Stat 6390.001: High Dimensional Probability and Statistics, Fall 2026 **Topic 5**

Sub-exponential Variables and the Azuma–Hoeffding Inequality

Lecturer: Yun Wei *Scribe: Fariha Taskin*

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications. They may be distributed outside this class only with the permission of the Instructor.*

1 Recall

We begin by recalling the Johnson–Lindenstrauss lemma and equivalent characterizations of sub-exponential random variables. These results will be used to derive concentration inequalities for martingale difference sequences.

Lemma 1.1 (Johnson–Lindenstrauss lemma). *Suppose that we are given $N \geq 2$ distinct vectors*

$$\left\{u^1, u^2, \ldots, u^N\right\} \subset \mathbb{R}^d.$$

Consider the random map

$$F : \mathbb{R}^d \longrightarrow \mathbb{R}^m, \qquad F(u) = \frac{Xu}{\sqrt{m}},$$

where $X \in \mathbb{R}^{m \times d}$ has independent and identically distributed $N(0,1)$ entries. If

$$m > \frac{16}{\delta^2} \log\left(\frac{N}{\varepsilon}\right),$$

then, with probability at least $1 - \varepsilon$,

$$1 - \delta \leq \frac{\left\|F(u^i) - F(u^j)\right\|_2^2}{\left\|u^i - u^j\right\|_2^2} \leq 1 + \delta \qquad \textit{for every } i \neq j.$$

Theorem 1.2 (Equivalent characterizations of sub-exponential random variables). *For a mean-zero random variable X, the following conditions are equivalent.*

(i) There exist non-negative parameters (ν, α) such that

$$\mathbb{E}\left[e^{\lambda X}\right] \leq e^{\nu^2\lambda^2/2} \qquad \textit{for all } |\lambda| \leq \frac{1}{\alpha}.$$

(ii) There exist constants $c_0, c_0 > 0$ such that

$$\mathbb{E}\left[e^{\lambda X}\right] \leq c_0 \qquad \textit{for all } |\lambda| \leq c_0.$$

(iii) There exist constants $c_1, c_2 > 0$ such that

$$\mathbb{P}\left(|X| \geq t\right) \leq c_1 e^{-c_2 t} \qquad \textit{for all } t \geq 0.$$

(iv) The quantity

$$\gamma := \sup_{k \geq 2} \left(\frac{\mathbb{E}\left[|X|^k\right]}{k!}\right)^{1/k}$$

is finite.

Remark 1.3. A sub-exponential tail bound decays more slowly than a sub-Gaussian tail bound for large t, so the sub-exponential class permits heavier tails.

2 Martingale-Based Methods

Definition 2.1 (Martingale difference sequence)**.** An adapted sequence $\{(D_k, \mathcal{F}_k)\}_{k=0}^{\infty}$ is called a *martingale difference sequence* if, for every $k \geq 1$,

$$\mathbb{E}\left[|D_k|\right] < \infty \qquad \text{and} \qquad \mathbb{E}\left[D_k \mid \mathcal{F}_{k-1}\right] = 0.$$

Equivalently, after shifting the index, $\mathbb{E}\left[D_{k+1} \mid \mathcal{F}_k\right] = 0$.

Remark 2.2. Here, D_k denotes the difference and $\mathcal{F}_k$ denotes the associated sigma-algebra.

Example 2.3. Let $\{(Y_k, \mathcal{F}_k)\}_{k\geq 0}$ be a martingale and define

$$D_k := Y_k - Y_{k-1}, \qquad k \geq 1.$$

Then D_k is $\mathcal{F}_k$-measurable and

$$\mathbb{E}\left[D_{k+1} \mid \mathcal{F}_k\right] = \mathbb{E}\left[Y_{k+1} \mid \mathcal{F}_k\right] - Y_k = 0 \qquad \text{a.s.}$$

Moreover,

$$Y_k - Y_0 = \sum_{i=1}^{k} D_i.$$

Theorem 2.4. *Let $\{(D_k, \mathcal{F}_k)\}$ be a martingale difference sequence. Suppose that, for every $k = 1, \ldots, n$,*

$$\mathbb{E}\left[e^{\lambda D_k} \mid \mathcal{F}_{k-1}\right] \leq e^{\lambda^2 \nu_k^2/2} \qquad \textit{for all } |\lambda| < \frac{1}{\alpha_k}.$$

Define

$$\alpha_* := \max_{1\leq k\leq n} \alpha_k.$$

Then the following statements hold.

(a) $\sum_{k=1}^{n} D_k$ *is sub-exponential with parameters* $\left(\sqrt{\sum_{k=1}^{n} \nu_k^2},\, \alpha_*\right)$.

