Lecture 2: Heavier-tailed r.v.'s, random vectors in high dimensions

Recall: X is σ^2-subgaussian ($\mathbb{E}X=0$) if $\psi(\lambda) = \log \mathbb{E}e^{\lambda X} \le e^{\frac{\lambda^2\sigma^2}{2}}$.

$$\Leftrightarrow \mathbb{P}[|X| \ge t] \le 2e^{-t^2/k^2} \quad \forall t \ge 0.$$

$X_1, \dots, X_n$ independent, $\mathbb{E}X_i = 0$, σ^2-subgaussian

$$\Rightarrow \mathbb{P}\left[\sum_{i=1}^n X_i \ge t\right] \le e^{-\frac{t^2}{2n\sigma^2}} \quad \text{(Hoeffding)}$$

What if X_i's have heavier tails?

- By CLT, for $t \simeq \sqrt{n}$, $\mathbb{P}\left[\sum_{i=1}^n X_i \ge t\right] \approx e^{-O(t^2/n)}$
- However for $t \gg \sqrt{n}$, a lower bound is

$$\mathbb{P}\left[\sum_{i=1}^n X_i \ge t\right] \ge \mathbb{P}[X_1 \ge (1+\varepsilon)t] \cdot \underbrace{\mathbb{P}\left[\left|\sum_{i=2}^n X_i\right| \le \varepsilon t\right]}_{\approx 1}$$

$$\approx \mathbb{P}[X_1 \ge (1+\varepsilon)t]$$

Expect to see two types of bounds, for moderate vs. large t.

Def: A mean-zero r.v. X is (σ^2, b)-subexponential if $\psi(\lambda) \le \frac{\lambda^2\sigma^2}{2}$ $\forall |\lambda| \le \frac{1}{b}$.

Prop: If $\mathbb{E}X=0$ and X is (σ^2, b)-subexponential, then

$$\mathbb{P}[X \ge t] \le e^{-\frac{1}{2}\min(t^2/\sigma^2,\, t/b)}$$

[Gaussian tail $e^{-t^2/2\sigma^2}$ for $t \le \frac{\sigma^2}{b}$, exp. tail $e^{-t/2b}$ for $t > \frac{\sigma^2}{b}$.]

Proof: $\mathbb{P}[X \ge t] \le e^{-\lambda t}\, \mathbb{E} e^{\lambda X} \le e^{-\lambda t + \frac{\lambda^2 \sigma^2}{2}}$ $\forall |\lambda| \le \frac{1}{b}$.

If $t < \frac{\sigma^2}{b}$: Pick $\lambda = \frac{t}{\sigma^2} \Rightarrow \mathbb{P}[X \ge t] \le e^{-\frac{t^2}{2\sigma^2}}$.

If $t \ge \frac{\sigma^2}{b}$: Pick $\lambda = \frac{1}{b} \Rightarrow \mathbb{P}[X \ge t] \le e^{-\frac{t}{b} + \frac{\sigma^2}{2b^2}} \le e^{-\frac{t}{2b}}$. ∎

Example: $X = Z^2$, $Z \sim N(0,1)$. $\mathbb{E} e^{\lambda(X-1)} = \frac{1}{\sqrt{1-2\lambda}} e^{-\lambda}$ $\forall \lambda < 1/2$.

$\Rightarrow \psi(\lambda) = -\frac{1}{2}\log(1-2\lambda) - \lambda \le 2\lambda^2$ $\forall |\lambda| \le 1/4$.

$\Rightarrow X-1$ is $(4,4)$-subexponential.

Example: $X \sim \chi^2_n$. $\psi(\lambda) = \log \mathbb{E} e^{\lambda(X-n)} \le 2n\lambda^2$ $\forall |\lambda| \le 1/4$.

