Lecture 8

Last time

- McDiarmid's cont.
- Lipschitz functions of Gaussians.

Today

- Missing Claim
- Concentration of the norm

Last time we used the following claim.

Claim (⊙̲⊙): We have that for convex $\psi : \mathbb{R} \to \mathbb{R}$,

$$\mathbb{E}\left[\psi\left(f(X) - \mathbb{E}f(X)\right)\right] \leq \mathbb{E}\, \psi\left(\frac{\pi}{2}\langle \nabla f(X), Y\rangle\right),$$

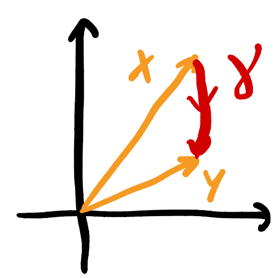
where X, Y are iid $N(0,1)$. $\dashv$

Proof of Claim (⊙̲⊙): We can introduce Y

$$\mathbb{E}_X\, \psi\left(f(X) - \mathbb{E}_Y f(Y)\right) \underset{\text{Jensen's}}{\leq} \mathbb{E}_{XY}\, \psi\left(\left(f(X) - f(Y)\right)\right)$$

For each $\theta \in [0, \pi/2]$, let

$$\gamma(\theta) := X\cos\theta + Y\sin\theta$$
$$\dot{\gamma}(\theta) = -X\sin\theta + Y\cos\theta$$



By the fundamental Theorem of Calculus

$$f(X) - f(Y) = \int_0^{\pi/2} (f\circ\gamma)'(\theta)\, d\theta$$

$$\overset{\text{Chain rule}}{=} \int_0^{\pi/2} \langle \nabla f(\gamma(\theta)), \dot{\gamma}(\theta)\rangle\, d\theta$$

Therefore,

$$\Psi\left(f(Y)-f(X)\right) = \Psi\Bigg(\underbrace{\frac{2}{\pi}\int_0^{\pi/2} \frac{\pi}{2}\langle \nabla f(\gamma(\theta)), \dot\gamma(\theta)\rangle \, d\theta}_{\text{Uniform expectation on } \theta}\Bigg)$$

$$\overset{\text{Jensen's}}{\leq} \frac{2}{\pi}\int_0^{\pi} \Psi\left(\frac{\pi}{2}\langle \nabla f(\gamma(\theta)), \dot\gamma(\theta)\rangle\right) d\theta.$$

Hence,

$$\mathbb{E}\,\Psi(f(Y)-f(X)) \leq \frac{2}{\pi}\int_0^{1} \mathbb{E}\,\Psi\left(\frac{\pi}{2}\langle \nabla f(\gamma(\theta)), \dot\gamma(\theta)\rangle\right) d\theta.$$

It turns out that the integrant is independent of θ.

Fact (4): Let Z be a random vector with $Z \sim N(0, I)$. Then, for any $\mathcal{Q}$ a matrix s.t. $\mathcal{Q}\mathcal{Q}^T = \mathcal{Q}^T\mathcal{Q} = I$, we have

$$\mathcal{Q} Z \sim N(0, I).$$

Notice that $Z = (\gamma(0), \dot\gamma(0)) = (X, Y)^\top$ and $(\gamma(\theta), \dot\gamma(\theta)) = \mathcal{Q}_\theta Z$ with $\mathcal{Q}_\theta$ a rotation matrix. Thus,

$$\mathbb{E}\,\psi(f(X) - \mathbb{E} f(X)) \leq \mathbb{E}\,\Psi\left(\frac{\pi}{2}\langle \nabla f(X), Y\rangle\right). \quad \square$$

Example (Order Statistics): Suppose we are given a sample $X_1, \dots, X_n$. Its order statistics are given by reordering

$$X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}.$$

In HW 1 we studied the expected value of $X_{(n)} = \max_i X_i$. Further, we have.

Fact (HW 2): For any $X, Y \in \mathbb{R}^n$,

$$|X_{(k)} - Y_{(k)}| \leq \|X - Y\|_2 \quad \forall k \in [n].$$

Thus, if $X_1, \dots, X_n$ are iid $N(0,1)$, we obtain that

$$\mathbb{P}(|X_{(k)} - \mathbb{E} X_{(k)}| \geq t) \leq 2e^{-t^2/2}.$$

Concentration of the norm.

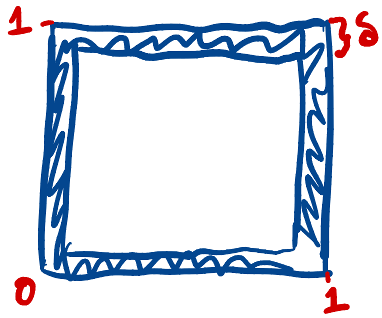
Random vectors in high dimensions are very different from what you would expect.