(b) $\mathbb{P}\left(\left|\sum_{k=1}^{n} D_k\right| \geq t\right) \leq \begin{cases} 2\exp\left(-\dfrac{t^2}{2\sum_{k=1}^{n} \nu_k^2}\right), & 0 \leq t \leq \dfrac{\sum_{k=1}^{n} \nu_k^2}{\alpha_*}, \\ 2\exp\left(-\dfrac{t}{2\alpha_*}\right), & t > \dfrac{\sum_{k=1}^{n} \nu_k^2}{\alpha_*}. \end{cases}$

Remark 2.5. The moment-generating-function assumption in Theorem 2.4 is a conditional sub-exponential condition. Part (a) is the martingale-difference analogue of the standard result that a sum of independent sub-exponential random variables is sub-exponential.

Proof. First, the tower property gives

$$\mathbb{E}[D_k] = \mathbb{E}\left[\mathbb{E}\left[D_k \mid \mathcal{F}_{k-1}\right]\right] = 0,$$

and hence $\mathbb{E}[\sum_{k=1}^{n} D_k] = 0$.

For $|\lambda| < 1/\alpha_*$, condition successively on the filtration:

$$\begin{aligned}
\mathbb{E}\left[e^{\lambda \sum_{k=1}^{n} D_k}\right] &= \mathbb{E}\left[\mathbb{E}\left[e^{\lambda \sum_{k=1}^{n} D_k} \mid \mathcal{F}_{n-1}\right]\right] \\
&= \mathbb{E}\left[e^{\lambda \sum_{k=1}^{n-1} D_k} \mathbb{E}\left[e^{\lambda D_n} \mid \mathcal{F}_{n-1}\right]\right] \\
&\le \mathbb{E}\left[e^{\lambda \sum_{k=1}^{n-1} D_k} e^{\lambda^2 \nu_n^2/2}\right], \quad |\lambda| < \frac{1}{\alpha_n} \\
&= e^{\lambda^2 \nu_n^2/2} \mathbb{E}\left[e^{\lambda \sum_{k=1}^{n-1} D_k}\right], \quad |\lambda| < \frac{1}{\alpha_n} \\
&\le e^{\lambda^2 \nu_n^2/2} e^{\lambda^2 \nu_{n-1}^2/2} \mathbb{E}\left[e^{\lambda \sum_{k=1}^{n-2} D_k}\right], \quad |\lambda| < \frac{1}{\alpha_n}, \ |\lambda| < \frac{1}{\alpha_{n-1}} \\
&\le \cdots \le e^{\lambda^2 \nu_n^2/2} \ldots e^{\lambda^2 \nu_1^2/2}, \quad |\lambda| < \frac{1}{\alpha_n}, \ldots, \ |\lambda| < \frac{1}{\alpha_1} \\
&= \exp\left(\frac{\lambda^2}{2} \sum_{k=1}^{n} \nu_k^2\right), \qquad |\lambda| < \frac{1}{\max_{1 \le k \le n} \alpha_k}.
\end{aligned}$$

This proves part (a). Part (b) follows from the standard two-regime sub-exponential tail bound. □

Remark 2.6. If the D_k were independent, the MGF would factor as $\prod_{k=1}^{n} \mathbb{E}[e^{\lambda D_k}]$. For a martingale difference sequence, conditional expectation replaces this factorization. Given $\mathcal{F}_{n-1}$, the partial sum $\sum_{k=1}^{n-1} D_k$ is $\mathcal{F}_{n-1}$-measurable and therefore acts as a constant inside the conditional expectation.

2.1 Azuma–Hoeffding Inequality

Corollary 2.7 (Azuma–Hoeffding)**.** *Let $\{(D_k, \mathcal{F}_k)\}$ be a martingale difference sequence. Suppose there are constants $\{(a_k, b_k)\}_{k=1}^{\infty}$ such that*

$$D_k \in [a_k, b_k] \qquad \textit{a.s. for every } k.$$

Then, for every $t \ge 0$,

$$\mathbb{P}\left(\left|\sum_{k=1}^{n} D_k\right| \ge t\right) \le 2 \exp\left(-\frac{2t^2}{\sum_{k=1}^{n} (b_k - a_k)^2}\right).$$

Proof. Conditioned on $\mathcal{F}_{k-1}$, the random variable D_k still lies in $[a_k, b_k]$ almost surely. By the conditional form of Hoeffding's lemma, D_k is conditionally $(b_k - a_k)/2$-sub-Gaussian and is therefore conditionally sub-exponential with parameters

$$\left(\frac{b_k - a_k}{2}, 0\right).$$

By Theorem 2.4(a),

$$\sum_{k=1}^{n} D_k \text{ is } \left(\sqrt{\sum_{k=1}^{n} \frac{(b_k - a_k)^2}{4}}, 0\right) \text{ sub-exponential } \longrightarrow \left(\sqrt{\sum_{k=1}^{n} \frac{(b_k - a_k)^2}{4}}\right) \text{ sub-Gaussian.}$$

The stated tail bound follows. □

Remark 2.8. Conditional expectation preserves inequalities: if $X \le b$, then $\mathbb{E}[X \mid \mathcal{F}] \le \mathbb{E}[b \mid \mathcal{F}] = b$. Also, every sub-Gaussian random variable is sub-exponential.

