MATH 7251: High-Dimensional Probability	**Spring 2022**
Lecture 4	
Name: Tyler LaBonte	**Professor:** Konstantin Tikhomorov

1 Subgaussian and subexponential random variables

The function $\Psi : [0, \infty) \to [0, \infty)$ is called an Orlicz function if $\Psi(0) = 0$ and Ψ is convex and strictly increasing. Given a random variable X, we define the Orlicz norm of X wrt Ψ as

$$\|X\|_\Psi \coloneqq \inf \{\lambda > 0 : \mathbb{E}[\Psi(|X|/\lambda)] \leq 1\}. \tag{1}$$

Note that for some random variables and choices of Ψ, $\|X\|_\Psi$ can be undefined ($+\infty$). To avoid this, we restrict ourselves to Orlicz spaces L_Ψ, the collection of all random variables on a common probability space $(\Omega, \Sigma, \mathbb{P})$ such that the Orlicz norm is finite. Two important examples are:

1. A variable X is subgaussian if $\|X\|_{\Psi_2} < \infty$, where $\Psi_2(t) \coloneqq \exp(t^2) - 1$.
2. A variable X is subexponential if $\|X\|_{\Psi_1} < \infty$, where $\Psi_1(t) \coloneqq \exp(t) - 1$.

Note that any Gaussian random variable $\mathcal{N}(\mu, \sigma^2)$ is also subgaussian, and any subgaussian random variable is also subexponential. Additionally, any uniformly bounded variable (*e.g.,* Bernoulli or uniform on an interval) is subgaussian. Thus, subgaussian variables are important for unifying uniformly bounded and Gaussian random variables. In particular, if $X_1, X_2, \ldots$ are iid uniformly bounded random variables of zero mean, then by CLT, we know that $\frac{1}{\sqrt{n}} \sum_{i=1}^n X_i$ converge to a Gaussian variable $\mathcal{N}(0, \sigma^2)$. So, if we start with a uniformly bounded variable, then in the limit, these partial sums are not uniformly bounded, but they are uniformly subgaussian. The class of subgaussian variables appears in connection with CLT and taking linear combinations of independent random variables.

If X is subgaussian, then X is subexponential, but also X^2 is subexponential:

$$\|X^2\|_{\Psi_1} = \|X\|_{\Psi_2}^2. \tag{2}$$

The most natural example of a subexponential random variable is a χ^2 variable (*i.e.,* the square of a Gaussian variable). Also, the standard exponential random variable Y with the distribution density

$$\rho(t) = \begin{cases} 0 & t \leq 0, \\ \exp(-t) & t > 0 \end{cases} \tag{3}$$

is subexponential.

A natural way to think of subgaussian and subexponential random variables is that they are the random variables with lighter tails (*i.e.,* faster decaying tails) than Gaussian and exponential random variables, respectively.

An Orlicz space is called a "space" because it is a normed space, *i.e.,* it is closed under linear combination. In particular, subgaussian and subexponential random variables are closed under linear combination, even if they are not independent. This gives a trivial bound on the size of the norm of the sum:

$$\left\| \sum_{i=1}^{m} X_i \right\|_{\Psi_2} \le \sum_{i=1}^{m} \|X_i\|_{\Psi_2}. \tag{4}$$

Also, for a subgaussian variable X and $\lambda \in \mathbb{R}$,

$$\|\lambda X\|_{\Psi_2} = |\lambda| \, \|X\|_{\Psi_2}. \tag{5}$$

Lemma 1.1: Characterization of subgaussian random variables

For every random variable X with $\|X\|_{\Psi_2} \le 1$, we have for all $p \ge 1$

$$\mathbb{E}[|X|^p] \le (C\sqrt{p})^p, \tag{6}$$

where C is a universal constant (like 10). Conversely, if Y is a random variable such that for all $p \ge 1$ integer we have

$$\mathbb{E}[|Y|^p] \le (\sqrt{p})^p \tag{7}$$

then $\|Y\|_{\Psi_2} \le C$. That is, Y is subgaussian.

Proof: Consider the first part of the lemma where X is a subgaussian random variable with $\|X\|_{\Psi_2} \le 1$. By definition,

$$\mathbb{E}[|X|^p] = p \int_0^\infty t^{p-1} \Pr[|X| \ge t] \, dt. \tag{8}$$

Recall also Markov's inequality: if Z is a random variable, then for all $s > 0$,

$$\Pr[Z \ge s] \le \frac{\mathbb{E}[Z]}{s}. \tag{9}$$

Applying the inequality,

$$\begin{aligned} \Pr[|X| \ge t] &= \Pr\big[\exp\big(X^2\big) \ge \exp\big(t^2\big)\big] && (10) \\ &\le \frac{\mathbb{E}\big[\exp\big(X^2\big)\big]}{\exp(t^2)} && (11) \\ &\le \frac{2}{\exp(t^2)}, && (12) \end{aligned}$$

where the last step follows from $\|X\|_{\Psi_2} \le 1 \implies \mathbb{E}\big[\exp\big(X^2\big) - 1\big] \le 1$. Returning to moment computations, we get

