_n$, and $\boldsymbol{x} \sim \Omega$ be chosen randomly with independent coordinates. Then

$$\Pr[\boldsymbol{x} \in A] \Pr[d_T(\boldsymbol{x}, A) \geq t] \leq e^{-t^2/4}$$

Talagrand's Inequality: Convex Distance

Theorem (Talagrand's inequality, functions with weighted certificates)**.**

Let $\boldsymbol{x} \sim \Omega$ with independent coordinates. Suppose that

$$f(y) \geq f(x) - \sum_{i=1}^{n} \alpha_i(x)\, \mathbf{1}[x_i \neq y_i] \qquad \forall x, y \in \Omega$$

Then,

$$\Pr\left[\left|f(\boldsymbol{x}) - \mathrm{med}\big(f(\boldsymbol{x})\big)\right| \geq t\right] \leq 4e^{-t^2/\nu^2}, \qquad \nu \coloneqq 2 \sup_{x \in \Omega} \|\alpha(x)\|_2$$

Talagrand's Inequality: Convex Distance

Theorem (Talagrand's inequality, functions with weighted certificates)**.**

Let $\boldsymbol{x} \sim \Omega$ with independent coordinates. Suppose that $f(y) \geq f(x) - \sum_{i=1}^{n} \alpha_i(x)\, \mathbf{1}[x_i \neq y_i]$ for any $x, y \in \Omega$. Then, for $\nu^2 := 4 \sup_{x \in \Omega} \|\alpha(x)\|_2^2$,

$$\Pr\big[\big|f(x) - \mathrm{med}\big(f(x)\big)\big| \geq t\big] \leq 4e^{-t^2/\nu^2}$$

Proof.

- For $r \in \mathbb{R}$, let $A := \{y : f(y) \leq r - t\}$
- For any $x \in \Omega$ such that $f(x) \geq r$, the assumption gives that
$$\exists \alpha(x) \in \mathbb{R}_{\geq 0}^{n}, \forall y \in A, \qquad d_{\alpha(x)}(x, y) \geq f(x) - f(y) \geq r - (r - t) = t$$
- Then, we have $d_{\alpha(x)}(x, A) \geq t$ and $d_T(x, A) \geq t / \|\alpha(x)\|_2 \geq 2t/\nu$

Talagrand's Inequality: Convex Distance

Theorem (Talagrand's inequality, functions with weighted certificates)**.**

Let $\boldsymbol{x} \sim \Omega$ with independent coordinates. Suppose that $f(y) \geq f(x) - \sum_{i=1}^{n} \alpha_i(x)\, \mathbf{1}[x_i \neq y_i]$ for any $x, y \in \Omega$. Then, for $\nu^2 := 4 \sup_{x \in \Omega} \|\alpha(x)\|_2^2$,

$$\Pr\left[\left|f(x) - \operatorname{med}\left(f(x)\right)\right| \geq t\right] \leq 4e^{-t^2/\nu^2},$$

Proof.

- Then, we have $d_{\alpha(x)}(x, A) \geq t$ and $d_T(x, A) \geq t/\|\alpha(x)\|_2 \geq 2t/\nu$
- By Talagrand's inequality (general form),

$$\Pr[f \leq r - t] \Pr[f \geq r] \leq \Pr[\boldsymbol{x} \in A] \Pr\left[d_T(\boldsymbol{x}, A) \geq \frac{2t}{\nu}\right] \leq e^{-t^2/\nu^2}$$

- The result then follows by taking $r := \operatorname{med}(f) + t$ and $r := \operatorname{med}(f)$

∎

Median vs. Mean

For any real random variable X satisfying

$$\Pr[|X - m| \geq t] \leq 2e^{-t^2/\sigma^2} \qquad \forall t \geq 0$$

for some $m \in \mathbb{R}$ and $\sigma > 0$, then the followings hold:

1. $|\mathrm{med}(X) - m| \leq C\sigma$
2. $|\mathbb{E}[X] - m| \leq C\sigma$
3. For every constant A, if $|m' - m| \leq A\sigma$, then

$$\Pr[|X - m'| \geq t] \leq 2e^{-\Omega(t^2/\sigma^2)} \qquad \forall t \geq 0$$

Median vs. Mean (Proof)

