Stat 6390.001: High Dimensional Probability and Statistics, Fall 2026 **Topic 3**

Symmetrization and Sub-exponential Variables

Lecturer: Yun Wei *Scribe: Hanpei Xu*

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications. They may be distributed outside this class only with the permission of the Instructor.*

1 Symmetrization

Lemma 1.1. *Let $Y \perp\!\!\!\perp Z$ with $\mathbb{E}[Z] = 0$. Then for any convex function F,*

$$\mathbb{E}\left[F(Y)\right] \le \mathbb{E}\left[F(Y+Z)\right]. \tag{1}$$

Example 1.2. The function $F(x) = e^{\lambda x}$ is convex. Suppose $\mathbb{E}[X] = 0$. Applying Lemma 1.1 with $Y = X$ and $Z = -X'$, where $X' \perp\!\!\!\perp X$ and $X' \sim \mathrm{Law}(X)$, we obtain

$$\mathbb{E}_X\left[e^{\lambda X}\right] \le \mathbb{E}_{X,X'}\left[e^{\lambda(X-X')}\right]. \tag{2}$$

Proof of Lemma 1.1. For every $y \in \mathbb{R}$, using $\mathbb{E}[Z] = 0$ and Jensen's inequality,

$$F(y) = F\left(y + \mathbb{E}_Z[Z]\right) = F\left(\mathbb{E}_Z\left[y+Z\right]\right) \le \mathbb{E}_Z\left[F(y+Z)\right].$$

Substituting $y = Y$ gives $F(Y) \le \mathbb{E}_Z\left[F(Y+Z)\right]$, and taking expectations over Y yields

$$\mathbb{E}_Y\left[F(Y)\right] \le \mathbb{E}_Y\mathbb{E}_Z\left[F(Y+Z)\right] = \mathbb{E}_{Y,Z}\left[F(Y+Z)\right],$$

where the last equality uses $Y \perp\!\!\!\perp Z$. $\square$

Definition 1.3. A random variable X is called *symmetric* if it has the same distribution as $-X$.

Remark 1.4. If X has pdf $f(x)$, then X is symmetric if and only if $f(x) = f(-x)$.

Lemma 1.5 (Constructing symmetric distributions)**.** *Let X be a random variable and let ε be an independent Rademacher random variable.*

(a) εX and $\varepsilon\,|X|$ are identically distributed and symmetric.

(b) If X is symmetric, then both εX and $\varepsilon\,|X|$ have the same distribution as X.

(c) If X' is an independent copy of X, then $X - X'$ is symmetric.

Proof. We only prove that εX is symmetric. For every event $A \subseteq \mathbb{R}$, conditioning on the value of ε gives

$$\mathbb{P}\left[\varepsilon X \in A\right] = \tfrac{1}{2}\mathbb{P}\left[X \in A\right] + \tfrac{1}{2}\mathbb{P}\left[-X \in A\right] = \tfrac{1}{2}\mathbb{P}\left[-X \in A\right] + \tfrac{1}{2}\mathbb{P}\left[X \in A\right] = \mathbb{P}\left[-\varepsilon X \in A\right].$$ $\square$

Lemma 1.6 (Symmetrization)**.** *Let $X_1, \dots, X_N$ be independent, mean-zero random vectors in some normed space, and let $\varepsilon_1, \dots, \varepsilon_N$ be independent Rademacher random variables. Then*

$$\frac{1}{2}\,\mathbb{E}\left\|\sum_{i=1}^N \varepsilon_i X_i\right\| \;\le\; \mathbb{E}\left\|\sum_{i=1}^N X_i\right\| \;\le\; 2\,\mathbb{E}\left\|\sum_{i=1}^N \varepsilon_i X_i\right\|. \tag{3}$$

Proof. Let $(X_1', \dots, X_N')$ be independent copies of $(X_1, \dots, X_N)$. By Lemma 1.5(c) the vectors $X_i - X_i'$ are independent and symmetric (this step uses the assumption that X_i are independent), so that

$$(X_1 - X_1', \dots, X_N - X_N') \stackrel{d}{=} (\varepsilon_1(X_1 - X_1'), \dots, \varepsilon_N(X_N - X_N'))\,. \tag{4}$$