$\Rightarrow X-n$ is $(4n, 4)$-subexponential

$\Rightarrow \mathbb{P}[|X-n| \ge t] \le 2 e^{-\frac{1}{2}\min\left(\frac{t^2}{4n}, \frac{t}{4}\right)}$

Thm (Bernstein's inequality): If $X_1, \dots, X_n$ are independent, $\mathbb{E}X_i = 0$, X_i is (σ_i^2, b_i)-subexponential, then $\sum_{i=1}^n X_i$ is $\left(\sum_{i=1}^n \sigma_i^2, \max_{i=1}^n b_i\right)$-subexponential. Thus,

$$\mathbb{P}\left[\sum_{i=1}^n X_i \ge t\right] \le e^{-\frac{1}{2}\min\left(\frac{t^2}{\sum_i \sigma_i^2}, \frac{t}{\max_i b_i}\right)}$$

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

$$\left(\forall |\lambda| \le \frac{1}{\max_i b_i}\right)$$

$\blacksquare$

Lemma (Bernstein): If $\mathbb{E} X = 0$, $\mathrm{Var}\, X = \sigma^2$, $|X| \le b$ a.s., then

$$\psi(\lambda) \le \frac{\lambda^2 \sigma^2}{2(1 - |\lambda| b/3)} \qquad \forall |\lambda| < \frac{3}{b}$$

In particular, X is $(2\sigma^2, \frac{2b}{3})$-subexponential.

Proof: $\mathbb{E} e^{\lambda X} = \sum_{k \ge 0} \frac{\lambda^k}{k!} \mathbb{E} X^k \le 1 + \frac{\lambda^2 \sigma^2}{2} + \sum_{k \ge 3} \frac{\lambda^2 \sigma^2}{1 \cdot 2 \cdot 3^{k-2}} (|\lambda| b)^{k-2}$

$$= 1 + \frac{\lambda^2 \sigma^2}{2} \cdot \frac{1}{1 - |\lambda| b/3} \le e^{\frac{\lambda^2 \sigma^2}{2(1 - |\lambda| b/3)}} \quad \forall |\lambda| < \frac{3}{b}.$$

Thus $\psi(\lambda) \le \frac{\lambda^2 \sigma^2}{2(1 - |\lambda| b/3)}$. Further bound by $\lambda^2 \sigma^2$ for $|\lambda| \le \frac{3}{2b}$. $\blacksquare$

Cor: If $\mathbb{E} X_i = 0$, $|X_i| \le b$ a.s., $\mathrm{Var}\, X_i \le \sigma^2$, then

$$\mathbb{P}\left[\sum_{i=1}^n X_i \ge t\right] \le e^{-\frac{t^2/2}{n\sigma^2 + bt/3}}.$$

Proof: X_i is $(2\sigma^2, \frac{2b}{3})$-subexponential $\Rightarrow \mathbb{P}\left[\sum_{i=1}^n X_i \ge t\right] \le e^{-\min\left(\frac{t^2}{4n\sigma^2}, \frac{3t}{4b}\right)}$.

To improve the constants: Use $\psi(\lambda) \le \frac{\lambda^2 \sigma^2}{2(1 - |\lambda| b/3)}$ $\forall |\lambda| < \frac{3}{b}$ from Bernstein's Lemma. Pick $\lambda = \frac{t}{n\sigma^2 + bt/3}$:

$$\mathbb{P}\left[\sum_{i=1}^n X_i \ge t\right] \le e^{-\lambda t + n\psi(\lambda)} \le e^{-\lambda t + \frac{\lambda^2 n \sigma^2}{2(1 - \lambda b/3)}} = e^{-\frac{t^2/2}{n\sigma^2 + bt/3}}$$

$\blacksquare$

Prop: The following are equivalent:

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

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

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

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

(d) $\exists K_4 > 0$ s.t. $\psi(\lambda) \leq K_4^2 \lambda^2$ $\forall |\lambda| \leq \frac{1}{K_4}$.

Proof: Similar to subgaussian case.

[Note: (d) $\Rightarrow$ $\mathbb{P}[|X| \geq t] \leq 2e^{-\frac{1}{4}\min\left(\frac{t^2}{K_4^2}, \frac{t}{K_4}\right)}$ by the above.

For $|t| \geq K_4$, this shows (a) w/ $K_1 = K_4$.

