The Bounded Difference Inequality

1 Introduction

The goal of this lecture is to introduce and prove the bounded difference inequality (BDI). This is a concentration inequality that generalizes Hoeffding’s and that has found many uses in learning theory. Our first use of it will be in the development of Rademacher complexity. The BDI was first proved by McDiarmid [1], and its proof leverages techniques that had been previously developed by Hoeffding [2] and Azuma [3].

2 The Bounded Difference Inequality

Definition 1. Let $A$ be some set and $\phi : A^n \to \mathbb{R}$. We say $\phi$ satisfies the bounded difference assumption if there exist $c_1, \ldots, c_n \geq 0$ such that

$$\sup_{x_1, \ldots, x_n, x'_i \in A} |\phi(x_1, \ldots, x_i, \ldots, x_n) - \phi(x_1, \ldots, x'_i, \ldots, x_n)| \leq c_i.$$

That is, if we substitute $x_i$ to $x'_i$, while keeping other $x_j$ fixed, $\phi$ changes by at most $c_i$.

Theorem 1. Let $X_1, \ldots, X_n$ be arbitrary independent random variables on set $A$ and $\phi : A^n \to \mathbb{R}$ satisfy the bounded difference assumption. Then for all $t > 0$

$$\Pr\{\phi(X_1, \ldots, X_n) - \mathbb{E}[\phi(X_1, \ldots, X_n)] \geq t\} \leq e^{-\frac{2t^2}{\sum_{i=1}^{n} c_i^2}}$$

and

$$\Pr\{\phi(X_1, \ldots, X_n) - \mathbb{E}[\phi(X_1, \ldots, X_n)] \leq -t\} \leq e^{-\frac{2t^2}{\sum_{i=1}^{n} c_i^2}}.$$

Remark. By combining the above two inequalities, we obtain:

$$\Pr\{|\phi(X_1, \ldots, X_n) - \mathbb{E}[\phi(X_1, \ldots, X_n)]| \geq t\} \leq 2e^{-\frac{2t^2}{\sum_{i=1}^{n} c_i^2}}.$$

Remark. The bounded difference inequality recovers Hoeffding’s inequality [2]. Assume $X_i \in [a_i, b_i]$ and take $\phi(X_1, \ldots, X_n) = \sum_{i=1}^{n} X_i$. Then $c_i = b_i - a_i$. Plugging everything into the BDI gives

$$\Pr\{|\sum_{i=1}^{n} X_i - \mathbb{E}[\sum_{i=1}^{n} X_i]| \geq t\} \leq 2e^{-\frac{2t^2}{\sum_{i=1}^{n} (b_i - a_i)^2}}.$$
We omit the proof, which is similar to the proof of the lemma used for proving Hoeffding’s inequality. There $Z$ was independent of $V$, $\psi(Z) = a_i$ and $c = b_i - a_i$.

**Proof of bounded difference inequality.** Denote

$$V = \phi(X_1, \ldots, X_n) - \mathbb{E}[\phi(X_1, \ldots, X_n)]$$

and

$$V_i = \mathbb{E}[\phi(X_1, \ldots, X_n)|X_1, \ldots, X_i] - \mathbb{E}[\phi(X_1, \ldots, X_n)|X_1, \ldots, X_{i-1}].$$

Observe that $V = \sum_{i=1}^{n} V_i$ (telescoping series).

**Note.** ($V_i$) is an example of what is called a **martingale difference sequence**.

**Claim:** Each $V_i$ satisfies Lemma 1 with $c = c_i$ and $Z = (X_1, \ldots, X_{i-1})$.

Let’s first assume the above claim holds; we’ll prove it later. Applying Chernoff’s bounding technique, $\forall s > 0$ we have

$$\Pr\{\phi(X_1, \ldots, X_n) - \mathbb{E}[\phi(X_1, \ldots, X_n)] \geq t\} \leq e^{-st} \mathbb{E}[e^{s\sum_{i=1}^{n} V_i}]$$

(Markov’s inequality)

$$= e^{-st} \mathbb{E}[e^{s\sum_{i=1}^{n-1} V_i + sV_n}]$$

(by only $V_n$ depends on $X_n$)

$$= e^{-st} \mathbb{E}_{X_1, \ldots, X_{n-1}} \mathbb{E}_{X_n|X_1, \ldots, X_{n-1}} [e^{s\sum_{i=1}^{n-1} V_i + sV_n}|X_1, \ldots, X_{n-1}]$$

where the last step uses the lemma. If we repeatedly apply the above inequality to $V_{n-1}, \ldots, V_1$, we obtain:

$$\Pr\{\phi(X_1, \ldots, X_n) - \mathbb{E}[\phi(X_1, \ldots, X_n)] \geq t\} \leq e^{-st} \frac{e^{s\sum_{i=1}^{n} c_i}}{\sum_{i=1}^{n} c_i}.$$

