

Lecture 10 — October 8, 2015

Prof. Jelani Nelson

Today we will prove the distributional JL lemma from last lecture. First we collect some notation and basic lemmas we will use.

Throughout, for a random variable X , $\|X\|_p$ denotes $(\mathbb{E}|X|^p)^{1/p}$. It is known that $\|\cdot\|_p$ is a norm for any $p \geq 1$ (Minkowski's inequality). It is also known $\|X\|_p \leq \|X\|_q$ whenever $p \leq q$. Henceforth, whenever we discuss $\|\cdot\|_p$, we will assume $p \geq 1$.

We often use Jensen's inequality below, especially for $F(x) = |x|^p$ ($p \geq 1$).

Lemma 1 (Jensen's inequality). *For F convex, $F(\mathbb{E} X) \leq \mathbb{E} F(X)$.*

Lemma 2. *For $1 \leq p < q < \infty$, $\|X\|_p \leq \|X\|_q$.*

Proof. Define $f(x) = |x|^{q/p}$. Then f is convex. Thus by Jensen's inequality,

$$(\mathbb{E}|X|^p)^{q/p} \leq \mathbb{E}|X|^q.$$

Now raise both sides of the inequality to the $1/p$. □

Definition 3. *The gaussian distribution $\mathcal{N}(0, \sigma^2)$ has density function $f(x) = (2\pi\sigma^2)^{-1/2} e^{-x^2/(2\sigma^2)}$.*

Fact 4. *If $g \sim \mathcal{N}(0, \sigma^2)$ then $\mathbb{E} g^p$ for integer p is 0 for p odd and is $\sigma^p(p-1) \cdot (p-3) \cdots 1 < (\sigma\sqrt{p})^p$ for p even.*

Lemma 5 (Khintchine inequality). *For any $p \geq 1$, $x \in \mathbb{R}^n$, and (σ_i) independent Rademachers,*

$$\left\| \sum_i \sigma_i x_i \right\|_p \lesssim \sqrt{p} \cdot \|x\|_2$$

Proof. Without loss of generality we can assume p is an even integer. (If not, let q be the smallest even integer larger than p then it suffices to have $\|\sum_i \sigma_i x_i\|_q \lesssim \sqrt{p} \|x\|_2$ since $\|\cdot\|_p \leq \|\cdot\|_q$ by Lemma 2.) Consider (g_i) independent gaussians of mean zero and variance 1. Expand $\mathbb{E}(\sum_i \sigma_i x_i)^p$ into a sum of monomials.

$$\begin{aligned} \mathbb{E}(\sum_i \sigma_i x_i)^p &= \sum_{t=1}^{\min\{p,n\}} \sum_{i_1 < i_2 < \dots < i_t} \sum_{\substack{d_1, \dots, d_t \geq 1 \\ d_1 + \dots + d_t = p}} \binom{p}{d_1, \dots, d_t} \left(\prod_{j=1}^t x_{i_j}^{d_j} \right) \left(\mathbb{E}_\sigma \prod_{j=1}^t \sigma_{i_j}^{d_j} \right) \\ &= \sum_{t=1}^{\min\{p,n\}} \sum_{i_1 < i_2 < \dots < i_t} \sum_{\substack{d_1, \dots, d_t \geq 1 \\ d_1 + \dots + d_t = p}} \binom{p}{d_1, \dots, d_t} \left(\prod_{j=1}^t x_{i_j}^{d_j} \right) \left(\prod_{j=1}^t \mathbb{E} \sigma_{i_j}^{d_j} \right) \end{aligned}$$

Any monomial with odd exponents (i.e. odd d_j) vanishes, as in the gaussian case. Meanwhile, monomials with all d_j being even have $\prod_{j=1}^t x_{i_j}^{d_j}$ nonnegative and $\prod_{j=1}^t \mathbb{E} \sigma_{i_j}^{d_j} = 1$. Meanwhile if

the σ_{i_j} are replaced by gaussians g_{i_j} , then $\prod_{j=1}^t \mathbb{E}_{\sigma_{i_j}} g_{i_j}^{d_j} \geq 1$. Thus the Rademacher p th moment is term-by-term dominated by the gaussian case and thus $\|\sum_i \sigma_i x_i\|_p \leq \|\sum_i g_i x_i\|_p$. But $\sum_i g_i x_i$ is a gaussian with mean zero and variance $\|x\|_2^2$, and we apply Fact 4.

The point of the above argument was to show that the Rademacher case is bounded by the gaussian case, after which point we concluded. Another way to show this is as follows. Now we will not require p to be an even integer.

$$\begin{aligned} \left\| \sum_i \sigma_i x_i \right\|_p &= \sqrt{\frac{\pi}{2}} \cdot \left\| \mathbb{E}_g \sum_i \sigma_i |g_i| x_i \right\|_p \text{ (since } \mathbb{E} |g| = \sqrt{2/\pi}) \\ &\leq \sqrt{\frac{\pi}{2}} \cdot \left\| \sum_i \sigma_i |g_i| x_i \right\|_p \text{ (Jensen)} \\ &= \sqrt{\frac{\pi}{2}} \cdot \left\| \sum_i g_i x_i \right\|_p \text{ (}\sigma_i |g_i| \text{ is distributed as } g_i\text{)} \end{aligned}$$

□

We now prove a decoupling inequality which will be useful for our proof of Hanson-Wright, which we will use to prove distributional JL. We use $\|\cdot\|_{L^p(X)}$ to denote $(\mathbb{E}_X |\cdot|^p)^{1/p}$ when we want to make it clear which random variable we are taking the expectation over.

