MATH888: High-dimensional probability and statistics Fall 2021

Lecture 9 — September 27, 2021

Sebastien Roch, UW-Madison *Scribe: Chenyang Jiang*

1 Overview

In the last lecture we introduced the properties of sub-exponential variables and its relation with sub-gaussian variables. We proved the concentration of norm theorem.

In this lecture we review the properties of conditional expectations, introduce the symmetrization method with corresponding lemma, and discuss the main idea of proving Hanson-Wright inequality.

2 Main Section

We begin by describing the properties of conditional expectations, the slides for this part can be found in `https://people.math.wisc.edu/~roch/hdps/roch-hdps-slides9.pdf`

2.1 A Useful Fact

Lemma 1. *(lemma 6.1.2 in[1]) Let Y and Z be independent random variables such that $\mathbb{E}Z = 0$. Then, for every convex function F, one has*

$$\mathbb{E}F(Y) \leq \mathbb{E}F(Y+Z)$$

Proof.

$$\begin{aligned}
\mathbb{E}F(Y+Z) &= \mathbb{E}(\mathbb{E}(F(Y+Z)|Y)) \quad \text{(Tower property)} \\
&\geq \mathbb{E}(F(\mathbb{E}(Y+Z|Y))) \quad \text{(Jensen's inequality)} \\
&= \mathbb{E}(F(\mathbb{E}(Y|Y)+\mathbb{E}(Z|Y))) \quad \text{(Linearity)} \\
&= \mathbb{E}(F(Y+\mathbb{E}(Z|Y))) \quad \text{(Property (b) in slides 9)} \\
&= \mathbb{E}(F(Y+EZ)) \quad \text{(Property (k) in slides 9)} \\
&= \mathbb{E}(F(Y))
\end{aligned}$$

□

2.2 Symmetrization

Lemma 2. *Let X be a mean-zero random variable with $|X| \leq c$ a.s. Then*

$$\mathbb{E}(exp(\lambda X)) \leq exp(2c^2\lambda^2)$$

Proof. We use symmetrization to proof this lemma. Let Z be a Rademacher variable, i.e. $\mathbb{P}(Z = 1) = \mathbb{P}(Z = -1) = \frac{1}{2}$ and independent of X.
Let $X^{'}$ be an independent copy of X.

$$\begin{aligned}
\mathbb{E}(exp(\lambda X)) &\leq \mathbb{E}(exp(\lambda(X - X^{'}))) \quad \text{(Previous lemma)} \\
&= \mathbb{E}(exp(\lambda Z(X - X^{'}))) \quad (X - X^{'} \text{ is symmetric}) \\
&= \mathbb{E}(\mathbb{E}(exp(\lambda Z(X - X^{'}))|X, X^{'})) \quad \text{(Tower property)} \\
&\leq \mathbb{E}(exp(\frac{1}{2}\lambda^2(X - X^{'})^2)) \\
&\leq exp(2\lambda^2 c^2) \quad (X, X^{'} \text{ are bounded})
\end{aligned}$$

where the fourth inequality follows from the fact that

$$\begin{aligned}
\mathbb{E}(exp(tZ)) &= \frac{e^t + e^{-t}}{2} \quad \text{(Definition of Z)} \\
&= \frac{1}{2}\sum_{k=0}^{\infty} \frac{t^k + (-t)^k}{k!} \quad \text{(Taylor's expansion)} \\
&= \sum_{k=0}^{\infty} \frac{t^{2k}}{(2k)!} \\
&\leq \sum_{k=0}^{\infty} \frac{t^{2k}}{2^k k!} \\
&= exp(\frac{t^2}{2})
\end{aligned}$$

□

2.3 Hanson-Wright Inequality

Theorem 3. *(Theorem 6.2.1 in [1] Hanson-Wright inequality) Let $X = (X_1, X_2, ...X_n) \in \mathbb{R}^n$ be a random vector with independent, mean-zero, sub-gaussian coordinates. Let A be an $n \times n$ deterministic matrix. Then, for every $t \geq 0$, we have*

$$\mathbb{P}\{|X^T AX - \mathbb{E}X^T AX| \geq t\} \leq 2exp[-c\ min(\frac{t^2}{K^4||A||_F^2}, \frac{t}{K^2||A||})]$$

where $K = max_i ||X_i||_{\psi_2}$

We will prove this theorem in next lecture using the following ideas.

2.3.1 Decoupling

Let $A = (a_{ij})_{i,j}$, assume X_i are mean-zero and have unit varianc. Notice that

$$X^T AX = \sum_{i,j} a_{ij} X_i X_j \qquad \mathbb{E}(X^T AX) = \sum_i a_{ii} = tr(A)$$

The first term is hard to handle, we replace it by

$$X^T A X' = \sum_{i,j} a_{ij} X_i X_j'$$

where X' is an independent copy of X. This is better since it is a linear combination of independent random variables if we conditioning on X'.

2.3.2 Compare to gaussians

Then we will compare $X^T A X'$ to gAg', where $g, g' \sim N(0, I_n)$.

2.3.3 Compute explicitly

We can compute the moment generating functions of gaussians explicitly, this will complete the proof.

References

[1] R.Vershynin, *High-dimensional probability: An introduction with applications in data science*, Cambridge university press, 2008.