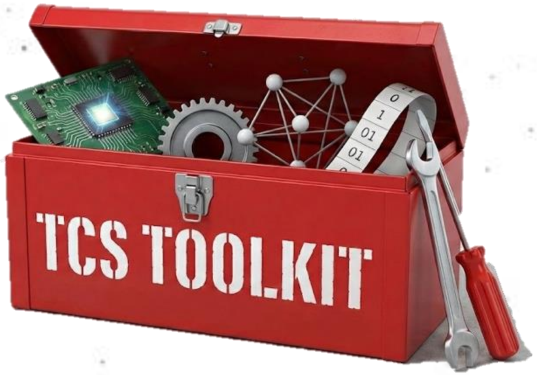
CS 58500 – Theoretical Computer Science Toolkit

Lecture 2 (01/22)

Concentration Inequality I



https://ruizhezhang.com/course_spring_2026.html

Today's Lecture

- Introduction to High-Dimensional Probability
- Markov Inequality
- Chebyshev Inequality

High-Dimensional Probability

HDP focuses on probabilistic models involving either many random variables or variables that take values in high-dimensional spaces, analyzed in a quantitative and non-asymptotic manner

Theorem (Weak law of large numbers (WLLN)).

Let X be a real random variable with expectation $\mathbb{E}[X] = m$. Consider an *iid* sequence $\{X_i\}_{i \in \mathbb{N}}$ of copies of X, and let

$$\overline{X}_n := \frac{1}{n} \sum_{i=1}^{n} X_i \quad \text{for } n \in \mathbb{N}$$

Then, for any $t > 0$, we have the limit

$$\Pr\left[\left|\overline{X}_n - m\right| \geq t\right] \to 0 \quad \text{as } n \to \infty$$

Why WLLN is not useful?

High-Dimensional Probability

HDP focuses on probabilistic models involving either many random variables or variables that take values in high-dimensional spaces, analyzed in a quantitative and non-asymptotic manner

Example:

- Flip n unbiased coins
- Let $X_1, \dots, X_n$ be the outcomes (i.e., $X_i \coloneqq 1$ for "head" and $X_i \coloneqq 0$ for "tail")
- By the linearity of expectation, $\mathbb{E}[\overline{X}_n] = \frac{1}{n}(\mathbb{E}[X_1] + \cdots + \mathbb{E}[X_n]) = \frac{1}{2}$
- WLLN $\Longrightarrow \Pr\left[\left|\overline{X}_n - \frac{1}{2}\right| \geq t\right] \to 0 \quad \text{as} \;\; n \to \infty$
- However, we only have a finite number of coins (say, 100). What is the probability of getting 60 heads?

High-Dimensional Probability

Phenomena in HDP

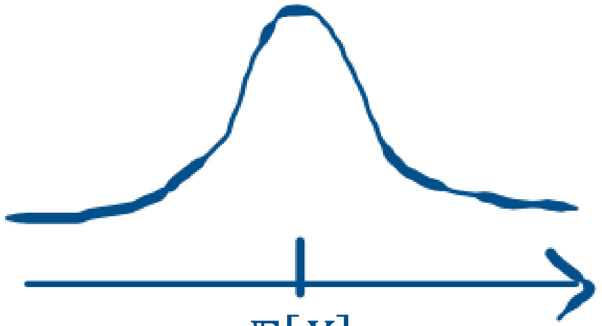
- Concentration
- Suprema
- Universality
- Phase transition

High-Dimensional Probability

Concentration

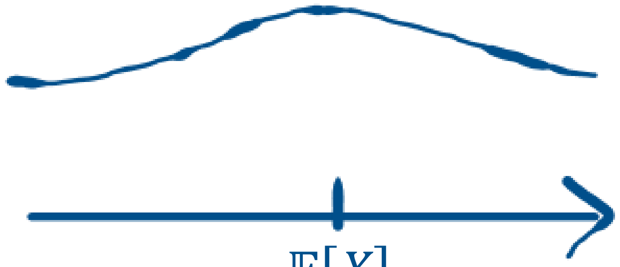
- Consider a collection of random variables $(X_1, \dots, X_n)$ and a measurable function f
- Define a random variable $Z \coloneqq f(X_1, \dots, X_n)$
- Concentration inequalities take the following forms:

$$\Pr[|Z - \mathrm{median}(Z)| \geq t] \leq ??? \quad \text{or} \quad \Pr[|Z - \mathbb{E}[Z]| \geq t] \leq ???$$



$\mathbb{E}[X]$

Strong concentration



$\mathbb{E}[X]$

Weak concentration

High-Dimensional Probability

Suprema

- For a (possibly uncountable) index set T, consider a real-valued random process $\mathcal{X} \coloneqq (X_t : t \in \mathrm{T})$
- Think about $\mathcal{X}$ as a map from each point $t \in \mathrm{T}$ to a random variable X_t
- Define a random variable $Z \coloneqq \sup\{X_t : t \in \mathrm{T}\}$
- We can apply concentration inequalities to Z, but what is $\mathbb{E}[Z]$?