- We can rescale X so that $\sigma = 1$
- For the median, take $t > \sqrt{2\log 2}$:

$$\Pr\left[|X - m| > \sqrt{2\log 2}\right] \le e^{-t^2} < 1/2$$

  Thus, $\mathrm{med}(X)$ is within $\sqrt{2\log 2}$ of m
- For the mean,

$$|\mathbb{E}[X] - m| \le \mathbb{E}[|X - m|] = \int_0^\infty \Pr[|X - m| \ge t]\mathrm{d}t \le 2\int_0^\infty e^{-t^2}\mathrm{d}t = \sqrt{\pi}$$

- For the last inequality, since A is constant, by choosing a sufficiently small $c > 0$, we can let $2e^{-ct^2} \ge 1$ when $t \le 2A$ (e.g., $c = \frac{1}{10A^2}$). Then, for $t > 2A$, we have

$$\Pr[|X - m'| \ge t] \le \Pr[|X - m| \ge t/2] \le e^{-t^2/4}$$

■

Today's Lecture

- Subgaussian and Subgamma Random Variable
- The Entropy Method
- Talagrand's Inequality
- Applications: LIS and TSP

Application 1: Longest Increasing Subsequence

An increasing subsequence of a permutation $\sigma = (\sigma_1, \dots, \sigma_n)$ is defined to be some $\sigma_{i_1} < \cdots < \sigma_{i_\ell}$ for some $i_1 < \cdots < i_\ell$.

How well is the length X of the longest increasing subsequence of uniformly random permutation concentrated?

Example: $\sigma = (2, 1, 8, 7, 4, 5, 6, 3)$

- You can show that $\mathbb{E}[X] = \Theta(\sqrt{n})$ (A good exercise of binomial coefficient approximation)
- For concentration, there is one problem: $\sigma_1, \dots, \sigma_n$ are not independent
- You can sample $X_1, \dots, X_n \sim_{iid} \mathrm{Unif}[0,1]$, and their ordering gives a random permutation
- Talagrand's inequality $\Longrightarrow X = \Theta(\sqrt{n}) \pm \mathcal{O}\left(n^{1/4}\right)$ w.h.p.

Application 1: Longest Increasing Subsequence

Let $\Omega = \Omega_1 \times \cdots \times \Omega_n$ and $A \subseteq \Omega$. We say A is s-certifiable if for every $x \in A$, there exists a subset $I \subset [n]$ with $|I| \leq s$ such that for every $y \in \Omega$, if $y_I = x_I$, then $y \in A$

- For LIS, $\Omega = [0,1]^n$ and $A = \{x \in [0,1]^n : \mathrm{LIS}(x) \geq k\}$. Then A is k-certifiable since we just need to check k coordinates to determine an increasing subsequence of length k

Theorem (Talagrand's inequality for certifiable functions).

Let $\boldsymbol{x} \sim \Omega$ with independent coordinates. Let $f: \Omega \to \mathbb{R}$ be 1-Lipschitz with respect to Hamming distance on Ω. Suppose that $\{x \in \Omega : f(x) \geq r\}$ is r-certifiable. Then, for $m = \mathrm{med}\big(f(x)\big)$,

$$\Pr[f(\boldsymbol{x}) \leq m - t] \leq 2e^{-t^2/(4m)}$$
$$\Pr[f(\boldsymbol{x}) \geq m + t] \leq 2e^{-t^2/\left(4(m+t)\right)}$$

- For $x \in [0,1]^n$, let $f(x) \coloneqq \mathrm{LIS}(x)$. Then f is 1-Lipschitz (since changing one coordinate can change the LIS by at most 1). It is easy to show that $m = \Theta(\sqrt{n})$. The above theorem implies the desired concentration bound.

Theorem (Talagrand's inequality for certifiable functions).