For $|t| \leq K_4$, pick $K_1 = K_4/\log 2$ $\Rightarrow$ (a) is vacuous as $\mathbb{P}[|X| \geq t] \leq 1$.

So Gaussian part of tail can be absorbed into the constants.]

Def: $\|X\|_{\psi_1} = \inf\{K > 0 : \mathbb{E} e^{|X|/K} \leq 2\}$ is the subexponential norm

(Sufficient to characterize (σ^2, b)-subexponential r.v.'s in this way if σ, b are of the same order.)

Cor: If $\mathbb{E}X_i = 0$, $\|X_i\|_{\psi_1} \le K$, then for an absolute constant $c > 0$,

$$\mathbb{P}\left[\sum_{i=1}^n X_i \ge t\right] \le e^{-c\cdot\min\left(\frac{t^2}{nK^2}, \frac{t}{K}\right)}.$$

Prop: $\|X - \mathbb{E}X\|_{\psi_1} \le C\|X\|_{\psi_1}$ for a universal constant $C > 0$.

Proof: Let $K = \|X\|_{\psi_1}$. By Prop. $(c)\Rightarrow(b)$, $|\mathbb{E}X| \le CK$

$$\Rightarrow \|X - \mathbb{E}X\|_{\psi_1} \le \|X\|_{\psi_1} + \|\mathbb{E}X\|_{\psi_1} \le C'K.$$

Prop: $\|X^2\|_{\psi_1} = \|X\|_{\psi_2}^2$.

Proof: Immediate from defs.

$$\|X^2\|_{\psi_1} = \inf\{K>0 : \mathbb{E}e^{\frac{X^2}{K}} \le 2\} = \left(\inf\{L>0 : \mathbb{E}e^{\frac{X^2}{L^2}} \le 2\}\right)^2 = \|X\|_{\psi_2}^2. \quad \blacksquare$$

Prop: $\|XY\|_{\psi_1} \le \|X\|_{\psi_2} \cdot \|Y\|_{\psi_2}$ for any (not necessarily independent) X, Y

Proof: Suppose $\|X\|_{\psi_2} = K$, $\|Y\|_{\psi_2} = L$. Then

$$\mathbb{E}e^{\frac{|XY|}{KL}} \le \mathbb{E}e^{\frac{1}{2}\left(\frac{X^2}{K^2} + \frac{Y^2}{L^2}\right)} \le \mathbb{E}\left[\frac{1}{2}\left(e^{\frac{X^2}{K^2}} + e^{\frac{Y^2}{L^2}}\right)\right] \le 2.$$

So $\|XY\|_{\psi_1} \le KL$.

$\blacksquare$

Random vectors in high dimensions

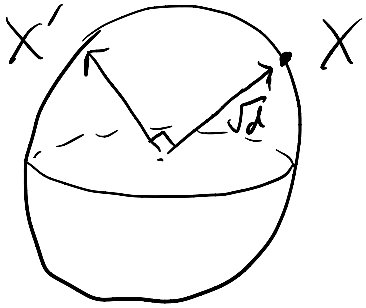
Prop: Let $X = (X_1, \ldots, X_d) \in \mathbb{R}^d$ have independent entries, $\mathbb{E}X_i = 0$, $\operatorname{Var} X_i = 1$, X_i σ^2-subgaussian. Then for a universal constant $c > 0$:

(a) $\mathbb{P}\left[\,\left|\|X\|_2 - \sqrt{d}\right| \geq t\right] \leq 2e^{-\frac{ct^2}{\sigma^4}}$

(b) If X' is an independent copy of X,

$$\mathbb{P}\left[\frac{|X^T X'|}{\|X\|_2 \cdot \|X'\|_2} \geq t\right] \leq 2e^{-\frac{cd}{\sigma^4}\min(t^2, t)} + 2e^{-\frac{cd}{\sigma^4}}.$$

For $\sigma^2 = O(1)$: $\|X\|_2 = \sqrt{d} \pm O_{\mathbb{P}}(1)$, $\dfrac{|X^T X'|}{\|X\|_2 \cdot \|X'\|_2} = O_{\mathbb{P}}\left(\dfrac{1}{\sqrt{d}}\right)$



Proof: (a) Note: $\dfrac{\sigma^2}{2} \geq \lim_{\lambda \to 0} \dfrac{\psi(\lambda)}{\lambda^2} = \lim_{\lambda \to 0} \dfrac{\log\left(1 + \frac{\lambda^2}{2}\mathbb{E}X_i^2 + \ldots\right)}{\lambda^2} = \dfrac{1}{2}$, so $\sigma^2 \geq 1$.