Example:

- For a random matrix $\boldsymbol{X} \in \mathbb{R}^{m \times n}$, its spectral norm is

$$\|\boldsymbol{X}\| = \sup\{\boldsymbol{u}^\top \boldsymbol{X} \boldsymbol{v} : \|\boldsymbol{u}\|_2 = 1, \|\boldsymbol{v}\|_2 = 1\}$$

- $\mathrm{T} = \mathbb{S}^{m-1} \times \mathbb{S}^{n-1}$, and for $t = (\boldsymbol{u}, \boldsymbol{v}) \in T$ the random variable $X_t \coloneqq \boldsymbol{u}^\top \boldsymbol{X} \boldsymbol{v}$
- $\mathbb{E}[\|\boldsymbol{X}\|] = \mathbb{E}[\sup\{X_t : t \in \mathrm{T}\}]$

High-Dimensional Probability

Universality / Invariance principle

- If $X_1, \dots, X_n$ are independent (or weakly dependent) random variables, then the expectation $\mathbb{E}[f(X_1, \dots, X_n)]$ is "insensitive" to the distribution of $X_1, \dots, X_n$ when the function f is "sufficiently smooth.

Example 1: central-limit theorem

- Let $X_1, X_2, \dots$ be a sequence of *iid* random variables with $\mathbb{E}[X_i] = \mu$ and $\mathrm{Var}[X_i] = \sigma^2$. Then

$$\sqrt{n}\left(\overline{X}_n - \mu\right) \to \mathcal{N}(0, \sigma^2) \quad \text{as } n \to \infty$$

High-Dimensional Probability

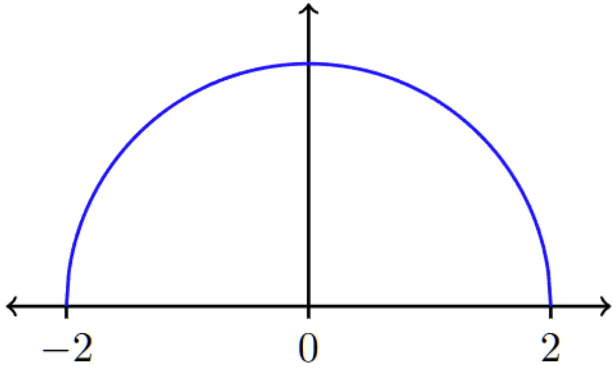
Universality / Invariance principle

- If $X_1, \dots, X_n$ are independent (or weakly dependent) random variables, then the expectation $\mathbb{E}[f(X_1, \dots, X_n)]$ is "insensitive" to the distribution of $X_1, \dots, X_n$ when the function f is "sufficiently smooth.

Example 2: Wigner matrix

- Let $\boldsymbol{X} \in \mathbb{R}^{n\times n}$ be a symmetric matrix whose entries $\boldsymbol{X}_{ij}$ are independent random variables with $\mathbb{E}[\boldsymbol{X}_{ij}] = 0$, $\mathbb{E}[\boldsymbol{X}_{ij}^2] = 1$, and $\mathbb{E}[|\boldsymbol{X}_{ij}^3|] \leq C$ for $i \geq j$

- The (averaged) histogram of the eigenvalues of the matrix $\frac{\boldsymbol{X}}{\sqrt{n}}$ looks, when n is large, like a semicircle



$$\mu_n := \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} \delta_{\lambda_i(\boldsymbol{X}/\sqrt{n})}\right] \to \mu_{\mathrm{sc}}(\mathrm{d}x) := \frac{1}{2\pi}\sqrt{4-x^2}\,\mathbf{1}_{|x|\leq 2}$$