Upper bound. Set $p := \mathbb{E}\left\|\sum_{i=1}^N X_i\right\|$. Since $\mathbb{E}\left[\sum_{i=1}^N X_i'\right] = 0$, we may apply Lemma 1.1 with $F(x) = \|x\|$, $Y = \sum_{i=1}^N X_i$ and $Z = -\sum_{i=1}^N X_i'$ to get

$$p = \mathbb{E}\left\|\sum_{i=1}^N X_i - \mathbb{E}\left[\sum_{i=1}^N X_i'\right]\right\| \le \mathbb{E}\left\|\sum_{i=1}^N (X_i - X_i')\right\|.$$

Then by (4) and the triangle inequality,

$$\begin{aligned}\mathbb{E}\left\|\sum_{i=1}^N (X_i - X_i')\right\| &= \mathbb{E}\left\|\sum_{i=1}^N \varepsilon_i\,(X_i - X_i')\right\| = \mathbb{E}\left\|\sum_{i=1}^N \varepsilon_i X_i - \sum_{i=1}^N \varepsilon_i X_i'\right\| \\ &\le \mathbb{E}\left\|\sum_{i=1}^N \varepsilon_i X_i\right\| + \mathbb{E}\left\|\sum_{i=1}^N \varepsilon_i X_i'\right\| = 2\,\mathbb{E}\left\|\sum_{i=1}^N \varepsilon_i X_i\right\|.\end{aligned}$$

Lower bound. Conditionally on $\varepsilon = (\varepsilon_1, \dots, \varepsilon_N)$, the vectors $\sum_{i=1}^N \varepsilon_i X_i$ and $-\sum_{i=1}^N \varepsilon_i X_i'$ are independent and the latter has mean zero, so Lemma 1.1 applies conditionally and gives, after averaging over ε,

$$\mathbb{E}\left\|\sum_{i=1}^N \varepsilon_i X_i\right\| \le \mathbb{E}\left\|\sum_{i=1}^N \varepsilon_i X_i - \sum_{i=1}^N \varepsilon_i X_i'\right\| = \mathbb{E}\left\|\sum_{i=1}^N \varepsilon_i\,(X_i - X_i')\right\|.$$

Applying (4) once more and then the triangle inequality,

$$\mathbb{E}\left\|\sum_{i=1}^N \varepsilon_i\,(X_i - X_i')\right\| = \mathbb{E}\left\|\sum_{i=1}^N (X_i - X_i')\right\| \le \mathbb{E}\left\|\sum_{i=1}^N X_i\right\| + \mathbb{E}\left\|\sum_{i=1}^N X_i'\right\| = 2\,\mathbb{E}\left\|\sum_{i=1}^N X_i\right\|.$$ □

Remark 1.7. The zero-mean assumption is needed for both bounds: in the upper bound to subtract $\mathbb{E}\left[\sum_i X_i'\right]$, and in the lower bound to guarantee that $\mathbb{E}\left[\sum_i \varepsilon_i X_i' \mid \varepsilon\right] = 0$.

2 Sub-exponential Random Variables

Definition 2.1 ([1, Definition 2.7])**.** A random variable X with mean $\mu = \mathbb{E}[X]$ is *sub-exponential* if there are non-negative parameters (ν, α) such that

$$\mathbb{E}\left[e^{\lambda(X-\mu)}\right] \le e^{\frac{\nu^2\lambda^2}{2}} \qquad \text{for all } |\lambda| < \frac{1}{\alpha}. \tag{5}$$

Remark 2.2. A σ-sub-Gaussian variable is $(\sigma, 0)$-sub-exponential, where we interpret $1/0$ as $+\infty$.

Example 2.3 (Sub-exponential but not sub-Gaussian)**.** Let $Z \sim N(0,1)$ and consider $X = Z^2$, so that $\mathbb{E}[X] = 1$. For $\lambda < \frac{1}{2}$,

$$\begin{aligned}\mathbb{E}\left[e^{\lambda(X-1)}\right] &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{\lambda(z^2-1)} e^{-\frac{z^2}{2}}\,dz \\ &= \frac{e^{-\lambda}}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{-\frac{(1-2\lambda)z^2}{2}}\,dz = \frac{e^{-\lambda}}{\sqrt{1-2\lambda}},\end{aligned} \tag{6}$$

