Concentration inequalities

Saverio Salzo[*]

March 28, 2021

1 Preliminaries

1.1 Basic notation and facts

We recall the concept of Legendre-Fenchel transformation. Let $\phi\colon \mathbb{R} \to]-\infty, +\infty]$ be a proper function. The *Fenchel conjugate* of ϕ is $\phi^*(t) = \sup_{\lambda\in\mathbb{R}} \lambda t - \phi(\lambda)$. The Fenchel-Moreau theorem ensures that if ϕ is proper, convex and lower semicontinuous, then ϕ^* is proper convex and lower semicontinuous and $(\phi^*)^* = \phi$. Moreover, if $\psi(\lambda) = \mu\phi(c\lambda)$, then

$$\psi^*(t) = \sup_\lambda \lambda t - \mu\phi(c\lambda) = \mu \sup_\lambda (c\lambda)\frac{t}{c\mu} - \phi(c\lambda) = \mu\phi^*\Big(\frac{t}{c\mu}\Big). \tag{1}$$

Then we continue by recalling the definition and the main properties of the *Gamma function*. $\Gamma\colon \mathbb{R}_+ \to \mathbb{R}$, $\Gamma(p) = \int_0^\infty e^{-u}u^{p-1}du$. Then we have, $\Gamma(1) = 1$ and $p\Gamma(p) = \Gamma(p+1)$, so that $\Gamma(n+1) = n\Gamma(n) = n!$. Moreover, $\Gamma(1/2) = \sqrt{\pi}$ and Γ is increasing on the interval $[3/2, +\infty[$.

We end this section by giving a formula for integration.

Lemma 1 (Layer cake representation)**.** *Let X be a positive random variable. Then*

$$\mathbb{E}[X] = \int_0^\infty \mathbb{P}(\{X > t\})dt. \tag{2}$$

[*]Istituto Italiano di Tecnologia, Via E. Melen 83, 16152 Genova. Email: `saverio.salzo@iit.it`

Proof. Note that, for every $x \in \mathbb{R}$, $x = \int_0^x dt = \int_0^\infty \chi_{[0,x[}(t)dt$. Hence, for every $\omega \in \Omega$, $X(\omega) = \int_0^\infty \chi_{[0,X(\omega)[}(t)dt$. Therefore, using Fubini's theorem, we have

$$\begin{aligned}
\mathbb{E}[X] &= \int_\Omega X(\omega)d\mathbb{P}(\omega) \\
&= \int_\Omega \int_0^\infty \chi_{[0,X(\omega)[}(t)dtd\mathbb{P}(\omega) \\
&= \int_0^\infty \int_\Omega \chi_{[0,X(\omega)[}(t)d\mathbb{P}(\omega)dt \\
&= \int_0^\infty \int_\Omega \chi_{\{X>t\}}(\omega)d\mathbb{P}(\omega)dt
\end{aligned}$$

and the statement follows. □

1.2 Orlicz spaces

We now introduce the Minkowski functional and its main properties. Let E be a real vector space and let $U \subset E$. We define

$$p_U \colon E \to [0, +\infty], \quad p_U(f) = \inf\{r > 0 \,|\, f \in rU\}, \tag{3}$$

which is called the *Minkowski functional* associated to U or the *gauge* of U.

Proposition 1. *Let E be a real vector space and let $U \subset E$ be a nonempty set satisfying the following properties*

(a) $0 \in U$ *and* $U = -U$

(b) U *is convex,*

Then p_U is a seminorm, that is, $p_U(0) = 0$, $p_U(\lambda f) = |\lambda| p_U(f)$ (homogeneity), $p_U(f + g) \leq p_U(f) + p_U(g)$(subadditivity), and

$$\{f \in E \,|\, p_U(f) < 1\} \subset U \subset \{f \in E \,|\, p_U(f) \leq 1\}. \tag{4}$$

Moreover, if in addition U is linearly bounded with respect to 0, meaning that for every $f \in U \setminus \{0\}$ the set $\{\alpha \in \mathbb{R}_+ \,|\, \alpha f \in U\}$ is bounded, then p_U is a norm on the vector space $M = \{f \in E \,|\, p_U(f) < +\infty\}$. Finally, we suppose that

(c) *for every increasing sequence* $(\alpha_k)_{k\in\mathbb{N}} \in \mathbb{R}_{++}^{\mathbb{N}}$ *such that* $\alpha_k \to 1$*, we have* $(\forall\, k \in \mathbb{N}\ \alpha_k f \in U) \ \Rightarrow\ f \in U$.

Then $U = \{f \in E \,|\, p_U(f) \leq 1\}$.

Proof. We first note that, the assumptions (a) and (b) yield that U is balanced, meaning that for every $\lambda \in \mathbb{R}$ with $|\lambda| \leq 1$, we have $\lambda U \subset U$. Indeed, let $\lambda \in \mathbb{R}$ with $|\lambda| \leq 1$. Then, since $0 \in U$, we have $|\lambda|U = (1-|\lambda|)0+|\lambda|U \subset (1-|\lambda|)U + |\lambda|U \subset U$. Moreover, since U is symmetric $\lambda U = |\lambda|(\operatorname{sign}\lambda)U = |\lambda|U \subset U$. Now let $f \in E$ and define $A(f) = \{r > 0 \,|\, f \in rU\}$. We prove that

$$s > r \text{ and } r \in A(f) \ \Rightarrow\ s \in A(f). \tag{5}$$

Indeed, $rU = s(r/s)U \subset sU$. Property (5) implies that $A(f)$ in an interval of $\mathbb{R}$ which is unbounded from above, and since $p_U(f) = \inf A(f)$, this yields that $A(f) = [p_U(f), +\infty[$ or $A(f) =]p_U(f), +\infty[$. Now, $p_U(0) = \inf A(0) = \inf \mathbb{R}_{++} = 0$. Let $f \in E$ and $\lambda \in \mathbb{R} \setminus \{0\}$. Then, since U is symmetric, $\lambda f \in rU \ \Leftrightarrow\ |\lambda| f \in rU$. Therefore, $A(\lambda f) = |\lambda|A(f)$ and hence $p_U(\lambda f) = |\lambda|p_U(f)$. Now let $f, g \in E$. We prove that $A(f) + A(g) \subset A(f+g)$. Indeed, let $r, s > 0$ be such that $f \in rU$ and $g \in sU$. Then, $f+g \in rU+sU = (r+s)[r(r+s)^{-1}U + s(r+s)^{-1}U] \subset (r+s)U$ and hence $r+s \in A(f+g)$. Therefore, $p_U(f+g) = \inf A(f+g) \leq \inf A(f)+A(g) = \inf A(f)+\inf A(g) = p_U(f)+p_U(g)$. Concerning (4), we note that by definition of $A(f)$ we have $f \in U \ \Leftrightarrow\ 1 \in A(f)$. Therefore, if $f \in U$, then $p_U(f) \leq 1$. Vice versa, if $p_U(f) < 1$, then $1 \in A(f)$ and hence $f \in U$. We now prove the second part of the statement. Suppose that U is linearly bounded. It is clear that M is a vector space. Let $f \in M \setminus \{0\}$ and let $r_0 > 0$ be such that $f/r_0 \in U$. Since U is linearly bounded we have $\alpha_0 := \sup\{\alpha \,|\, \alpha f/r_0 \in U\} < +\infty$. Then for every $\alpha > \alpha_0$ we have $f \notin (r_0/\alpha)U$ and hence for every $r > 0$ such that $f \in rU$ we have that $r_0/\alpha_0 \leq r$, and hence $p_U(f) \geq \alpha_0/r_0 > 0$. Finally, suppose that $p_U(f) = 1$. Since $A(f)$ is an interval unbounded from above and 1 is the left end point of it, there exists $(r_k)_{k\in\mathbb{N}}$ a decreasing sequence in $\mathbb{R}_{++}$ such that, for every $k \in \mathbb{N}$, $r_k \in A(f)$ and $r_k \to 1$. Then $r_k^{-1} f \in U$, $(r_k^{-1})_{k\in\mathbb{N}}$ is increasing and $r_k^{-1} \to 1$. Thus, (c) yields $f \in U$. $\square$

We now introduce Orlicz spaces. We start with the following result.

Proposition 2. *Let* $\psi\colon \mathbb{R}_+ \to [0, +\infty]$ *be a convex function such that* $\psi(0) = 0$*. Then the following hold.*

(i) *the map* $t \in \mathbb{R}_{++} \mapsto \psi(t)/t$ *is increasing.*

(ii) ψ *is increasing and* $\operatorname{dom}\psi := \{t \in \mathbb{R}_+ \mid \psi(t) < +\infty\}$ *is an interval of type* $[0, b[$ *with* $b \in \mathbb{R}_{++} \cup \{+\infty\}$ *or* $[0, b]$ *with* $b \in \mathbb{R}_+$.

(iii) $(\forall\, t \in \mathbb{R}_+)(\forall\, \alpha \in [0, 1])\ \psi(\alpha t) \leq \alpha\psi(t)$.

(iv) *If* ψ *is lower semicontinuous, then* $\psi_{|\operatorname{dom}\psi}$ *is continuous.*

(v) *If* $\psi \neq 0$*, then* $\lim_{t\to+\infty} \psi(t) = +\infty$.

Proof. (i)-(ii): Let $s, t \in \mathbb{R}_+$ with $0 \leq s < t$. Then $s = (1-s/t)0+(s/t)t$ and hence by the convexity of ψ, $\psi(s) \leq (1 - s/t)\psi(0) + (s/t)\psi(t) = (s/t)\psi(t) \leq \psi(t)$ and the first and second statements follow.

(iii): It follows by noting that $\alpha t \leq t$ and by applying (i).

(iv): We can assume that $\psi \neq 0$. Let $t > 0$ be such that $\psi(t) > 0$. Let $(\alpha_n)_{n\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}}_{++}$ be such that $\alpha_n \to 0$. Then it follows from (iii) that for n sufficiently large $\psi(\alpha_n t) \leq \alpha_n\psi(t)$ and hence $\psi(\alpha_n t) \to 0$. This shows that ψ is continuous at 0. Moreover, it is known that ψ is continuous in the interior of $\operatorname{dom}\psi$. So, we need only to analyze the case that $\operatorname{dom}\psi = [0, b]$ with $b \in \mathbb{R}_+$ and prove that ψ is continuous at b. Let $(t_n)_{n\in\mathbb{N}}$ be an increasing sequence converging to b. Since ψ is lower semicontinuous and increasing we have $\limsup_n \psi(t_n) \leq \psi(b) \leq \liminf_n \psi(t_n)$. Thus, $\psi(t_n) \to \psi(b)$. The statement follows.

(v): Since $\psi \neq 0$ and ψ is increasing, $t_0 := \sup\{t \in \mathbb{R}_+ \mid \psi(t) = 0\} < +\infty$. Let $t > t_0 + 1$. Then $\psi(t_0 + 1) > 0$ and it follows from (i) that $\psi(t)/t \geq \psi(t_0 + 1)/(t_0 + 1)$ and hence $\psi(t) \geq t\psi(t_0 + 1)/(t_0 + 1)$. The statement follows. □

A function $\psi\colon \mathbb{R}_+ \to [0, +\infty]$ which is nonzero, convex and with $\psi(0) = 0$ is called a *Young function*. Let $\psi\colon \mathbb{R}_+ \to [0, +\infty]$ be a Young function and let $(\Omega, \mathfrak{A}, \mu)$ be a σ-finite measure space and let $\mathcal{M}(\mathfrak{A})$ be the set of real-valued functions measurable with respect to $\mathfrak{A}$. Let E be the quotient space of $\mathcal{M}(\mathfrak{A})$ with respect to the relation of equality μ-a.e. The *Orlicz space* is defined as

$$L_\psi(\mu) = \Big\{ f \in E \,\Big|\, \exists\, r > 0 \int_\Omega \psi(|f|/r)d\mu < +\infty \Big\}. \tag{6}$$

Moreover, we define, for every $f \in E$,

$$\|f\|_\psi = \inf \Big\{ r > 0 \,\Big|\, \int_\Omega \psi(|f|/r)d\mu \leq 1 \Big\}. \tag{7}$$

Then we have

$$f \in L_\psi(\mu) \;\Leftrightarrow\; \exists r > 0 \text{ such that } \int_\Omega \psi(|f|/r)d\mu \le 1 \;\Leftrightarrow\; \|f\|_\psi < +\infty. \quad (8)$$

The second equivalence is immediate. We need just to prove the implication $\Rightarrow$ in the first equivalence. Suppose that $R := \int_\Omega \psi(|f|/r)d\mu < +\infty$ for some $r > 0$. Clearly, we can consider only the case that $R > 1$. Then, by Proposition 2(iii), we have $\psi(R^{-1}|f|/r) \le R^{-1}\psi(|f|/r)$ and hence $\int_\Omega \psi(|f|/(rR))d\mu \le R^{-1}\int_\Omega \psi(|f|/r)d\mu = 1$ and the statement follows. Moreover, $\|\cdot\|_\psi$ is the Minkowski functional of the set

$$U_\psi = \Big\{ f \in E \,\Big|\, \int_\Omega \psi(|f|)d\mu \le 1 \Big\}. \quad (9)$$

The following result establishes that $(L_\psi(\mu), \|\cdot\|_\psi)$ is a normed space.