High-Dimensional Probability

Phase transition

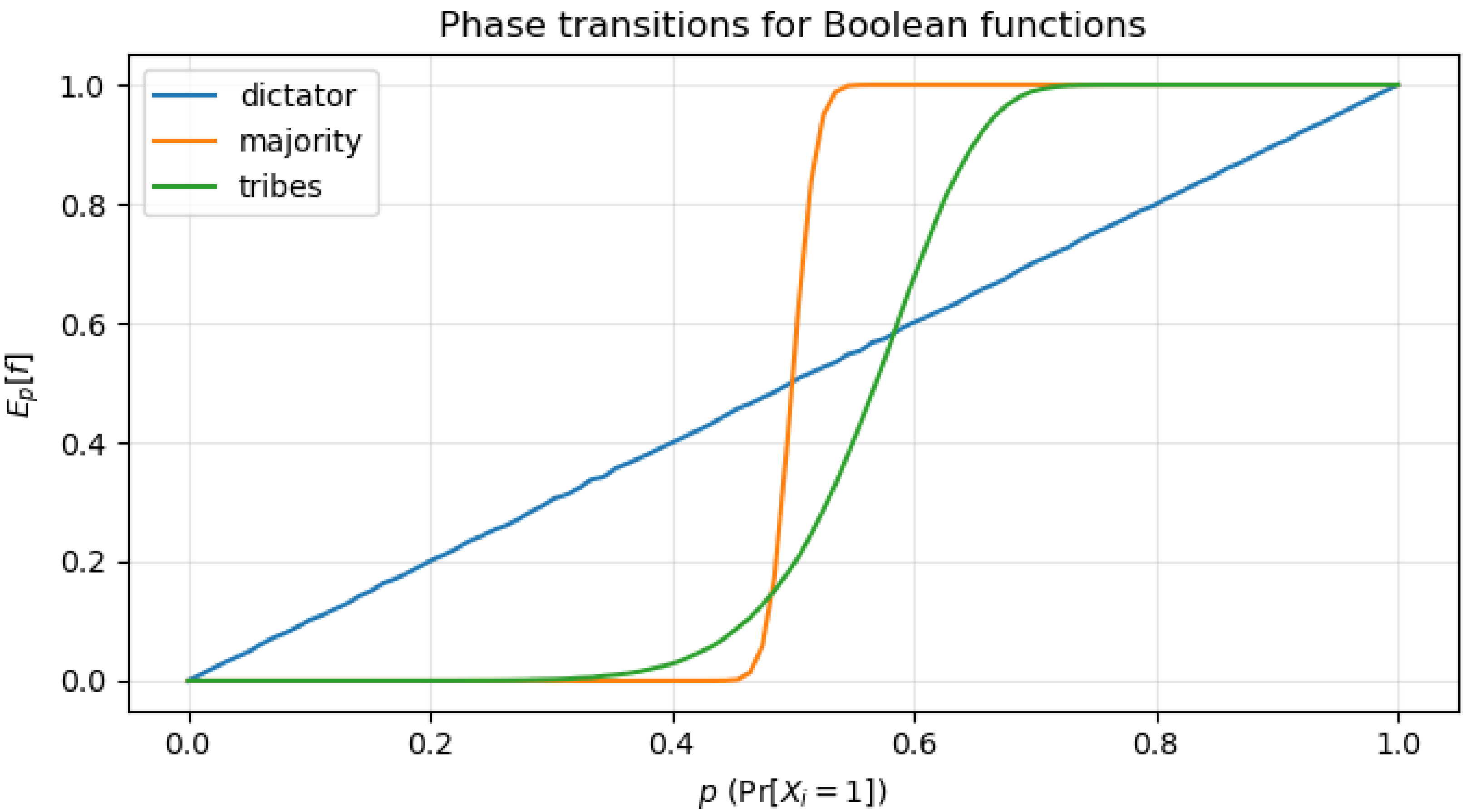
- The behavior of probabilistic models tends to undergo abrupt changes when the model parameters cross some threshold value.

Example:

- Let $X_1, \dots, X_n$ be iid Bernoulli random variables, that is, $\Pr[X_i = 1] = p$ and $\Pr[X_i = 0] = 1 - p$
- Let $f: \{0,1\}^n \to \{0,1\}$ be a Boolean function
- For a large family of f, the expectation $\mathbb{E}_p[f(X_1, \dots, X_n)]$ exhibits a sharp transition as p changes
- Concretely, you can view f as a voting rule that takes n votes and outputs the winner
  - → Majority: $f(x_1, \dots, x_n) = \mathbf{1}_{x_1 + \cdots + x_n \geq n/2}$
  - → Dictatorship: $f(x_1, \dots, x_n) = x_1$
  - → Tribes: $f(x_1, \dots, x_n) = (x_1 \wedge \cdots \wedge x_w) \vee (x_{w+1} \wedge \cdots \wedge x_{2w}) \vee \cdots \vee (x_{n-w+1} \wedge \cdots \wedge x_n)$

High-Dimensional Probability

Phase transition



Today's Lecture

- Introduction to High-Dimensional Probability
- Markov Inequality
- Chebyshev Inequality

Markov Inequality

Let X be a non-negative random variable. Then, we have

$$\Pr[X \geq t] \leq \frac{\mathbb{E}[X]}{t}$$

Proof.

- Let $\mu(x)$ be the probability density function of X
- The expectation $\mathbb{E}[X]$ can be expressed as:

$$\mathbb{E}[X] = \int_0^\infty x\mu(x)\mathrm{d}x = \int_0^\infty \left(\int_0^x \mu(x)\mathrm{d}y\right)\mathrm{d}x = \int_0^\infty \left(\int_y^\infty \mu(x)\mathrm{d}x\right)\mathrm{d}y = \int_0^\infty \Pr[X \geq y]\mathrm{d}y$$

- $\Pr[X \geq y]$ is a non-increasing function of y, which implies that

$$\mathbb{E}[X] = \int_0^\infty \Pr[X \geq y]\mathrm{d}y \geq \int_0^t \Pr[X \geq y]\mathrm{d}y \geq \int_0^t \Pr[X \geq t]\mathrm{d}y = t\Pr[X \geq t]$$

■

Markov Inequality

Let X be a non-negative random variable. Then, we have

$$\Pr[X \geq t] \leq \frac{\mathbb{E}[X]}{t}$$

Proof (discrete case).

