# 2 27/12/2019 - High-dimensional probability. An Introduction with applications in Data Science (Vershynin R.)

На семинаре обсуждались случайные величины с субгауссовским распределением и их различные свойства. В частности были разобраны многие моменты доказательства следующего предложения:

Theorem 4. (Sub-gaussian properties) Let $X$ be a random variable. Then the following properties are equivalent; the parameters $K_i > 0$ appearing in these properties differ from each other by at most an absolute constant factor.

- The tails of X satisfy
$$\mathbb{P}\{|X| \geq t\} \leq 2\exp\left(-t^2/K_1^2\right) \quad \text{for all } t \geq 0$$
- The moments of $X$ satisfy
$$\|X\|_p = (\mathbb{E}|X|^p)^{1/P} \leq K_2\sqrt{p} \quad \text{for all } p \geq 1$$
- The $MGF$ of $X^2$ is finite at some point, namely
$$\mathbb{E}\exp\left(X^2/K_3^2\right) \leq 2$$
Moreover, if $\mathbb{E}X = 0$ then properties $1-3$ are also equivalent to the following one.
- The MGF of X satisfies
$$\mathbb{E}\exp(\lambda X) \leq \exp\left(\lambda^2 K_4^2\right) \quad \text{for all } \lambda \in \mathbb{R}$$

Замечание: одним из самых затруднительных моментов, который внёс некоторую путаницу, был момент про абсолютную константу $C* > 1$, связывающую значения $K_1$–$K_4$. Есть предположение, что она не зависит от распределения и позволяет при фиксации $K_1$ сказать, что $K_i \in K_1[\frac{1}{C*}, C*]$, при $i = 2, 3, 4$.

На основе данных свойств была введена норма:
$$\|X\|_{\psi_2} = \inf\{t > 0 : \ \mathbb{E}\exp(X^2/t^2) \leq 2\}.$$

Замечание: Для мистической двойки строгого обоснования не было найдено. Тем не менее эта двойка позволяет получать первое свойство последней теоремы из третьего без изменения констант.

Была упомянута связь с пространствами Орлича с функцией Орлича $\psi(x) = \exp(x^2) - 1$. В связи с этим мы обсудили идею доказательства того, что упомянутая норма является нормой.

Наконец, были обсуждены следующие утверждения:

Theorem 5. (Sums of independent sub-gaussians) Let $X_1, \dots, X_N$ be independent, mean zero, sub-gaussian random variables. Then $\sum_{i=1}^N X_i$ is also a sub-gaussian random variable, and

$$\left\| \sum_{i=1}^N X_i \right\|_{\psi_2}^2 \leq C \sum_{i=1}^N \|X_i\|_{\psi_2}^2$$

where $C$ is an absolute constant.

Theorem 6. (General Hoeffding's inequality) Let $X_1, \dots, X_N$ be independent, mean zero, sub-gaussian random variables. Then, for every $t \geq 0$, we have

$$\mathbb{P}\left\{ \left| \sum_{i=1}^N X_i \right| \geq t \right\} \leq 2 \exp\left( - \frac{ct^2}{\sum_{i=1}^N \|X_i\|_{v_2}^2} \right)$$

Как следствие последнего мы получили класс субгауссовских распределений эквивалентен классу распределений, для которых выполнено неравенство Хёфдинга.