Proposition 3. *Let ψ be a Young function. Then the set U_ψ satisfies the properties (a)-(b)-(c) of Proposition 1 and moreover, it is linearly bounded. Therefore, $L_\psi(\mu)$ is the vector space defined in Proposition 1 and $\|\cdot\|_\psi$ is the associated norm (the Luxemburg norm). Moreover, if ψ is lower semicontinuous, then*

$$\{f \in E \mid \|f\|_\psi \le 1\} = \Big\{ f \in E \,\Big|\, \int_\Omega \psi(|f|)d\mu \le 1 \Big\}. \quad (10)$$

Proof. Clearly (a) holds. Let $f, g \in U_\psi$ and $\alpha \in [0,1]$. Then, since ψ is increasing and convex we have

$$\begin{aligned}
\int_\Omega \psi(|(1-\alpha)f + \alpha g|)d\mu &\le \int_\Omega \psi((1-\alpha)|f| + \alpha|g|)d\mu \\
&\le \int_\Omega (1-\alpha)\psi(|f|) + \alpha\psi(|g|)d\mu \\
&= (1-\alpha)\int_\Omega \psi(|f|) + \alpha \int_\Omega \psi(|g|)d\mu \\
&\le 1.
\end{aligned}$$

Thus, (b) holds. We now prove that U_ψ is linearly bounded. Let $f \in U_\psi$ and suppose that $f \ne 0$ as equivalence class. This implies that there exists $B \in \mathfrak{A}, \mu(B) > 0$ such that $|f| > 0$ on B. Let $A_n = \{\omega \in \Omega \,|\, |f(\omega)| \ge$

$1/(n+1)\}$. Then $\omega \in B \Rightarrow |f(\omega)| > 0 \Rightarrow$ there exists $n \in \mathbb{N}$ such that $|f(\omega)| \geq 1/(n+1) \Rightarrow$ there exists $n \in \mathbb{N}$ such that $\omega \in A_n$. Thus, $B = \bigcup_{n\in\mathbb{N}}(A_n \cap B)$ and hence $0 < \mu(B) = \sup_{n\in\mathbb{N}} \mu(A_n \cap B)$, which yields that $0 < \mu(A_n \cap B)$ for some $n \in \mathbb{N}$. So, we proved that there exists $A \in \mathfrak{A}$ with $\mu(A) > 0$ and $\delta > 0$ such that $|f| \geq \delta > 0$ on A. Now, let $\alpha > 0$ and suppose that $\alpha f \in U_\psi$. Then

$$1 \geq \int_\Omega \psi(\alpha|f(\omega)|)d\mu(\omega) \geq \psi(\alpha\delta)\mu(A) \tag{11}$$

which yields, $\alpha\delta \in \psi^{-1}([1/\mu(A), +\infty[)$. Since $\lim_{t\to+\infty} \psi(t) = +\infty$, ψ has bounded sublevel sets and we have $\alpha \leq \delta^{-1} \sup \psi^{-1}([1/\mu(A), +\infty[) < +\infty$. Therefore, U_ψ is linearly bounded. Finally, let $(\alpha_k)_{k\in\mathbb{N}}$ be an increasing sequence in $\mathbb{R}_{++}$ converging to 1, let $f \in U_\psi$, and suppose that for every $n \in \mathbb{N}$, $\alpha_n f \in U_\psi$, i.e.,

$$(\forall\, n \in \mathbb{N}) \qquad \int_\Omega \psi(\alpha_n|f|)d\mu \leq 1 \tag{12}$$

This implies that $\psi(\alpha_n|f|) < +\infty$ and hence $\alpha_n|f| \in \operatorname{dom}\psi$ μ-a.e. Since ψ is increasing and continuous on $\operatorname{dom}\psi$, we have that $\psi(\alpha_k|f|)_{k\in\mathbb{N}}$ is an increasing sequence of measurable functions converging to $\psi(|f|)$. Therefore, the monotone convergence theorem ensures that $\int_\Omega \psi(\alpha_k|f|)d\mu \to \int_\Omega \psi(|f|)d\mu$ and hence $\int_\Omega \psi(|f|)d\mu \leq 1$. $\square$

1.3 Some useful Fenchel conjugate computations

Proposition 4. *The following hold.*

(i) *Let* $\phi_0 \colon \mathbb{R} \to \,]-\infty, +\infty]$ *be defined as*

$$\phi_0(x) = e^x - x - 1. \tag{13}$$

Then, the Fenchel conjugate is $\phi_0^* \colon \mathbb{R} \to \,]-\infty, +\infty]$ *with*

$$\phi_0^*(u) = \begin{cases} (u+1)\log(u+1) - t & \text{if } u > -1 \\ 1 & \text{if } u = -1 \\ +\infty & \text{otherwise.} \end{cases}$$

$$\geq \frac{u^2/2}{u/3+1}.$$

(ii) *Let* $\phi_1\colon \mathbb{R} \to \left]-\infty, +\infty\right]$ *be defined as*

$$\phi_1(x) = \begin{cases} \dfrac{x^2}{2(1-x)} & \text{if } x < 1 \\ +\infty & \text{otherwise.} \end{cases} \tag{14}$$

Then the Fenchel conjugate is $\phi_1^*\colon \mathbb{R} \to \left]-\infty, +\infty\right]$ *with*

$$\phi_1^*(u) = \begin{cases} u + 1 - \sqrt{1+2u} & \text{if } u > -1/2 \\ +\infty & \text{otherwise.} \end{cases} \tag{15}$$

(iii) *Let* $\phi_2\colon \mathbb{R} \to [0, +\infty]$ *be defined as*

$$\phi_2(x) = \begin{cases} \dfrac{x^2}{2(1-|x|)} & \text{if } |x| < 1 \\ +\infty & \text{otherwise.} \end{cases} \tag{16}$$

Then the Fenchel conjugate is $\phi_2^*\colon \mathbb{R} \to \mathbb{R}$ *with*

$$\phi_2^*(u) = |u| + 1 - \sqrt{1 + 2|u|} \geq \frac{u^2}{2(1+|u|)}. \tag{17}$$

(iv) *Let* $\phi_3\colon \mathbb{R} \to [0, +\infty]$ *be defined as*

$$\phi_3(\lambda) = \frac{\lambda^2\sigma^2}{2} + \iota_{[-a,a]}(\lambda) \tag{18}$$

Then

$$(\forall\, t \in \mathbb{R}) \quad \phi_3^*(t) = \begin{cases} \dfrac{t^2}{2\sigma^2} & \text{if } |t| \leq a\sigma^2 \\ a|t| - \dfrac{a^2\sigma^2}{2} & \text{if } |t| > a\sigma^2 \end{cases}$$
$$\geq \min\left\{\frac{t^2}{2\sigma^2}, \frac{a|t|}{2}\right\}.$$

Proof. (i): The second inequality follows from [3, Lemma 6.11].

(ii): Let $u \in \mathbb{R}$. We have

$$\phi_1^*(u) = \sup_{x<1}\left(xu - \frac{x^2}{2(1-x)}\right) = \sup_{x<1} \frac{2xu - (1+2u)x^2}{2(1-x)}.$$

Then,

$$0 \le D_x\Big[xu - \frac{x^2}{2(1-x)}\Big] = u - \frac{1}{2}\frac{2x - x^2}{(1-x)^2}$$
$$\Leftrightarrow \ 0 \le (1+2u)x^2 - 2(1+2u)x + 2u = (1+2u)(x-1)^2 - 1.$$

Hence, if $1+2u \le 0$ the condition is never satisfied and so, the function inside the sup is decreasing and $\phi_1^*(u) = +\infty$. Otherwise suppose that $1 + 2u > 0$. Then, the solution (strictly less than 1) is $x = 1 - 1/\sqrt{1+2u}$ which corresponds to a maximum and hence, since $u = (1/2)x(2-x)/(1-x)^2$, we have

$$\begin{aligned}
\phi_1^*(u) &= \frac{1}{2}\frac{x^2(2-x)}{(1-x)^2} - \frac{x^2}{2(1-x)} \\
&= \frac{x^2}{2(1-x)}\Big[\frac{2-x}{1-x} - 1\Big] \\
&= \frac{x^2}{2(1-x)^2} \\
&= \frac{1}{2}\Big(1 - \frac{1}{\sqrt{1+2u}}\Big)^2(1+2u) \\
&= \frac{1}{2}(\sqrt{1+2u} - 1)^2 \\
&= \frac{1}{2}(2 + 2u - 2\sqrt{1+2u}) \\
&= 1 + u - \sqrt{1+2u}
\end{aligned}$$

and the statement follows.

(iii): Let $u \in \mathbb{R}$. We consider two cases. We first suppose that $u \ge 0$. Then,

$$x \in \,]-1, 0] \ \Rightarrow \ xu - \frac{x^2}{2(1-|x|)} \le (-x)u - \frac{(-x)^2}{2(1-|-x|)}$$

and hence

$$\phi_2^*(u) = \sup_{x\in[0,1[} \Big(xu - \frac{x^2}{2(1-x)}\Big) = \phi_1^*(u) = 1 + u - \sqrt{1+2u}.$$

Now, suppose that $u < 0$. Then

$$x \in [0,1[\;\Rightarrow\; xu - \frac{x^2}{2(1-|x|)} \le (-x)u - \frac{(-x)^2}{2(1-|-x|)}$$

and hence

$$\phi_2^*(u) = \sup_{x \in]-1,0]} \Big(xu - \frac{x^2}{2(1+x)} \Big) = \sup_{x\in[0,1[} \Big(x(-u) - \frac{x^2}{2(1-x)} \Big) = \phi_1^*(-u)$$

and the statement follows.

(iv): Let $t \in \mathbb{R}$. We have

$$\phi_3^*(t) = \sup_{\lambda \in [-a,a]} \lambda t - \frac{\lambda^2\sigma^2}{2}. \tag{19}$$

Then computing the derivative w.r.t. λ, we have $0 \le t - \lambda\sigma^2 \;\Leftrightarrow\; \lambda \le t/\sigma^2$. Therefore, if $t/\sigma^2 \in [-a,a]$, we have $\phi_3^*(t) = t^2/(2\sigma^2)$. If $t/\sigma^2 < -a$, then $\phi_3^*(t) = \phi_3^*(-a) = -at - a^2\sigma^2/2$. Finally, if $t/\sigma^2 > a$, then $\phi_3^*(t) = \phi_3^*(a) = at - a^2\sigma^2/2$. The first equality follows. Concerning the inequality we note that $t^2/(2\sigma^2) > a|t|/2 \;\Leftrightarrow\; |t| > a\sigma^2 \;\Leftrightarrow\; a|t|/2 > a^2\sigma^2/2$ and hence, in such case, $a|t| - a^2\sigma^2/2 > a|t| - a|t|/2 = a|t|/2 = \min\{t^2/(2\sigma^2), a|t|/2\}$. the statement follows. $\square$

Proposition 5. *Let $\sigma > 0$. Then we have*

$$\frac{\lambda^2\sigma^2}{2} + \iota_{[-\frac{1}{\sigma},\frac{1}{\sigma}]}(\lambda) \le \frac{\lambda^2\sigma^2}{2(1-\sigma|\lambda|)} + \iota_{]-\frac{1}{\sigma},\frac{1}{\sigma}[}(\lambda) \le \frac{\lambda^2(2\sigma)^2}{2} + \iota_{[-\frac{1}{2\sigma},\frac{1}{2\sigma}]}(\lambda). \tag{20}$$

Proof. The first inequality follows from the fact that, if $\sigma|\lambda| < 1$, then $0 < 1 - \sigma|\lambda| \le 1$. Whereas, the second inequality follows from the fact that if $\sigma|\lambda| \le 1/2$, then $1/(1-\sigma|\lambda|) \le 2$. $\square$

2 The Cramér-Chernoff method

Definition 1. Let X be a real random variable. The *moment generating function of X* is

$$\mathbb{M}_X \colon \mathbb{R} \to [0,+\infty], \qquad \mathbb{M}_X(\lambda) = \mathbb{E}[e^{\lambda X}] = \int_{\mathbb{R}} e^{\lambda x} d\,\mathbb{P}_X(x) \tag{21}$$

Proposition 6. *The following hold.*

(i) *The function $\mathbb{M}_X$ is convex and lower semicontinuous.*

(ii) $\mathbb{M}_X(0) = 1$

(iii) *For every $\lambda \in \mathbb{R}$, $\mathbb{M}_X(\lambda) > 0$.*

Proof. (i): Let $\theta \in [0,1]$. Since $x \mapsto e^x$ is convex, we have

$$e^{((1-\theta)\lambda_1+\theta\lambda_2)X} = e^{(1-\theta)\lambda_1 X+\theta\lambda_2 X} \leq (1-\theta)e^{\lambda_1 X} + \theta e^{\lambda_2 X}$$

and hence $\mathbb{E}[e^{((1-\theta)\lambda_1+\theta\lambda_2)X}] \leq (1-\theta)\mathbb{E}[e^{\lambda_1 X}] + \theta\mathbb{E}[e^{\lambda_2 X}]$. Now, let $\lambda \in \mathbb{R}$ and $(\lambda_k)_{k\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$ be such that $\lambda_k \to \lambda$. Then clearly $e^{\lambda_k X} \to e^{\lambda X}$ pointwise. Therefore, by Fatou's lemma we have $\mathbb{E}[e^{\lambda X}] \leq \liminf_k \mathbb{E}[e^{\lambda_k X}]$ and the statement follows.