- Let's prove it by contradiction. Suppose $\exists t > 0$ such that $\Pr[X \geq t] > \mathbb{E}[X]/t$

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{i \in \Omega} i \Pr[X = i] = \sum_{i<t} i \Pr[X = i] + \sum_{i \geq t} i \Pr[X = i] \\
&\geq \sum_{i \geq t} i \Pr[X = i] \geq \sum_{i \geq t} t \Pr[X = i] = t \sum_{i \geq t} \Pr[X = i] = t \Pr[X \geq t] \\
&> \mathbb{E}[X]
\end{aligned}$$

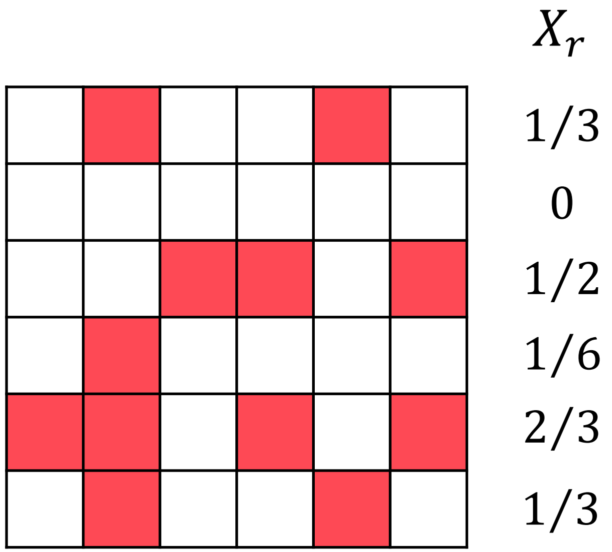
- We get $\mathbb{E}[X] > \mathbb{E}[X]$, contradiction!

$\blacksquare$

Markov Inequality

Pigeon-hole principle / Averaging argument

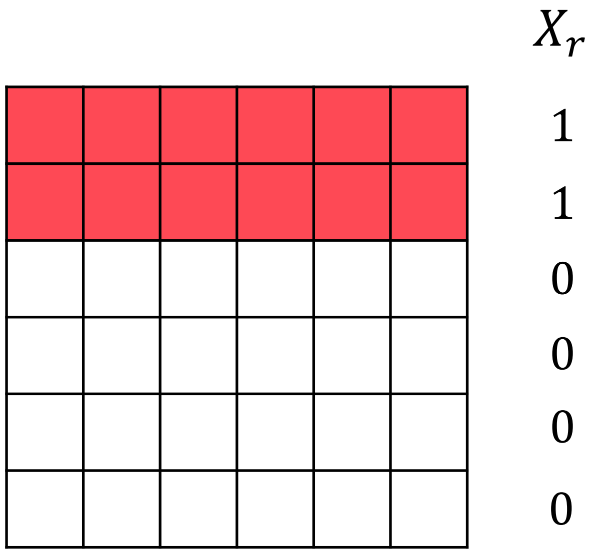
- Suppose (R, C) is a joint distribution over $[m] \times [n]$.
  Intuitively, think of a matrix with m-rows and n-columns
- Let $\mathcal{E} \subset [m] \times [n]$ be a subset of cells such that $\Pr[(R, C) \in \mathcal{E}] \geq \epsilon$
- Let $X_r := \Pr[(R, C) \in \mathcal{E} | R = r]$
- Then, for $\epsilon \in (0,1)$, $\alpha \in [0, \epsilon]$, we have $\Pr[X_R \geq \alpha] > \frac{\epsilon - \alpha}{1 - \alpha}$



Markov Inequality

Pigeon-hole principle / Averaging argument

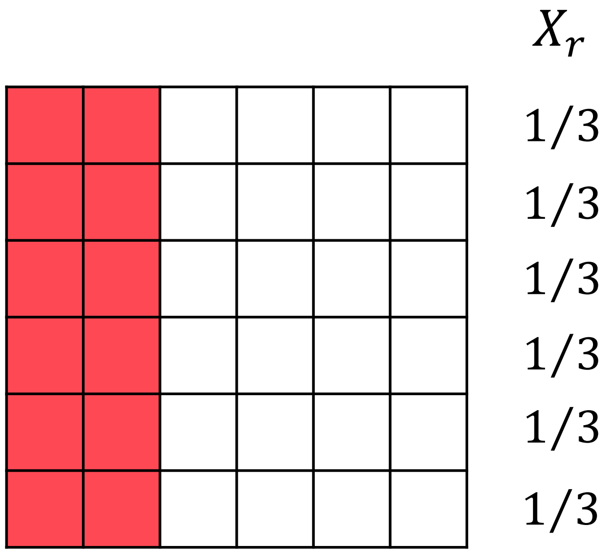
- Suppose (R, C) is a joint distribution over $[m] \times [n]$.
  Intuitively, think of a matrix with m-rows and n-columns
- Let $\mathcal{E} \subset [m] \times [n]$ be a subset of cells such that $\Pr[(R, C) \in \mathcal{E}] \geq \epsilon$
- Let $X_r \coloneqq \Pr[(R, C) \in \mathcal{E} | R = r]$
- Then, for $\epsilon \in (0,1)$, $\alpha \in [0, \epsilon]$, we have $\Pr[X_R \geq \alpha] > \frac{\epsilon - \alpha}{1 - \alpha}$



Markov Inequality

Pigeon-hole principle / Averaging argument

- Suppose (R, C) is a joint distribution over $[m] \times [n]$.
  Intuitively, think of a matrix with m-rows and n-columns
- Let $\mathcal{E} \subset [m] \times [n]$ be a subset of cells such that $\Pr[(R, C) \in \mathcal{E}] \geq \epsilon$
- Let $X_r \coloneqq \Pr[(R, C) \in \mathcal{E} | R = r]$
- Then, for $\epsilon \in (0,1)$, $\alpha \in [0, \epsilon]$, we have $\Pr[X_R \geq \alpha] > \frac{\epsilon - \alpha}{1 - \alpha}$



Proof.

