Topics in High-Dimensional Probability and Statistics*

Lecture 1: Sub-gaussian distributions

Contents

This is the first in a series of lectures on High-Dimensional Probability and Statistics. The topics selected in these lectures do not claim to represent the field and are biased by the personal choices of the author. The lectures mainly follow the presentation of [7]. Additional recommended references are [2, 5, 1, 4, 6, 8].

1 On the gaussian distribution

The goal of the first few lectures will be to have a first look at the notion of concentration of random variables, i.e., to quantify the fact that certain random variables stay close to their expectation. Due to its universality, illustrated by the central limit theorem, the gaussian distribution is a natural first example that we'll shortly focus on. The results provided in this section will serve as a benchmark for further discussions. Let

$$\phi(t) := \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right),$$

be the density function of the gaussian $\mathcal{N}(0,1)$ distribution.

Theorem 1.1. *Suppose X follows the gaussian distribution $\mathcal{N}(0,1)$. Then, for all $t > 0$,*

$$\frac{t\phi(t)}{1+t^2} \le \mathbb{P}\{X > t\} \le \frac{\phi(t)}{t}. \tag{1.1}$$

Proof. Observe first that for any integrable function $f : \mathbb{R} \to (0, +\infty)$, and any $t > 0$,

$$\int_t^{+\infty} f(x)\,\mathrm{d}x \le \frac{1}{t}\int_t^{+\infty} x f(x)\,\mathrm{d}x.$$

Applying this inequality to ϕ yields, for all $t > 0$,

$$\begin{aligned}
\mathbb{P}\{X > t\} &= \int_t^{+\infty} \phi(x)\,\mathrm{d}x \\
&\le \frac{1}{t}\int_t^{+\infty} x\phi(x)\,\mathrm{d}x \\
&= \frac{\phi(t)}{t}.
\end{aligned}$$

For the second part, we integrate by parts the inequality that we have just obtained,

$$0 \le -x\mathbb{P}\{X > x\} + \phi(x), \quad x > 0,$$

to get

$$\begin{aligned}
0 &\le \int_t^{+\infty} -x\mathbb{P}\{X > x\} + \phi(x)\,\mathrm{d}x \\
&= \frac{t^2}{2}\mathbb{P}\{X > t\} - \int_t^{+\infty} \frac{x^2}{2}\phi(x)\,\mathrm{d}x + \int_t^{+\infty} \phi(x)\,\mathrm{d}x \\
&= \frac{t^2}{2}\mathbb{P}\{X > t\} - \frac{t}{2}\phi(t) + \frac{1}{2}\int_t^{+\infty} \phi(x)\,\mathrm{d}x \\
&= \frac{(1+t^2)}{2}\mathbb{P}\{X > t\} - \frac{t}{2}\phi(t),
\end{aligned}$$

where we have used that $\phi'(t) = -t\phi(t)$. □

The next inequality improves the previous upper bound for small values of $t > 0$.

Theorem 1.2. *Suppose X follows the gaussian distribution $\mathcal{N}(0,1)$. Then, for all $t > 0$,*

$$\mathbb{P}\{X > t\} \le \frac{1}{2}\exp\left(-\frac{t^2}{2}\right). \tag{1.2}$$

Proof. Consider the function $F : \mathbb{R}_+ \to \mathbb{R}$, defined by

$$F(t) = \frac{1}{2}\exp\left(-\frac{t^2}{2}\right) - \mathbb{P}\{X > t\}.$$

We need to show that F takes only positive values. To that aim, observe that F is continuously differentiable on $\mathbb{R}_+$, that $F(0) = 0$, and that

$$\begin{aligned}
F'(t) &= -\frac{t}{2}\exp\left(-\frac{t^2}{2}\right) - \frac{\mathrm{d}}{\mathrm{d}t}\frac{1}{\sqrt{2\pi}}\int_t^{+\infty} \exp\left(-\frac{x^2}{2}\right)\mathrm{d}x \\
&= \left\{\frac{1}{\sqrt{2\pi}} - \frac{t}{2}\right\}\exp\left(-\frac{t^2}{2}\right).
\end{aligned}$$