Lemma 6 (Decoupling [dlPnG99]). *Let x_1, \dots, x_n be independent and mean zero, and x'_1, \dots, x'_n identically distributed as the x_i and independent of them. Then for any $(a_{i,j})$ and for all $p \geq 1$*

$$\left\| \sum_{i \neq j} a_{i,j} x_i x_j \right\|_p \leq 4 \left\| \sum_{i,j} a_{i,j} x_i x'_j \right\|_p$$

Proof. Let η_1, \dots, η_n be independent Bernoulli random variables each of expectation $1/2$. Then

$$\begin{aligned} \left\| \sum_{i \neq j} a_{i,j} x_i x_j \right\|_{L^p(x)} &= 4 \cdot \left\| \mathbb{E}_\eta \sum_{i \neq j} a_{i,j} x_i x_j |\eta_i| |1 - \eta_j| \right\|_{L^p(x)} \\ &\leq 4 \cdot \left\| \sum_{i \neq j} a_{i,j} x_i x_j \eta_i (1 - \eta_j) \right\|_{L^p(x, \eta)} \text{ (Jensen)} \end{aligned} \tag{1}$$

Hence there must be some fixed vector $\eta' \in \{0, 1\}^n$ which achieves

$$\left\| \sum_{i \neq j} a_{i,j} x_i x_j \eta_i (1 - \eta_j) \right\|_{L^p(x, \eta)} \leq \left\| \sum_{i \in S} \sum_{j \notin S} a_{i,j} x_i x_j \right\|_{L^p(\eta)}$$

where $S = \{i : \eta'_i = 1\}$. Let x_S denote the $|S|$ -dimensional vector corresponding to the x_i for $i \in S$. Then

$$\begin{aligned} \left\| \sum_{i \in S} \sum_{j \notin S} a_{i,j} x_i x_j \right\|_{L^p(x)} &= \left\| \sum_{i \in S} \sum_{j \notin S} a_{i,j} x_i x'_j \right\|_{L^p(x_S, x'_S)} \\ &= \left\| \mathbb{E}_{x_S} \mathbb{E}_{x'_S} \sum_{i,j} a_{i,j} x_i x'_j \right\|_{L^p(x_S, x'_S)} \text{ (}\mathbb{E} x_i = \mathbb{E} x'_j = 0\text{)} \\ &\leq \left\| \sum_{i,j} a_{i,j} x_i x'_j \right\|_{L^p(x, x')} \text{ (Jensen)} \end{aligned}$$

□

The following proof of the Hanson-Wright was shared to me by Sjoerd Dirksen (personal communication). See also a recent proof in [RV13].

Recall that by problem set 1, problem 1, the statement of the Hanson-Wright inequality below is equivalent to the statement that there exists a constant $C > 0$ such that for all $\lambda > 0$

$$\mathbb{P}_{\sigma}(|\sigma^T A \sigma - \mathbb{E} \sigma^T A \sigma| > \lambda) \lesssim e^{-C\lambda^2/\|A\|_F^2} + e^{-C\lambda/\|A\|}. \quad (2)$$

Theorem 7 (Hanson-Wright inequality [HW71]). *For $\sigma_1, \dots, \sigma_n$ independent Rademachers and $A \in \mathbb{R}^{n \times n}$ real and symmetric, for all $p \geq 1$*

$$\|\sigma^T A \sigma - \mathbb{E} \sigma^T A \sigma\|_p \lesssim \sqrt{p} \cdot \|A\|_F + p \cdot \|A\|.$$

Proof. Without loss of generality we assume in this proof that $p \geq 2$ (so that $p/2 \geq 1$). Then

$$\|\sigma^T A \sigma - \mathbb{E} \sigma^T A \sigma\|_p \lesssim \|\sigma^T A \sigma'\|_p \text{ (decoupling)} \quad (3)$$

$$\lesssim \sqrt{p} \cdot \|\|Ax\|_2\|_p \text{ (Khintchine)} \quad (4)$$

$$= \sqrt{p} \cdot \|\|Ax\|_2^2\|_{p/2}^{1/2} \quad (5)$$

$$\leq \sqrt{p} \cdot \|\|Ax\|_2^2\|_p^{1/2}$$

$$\leq \sqrt{p} \cdot (\|A\|_F^2 + \|\|Ax\|_2^2 - \mathbb{E} \|Ax\|_2^2\|_p)^{1/2} \text{ (triangle inequality)}$$

$$\leq \sqrt{p} \cdot \|A\|_F + \sqrt{p} \cdot \|\|Ax\|_2^2 - \mathbb{E} \|Ax\|_2^2\|_p^{1/2}$$

$$\lesssim \sqrt{p} \cdot \|A\|_F + \sqrt{p} \cdot \|x^T A^T A x'\|_p^{1/2} \text{ (decoupling)}$$

$$\lesssim \sqrt{p} \cdot \|A\|_F + p^{3/4} \cdot \|\|A^T A x\|_2\|_p^{1/2} \text{ (Khintchine)}$$

$$\lesssim \sqrt{p} \cdot \|A\|_F + p^{3/4} \cdot \|A\|^{1/2} \cdot \|\|Ax\|_2\|_p^{1/2} \quad (6)$$