- We know that $\mathbb{E}[X_R] \geq \epsilon$ and $X_R \leq 1$
- By the law of total expectation, we have

$$\mathbb{E}[X_R] = \underbrace{\mathbb{E}[X_R | X_R \geq \alpha]}_{\leq 1} \cdot \Pr[X_R \geq \alpha] + \underbrace{\mathbb{E}[X_R | X_R < \alpha]}_{< \alpha} \cdot (1 - \Pr[X_R \geq \alpha]) \geq \epsilon$$

Markov Inequality

Pigeon-hole principle / Averaging argument

- Suppose (R, C) is a joint distribution over $[m] \times [n]$.
  Intuitively, think of a matrix with m-rows and n-columns
- Let $\mathcal{E} \subset [m] \times [n]$ be a subset of cells such that $\Pr[(R, C) \in \mathcal{E}] \geq \epsilon$
- Let $X_r \coloneqq \Pr[(R, C) \in \mathcal{E} | R = r]$
- Then, for $\epsilon \in (0,1)$, $\alpha \in [0, \epsilon]$, we have $\Pr[X_R \geq \alpha] > \frac{\epsilon - \alpha}{1 - \alpha}$

Proof.

- We get that

$$\Pr[X_R \geq \alpha] + \alpha(1 - \Pr[X_R \geq \alpha]) > \epsilon$$

$$\Pr[X_R \geq \alpha] > \frac{\epsilon - \alpha}{1 - \alpha}$$

■

Markov Inequality

First-moment principle

Let X be a random variable with $\mathbb{E}[X] = \mu$. Then, we have

$$\Pr[X \geq \mu] > 0\,, \qquad \Pr[X \leq \mu] > 0$$

It is a special case of the pigeon-hole principle (in the previous slide) with $\alpha = \epsilon = \mu$

Proof.

- We'll prove by contradiction. Suppose $\Pr[X \geq \mu] = 0$
- Then, the expectation becomes:

$$\mu = \mathbb{E}[X] = \sum_{x} x \Pr[X = x] = \sum_{x<\mu} x \Pr[X = x] < \mu \sum_{x<\mu} \Pr[X = x] = \mu$$

- It implies $\mu < \mu$, which is a contradiction

■

Derandomization

- To what extent is randomness essential for efficient algorithms?
- $\text{BPP} \stackrel{?}{=} \text{P}$

Basic idea of derandomization

- Every randomized algorithm can be expressed as a deterministic function $\mathcal{A}(x, r)$
- $x \in \{0,1\}^n$ is the input and $r \in \{0,1\}^m$ is the random string
- BPP guarantees that, for any $x \in \{0,1\}^n$,

$$\Pr_{r \sim \{0,1\}^n}[\mathcal{A}(x, r) = f(x)] \geq 2/3$$

i.e. at least $\frac{2}{3}$-fraction of r's are "good" for x

- If there is a deterministic algorithm $\mathcal{B}(x)$ that outputs "good" random string r
- Then $\mathcal{A}(x, \mathcal{B}(x))$ is a deterministic algorithm that computes f

Derandomization

Max-3SAT:

- $(x_1 \vee x_2 \vee \overline{x_3}) \wedge (x_2 \vee x_4 \vee x_7) \wedge \cdots$ is a 3-CNF formula with n variables and m clauses of size 3
- Find an assignment of $x_1, \ldots, x_n \in \{0,1\}$ to satisfy the most clauses

Randomized algorithm

- Set each $x_i = 0$ or 1 with probability $1/2$ independently
- Let $Z_i \in \{0,1\}$ indicate whether the i-th clause is satisfied
- $\mathbb{E}[Z_i] = 7/8$ (check by yourself)
- Let $Z \coloneqq Z_1 + \cdots + Z_m$ be the total number of satisfied clauses
- By the linearity of expectation, $\mathbb{E}[Z] = \mathbb{E}[Z_1] + \cdots + \mathbb{E}[Z_m] = \frac{7}{8}m$

Derandomization

Max-3SAT:

- $(x_1 \vee x_2 \vee \overline{x_3}) \wedge (x_2 \vee x_4 \vee x_7) \wedge \cdots$ is a 3-CNF formula with n variables and m clauses of size 3
- Find an assignment of $x_1, \ldots, x_n \in \{0,1\}$ to satisfy the most clauses

Derandomization via conditional expectation

- By the first moment principle, $\mathbb{E}[Z] \geq (7/8)m$ implies that $\Pr[Z \geq (7/8)m] > 0$
- Our goal is to find such an assignment
- By the law of total expectation,

$$\mathbb{E}[Z] = \mathbb{E}[Z|x_1 = 0] \cdot \frac{1}{2} + \mathbb{E}[Z|x_1 = 1] \cdot \frac{1}{2} \geq \frac{7}{8} m$$

- Either $\mathbb{E}[Z|x_1 = 0] \geq (7/8)m$ or $\mathbb{E}[Z|x_1 = 1] \geq (7/8)m$. Then, we can fix the value of x_1
- How to calculate the conditional expectation?