(iii): Let $\mathbb{P}_X$ be the distribution of X and let $\lambda \in \mathbb{R}$. Since

$$1 = \mathbb{P}_X(\mathbb{R}) = \mathbb{P}_X\Big(\bigcup_{n\in\mathbb{Z}}[n,n+1]\Big) \leq \sum_{n\in\mathbb{N}} \mathbb{P}_X([n,n+1]),$$

there exists $n \in \mathbb{Z}$ such that $0 < \mathbb{P}_X([n,n+1])$ and $\alpha := \min_{x\in[n,n+1]} e^{\lambda x} > 0$. Therefore, $\mathbb{M}_X(\lambda) \geq \int_{[n,n+1]} e^{\lambda x} d\,\mathbb{P}_X(x) \geq \alpha\mathbb{P}_X([n,n+1]) > 0$. □

Proposition 7. *Suppose that for some $\bar{\lambda} > 0$ $\mathbb{E}[e^{\bar{\lambda}|X|}] < +\infty$. Then, for every $n \in \mathbb{N}$, $\mathbb{E}[|X|^n] < +\infty$, $[-\bar{\lambda},\bar{\lambda}] \subset \operatorname{dom}\mathbb{M}_X$ and*

$$(\forall\,\lambda \in [-\bar{\lambda},\bar{\lambda}]) \quad \mathbb{M}_X(\lambda) = \sum_{n=0}^{+\infty} \frac{\mathbb{E}[X^n]}{n!}\lambda^n. \tag{22}$$

Hence if $0 \in \operatorname{int}(\operatorname{dom}\mathbb{M}_X)$, $\mathbb{M}_X$ is analytic (and hence infinitely times differentiable) on $]\delta,\delta[$ where $\delta = \sup\{\lambda \in \mathbb{R}_{++} \,|\, [-\lambda,\lambda] \subset \operatorname{dom}\mathbb{M}_X\}$.

Proof. Let $\lambda \in [-\bar{\lambda},\bar{\lambda}]$. Then $\mathbb{E}[e^{\lambda X}] \leq \mathbb{E}[e^{|\lambda||X|}] \leq \mathbb{E}[e^{\bar{\lambda}|X|}] < +\infty$ and

$$+\infty > \mathbb{E}[e^{|\lambda||X|}] = \mathbb{E}\bigg[\sum_{n\in\mathbb{N}} \frac{|\lambda|^n|X|^n}{n!}\bigg] = \sum_{n\in\mathbb{N}} \frac{|\lambda|^n}{n!}\mathbb{E}[|X|^n].$$

Thus, for every $n \in \mathbb{N}$, $\mathbb{E}[|X|^n] < +\infty$. Moreover, let $f_n = \sum_{k=0}^n \lambda^k X^k/k!$. Then $f_n \to e^{\lambda X}$ pointwise and $|f_n| \leq \sum_{k=0}^n |\lambda|^k|X|^k/k! \leq e^{|\lambda||X|}$, which is summable. Therefore, by Lebesgue's dominated convergence theorem we have $\mathbb{E}[f_n] \to \mathbb{E}[e^{\lambda X}]$ and hence

$$\mathbb{E}[e^{\lambda X}] = \mathbb{E}\bigg[\sum_{n\in\mathbb{N}} \frac{\lambda^n X^n}{n!}\bigg] = \sum_{n\in\mathbb{N}} \frac{\lambda^n}{n!}\mathbb{E}[X^n].$$ □

Remark 1. Let $\bar{\lambda} \in \mathbb{R}$ and let $A = \{X \geq 0\}$. Then

$$\begin{aligned}\mathbb{E}[e^{\bar{\lambda}|X|}] &= \int_A e^{\bar{\lambda}X} d\mathbb{P} + \int_{\Omega\setminus A} e^{-\bar{\lambda}X} d\mathbb{P} \\ &\leq \mathbb{E}[e^{\bar{\lambda}X}] + \mathbb{E}[e^{-\bar{\lambda}X}] \\ &\leq 2\mathbb{E}[e^{\bar{\lambda}|X|}].\end{aligned}$$

Thus, $\mathbb{E}[e^{\bar{\lambda}|X|}] < +\infty \;\Leftrightarrow\; \mathbb{E}[e^{\bar{\lambda}X}] < +\infty$ and $\mathbb{E}[e^{-\bar{\lambda}X}] < +\infty \;\Leftrightarrow\; [-\bar{\lambda}, \bar{\lambda}] \subset \operatorname{dom} \mathbb{M}_X$.

Example 1. Let X be Bernoulli random variable with parameter $p \in]0,1[$. Then $\mathbb{E}[e^{\lambda X}] = pe^{\lambda} + (1-p) = 1 + p(e^{\lambda} - 1)$. Let X be a Rademacher random variable. Then $\mathbb{E}[e^{\lambda X}] = (1/2)e^{\lambda} + (1/2)e^{-\lambda} = \cosh(\lambda)$.

Definition 2. Let X be a real-valued random variable. The *logarithm of the moment generating function of X* is

$$\psi_X \colon \mathbb{R} \to \left]-\infty, +\infty\right], \quad \psi_X(\lambda) = \log \mathbb{M}_X(\lambda) = \log \mathbb{E}[e^{\lambda X}]. \tag{23}$$

Proposition 8. *The following hold.*

(i) *The function ψ_X is convex lower semicontinuous and $\psi_X(0) = 0$.*

(ii) *For all $\lambda \in \mathbb{R}$, $\psi_X(\lambda) \geq \lambda \mathbb{E}[X]$.*

(iii) $\psi_X(-\lambda) = \psi_{-X}(\lambda)$.

(iv) *For all $\lambda \in \mathbb{R}$, $\psi_{X-\mathbb{E}[X]}(\lambda) = \psi_X(\lambda) - \lambda \mathbb{E}[X]$.*

(v) *Let $(X_i)_{1\leq i\leq N}$ be N independent random variables and let $S = \sum_{i=1}^N X_i$. Then $\psi_S(\lambda) = \sum_{i=1}^N \psi_{X_i}(\lambda)$.*

Proof. (i): Let $\lambda_1, \lambda_2 \in \mathbb{R}$ and $\theta \in [0,1]$. Then

$$\mathbb{E}[e^{((1-\theta)\lambda_1 + \theta\lambda_2)X}] = \mathbb{E}[e^{(1-\theta)\lambda_1 X + \theta\lambda_2 X}] = \mathbb{E}[UV], \tag{24}$$

where we set $U = e^{(1-\theta)\lambda_1 X}$ and $V = e^{\theta\lambda_2 X}$. Then, using Hölder's inequality, we have

$$\mathbb{E}[UV] \leq \mathbb{E}[U^{\frac{1}{1-\theta}}]^{1-\theta}\mathbb{E}[V^{\frac{1}{\theta}}]^{\theta} = \mathbb{E}[e^{\lambda_1 X}]^{1-\theta}\mathbb{E}[e^{\lambda_2 X}]^{\theta}.$$

Therefore, $\psi_X((1-\theta)\lambda_1 + \theta\lambda_2) \leq \log(\mathbb{E}[e^{\lambda_1 X}]^{1-\theta}\mathbb{E}[e^{\lambda_2 X}]^{\theta}) = (1-\theta)\psi_X(\lambda_1) + \theta\psi_X(\lambda_2)$ and the convexity follows. Now we prove that ψ_X has closed sublevel sets (which implies lower semicontinuity). Let $\alpha \in \mathbb{R}$. Then $\{\psi_X \leq \alpha\} = \{\mathbb{M}_X \leq e^{\alpha}\}$ which is closed because of Proposition 6(i).

(ii): Since $e^{\lambda \cdot}$ is convex, Jensen's inequality yields that, for all $\lambda \in \mathbb{R}$, $e^{\lambda \mathbb{E}[X]} \leq \mathbb{E}[e^{\lambda X}]$. Taking the logarithm of both sides the statement follows.

(iii): It is immediate.

(iv): $\mathbb{E}[\exp(\lambda(X - \mathbb{E}[X]))] = \mathbb{E}[e^{\lambda \mathbb{E}[X]} e^{-\lambda X}] = e^{-\lambda \mathbb{E}[X]} e^{\psi_X(\lambda)} = e^{\psi_X(\lambda) - \lambda \mathbb{E}[X]}$. Taking the logarithm, the statement follows.

(v): We have

$$\mathbb{E}\big[e^{\lambda S}\big] = \mathbb{E}\big[e^{\sum_{i=1}^n \lambda X_i}\big] = \mathbb{E}\Bigg[\prod_{i=1}^n e^{\lambda X_i}\Bigg] = \prod_{i=1}^n \mathbb{E}[e^{\lambda X_i}] \tag{25}$$

and hence, taking the logarithm of both sides, the statement follows. □

Remark 2. Since $\lambda \mapsto \log \mathbb{M}_X(\lambda)$ is convex, one says also that $\mathbb{M}_X$ is logarithmically convex.

Proposition 9 (Chernoff's inequality)**.** Let X be a centered real-valued random variable ($\mathbb{E}[X] = 0$). Then

$$(\forall\, t > 0) \quad \mathbb{P}(\{X > t\}) \leq e^{-\psi_X^*(t)}, \tag{26}$$

where ψ_X^* is the Fenchel conjugate of ψ_X.

Proof. Let $t > 0$ and $\lambda \in \mathbb{R}$. We consider two cases. Suppose that $\lambda > 0$. Then, since $z \mapsto e^{\lambda z}$ is strictly increasing, Markov's inequality yields

$$\mathbb{P}(\{X > t\}) = \mathbb{P}(\{e^{\lambda X} > e^{\lambda t}\}) \leq \frac{\mathbb{E}[e^{\lambda X}]}{e^{\lambda t}} = e^{\psi_X(\lambda) - \lambda t}. \tag{27}$$

Now, suppose that $\lambda \leq 0$. Then, since, in virtue of Proposition 8(ii), $\psi_X(\lambda) \geq 0$, we have $e^{\psi_X(\lambda) - \lambda t} \geq 1 \geq \mathbb{P}(\{X > t\})$. Thus, (27) holds for every $\lambda \in \mathbb{R}$ and hence

$$\mathbb{P}(\{X > t\}) \leq \inf_{\lambda \in \mathbb{R}} e^{\psi_X(\lambda) - \lambda t} = e^{\inf_{\lambda \in \mathbb{R}} \psi_X(\lambda) - \lambda t} = e^{-\sup_{\lambda \in \mathbb{R}} \lambda t - \psi_X(\lambda)}$$

and recalling that $\psi_X^*(t) = \sup_{\lambda \in \mathbb{R}} t\lambda - \psi_X(\lambda)$, the statement follows. □

Proposition 10. *Let X be a real-valued random variable. Then the following hold.*

(i) $\psi_X^* \geq 0$.

(ii) *For every* $t \in \mathbb{R}$, $\psi_X^*(-t) = \psi_{-X}^*(t)$.

(iii) *For every* $t \in \mathbb{R}$, $\psi_{X-\mathbb{E}[X]}^*(t) = \psi_X^*(t + \mathbb{E}[X])$.

(iv) $\psi_X^*(\mathbb{E}[X]) = 0$ *and* ψ_X^* *is increasing on the interval* $[\mathbb{E}[X], +\infty[$ *and decreasing on the interval* $]-\infty, \mathbb{E}[X]]$.

(v) *If* $t \geq \mathbb{E}[X]$, *then* $\psi_X^*(t) = \sup_{\lambda \geq 0} \lambda t - \psi_X(\lambda)$. *The right hand side of the above equality defines the so called* Cramer transform *of* ψ_X. *Therefore, the Fenchel conjugate of* ψ_X *coincides with its Cramer transform for* $t \geq \mathbb{E}[X]$.

Proof. (i): Since $\psi_X(0) = 0$, we have $\psi_X^*(t) \geq 0t - \psi_X(0) = 0$.

(iv): Let $\mu = \mathbb{E}[X]$. It follows from Proposition 8(ii) that $\psi_X^*(\mu) \leq 0$ and since $\psi_X^* \geq 0$, we have $\psi_X^*(\mu) = 0$. Let $s > t \geq \mu$. Then $t = (1-\alpha)\mu + \alpha s$ with $\alpha = (t-\mu)/(s-\mu) \in [0, 1[$ and hence $\psi_X^*(t) \leq (1-\alpha)\psi_X^*(\mu) + \alpha\psi_X^*(s) = \alpha\psi_X^*(s) \leq \psi_X^*(s)$.

(iii): It follows from the rule $\varphi(\lambda) = \psi(\lambda) - \lambda a$ implies $\varphi^*(t) = \psi^*(t + a)$.

