McDiarmid's inequality

COMS 4995-1 Spring 2020 (Daniel Hsu)

Theorem (McDiarmid's inequality). Let $X_1, \dots, X_n$ be independent random variables, where X_i has range $\mathcal{X}_i$. Let $f \colon \mathcal{X}_1 \times \cdots \times \mathcal{X}_n \to \mathbb{R}$ be any function with the $(c_1, \dots, c_n)$-*bounded differences property*: for every $i = 1, \dots, n$ and every $(x_1, \dots, x_n), (x'_1, \dots, x'_n) \in \mathcal{X}_1 \times \cdots \times \mathcal{X}_n$ that differ only in the i-th corodinate ($x_j = x'_j$ for all $j \neq i$), we have

$$|f(x_1, \dots, x_n) - f(x'_1, \dots, x'_n)| \leq c_i.$$

For any $t > 0$,

$$\Pr(f(X_1, \dots, X_n) - \mathbb{E}[f(X_1, \dots, X_n)] \geq t) \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^n c_i^2}\right).$$

Proof. Write $\mathbb{E}_i[\cdot]$ to denote expectation conditioned on $X_1^i := (X_1, \dots, X_i)$. Therefore, $g_i(X_1^i) := \mathbb{E}_i[f(X_1^n)]$ is, as the notation suggests, a function of X_1^i. Now define the following random variables:

$$\begin{aligned}
Y_i &:= g_i(X_1^i) - g_{i-1}(X_1^{i-1}), \\
A_i &:= \inf_{x_i \in \mathcal{X}_i} g_i(X_1^{i-1}, x_i) - g_{i-1}(X_1^{i-1}), \\
B_i &:= \sup_{x_i \in \mathcal{X}_i} g_i(X_1^{i-1}, x_i) - g_{i-1}(X_1^{i-1}).
\end{aligned}$$

These random variables satisfy

$$\begin{aligned}
&Y_i \in [A_i, B_i], \\
&\mathbb{E}_{i-1}[Y_i] = 0, \\
&\sum_{i=1}^n Y_i = f(X_1^n) - \mathbb{E}[f(X_1^n)].
\end{aligned}$$

Furthermore, we claim that $[A_i, B_i]$ is always an interval of length at most c_i. We defer the proof of this claim to later.

To bound the probability that the sum of the Y_i's is at least t, we use the Chernoff bounding method. Let $S_i := Y_1 + \cdots + Y_i$. Then

$$\begin{aligned}
\mathbb{E}[\exp(\lambda S_n)] &= \mathbb{E}[\exp(\lambda(Y_n + S_{n-1}))] \\
&= \mathbb{E}[\mathbb{E}_{n-1}[\exp(\lambda Y_n)] \exp(\lambda S_{n-1})] \\
&\leq \exp(\lambda^2 c_n^2/8)\mathbb{E}[\exp(\lambda S_{n-1})] \\
&= \exp(\lambda^2 c_n^2/8)\mathbb{E}[\exp(\lambda(Y_{n-1} + S_{n-2}))] \\
&= \exp(\lambda^2 c_n^2/8)\mathbb{E}[\mathbb{E}_{n-2}[\exp(\lambda Y_{n-1})] \exp(\lambda S_{n-2})] \\
&\leq \exp(\lambda^2 c_n^2/8) \exp(\lambda^2 c_{n-1}^2/8)\mathbb{E}[\exp(\lambda S_{n-2})] \\
&\ \ \vdots \\
&\leq \exp(\lambda^2 c_n^2/8) \cdots \exp(\lambda^2 c_1^2/8) \\
&= \exp\left(\frac{\sum_{i=1}^n c_i^2}{4} \cdot \frac{\lambda^2}{2}\right).
\end{aligned}$$

Above, the inequalities all follow from Hoeffding's inequality. Therefore, S_n is $\sum_{i=1}^n c_i^2/4$-subgaussian. The probability bound now follows from the Cramer-Chernoff inequality for subgaussian random variables.

It remains to prove that $[A_i, B_i]$ is an interval of length at most c_i. We show that $B_i - A_i \leq c_i$:

$$\begin{aligned}
B_i - A_i &= \sup_{x_i' \in \mathcal{X}_i} g_i(X_1^{i-1}, x_i') - \inf_{x_i \in \mathcal{X}_i} g_i(X_1^{i-1}, x_i) \\
&= \sup_{x_i, x_i' \in \mathcal{X}_i} g_i(X_1^{i-1}, x_i') - g_i(X_1^{i-1}, x_i) \\
&= \sup_{x_i, x_i' \in \mathcal{X}_i} \mathbb{E}_i[f(X_1^{i-1}, x_i', X_{i+1}^n) - f(X_1^{i-1}, x_i, X_{i+1}^n)] \\
&\leq \mathbb{E}_i\left[\sup_{x_i, x_i' \in \mathcal{X}_i} |f(X_1^{i-1}, x_i', X_{i+1}^n) - f(X_1^{i-1}, x_i, X_{i+1}^n)|\right] \\
&\leq c_i.
\end{aligned}$$

The last step above uses the bounded differences property of f. ∎