This already proves that F is positive on $[0, \sqrt{2/\pi}]$. Finally, for $t > \sqrt{2/\pi}$, using the same trick as for the upper bound in Theorem 1.1, we obtain

$$\begin{aligned}
\mathbb{P}\{X > t\} &= \frac{1}{\sqrt{2\pi}}\int_t^{+\infty} \exp\left(-\frac{x^2}{2}\right)\mathrm{d}x \\
&\le \frac{1}{\sqrt{2\pi}}\sqrt{\frac{\pi}{2}}\int_t^{+\infty} x\exp\left(-\frac{x^2}{2}\right)\mathrm{d}x \\
&= \frac{1}{2}\exp\left(-\frac{t^2}{2}\right),
\end{aligned}$$

which implies that F is positive for $t > \sqrt{2/\pi}$ and completes the proof. □

*Teaching material can be found at https://www.qparis-math.com/teaching.

Remark 1.3. *One can show that the constant 1/2 in Theorem 1.2 is optimal in the sense that*

$$\sup_{t>0}\left\{\exp\left(\frac{t^2}{2\sigma^2}\right)\mathbb{P}\{X>t\}\right\}=\frac{1}{2}.$$

Remark 1.4. *Theorems 1.1 and 1.2 apply to the $\mathcal{N}(m,\sigma^2)$ distribution for any $m\in\mathbb{R}$ and $\sigma^2>0$ by a simple scaling argument: If $X\sim\mathcal{N}(m,\sigma^2)$ then*

$$\frac{X-m}{\sigma}\sim\mathcal{N}(0,1).$$

2 Moment generating functions

The results derived in the previous section rely heavily on the specific form of the gaussian density. In particular, one cannot reproduce the previous arguments in situations where the information on the distribution of X is of more general nature. The present section develops a tool to deal with more complex scenarios. We start with a very basic result.

Theorem 2.1 (Markov's inequality)**.** *Let X be a nonnegative random variable such that $\mathbb{E}X<+\infty$. Then, for all $t\geq 0$,*

$$t\mathbb{P}\{X\geq t\}\leq\mathbb{E}X.$$

Proof. We simply notice that,

$$t\mathbb{P}\{X\geq t\}=t\mathbb{E}\mathbf{1}\{X\geq t\}\leq\mathbb{E}[X\mathbf{1}\{X\geq t\}]\leq\mathbb{E}X,$$

for any $t\geq 0$. □

Exercise 2.2 (Chebychev's inequality)**.** *Using Markov's inequality, show that for any $\mathbb{R}$-valued random variable X such that $\mathbb{E}X^2<+\infty$ and any $t\geq 0$,*

$$t^2\mathbb{P}\{|X-\mathbb{E}X|\geq t\}\leq\mathbb{V}(X).$$

The simple idea used in the proof of Markov's inequality can be generalised and turned into a powerful method known as the Cramér-Chernoff method. To describe this method, consider a random variable X and any nonnegative and increasing[1] $\varphi:\mathbb{R}\to\mathbb{R}_+$. Then, for all $t\geq 0$,

$$\varphi(t)\mathbb{P}\{X\geq t\}=\varphi(t)\mathbb{P}\{\varphi(X)\geq\varphi(t)\}\leq\mathbb{E}\varphi(X),$$

by Markov's inequality. In particular, since $\mathbb{P}\{X\geq t\}$ does not depend on the choice of φ, it follows that

$$\mathbb{P}\{X\geq t\}\leq\inf_{\varphi\in\Phi}\varphi(t)^{-1}\mathbb{E}\varphi(X),\tag{2.1}$$

for any collection Φ of nonnegative and increasing functions $\varphi:\mathbb{R}\to\mathbb{R}_+$. Due to the specific algebraic properties of the exponential function, a very convenient choice for the class Φ is provided by the collection of all functions $x\mapsto e^{\theta x},\theta>0$.

