S&DS 602: High-Dimensional Probability and Applications

HDP finds application to many fields:

- Statistics/ML: high-dim. data, parameter estimation
- CS: randomized algorithms, average-case complexity
- Engineering: multi-agent systems, randomized codes
- Math: combinatorial structures, Markov chains, high-dim. geometry

This course: common principles, methods, techniques.

I. Concentration of measure

$X_1, \ldots, X_n$ independent r.v.'s. If $f(X_1, \ldots, X_n)$ depends "not too much" on any individual X_i, then it is "typically close to" its expectation.

Ex: (CLT) $X_1, \ldots, X_n$ iid, $\mathbb{E}X_i = 0$, $\operatorname{Var} X_i = \sigma^2$.

$$\mathbb{P}\left[\left|\frac{1}{n}\sum_{i=1}^{n} X_i\right| > t \cdot \frac{\sigma}{\sqrt{n}}\right] \longrightarrow 2(1 - \underbrace{\Phi(t)}_{\text{standard } N(0,1) \text{ CDF}}) \quad \text{as } n \to \infty.$$

Goals: (1) Non-asymptotic tail bounds

(2) Beyond sums of i.i.d. r.v.'s:

- sums of i.i.d. vectors and matrices
- U-statistics $\frac{1}{\binom{n}{2}} \sum_{i<j} k(X_i, X_j)$
- polynomials $p(X_1, \dots, X_n)$
- bounded differences

$$|f(x_1, \dots, x_i, \dots, x_n) - f(x_1, \dots, x_i', \dots, x_n)| \leq C.$$

- Lipschitz functions, i.e. $\|\nabla f(x_1, \dots, x_n)\|_2 \leq C$.

II. Suprema of random processes, $\sup_{t \in T} X_t$

If $\{X_t\}_{t \in T}$ "sufficiently continuous," then size of $\sup_{t \in T} X_t$ is controlled by "complexity" of T.

Examples:

(1) Norms of random vectors and matrices

$$\|v\|_2 = \sup_{u : \|u\|_2 = 1} u^T v, \quad \|A\|_{op} = \sup_{u, v : \|u\|_2 = \|v\|_2 = 1} u^T A v$$

(2) Gaussian processes, e.g.

Gaussian complexity: $\sup_{t \in T} g^T t$, $g_i \overset{iid}{\sim} N(0,1)$

Min-cut: $\sup_{x \in \{+1,-1\}^n} x^T A x$, $(A_{ij})_{i<j} \overset{iid}{\sim} N(0,1)$

(3) Empirical risk minimization:

$$\sup_{\theta \in \Theta} \left| \frac{1}{n} \sum_{i=1}^{n} \ell(\theta, X_i) - \mathbb{E}\ell(\theta, X) \right|$$

Lecture 1: Chernoff bound, subgaussian r.v.'s, martingale method

Example: $X_1, \ldots, X_n \in \{0,1\}$, iid Bernoulli$(p)$, $q > p$.

CLT: $\mathbb{P}\left[\frac{1}{n}\sum_i X_i \geq q\right] \approx 1 - \Phi\left(q-p \big/ \sqrt{\frac{p(1-p)}{n}}\right) \approx e^{-\frac{n(q-p)^2}{2p(1-p)}}$

for $q - p \approx \sqrt{\frac{p(1-p)}{n}}$, large n.

Cramer's Thm: $\mathbb{P}\left[\frac{1}{n}\sum_i X_i \geq q\right] \approx e^{-n D(q \| p)}$,

$D(q\|p) = (1-q)\log\frac{1-q}{1-p} + q \log \frac{q}{p}$, for $q - p \approx 1$, large n

$\left[\overset{\text{(Taylor exp.)}}{\approx} \frac{(q-p)^2}{2p(1-p)} \text{ for } q \approx p. \right]$

Non-asymptotic bound, valid for all $q > p$?