(v): Let $t \geq \mathbb{E}[X]$. In virtue of Proposition 8(ii), we have that, for all $\lambda < 0$, $t\lambda - \psi_X(\lambda) \leq \mathbb{E}[X]\lambda - \psi_X(\lambda) \leq 0$, hence

$$\psi_X^*(t) = \sup_{\lambda \geq 0} \lambda t - \psi_X(\lambda). \qquad \square$$

Example 2. Let $X \sim \mathcal{N}(0, 1)$ and let f be its density, i.e., $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$. Then

$$\mathbb{M}_X(\lambda) = \mathbb{E}[e^{\lambda X}] = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{\lambda x} e^{-\frac{x^2}{2}} dx = e^{\frac{\lambda^2}{2}} \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-\frac{(x-\lambda)^2}{2}} dx = e^{\frac{\lambda^2}{2}}.$$

Now, suppose that $X \sim \mathcal{N}(\mu, \sigma^2)$. Then $Z := (X - \mu)/\sigma \sim \mathcal{N}(0, 1)$ and $X = \sigma Z + \mu$. Thus,

$$\mathbb{M}_X(\lambda) = \mathbb{E}[e^{\lambda\sigma Z} e^{\lambda\mu}] = e^{\lambda\mu}\psi_Z(\lambda\sigma) = e^{\lambda\mu} e^{\frac{\lambda^2\sigma^2}{2}} = e^{\frac{\lambda^2\sigma^2}{2} + \lambda\mu}. \tag{28}$$

This implies that if $X \sim \mathcal{N}(0, \sigma^2)$, we have

$$\mathbb{M}_X(\lambda) = e^{\frac{\lambda^2\sigma^2}{2}} \quad \text{and} \quad \psi_X(\lambda) = \frac{\lambda^2\sigma^2}{2}. \tag{29}$$

Therefore, $\psi_X^*(t) = t^2/(2\sigma^2)$ and Chernoff's inequality yields

$$(\forall\, t > 0) \quad \mathbb{P}(\{X > t\}) \leq \exp\Big(-\frac{t^2}{2\sigma^2}\Big). \tag{30}$$

Example 3 (Exercise 2.3.3 in [4])**.** Let X be a Poisson random variable with parameter $\mu > 0$. Its distribution is $\mathbb{P}(X = k) = e^{-\mu}\mu^k/k!$ and $\mathbb{E}[X] = \mu$. Then

$$\mathbb{E}[e^{\lambda X}] = \sum_{k=0}^{+\infty} e^{\lambda k} e^{-\mu} \frac{\mu^k}{k!} = \frac{(e^\lambda \mu)^k}{k!} = e^{e^\lambda \mu - \mu}$$

and hence $\psi_X(\lambda) = \mu(e^\lambda - 1)$. Now, we set $Z = X - \mu$. Then, using Proposition 8(iv), we have $\psi_Z(\lambda) = \psi_X(\lambda) - \lambda\mu = \mu\phi_0(\lambda)$, where $\phi_0(\lambda) = e^\lambda - \lambda - 1$ (following the notation of Proposition 4(i)). Note that, recalling formula (1), $\psi_Z^*(t) = \mu\phi_0^*(t/\mu)$ and, by Proposition 4(i),

$$\phi_0^*(t) = \begin{cases} (t+1)\log(t+1) - t & \text{if } t > -1 \\ 1 & \text{if } t = -1 \\ +\infty & \text{otherwise.} \end{cases} \tag{31}$$

Therefore, for every $t > -1$,

$$\begin{aligned} \psi_Z^*(t) &= \mu\phi_0^*(t/\mu) \\ &= \mu\big[(t/\mu + 1)\log(t/\mu + 1) - t/\mu\big] \\ &= (t+\mu)\log[(t+\mu)/\mu] - t \end{aligned} \tag{32}$$

and Chernoff's inequality yields, for every $t > 0$

$$\mathbb{P}(\{X - \mu > t\}) \leq e^{-(t+\mu)\log[(t+\mu)/\mu]+t} = \frac{e^t}{[(t+\mu)/\mu]^{t+\mu}} = e^{-\mu}\Big(\frac{e\mu}{t+\mu}\Big)^{t+\mu} \tag{33}$$

Moreover, by Proposition 4(i), we have $\phi_0^*(t) \geq \frac{t^2/2}{t/3+1}$, therefore we have also

$$\begin{aligned} \mathbb{P}(\{X - \mu > t\}) &\leq \exp(-\psi_Z^*(t)) \\ &= \exp(-\mu\phi_0^*(t/\mu)) \\ &\leq \exp\Big(-\frac{\mu}{2}\frac{t^2/\mu^2}{t/(3\mu)+1}\Big) \\ &= \exp\Big(-\frac{t^2/2}{t/3+\mu}\Big). \end{aligned}$$

Corollary 1 (Bernstein's inequality for bounded random variables)**.** *Let $(X_i)_{1\leq i\leq N}$ be an independent sequence of random variable such that $\mathbb{E}[X_i] =$*

0, $|X_i| \leq M$ $\mathbb{P}$-a.s., and $\sum_{i=1}^N \mathbb{E}[X_i^2] \leq \sigma^2$. Let $S = \sum_{i=1}^N X_i$. Then

$$(\forall\, t > 0) \quad \mathbb{P}(\{S > t\}) \leq \Big(\frac{e\sigma^2}{Mt + \sigma^2}\Big)^{t/M + \sigma^2/M^2} e^{-\sigma^2/M^2} \tag{34}$$

$$\leq \exp\Big(-\frac{t^2/2}{Mt/3 + \sigma^2}\Big) \tag{35}$$

Proof. We know that $\psi_S(\lambda) = \sum_{i=1}^N \psi_{X_i}(\lambda)$. Therefore, we look at ψ_{X_i}. Since $\log x \leq x - 1$ for $x > 0$ and $\mathbb{E}[X_i] = 0$, we have

$$\begin{aligned}
\psi_{X_i}(\lambda) &= \log \mathbb{E}[e^{\lambda X_i}] \\
&\leq \mathbb{E}[e^{\lambda X_i}] - 1 \\
&= \sum_{n=2}^{\infty} \frac{\lambda^n \mathbb{E}[X_i^n]}{n!} \\
&= \sum_{n=2}^{\infty} \frac{\lambda^n \mathbb{E}[X_i^2 X^{n-2}]}{n!} \\
&= \sum_{n=2}^{\infty} \frac{\lambda^n M^{n-2} \mathbb{E}[X_i^2]}{n!} \\
&= \frac{\mathbb{E}[X_i^2]}{M^2} \sum_{n=2}^{\infty} \frac{\lambda^n M^n}{n!} \\
&= \frac{\mathbb{E}[X_i^2]}{M^2} (e^{\lambda M} - \lambda M - 1).
\end{aligned}$$

Therefore, setting $\mu = \sigma^2/M$, we have

$$\psi_S(\lambda) \leq \frac{\sigma^2}{M^2}(e^{\lambda M} - \lambda M - 1) = \mu \frac{1}{M} \phi_0(\lambda M)$$

and hence since, by (1), $[(\mu/M)\phi_0(\cdot M)]^*(t) = (\mu/M)\phi_0^*(t/\mu)$ we have

$$-\psi_S^*(t) \leq -\frac{1}{M}\mu\phi_0^*(t/\mu) \tag{36}$$

Now, using (32), we get

$$\begin{aligned}
e^{-\psi_S^*(t)} &\le e^{-(1/M)(t+\mu)\log[(t+\mu/\mu)]+t/M} \\
&= \frac{e^{t/M}}{[(t+\mu)/\mu]^{(t+\mu)/M}} \\
&= \frac{e^{(t+\mu)/M}}{[(t+\mu)/\mu]^{(t+\mu)/M}} e^{-\mu/M} \\
&= \left(\frac{e\mu}{t+\mu}\right)^{(t+\mu)/M} e^{-\mu/M} \\
&= \left(\frac{e\sigma^2}{Mt+\sigma^2}\right)^{t/M+\sigma^2/M^2} e^{-\sigma^2/M^2}
\end{aligned}$$

and (34) follows from Chernoff's inequality (26). Concerning (35), we just use the inequality $-\phi_0^*(t) \le -\frac{1}{2}\frac{t^2}{t/3+1}$ and similarly to what we get in Example 3, we have

$$-\psi_S^*(t) \le -\frac{\mu}{M}\phi_0^*(t/\mu) \le -\frac{t^2/2}{Mt/3+M\mu} = -\frac{t^2/2}{Mt/3+\sigma^2}$$

and (35) follows again from Chernoff's inequality (26). □

Corollary 2 (Theorem 2.3.1 in [4]). *Let $(X_i)_{1\le i\le N}$ be a sequence of independent Bernoulli random variables with parameters $(p_i)_{1\le i\le N}$ and let $S = \sum_{i=1}^N X_i$ and $\mu = \mathbb{E}[S]$. Then for every $t > 0$*

$$\mathbb{P}(\{S-\mu > t\}) \le \left(\frac{e\mu}{t+\mu}\right)^{t+\mu} e^{-\mu}. \tag{37}$$

Proof. Let, for every $i = 1,\dots,N$, $Z_i = X_i - p_i$. Clearly $\mathbb{E}[X_i] = p_i$ and $\mathbb{E}[Z_i^2] = \mathbb{V}\mathrm{ar}(X_i) = \mathbb{E}[|X_i - p_i|^2] = p_i - p_i^2 \le p_i$, so that $\sum_{i=1}^N \mathbb{E}[Z_i^2] \le \sum_{i=1}^N p_i = \mathbb{E}[S] = \mu$. Moreover, $|Z_i| \le \max\{1-p_i, p_i\} \le 1$. Therefore, the statement follows from (34). □

3 Sub-Gaussian random variables

Theorem 1. Let X be a real-valued random variable, let $K_i > 0$, $i = 1, \dots, 6$, and consider the following statements.

(a) For every $t > 0$, $\mathbb{P}(\{|X| \geq t\}) \leq 2\exp\left(-\dfrac{t^2}{2K_1^2}\right)$

(b) For every $p \geq 1$, $\mathbb{E}[|X|^p] \leq K_2^p\, p\Gamma(p/2)$.

(c) For every integer $n \geq 1$, $\mathbb{E}[|X|^{2n}] \leq K_3^{2n} n!$

(d) $\mathbb{E}[\exp(X^2/K_4^2) \leq 2$.

Then the following hold.

(i) (a) $\Rightarrow$ (b) with $K_2 = \sqrt{2}K_1$.

(ii) (b) $\Rightarrow$ (c) with $K_3 = \sqrt{2}K_2$.

(iii) (c) $\Rightarrow$ (d) with $K_4 = \sqrt{2}K_3$.

(iv) (d) $\Rightarrow$ (a) (c) with $K_1 = K_4/\sqrt{2}$ and $K_3 = K_4$ respectively.

Moreover, suppose that $\mathbb{E}[X] = 0$ and consider the statements

(e) For every $\lambda \in \mathbb{R}$, $\psi_X(\lambda) \leq \dfrac{\lambda^2 K_5^2}{2}$.

(f) For every $t > 0$, $\max\{\mathbb{P}(\{X \geq t\}), \mathbb{P}(\{X \leq -t\})\} \leq \exp\left(-\dfrac{t^2}{2K_6^2}\right)$

Then the following hold.

(v) (e) $\Rightarrow$ (f) with $K_6 = K_5$.

(vi) (f) $\Rightarrow$ (a) with $K_1 = K_6$.

(vii) (b) $\Rightarrow$ (e) with $K_5^2 = 3K_2^2$.

(viii) (c) $\Rightarrow$ (e) with $K_5 = 2K_3$.

Proof. (i): Suppose that (a) holds and let $p \geq 1$. Then

$$
\begin{aligned}
\mathbb{E}[|X|^p] &= \int_0^{+\infty} \mathbb{P}(\{|X|^p > t\})dt \\
&= \int_0^{+\infty} \mathbb{P}(\{|X| > t^{1/p}\})dt \\
&\leq 2\int_0^{+\infty} \exp\Big(-\frac{t^{2/p}}{2K_1^2}\Big)dt.
\end{aligned} \tag{38}
$$

Now, we have $u = t^{2/p}/(2K_1^2) \Rightarrow t = (\sqrt{2}K_1)^p u^{p/2} \Rightarrow dt = (\sqrt{2}K_1)^p (p/2)u^{p/2-1}du$, and hence

$$
\mathbb{E}[|X|^p] \leq (\sqrt{2}K_1)^p\, p \int_0^{+\infty} e^{-u}u^{p/2-1}du
$$

and the statement follows.

(ii): Suppose that (b) holds and let n be an integer $n \geq 1$. Then

$$
\mathbb{E}[|X|^{2n}] \leq K_2^{2n}\, 2n\Gamma(n) \leq (2K_2^2)^n n! \tag{39}
$$

(iii): Suppose that (c) holds. Then

$$
\mathbb{E}\bigg[\exp\Big(\frac{X^2}{2K_3^2}\Big)\bigg] = \mathbb{E}\bigg[\sum_{n=0}^{+\infty} \frac{X^{2n}}{2^n K_3^{2n}\, n!}\bigg] = \sum_{n=0}^{+\infty} \frac{\mathbb{E}[X^{2n}]}{2^n K_3^{2n}\, n!} \leq \sum_{n=0}^{+\infty} \frac{1}{2^n} = 2.
$$

(iv): Suppose that (d) holds. Then, using Markov's inequality,

$$
\begin{aligned}
\mathbb{P}(\{|X| \geq t\}) &= \mathbb{P}(\{\exp(X^2/K_4^2) \geq \exp(t^2/K_4^2)\}) \\
&\leq \mathbb{E}[\exp(X^2/K_4^2)]\exp(-t^2/K_4^2) \\
&\leq 2\exp(-t^2/K_4^2),
\end{aligned}
$$

so that (a) holds with $K_1^2 = K_4^2/2$. The second part of the statement follows by the fact that

$$
1 + \sum_{n=1}^{+\infty} \frac{\mathbb{E}[X^{2n}]}{K_4^{2n} n!} = \mathbb{E}[\exp(X^2/K_4^2)] \leq 2, \tag{40}
$$

which implies that each term of the series is less or equal than 1 and hence that $\mathbb{E}[X^{2n}] \leq K_4^{2n} n!$.