Definition 2.3 (Moment generating function)**.** *The moment generating function[2] (MGF) of X, is defined, for all $\theta\in\mathbb{R}$, by*

$$M_X(\theta):=\mathbb{E}\exp(\theta X).$$

We'll denote, for $\theta\in\mathbb{R}$,

$$\Lambda_X(\theta)=\log M_X(\theta).$$

[1] By increasing we mean that $x<y$ implies $\varphi(x)<\varphi(y)$.

[2] Also referred to as the Laplace transform of the distribution of X

The most important insight, relative to the MGF, is summarised in the following result.

Theorem 2.4 (Cramér-Chernoff)**.** *For any real-valued random variable X and any $t\in\mathbb{R}$, defining*

$$\Lambda_X^*(t):=\sup_{\theta>0}\{\theta t-\Lambda_X(\theta)\},$$

we have

$$\mathbb{P}\{X>t\}\leq e^{-\Lambda_X^*(t)}.$$

Proof. For all $\theta>0$, using that thet function $x\mapsto e^{\theta x}$ is increasing, we deduce that for all $t\in\mathbb{R}$,

$$\begin{aligned}\mathbb{P}\{X>t\}&=\mathbb{P}\{\exp(\theta X)>\exp(\theta t)\}\\&\leq\exp(-\theta t)\,\mathbb{E}\exp(\theta X)\\&=\exp(-\theta t+\Lambda_X(\theta)),\end{aligned}$$

where the second line follows from Markov's inequality. The result follows by optimising the bound in $\theta>0$. □

Exercise 2.5 (Bernoulli distribution)**.** *For $p\in(0,1)$, consider a random variable ξ such that*

$$\mathbb{P}\{\xi=0\}=1-p\quad\textit{and}\quad\mathbb{P}\{\xi=1\}=p,$$

and let $X=\xi-p$. For all $t\in(0,1-p)$, show that

$$\mathbb{P}\{X>t\}\leq e^{-h_p(t+p)},$$

where

$$h_p(u):=u\log\frac{u}{p}+(1-u)\log\frac{1-u}{1-p}.$$

Note that, for $t\geq 1-p$, we obviously have $\mathbb{P}\{X>t\}=0$.

The algebraic property $e^{x+y}=e^xe^y$ of the exponential function implies that the MGF behaves favorably in the context of independent random variables as described in the next exercise.

Exercise 2.6. *Let $X_1,\dots,X_n$ be independent random variables and set $X=X_1+\cdots+X_n$. Then, for all $\theta\in\mathbb{R}$,*

$$\Lambda_X(\theta)=\sum_{i=1}^n\Lambda_{X_i}(\theta).$$

In particular, if the variables $X_1,\dots,X_n$ are i.i.d., then for all $t\in\mathbb{R}$,

$$\Lambda_X^*(t)=n\Lambda_{X_1}^*\left(\frac{t}{n}\right).$$

Exercise 2.7 (Binomial distribution)**.** *Consider a variable ξ distributed according to the Binomial distribution with parameters $n\geq 1$ and $p\in(0,1)$, i.e.*

$$\mathbb{P}\{\xi=k\}=\binom{n}{k}p^k(1-p)^{n-k},$$

and let X be the centered random variable $X=\xi-np$. Using Exercises 2.5 and 2.6, show that for all $t\in(0,n(1-p))$,

$$\mathbb{P}\{X>t\}\leq e^{-nh_p\left(\frac{t}{n}+p\right)},$$

where h_p is as in Exercise 2.5. Note that for $t\geq n(1-p)$, we have $\mathbb{P}\{X>t\}=0$.

We end the section by studying the case of the gaussian distribution.

