LECTURE 4: MCDIARMID'S INEQUALITY & APPLICATIONS

This lecture introduces McDiarmid's Inequality, a powerful concentration inequality that bounds the deviation of a function of many independent random variables from its expected value. Concentration inequalities provide guarantees on how much a random variable can deviate from its expectation. McDiarmid's inequality is particularly useful because it applies to general functions, as long as they satisfy a stability condition known as the bounded difference property. It is a generalization of Hoeffding's inequality. Before delving into McDiarmid's Inequality, let us define Conditional Expectation.

Conditional Expectation. Let $\mathbb{E}_i[\cdot]$ denote the expectation conditioned on $X_{1:i}$. For any measurable function $h : \mathcal{X}^n \to \mathbb{R}$, we have

$$\mathbb{E}_i[h(X_{1:n})] = \mathbb{E}[h(X_{1:n}) \mid X_{1:i}] = \int_{\mathcal{X}_{i+1:n}} h(x_{1:n})\, p(x_{i+1:n} \mid x_{1:i})\, dx_{i+1:n},$$

where $p(x_{i+1:n} \mid x_{1:i})$ is the conditional density of $X_{i+1:n}$ given $X_{1:i}$.

Property (Law of Total Expectation). For any random variables X and Y, we have

$$\mathbb{E}_{X,Y}[h(X,Y)] \;=\; \mathbb{E}_X[\,\mathbb{E}_Y[h(X,Y) \mid X]\,]\,.$$

Proof. By the definition of conditional expectation

$$\begin{aligned}
\mathbb{E}_Y[h(X,Y)|X] &= \int_{y\in\mathcal{Y}} h(X,Y=y)p(Y=y|X)dy \\
&= \int_{y\in\mathcal{Y}} h(X,Y=y)\frac{p(X,Y=y)}{p(X)}dy
\end{aligned}$$

Thus,

$$\begin{aligned}
\mathbb{E}_X\left[\mathbb{E}_Y[h(X,Y)|X]\right] &= \int_{x\in\mathcal{X}}\int_{y\in\mathcal{Y}} h(X=x,Y=y)\frac{p(X=x,Y=y)}{p(X=x)}p(X=x)dydx \\
&= \int_{x\in\mathcal{X},y\in\mathcal{Y}} h(X=x,Y=y)p(X=x,Y=y)dydx \\
&= \mathbb{E}_{X,Y}[h(X,Y)]\,.
\end{aligned}$$

□

Now let's state the MdDiarmid's Inequality.

Theorem 1 (McDiarmid's Inequality)**.** *Let $X_1, \ldots, X_n$ be independent random variables, with $X_i \in \mathcal{X}_i$. Let $f : \mathcal{X}_1 \times \cdots \times \mathcal{X}_n \to \mathbb{R}$ be a function such that for every $i = 1, \ldots, n$, it satisfies the bounded difference property:*

$$\sup_{x_1,\ldots,x_n,x_i'} |f(x_1,\ldots,x_n) - f(x_1,\ldots,x_{i-1},x_i',x_{i+1},\ldots,x_n)| \le c_i$$

Then for any $\epsilon > 0$,

$$\mathbb{P}(f(X_1,\ldots,X_n) - \mathbb{E}[f(X_1,\ldots,X_n)] \ge \epsilon) \le \exp\left(-\frac{2\epsilon^2}{\sum_{i=1}^n c_i^2}\right)$$

The core idea is that if changing one of the input variables X_i does not change the value of the function f by too much (at most by c_i), then the function's value will be concentrated around its mean.

Proof. The proof proceeds by constructing a martingale difference sequence and applying Hoeffding's Lemma. This technique, often called the method of bounded differences, breaks down the total deviation of the function into a sum of smaller, more manageable terms.

