Math 547: Mathematical Foundations of Statistical Learning Theory Fall 2017

Lecture 26 — October 20

Instructor: Stas Minsker *Scribe: Mose Wintner*

26.1 Bounded Difference Inequality

Definition 1. $(Y_j, \mathcal{F}_j)$ is a martingale if

1. $\mathbb{E}Y_j < \infty$
2. Y_j is $\mathcal{F}_j$-measurable
3. $\mathbb{E}[Y_j|F_{j-1}] = Y_{j-1}$.

Example 1. $X_1, \ldots, X_n$ iid with finite mean, $S_k = \sum_1^k X_j - \mathbb{E}X_j$, $\mathcal{F}_k = \sigma(X_1, \ldots, X_k)$ is a martingale.

$$\mathbb{E}[S_k|\mathcal{F}_{k-1}] = \mathbb{E}[S_{k-1} + X_k|\mathcal{F}_{k-1}] = S_{k-1} + \mathbb{E}[X_k - \mathbb{E}X_k|\mathcal{F}_{k-1}] = S_{k-1}.$$

Theorem 1. Let $X_1, \ldots, X_n$ be iid. Assume that $Z(x_1, \ldots, x_n)$ is such that

$$Z(x_1, \ldots, x_j, \ldots, x_n) = Z(x_1, \ldots, x_j', \ldots, x_n)| \le c_1$$

for all $x_1, \ldots, x_j, x_j', \ldots, x_n$. Then

$$P(|Z(x_1, \ldots, x_n) - \mathbb{E}Z(x_1, \ldots, x_n)| \ge t) \le 2\exp\left(-\frac{t^2}{2\sum c_j^2}\right).$$

Proof: First, we will find $V_1, \ldots, V_n$ such that (letting $Z(x_1, \ldots, x_n) = Z$),

1. $Z - \mathbb{E}Z = \sum V_j$
2. $\mathbb{E}[V_j|V_1, \ldots, V_{j-1}] = 0$
3. $V_j = V_j(X_1, \ldots, X_j)$

This is called the martingale difference decomposition; the V_j are called martingale differences because $S_k = \sum V_j$ is a martingale. Let Y_j be independent copies of the X_j, and let

$$V_j = \mathbb{E}_Y Z(X_1, \ldots, X_j, Y_{j+1}, \ldots, Y_n) - \mathbb{E}_Y[Z(X_1, \ldots, X_{j-1}, Y_j, \ldots, Y_n)].$$

Note that

$$V_j = \mathbb{E}[Z|X_1, \ldots, X_j, Y_{j+1}, \ldots, Y_n] - \mathbb{E}[Z|X_1, \ldots, X_{j-1}, Y_j, \ldots, Y_n].$$

This means

$$\begin{aligned}
V_1 + \cdots + V_n = \mathbb{E}[Z|X_1] - \mathbb{E}Z + \mathbb{E}[Z|X_1, X_2] - \mathbb{E}[Z|X_1] \\
+ \mathbb{E}[Z|X_1, X_2, X_3] - \mathbb{E}[Z|X_1, X_2] + \cdots \\
+ \mathbb{E}[Z|X_1, \ldots, X_n] - \mathbb{E}[Z|X_1, \ldots, X_{n-1}] \\
= Z - \mathbb{E}Z.
\end{aligned}$$

To show 2. note that by integrating out X_j,

$$\mathbb{E}[\mathbb{E}[Z|X_1, \ldots, X_j] - \mathbb{E}[Z|X_1, \ldots, X_{j-1}]|X_1, \ldots, X_{j-1}] = \mathbb{E}[Z|X_1, \ldots, X_{j-1}] - \mathbb{E}Z|X_1, \ldots, X_{j-1} = 0.$$

Next, for 3.,

$$\begin{aligned}
P(Z - \mathbb{E}Z \geq t) &= P(e^{\lambda(Z-\mathbb{E}Z)} \geq e^{\lambda t}) \\
&\leq e^{-\lambda t}\mathbb{E}e^{\lambda(Z-\mathbb{E}Z)} \\
&= e^{-\lambda t}\mathbb{E}e^{\lambda \sum V_j} \\
&= e^{-\lambda t}\mathbb{E}\mathbb{E}[e^{\lambda \sum^{n-1} V_j} e^{\lambda V_n}|X_1, \ldots, X_{n-1}] \\
&= \mathbb{E}e^{\lambda \sum^{n-1} V_j}\mathbb{E}[e^{\lambda V_n}|X_1, \ldots, X_{n-1}] \\
|V_n| &= |\mathbb{E}_Y[Z|X_1, \ldots, X_n] - \mathbb{E}_Y[Z|X_1, \ldots, X_{n-1}, Y_n]| \leq |c_n| \text{ a.s., so } V_n \in SG(c_n^2) \\
&\leq \mathbb{E}^{\lambda c_n^2/2}\mathbb{E}e^{\lambda \sum^{n-1} V_j} \\
\cdots &\leq e^{\lambda^2/2 \sum c_j^2}.
\end{aligned}$$

Hence

$$P(Z - \mathbb{E}Z \geq t) \leq e^{-\lambda t + \lambda^2/2 \sum c_j^2} \leq e^{-\frac{t^2}{2\sum c_j^2}}.$$

Using the same technique to get a bound for $P(Z - \mathbb{E}Z \leq t)$, we finish.