Derandomization

Max-3SAT:

- $(x_1 \lor x_2 \lor \overline{x_3}) \land (x_2 \lor x_4 \lor x_7) \land \cdots$ is a 3-CNF formula with n variables and m clauses of size 3
- Find an assignment of $x_1, \dots, x_n \in \{0,1\}$ to satisfy the most clauses

Derandomization via conditional expectation

- We can iterate this process to find the values of $x_2, x_3, \dots, x_n$
- During this process, we always maintain that

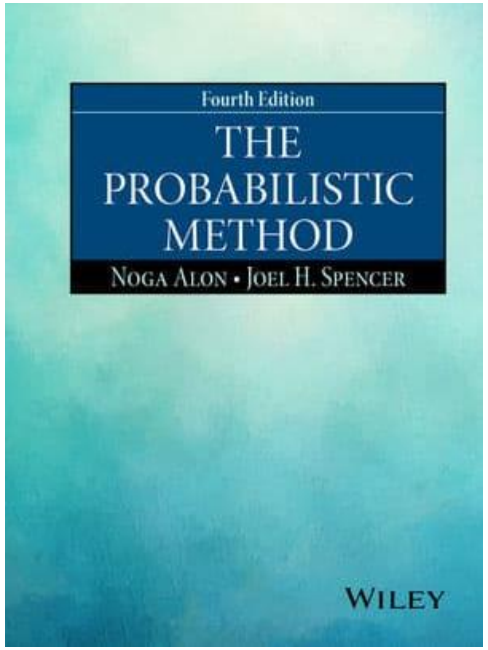
$$\mathbb{E}[Z | x_1 = a_1, x_2 = a_2, \dots, x_k = a_k] \geq \frac{7}{8} m$$

- Thus, after the final round, we obtain an assignment can satisfy at least $(7/8)m$ clauses

Probabilistic methods

One of the most powerful method in combinatorics to show the existence of some object without giving an explicit construction

1. Construct a probability space over the possible objects and their attributes (e.g. graphs)
2. Show that $\Pr[\text{sampled object having the properties}] > 0$



Today's Lecture

- Introduction to High-Dimensional Probability
- Markov Inequality
- Chebyshev Inequality

Chebyshev Inequality

Variance

Let X be a real random variable. The variance of X is

$$\mathrm{Var}[X] \coloneqq \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 \geq 0$$

(Why?)

Chebyshev Inequality

Let X be a real random variable. Then

$$\Pr[|X - \mathbb{E}[X]| \geq t] \leq \frac{\mathrm{Var}[X]}{t^2} \quad \forall\, t > 0$$

Equivalently, if $\mathrm{Var}[X] = \sigma^2$, then

$$\Pr[|X - \mathbb{E}[X]| \geq t\sigma] \leq \min\{t^{-2}, 1\}$$

Chebyshev Inequality

Let X be a real random variable. Then

$$\Pr[|X - \mathbb{E}[X]| \geq t] \leq \frac{\mathrm{Var}[X]}{t^2} \quad \forall\, t > 0$$

Proof.

- $\Pr[|X - \mathbb{E}[X]| \geq t] = \Pr[|X - \mathbb{E}[X]|^2 \geq t^2]$
- Apply Markov inequality to $|X - \mathbb{E}[X]|^2$:

$$\Pr[|X - \mathbb{E}[X]|^2 \geq t^2] \leq \frac{\mathbb{E}[|X - \mathbb{E}[X]|^2]}{t^2} = \frac{\mathrm{Var}[X]}{t^2}$$

■

Chebyshev Inequality

Let X be a real random variable. Then

$$\Pr[|X - \mathbb{E}[X]| \geq t] \leq \frac{\mathrm{Var}[X]}{t^2} \quad \forall\, t > 0$$

How to estimate the variance?

- Range bound
- Tensorization of variance

Chebyshev Inequality

Range bound

Let X be a random variable whose support is contained in $[a, b]$. Then

$$\mathrm{Var}[X] \leq \frac{1}{4}(b-a)^2$$

Proof.