Step 1: Defining the Martingale Sequence

Let $g_i(x_1,\ldots,x_i) = \mathbb{E}[f(X_1,\ldots,X_n)|X_1 = x_1,\ldots,X_i = x_i]$. This is a function of the first i variables that represents the expected value of f after revealing the first i random variables. The sequence $g_0, g_1(X_1), g_2(X_{1:2}),\ldots,g_n(X_{1:n})$ forms a special type of stochastic process known as a Doob martingale. Each term in the sequence represents our best estimate of the final value of f given the information we have seen so far.

Step 2: Introducing Dummy Variables

We can think of the sequence of conditional expectations in terms of a set of dummy and independent random variables X_0 and X_{n+1} bracketing the original sequence $X_{1:n}$. This helps define the boundary conditions for our sequence g_i. This allows us to treat the first and last steps of the process ($i = 0$ and $i = n$) uniformly with the intermediate steps.

- **Without dummy variables:** The function g_i is defined as:

$$g_i(x_{1:i}) = \int_{\mathcal{X}_{i+1:n}} f(x_{1:n})p(x_{i+1:n}|x_{1:i})dx_{i+1:n} = \int_{\mathcal{X}_{i+1:n}} f(x_{1:n})p(x_{i+1:n})dx_{i+1:n}$$

- **After introducing dummy variables:**

$$g_i(x_{1:i}) = \int_{\mathcal{X}_{i+1:n+1}} f(x_{1:n})p(x_{i+1:n+1}|x_{0:i})dx_{i+1:n+1} = \int_{\mathcal{X}_{i+1:n}} f(x_{1:n})p(x_{i+1:n})dx_{i+1:n}$$

Observe that both definitions of $g_i(x_{1:i})$ are equivalent as X_0 and X_{n+1} are assumed to be independent.

Step 3: Fixing $X_i = x_i$

For a fixed set of values $x_1,\ldots,x_i$, the function g_i is the expectation over the remaining random variables $X_{i+1},\ldots,X_n$:

$$\begin{aligned}
g_i(X_{1:i-1}, x_i) &= \int_{\mathcal{X}_{i+1:n}} f(x_{1:i-1}, x_i, x_{i+1:n})p(x_{i+1:n})dx_{i+1:n} \\
&= \int_{\mathcal{X}_{i:n}} f(X_{1:i-1}, x_i, X_{i+1:n})p(X_{i:n})dx_{i:n} \\
&= \mathbb{E}_{X_{i:n}}[f(X_{1:i-1}, x_i, X_{i+1:n})] \\
&= \mathbb{E}_{i-1}[f(X_{1:i-1}, x_i, X_{i+1:n})]
\end{aligned}$$

Since the variables $X_1,\ldots,X_n$ are independent, $p(x_{i+1:n}|x_{1:i}) = p(x_{i+1:n})$.

Step 4: Defining the Difference Sequence

We now define a sequence of random variables, called a martingale difference sequence, for $i = 1, \ldots, n$:

$$Y_i = g_i(X_{1:i}) - g_{i-1}(X_{1:i-1})$$

This represents the change in expectation upon revealing X_i. Each Y_i represents the new information gained by revealing the random variable X_i. For a fixed realization of $X_{1:i-1}$, we also define the bounds for Y_i:

$$A_i = \inf_{x_i \in \mathcal{X}_i} g_i(X_{1:i-1}, x_i) - g_{i-1}(X_{1:i-1})$$
$$B_i = \sup_{x_i \in \mathcal{X}_i} g_i(X_{1:i-1}, x_i) - g_{i-1}(X_{1:i-1})$$

Properties of the Difference Sequence

We now establish several key properties of this sequence $\{Y_i\}$ which are essential for completing the proof.

(i) $Y_i \in [A_i, B_i]$.

Proof. For any realization of $X_{1:i}$, the value of Y_i is $g_i(X_{1:i}) - g_{i-1}(X_{1:i-1})$. By definition, A_i and B_i are the infimum and supremum of this quantity over all possible values of X_i, holding $X_{1:i-1}$ fixed. Thus, any realized value of Y_i must lie within these bounds. □

(ii) $\mathbb{E}[Y_i] = 0$.