Lemma 2.8. *Let $m \in \mathbb{R}$ and $\sigma^2 > 0$. Suppose that X follows the gaussian distribution $\mathcal{N}(m, \sigma^2)$. Then, for all $\theta \in \mathbb{R}$,*

$$\log \mathbb{E} \exp(\theta\{X - m\}) = \frac{\theta^2\sigma^2}{2}.$$

Proof. Without loss of generality, suppose $m = 0$. Then, for $\theta \in \mathbb{R}$, we obtain

$$\begin{aligned}
\mathbb{E} \exp(\theta X) &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp\left(\theta x - \frac{x^2}{2\sigma^2}\right) \mathrm{d}x \\
&= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp\left(\frac{\theta^2\sigma^2}{2} - \frac{(x-\sigma^2\theta)^2}{2\sigma^2}\right) \mathrm{d}x \\
&= \exp\left(\frac{\theta^2\sigma^2}{2}\right) \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp\left(-\frac{(x-\sigma^2\theta)^2}{2\sigma^2}\right) \mathrm{d}x \\
&= \exp\left(\frac{\theta^2\sigma^2}{2}\right),
\end{aligned}$$

which concludes the proof. □

Using the computation of the previous Lemma, applying Theorem 2.4 to the gaussian distribution yields that, for all $t > 0$,

$$\mathbb{P}\{X - m > t\} = \mathbb{P}\{X - m < -t\} \le \exp\left(-\frac{t^2}{2\sigma^2}\right),$$

whenever $X \sim \mathcal{N}(m, \sigma^2)$. Comparing this bound with Theorems 1.1 and 1.2, we observe that the method presented in Theorem 2.4 gives a result only slightly weaker than if we had used the specific form of the gaussian density as in Section 1. This advocates for a use of Theorem 2.4 in more general situations. A case of particular interest is the case of subgaussian random variables.

3 Sub-gaussian distributions

In the previous section we have shown how the deviations of a random variable X can be investigated by studying the behavior of its MGF. Motivated by the result of Lemma 2.8, we introduce a specific class of distributions that have lighter tails than gaussian distributions.

Definition 3.1. *A real-valued random variable X (or its distribution) is said to be subgaussian if there exists $s^2 > 0$ such that,*

$$\forall \theta \in \mathbb{R}: \quad \log \mathbb{E} \exp(\theta\{X - \mathbb{E}X\}) \le \frac{\theta^2 s^2}{2}.$$

Whenever this holds, we'll denote $X \in \mathrm{SG}(s^2)$. The smallest $s^2 > 0$ for which $X \in \mathrm{SG}(s^2)$ is called the variance proxy of X, sometimes denoted $\|X\|_{\mathrm{vp}}^2$, and given by

$$\|X\|_{\mathrm{vp}}^2 = \sup_{\theta \ne 0} \frac{2}{\theta^2} \log \mathbb{E} \exp(\theta\{X - \mathbb{E}X\}).$$

According to Lemma 2.8, a gaussian variable is clearly subgaussian. Other examples of subgaussian variables are discussed below.

Remark 3.2. *The notation $\|X\|_{\mathrm{vp}}^2$ relates to the fact that the variance proxy is a squared semi-norm on the set of subgaussian random variables. More precisely, the reader may show as an exercise that the set of subgaussian random variables is indeed an $\mathbb{R}$-vector space and that the following properties hold.*

(1) For any subgaussian variable X and any $\alpha \in \mathbb{R}$,

$$\|\alpha X\|_{\mathrm{vp}} = |\alpha|\, \|X\|_{\mathrm{vp}}.$$

(2) For any subgaussian variables X, Y (non necessarily independent)

$$\|X + Y\|_{\mathrm{vp}} \le \|X\|_{\mathrm{vp}} + \|Y\|_{\mathrm{vp}}.$$

(3) For any subgaussian variable X,

$$\|X\|_{\mathrm{vp}} = 0 \quad \Leftrightarrow \quad X = \mathbb{E}X \ a.s..$$

In particular, $\|.\|_{\mathrm{vp}}$ defines a norm on the space of centered subgaussian random variables. For more details in this direction, we refer the reader to [3].