Let $\boldsymbol{x} \sim \Omega$ with independent coordinates. Let $f : \Omega \to \mathbb{R}$ be 1-Lipschitz with respect to Hamming distance on Ω. Suppose that $\{x \in \Omega : f(x) \geq r\}$ is r-certifiable. Then, for $m = \mathrm{med}(f(x))$,

$$\Pr[f(\boldsymbol{x}) \leq m - t] \leq 2e^{-t^2/(4m)}$$
$$\Pr[f(\boldsymbol{x}) \geq m + t] \leq 2e^{-t^2/(4(m+t))}$$

Proof.

- Let $A := \{f \leq r - t\}$ and $B := \{f \geq r\}$. We first show that $\Pr[f \leq r - t] \Pr[f \geq r] \leq e^{-t^2/(4r)}$
- B is r-certifiable, so for every $y \in B$, let $I(y)$ denote its certificate with $|I(y)| \leq r$
- By the 1-Lipschitzness of f, for every $x \in A$, $t \leq |f(x) - f(y)| \leq d_H(x, y)$
- We want to apply Talagrand's inequality (the general form):

$$\Pr[\boldsymbol{x} \in A] \Pr[d_T(\boldsymbol{x}, A) \geq t] \leq e^{-t^2/4}$$

- For $i \in [n]$, define $\alpha_i(y) := 1/\sqrt{|I(y)|}$ for $i \in I(y)$ and $\alpha_i(y) := 0$ otherwise. Then,

$$\|\alpha(y)\|_2 = 1, \qquad d_\alpha(x, y) \geq t/\sqrt{|I(y)|} \geq t/\sqrt{r} \quad \forall x \in A$$

Theorem (Talagrand's inequality for certifiable functions).

Let $\boldsymbol{x} \sim \Omega$ with independent coordinates. Let $f: \Omega \to \mathbb{R}$ be 1-Lipschitz with respect to Hamming distance on Ω. Suppose that $\{x \in \Omega : f(x) \geq r\}$ is r-certifiable. Then, for $m = \text{med}(f(x))$,

$$\Pr[f(\boldsymbol{x}) \leq m - t] \leq 2e^{-t^2/(4m)}$$

$$\Pr[f(\boldsymbol{x}) \geq m + t] \leq 2e^{-t^2/(4(m+t))}$$

Proof.

- Let $A \coloneqq \{f \leq r - t\}$ and $B \coloneqq \{f \geq r\}$. We first show that $\Pr[f \leq r - t] \Pr[f \geq r] \leq e^{-t^2/(4r)}$
- We want to apply Talagrand's inequality (the general form):

$$\Pr[\boldsymbol{x} \in A] \Pr[d_T(\boldsymbol{x}, A) \geq t] \leq e^{-t^2/4}$$

- For $i \in [n]$, define $\alpha_i(y) \coloneqq 1/\sqrt{|I(y)|}$ for $i \in I(y)$ and $\alpha_i(y) \coloneqq 0$ otherwise. Then,

$$\|\alpha(y)\|_2 = 1, \qquad d_\alpha(x, y) \geq t/\sqrt{|I(y)|} \geq t/\sqrt{r} \quad \forall x \in A$$

- Thus, $d_T(y, A) \geq t/\sqrt{r}$ for every $y \in B$

$$\Pr[f \leq r - t] \Pr[f \geq r] \leq \Pr[\boldsymbol{x} \in A] \Pr[d_T(\boldsymbol{x}, A) \geq t/\sqrt{r}] \leq e^{-t^2/(4r)}$$

Theorem (Talagrand's inequality for certifiable functions)**.**

Let $\boldsymbol{x} \sim \Omega$ with independent coordinates. Let $f: \Omega \to \mathbb{R}$ be 1-Lipschitz with respect to Hamming distance on Ω. Suppose that $\{x \in \Omega : f(x) \geq r\}$ is r-certifiable. Then, for $m = \mathrm{med}\big(f(x)\big)$,

$$\Pr[f(\boldsymbol{x}) \leq m - t] \leq 2e^{-t^2/(4m)}$$
$$\Pr[f(\boldsymbol{x}) \geq m + t] \leq 2e^{-t^2/\left(4(m+t)\right)}$$

Proof.

$$\Pr[f \leq r - t] \Pr[f \geq r] \leq e^{-t^2/(4r)}$$

- The lower tail in the theorem follows from taking $r = m$
- The upper tail in the theorem follows from taking $r = m + t$

■

Application 1: Longest Increasing Subsequence

An increasing subsequence of a permutation $\sigma = (\sigma_1, \ldots, \sigma_n)$ is defined to be some $\sigma_{i_1} < \cdots < \sigma_{i_\ell}$ for some $i_1 < \cdots < i_\ell$.

