Lecture 6

Last time

- Johnson-Lindenstrauss Lemma.
- Orlicz norms

Today

- Orlicz norms cont.
- McDiarmid's Ineq.

Orlicz norms

We finished with:

Proposition: Let X be a r.v. the following are equivalent (modulo const. factors):

1) $\exists K_1 > 0$ s.t. $\mathbb{P}(|X| > t) \leq 2e^{-t/K_1}$ $\forall t \geq 0$.
2) $\exists K_2 > 0$ s.t. $\|X\|_{\mathcal{L}_p} := (\mathbb{E}|X|^p)^{1/p} \leq K_2 \, p$ $\forall p \geq 1$.
3) $\exists K_3 > 0$ s.t. $\mathbb{E}\exp(|X|/K_3) \leq 2$.

Moreover, if $\mathbb{E}X = 0$ then, these are quivalent to

4) $\exists K_4 > 0$ s.t. $\mathbb{E}\exp(\lambda X) \leq \exp(K_4^2 \lambda^2)$ $\forall |\lambda| \leq \frac{1}{K_4}$.

$\dashv$

This motivates the following.

Def: The subexponential norm of a r.v. is

$$\|X\|_{\psi_1} := \inf\{K > 0 : \mathbb{E}\exp(|X|/K) \leq 2\}.$$

$\dashv$

Just as before $\|\cdot\|_{\psi_1}$ is a norm over the set of subexponentials. Moreover

$$\|X - \mathbb{E}X\|_{\psi_1} \leq C\|X\|_{\psi_1}.$$

We motivated subexponential via χ^2 distributions, in turn products of sub-Gaussians are always subexponential.

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

$$\|XY\|_{\psi_1} \leq \|X\|_{\psi_2}\|Y\|_{\psi_2}.$$

Proof: WLOG $\|X\|_{\psi_2} = \|Y\|_{\psi_2} = 1$. Then

$$\mathbb{E}\exp(|XY|) \leq \mathbb{E}\exp(X^2/2 + Y^2/2)$$

Young's ineq. $|ab| \leq a^2/2 + b^2/2$

$$= \mathbb{E}\exp(X^2/2)\exp(Y^2/2)$$

$$\leq \frac{1}{2}\left(\mathbb{E}\exp(X^2) + \mathbb{E}\exp(Y^2)\right)$$

$$\leq \frac{1}{2}(2+2) = 2.$$

□

It is natural to wonder whether other functions besides exponentials define other norms capturing different growth/tails. Indeed, this is the case

Def: Given a convex, nondecreasing function $\psi : \mathbb{R}_+ \to \mathbb{R}_+$ s.t. $\psi(0) = 0$ with $\psi(t) \xrightarrow{t \to \infty} \infty$, define the Orlicz norm of a r.v. X as

$$\|X\|_\psi = \inf \{ K > 0 \mid \mathbb{E}\, \psi(|X|/K) \le 1 \}.$$

One can show that this defines a norm on $\{ X \mid \|X\|_\psi < \infty \}$.

Example: For $\psi(t) = |t|^p$ with $p \ge 1$ defines $\mathcal{L}_p$. While $\psi_2(t) = e^{t^2} - 1$ and $\psi_1(t) = e^t - 1$ define sub-Gaussians and sub exponentials, respectively.

Concentration of functions of iid r.v.

So far we have studied concentration of sums. However, this is a more general phenomenon. The following priciple it's good to have in mind

If $X_1, \dots, X_n$ are independent r.v. then $f(X_1, \dots, X_n)$ concentrates near $\mathbb{E} f(X_1, \ldots, X_n)$ provided f is not to sensitive to any coordinate.

We will instantiate this principles for two notions of "sensitivity."

Our goal is to prove the following.

Theorem (McDiarmid): Let $X_1, \ldots, X_n$ be ind. r.vs and $f : \mathbb{R}^n \to \mathbb{R}$ be a function s.t. $\forall j \in [n]\ \exists c_j > 0$ with

$$|f(z_1, \ldots, z_j, \ldots, z_n) - f(z_1, \ldots, \tilde{z}_j, \ldots, z_n)| \le c_j$$

$\forall z_1, z_2, \ldots, z_n, \tilde{z}_j \in \mathbb{R}$. Then,

$$\mathbb{P}(|f(X) - \mathbb{E} f(X)| > t) \le 2e^{-2t^2/\|c\|_2^2}.$$

To prove this result we will use the so-called Martingale method, which is useful beyond this proof.

Def (Martingale): We say that a sequence of r.v. $Y_0, Y_1, \ldots$ is a r.v. with respect to another sequence of r.v.

$X_0, X_1, \ldots$ if $\forall n$ we have

- $\mathbb{E}|Y_n| < \infty$ $Y_n = f(X_0, \ldots, X_n)$
- Y_n is measurable w.r.t. $X_0, \ldots, X_n$)
- $\mathbb{E}(Y_{n+1} \mid X_0, \ldots, X_n) = Y_n$ $\dashv$

Martingales model fair games. They are helpful to derive results when full independence fails (CLTs, concentration). They are covered in Prob. Theory II. We will need to remember a few facts.

Fact (Tower law): If $j < k$

$$\mathbb{E}\big[\mathbb{E}[Y \mid X_1, \ldots, X_k] \mid X_1, \ldots, X_j\big] = \mathbb{E}[Y \mid X_1, \ldots, X_j].$$

Fact: If Y is measurable w.r.t. $X_1, \ldots, X_k$,

$$\mathbb{E}[Y \mid X_1, \ldots, X_k] = Y.$$

The idea for the proof is to consider $Y_0 = \mathbb{E} f(X)$ and $Y_j = \mathbb{E}[f(X) \mid X_1, \ldots, X_j]$ $\forall j$.

Then, we can decompose

$$f(X) - \mathbb{E} f(X) = Y_n - Y_0 = \sum_{j=0}^{n-1} (Y_{j+1} - Y_j). \quad (\star)$$

It is not hard to see that $\{Y_k\}$

is a Martingale with respect to $\{X_k\}$:

$$\mathbb{E}[Y_{j+1} \mid X_1, \dots, X_j] = \mathbb{E}[\mathbb{E}[f(X) \mid X_1, \dots, X_{j+1}] \mid X_1, \dots, X_j]$$

$$\overset{\text{Tower law}}{=} \mathbb{E}[f(X) \mid X_1, \dots, X_j] \qquad (\heartsuit)$$

$$= Y_j.$$

Thus, in order to control the difference $|f(X) - \mathbb{E}f(X)|$ it suffices to control sums of Martingale differences.

Lemma (Azuma): Suppose that $\{Y_k\}$ is a Martingale w.r.t. $\{X_k\}$ and set $\Delta_k = Y_k - Y_{k-1}$. Further, assume $\forall k$

$$\mathbb{E}[e^{\lambda \Delta_{k+1}} \mid X_1, \dots, X_k] \le e^{\lambda^2 \sigma_k^2 / 2} \quad \text{a.s.} \; (\ddot{o})$$

Then, the sum $\sum_{k=1}^{n} \Delta_k$ is $\|\sigma\|_2^2$-sub-Gaussian. $\dagger$

We will come back to the proof of this result next time.