Theorem 3.3. *Suppose that X is subgaussian. Then, for all $t \in \mathbb{R}$,*

$$\mathbb{P}\{X - \mathbb{E}X > t\} \vee \mathbb{P}\{X - \mathbb{E}X < -t\} \le \exp\left(-\frac{t^2}{2\|X\|_{\mathrm{vp}}^2}\right).$$

Proof. By definition, X is subgaussian if and only if $-X$ is subgaussian and $\|X\|_{\mathrm{vp}}^2 = \|-X\|_{\mathrm{vp}}^2$. As a result, it is enough to prove the first inequality. Now, combining Theorem 2.4 and Definition 3.1, we obtain

$$\begin{aligned}
\mathbb{P}\{X - \mathbb{E}X > t\} &\le \exp\left(-\sup_{\theta \ge 0}\left\{\theta t - \frac{\theta^2\|X\|_{\mathrm{vp}}^2}{2}\right\}\right) \\
&= \exp\left(-\frac{t^2}{2\|X\|_{\mathrm{vp}}^2}\right),
\end{aligned}$$

which completes the proof. □

The class of subgaussian random variables is much wider than the class of gaussian variables. The next result shows, for instance, that any bounded random variable is subgaussian.

Lemma 3.4 (Hoeffding's lemma)**.** *Let X be an $[a, b]$-valued random variable for $-\infty < a < b < +\infty$. Then, for all $\theta \in \mathbb{R}$,*

$$\log \mathbb{E} \exp(\theta\{X - \mathbb{E}X\}) \le \frac{\theta^2(b-a)^2}{8}.$$

In other words $\|X\|_{\mathrm{vp}}^2 \le (b-a)^2/4$.

Proof. Note that, by the convexity of the exponential function,

$$e^{\theta x} \le \frac{x-a}{b-a} e^{\theta b} + \frac{b-x}{b-a} e^{\theta a},$$

for all $a \le x \le b$. Exploiting the fact that $\mathbb{E}[X - \mathbb{E}X] = 0$, and introducing the notation $p = -a/(b-a)$, we deduce that

$$\begin{aligned}
\mathbb{E} \exp(\theta\{X - \mathbb{E}X\}) &\le \frac{b}{b-a} e^{\theta a} - \frac{a}{b-a} e^{\theta b} \\
&= (1 - p + p e^{\theta(b-a)}) e^{-\theta p(b-a)} \\
&= e^{f(u)},
\end{aligned}$$

where $u = \theta(b-a)$ and $f(u) = -pu + \log(1 - p + pe^u)$. By straightforward computations, we get

$$f'(u) = -p + \frac{p}{p + (1-p)e^{-u}},$$

so that $f(0) = f'(0) = 0$. Moreover, for all $c \geq 0$,

$$f''(c) = \frac{p(1-p)e^{-c}}{(p + (1-p)e^{-c})^2} \leq \frac{1}{4}.$$

Thus, by the Taylor-Lagrange theorem, there exists $c \in [0, u]$ such that,

$$f(u) = f(0) + uf'(0) + \frac{u^2}{2} f''(c) \leq \frac{u^2}{8} = \frac{\theta^2(b-a)^2}{8},$$

which concludes the proof. $\square$

Remark 3.5 (Variance vs variance proxy)**.** *As proven in Lemma 2.8, any gaussian random variable is subgaussian with variance proxy equal to its variance. However a random variable may be subgaussian with variance proxy strictly larger than its variance. For example, if X follows the Bernoulli distribution with parameter $p \in (0,1)$, $p \neq 1/2$, the variance of X is $p(1-p)$ while its variance proxy is*

$$\sup_{\theta \neq 0} \frac{2}{\theta^2} \log(pe^{\theta(1-p)} + (1-p)e^{-\theta p}) = \frac{1-2p}{2(\log(1-p) - \log p)},$$

which is strictly larger than $p(1-p)$. More generally, it may be proven as an exercise that inequality

$$\operatorname{var}(X) \leq \|X\|_{\mathrm{vp}}^2,$$

always holds.