How well is the length X of the longest increasing subsequence of uniformly random permutation concentrated?

- Talagrand's inequality $\Longrightarrow X = \Theta(\sqrt{n}) \pm \mathcal{O}(n^{1/4})$ w.h.p.
- Final remark: *the correct order of the fluctuation is $n^{1/6}$ (Baik-Deift-Johansson '99). They showed that $n^{-1/6}(X - 2\sqrt{n})$ converges to the Tracy–Widom distribution*

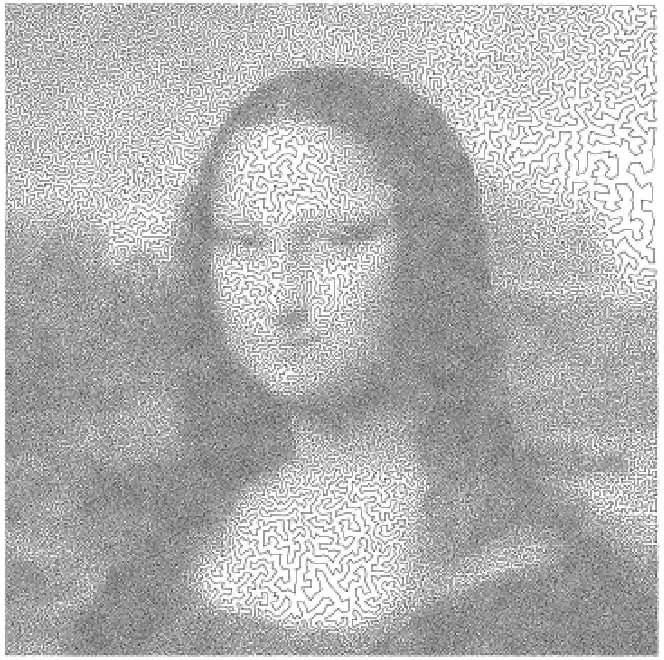
Application 2: Euclidean TSP

Let $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n \in [0,1]^2$ be uniformly random points in the unit square. The travelling salesman problem (TSP) is to find a tour through all the n points with the shortest possible length:

$$\mathrm{TSP}(x_1, \dots, x_n) \coloneqq \min_{\pi \in \mathcal{S}_n} \sum_{i=1}^{n} \mathrm{d}\big(x_{\pi(i)}, x_{\pi(i+1)}\big), \qquad x_{\pi(n+1)} \coloneqq x_{\pi(1)}$$

Here, $\mathrm{d}(x, y) = \|x - y\|_2$ is the Euclidean distance

- Let $L_n \coloneqq \mathrm{TSP}(\boldsymbol{x}_1, \dots, \boldsymbol{x}_n)$ be the random variable of TSP length
- It is known that $\mathbb{E}[L_n] = \Theta(\sqrt{n})$
- We can show that L_n is 16-subgaussian



Mona Lisa TSP Challenge

Application 2: Euclidean TSP

Our plan is to apply the following version of Talagrand's inequality:

Theorem. Let $\boldsymbol{x} \sim \Omega$ with independent coordinates. Suppose that $f(y) \geq f(x) - \sum_{i=1}^{n} \alpha_i(x)\, \mathbf{1}[x_i \neq y_i]$ for any $x, y \in \Omega$. Then, for $\nu^2 := 4 \sup_{x \in \Omega} \|\alpha(x)\|_2^2$, $\Pr\left[\left|f(\boldsymbol{x}) - \mathrm{med}\big(f(\boldsymbol{x})\big)\right| \geq t\right] \leq 4e^{-t^2/\nu^2}$

