MATH888: High-dimensional probability and statistics Fall 2021

Lecture 6 — September 20, 2021

Sebastien Roch, UW-Madison *Scribe: Kartik Sreenivasan*

1 Overview

In the last lecture we completed the proof of the Cramer-Rao bound and discussed a few examples.

In this lecture we will begin our foray into *High Dimensional Probability* by Roman Vershynin [1] and discuss sub-gaussian random variables and some of their properties.

2 Sub-Gaussian Random Variables

These are a very useful class of random variables whose tails decay atleast as fast as gaussian random variables.

Notation:

- $a \lesssim b : \exists C > 0$ s.t $a \leq Cb$
- $a \vee b := \max\{a, b\}$

Proposition 1. ***(2.12 in [1])*** *Let $g \sim \mathcal{N}(0, 1)$ i.e., $f_g(t) = \frac{1}{\sqrt{2\pi}} \exp(-t^2/2)$. Then, for any $t \geq 0$,*

$$\left(\frac{1}{t} - \frac{1}{t^3}\right) \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \leq \mathbb{P}(g \geq t) \leq \left(\frac{1}{t}\right) \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \tag{1}$$

Definition 2. *The **sub-gaussian norm** of $X \in \mathbb{R}$ is*

$$\|X\|_{\psi_2} := \inf\left\{t \geq 0 : \mathbb{E}\left[\psi_2\left(\frac{|X|}{t}\right)\right] \leq 1\right\} \tag{2}$$

where $\psi_2(x) = e^{x^2} - 1$

Remarks:

- The sub-gaussian norm is a valid norm and therefore obeys useful properties such as absolute homogeneity and the triangle inequality.
- The "-1" is a convention chosen so that $\psi_2(0) = 0$
- Refer [1] for a discussion on the connection with ℓ_p spaces.

2.1 Sub-gaussian norms of some examples

Gaussian case: $g \sim \mathcal{N}(0,1)$.
First note that we can push the "-1" to the other side of the inequality in Equation (2) and therefore, it suffices to consider

$$\begin{aligned}\mathbb{E}\left[\psi_2\left(\frac{|g|}{t}\right)\right] &= \mathbb{E}\left[\exp\left(\frac{g^2}{t^2}\right)\right] \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} exp\left(\frac{-x^2}{2} + \frac{x^2}{t^2}\right) dx\end{aligned}$$

It is clear that the above integral is finite if and only if $t \geq \sqrt{2}$. Moreover, $\mathbb{E}[\psi_2(|g|/t)] \to 1$ as $t \to \infty$ as the integral reduces to the integral of the gaussian pdf. Therefore, $\exists t^* > \sqrt{2}$ such that $\mathbb{E}[\psi_2(|g|/t)] \leq 2$ so that $\|g\|_{\psi_2} = t^*$.

Bounded case: $X \in [a,b]$ a.s.
Then, $\|X\|_{\psi_2} \lesssim |a| \vee |b|$

2.2 Useful properties of the sub-gaussian norm

Exercise (2.5.7 in [1]) Show that

$$\|X+Y\|_{\psi_2} \leq \|X\|_{\psi_2} + \|Y\|_{\psi_2}$$

Proof. Note that,

$$\begin{aligned}\psi_2\left(\frac{|X+Y|}{a+b}\right) &\leq \psi_2\left(\frac{|X|+|Y|}{a+b}\right) \\ &\leq \frac{a}{a+b}\psi_2\left(\frac{|X|}{b}\right) + \frac{b}{a+b}\psi_2\left(\frac{|Y|}{b}\right)\end{aligned}$$

where the first inequality holds since ψ_2 is an increasing function of its argument and the second is from applying Jensen's inequality since ψ_2 is convex.
Therefore, for any $a > \|X\|_{\psi_2}$, $b > \|Y\|_{\psi_2}$

$$\begin{aligned}\mathbb{E}\left[\left(\frac{|X+Y|}{a+b}\right)\right] &\leq \frac{a}{a+b}\mathbb{E}\left[\psi_2\left(\frac{|X|}{a}\right)\right] + \frac{b}{a+b}\mathbb{E}\left[\psi_2\left(\frac{|Y|}{b}\right)\right] \\ &\leq 1\end{aligned}$$

where both expectations are bounded by 1 simply by applying the definition of sub-gaussian norm.

Theorem 3. ***(2.5.2 in [1]).*** *If $\|X\|_{\psi_2} < +\infty$, then the following properties are equivalent*

(i) $\mathbb{P}(|X| > t) \leq 2\exp\left(\frac{-ct^2}{K^{@}}\right)$, *for all* $t \geq 0$.

(ii) $\|X\|_p = (\mathbb{E}(|\mathbb{X}|^{'}))^{\frac{1}{p}} \leq C \cdot K\sqrt{p}$, *for all* $p \geq 1$.

(v) If $\mathbb{E}X = 0$*, then* $\mathbb{E}[\exp(\lambda X)] \le \exp(C\lambda^2 K^2)$*, for all* $\lambda \in \mathbb{R}$

where $K = \|X\|_{\psi_2}$*,* c *is a small universal constant and* C *is a large universal constant. Vice Versa, if any of the above properties is true, then* $\|X\|_{\psi_2} \lesssim K$*.*

Proof ideas.
(i) $\Rightarrow$ **(ii)**: $\mathbb{E}|X|^p = \int_0^\infty \mathbb{P}(|X| \ge t)pt^{p-1}dt = \dots$ The $\sqrt{p}$ should come naturally out of the calculus.

(ii) $\Rightarrow \|\mathbf{X}\|_{\psi_2} < +\infty$: Using the taylor's expansion, $\mathbb{E}[\exp(\lambda^2 x^2)] = 1 + \sum_{p=1}^\infty \frac{\lambda^{2p}\mathbb{E}[X^{2p}]}{p} = \dots$. Applying property (ii) and massaging the terms should give the result.

$\|\mathbf{X}\|_{\psi_2} < +\infty \Rightarrow$ **(i)**: For simplicity, assume that $\|X\|_{\psi_2} = 1$. The proof is identical otherwise with just appropriate scaling.

$$\begin{aligned}
\mathbb{P}(|X| \ge t) &= \mathbb{P}(e^{X^2} \ge e^{t^2}) \\
&\le e^{-t^2}\mathbb{E}[e^{X^2}] \quad \text{(Markov's inequality)} \\
&\le 2e^{-t^2}
\end{aligned}$$

(v) $\Rightarrow$ **(i)**:

$$\begin{aligned}
\mathbb{P}(X \ge t) &= \mathbb{P}(e^{\lambda x} \ge e^{\lambda t}),\ \lambda > 0 \\
&\le e^{-\lambda t}\mathbb{E}[e^{\lambda x}] \quad \text{(Markov's inequality)} \\
&\le e^{-\lambda t + C\lambda^2} \quad \text{(applying (v))} \\
&= e^{-t^2/4c}
\end{aligned}$$

where the last equality follows by choosing a good λ.

References

[1] Vershynin, Roman, *High-dimensional probability: An introduction with applications in data science*. Vol. 47. Cambridge university press, 2018,