4 Equivalent characterizations

Next, we prove an equivalent characterisation of subgaussianity based on moments. We show that a random variable X is subgaussian if, and only if, there exists an absolute constant $C > 0$ such that

$$\forall k \geq 1, \quad (\mathbb{E}|X - \mathbb{E}X|^k)^{\frac{1}{k}} \leq C\sqrt{k}.$$

The following result puts a slight emphasis on constants.

Theorem 4.1. (1) *Let X be a sub-gaussian random variable. Then,*

$$\forall k \geq 1, \quad (\mathbb{E}|X - \mathbb{E}X|^k)^{\frac{1}{k}} \leq c_k\sqrt{k}\|X\|_{\mathrm{vp}},$$

where $c_k = \sqrt{2}k^{\frac{1}{k}-\frac{1}{2}}\Gamma(k/2)^{\frac{1}{k}}$. In particular $(c_k)_{k\geq 1}$ is a decreasing sequence.
(2) *Conversely, suppose there exists $C > 0$ such that,*

$$\forall k \geq 1, \quad (\mathbb{E}|X - \mathbb{E}X|^k)^{\frac{1}{k}} \leq C\sqrt{k}.$$

Then X is sub-gaussian and $\|X\|_{\mathrm{vp}} \leq 2C\sqrt{2e}$.

Remark 4.2 (Constant c_k)**.** *In some cases, it will be useful to evaluate the magnitude of constants. To this aim, we notice that the first terms of the decreasing sequence $(c_k)_{k\geq 1}$ displayed in Theorem 4.1 are given by*

$$c_k = \begin{cases} \sqrt{2\pi} & \approx 2.507 \quad \text{for } k = 1, \\ \sqrt{2} & \approx 1.414 \quad \text{for } k = 2, \\ (2\pi/3)^{1/6} & \approx 1.131 \quad \text{for } k = 3, \\ 1 & \quad \text{for } k = 4. \end{cases}$$

Proof of Theorem 4.1. Without loss of generality we may suppose that $\mathbb{E}X = 0$.

(1) Denote $s^2 = \|X\|_{\mathrm{vp}}^2$ for brevity. Then, for all $k \geq 1$,

$$\begin{aligned} \mathbb{E}|X|^k = \int_0^{+\infty} \mathbb{P}\{|X|^k > t\}\,\mathrm{d}t &= \int_0^{+\infty} \mathbb{P}\{|X| > t^{1/k}\}\,\mathrm{d}t \\ &\leq 2\int_0^{+\infty} \exp(-t^{\frac{2}{k}}/2s^2)\,\mathrm{d}t \qquad (4.1) \\ &= k(2s^2)^{\frac{k}{2}} \int_0^{+\infty} u^{\frac{k}{2}-1}e^{-u}\,\mathrm{d}u \\ &= (2s^2)^{\frac{k}{2}} k\Gamma(k/2), \end{aligned}$$

where inequality (4.1) follows from Theorem 3.3. Hence, for $k \geq 1$,

$$k^{-\frac{1}{2}}(\mathbb{E}|X|^k)^{\frac{1}{k}} \leq \sqrt{2s^2}k^{\frac{1}{k}-\frac{1}{2}}\Gamma(k/2)^{\frac{1}{k}} = c_k\|X\|_{\mathrm{vp}}. \qquad (4.2)$$

The fact that $c_1 \leq \sqrt{2\pi}$ and $c_2 \leq \sqrt{2}$ follows from the fact that $\Gamma(1/2) = \sqrt{\pi}$ and $\Gamma(1) = 1$. The other statements in Remark 4.2 follow from direct computations using the fact that $\Gamma(1+x) = x\Gamma(x)$.