- Let $\Omega = \{X := (x_1, \dots, x_n) : x_i \in [0,1]^2\}$ and $f(X) := \mathrm{TSP}(x_1, \dots, x_n)$
- We need to construct a certificate $\alpha(X)$ such that for any two inputs X and Y,

$$f(X) \leq f(Y) + \sum_{i=1}^{n} \alpha_i(X)\, \mathbf{1}[x_i \neq y_i]$$

- We'll show how to merge a tour of x and the optimal tour of y and obtain a tour of $x \cup y$ of length $\ell_{X \cup Y} \leq f(Y) + \sum_{i=1}^{n} \alpha_i(X)\, \mathbf{1}[x_i \neq y_i]$
- Then, by removing the points not in x, the length is non-increasing. Thus, $f(X) \leq \ell_{X \cup Y}$

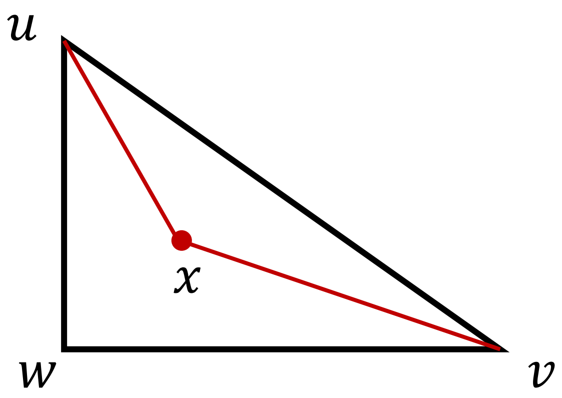
Application 2: Euclidean TSP

We need the following geometric lemma (related to the Sierpiński curve):

Lemma. For any $x_1, x_2, \ldots, x_n \in [0,1]^2$, there exists a permutation σ such that

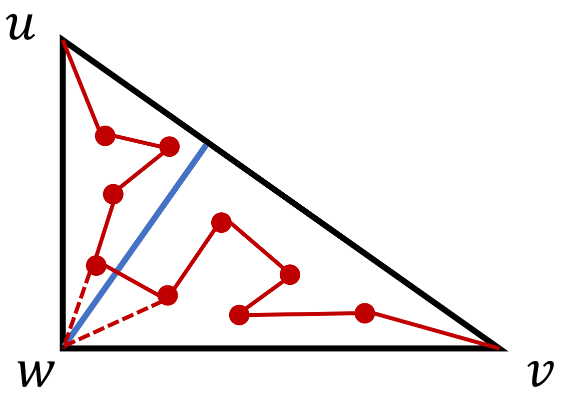
$$\mathrm{d}\left(x_{\sigma(1)}, x_{\sigma(2)}\right)^2 + \mathrm{d}\left(x_{\sigma(2)}, x_{\sigma(3)}\right)^2 + \cdots + \mathrm{d}\left(x_{\sigma(n-1)}, x_{\sigma(n)}\right)^2 + \mathrm{d}\left(x_{\sigma(n)}, x_{\sigma(1)}\right)^2 \le 4$$

Proof.

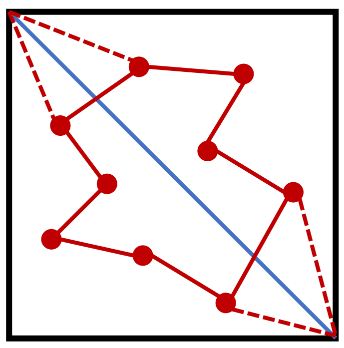


Pythagorean Inequality:

$$\mathrm{d}(x,u)^2 + \mathrm{d}(x,v)^2 \le \mathrm{d}(u,v)^2$$



$$\mathrm{d}\left(u, x_{\tau(1)}\right)^2 + \sum_{i=1}^{m-1} \mathrm{d}\left(x_{\tau(i)}, x_{\tau(i+1)}\right)^2 + \mathrm{d}\left(x_{\tau(m)}, v\right)^2 \le \mathrm{d}(u,v)^2$$



Merge two triangles

Application 2: Euclidean TSP

We need the following geometric lemma (related to the Sierpiński curve):

Lemma. For any $x_1, x_2, \dots, x_n \in [0,1]^2$, there exists a permutation σ such that