① Chebyshev/Markov inequality:

$$\mathbb{P}\left[\frac{1}{n}\sum_i X_i \geq q\right] \leq \mathbb{P}\left[\left(\frac{1}{n}\sum_i X_i - p\right)^2 \geq (q-p)^2\right]$$

$$\leq \frac{\mathbb{E}\left[\left(\frac{1}{n}\sum_i X_i - p\right)^2\right]}{(q-p)^2} \leq \frac{p(1-p)}{n(q-p)^2}$$

② k^{th}-moment:

$$\mathbb{P}\left[\frac{1}{n}\sum_i X_i \geq q\right] \leq \mathbb{P}\left[\left|\frac{1}{n}\sum_i X_i - p\right|^k \geq (q-p)^k\right]$$

$$\leq \frac{\mathbb{E}\left|\frac{1}{n}\sum_i X_i - p\right|^k}{(q-p)^k}$$ Sharp (for $k(n) \gg 1$) but harder to bound.

③ Moment/cumulant generating function:

$$\psi(\lambda) := \log \mathbb{E} e^{\lambda(X_i - p)}$$

$$\mathbb{P}\left[\frac{1}{n}\sum_i X_i \geq q\right] = \mathbb{P}\left[e^{\lambda\sum_i(X_i - p)} \geq e^{\lambda n(q-p)}\right] \quad \forall \lambda \geq 0$$

$$\leq e^{-\lambda n(q-p)}\, \mathbb{E} e^{\lambda \sum_i (X_i - p)} = e^{-n(\lambda(q-p) - \psi(\lambda))}$$

Here: $\psi(\lambda) = \log \mathbb{E} e^{\lambda X_i} - \lambda p = \log(1-p+pe^{\lambda}) - \lambda p.$

$$\sup_{\lambda \geq 0} \lambda(q-p) - \psi(\lambda) = D(q\|p).$$

$$\Rightarrow \mathbb{P}\left[\frac{1}{n}\sum_i X_i \geq q\right] \leq e^{-n \cdot D(q\|p)}$$

Abstracting this argument:

Thm (Chernoff bound): $X_1, \dots, X_n$ iid, $\mathbb{E}X_i = 0$. Let $\psi(\lambda) = \log \mathbb{E} e^{\lambda X_i}$.

Define the Fenchel-Legendre transform $\psi^*(t) = \sup_{\lambda \geq 0} \lambda t - \psi(\lambda)$

Then $\mathbb{P}\left[\frac{1}{n}\sum_i X_i \geq t\right] \leq e^{-n\psi^*(t)}$ for any $t \geq 0$.

Examples:

- $X_i \sim \text{Bern}(p) - p$. $\psi(\lambda) = \log(1-p+pe^{\lambda}) - \lambda p$, $\psi^*(t) = D(t+p \,\|\, p)$
- $X_i \sim \text{Pois}(\theta) - \theta$. $\psi(\lambda) = \theta(e^{\lambda} - \lambda - 1)$, $\psi^*(t) = (t+\theta)\log\left(1+\frac{t}{\theta}\right) - t$
- $X_i \sim \mathcal{N}(0, \sigma^2)$. $\psi(\lambda) = \frac{\lambda^2\sigma^2}{2}$, $\psi^*(t) = \frac{t^2}{2\sigma^2}$.

Def: A mean-zero r.v. X is $\underline{\sigma^2\text{-subgaussian}}$ if $\mathbb{E}e^{\lambda X} \leq e^{\frac{\lambda^2\sigma^2}{2}}$

(i.e. $\psi(\lambda) \leq \frac{\lambda^2\sigma^2}{2}$) for all $\lambda \in \mathbb{R}$.

Check: $\mathcal{N}(0,\sigma^2)$, $\text{Bern}(p)-p$ are subgaussian. $\text{Pois}(\theta)-\theta$ is not.

Lemma (Hoeffding): If $a \leq X \leq b$ a.s., then $X - \mathbb{E}X$ is $\frac{(b-a)^2}{4}$-subgaussian.