2.2 Bounded Differences

Definition 2.9 (Bounded difference property)**.** We say that a function f satisfies the bounded difference property with parameters $(L_1, \dots, L_n)$ if, for every $k = 1, \dots, n$,

$$|f(x_1, \dots, x_n) - f(x_1, \dots, x_{k-1}, x_k', x_{k+1}, \dots, x_n)| \le L_k$$

for all $x_1, \dots, x_n, x_k'$.

Example 2.10. For $x, y \in \mathbb{R}^n$, define the Hamming distance by

$$d_H(x, y) := \sum_{i=1}^{n} \mathbf{1}\{x_i \neq y_i\}.$$

If f is L-Lipschitz with respect to d_H, meaning that

$$|f(x) - f(y)| \le L\, d_H(x, y),$$

then changing only the kth coordinate gives

$$|f(x_1, \dots, x_n) - f(x_1, \dots, x_{k-1}, x_k', x_{k+1}, \dots, x_n)| \le L \cdot 1 = L.$$

Thus, f has the bounded difference property with $L_k = L$ for every k.

Corollary 2.11 (Bounded difference inequality)**.** *Suppose f satisfies the bounded difference property with parameters $(L_1, \dots, L_n)$ and that $X = (X_1, \dots, X_n)$ has independent components. Then, for every $t \ge 0$,*

$$\mathbb{P}\left(|f(X) - \mathbb{E}[f(X)]| \ge t\right) \le 2 \exp\left(-\frac{2t^2}{\sum_{k=1}^{n} L_k^2}\right).$$

Partial proof. Let

$$\mathcal{F}_k := \sigma(X_1, \dots, X_k), \qquad Y_k := \mathbb{E}\left[f(X) \mid \mathcal{F}_k\right].$$

Since $\mathcal{F}_n$ is generated by all the components of X,

$$Y_n = \mathbb{E}\left[f(X) \mid \mathcal{F}_n\right] = f(X),$$

while

$$Y_0 = \mathbb{E}\left[f(X) \mid \mathcal{F}_0\right] = \mathbb{E}[f(X)].$$

Therefore,

$$f(X) - \mathbb{E}[f(X)] = Y_n - Y_0 = \sum_{k=1}^{n} D_k, \qquad D_k := Y_k - Y_{k-1}.$$

The sequence $\{(D_k, \mathcal{F}_k)\}$ is a martingale difference sequence.

To obtain conditional bounds for D_k, define

$$\begin{aligned}
A_k &:= \inf_x \mathbb{E}\left[f(X) \mid X_1, \dots, X_{k-1}, X_k = x\right] - \mathbb{E}\left[f(X) \mid X_1, \dots, X_{k-1}\right] \\
&= \inf_x \mathbb{E}\left[f(X_1, \dots, X_{k-1}, x, X_{k+1}, \dots, X_n) \mid X_1, \dots, X_{k-1}\right] - \mathbb{E}\left[f(X) \mid X_1, \dots, X_{k-1}\right], \\
B_k &:= \sup_x \mathbb{E}\left[f(X) \mid X_1, \dots, X_{k-1}, X_k = x\right] - \mathbb{E}\left[f(X) \mid X_1, \dots, X_{k-1}\right] \\
&= \sup_x \mathbb{E}\left[f(X_1, \dots, X_{k-1}, x, X_{k+1}, \dots, X_n) \mid X_1, \dots, X_{k-1}\right] - \mathbb{E}\left[f(X) \mid X_1, \dots, X_{k-1}\right].
\end{aligned}$$

Conditioning on $X_1, \dots, X_{k-1}$ is equivalent to conditioning on $\mathcal{F}_{k-1}$.

Because $X_1, \ldots, X_n$ are independent, these bounds can also be written as

$$
\begin{aligned}
A_k &= \inf_x \mathbb{E}_{X_{k+1},\ldots,X_n}\left[f(X_1,\ldots,X_{k-1},x,X_{k+1},\ldots,X_n)\right] - \mathbb{E}_{X_k,\ldots,X_n}\left[f(X_1,\ldots,X_n)\right], \\
B_k &= \sup_x \mathbb{E}_{X_{k+1},\ldots,X_n}\left[f(X_1,\ldots,X_{k-1},x,X_{k+1},\ldots,X_n)\right] - \mathbb{E}_{X_k,\ldots,X_n}\left[f(X_1,\ldots,X_n)\right].
\end{aligned}
$$

Both A_k and B_k are random because they depend on $X_1, \ldots, X_{k-1}$.

The proof will be continued in the next lecture.

References

[1] M. J. Wainwright, *High-Dimensional Statistics: A Non-Asymptotic Viewpoint*, Cambridge University Press, 2019.