Lecture 7

Last time

- Orlicz norms cont.
- McDiarmid's Ineq.

Today

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

McDiarmid's Inequality continued

Last time we finished with

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 $\forall k$

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

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

Proof: For any $k \in [n]$, we bound

$$\mathbb{E}\left[e^{\lambda \sum_{k=1}^n \Delta_k}\right] \overset{\text{Tower law}}{=} \mathbb{E}\left[\mathbb{E}\left[e^{\lambda \sum_{k=1}^n \Delta_k} \mid X_1, \dots, X_{n-1}\right]\right]$$

$$= \mathbb{E}\left[e^{\lambda \sum_{k=1}^{n-1} \Delta_k} \mathbb{E}\left[e^{\lambda \Delta_n} \mid X_1, \dots, X_{n-1}\right]\right]$$

$$\le e^{\lambda^2 \sigma_n^2/2}\, \mathbb{E}\left[e^{\lambda \sum_{k=1}^{n-1} \Delta_k}\right]$$

Repeat $\leq e^{\lambda \sum_{k=1}^{n} \sigma_k^2 / 2}$. $\square$

Note that we didn't use the fact that $\mathbb{E}[\Delta_k \mid X_1, \dots, X_{k-1}] = 0$ (Martingale property explicitly, but we cannot get (:o) without it; see HW 1.

Proof of McDiarmid's:

Note that we have that $\{Y_k\}$ is a Martingale w.r.t. $\{X_k\}$ thanks to

Lecture 6 → (♡). To apply Azuma's we need to show (:o). Recall

$$
\begin{aligned}
Y_k - Y_{k-1} &= \mathbb{E}[f(X_1, \dots, X_n) \mid X_1, \dots, X_k] - \mathbb{E}[f(X_1, \dots, X_n) \mid X_1, \dots, X_{k-1}] \\
&\geq \underbrace{\mathbb{E}[\inf_t f(x_1, \dots, X_{k-1}, t, X_{k+1}, \dots, X_n) - f(X_1, \dots, X_n) \mid X_1, \dots, X_{k-1}]}_{A_k}
\end{aligned}
$$

Similarly

$$
\begin{aligned}
Y_k - Y_k &\leq \underbrace{\mathbb{E}[\sup_t f(x_1, \dots, X_{k-1}, t, X_{k+1}, \dots, X_n) - f(X_1, \dots, X_n) \mid X_1, \dots, X_{k-1}]}_{B_k}
\end{aligned}
$$

Thus, conditioned on $X_1, \dots, X_{k-1}$ Δ_k lands on $[A_k, B_k]$ and moreover thanks to the bounded differences assumption

$$B_k - A_k \le \mathbb{E}\left[\sup_t f(X_1, \dots, t, X_n) - \inf_t f(X_1, \dots, t, \dots, X_n) \mid X_1, \dots, X_{k-1}\right]$$
$$\le c_k$$

Thus, (using the same argument as in lect. 2)

$$\mathbb{E}\left[e^{\lambda \Delta_k} \mid X_1, \dots, X_{k-1}\right] \le e^{\lambda^2 c_k^2 / 8}.$$

By Azuma's we have that $\sum_{k=1}^{n} \Delta_k$ is $\|c\|^2/4$-sub-Gaussian and so

$$\mathbb{P}\left(|f(X) - \mathbb{E} f(X)| \ge t \right) \le 2 \exp\left(-2t^2 / \|c\|_2^2\right).$$

$\square$

McDiarmid's inequality is specially useful when we lack independence. Let's see an example.

Example (U-statistics) Suppose we want to estimate $\mathbb{E}\, g(X,Y)$ where X, Y are iid rvs and $g : \mathbb{R}^2 \to \mathbb{R}$ bounded by C with access to a sample $X_1, \dots, X_n$. A natural approach is

$$U(X) = \frac{1}{\binom{n}{2}} \sum_{i<j} g(X_i, X_j).$$

Notice that the elements in the sum are not independent. However, they are only weakly-dependent.
We can bound

$$|U(x_1, \dots, x_n) - U(\underbrace{x_1, \dots, \hat{x}_j, \dots x_n}_{\tilde{x} :=})|$$

$$= \frac{1}{\binom{n}{2}} \left| \sum_{i<j} g(x_i, x_j) - g(\tilde{x}_i, \tilde{x}_j) \right|$$

$$\leq \frac{1}{\binom{n}{2}} \sum_{i<j} \left| g(x_i, x_j) - g(\tilde{x}_i, \tilde{x}_j) \right|$$

$$\leq \frac{1}{\binom{n}{2}} (n-1) 2C = \frac{4C}{n}.$$

Thus the function U satisfies the bounded differences property and McDiarmid's inequality yields

$$\mathbb{P}(|U - \mathbb{E}U| \geq t) \leq 2\exp(-nt^2/(8c^2)). \dashv$$

Lipschitz functions of Gaussians.

Next we see another instantiation of the principle from the previous lecture.

Theorem: Let $X_1, \dots, X_n$ be iid rvs with $X_1 \sim N(0,1)$. Let $f: \mathbb{R}^n \to \mathbb{R}$ be an L-Lipschitz function, i.e.,

$$|f(x) - f(y)| \leq L \|x-y\|_2 \quad \forall x, y \in \mathbb{R}^n.$$

Then,

$$\mathbb{P}(|F(X) - \mathbb{E}F(X)| \geq t) \leq 2e^{-\frac{t^2}{2L^2}}. \dashv$$

Proof: We will prove a weaker version of this result with

$$\mathbb{P}(|F(X) - \mathbb{E}F(X)| \geq t) \leq 2e^{-\frac{2}{\pi^2}\frac{t^2}{L^2}}.$$

For the best constant see the proof

in Vershynin's (it uses deep results that we will not cover).

WLOG assume F is C^1-smooth
Radamacher's Theorem

Claim (☺): We have that for convex $\psi: \mathbb{R} \to \mathbb{R}$,
$$\mathbb{E}[\psi(f(X) - \mathbb{E}f(X))] \le \mathbb{E}\,\psi\left(\frac{\pi}{2}\langle \nabla f(X), Y\rangle\right),$$
where X, Y are iid $N(0,1)$. ⊣

Before proving this claim, let us show how it implies the result. Notice that it suffices to show that $f(X) - \mathbb{E}f(X)$ is $\left(\frac{\pi L}{2}\right)^2$-sub-Gaussian. Applying the Claim with $t \mapsto e^{\lambda t}$

$$\mathbb{E}[\exp(\lambda(f(X) - \mathbb{E}f(X)))] \le \mathbb{E}_{X,Y} \exp\left(\lambda \frac{\pi}{2}\langle \nabla f(X), Y\rangle\right)$$

For fixed X, $\langle \nabla f(X), Y\rangle$ is a r.v. with dist. $N(0, \|\nabla f(X)\|^2)$

$$= \mathbb{E}_X \exp\left(\frac{\lambda^2 \pi^2}{8}\|\nabla f(X)\|^2\right)$$

For Lipschitz f, $\|\nabla f(X)\| \le L$

$$\le \exp\left(\frac{\lambda^2\pi^2}{8} L^2\right)$$

Therefore, $f(X) - \mathbb{E}f(X)$ is sub-Gaussian with $\sigma^2 = \pi^2 L^2/4$ as we wanted.

Next we establish Claim (☺).