where the last equality uses $\int_{-\infty}^{+\infty} e^{-az^2/2}dz = \sqrt{2\pi/a}$ with $a = 1-2\lambda > 0$. For $\lambda \geq \frac{1}{2}$ the moment generating function is infinite, so X is not sub-Gaussian. On the other hand,

$$\frac{e^{-\lambda}}{\sqrt{1-2\lambda}} \leq e^{2\lambda^2} = e^{\frac{2^2\lambda^2}{2}} \qquad \text{where } |\lambda| < \tfrac{1}{4} \qquad \text{(check this!)}, \tag{7}$$

so that $\mathbb{E}\left[e^{\lambda(X-1)}\right] \leq e^{\frac{2^2\lambda^2}{2}}$ for all $|\lambda| < \frac{1}{4}$, i.e. X is $(2,4)$-sub-exponential.

Proposition 2.4 (Sub-exponential tail bound)**.** *Suppose X is sub-exponential with parameters (ν, α) and mean μ. Then*

$$\mathbb{P}[X - \mu \geq t] \leq \begin{cases} e^{-\frac{t^2}{2\nu^2}}, & \text{if } 0 \leq t \leq \frac{\nu^2}{\alpha}, \\ e^{-\frac{t}{2\alpha}}, & \text{for } t > \frac{\nu^2}{\alpha}, \end{cases} \quad = \quad e^{-\min\left\{\frac{t^2}{2\nu^2}, \frac{t}{2\alpha}\right\}}. \tag{8}$$

Proof. Assume $\mu = 0$ without loss of generality. By the Chernoff bound from Lecture 1 and Definition 2.1,

$$\begin{aligned} \log \mathbb{P}[X \geq t] &\leq \inf_{0 \leq \lambda < \frac{1}{\alpha}} \left\{\log \mathbb{E}\left[e^{\lambda X}\right] - \lambda t\right\} \leq \inf_{0 \leq \lambda < \frac{1}{\alpha}} \left\{\log e^{\frac{\nu^2\lambda^2}{2}} - \lambda t\right\} \\ &= \inf_{0 \leq \lambda < \frac{1}{\alpha}} \underbrace{\left\{\frac{\nu^2\lambda^2}{2} - \lambda t\right\}}_{=:\, g(\lambda)}. \end{aligned} \tag{9}$$

The function g is a convex parabola (see Figure 1) with

$$g'(\lambda) = \nu^2\lambda - t = 0 \qquad \Longleftrightarrow \qquad \lambda = \frac{t}{\nu^2}.$$

① If $\frac{t}{\nu^2} < \frac{1}{\alpha}$ (i.e. $t < \frac{\nu^2}{\alpha}$), the unconstrained minimizer lies in the feasible region, so

$$\inf_{0 \leq \lambda < \frac{1}{\alpha}} g(\lambda) = g\left(\frac{t}{\nu^2}\right) = \frac{\nu^2}{2} \cdot \frac{t^2}{\nu^4} - \frac{t}{\nu^2} \cdot t = -\frac{t^2}{2\nu^2}.$$

② If $\frac{t}{\nu^2} \geq \frac{1}{\alpha}$ (i.e. $t \geq \frac{\nu^2}{\alpha}$), then g is decreasing on $[0, \frac{t}{\nu^2})$ and the infimum is approached at $\lambda = \frac{1}{\alpha}$, giving

$$\inf_{0 \leq \lambda < \frac{1}{\alpha}} g(\lambda) = g\left(\frac{1}{\alpha}\right) = \frac{\nu^2}{2} \cdot \frac{1}{\alpha^2} - \frac{1}{\alpha}t = \frac{\nu^2}{2\alpha^2} - \frac{t}{\alpha} \leq -\frac{t}{2\alpha},$$

where the last step uses $\nu^2 < \alpha t$. □

3 Bernstein's Condition

We say that X satisfies *Bernstein's condition with parameter b* if

$$\left|\mathbb{E}\left[(X-\mu)^k\right]\right| \leq \frac{1}{2} k!\, \mathrm{Var}(X)\, b^{k-2}, \qquad \text{for any } k = 2, 3, \ldots, \tag{10}$$

where $\mu = \mathbb{E}[X]$.

