EECS 598: Statistical Learning Theory, Winter 2014 **Topic 9**

The Bounded Difference Inequality

Lecturer: Clayton Scott *Scribe: Cheng Zhang*

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 $\phi : A^n \to R$. We say ϕ satisfies the bounded difference assumption if $\exists c_1, \dots, c_n \geqslant 0$ s.t. $\forall i, 1 \leqslant i \leqslant n$*

$$\sup_{x_1,\dots,x_n,x_i' \in A} |\phi(x_1,\dots,x_i,\dots,x_n) - \phi(x_1,\dots,x_i',\dots,x_n)| \leqslant c_i$$

That is, if we subsitute x_i to x_i', while keeping other x_j fixed, ϕ changes by at most c_i.

Theorem 1. *Let $X_1, \dots, X_n$ be arbitrary independent random variables on set A and $\phi : A^n \to R$ satisfy the bounded difference assumption. Then $\forall t > 0$*

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

and

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

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

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

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

$$\Pr\{|\sum_{i=1}^n X_i - \mathbb{E}[\sum_{i=1}^n X_i]| \geqslant t\} \leqslant 2e^{-\frac{2t^2}{\sum_{i=1}^n (b_i - a_i)^2}}.$$

The following lemma is used in establishing the BDI.

Lemma 1. *Let V, Z be random variables s.t. $\mathbb{E}[V|Z] = 0$ w.p. 1. Assume $\exists$ a function ψ and constant $c > 0$ s.t. $\psi(Z) \leqslant V \leqslant \psi(Z) + c$. Then $\forall s > 0$*

$$\mathbb{E}[e^{sV}|Z] \leqslant e^{\frac{s^2c^2}{8}}.$$

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 statisfies 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

$$\begin{aligned}
&\Pr\{\phi(X_1, \ldots, X_n) - \mathbb{E}[\phi(X_1, \ldots, X_n)] \geqslant t\} \\
&\quad = \Pr\{\sum_{i=1}^n V_i \geqslant t\} \\
&\quad = \Pr\{e^{s\sum_{i=1}^n V_i} \geqslant e^{st}\} \\
&\quad \leqslant e^{-st}\mathbb{E}[e^{s\sum_{i=1}^n V_i}] \\
&\qquad \text{(Markov's inequality)} \\
&\quad = e^{-st}\mathbb{E}[e^{s\sum_{i=1}^{n-1} V_i + sV_n}] \\
&\quad = e^{-st}\mathbb{E}_{X_1,\ldots,X_{n-1}}\mathbb{E}_{X_n|X_1,\ldots,X_{n-1}}[e^{s\sum_{i=1}^{n-1} V_i + sV_n}|X_1, \ldots, X_{n-1}] \\
&\qquad \text{(only } V_n \text{ depends on } X_n) \\
&\quad = e^{-st}\mathbb{E}_{X_1,\ldots,X_{n-1}}\{e^{s\sum_{i=1}^{n-1} V_i}\mathbb{E}_{X_n|X_1,\ldots,X_{n-1}}[e^{sV_n}|X_1, \ldots, X_{n-1}]\} \\
&\quad \leqslant e^{-st + \frac{s^2c_n^2}{8}}\mathbb{E}_{X_1,\ldots,X_{n-1}}[e^{s\sum_{i=1}^{n-1} V_i}]
\end{aligned}$$

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

$$\Pr\{\phi(X_1, \ldots, X_n) - \mathbb{E}[\phi(X_1, \ldots, X_n)] \geqslant t\} \leqslant e^{-st + \frac{s^2\sum_{i=1}^n c_i^2}{8}}.$$

Now take $s = \frac{4t}{\sum_{i=1}^n c_i^2}$ to minimize the upper bound, giving

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

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

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

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

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

$$
\begin{aligned}
&\mathbb{E}[\mathbb{E}[\phi(X_1,\dots,X_n)|X_1,X_2,X_3]|X_1,X_2] \\
&\qquad = \mathbb{E}_{X_3|X_1,X_2}[\int \phi(X_1,X_2,X_3,x_4,\dots,x_n)dP(x_4,\dots,x_n|X_1,X_2,X_3)] \\
&\qquad = \int\int \phi(X_1,X_2,x_3,\dots,x_n)\underbrace{dP(x_4,\dots,x_n|X_1,X_2,X_3)dP(x_3|X_1,X_2)}_{dP(x_3,\dots,x_n|X_1,X_2)} \\
&\qquad = \int\int \phi(X_1,X_2,x_3,\dots,x_n)dP(x_3,\dots,x_n|X_1,X_2) \\
&\qquad = \mathbb{E}[\phi(X_1,\dots,X_n)|X_1,X_2].
\end{aligned}
$$

Therefore $\mathbb{E}[V_3|X_1,X_2]=0$.

To show the second part of the claim, define

$$
L_i = \inf_x \mathbb{E}[\phi(X_1,\dots,X_n)|X_1,\dots,X_{i-1},x] - \mathbb{E}[\phi(X_1,\dots,X_n)|X_1,\dots,X_{i-1}]
$$

and

$$
U_i = \sup_{x'} \mathbb{E}[\phi(X_1,\dots,X_n)|X_1,\dots,X_{i-1},x'] - \mathbb{E}[\phi(X_1,\dots,X_n)|X_1,\dots,X_{i-1}].
$$

Clearly, $L_i \leqslant V_i \leqslant U_i$. Furthermore,

$$
\begin{aligned}
U_i - L_i &= \sup_{x'} \mathbb{E}[\phi(X_1,\dots,X_n)|X_1,\dots,X_{i-1},x'] - \inf_x \mathbb{E}[\phi(X_1,\dots,X_n)|X_1,\dots,X_{i-1},x] \\
&= \sup_{x'} \int \phi(X_1,\dots,X_{i-1},x',x_{i+1},\dots,x_n)dP(x_{i+1},\dots,x_n|X_1,\dots,X_{i-1},x') \\
&\quad - \inf_x \int \phi(X_1,\dots,X_{i-1},x,x_{i+1},\dots,x_n)dP(x_{i+1},\dots,x_n|X_1,\dots,X_{i-1},x) \\
&= \sup_{x'} \int \phi(X_1,\dots,X_{i-1},x',x_{i+1},\dots,x_n)dP(x_{i+1},\dots,x_n) \\
&\quad - \inf_x \int \phi(X_1,\dots,X_{i-1},x,x_{i+1},\dots,x_n)dP(x_{i+1},\dots,x_n) \\
&= \sup_{x,x'} \int [\phi(X_1,\dots,X_{i-1},x',x_{i+1},\dots,x_n) - \phi(X_1,\dots,X_{i-1},x,x_{i+1},\dots,x_n)]dP(x_{i+1},\dots,x_n) \\
&\leqslant c_i \int dP(x_{i+1},\dots,x_n) \\
&= c_i,
\end{aligned}
$$

where we used independence of $X_1,\dots,X_n$ in the third equality, and the bounded difference assumption in the inequality. □

References

[1] Colin McDiarmid, "On the method of bounded differences", *Surveys in Combinatorics, 1989*, J. Siemons ed., London Mathematical Society Lecture Note Series 141, Cambridge University Press, 1989, 148-188.

[2] W. Hoeffding, "Probability inequalities for sums of bounded random variables", *Journal of the American Statistical Association*, vol.58, no. 301, pp. 13-30, 1963.

[3] K. Azuma, "Weighted Sums of Certain Dependent Random Variables," *Tohoku Mathematical Journal*, vol. 19, pp. 357-367, 1967.

[4] S. Resnick, *A Probability Path*, Birkhäuser, Boston, 1999.