We have $\|X_i^2 - 1\|_{\psi_1} \leq C\|X_i^2\|_{\psi_1} = C\|X_i\|_{\psi_2}^2 \leq C'\sigma^2$.

$$\Rightarrow \mathbb{P}\left[\left|\frac{1}{d}\|X\|_2^2 - 1\right| \geq u\right] = \mathbb{P}\left[\left|\sum_{i=1}^d (X_i^2 - 1)\right| \geq du\right]$$

$$\leq 2e^{-c\min\left(\frac{du^2}{\sigma^4}, \frac{du}{\sigma^2}\right)} \leq 2e^{-\frac{cd}{\sigma^4}\min(u^2, u)}$$

If $z \geq 0$ and $|z-1| \geq \delta$, then $|z^2-1| \geq \max(\delta, \delta^2)$. So

$$\mathbb{P}\left[\left|\tfrac{1}{\sqrt{d}}\|X\|_2 - 1\right| \geq \delta\right] \leq \mathbb{P}\left[\left|\tfrac{1}{d}\|X\|_2^2 - 1\right| \geq \max(\delta, \delta^2)\right] \leq 2e^{-\frac{cd\delta^2}{\sigma^4}}$$

Take $\delta = t/\sqrt{d}$

(b) Similarly $\|X_i X_i'\|_{\psi_1} \leq \|X_i\|_{\psi_2} \|X_i'\|_{\psi_2} \leq C\sigma^2$

$$\Rightarrow \mathbb{P}\left[\left|\tfrac{1}{d} X^T X'\right| \geq u\right] \leq 2e^{-\frac{cd}{\sigma^4}\min(u^2, u)}$$

$$\Rightarrow \mathbb{P}\left[|X^T X'| / \|X\|_2 \|X'\|_2 \geq t\right]$$

$$\leq \mathbb{P}\left[\left|\tfrac{1}{d} X^T X'\right| \geq \tfrac{t}{4}\right] + \mathbb{P}\left[\|X\|_2 < \tfrac{\sqrt{d}}{2}\right] + \mathbb{P}\left[\|X'\| < \tfrac{\sqrt{d}}{2}\right]$$

$$\leq 2e^{-\frac{c'd}{\sigma^4}\min(t^2, t)} + 2e^{-\frac{c'd}{\sigma^4}}$$

$\blacksquare$

Thm (Johnson-Lindenstrauss): Let $\mathcal{A} \subset \mathbb{R}^D$ with $|\mathcal{A}| = n$ arbitrary. There is an absolute constant $C > 0$ s.t. for any $\varepsilon > 0$, if $d \geq \frac{C}{\varepsilon^2} \log n$ (no dependence on D) then there exists $P \in \mathbb{R}^{d \times D}$ for which

$$(1-\varepsilon)\|a - a'\|_2 \leq \|Pa - Pa'\|_2 \leq (1+\varepsilon)\|a - a'\|_2 \quad \forall a, a' \in \mathcal{A}.$$

Proof (probabilistic method):

Let Z_{ij} be iid $\mathbb{E} Z_{ij} = 0$, $\operatorname{Var} Z_{ij} = 1$, Z_{ij} σ^2-subgaussian for a constant $\sigma^2 > 0$, and take $P = \frac{1}{\sqrt{d}} Z$.

Fix any $u \in \left\{ \frac{a - a'}{\|a - a'\|_2} : a, a' \in A \right\}$, set $X = Zu$.