- Variational formulation (check by yourself): for *any* random variable X,

  $$\mathrm{Var}[X] = \inf_{\tau} \mathbb{E}[(X-\tau)^2]$$

  And the infimum is attained at $\tau = \mathbb{E}[X]$

- Set $\tau = (b+a)/2$. Then we have $|X - \tau| \leq (b-a)/2$

$\blacksquare$

Think: when the range bound is tight

Chebyshev Inequality

Range bound

Let X be a random variable whose support is contained in $[a, b]$. Then

$$\mathrm{Var}[X] \leq \frac{1}{4}(b-a)^2$$

Example:

- Let $X_1, \dots, X_n$ be independent random variables such that $\mathrm{supp}(X_i) \subset [a_i, b_i]$
- Let $Z = X_1 + \cdots + X_n$
- By the independence, we know that

$$\mathrm{Var}[Z] = \sum_{i=1}^{n} \mathrm{Var}[X_i] \leq \frac{1}{4}\sum_{i=1}^{n}(b_i - a_i)^2 =: \frac{1}{4}\sum_{i=1}^{n} c_i^2 = \frac{1}{4}\|\boldsymbol{c}\|_2^2$$

Chebyshev Inequality

Range bound

Let X be a random variable whose support is contained in $[a, b]$. Then

$$\mathrm{Var}[X] \leq \frac{1}{4}(b-a)^2$$

Example:

- Let $X_1, \dots, X_n$ be independent random variables such that $\mathrm{supp}(X_i) \subset [a_i, b_i]$
- Let $Z = X_1 + \cdots + X_n$
- On the other hand, it is easy to see that $\mathrm{supp}(Z) \subset [\sum_{i=1}^n a_i, \sum_{i=1}^n b_i]$
- Thus,

$$\mathrm{Var}[Z] \leq \frac{1}{4}\left(\sum_{i=1}^n b_i - \sum_{i=1}^n a_i\right)^2 = \frac{1}{4}\left(\sum_{i=1}^n c_i\right)^2 = \frac{1}{4}\|\boldsymbol{c}\|_1^2$$

Chebyshev Inequality

Range bound

Let X be a random variable whose support is contained in $[a, b]$. Then

$$\mathrm{Var}[X] \leq \frac{1}{4}(b-a)^2$$

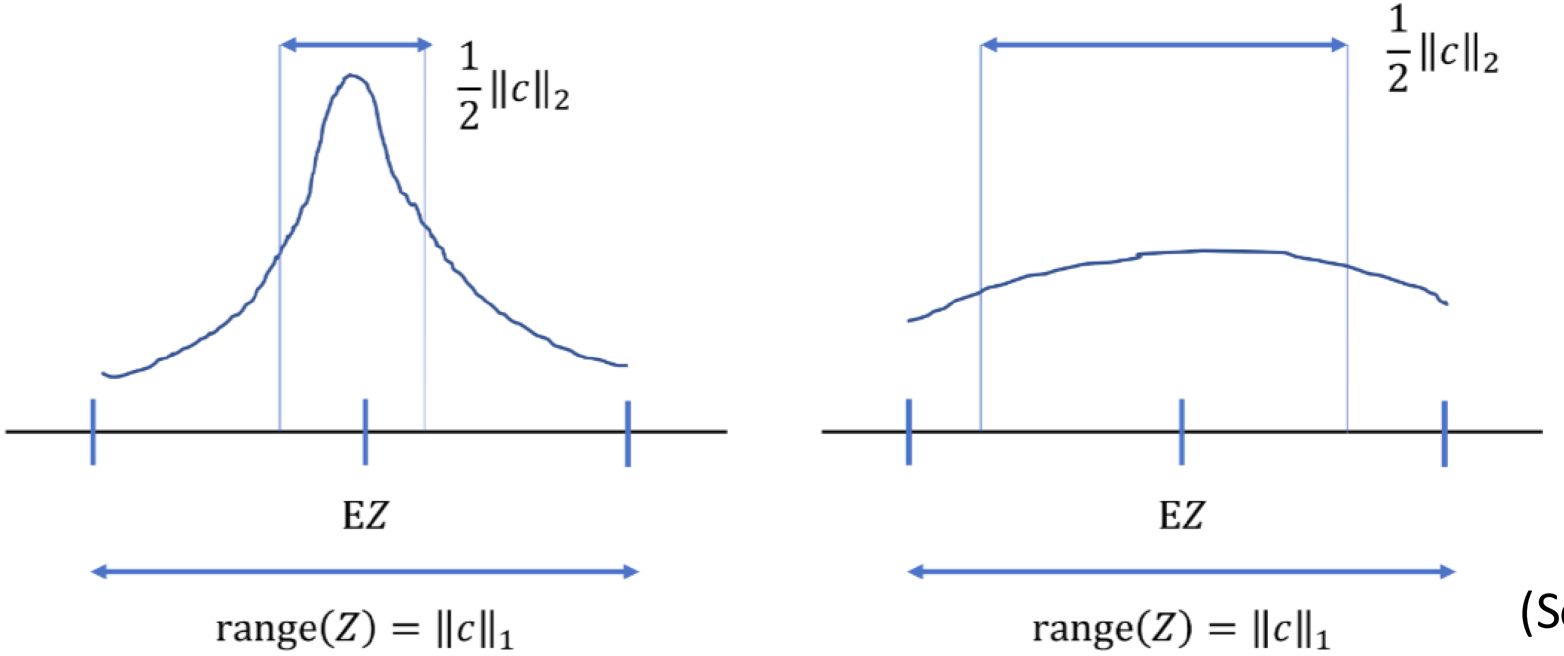
Example:

- Let $X_1, \dots, X_n$ be independent random variables such that $\mathrm{supp}(X_i) \subset [a_i, b_i]$
- Let $Z = X_1 + \cdots + X_n$

$$\mathrm{Var}[Z] \leq \frac{1}{4}\min\{\|\boldsymbol{c}\|_1^2, \|\boldsymbol{c}\|_2^2\}$$

- Think: when $\|\boldsymbol{c}\|_2 \ll \|\boldsymbol{c}\|_1$ and when $\|\boldsymbol{c}\|_2 \approx \|\boldsymbol{c}\|_1$

Chebyshev Inequality



(Source: Joel A. Tropp)

"A random variable that depends (in a 'smooth' way) on the influence of many independent variables (but not too much on any of them) is essentially constant."