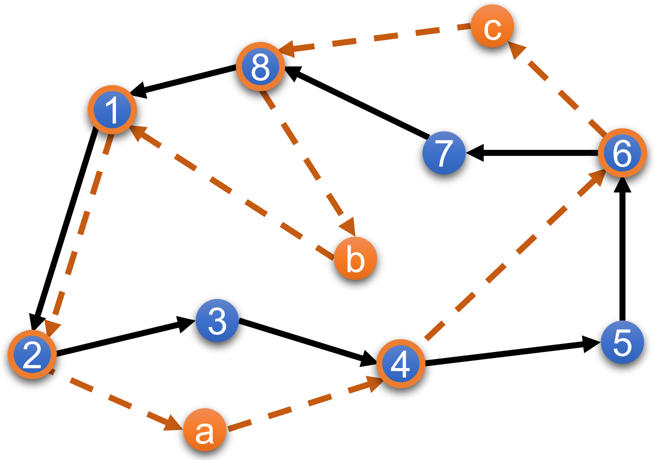
$$\mathrm{d}\big(x_{\sigma(1)}, x_{\sigma(2)}\big)^2 + \mathrm{d}\big(x_{\sigma(2)}, x_{\sigma(3)}\big)^2 + \cdots + \mathrm{d}\big(x_{\sigma(n-1)}, x_{\sigma(n)}\big)^2 + \mathrm{d}\big(x_{\sigma(n)}, x_{\sigma(1)}\big)^2 \le 4$$

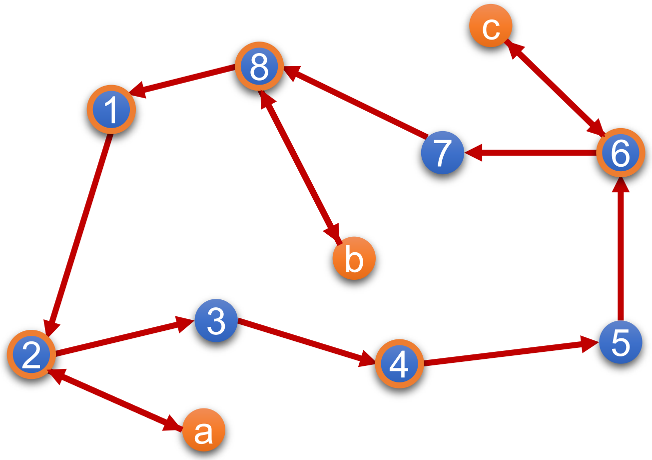
- For simplicity, we can consider σ as a function $\sigma: \mathbb{R}^2 \to \mathbb{R}^2$ such that $\sigma\big(x_{\sigma(i)}\big) \coloneqq x_{\sigma(i-1)}$ for any $i \in [n]$ and $x_{\sigma(0)} \coloneqq x_{\sigma(n)}$, i.e., the predecessor function
- Then, the lemma is equivalent to

$$\sum_{i=1}^{n} \mathrm{d}\big(x_i, \sigma(x_i)\big)^2 \le 4$$

Application 2: Euclidean TSP

How to merge the tours





$\boldsymbol{y} = \{1,2,3,4,5,6,7,8\}$

$f(\boldsymbol{y}) = \text{dist}(1 \to 2 \to \cdots \to 8 \to 1)$ (optimal)

$\boldsymbol{x} = \{1,2,4,6,8,a,b,c\}$

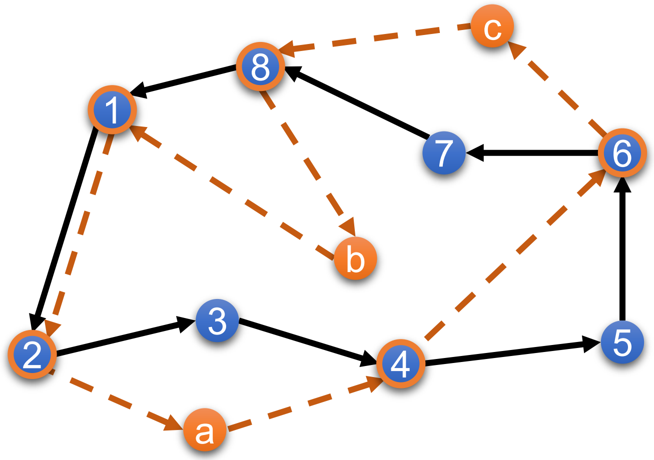
Lemma $\Longrightarrow \sigma = \begin{pmatrix} 1 & 2 & 4 & 6 & 8 & a & b & c \\ b & 1 & a & 4 & c & 2 & 8 & 6 \end{pmatrix}$