Proof. Recall that

$$\mathbb{E}[Y_i] \;=\; \mathbb{E}_{i-1}[g_i(X_{1:i})] - \mathbb{E}_{i-1}[g_{i-1}(X_{1:i-1})].$$

Thus, it suffices to show that

$$\mathbb{E}[g_i(X_{1:i}) \mid X_{1:i-1}] = g_{i-1}(X_{1:i-1}),$$

since clearly $\mathbb{E}[g_{i-1}(X_{1:i-1}) \mid X_{1:i-1}] = g_{i-1}(X_{1:i-1})$.

Now, for fixed $x_{1:i-1}$, we compute

$$\begin{aligned}
\mathbb{E}[g_i(X_{1:i}) \mid X_{1:i-1} = x_{1:i-1}] &= \int_{\mathcal{X}_i} g_i(x_{1:i-1}, x_i)\, p(x_i)\, dx_i \\
&= \int_{\mathcal{X}_i} \left(\int_{\mathcal{X}_{i+1:n}} f(x_{1:n})\, p(x_{i+1:n})\, dx_{i+1:n} \right) p(x_i)\, dx_i \\
&= \int_{\mathcal{X}_{i:n}} f(x_{1:n})\, p(x_{i:n})\, dx_{i:n} \\
&= \mathbb{E}[f(X_{1:n}) \mid X_{1:i-1} = x_{1:i-1}] \\
&= g_{i-1}(x_{1:i-1}).
\end{aligned}$$

Therefore, $\mathbb{E}[Y_i | X_{1:i-1}] = \mathbb{E}[g_i(X_{1:i}) | X_{1:i-1}] - \mathbb{E}[g_{i-1}(X_{1:i-1}) | X_{1:i-1}] = g_{i-1}(X_{1:i-1}) - g_{i-1}(X_{1:i-1}) = 0$. This property is the formal definition of a martingale difference sequence. It means that, on average, each new piece of information does not systematically shift our expectation up or down. □

(iii) Y_i only depends on $X_1, \ldots, X_i$.

Proof. By definition, g_i is a function of $X_{1:i}$ and g_{i-1} is a function of $X_{1:i-1}$. Therefore, their difference Y_i can only depend on variables up to and including index i. □

(iv) $\sum_{i=1}^{n} Y_i = f(X_{1:n}) - \mathbb{E}[f(X_{1:n})]$.

Proof. By definition,

$$\sum_{i=1}^{n} Y_i = \sum_{i=1}^{n} \big(g_i - g_{i-1}\big) = \sum_{i=1}^{n} g_i(X_{1:i}) - \sum_{i=1}^{n} g_{i-1}(X_{1:i-1}) = g_n(X_{1:n}) - g_0(X_{1:0})$$

Now note that $g_n(X_{1:n}) = \mathbb{E}[f(X_{1:n}) \mid \mathcal{F}_n]$. Since $f(X_{1:n})$ is $\mathcal{F}_n$-measurable, it follows that

$$g_n(X_{1:n}) = f(X_{1:n}).$$

Similarly, $g_0(X_{1:0}) = \mathbb{E}[f(X_{1:n}) \mid \mathcal{F}_0]$. Because $\mathcal{F}_0$ is the trivial σ-algebra, we obtain

$$g_0(X_{1:0}) = \mathbb{E}[f(X_{1:n})].$$

Therefore,

$$\sum_{i=1}^{n} Y_i \;=\; g_n(X_{1:n}) - g_0(X_{1:0}) \;=\; f(X_{1:n}) - \mathbb{E}[f(X_{1:n})].$$

□

(v) $B_i - A_i \leq c_i$.