(2) Let us start by observing that, since X is centered, the convexity of $x \mapsto \exp(-\theta x)$ for any $\theta \in \mathbb{R}$ implies that

$$\mathbb{E}\exp(-\theta X) \geq 1, \qquad (4.3)$$

by Jensen's inequality. Now, introduce an independent copy Y of X (that is Y is independent of X and has same distribution). Then, it follows by (4.3) that, for all $\theta \in \mathbb{R}$,

$$\begin{aligned} \mathbb{E}\exp(\theta X) &\leq \mathbb{E}\exp(\theta\{X - Y\}) \\ &= \sum_{k=0}^{+\infty} \frac{\theta^{2k}\mathbb{E}(X-Y)^{2k}}{(2k)!}, \qquad (4.4) \end{aligned}$$

where all terms of odd order cancel due to the symmetry of $X - Y$. Next, by convexity of $x \mapsto x^{2k}$, it follows from Jensen's inequality that

$$\mathbb{E}(X-Y)^{2k} \leq 2^{2k-1}(\mathbb{E}X^{2k} + \mathbb{E}Y^{2k}) = 2^{2k}\mathbb{E}X^{2k}.$$

Combining this inequality with (4.4) and the assumption on the moments of X, we deduce that

$$\begin{aligned} \mathbb{E}\exp(\theta X) &\leq \sum_{k=0}^{+\infty} \frac{(2\theta)^{2k}\mathbb{E}X^{2k}}{(2k)!} \\ &\leq \sum_{k=0}^{+\infty} \frac{(2\theta C)^{2k}(2k)^k}{(2k)!} \\ &= \sum_{k=0}^{+\infty} \frac{(4\theta^2C^2)^k}{k!}\frac{k!(2k)^k}{(2k)!}. \end{aligned}$$

Using finally that

$$2^k(k!)^2 \leq (2k)! \quad \text{and} \quad k^k \leq k!e^k,$$

it follows that

$$\frac{k!(2k)^k}{(2k)!} \leq e^k,$$

which implies that

$$\mathbb{E}\exp(\theta X) \leq \sum_{k=0}^{+\infty} \frac{(4\theta^2C^2e)^k}{k!} = \exp\left(4\theta^2C^2e\right).$$

The result follows by definition of the variance proxy. $\square$

Remark 4.3. *In [7], the author uses the notion of subgaussian norm, defined by*

$$\|X\|_{\psi_2} := \inf\{t > 0 : \mathbb{E}\exp(X^2/t^2) \leq 2\},$$

and defines a subgaussian r.v. as a random variable such that $\|X\|_{\psi_2} < +\infty$*. The reader may check as an exercise that this definition is equivalent to the definition given here and that, for any subgaussian* X*, there exists constants* $0 < c \leq C$ *such that*

$$c\|X\|_{\mathrm{vp}} \leq \|X - \mathbb{E}X\|_{\psi_2} \leq C\|X\|_{\mathrm{vp}}.$$

Note in particular that our definition of the variance proxy automatically includes a notion of centering.

5 Reading assignment

We recommend reading Chapter 1 as well as the beginning of Chapter 2 in [7]. Our use of the variance proxy, rather than the sub-gaussian norm, seems widespread. The reader may have a look at [5, Section 3.1] and [4, Chapter 1] for similar definitions and results.

References

[1] A. Bandeira. Ten lectures and forty-two open problems in the mathematics of data science. Lecture notes, 2016.

[2] S. Boucheron, G. Lugosi, and P. Massart. *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press, 2013.

[3] V. Buldygin and Y. Kozachenko. Sub-gaussian random variables. *Ukrainian Mathematical Journal*, 32(6):483–489, 1980.

[4] P. Rigollet and J.-C. Hütter. High-dimensional statistics. Lecture notes, 2017.

[5] R. van Handel. Probability in high dimension. Lecture notes, 2016.

[6] R. Vershynin. Four lectures on probabilistic methods for data science. In *The mathematics of data*, volume 25 of *IAS/Park City Math. Ser.*, pages 231–271. Amer. Math. Soc., Providence, RI, 2018.

[7] R. Vershynin. *High-dimensional probability: An introduction with applications in data science*, volume 47 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 2018.

[8] M. J. Wainwright. *High-dimensional statistics: A non-asymptotic viewpoint*, volume 48 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 2019.