Michel Talagrand '96

Chebyshev Inequality

Tensorization

- The motivation comes from the additivity law of variance for independent sum:

$$\mathrm{Var}[Z] = \mathrm{Var}[X_1] + \cdots + \mathrm{Var}[X_n]$$

- What if $Z = f(X_1, \ldots, X_n)$ a general function?

Chebyshev Inequality

Tensorization

- We define the coordinatewise expectation: for any $i \in [n]$,

$$\mathbb{E}_i[Z] \coloneqq \mathbb{E}[Z | X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n]$$

  Note that $\mathbb{E}_i[Z]$ is a random variable depending on $\{X_j : j \neq i\}$

$$\mathbb{E}_i[Z] \equiv \mathbb{E}_i[Z](X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n)$$

- We can also define the coordinatewise variance: for any $i \in [n]$,

$$\mathrm{Var}_i[Z] \coloneqq \mathbb{E}_i[(Z - \mathbb{E}_i[Z])^2] = \mathbb{E}_i[Z^2] - \mathbb{E}_i[Z]^2$$

Theorem. Suppose $X_1, \dots, X_n$ are independent random variables. Let $Z = f(X_1, \dots, X_n)$.

$$\mathrm{Var}[Z] \leq \mathbb{E}\left[\sum_{i=1}^{n} \mathrm{Var}_i[Z]\right]$$

Chebyshev Inequality

Tensorization

Theorem. Suppose $X_1, \dots, X_n$ are independent random variables. Let $Z = f(X_1, \dots, X_n)$.

$$\mathrm{Var}[Z] \le \mathbb{E}\left[\sum_{i=1}^{n} \mathrm{Var}_i[Z]\right]$$

- You can check that when $Z = X_1 + \cdots + X_n$, $\mathrm{Var}_i[Z] = \mathrm{Var}[X_i]$. The tensorization inequality for variances holds with equality.
- The proof uses Doob's martingale (we'll discuss it later in this course)

$$\mathbb{E}[\mathrm{Var}_i[f(X_1, \dots, X_n)]] = \sum_{x_1} \cdots \sum_{x_{i-1}} \sum_{x_{i+1}} \cdots \sum_{x_n} \mathrm{Var}[f(x_1, \dots, x_{i-1}, X_i, x_{i+1}, \dots, x_n)] \prod_{j \ne i} \Pr[X_j = x_j]$$

Chebyshev Inequality

Tensorization

Theorem. Suppose $X_1, \dots, X_n$ are independent random variables. Let $Z = f(X_1, \dots, X_n)$.

$$\mathrm{Var}[Z] \le \mathbb{E}\left[\sum_{i=1}^{n} \mathrm{Var}_i[Z]\right]$$

Corollary. For $Z = f(X_1, \dots, X_n)$, define the i-th discrete partial derivative as:

$$(D_i f)(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n) \coloneqq \sup_{z \in \mathrm{supp}(X_i)} f(x_1, \dots, x_{i-1}, z, x_{i+1}, \dots, x_n) - \inf_{z \in \mathrm{supp}(X_i)} f(x_1, \dots, x_{i-1}, z, x_{i+1}, \dots, x_n)$$

Then,

$$\mathrm{Var}[Z] \le \frac{1}{4} \sum_{i=1}^{n} \mathbb{E}[(D_i f)^2]$$

Proof.

- Range bound for $\mathrm{Var}_i[Z]$:

$$\mathrm{Var}_i[Z] \le \frac{1}{4}(D_i f)^2$$

Chebyshev Inequality

Tensorization

Corollary. For $Z = f(X_1, \dots, X_n)$, define the i-th discrete partial derivative as:

$$(D_i f)(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n) \coloneqq \sup_{z \in \mathrm{supp}(X_i)} f(x_1, \dots, x_{i-1}, z, x_{i+1}, \dots, x_n) - \inf_{z \in \mathrm{supp}(X_i)} f(x_1, \dots, x_{i-1}, z, x_{i+1}, \dots, x_n)$$

Then,

$$\mathrm{Var}[Z] \leq \frac{1}{4} \sum_{i=1}^{n} \mathbb{E}[(D_i f)^2]$$

- Can we further generalize it to dependent random variables?
- Related to the Poincaré inequality:

$$\mathrm{Var}[f(\boldsymbol{X})] \leq C_{\mathrm{PI}} \cdot \mathbb{E}[\|\nabla f(\boldsymbol{X})\|_2^2]$$