Then $X_1, \ldots, X_d$ are independent, $X_i = Z_{i1} u_1 + \ldots + Z_{iD} u_D$

$\Rightarrow \mathbb{E} X_i = 0$, $\operatorname{Var} X_i = u_1^2 + \ldots + u_D^2 = \|u\|_2^2 = 1$,

X_i is $(u_1^2 \sigma^2 + \ldots + u_D^2 \sigma^2) = \sigma^2$-subgaussian by Hoeffding's ineq.

So $\mathbb{P}\left[\|Pu\|_2 \notin (1-\varepsilon, 1+\varepsilon) \right]$

$$= \mathbb{P}\left[\left| \frac{1}{\sqrt{d}} \|X\|_2 - 1 \right| \geq \varepsilon \right] \leq 2 e^{-\frac{c d \varepsilon^2}{\sigma^4}}$$

$$\Rightarrow \mathbb{P}\left[(1-\varepsilon)\|a - a'\|_2 \leq \|Pa - Pa'\|_2 \leq (1+\varepsilon)\|a - a'\|_2 \ \forall a, a' \in A \right]$$

$$\geq 1 - \binom{n}{2} \cdot 2 e^{-\frac{c d \varepsilon^2}{\sigma^4}} \geq 1 - e^{-\frac{c d \varepsilon^2}{\sigma^4} + 2 \log n}$$

This is positive for $d \geq \frac{C}{\varepsilon^2} \log n$, so such a P exists. ∎

Sums of heavy-tailed r.v.'s

For X with tails heavier than exponential, its MGF may not exist.

Two common approaches for tail bounds:

(1) Truncation

(2) Moment inequalities.

To illustrate (1):

Thm: Suppose $X_1, \ldots, X_n$ are independent, $\mathbb{E}X_i = 0$, $\mathbb{P}[|X_i| \geq t] \leq 2e^{-(\frac{t}{K})^\alpha}$ $\forall t \geq 0$, some $K > 0$, $\alpha \in (0,1)$. Then there exist $C, c(\alpha) > 0$ s.t.

$$\mathbb{P}\left[\left|\sum_{i=1}^n X_i\right| \geq t\right] \leq Ce^{-c(\alpha)\min\left(\frac{t^2}{K^2 n}, \left(\frac{t}{K}\right)^\alpha\right)}$$

[For more general statements, see Bakhshizadeh, Maleki, de la Pena '23.]

Proof: Let $Z_i = X_i \mathbb{1}\{X_i \leq L\}$

$$\mathbb{P}\left[\sum_{i=1}^n X_i \geq t\right] \leq \mathbb{P}\left[\sum_{i=1}^n Z_i \geq t\right] + \mathbb{P}[X_i > L \text{ for some } i \in \{1, \ldots, n\}]$$

$$\leq \mathbb{P}\left[\sum_{i=1}^n Z_i \geq t\right] + 2ne^{-(L/K)^\alpha}$$

$$\mathbb{E}e^{\lambda Z_i} = 1 + \lambda \underbrace{\mathbb{E}Z_i}_{\leq \mathbb{E}X_i = 0} + \frac{\lambda^2}{2}\mathbb{E}Z_i^2 e^{\lambda Y_i} \text{ for } Y_i \text{ between } 0 \text{ and } Z_i$$

$$\leq 1 + \frac{\lambda^2}{2}\mathbb{E}\left[Z_i^2 \mathbb{1}\{Z_i \leq 0\} + Z_i^2 e^{\lambda Z_i}\mathbb{1}\{Z_i > 0\}\right].$$

$$\bullet\ \mathbb{E} Z_i^2 \mathbb{I}\{Z_i \le 0\} \le \mathbb{E} Z_i^2 \le \mathbb{E} X_i^2 = \int_0^\infty \mathbb{P}[X_i^2 \ge u]\,du$$

$$= \int_0^\infty \mathbb{P}[|X_i| \ge t]\, 2t\,dt \le \int_0^\infty 4t e^{-(t/K)^\alpha}\,dt \le C(\alpha) K^2.$$

$$\bullet\ \mathbb{E} Z_i^2 e^{\lambda Z_i} \mathbb{I}\{Z_i > 0\} = \int_0^\infty \mathbb{P}[X_i^2 e^{\lambda X_i} \mathbb{I}\{0 < X_i \le L\} \ge u]\,du$$