Proof: Assume WLOG $\mathbb{E}X = 0$.

Define $\tilde{X}$ w/ law $\mathbb{P}[\tilde{X} \in A] = \mathbb{E}[\mathbb{1}_{X \in A}\, e^{\lambda X}] / \mathbb{E}e^{\lambda X}$

$$\Rightarrow \psi(\lambda) = \log \mathbb{E} e^{\lambda X}, \quad \psi(0) = 0$$

$$\psi'(\lambda) = \frac{\mathbb{E} X e^{\lambda X}}{\mathbb{E} e^{\lambda X}} = \mathbb{E} \tilde{X}, \quad \psi'(0) = \mathbb{E} X = 0.$$

$$\psi''(\lambda) = \frac{\mathbb{E} X^2 e^{\lambda X}}{\mathbb{E} e^{\lambda X}} - \left(\frac{\mathbb{E} X e^{\lambda X}}{\mathbb{E} e^{\lambda X}} \right)^2 = \mathbb{E} \tilde{X}^2 - (\mathbb{E} \tilde{X})^2 = \operatorname{Var} \tilde{X} \leq \frac{(b-a)^2}{4}.$$

So $\psi(\lambda) = \int_0^\lambda \psi'(t)\, dt = \int_0^\lambda \int_0^t \psi''(s)\, ds\, dt \leq \frac{(b-a)^2}{4} \int_0^\lambda \int_0^t ds\, dt = \frac{\lambda^2 (b-a)^2}{8}$. ∎

Prop: If X is mean-zero and σ^2-subgaussian, then

$$\mathbb{P}[X \geq t] \leq e^{-\frac{t^2}{2\sigma^2}} \quad \forall t \geq 0.$$

Proof: $\mathbb{P}[X \geq t] \leq e^{-\lambda t} \mathbb{E} e^{\lambda X} \leq e^{-\lambda t + \frac{\lambda^2 \sigma^2}{2}}$. Pick $\lambda = \frac{t}{\sigma^2}$. ∎

Prop: The following are equivalent:

(a) $\exists K_1 > 0$ s.t. $\mathbb{P}[|X| \geq t] \leq 2 e^{-t^2/K_1^2} \quad \forall t \geq 0$

(b) $\exists K_2 > 0$ s.t. $\|X\|_{L^p} := (\mathbb{E}|X|^p)^{1/p} \leq K_2 \sqrt{p} \quad \forall p \geq 1.$

(c) $\exists K_3 > 0$ s.t. $\mathbb{E} e^{\frac{X^2}{K_3^2}} \leq 2.$

If $\mathbb{E} X = 0$, then these are also equivalent to

(d) $\exists K_4 > 0$ s.t. $\psi(\lambda) \leq \lambda^2 K_4^2 \quad \forall \lambda \in \mathbb{R}.$

Proof: (a)$\Rightarrow$(b): [Lemma: If $Z \geq 0$ a.s., then $\mathbb{E}Z = \int_0^\infty \mathbb{P}[Z \geq t]\,dt$.

Proof: $Z = \int_0^Z dt = \int_0^\infty \mathbb{1}\{t \leq Z\}\,dt$. Take $\mathbb{E}$ on both sides.]

$$\mathbb{E}|X|^p = \int_0^\infty \mathbb{P}[|X|^p \geq t]\,dt = \int_0^\infty \mathbb{P}[|X|^p \geq u^p] \cdot p u^{p-1}\,du$$

$$\overset{(a)}{\leq} \int_0^\infty 2e^{-u^2/K_1^2} \cdot p u^{p-1}\,du \leq 2p\left(\frac{K_1}{\sqrt{2}}\right)^p \int_0^\infty e^{-v^2/2} v^{p-1}\,dv \leq (CK_1\sqrt{p})^p$$