(v): Suppose that $\mathbb{E}[X]=0$ and (e) holds. Since $\psi \leq \varphi \Rightarrow \varphi^* \leq \psi^*$, we have

$$(\forall\, t \in \mathbb{R}) \quad \frac{t^2}{2K_5^2} \leq \psi_X^*(t). \tag{41}$$

Therefore, by Cernoff's inequality (26), we have

$$(\forall\, t > 0) \quad \mathbb{P}(\{X \geq t\}) \leq \exp(-\psi_X^*(t)) \leq \exp\Big(-\frac{t^2}{2K_5^2}\Big). \tag{42}$$

Now, since $\psi_X(-\lambda) = \log \mathbb{E}[e^{-\lambda X}] = \psi_{-X}(\lambda)$, we have

$$(\forall\, t \in \mathbb{R}) \quad \psi_{-X}^*(t) = \sup_{\lambda \in \mathbb{R}} \lambda t - \psi_X(-\lambda) = \sup_{\lambda \in \mathbb{R}} -\lambda t - \psi_X(\lambda) = \psi_X^*(-t)$$

and hence, for every $t > 0$, $\mathbb{P}(\{X < -t\}) = \mathbb{P}(\{-X > t\}) \leq \exp(-\psi_{-X}^*(t)) = \exp(-\psi_X^*(-t))$. Then, recalling (41), we have

$$\mathbb{P}(\{X < -t\}) \leq \exp(-\psi_X^*(-t)) \leq \exp\Big(-\frac{t^2}{2K_5^2}\Big).$$

(vi): We have

$$\mathbb{P}(\{|X| > t\}) \leq \mathbb{P}(\{X > t\}) + \mathbb{P}(\{X < -t\}) \leq 2\exp\Big(-\frac{t^2}{2K_6^2}\Big). \tag{43}$$

(vii): Suppose that (b) holds. Then, for every $n \in \mathbb{N}$, $\mathbb{E}[|X|^n] \leq K_2^n\, n\Gamma(n/2)$. Moreover,

$$\mathbb{E}[e^{\lambda X}] \leq \mathbb{E}[e^{|\lambda||X|}] \leq 1 + |\lambda|\mathbb{E}[|X|] + \sum_{n=2}^{+\infty} \frac{|\lambda|^n \mathbb{E}[|X|^n]}{n!}. \tag{44}$$

Then, using the property of the Gamma function and the fact that $n!/(2n!) \leq 1/2n!$, we have

$$\begin{aligned}\sum_{n=2}^{+\infty} \frac{|\lambda|^n \mathbb{E}[|X|^n]}{n!} &\leq \sum_{n=2}^{+\infty} \frac{(K_2^2\lambda^2)^{n/2} n\Gamma(n/2)}{n!} \\ &= \sum_{n=1}^{+\infty} \frac{(K_2^2\lambda^2)^n 2n\Gamma(n)}{(2n)!} + \sum_{n=1}^{+\infty} \frac{(K_2^2\lambda^2)^{n+1/2}(2n+1)\Gamma(n+1/2)}{(2n+1)!}\end{aligned}$$

$$
\begin{aligned}
&\le \sum_{n=1}^{+\infty} \frac{(K_2^2\lambda^2)^n 2n\Gamma(n)}{(2n)!} + K_2|\lambda| \sum_{n=1}^{+\infty} \frac{(K_2^2\lambda^2)^n \Gamma(n+1)}{(2n)!} \\
&= \sum_{n=1}^{+\infty} \frac{(K_2^2\lambda^2)^n 2n!}{(2n)!} + K_2|\lambda| \sum_{n=1}^{+\infty} \frac{(K_2^2\lambda^2)^n n!}{(2n)!} \\
&= (2 + K_2|\lambda|) \sum_{n=1}^{+\infty} \frac{(K_2^2\lambda^2)^n n!}{(2n)!} \\
&\le (1 + K_2|\lambda|/2) \sum_{n=1}^{+\infty} \frac{(K_2^2\lambda^2)^n}{n!} \\
&= \big(1 + K_2|\lambda|/2\big)(\exp(K_2^2\lambda^2) - 1) \\
&= \exp(K_2^2|\lambda|^2) - 1 + \sqrt{\frac{K_2^2\lambda^2}{4}}(\exp(K_2^2\lambda^2) - 1).
\end{aligned}
$$

Therefore, $\mathbb{E}[e^{|\lambda||X|}] < +\infty$ and, since $\mathbb{E}[X] = 0$,

$$
\begin{aligned}
\mathbb{E}[e^{\lambda X}] &= 1 + \sum_{n=2}^{+\infty} \frac{\lambda^n \mathbb{E}[X^n]}{n!} \\
&\le 1 + \sum_{n=2}^{+\infty} \frac{|\lambda|^n \mathbb{E}[|X|^n]}{n!} \\
&\le \exp(K_2^2\lambda^2) + \sqrt{\frac{K_2^2\lambda^2}{4}}(\exp(K_2^2\lambda^2) - 1).
\end{aligned}
$$

Now, since $1 - e^{-t} \le \sqrt{t}$,[1] we have $e^t + \sqrt{t/4}(e^t - 1) = e^t(1 + \sqrt{t/4}(1 - e^{-t})) \le e^t(1+t/2) \le e^t e^{t/2} = e^{3t/2}$ and hence $\mathbb{E}[e^{\lambda X}] \le \exp(3K_2^2\lambda^2/2)$. Thus, (e) holds with $K_5^2 = 3K_2^2$.

(viii): Suppose that (c) holds. We first note that for every $n \in \mathbb{N}$, $\mathbb{E}[|X|^{2n+1}] < +\infty$. This follows from the fact that $L^p(\Omega, \mathbb{P}) \subset L^q(\Omega, \mathbb{P})$ if $p \ge q$. Let X' be an independent copy of X. Then, since $Y := X - X'$ is symmetric (meaning that Y and $-Y$ have the same distribution), we have

$$
(\forall\, n \in \mathbb{N}) \quad \mathbb{E}[Y^{2n+1}] = \mathbb{E}[(-Y)^{2n+1}] = -\mathbb{E}[Y^{2n+1}]
$$

[1] Let $f(t) = 1 - e^{-t} - \sqrt{t}$. Then $f(0) = 0$ and $f'(t) = e^{-t} - 1/2\sqrt{t} \le 0$, since $e^t \ge 1 + t \ge 2\sqrt{t}$.

and hence $\mathbb{E}[Y^{2n+1}] = 0$. Moreover, for every integer $n \geq 1$, since $x \mapsto x^{2n}$ is convex, we have

$$\mathbb{E}[Y^{2n}] = \mathbb{E}\Big[2^{2n}\Big(\frac{X - X'}{2}\Big)^{2n}\Big] \leq 2^{2n-1}\big(\mathbb{E}[X^{2n}] + \mathbb{E}[(X')^{2n}]\big) = 2^{2n}\mathbb{E}[X^{2n}].$$

Then, since $-X'$ is centered, we have $1 \leq \mathbb{E}[e^{-\lambda X'}]$ and hence, using the independence between X and $-X'$, we have

$$\begin{aligned}
\mathbb{E}[e^{\lambda X}] &\leq \mathbb{E}[e^{\lambda X}]\mathbb{E}[e^{-\lambda X'}] \\
&= \mathbb{E}[e^{\lambda(X-X')}] \\
&= \sum_{n=1}^{+\infty} \frac{\lambda^{2n}\mathbb{E}[Y^{2n}]}{(2n)!} \\
&\leq \sum_{n=1}^{+\infty} \frac{\lambda^{2n}2^{2n}\mathbb{E}[X^{2n}]}{(2n)!} \\
&\leq \sum_{n=1}^{+\infty} \frac{\lambda^{2n}2^{2n}(K_3^2)^n n!}{(2n)!}.
\end{aligned}$$

Now, since

$$\frac{(2n)!}{n!} = \prod_{j=1}^{n}(n+j) \geq \prod_{j=1}^{n}(2j) = 2^n n!, \tag{45}$$

we have

$$\mathbb{E}[e^{\lambda X}] \leq \sum_{n=1}^{+\infty} \frac{\lambda^{2n}2^{2n}(K_3^2)^n}{2^n n!} \leq \sum_{n=0}^{+\infty} \frac{(2K_3^2\lambda^2)^n}{n!} = e^{2K_3^2\lambda^2}.$$

The statement follows. □

Remark 3. Suppose that (b) holds. Then using the Stirling's approximation $\Gamma(x) \leq x^x$ (which holds for $x \geq 1$), we have, for $p \geq 2$, $\|X\|_p^p \leq K_2^p p\Gamma(p/2) \leq K_2^p p(p/2)^{p/2}$ and hence $\|X\|_p \leq K_2 p^{1/p}\sqrt{p/2}$. Since $p^{1/p} = \exp((\log p)/p) \leq \exp(1/e)$, we have

$$(\forall\, p \geq 2) \quad \|X\|_p \leq \frac{e^{1/e}}{\sqrt{2}} K_2\sqrt{p}. \tag{46}$$

Note also that $\|X\|_1 \leq K_2\Gamma(1/2) = \sqrt{\pi}K_2$.

Remark 4 (Excercise 2.5.4 in [4]). Suppose that X is a real-valued random variable and that condition (e) of Theorem 1 holds. Then necessarily $\mathbb{E}[X] = 0$. Indeed, it follows from Proposition 8(ii) that, for every $\lambda \in \mathbb{R}$, $\lambda\mathbb{E}[X] \leq \psi_X(\lambda) \leq \lambda^2 K_5^2/2$ and hence

$$(\forall\, \lambda \in \mathbb{R}) \quad \lambda^2 K_5^2/2 - \mathbb{E}[X]\lambda \geq 0.$$

Since the discriminant of the quadratic polynomial in λ is $\mathbb{E}[X]^2$ and the inequality holds for every $\lambda \in \mathbb{R}$, then necessarily $\mathbb{E}[X] = 0$.

Definition 3 ([1, 2]). Let X be a centered real-valued random variable and let $\sigma^2 \geq 0$. Then X is *sub-Gaussian with variance factor (or proxy)* σ^2 if

$$(\forall\, \lambda \in \mathbb{R}) \quad \psi_X(\lambda) \leq \frac{\lambda^2\sigma^2}{2} \tag{47}$$

The class of centered sub-Gaussian variables with variance factor σ^2 is denoted by $\mathsf{subG}(\sigma^2)$.

Remark 5. Suppose that X is centered and $X \sim \mathsf{subG}(\sigma^2)$. Then it follows from Theorem 1 that the following hold.

(i) $(\forall\, t \in \mathbb{R}_{++}) \quad \max\{\mathbb{P}(\{X < -t\}), \mathbb{P}(\{X > t\})\} \leq \exp\Big(-\dfrac{t^2}{2\sigma^2}\Big)$.

(ii) $(\forall\, p \geq 2) \quad \|X\|_p \leq e^{1/e}\sigma\sqrt{p}$ and $\|X\|_1 \leq \sqrt{2\pi}\sigma$.

(iii) $(\forall\, n \in \mathbb{N}) \quad \mathbb{E}[|X|^{2n}] \leq 2^{n+1}\sigma^{2n} n!$.

(iv) $\mathbb{E}\Big[\exp\Big(\dfrac{X^2}{8\sigma^2}\Big)\Big] \leq 2$.

Proposition 11 (Linear combination of independent sub-Gaussian random variables). *Let $(X_i)_{1\leq i\leq n}$ be N independent zero-mean random variables such that $X_i \sim \mathsf{subG}(\sigma^2)$, let $(a_i)_{1\leq i\leq N} \in \mathbb{R}^N$, and set $S = \sum_{i=1}^N a_i X_i$. Then $S \sim \mathsf{subG}(\sigma^2 \|a\|_2^2)$.*

Proof. We have $\psi_S(\lambda) = \sum_{i=1}^n \psi_{X_i}(\lambda a_i) \leq \sum_{i=1}^n \lambda^2\sigma^2 a_i^2/2 = \lambda^2\sigma^2 \|a\|_2^2/2$. □

Lemma 2 (Hoeffding's Lemma). Let X be a random variable with $\mathbb{E}[X] = 0$, taking values in a bounded interval $[a, b]$. Then $X \in \mathsf{subG}((b-a)^2/4)$.

Proof. Since $a \le X \le b$, we have $a \le \mathbb{E}[X] \le b$ and hence $a \le 0 \le b$. Let $x \in [a, b]$. Then x can be written as a convex combination of a and b, that is, $x = (1-\alpha)a + \alpha b = a + \alpha(b-a)$, with $\alpha \in [0,1]$. Therefore, necessarily $\alpha = (x-a)/(b-a)$ and hence

$$x = \Big(1 - \frac{x-a}{b-a}\Big)a + \frac{x-a}{b-a}. \tag{48}$$

Now, let $\lambda \in \mathbb{R}$. Since $e^{\lambda \cdot}$ is convex we have

$$e^{\lambda x} \le \Big(1 - \frac{x-a}{b-a}\Big)e^{\lambda a} + \frac{x-a}{b-a}e^{\lambda b} \tag{49}$$

and hence

$$\begin{aligned} \mathbb{E}[e^{\lambda X}] &\le \Big(1 - \frac{-a}{b-a}\Big)e^{\lambda a} + \frac{-a}{b-a}e^{\lambda b} \\ &= e^{\lambda a}\big(1 - p + pe^{\lambda(b-a)}\big), \end{aligned}$$

where $p := -a/(b-a)$. Note that, since $a \le 0 \le b$, we have $0 \le p \le 1$. Then, since $a = -p(b-a)$,

$$\begin{aligned} \psi_X(\lambda) &= \log \mathbb{E}[e^{\lambda X}] \\ &\le \lambda a + \log(1 - p + pe^{\lambda(b-a)}) \\ &= -\lambda p(b-a) + \log(1 - p + pe^{\lambda(b-a)}) \\ &= \phi_p(\lambda(b-a)), \end{aligned}$$

where $\phi_p(s) = \log(1 - p + pe^s) - ps$. We now compute the derivatives of ϕ_p. We have

$$\phi_p'(s) = \frac{pe^s}{1 - p + pe^s} - p,$$

and

$$\phi_p''(s) = \frac{pe^s(1-p+pe^s) - (pe^s)^2}{(1-p+pe^s)^2} = \frac{pe^s(1-p)}{(1-p+pe^s)^2} \le \frac{pe^s(1-p)}{4pe^s(1-p)} \le \frac{1}{4},$$

where we used that $(x+y)^2 \ge 4xy$. This shows that ϕ_p is convex and with Lipschitz continuous derivative with constant $1/4$. Therefore, the descent lemma yields

$$\phi_p(s) \le \phi_p(0) + (s-0)\phi_p'(0) + \frac{1}{8}(s-0)^2 = \frac{s^2}{8}, \tag{50}$$

where we used that $\phi_p(0) = \phi_p'(0) = 0$. In the end, $\psi_X(\lambda) \le \phi_p(\lambda(b-a)) \le \lambda^2(b-a)^2/8$ and the statement follows. $\square$