Writing $E = \|\|Ax\|_2\|_p^{1/2}$ and comparing (4) and (6), we see that for some constant $C > 0$,

$$E^2 - Cp^{1/4}\|A\|^{1/2}E - C\|A\|_F \leq 0.$$

Thus E must be smaller than the larger root of the above quadratic equation, implying our desired upper bound on E^2 . \square

Remark 1. The “square root trick” in the proof of the Hanson-Wright inequality above is quite handy and can be used to prove several moment inequalities (for example, you will see how to prove the Bernstein inequality with it in tomorrow’s lecture). As far as I am aware, the trick was first used in a work of Rudelson [Rud99].

Remark 2. We could have upper bounded Eq. (5) by

$$\sqrt{p} \cdot \|A\|_F + \sqrt{p} \cdot \|\|Ax\|_2^2 - \mathbb{E} \|Ax\|_2^2\|_{p/2}^{1/2}$$

by the triangle inequality. Now notice we have bounded the p th central moment of a symmetric quadratic form (3) by the $p/2$ th moment also of a symmetric quadratic form. Writing $p = 2^k$, this observation leads to a proof by induction on k , which was the approach used in [DKN10].

Distributional Johnson-Lindenstrauss (DJL) lemma Now we finally prove the *Distributional JL Lemma (DJL)* stated in last lecture (which implies the JL lemma itself).

Lemma 8. DJL Lemma For any integer $n > 1$ and $\varepsilon, \delta \in (0, 1/2)$, there exists a distribution $\mathcal{D}_{\varepsilon, \delta}$ over $\mathbb{R}^{m \times n}$ for $m \lesssim \varepsilon^{-2} \log(1/\delta)$ such that for any $x \in \mathbb{R}^n$ of unit Euclidean norm,

$$\mathbb{P}_{\Pi \sim \mathcal{D}_{\varepsilon, \delta}} (|\|\Pi x\|_2^2 - 1| > \varepsilon) < \delta$$

Proof. Write $\Pi_{i,j} = \sigma_{i,j}/\sqrt{m}$, where the $\sigma_{i,j}$ are independent Rademachers. Also overload σ to mean these Rademachers arranged as a vector of length mn , by concatenating rows of Π . Note

$$\Pi x = A_x \sigma, \text{ implying } \|\Pi x\|_2^2 = \|A_x \sigma\|_2^2$$

where

$$A_x = \frac{1}{\sqrt{m}} \cdot \begin{bmatrix} -x^T - & 0 & \cdots & 0 \\ 0 & -x^T - & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & -x^T - \end{bmatrix}. \quad (7)$$

Thus

$$\mathbb{P}(|\|\Pi x\|_2^2 - 1| > \varepsilon) = \mathbb{P}(|\|A_x \sigma\|_2^2 - \mathbb{E} \|A_x \sigma\|_2^2| > \varepsilon),$$

where we see that the right-hand side is readily handled by the Hanson-Wright inequality with $A = A_x^T A_x$ (using (2)). Observe A is a block-diagonal matrix with each block equaling $(1/m)xx^T$, and thus $\|A\| = \|x\|_2^2/m = 1/m$. We also have $\|A\|_F^2 = 1/m$. Thus Hanson-Wright yields

$$\mathbb{P}(|\|\Pi x\|_2^2 - 1| > \varepsilon) \lesssim e^{-C\varepsilon^2 m} + e^{-C\varepsilon m},$$

which for $\varepsilon < 1$ is at most δ for $m \gtrsim \varepsilon^{-2} \log(1/\delta)$. \square

References

- [DKN10] Ilias Diakonikolas, Daniel M. Kane, and Jelani Nelson. Bounded independence fools degree-2 threshold functions. In *51th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 11–20, 2010.
- [dlPnG99] Victor de la Peña and Evarist Giné. *Decoupling: From dependence to independence*. Probability and its Applications. Springer-Verlag, New York, 1999.
- [HW71] David Lee Hanson and Farroll Tim Wright. A bound on tail probabilities for quadratic forms in independent random variables. *Ann. Math. Statist.*, 42:1079–1083, 1971.
- [Rud99] Mark Rudelson. Random vectors in the isotropic position. *J. Functional Analysis*, 164(1):60–72, 1999.
- [RV13] Mark Rudelson and Roman Vershynin. Hanson-Wright inequality and sub-gaussian concentration. *Electronic Communications in Probability*, 18:1–9, 2013.