(b)$\Rightarrow$(c): $$\mathbb{E}e^{\lambda^2 X^2} = \sum_{k\geq 0} \frac{\lambda^{2k}}{k!}\mathbb{E}X^{2k} \overset{(b)}{\leq} \sum_{k\geq 0} \frac{\lambda^{2k}}{k!}(K_2\sqrt{2k})^{2k}$$

Apply $k! \geq (k/e)^k$: $\mathbb{E}e^{\lambda^2X^2} \leq \sum_{k\geq 0} (2eK_2^2\lambda^2)^k \leq 2$ for $\lambda = \frac{1}{CK_2}$

(c)$\Rightarrow$(a): $$\mathbb{P}[|X| \geq t] \leq \mathbb{P}[e^{X^2/K_3^2} \geq e^{t^2/K_3^2}]$$

$$\leq e^{-t^2/K_3^2} \cdot \mathbb{E}e^{X^2/K_3^2} \overset{(c)}{\leq} 2e^{-t^2/K_3^2}.$$

If $\mathbb{E}X = 0$:

(b)$\Rightarrow$(d): $$\mathbb{E}e^{\lambda X} = 1 + \sum_{k\geq 2} \frac{\lambda^k}{k!}\mathbb{E}X^k.$$

For $k \geq 1$: $\mathbb{E}|\lambda X|^{2k+1} \leq \frac{1}{2}\left(\mathbb{E}|\lambda X|^{2k} + \mathbb{E}|\lambda X|^{2k+2}\right)$

$$\Rightarrow \mathbb{E}e^{\lambda X} \leq \sum_{k\geq 0} \frac{2^k \lambda^{2k}\,\mathbb{E}X^{2k}}{(2k)!} \overset{(b)}{\leq} \sum_{k\geq 0} \frac{2^k \lambda^{2k}(K_2\sqrt{2k})^{2k}}{(2k)!}$$

Apply $k^k \leq e^k \cdot k!$, $(2k)! \geq (2^k \cdot k!)^2$: $\mathbb{E}e^{\lambda X} \leq \sum_{k\geq 0} \frac{(CK_2^2\lambda^2)^k}{k!} = e^{CK_2^2\lambda^2}$.

(d)$\Rightarrow$(a): Apply previous prop. to X and $-X$.

$\blacksquare$

Def: $\|X\|_{\psi_2} = \inf\{K>0 : \mathbb{E}e^{X^2/K^2} \le 2\}$ is the <u>subgaussian norm</u>.

[Vershynin takes $\|X\|_{\psi_2} < \infty$ to be definition of subgaussian if $\mathbb{E}X \neq 0$.]

Thm (Hoeffding's inequality): If $X_1, \dots, X_n$ are independent, $\mathbb{E}X_i = 0$, X_i is σ_i^2-subgaussian, then $\sum_{i=1}^n X_i$ is $\sum_{i=1}^n \sigma_i^2$-subgaussian.

Thus, $\mathbb{P}\left[\sum_{i=1}^n X_i \ge t\right] \le e^{-\frac{t^2}{2\sum_{i=1}^n \sigma_i^2}}$ for all $t \ge 0$.

Proof: $\mathbb{E}e^{\lambda \sum_{i=1}^n X_i} = \prod_{i=1}^n \mathbb{E}e^{\lambda X_i} \le \prod_{i=1}^n e^{\frac{\lambda^2 \sigma_i^2}{2}} = e^{\frac{\lambda^2}{2}\sum_{i=1}^n \sigma_i^2}$ ∎

Cor: If $a_i \le X_i \le b_i$ a.s., then

$$\mathbb{P}\left[\sum_{i=1}^n X_i - \mathbb{E}X_i \ge t\right] \le e^{-\frac{2t^2}{\sum_{i=1}^n (b_i - a_i)^2}} \text{ for all } t \ge 0.$$

Proof: Follows from Hoeffding's Lemma. ∎

Example: $X_1, \dots, X_n \sim \text{Bernoulli}(p)$. Then $\mathbb{P}\left[\frac{1}{n}\sum_i X_i - p \ge t\right] \le e^{-2nt^2}$.