Another proof. Since X is bounded, clearly $\mathbb{M}_X$ and hence ψ_X is everywhere finite and infinitely times differentiable (actually, since $\operatorname{dom}\mathbb{M}_X = \mathbb{R}$, by Proposition 7 is it analytic). Moreover, we have, for every $\lambda \in \mathbb{R}$,

$$\psi_X'(\lambda) = \frac{\mathbb{E}[Xe^{\lambda X}]}{\mathbb{E}[e^{\lambda X}]} \quad \text{and} \quad \psi_X''(\lambda) = \frac{\mathbb{E}[X^2e^{\lambda X}]}{\mathbb{E}[e^{\lambda X}]} - \left(\frac{\mathbb{E}[Xe^{\lambda X}]}{\mathbb{E}[e^{\lambda X}]}\right)^2. \tag{51}$$

Now, let $\lambda \in \mathbb{R}$ and define the probability measure

$$\mathbb{Q}_\lambda = \frac{e^{\lambda X}}{\mathbb{E}[e^{\lambda X}]} d\mathbb{P}. \tag{52}$$

We then consider the random variable X under the probability space $(\Omega, \mathfrak{A}, \mathbb{Q}_\lambda)$. We have

$$\mathbb{E}_{\mathbb{Q}_\lambda}[X] = \int_\Omega X d\mathbb{Q}_\lambda = \frac{1}{\mathbb{E}[e^{\lambda X}]} \int_\Omega X e^{\lambda X} d\mathbb{P} = \psi_X'(\lambda),$$

$$\mathbb{E}_{\mathbb{Q}_\lambda}[X^2] = \int_\Omega X^2 d\mathbb{Q}_\lambda = \frac{1}{\mathbb{E}[e^{\lambda X}]} \int_\Omega X^2 e^{\lambda X} d\mathbb{P} = \frac{\mathbb{E}[X^2e^{\lambda X}]}{\mathbb{E}[e^{\lambda X}]},$$

and

$$\mathbb{V}\mathrm{ar}_{\mathbb{Q}_\lambda}[X] = \mathbb{E}_{\mathbb{Q}_\lambda}[X^2] - \mathbb{E}_{\mathbb{Q}_\lambda}[X]^2 = \psi_X''(\lambda)$$

Since $X \in [a,b]$ $\mathbb{P}$-a.s., setting $A = \{X \in [a,b]\}$, we have $\mathbb{P}(A) = 1$. Moreover, $\mathbb{Q}_\lambda(A) = \mathbb{E}[e^{\lambda X}]^{-1} \int_A e^{\lambda X} d\mathbb{P} = 1$. Therefore, $X \in [a,b]$ $\mathbb{Q}_\lambda$-a.s. and hence

$$\left|X - \frac{b+a}{2}\right| \leq \frac{b-a}{2} \quad \mathbb{Q}_\lambda\text{-a.s.}$$

Then

$$\psi_X''(\lambda) = \mathbb{V}\mathrm{ar}_{\mathbb{Q}_\lambda}[X] = \mathbb{V}\mathrm{ar}_{\mathbb{Q}_\lambda}\left[X - \frac{a+b}{2}\right] \leq \mathbb{E}_{\mathbb{Q}_\lambda}\left[\left|X - \frac{a+b}{2}\right|^2\right] \leq \frac{(b-a)^2}{4}.$$

Finally note that $\psi_X(0) = \psi_X'(0) = 0$ and using Taylor's formula, there exists θ (between 0 and λ) such that $\psi_X(\lambda) = \psi_X(0) + \lambda\psi_X'(0) + (\lambda^2/2)\psi_X''(\theta) \leq \lambda^2(b-a)^2/8$. □

Corollary 3. *Let X be a random variable with $\mathbb{E}[X] = 0$ and $|X| \leq M$. Then $X \in \mathsf{subG}(M^2)$. If X is a* Rademacher random variable *(also called* symmetric Bernoulli random variable*), then $M = 1$ and hence $X \in \mathsf{subG}(1)$.*

Theorem 2 (Hoeffding's inequality)**.** *Let $(X_i)_{1\leq i\leq N}$ be N independent random variables such that almost surely $X_i \in [a_i, b_i]$. Set $S = \sum_{i=1}^N X_i$. Then*

$$S - \mathbb{E}[S] \sim \mathsf{subG}\Big(\frac{\sum_{i=1}^N (b_i - a_i)^2}{4}\Big) \tag{53}$$

and hence, for every $t > 0$,

$$\mathbb{P}(\{S - \mathbb{E}[S] > t\}) \leq \exp\bigg(-\frac{2t^2}{\sum_{i=1}^n (b_i - a_i)^2}\bigg). \tag{54}$$

Proof. Let $Z = S - \mathbb{E}[S]$ and $Z_i = X_i - \mathbb{E}[X_i]$. Then $Z = \sum_{i=1}^N Z_i$. Since $Z_i \in [a_i - \mathbb{E}[X_i], b_i - \mathbb{E}[X_i]]$, it follows from Proposition 8(v) and Lemma 2 that $\psi_Z(\lambda) = \sum_{i=1}^n \psi_{Z_i}(\lambda) \leq \lambda^2 \sum_{i=1}^n (b_i - a_i)^2/8$ and the first part of the statement follow. The second part follows from (42). $\square$

Theorem 3 (General Hoeffding's inequality)**.** *Let $(X_i)_{1\leq i\leq N}$ be N independent mean-zero sub-Gaussian random variables with $X_i \sim \mathsf{subG}(\sigma_i)$. Then*

$$\mathbb{P}\bigg(\Big\{\Big|\sum_{i=1}^N X_i\Big| \geq t\Big\}\bigg) \leq 2\exp\bigg(-\frac{t^2}{\sum_{i=1}^N \sigma_i^2}\bigg). \tag{55}$$

Proof. Let $S = \sum_{i=1}^N X_i$. Then, for every $\lambda \in \mathbb{R}$, $\psi_S(\lambda) = \sum_{i=1}^N \psi_{X_i}(\lambda) \leq \sum_{i=1}^N \lambda^2\sigma_i^2/2 = (\lambda^2/2)\sum_{i=1}^N \sigma_i^2$. The statement follows from Theorem 1. $\square$

Example 4 (A maximal inequality)**.** Let $(X_i)_{1\leq i\leq N}$ be real-valued random variable such that $X_i \in \mathsf{subG}(\sigma^2)$, for every $i = 1, \dots, N$. Then

$$\mathbb{E}[\max_{1\leq i\leq N} X_i] \leq \sqrt{2\sigma^2 N}. \tag{56}$$

Indeed let $X = \max_{1\leq i\leq N} X_i$. Then we have

$$\mathbb{E}[e^{\lambda X}] = \mathbb{E}[\max_{1\leq i\leq N} e^{\lambda X_i}] \leq \sum_{i=1}^N \mathbb{E}[e^{\lambda X_i}] \leq N\exp\Big(\frac{\lambda^2\sigma^2}{2}\Big) \tag{57}$$

and hence $\psi_X(\lambda) \leq \log N + \lambda^2\sigma^2/2$. Now, by Proposition 8(ii) we have

$$\mathbb{E}[X] \leq \inf_{\lambda>0} \frac{\psi_X(\lambda)}{\lambda} = \inf_{\lambda>0}\Big(\frac{N}{\lambda} + \frac{\lambda\sigma^2}{2}\Big) = \sqrt{2\sigma^2 N}. \tag{58}$$

Remark 6 (Sub-Gaussian class as an Orlicz space)**.** Recalling definition (6) of Orlicz spaces, one recognizes that if we define $\psi_2(t) = e^{t^2} - 1$ and consider the probability space $(\Omega, \mathfrak{A}, \mathbb{P})$, the corresponding Orlicz space is

$$L_{\psi_2}(\mathbb{P}) = \Big\{ X \colon \Omega \to \mathbb{R} \text{ r.v. } \mid \exists r > 0 \text{ such that } \mathbb{E}[\exp(X^2/r^2)] \leq 2 \Big\} \tag{59}$$

and $\|X\|_{\psi_2} = \inf\{r > 0 \mid \mathbb{E}[\exp(X^2/r^2)] \leq 2\}$. Therefore, $L_{\psi_2}(\mathbb{P})$ is the space of all random variables satisfying the condition (d) of Theorem 1 and sometimes such space is taken by definition as the class of sub-Gaussian random variables (which are not necessarily zero-mean). See [4]. Therefore, according to that definition the class of all sub-Gaussian random variables can be seen as an Orlicz space and the smallest constant K_4 satisfying Theorem 1(d) is the associated Luxemburg norm of X. We note that Theorem 1 yields also

(i) $\|X\|_{L^p} \leq \dfrac{e^{1/e}}{\sqrt{2}} \sqrt{p} \, \|X\|_{\psi_2}$ for all $p \geq 2$ and $\|X\|_1 \leq \sqrt{\pi} \, \|X\|_{\psi_2}$.

(ii) $\mathbb{P}(\{|X| > t\}) \leq 2 \exp(-t^2 / \|X\|_{\psi_2}^2)$

(iii) $\mathbb{E}[|X|^{2n}] \leq \|X\|_{\psi}^{2n} \, n!$

(iv) If $\mathbb{E}[X] = 0$, then $\psi_X(\lambda) \leq \dfrac{3}{2} \|X\|_{\psi_2}^2 \lambda^2$.

Example 5. Let $X \sim N(0,1)$. We compute the moment generating function of X^2. We have, for every $\lambda \in \mathbb{R}$,

$$\mathbb{E}[e^{\lambda X^2}] = \int_{\mathbb{R}} e^{\lambda x^2} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-\frac{x^2}{2}(1-2\lambda)} dx. \tag{60}$$

Therefore, if $\lambda \geq 1/2$, ww have $e^{-(x^2/2)(1-2\lambda)} \geq e^0 = 1$, and hence $\mathbb{E}[e^{\lambda X^2}] = +\infty$. Otherwise, suppose that $\lambda < 1/2$ and set $1/\sigma^2 = 1 - 2\lambda$. Then

$$\mathbb{E}[e^{\lambda X^2}] = \frac{\sigma}{\sqrt{2\pi}\sigma} \int_{\mathbb{R}} e^{-\frac{x^2}{2\sigma^2}} dx = \sigma = \frac{1}{\sqrt{1-2\lambda}}. \tag{61}$$

Therefore, for the logarithm of the moment generating function of X^2 we have

$$\psi_{X^2}(\lambda) = \begin{cases} -\frac{1}{2}\log(1-2\lambda) & \text{if } \lambda < 1/2 \\ +\infty & \text{otherwise.} \end{cases} \tag{62}$$

As a consequence suppose that Y is χ^2 with N degrees of freedom, meaning that $Y = \sum_{i=1}^{N} X_i^2$, where $(X_i)_{1\le i\le N}$ is a finite sequence of independent normal random variables. Then

$$\psi_Y(\lambda) = \sum_{i=1}^{N} \psi_{X_i^2}(\lambda) = \begin{cases} -\frac{N}{2}\log(1-2\lambda) & \text{if } \lambda < 1/2 \\ +\infty & \text{otherwise.} \end{cases} \tag{63}$$

Example 6 (Example 2.5.8 in [4])**.**

(i) Let $X \sim N(0,\sigma^2)$ and let $K_4 > 0$. Then, $Z := X/\sigma \sim N(0,1)$ and it follows from Example 5 that

$$\mathbb{E}\Big[\exp\Big(\frac{X^2}{K_4^2}\Big)\Big] = \mathbb{E}\Big[\exp\Big(\frac{\sigma^2}{K_4^2}Z^2\Big)\Big] = \frac{1}{\sqrt{1-2\sigma^2/K_4^2}} \tag{64}$$

Then, $\mathbb{E}[\exp(X^2/K_4^2)] \le 2 \ \Leftrightarrow\ 1 - 2\sigma^2/K_4^2 \ge 1/4 \ \Leftrightarrow\ 3/8 \ge \sigma^2/K_4^2 \ \Leftrightarrow\ K_4^2 \ge (8/3)\sigma^2$. Hence $\|X\|_{\psi_2} = \sqrt{8/3}\sigma$.

(ii) If X is a Rademacher random variable, then $\|X\|_{\psi_2} = 1/\sqrt{\log 2}$. Indeed $\mathbb{E}[\exp(X^2/K_4^2)] = (1/2)\exp(1/K_4^2) + (1/2)\exp(1/K_4^2) = \exp(1/K_4^2)$ and hence $\mathbb{E}[\exp(X^2/K_4^2)] \le 2 \ \Leftrightarrow\ 1/K_4^2 \le \log 2 \ \Leftrightarrow\ 1/\sqrt{\log 2} \le K_4$.