Now take $s = \frac{4t}{\sum_{i=1}^{n} c_i}$ to minimize the upper bound, giving

$$\Pr\{\phi(X_1, \ldots, X_n) - \mathbb{E}[\phi(X_1, \ldots, X_n)] \geq t\} \leq e^{-\frac{2t^2}{\sum_{i=1}^{n} c_i^2}}.$$

Apply the above inequality to $-\phi$ to get the other inequality.

Now let’s establish the claim. First observe $\mathbb{E}[V_i|X_1, \ldots, X_{i-1}] = 0 \forall i, i \leq n$. This follows immediately from the so-called “tower property” or “smoothing property” of conditional expectation [4], which implies

$$\mathbb{E}[\phi(X_1, \ldots, X_n)|X_1, \ldots, X_i]\mathbb{E}[X_1, \ldots, X_{i-1}] = \mathbb{E}[\phi(X_1, \ldots, X_n)|X_1, \ldots, X_{i-1}].$$

If this is unclear, consider as an example

$$V_3 = \mathbb{E}[\phi(X_1, \ldots, X_n)|X_1, X_2, X_3] - \mathbb{E}[\phi(X_1, \ldots, X_n)|X_1, X_2].$$
Now
\[
E[E[\phi(X_1, \ldots, X_n)|X_1, X_2, X_3]|X_1, X_2] \\
= E_{X_3|X_1, X_2} \int \phi(X_1, X_2, X_3, x_4, \ldots, x_n) dP(x_4, \ldots, x_n|X_1, X_2, X_3) \\
= \int \int \phi(X_1, X_2, x_3, \ldots, x_n) dP(x_3, \ldots, x_n|X_1, X_2) \frac{dP(x_4, \ldots, x_n|X_1, X_2, X_3)}{dP(x_3, \ldots, x_n|X_1, X_2)} \\
= \int \int \phi(X_1, X_2, x_3, \ldots, x_n) dP(x_3, \ldots, x_n|X_1, X_2) \\
= E[\phi(X_1, \ldots, X_n)|X_1, X_2].
\]
Therefore $E[V_3|X_1, X_2] = 0$.

To show the second part of the claim, define
\[
L_i = \inf_x E[\phi(X_1, \ldots, X_n)|X_1, \ldots, X_{i-1}, x] - E[\phi(X_1, \ldots, X_n)|X_1, \ldots, X_{i-1}]
\]
and
\[
U_i = \sup_x E[\phi(X_1, \ldots, X_n)|X_1, \ldots, X_{i-1}, x'] - E[\phi(X_1, \ldots, X_n)|X_1, \ldots, X_{i-1}].
\]
Clearly, $L_i \leq V_i \leq U_i$. Furthermore,
\[
U_i - L_i = \sup_x E[\phi(X_1, \ldots, X_n)|X_1, \ldots, X_{i-1}, x'] - \inf_x E[\phi(X_1, \ldots, X_n)|X_1, \ldots, X_{i-1}, x] \\
= \sup_{x'} \int \phi(X_1, \ldots, X_{i-1}, x', x_{i+1}, \ldots, x_n) dP(x_{i+1}, \ldots, x_n|X_1, \ldots, X_{i-1}, x') \\
- \inf_x \int \phi(X_1, \ldots, X_{i-1}, x, x_{i+1}, \ldots, x_n) dP(x_{i+1}, \ldots, x_n|X_1, \ldots, X_{i-1}, x) \\
= \sup_{x'} \int \phi(X_1, \ldots, X_{i-1}, x', x_{i+1}, \ldots, x_n) dP(x_{i+1}, \ldots, x_n) \\
- \inf_x \int \phi(X_1, \ldots, X_{i-1}, x, x_{i+1}, \ldots, x_n) dP(x_{i+1}, \ldots, x_n) \\
= \sup_{x,x'} [\phi(X_1, \ldots, X_{i-1}, x', x_{i+1}, \ldots, x_n) - \phi(X_1, \ldots, X_{i-1}, x, x_{i+1}, \ldots, x_n)] dP(x_{i+1}, \ldots, x_n) \\
\leq c_i \int dP(x_{i+1}, \ldots, x_n) \\
= c_i,
\]
where we used independence of $X_1, \ldots, X_n$ in the third equality, and the bounded difference assumption in the inequality.

References