(Cruder than the above Chernoff bound, esp. for p far from $1/2$.)

<u>Martingale method</u>

How to show concentration of <u>non-linear</u> functions $f(X_1, \dots, X_n)$?

Idea: Write $M_i = \mathbb{E}[f(X_1, \dots, X_n) \mid X_1, \dots, X_i]$

$$f(X_1,\dots,X_n) - \mathbb{E}f(X_1,\dots,X_n) = M_n - M_0 = \sum_{i=1}^{n} \underbrace{M_i - M_{i-1}}_{:=\Delta_i}$$

$\{M_i\}_{i=1}^{n}$ is a martingale (w.r.t. filtration $\mathcal{F}_i = \sigma(X_1,\dots,X_i)$)

i.e. $\mathbb{E}[M_i \mid \mathcal{F}_{i-1}] = M_{i-1}$.

Thm (Azuma-Hoeffding): Suppose $\{M_i\}_{i=1}^{n}$ is a martingale adapted to $\{\mathcal{F}_i\}_{i=1}^{n}$ and $\Delta_i := M_i - M_{i-1}$ satisfy $\mathbb{E}[e^{\lambda\Delta_i} \mid \mathcal{F}_{i-1}] \le e^{\frac{\lambda^2\sigma_i^2}{2}}$ a.s. (Δ_i is conditionally σ_i^2-subgaussian). Then $\sum_{i=1}^{n}\Delta_i = M_n - M_0$ is $\sum_{i=1}^{n}\sigma_i^2$-subgaussian, and

$$\mathbb{P}\left[\sum_{i=1}^{n}\Delta_i \ge t\right] \le e^{-\frac{t^2}{2\sum_{i=1}^{n}\sigma_i^2}}$$

Proof: $\mathbb{E}e^{\lambda\sum_{i=1}^{k}\Delta_i} = \mathbb{E}\left[\mathbb{E}\left[e^{\lambda\sum_{i=1}^{k}\Delta_i} \mid \mathcal{F}_{k-1}\right]\right]$

$$= \mathbb{E}\left[e^{\lambda\sum_{i=1}^{k-1}\Delta_i}\cdot\mathbb{E}\left[e^{\lambda\Delta_k} \mid \mathcal{F}_{k-1}\right]\right] \le e^{\frac{\lambda^2\sigma_k^2}{2}}\,\mathbb{E}e^{\lambda\sum_{i=1}^{k-1}\Delta_i}$$

Then by induction, $\mathbb{E}e^{\lambda\sum_{i=1}^{n}\Delta_i} \le e^{\frac{\lambda^2}{2}\sum_{i=1}^{n}\sigma_i^2}$, and tail bound follows as before. $\blacksquare$

Corollary (McDiarmid's bounded differences): Let

$$\|D_i f\|_\infty = \sup_{x_1,\dots,x_{i-1},x_{i+1},\dots,x_n} \Big[\sup_z f(x_1,\dots,x_{i-1},z,x_{i+1},\dots,x_n) - \inf_z f(x_1,\dots,x_{i-1},z,x_{i+1},\dots,x_n) \Big]$$

If $X_1,\dots,X_n$ are independent, then

$$\mathbb{P}\left[f(X_1,\dots,X_n) - \mathbb{E} f(X_1,\dots,X_n) \geq t \right] \leq e^{-\frac{2t^2}{\sum_{i=1}^n \|D_i f\|_\infty^2}}.$$

Proof: Take

$$\Delta_i = \mathbb{E}[f(X_1,\dots,X_n) \mid X_1,\dots,X_i] - \mathbb{E}[f(X_1,\dots,X_n) \mid X_1,\dots,X_{i-1}].$$

Here $\Delta_i \geq \underbrace{\mathbb{E}\left[\inf_z f(X_1,\dots,X_{i-1},z,X_{i+1},\dots,X_n) - f(X_1,\dots,X_n) \mid X_1,\dots,X_{i-1} \right]}_{:= A_i}$

$\Delta_i \leq \underbrace{\mathbb{E}\left[\sup_z f(X_1,\dots,X_{i-1},z,X_{i+1},\dots,X_n) - f(X_1,\dots,X_n) \mid X_1,\dots,X_{i-1} \right]}_{:= B_i}$

where A_i, B_i are $\mathcal{F}_{i-1}$-measurable, $|B_i - A_i| \leq \|D_i f\|_\infty$ a.s.