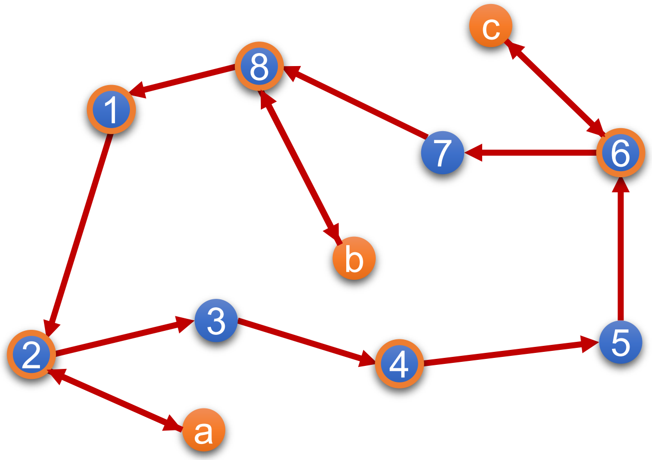
$1 \to 2 \to a \to 2 \to 3 \to 4 \to 5 \to 6 \to c \to 6 \to 7 \to 8 \to b \to 8 \to 1$

1. Traverses along the optimal order of Y
2. If the current point $y_k = x_{\sigma(i)}$ and $x_{\sigma(i+1)} \notin Y$:
   i. Traverse along X's tour just before it rejoins Y's tour
   ii. Traverse backward and return to y_j

Application 2: Euclidean TSP

How to merge the tours





$\boldsymbol{y} = \{1,2,3,4,5,6,7,8\}$

$f(\boldsymbol{y}) = \mathrm{dist}(1 \to 2 \to \cdots \to 8 \to 1)$ (optimal)

$\boldsymbol{x} = \{1,2,4,6,8,a,b,c\}$

Lemma $\Longrightarrow \sigma = \begin{pmatrix} 1 & 2 & 4 & 6 & 8 & a & b & c \\ b & 1 & a & 4 & c & 2 & 8 & 6 \end{pmatrix}$

$1 \to 2 \to a \to 2 \to 3 \to 4 \to 5 \to 6 \to c \to 6 \to 7 \to 8 \to b \to 8 \to 1$

$$f(X \cup Y) \le f(Y) + \sum_{i=1}^{n} 2\mathrm{d}\big(x_i, \sigma(x_i)\big)\mathbf{1}[x_i \notin Y]$$

$$\le f(Y) + \sum_{i=1}^{n} 2\mathrm{d}\big(x_i, \sigma(x_i)\big)\mathbf{1}[x_i \neq y_i]$$

Application 2: Euclidean TSP

$$f(X) \le f(Y) + \sum_{i=1}^{n} 2\mathrm{d}\big(x_i, \sigma(x_i)\big)\mathbf{1}[x_i \ne y_i]$$

- Let $\alpha_i(X) \coloneqq 2\mathrm{d}\big(x_i, \sigma(x_i)\big)$
- The geometric lemma gives that

$$\|\alpha(X)\|_2^2 = 4\sum_{i=1}^{n} \mathrm{d}\big(x_i, \sigma(x_i)\big)^2 \le 16$$

- Thus, by Talagrand's inequality,

$$\Pr\Big[\big|f(\boldsymbol{X}) - \mathrm{med}\big(f(\boldsymbol{X})\big)\big| \ge t\Big] \le 4e^{-t^2/16}$$

- That is, $L_n = f(\boldsymbol{X})$ is 16-subgaussian

∎