Proof. From the definitions of A_i and B_i:

$$\begin{aligned}
B_i - A_i &= \sup_{x_i} g_i(X_{1:i-1}, x_i) - \inf_{x_i'} g_i(X_{1:i-1}, x_i') \\
&= \sup_{x_i, x_i'} \big(g_i(X_{1:i-1}, x_i) - g_i(X_{1:i-1}, x_i')\big) \\
&= \sup_{x_i, x_i'} \mathbb{E}[f(\ldots, x_i, \ldots) - f(\ldots, x_i', \ldots) | X_{1:i-1}] \\
&\leq \mathbb{E}[\sup_{x_i, x_i'} |f(\ldots, x_i, \ldots) - f(\ldots, x_i', \ldots)| | X_{1:i-1}] \leq \mathbb{E}[c_i | X_{1:i-1}] = c_i.
\end{aligned}$$

□

Main Proof Continuation

Let $S = f(X_{1:n}) - \mathbb{E}[f(X_{1:n})] = \sum_{i=1}^{n} Y_i$. For any $t > 0$, we use the bounding technique starting with Markov's inequality:

$$\mathbb{P}(S \geq \epsilon) = \mathbb{P}(e^{tS} \geq e^{t\epsilon}) \leq e^{-t\epsilon}\mathbb{E}[e^{tS}] = e^{-t\epsilon}\mathbb{E}\left[\exp\left(t\sum_{i=1}^{n} Y_i\right)\right]$$

Our goal is to bound the moment-generating function $\mathbb{E}[e^{tS}]$. We can do this by iteratively applying the law of total expectation as defined in Property-1. We start by conditioning on $X_{1:n-1}$:

$$\begin{aligned}\mathbb{E}[\exp(t\sum_{i=1}^{n} Y_i)] &= \mathbb{E}[\mathbb{E}[\exp(t\sum_{i=1}^{n} Y_i)|X_{1:n-1}]] \\ &= \mathbb{E}[\exp(t\sum_{i=1}^{n-1} Y_i)\mathbb{E}[e^{tY_n}|X_{1:n-1}]]\end{aligned}$$

The inner expectation, $\mathbb{E}[e^{tY_n}|X_{1:n-1}]$, can now be bounded. This is the crucial step where the properties of the martingale difference sequence is used. Since we proved that $\mathbb{E}[Y_n|X_{1:n-1}] = 0$ and Y_n is bounded in an interval of length $B_n - A_n \leq c_n$, Hoeffding's Lemma gives $\mathbb{E}[e^{tY_n}|X_{1:n-1}] \leq \exp(\frac{t^2 c_n^2}{8})$. Substituting this back, we get:

$$\mathbb{E}[\exp(t\sum_{i=1}^{n} Y_i)] \leq \mathbb{E}[\exp(t\sum_{i=1}^{n-1} Y_i)]\exp(\frac{t^2 c_n^2}{8})$$

We can apply this argument iteratively, conditioning on $X_{1:n-2}$, then $X_{1:n-3}$, and so on. This process, starting from Y_n and working backwards to Y_1, effectively gives off one term at a time, applying the Hoeffding bound at each step. This gives the final bound on the function:

$$\mathbb{E}[\exp(t\sum_{i=1}^{n} Y_i)] \leq \exp\left(\frac{t^2\sum_{i=1}^{n} c_i^2}{8}\right)$$

Plugging this into the Markov inequality expression gives:

$$\mathbb{P}(S \geq \epsilon) \leq \exp(-t\epsilon + \frac{t^2 \sum c_i^2}{8})$$

To get the tightest bound, we minimize the right-hand side with respect to t. The minimum occurs at $t = \frac{4\epsilon}{\sum c_i^2}$, which gives:

$$\mathbb{P}(S \geq \epsilon) \leq \exp\left(-\frac{4\epsilon^2}{\sum c_i^2} + \frac{16\epsilon^2 \sum c_i^2}{8(\sum c_i^2)^2}\right) = \exp\left(-\frac{2\epsilon^2}{\sum_{i=1}^{n} c_i^2}\right)$$

This concludes the proof.

Application Taking $f(X_1, \ldots, X_n) = \frac{1}{n}\sum_{i=1}^{n} X_i$, where each $X_i \in [a_i, b_i]$ immediately gives us Hoeffding's inequality. Note that in this case $c_i = \frac{b_i - a_i}{n}$.