Then Δ_i is conditionally $\frac{\|D_i f\|_\infty^2}{4}$-subgaussian by Hoeffding Lemma, and result follows from Azuma-Hoeffding. ∎

Example (Rademacher complexity): Let $\varepsilon_1, \dots, \varepsilon_n$ be iid Rademacher variables, i.e. $\mathbb{P}[\varepsilon_i = \pm 1] = \frac{1}{2}$. Let $T \subseteq \mathbb{R}^n$,

$$f(\varepsilon_1, \dots, \varepsilon_n) = \sup_{t \in T} \varepsilon^T t.$$

Then $\|D_i f\|_\infty = 2 \sup_{t \in T} |t_i|$, so

$$\mathbb{P}\left[|f(\varepsilon_1, \dots, \varepsilon_n) - \mathbb{E} f(\varepsilon_1, \dots, \varepsilon_n)| \geq u \right] \leq 2 e^{-\frac{u^2}{2\sigma^2}}, \quad \sigma^2 := \sum_{i=1}^n \sup_{t \in T} t_i^2.$$

Later in the course: Improve this to $\sigma^2 = \sup_{t \in T} \sum_{i=1}^n t_i^2 = \sup_{t \in T} \|t\|_2^2$.

Example (U-statistic): Let $X_1, \dots, X_n$ be iid, $h: \mathbb{R}^2 \to \mathbb{R}$ with $\|h\|_\infty \leq B$, $h(x,y) = h(y,x)$. Consider

$$f(X_1, \dots, X_n) = \frac{1}{\binom{n}{2}} \sum_{i<j} h(X_i, X_j).$$

Then
$$\|D_i f\|_\infty \leq \sup_{x, z, z'} \frac{1}{\binom{n}{2}} \sum_{j: j \neq i} |h(z, x_j) - h(z', x_j)|$$

$$\leq \frac{(n-1) \cdot 2B}{\binom{n}{2}} = \frac{4B}{n}.$$

So $$\mathbb{P}\left[|f(X_1, \dots, X_n) - \mathbb{E} f(X_1, \dots, X_n)| \geq t \right] \leq 2 e^{-\frac{nt^2}{8B^2}}.$$

More on this example later.

Example (Shamir, Spencer '86): Let $G \sim \mathcal{G}(n,p)$ be an Erdős-Rényi graph, i.e. $\mathbb{P}[i \sim j] = p$ independently for all $i \neq j \in \{1, \dots, n\}$.

Let $f(G)$ be the "chromatic number" — minimal # colors needed to color all vertices s.t. no edge has 2 vertices of same color.

Let X_1 = all edges $1 \to \{2, 3, \dots, n\}$.

X_2 = all edges $2 \to \{3, \dots, n\}$

$\vdots$

X_{n-1} = edge $n-1 \to n$.

Fixing all but X_i, $f(G)$ smallest when $X_i = (0, 0, \dots, 0)$ largest when $X_i = (1, 1, \dots, 1)$, and $\|D_i f\|_\infty \leq 1$.

$$\Rightarrow \mathbb{P}\left[\,|f(G) - \mathbb{E}f(G)| \geq t\,\right] \leq 2e^{-\frac{2t^2}{n-1}}$$

For any $p \in (0,1)$, as $n \to \infty$, $f(G) = \mathbb{E}f(G) + O_p(\sqrt{n})$.

[It is known $\mathbb{E}f(G) \asymp \frac{n}{\log n}$.]