$$\le \int_0^\infty \mathbb{P}[L \ge X_i \ge t](2t + \lambda t^2) e^{\lambda t}\,dt \qquad (u = t^2 e^{\lambda t})$$

$$\le \int_0^L 2 e^{-(t/K)^\alpha} (2t + \lambda t^2) e^{\lambda t}\,dt$$

$$\left(\text{take } \lambda \le \frac{1}{2L}(L/K)^\alpha \le \frac{1}{2t}(t/K)^\alpha \ \forall t \in (0, L)\right)$$

$$\le \int_0^L \left(4 + (t/K)^\alpha\right) t e^{-\frac{1}{2}(t/K)^\alpha}\,dt \le C(\alpha) K^2.$$

$$\Rightarrow \mathbb{E} e^{\lambda Z_i} \le 1 + C(\alpha) K^2 \lambda^2 \le e^{C(\alpha) K^2 \lambda^2}$$

$$\Rightarrow \mathbb{P}\left[\sum_{i=1}^n Z_i \ge t\right] \le e^{-\lambda t} \prod_{i=1}^n e^{\lambda Z_i} \le e^{-\lambda t + C(\alpha) K^2 n \lambda^2}$$

$\forall \lambda \in \left[0, \frac{1}{2L}(L/K)^\alpha\right]$. Similarly for $-X_i$.

So $\mathbb{P}\left[\left|\sum_{i=1}^n X_i\right| \ge t\right] \le 2 e^{-\lambda t + C(\alpha) K^2 n \lambda^2} + 2n e^{-(L/K)^\alpha}$

$\forall L > 0,\ \lambda \in \left[0, \frac{1}{2L}(L/K)^\alpha\right]$.

Pick $L = t$, $\lambda = \begin{cases} \dfrac{t}{2C(\alpha)K^2 n} & \text{if } \dfrac{t}{2C(\alpha)K^2 n} \le \dfrac{1}{2t}\left(t/K\right)^\alpha \\ \dfrac{1}{2t}\left(t/K\right)^\alpha & \text{otherwise} \end{cases}$

$$\Rightarrow -\lambda t + C(\alpha)K^2 n \lambda^2$$

$$\le \begin{cases} -\dfrac{t^2}{4C(\alpha)K^2 n} & \text{if } \dfrac{t}{2C(\alpha)K^2 n} \le \dfrac{1}{2t}\left(t/K\right)^\alpha \\ -\dfrac{1}{4}\left(t/K\right)^\alpha & \text{otherwise} \end{cases}$$

So $\mathbb{P}\left[\left|\sum_{i=1}^n X_i\right| \ge t\right] \le 2e^{-\frac{1}{4}\min\left(\frac{t^2}{C(\alpha)K^2 n}, \left(\frac{t}{K}\right)^\alpha\right)} + 2ne^{-\left(\frac{t}{K}\right)^\alpha}.$

If $\frac{3}{4}\left(\frac{t}{K}\right)^\alpha > \log 2n$: $\quad 2ne^{-\left(\frac{t}{K}\right)^\alpha} = e^{-\left(\frac{t}{K}\right)^\alpha + \log 2n} < e^{-\frac{1}{4}\left(\frac{t}{K}\right)^\alpha}.$

If $\frac{3}{4}\left(\frac{t}{K}\right)^\alpha \le \log 2n$:

Choose $C(\alpha)$ s.t. $e^{-\frac{1}{4}\frac{t^2}{C(\alpha)K^2 n}} \ge e^{-\frac{1}{4}\left(\frac{4}{3}\log 2n\right)^{2/\alpha} \cdot \frac{1}{C(\alpha)n}} \ge 1$

$$\Rightarrow \mathbb{P}\left[\left|\sum_{i=1}^n X_i\right| \ge t\right] \le 3e^{-\frac{1}{4}\min\left(\frac{t^2}{C(\alpha)K^2 n}, \left(\frac{t}{K}\right)^\alpha\right)}$$

$\blacksquare$