$$\mathbb{E}[|X|^p] \leq p\int_0^\infty t^{p-1}\cdot\frac{2}{\exp(t^2)}dt \tag{13}$$

$$\leq 2p\int_0^{\sqrt{p}}\frac{t^{p-1}}{\exp(t^2)}dt + 2p\int_{\sqrt{p}}^\infty \frac{t^{p-1}}{\exp(t^2)}dt \tag{14}$$

$$\leq 2p\int_0^{\sqrt{p}} t^{p-1}dt + 2p\int_{\sqrt{p}}^\infty \frac{t^{p-1}}{\exp(t^2)}dt \tag{15}$$

$$\leq 2p(\sqrt{p})^p + 2p\int_{\sqrt{p}}^\infty \frac{t^{p-1}}{\exp(\sqrt{p}t)}dt. \tag{16}$$

We now need to analyze the function in the second integral. We can write it as

$$\exp(-\sqrt{p}t + (p-1)\log(t))\,. \tag{17}$$

This function achieves its maximum value at $\frac{p-1}{\sqrt{p}}$ and decays very fast to either side of the maximum. So, the value of the integral is more or less only affected by a narrow region around the maximum. The maximum value turns out to be $(c\sqrt{p})^p$. Since $2p \leq c^p$ for sufficiently large c, both terms are expressed in terms of $(c\sqrt{p})^p$. We can conclude that $\mathbb{E}[|X|]^p \leq (c\sqrt{p})^p$ for all $p \geq 1$. ■

Lemma 1.2: Characterization of subexponential random variables

For every random variable X with $\|X\|_{\Psi_1} \leq 1$, we have for all $p \geq 1$

$$\mathbb{E}[|X|]^p \leq (Cp)^p, \tag{18}$$

where C is a universal constant (like 10). Conversely, if Y is a random variable such that for all $p \geq 1$ integer we have

$$\mathbb{E}[|Y|]^p \leq p^p \tag{19}$$

then $\|Y\|_{\Psi_1} \leq C$. That is, Y is subexponential.

2 Concentration Inequalities

Theorem 2.1

If $X_1, X_2, \ldots, X_n$ are mutually independent zero mean subgaussian variables (not necessarily identically distributed), then

$$\left\|\sum_{i=1}^n X_i\right\|_{\Psi_2}^2 \leq C\sum_{i=1}^n \|X_i\|_{\Psi_2}^2\,. \tag{20}$$

Corollary 2.1: Kchintchine's Inequality

If $r_1, r_2, \ldots, r_n$ are iid symmetric sign variables (± 1 with probability $1/2$) and $\boldsymbol{a} \in \mathbb{R}^n$ has $\|a\|_2 = 1$, then $\sum_{i=1}^n a_i r_i$ is subgaussian with a universal constant. This implies that for every $p \geq 1$ we have

$$\left(\mathbb{E}\left[\left|\sum_{i=1}^n a_i r_i\right|^p\right]\right)^{1/p} \leq c\sqrt{p}, \tag{21}$$

where $c \leq 10$ is a universal constant.

Proof: Of the corollary. Note that $\|a_i r_i\|_{\Psi_2} = |a_i| \, \|r_i\|_{\Psi_2}$, and the second term is a universal constant $\tilde{c}$. Then,

$$\sum_{i=1}^n \|a_i r_i\|_{\Psi_2}^2 = \sum_{i=1}^n |a_i|^2 \tilde{c}^2 = \tilde{c}^2. \tag{22}$$

Applying the theorem yields the result. ∎

Corollary 2.2: Hoeffding's Inequality

If $X_1, X_2, \ldots, X_n$ are independent zero mean subgaussian random variables with $\|X_i\|_{\Psi_2} \leq 1$ for all i and $\boldsymbol{a} \in \mathbb{R}^n$ has $\|\boldsymbol{a}\|_2 \leq 1$, then $\sum_{i=1}^n a_i X_i$ is C-subgaussian.
Another form is as follows. If $X_1, X_2, \ldots, X_n$ are independent with X_i distributed within an interval $[a_i, b_i]$ for all i, then for $t > 0$ we have

$$\Pr\left[\left|\sum_{i=1}^n X_i - \mathbb{E}\left[\sum_{i=1}^n X_i\right]\right| > t\right] \leq 2\exp\left(-\frac{ct^2}{\sum_{i=1}^n (b_i - a_i)^2}\right), \tag{23}$$

where c is a small universal constant like $1/10$. Equivalently,

$$\left\|\sum_{i=1}^n X_i - \mathbb{E}\left[\sum_{i=1}^n X_i\right]\right\|_{\Psi_2} \leq \tilde{c}\sqrt{\sum_{i=1}^n (b_i - a_i)^2}. \tag{24}$$

The idea of the proof is to note that

$$\|X_i - \mathbb{E}[X_i]\|_{\Psi_2} \leq c_1(b_i - a_i) \tag{25}$$

and apply the theorem, then Markov's Inequality.