(iii) If X is a constant real-valued random variable, say $X = a, \mathbb{P}$-a.s. Then $\|X\|_{\psi_2} = |a|/\sqrt{\log 2}$. Indeed $\mathbb{E}[\exp(X^2/K_4^2)] = \exp(a^2/K_4^2)$ and hence $\mathbb{E}[\exp(X^2/K_4^2)] \le 2 \ \Leftrightarrow\ a^2/K_4^2 \le \log 2 \ \Leftrightarrow\ a^2/\log 2 \le K_4^2$.

(iv) Let X be a bounded random variable such that $|X| \le M$ $\mathbb{P}$-a.s. for some $M \ge 0$. Then $\|X\|_{\psi_2} \le M/\sqrt{\log 2}$. Indeed, $\mathbb{E}[\exp(\log 2X^2/M^2)] \le \mathbb{E}[\exp(\log 2)] = 2$.

Example 7.

(i) A Poisson random variable is not sub-Gaussian. Indeed, it follows from Example 3 that $\psi_X(\lambda) = \mathbb{E}[X](e^\lambda - 1)$, so that the logarithm of the MGF growths exponentially and hence it can not be bounded from above by a quadratic function.

(ii) A random variable with distribution $\chi^2(N)$ is not sub-Gaussian. Indeed it follows from Example 5 that the logarithm of the moment generating function assumes the infinite value. This prevents the random variable to be sub-Gaussian.

Lemma 3 (Centering)**.** *Let X a real valued random variable. Then X is sub-Gaussian if and only if $X - \mathbb{E}[X]$ is so and in such case we have* $\|X - \mathbb{E}[X]\|_{\psi_2} \leq C\,\|X\|_{\psi_2}$.

Proof. The first part follows from the fact that $L_{\psi_2}(\Omega, \mathbb{P})$ is a vector space that contains constant random variables. We prove the inequality. Triangular inequality yields $\|X - \mathbb{E}[X]\|_{\psi_2} \leq \|X\|_{\psi_2} + \|\mathbb{E}[X]\|_{\psi_2}$. Moreover, recalling Example 6(iii) and Remark 6, we have

$$\|\mathbb{E}[X]\|_{\psi_2} = \frac{1}{\sqrt{\log 2}}|\mathbb{E}[X]| \leq \frac{1}{\sqrt{\log 2}}\,\|X\|_1 \leq \sqrt{\frac{\pi}{\log 2}}\,\|X\|_{\psi_2}. \tag{65}$$

Hence $\|X - \mathbb{E}[X]\|_{\psi_2} \leq (1 + \sqrt{\pi/\log 2})\,\|X\|_{\psi_2}$. □

4 Sub-exponential random variables

We start with two examples.

Example 8.

(i) Let X be a random variable with exponential distribution with parameter $\nu > 0$. Then $\mathbb{P}_X = f dx$, where $f(x) = \nu e^{-\nu x}$ if $x \geq 0$ and $f(x) = 0$ if $x < 0$. It is well known that $\mathbb{E}[X] = 1/\nu$ and one can easily check that, for every $t > 0$, $\mathbb{P}(\{X > t\}) = \int_t^\infty \nu e^{-\nu x} dx = e^{-\nu t}$. Moreover,

$$\mathbb{M}_X(\lambda) = \mathbb{E}[e^{\lambda X}] = \int_{\mathbb{R}_+} \nu e^{-(\nu-\lambda)x} dx. \tag{66}$$

Therefore, if $\nu - \lambda \leq 0$, we have, for every $x \in \mathbb{R}_+$ $e^{-(\nu-\lambda)x} \geq e^0 = 1$ and hence $\mathbb{M}_X(\lambda) = +\infty$. Otherwise, if $\nu - \lambda > 0$, then

$$\begin{aligned}\mathbb{M}_X(\lambda) &= \frac{\nu}{\nu-\lambda}(\nu-\lambda)\int_{\mathbb{R}_+} e^{-(\nu-\lambda)x} dx\\ &= \frac{\nu}{\nu-\lambda} = \frac{1}{1-\lambda/\nu} \leq e^{2\lambda/\nu}.\end{aligned}$$

Thus, we have also $\operatorname{dom}\psi_X = \,]-\infty, \nu[$ and, for every $\lambda \in \operatorname{dom}\psi_X$, $\psi_X(\lambda) = -\log(1 - \lambda/\nu)$.

(ii) Let X be a random variable with Laplace distribution with parameters μ and b, that is, $\mathbb{P}_X = f dx$, where $f(x) = (2b)^{-1}\exp(-|x-\mu|/b)$. It is easy to see that, for $t > 0$,

$$\mathbb{P}(X > t + \mu) = \frac{1}{2} e^{-t/b}. \tag{67}$$

Indeed with the change of variables $u = x - \mu$, we have

$$\begin{aligned}
\mathbb{P}(\{X > t + \mu\}) &= \int_{t+\mu}^{\infty} \frac{1}{2b} e^{-|x-\mu|/b} dx \\
&= \int_{t}^{\infty} \frac{1}{2b} e^{-u/b} du = \frac{1}{2}\Big[-e^{-u/b}\Big]_t^{\infty} = \frac{1}{2} e^{-t/b}.
\end{aligned}$$

Moreover, suppose that $\mu = 0$ and $b = 1$. Then

$$\begin{aligned}
\mathbb{E}[e^{\lambda X}] &= \frac{1}{2b} \int_{\mathbb{R}} e^{\lambda x} e^{-|x|/b} dx \\
&= \frac{1}{2b} \int_{\mathbb{R}_+} e^{-(1/b-\lambda)x} dx + \frac{1}{2b} \int_{\mathbb{R}_-} e^{(1/b+\lambda)x} dx \\
&= \frac{1}{2b} \int_{\mathbb{R}_+} e^{-(1/b-\lambda)x} dx + \frac{1}{2b} \int_{\mathbb{R}_+} e^{-(1/b+\lambda)x} dx
\end{aligned}$$

Therefore, if $|\lambda| < 1/b$, we have

$$\begin{aligned}
\mathbb{E}[e^{\lambda X}] &= \frac{1}{2b}\Big[-\frac{1}{1/b-\lambda} e^{-(1/b-\lambda)x}\Big]_0^{\infty} + \frac{1}{2b}\Big[-\frac{1}{1/b+\lambda} e^{-(1/b+\lambda)x}\Big]_0^{\infty} \\
&= \frac{1}{2}\Big(\frac{1}{1-b\lambda} + \frac{1}{1+b\lambda}\Big) \\
&= \frac{1}{1-b^2\lambda^2}.
\end{aligned}$$

Otherwise if $|\lambda| \geq b$, we have $\mathbb{E}[e^{\lambda X}] = +\infty$. Now, we note that for $x \in [0, 1/2]$, we have $e^{2x} \geq 1 + 2x \geq 1/(1-x)$. Hence $\mathbb{E}[e^{\lambda X}] \leq \exp(2b^2\lambda^2)$ for $|\lambda| \leq 1/(\sqrt{2}b)$. In the end we have that

$$\psi_X(\lambda) \leq \begin{cases} 2b^2\lambda^2 & \text{if } |\lambda| \leq \dfrac{1}{\sqrt{2}b} \\ +\infty & \text{otherwise.} \end{cases} \tag{68}$$

Theorem 4. Let X be a real-valued random variable, let $K_i > 0$, $i = 1, \ldots, 6$, and consider the following statements.

(a) For every $t > 0$, $\mathbb{P}(\{|X| \geq t\}) \leq 2 \exp\Big(-\dfrac{t}{C_1}\Big)$

(b) For every $p \geq 1$, $\mathbb{E}[|X|^p] \leq C_2^p \, p\Gamma(p)$.

(c) For every integer $n \geq 1$, $\mathbb{E}[|X|^n] \leq C_3^n n!$

(d) $\mathbb{E}[\exp(|X|/C_4) \leq 2$.

Then the following hold.

(i) (a) $\Rightarrow$ (b) with $C_2 = 2C_1$.

(ii) (b) $\Rightarrow$ (c) with $C_3 = C_2$.

(iii) (c) $\Rightarrow$ (d) with $C_4 = 2C_3$.

(iv) (d) $\Rightarrow$ (a) with $C_1 = C_4$.

Moreover, suppose that $\mathbb{E}[X] = 0$ and consider the statements

(e) $(\forall\, \lambda \in \mathbb{R})$, $\psi_X(\lambda) \leq \dfrac{\lambda^2 C_5^2}{2} + \iota_{[-\frac{1}{C_5}, \frac{1}{C_5}]}(\lambda)$.

(f) $(\forall\, t > 0)$, $\max\{\mathbb{P}(\{X \geq t\}), \mathbb{P}(\{X \leq -t\})\} \leq \exp\Big(-\dfrac{1}{2}\min\Big\{\dfrac{t^2}{C_6^2}, \dfrac{t}{C_6}\Big\}\Big)$

Then the following hold.

(v) (c) $\Rightarrow$ (e) with $C_5 = 2C_3$.

(vi) (e) $\Rightarrow$ (f) with $C_6 = C_5$.

(vii) (f) $\Rightarrow$ (c) with $C_3 = 2C_6$.

Proof. In the following we set $Y = \sqrt{|X|}$.

(i): For all $t > 0$,

$$\mathbb{P}(\{|Y| > t\}) = \mathbb{P}(\{|X| > t^2\}) = 2e^{-t^2/C_1}.$$

Therefore, Theorem 1(a) holds with $K_1 = \sqrt{C_1}/\sqrt{2}$, meaning that Y is sub-Gaussian and hence in virtue of Theorem 1 we have, for every $p \geq 1$,

$\mathbb{E}[|Y|^p] \leq (\sqrt{2}K_1)^p p\Gamma(p/2)$. Then,

$$(\forall\, p \geq 1) \quad \mathbb{E}[|Y|^{2p}] \leq (\sqrt{C_1})^{2p} 2p\Gamma(p), \tag{69}$$

so that, for every $p \geq 1$, $\mathbb{E}[|X|^p] \leq C_1^p 2p\Gamma(p) \leq (2C_1)^p p\Gamma(p)$.

(iii): For every $n \in \mathbb{N}$, $\mathbb{E}[|Y|^{2n}] \leq C_3^n n!$ and hence Theorem 1(c) holds with $K_3 = \sqrt{C_3}$. Thus, we have $\mathbb{E}[\exp(Y^2/K_4^2)] \leq 2$ with $K_4 = \sqrt{2C_3}$.

(iv): We have $\mathbb{E}[\exp(Y^2/C_4)] \leq 2$ so that Theorem 1(a) holds for Y with $K_1 = \sqrt{C_4/2}$, meaning that for every $t > 0$,

$$\mathbb{P}(\{|X| > t\}) = \mathbb{P}(|Y| > \sqrt{t}\}) \leq e^{-t/(2K_1^2)} = e^{-t/C_4} \tag{70}$$

(v): We have $\mathbb{E}[\exp(|\lambda||X|)] = 1 + |\lambda|\mathbb{E}[|X|] + \sum_{n=2}^{+\infty}|\lambda|^n\mathbb{E}[|X|^n]/n!$ and, if $|\lambda|C_3 < 1$,

$$\sum_{n=2}^{+\infty} \frac{|\lambda|^n \mathbb{E}[|X|^n]}{n!} \leq \sum_{n=2}^{+\infty} (|\lambda|C_3)^n = \frac{\lambda^2 C_3^2}{1 - |\lambda|C_3} < +\infty. \tag{71}$$

Then, if $|\lambda|C_3 \leq 1/2$,

$$\mathbb{E}[\exp(\lambda X)] = 1 + \sum_{n=2}^{\infty} \frac{\lambda^n \mathbb{E}[X^n]}{n!} \leq 1 + \frac{\lambda^2 C_3^2}{1 - |\lambda|C_3} \leq 1 + 2\lambda^2 C_3^2 \leq e^{2\lambda^2 C_3^2}. \tag{72}$$

(vi): It follows from Chernoff's inequality and Proposition 4(iv).