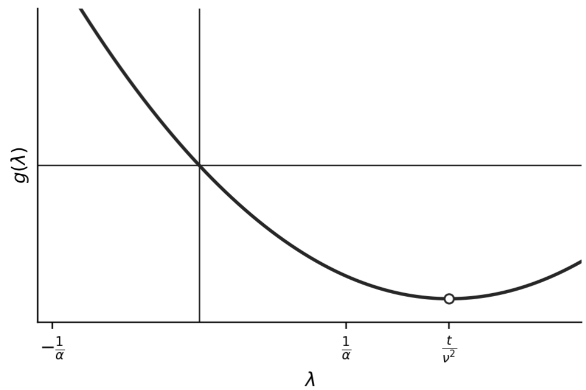

Figure 1: The function $g(\lambda) = \frac{\nu^2\lambda^2}{2} - \lambda t$ appearing in (9), in the case $\frac{t}{\nu^2} > \frac{1}{\alpha}$. The open circle marks the unconstrained minimizer $\frac{t}{\nu^2}$, which lies outside the feasible region $|\lambda| < \frac{1}{\alpha}$.

Example 3.1. If (10) is satisfied, then X is sub-exponential. Indeed, writing $\sigma^2 = \mathrm{Var}(X)$ and using the Taylor expansion of the exponential together with $\mathbb{E}[X-\mu] = 0$,

$$\begin{aligned}
\mathbb{E}\left[e^{\lambda(X-\mu)}\right] &= \mathbb{E}\left[\sum_{k=0}^{\infty} \frac{(\lambda(X-\mu))^k}{k!}\right] = 1 + 0 + \mathbb{E}\left[\sum_{k=2}^{\infty} \frac{\lambda^k (X-\mu)^k}{k!}\right] \\
&= 1 + \sum_{k=2}^{\infty} \frac{\lambda^k\, \mathbb{E}\left[(X-\mu)^k\right]}{k!} = 1 + \frac{\lambda^2\sigma^2}{2} + \sum_{k=3}^{\infty} \frac{\lambda^k\, \mathbb{E}\left[(X-\mu)^k\right]}{k!} \\
&\leq 1 + \frac{\lambda^2\sigma^2}{2} + \sum_{k=3}^{\infty} \frac{|\lambda|^k \cdot \frac{1}{2} k!\, \sigma^2 b^{k-2}}{k!} = 1 + \frac{\lambda^2\sigma^2}{2} + \frac{\lambda^2\sigma^2}{2} \sum_{k=3}^{\infty} (|\lambda|\, b)^{k-2} = 1 + \frac{\lambda^2\sigma^2}{2} \sum_{s=0}^{\infty} (|\lambda|\, b)^s .
\end{aligned} \tag{11}$$

For $|\lambda| < \frac{1}{b}$ the geometric series sums to $\frac{1}{1-|\lambda| b}$, so that

$$\mathbb{E}\left[e^{\lambda(X-\mu)}\right] \leq 1 + \frac{\lambda^2\sigma^2/2}{1 - b\,|\lambda|} \leq e^{\frac{\lambda^2\sigma^2/2}{1-b|\lambda|}}, \tag{12}$$

using $1 + t \leq e^t$. Restricting further to $|\lambda| < \frac{1}{2b}$ gives $1 - b\,|\lambda| > \frac{1}{2}$, and therefore

$$\mathbb{E}\left[e^{\lambda(X-\mu)}\right] \leq e^{\lambda^2\sigma^2} = e^{\frac{\lambda^2\left(\sqrt{2}\sigma\right)^2}{2}}, \qquad \text{for all } |\lambda| < \frac{1}{2b},$$

i.e. X is sub-exponential with parameters $(\nu, \alpha) = (\sqrt{2}\sigma, 2b)$.

Exercises

1. Verify the inequality (7), that is, show that $\frac{e^{-\lambda}}{\sqrt{1-2\lambda}} \leq e^{2\lambda^2}$ for all $|\lambda| < \frac{1}{4}$.

References

[1] M. J. Wainwright, *High-Dimensional Statistics: A Non-Asymptotic Viewpoint*, Cambridge University Press, 2019.

[2] R. Vershynin, *High-Dimensional Probability: An Introduction with Applications in Data Science*, Cambridge University Press, 2018.