For instance, consider $X \sim \text{Unif}([0,1]^d)$.

How much mass do we have in a thin shell of the hypercube?

Pick $\delta \in (0,1)$, then the shell is

$$[0,1]^d \setminus [\delta, 1-\delta]^d$$



The probability of this set is equal to

$$p_d = 1 - (1-2\delta)^d$$

At $d=1$, then $p_d = \delta$. But as $d \to \infty$, we have $p_d \to 1$.

Something similar happens with many high-dimensional quantities of random vectors.

Theorem: Suppose X is a random vector ($X \in \mathbb{R}^d$) with iid entries with X_i σ_i^2 sub-Gaussian, $\mathbb{E}X_i = 0$, and $\mathbb{E} X_i = 1$. Then,

$$\mathbb{P}\left(\left| \|X\|^2 - d \right| \geq td\right) \leq 2\exp\left(\frac{cd}{\sigma^2}(t \wedge t^2)\right),$$

$$\mathbb{P}\left(\left| \|X\| - \sqrt{d} \right| \geq t\sqrt{d}\right) \leq 2\exp\left(\frac{-cdt^2}{\sigma^2}\right).$$

Universal const.

Proof: First notice that

$$\frac{1}{d}\mathbb{E}\|X\|_2^2 = \frac{1}{d}\sum_{j=1}^{d}\mathbb{E}X_j^2 = 1$$

Furthermore we had this lemma from lecture 6:

Lemma: Suppose Y, Z are sub-Gaussian, then

$$\|YZ\|_{\psi_1} \leq \|Y\|_{\psi_2}\|Z\|_{\psi_2}. \qquad \dashv$$

Therefore

Different constants

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

Invoking Bernstein's ineq

$$\mathbb{P}\left(\left|\frac{1}{d}\|X\|_2^2 - 1\right| \geq t\right) = \mathbb{P}\left(\left|\frac{1}{d}\sum(X_i^2 - 1)\right| \geq t\right)$$

(☆)

$$\leq 2\exp\left(-c\left(\frac{t^2 d}{\sigma^4} \wedge \frac{td}{\sigma^2}\right)\right)$$

$1 \leq \sigma$

$$\leq 2\exp\left(-\frac{c}{\sigma^2}\left(t^2 \wedge t\right)\right)$$

This proves the same bound, we will now use it to prove the second one. Notice that

$$|z-1| > \delta \Rightarrow |z^2-1| = |z-1|\,|z+1| \geq \delta \cdot |z+1|$$

Further $|z+1| \geq 1$ and $|z+1| \geq |z-1| \geq \delta$.

$$|z-1| > \delta \Rightarrow |z^2-1| \geq \max\{\delta, \delta^2\}$$

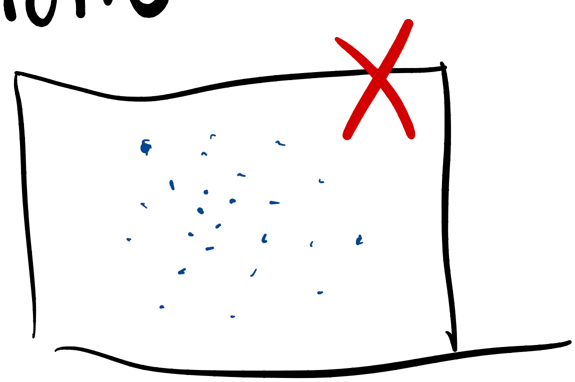
Therefore,

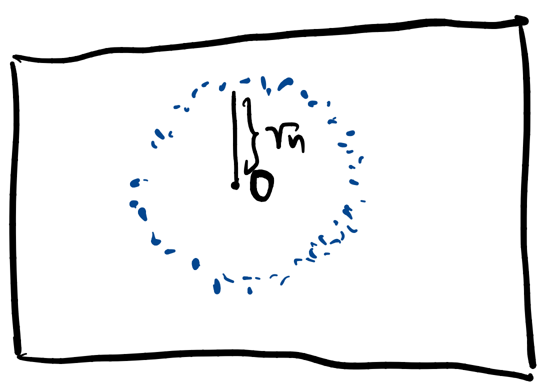
$$\mathbb{P}\left(\left|\tfrac{1}{\sqrt{d}}\|X\|_2 - 1\right| \geq \delta\right) \leq \mathbb{P}\left(\left|\tfrac{1}{d}\|X\|_2^2 - 1\right| \geq \delta \vee \delta^2\right)$$

$(\star)$ with $t = \delta \vee \delta^2$ $\leq 2\exp\left(\frac{cd}{\sigma^2}\delta^2\right)$.

$\square$

This means that in high dimensions





The norm of $\|X\| \sim \sqrt{d}$ with constant size deviations.