(vii): We note that, for every $t > 0$,

$$\mathbb{P}(\{|X| > t\}) \leq \mathbb{P}(\{X > t\}) + \mathbb{P}(\{X < -t\}) \leq 2\exp\Big(-\frac{1}{2}\min\Big\{\frac{t^2}{C_6^2}, \frac{t}{C_6}\Big\}\Big).$$

Therefore, for $n \in \mathbb{N}$, $n \geq 2$,

$$\begin{aligned}
\mathbb{E}[|X|^n] &= \int_0^\infty \mathbb{P}(\{|X|^n > t\})dt \\
&= \int_0^\infty \mathbb{P}(\{|X| > t^{1/n}\})dt \\
&\leq 2\int_0^\infty \exp\Big(-\frac{1}{2}\min\Big\{\frac{t^{2/n}}{C_6^2}, \frac{t^{1/n}}{C_6}\Big\}\Big)dt \\
&\leq 2\int_0^{C_6^n} \exp\Big(-\frac{1}{2}\frac{t^{2/n}}{C_6^2}\Big)dt + 2\int_{C_6^n}^\infty \exp\Big(-\frac{1}{2}\frac{t^{1/n}}{C_6}\Big)dt
\end{aligned}$$

$$
\begin{aligned}
&\leq 2\int_0^\infty \exp\Big(-\frac{1}{2}\frac{t^{2/n}}{C_6^2}\Big)dt + 2\int_0^\infty \exp\Big(-\frac{1}{2}\frac{t^{1/n}}{C_6}\Big)dt\\
&= n\int_0^\infty (\sqrt{2}C_6)^n e^{-u}u^{n/2-1}du + 2\int_0^\infty C_6^n n e^{-u}u^{n-1}du\\
&= (\sqrt{2}C_6)^n n\Gamma(n/2) + 2C_6^n n\Gamma(n).
\end{aligned}
$$

Then noting that, for $n \geq 2$, $\Gamma(n/2) \leq \Gamma(n)$, we have

$$
\mathbb{E}[|X|^n] \leq (2^{n/2}+2)C_6^n n\Gamma(n) \leq (2C_6)^n n!, \tag{73}
$$

where we used the fact that $2^{n/2}+2 \leq 2^n$ for $n \geq 2$. Therefore, (c) holds with $C_3 = 2C_6$. □

Definition 4. Let X be a real-valued random variable. Then X is *sub-exponential* if condition (d) of Theorem 4 holds for some $C_4 > 0$. Let $\psi_1 \colon \mathbb{R} \to \mathbb{R}$, $\psi_1(x) = e^x - 1$. Then condition (d) of Theorem 4 can be equivalently written as $\mathbb{E}[\psi_1(|X|/C_4)] \leq 1$. Hence the space of all sub-exponential random variables is the Orlicz space defined by ψ_1, that is

$$
\begin{aligned}
L_{\psi_1}(\Omega,\mathbb{P}) &= \big\{X\colon \Omega \to \mathbb{R}\text{ r.v.}\,\big|\,\exists\, C_4 > 0,\ \mathbb{E}[\psi_1(|X|/C_4)] \leq 1\big\}\\
&= \big\{X\colon \Omega \to \mathbb{R}\text{ r.v.}\,\big|\,\exists\, C_4 > 0,\ \mathbb{E}[\psi_1(|X|/C_4)] < +\infty\big\}
\end{aligned}
$$

endowed with the norm

$$
\begin{aligned}
\|X\|_{\psi_1} &= \inf\big\{C_4 > 0\,\big|\,\mathbb{E}[\psi_1(|X|/C_4)] \leq 1\big\}\\
&= \inf\big\{C_4 > 0\,\big|\,\mathbb{E}[\exp(|X|/C_4)] \leq 2\big\}.
\end{aligned}
$$

Motivated by Theorem 4(e) we denote by $\mathsf{subE}(\sigma)$ the class of all zero-mean random variables such that $\psi_X(\lambda) \leq \lambda^2\sigma^2/2 + \iota_{[-1/\sigma,1/\sigma]}(\lambda)$. Note that, the presence of the indicator function leads to a linear behavior of the dual function for large values ψ_X^* (see Propostion 4(iv)).

Remark 7. Suppose that condition (c) of Theorem 4 holds. Then, for every $\lambda \in \mathbb{R}$, such that $|\lambda| < 1/C_3$,

$$
\mathbb{E}[\exp(|\lambda||X|)] = 1 + \sum_{n=1}^\infty \frac{|\lambda|^n\mathbb{E}[|X|^n]}{n!} \leq \sum_{n=0}^\infty (|\lambda|C_3)^n = \frac{1}{1-|\lambda|C_3} \tag{74}
$$

and hence

$$
\mathbb{E}[\exp(\lambda|X|)] = 1 + \sum_{n=1}^\infty \frac{\lambda^n\mathbb{E}[|X|^n]}{n!} \leq \sum_{n=0}^\infty (\lambda C_3)^n = \frac{1}{1-\lambda C_3} \tag{75}
$$

Sub-exponential and sub-Gaussian random variables are closely related as the following result shows.

Proposition 12. *Let X be a real-valued random variable and let $Y = \sqrt{|X|}$. Then X is sub-exponential $\Leftrightarrow$ Y is sub-Gaussian, and in such case $\|X\|_{\psi_1} = \|Y\|_{\psi_2}^2$.*

Proof. The statement follows from the fact that $\mathbb{E}[\exp(|X|/C_4)] \leq 2 \Leftrightarrow \mathbb{E}[\exp(|Y|^2/C_4)] \leq 2 \Leftrightarrow Y$ is sub-Gaussian and $\|Y\|_{\psi_2} \leq \sqrt{C_4}$. □

Proposition 13. *Let X and Y be two real-valued sub-Gaussian random variables. Then XY is sub-exponential and $\|XY\|_{\psi_1} \leq \|X\|_{\psi_2} \|Y\|_{\psi_2}$.*

Proof. We may suppose that $\|X\|_{\psi_2} = \|Y\|_{\psi_2} = 1$. We apply the Young inequality $ab \leq (1/2)(a^2 + b^2)$. We have

$$\begin{aligned}
\mathbb{E}[\exp(XY)] &\leq \mathbb{E}\Big[\exp\Big(\frac{1}{2}(X^2 + Y^2)\Big)\Big] \\
&= \mathbb{E}[\exp(X^2/2)\exp(Y^2/2)] \\
&\leq \mathbb{E}\Big[\frac{1}{2}(\exp(X^2/2)^2 + \exp(Y^2/2)^2)\Big] \\
&= \frac{1}{2}\mathbb{E}[\exp(X^2)] + \frac{1}{2}\mathbb{E}[\exp(Y^2)] \\
&\leq 1 + 1 = 2,
\end{aligned}$$

so that XY is sub-exponential and $\|XY\|_{\psi_1} \leq 1 = \|X\|_{\psi_2} \|Y\|_{\psi_2}$. □

Example 9. We give a list of sub-exponential random variables.

(i) If X is a sub-Gaussian random variable, then X is sub-exponential and $\|X\|_{\psi_1} \leq \|X\|_{\psi_2} / \sqrt{\log 2}$. Indeed just take $Y = 1$ in Proposition 13 and use Example 6(iii). Note that this inequality is strict since it becomes an equality if one takes $X = 1$, since $\|1\|_{\psi_1} = 1/\log 2$.

(ii) If X is a Gaussian random variable, then X^2 is sub-exponential and $\|X^2\|_{\psi_1} = \|X\|_{\psi_2}^2 = 8\sigma^2/3$. This follows from Proposition 12 and Example 6(i).

(iii) If X is distributed as $\chi^2(N)$, then it is sub-exponential. Indeed, by Example 5 we have that $X = \sum_{i=1}^N X_i^2$, where $(X_i)_{1 \leq i \leq N}$ is a finite sequence of independent normal random variables. Note that X is positive, so condition (d) of Theorem 4 can be written as $\psi_X(1/C_4) \leq \log 2$. Therefore, recalling (63) in Example 5, we have $\psi_X(1/r) \leq \log 2 \Leftrightarrow 1/(1-2/r)^{N/2} \leq 2 \Leftrightarrow 1-2/r \geq 1/2^{2/N} \Leftrightarrow 1-1/2^{2/N} \geq 2/r \Leftrightarrow r \geq 2^{1+2/N}/(2^{2/N}-1)$. Hence, X is sub-exponential and $\|X\|_{\psi_1} = 2^{1+2/N}/(2^{2/N}-1)$.

(iv) Exponential distributed random variables are sub-exponential. Indeed, since $X \geq 0$ $\mathbb{P}$-a.s., it follows from Example 8(i) that for $r > 0$ we have

$$\mathbb{E}[\exp(|X|/r)] = \mathbb{E}[\exp(X/r)] = \frac{1}{1-1/(r\nu)} + \iota_{]-\infty,\nu[}\Big(\frac{1}{r}\Big). \tag{76}$$

Thus, by definition X is sub-exponential and $\|X\|_{\psi_1} = \nu/2$.

(v) Poisson distributed random variables are sub-exponential. Indeed, since $X \geq 0$ $\mathbb{P}$-a.s., it follow from Example 3 that, $\psi_X(\lambda) = \mu(e^\lambda - 1)$. Hence, for $r > 0$, we have $\mathbb{E}[\exp(|X|/r)] = \mathbb{E}[\exp(X/r)] = \exp(\psi_X(1/r)) = \exp(\mu(e^{1/r}-1))$. Therefore, the random variable X is sub-exponential and $\|X\|_{\psi_1} = \log\big((\log 2)/\mu + 1\big)$.

(vi) In [2] *sub-Gamma* random variables are defined as zero-mean random variables with logarithm of MGF satisfying

$$(\forall\,\lambda \in\,]-1/c, 1/c[) \quad \psi_X(\lambda) \leq \frac{\sigma^2\lambda^2}{2(1-c|\lambda|)} \tag{77}$$

for some $\sigma > 0$ and $c > 0$. Since for $c|\lambda| \leq 1/2$, $(1-c|\lambda|)^{-1} \leq 2$, we have that $\psi_X(\lambda) \leq \sigma^2\lambda^2 + \iota_{[-1/(2c),1/(2c)]}(\lambda)$. Therefore, X is sub-exponential, since it satisfies Theorem 4(e) with $C_5 = \max\{\sqrt{2}\sigma, 2c\}$. Vice versa if X satisfies Theorem 4(e), since $1 \leq (1-c|\lambda|)^{-1}$, then it satisfies (77) with $\sigma = c = C_5$. Therefore, the class of sub-Gamma random variables is the same of that of zero-mean sub-exponential random variables.

Theorem 5 (Bernstein's inequality)**.** *Let $(X_i)_{1\leq i\leq N}$ be a sequence of independent real-valued zero-mean random variables such that $X_i \in \mathsf{subE}(\sigma_i)$, meaning that, for every $\lambda \in \mathbb{R}$, $\psi_{X_i}(\lambda) \leq \lambda^2\sigma_i^2/2 + \iota_{[-1/\sigma_i,1/\sigma_i]}(\lambda)$. Then, for every $t > 0$*

$$\mathbb{P}\Big(\Big\{\sum_{i=1}^N X_i > t\Big\}\Big) \leq \exp\Big(-\frac{1}{2}\min\Big\{\frac{t^2}{\sum_{i=1}^N \sigma_i^2}, \frac{|t|}{\max_{1\leq i\leq N}\sigma_i}\Big\}\Big) \tag{78}$$

Proof. Let $S = \sum_{i=1}^{N} X_i$. Then

$$\psi_S(\lambda) = \sum_{i=1}^{N} \psi_{X_i}(\lambda) = \frac{\lambda^2}{2} \sum_{i=1}^{N} \sigma^2 + \iota_{[-\min \frac{1}{\sigma_i}, \min \frac{1}{\sigma_i}]}(\lambda). \tag{79}$$

Therefore, it follows from Proposition 4(iv) that

$$\psi_S^*(t) \geq \min \Big\{ \frac{t^2}{2 \sum_{i=1}^{N} \sigma_i^2}, \frac{|t|}{2 \max_{1 \leq i \leq N} \sigma_i} \Big\},$$

hence the statement follows. □

Theorem 6 (Bernstein's inequality)**.** *Let X be a real-valued random variable and suppose that,*

$$(\forall\, n \geq 2) \quad \mathbb{E}[|X|^n] \leq \frac{\sigma^2}{2} c^{n-2} n! \tag{80}$$

for some $\sigma \geq 0$ and $c \geq 0$. Then, for every $t > 0$

$$\mathbb{P}(\{X - \mathbb{E}[X] > t\}) \leq \exp\Big(- \frac{t^2}{2(\sigma^2 + ct)} \Big).$$

Proof. Let $\lambda \in \mathbb{R}$. It follows from Proposition 8(iv) that $\psi_{X-\mathbb{E}[X]}(\lambda) = \psi_X(\lambda) - \lambda \mathbb{E}[X] = \log \mathbb{E}[\exp(\lambda X)] - \lambda \mathbb{E}[X] \leq \mathbb{E}[\exp(\lambda X)] - 1 - \lambda \mathbb{E}[X]$, where we used that $\log u \leq u - 1$ for $u > 0$. Therefore, setting $\phi_0(u) = e^u - u - 1$, we have $\psi_{X-\mathbb{E}[X]}(\lambda) \leq \mathbb{E}[\phi_0(\lambda X)]$. Now, since $\phi_0(u) = \sum_{n=2}^{\infty} u^n/n! \leq \sum_{n=2}^{\infty} |u|^n/n!$, we have

$$\begin{aligned}
\psi_{X-\mathbb{E}[X]}(\lambda) &\leq \mathbb{E}[\phi_0(\lambda X)] \\
&\leq \sum_{n=2}^{\infty} \frac{|\lambda|^n \mathbb{E}[|X|^n]}{n!} \\
&\leq \frac{\sigma^2}{2} \sum_{n=2}^{\infty} |\lambda|^n c^{n-2} \\
&= \frac{\sigma^2 \lambda^2}{2(1 - c|\lambda|)} \\
&= \frac{\sigma^2}{c^2} \phi_2(c\lambda),
\end{aligned}$$

where ϕ_2 is defined in Proposition 4(iii). Then, using (1) we have

$$\left[\frac{\sigma^2}{c^2}\phi_2(c\,\cdot)\right]^*(t) = \frac{\sigma^2}{c^2}\phi_2^*\left(\frac{ct}{\sigma^2}\right) \tag{81}$$

and recalling Proposition 4(iii) and Chernoff's inequality given in Theorem 9 the statement follows. $\square$

Remark 8. Random variables satisfying inequality (80) are sub-exponential. Indeed, Let $C_3 = \max\{\sigma/\sqrt{2}, c\}$, then $2^{-1}\sigma^2 c^{n-2} \leq C_3^n$ and hence condition (c) in Theorem 4 holds (note that $\mathbb{E}[|X|] = \|X\|_1 \leq \|X\|_2 \leq \sigma$).

References

[1] P. Rigollet, J-C. Hütter, High Dimensional Statistics (Lecture Notes). 2017.

[2] S. Boucheron, G. Lugosi, P. Massart, Concentration Inequalities. Oxford University Press, 2013.

[3] I. Steinwart, A. Christmann, Support Vector Machines. Springer Verlag, 2008.

[4] R. Vershynin, High-Dimensional Probability. Cambridge University Press, 2018.