

# IDS.160 – Mathematical Statistics: A Non-Asymptotic Approach

Lecturer: P RIGOLLET

Lecture 2

Scribes: VARKEY ALUMOOTIL, SYLVIA KLOSIN AND BRICE HUANG

Feb. 6, 2020

**Goals:** Last time we introduced the class and gave a brief overview of the flavor of non-asymptotic statistics as opposed to classical asymptotic and high dimensional statistics. Today we will cover probabilistic tools in this field, especially for tail bounds. In particular, we will cover subGaussian random variables, Chernoff bounds, and Hoeffding's Inequality.

A tool in classical asymptotic analysis is the central limit theorem (CLT): If  $X_1, \dots, X_n \sim P$  iid with mean  $\mu$  and variance  $\sigma^2$ , then

$$\sqrt{n} (\bar{X}_n - \mu) \xrightarrow{d} \mathcal{N}(0, \sigma^2)$$

We note that if  $P = \mathcal{N}(\mu, \sigma^2)$ , then we actually have for all  $n$ ,

$$\sqrt{n} (\bar{X}_n - \mu) \sim \mathcal{N}(0, \sigma^2)$$

We would like to consider a class of distributions in which we can get non-asymptotic results somewhere between the generality of the central limit theorem and the specificity of just Gaussians. This class will be the class **subGaussian** random variables. For these we will have

$$\sqrt{n} (\bar{X}_n - \mu) \approx \mathcal{N}(0, \sigma^2)$$

Where  $\approx$  that the tails look similar, or the moment generating function are similar.

## 1. GAUSSIAN TAILS AND MOMENT GENERATING FUNCTIONS

Consider a standard normal random variable  $Z \sim \mathcal{N}(0, 1)$ . Then it has pdf given by

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right).$$

A key trait of this distribution is that the tail probability grows exponentially small.

**Proposition (Mills Ratio Inequality):**  $\forall t > 0$ ,

$$\mathbb{P}[|Z| > t] \leq \sqrt{\frac{2}{\pi}} \frac{e^{-\frac{t^2}{2}}}{t}.$$

*Proof.* We have

$$\mathbb{P}[Z > t] = \int_t^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

However, on the interval  $[t, \infty)$ ,  $x \geq t$ , so we can upper bound the right hand side by multiplying by  $\frac{x}{t} \geq 1$ .

$$\int_t^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \leq \int_t^\infty \left(\frac{x}{t}\right) \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \frac{1}{t\sqrt{2\pi}} \int_t^\infty x \exp(-\frac{x^2}{2}) dx = \frac{1}{\sqrt{2\pi}} \frac{e^{-\frac{t^2}{2}}}{t}.$$

Since  $Z$  is symmetric ( $Z$  has the same distribution as  $-Z$ ), we have that  $\mathbb{P}[Z < -t] = \mathbb{P}[Z > t]$ . Moreover,

$$\mathbb{P}[|Z| > t] = \mathbb{P}[Z > t] + \mathbb{P}[Z < -t] = 2\mathbb{P}[Z > t] = 2 \frac{1}{\sqrt{2\pi}} \frac{e^{-\frac{t^2}{2}}}{t}.$$

□

An immediate consequence is that if  $X \sim \mathcal{N}(\mu, \sigma^2)$ , then CLT says that

$$\lim_{n \rightarrow \infty} \mathbb{P}\left[\sqrt{n} \frac{|\bar{X}_n - \mu|}{\sigma} > t\right] \leq \sqrt{\frac{2}{\pi}} \frac{e^{-\frac{t^2}{2}}}{t}$$

To analyze the rate of convergence of the CLT, one applicable theorem is Berry-Esseen.

**Theorem (Berry-Esseen):** Given finite third moments of the i.i.d random variables  $X_i$ , then

$$\left| \mathbb{P}\left[\sqrt{n} \frac{|\bar{X}_n - \mu|}{\sigma} > t\right] - \mathbb{P}[|Z| > t] \right| \leq \frac{C}{\sqrt{n}}$$

for some constant  $C$ . Note that this compares the CDFs of our sample average distribution and the normal distribution.

We note that this is a 1D result, and this sort of result is still an active area of research for higher dimensions (c.f. V. Chernozhukov). Furthermore, Berry-Esseen does not give the sort of result we want. If we use triangle inequality on the Berry-Esseen result, then for a fixed  $n$ , we bound the tail probability above, but this bound is capped by a term depending on  $n$ . Restricting our random variables to only those with finite third moment gave some progress, but we will have to have more restrictions to get more useful results.

## 2. SUBGAUSSIAN RANDOM VARIABLES

Recall the definition of the Moment Generating Function.

**Definition (Moment Generating Function (MGF)):** The Moment Generating Function of a random variable  $X$  is the function  $s \mapsto M(s) = \mathbb{E}[e^{sX}]$  for  $s \in \mathbb{R}$ .

**Important Remark:** Note that this does not completely identify the random variable (i.e. log normal); we need  $s \in \mathbb{C}$  and the characteristic function to identify the random variable. This is good enough for our statistical purposes however.

This is called the MGF because

$$\frac{\partial^k}{\partial s^k} M(s) \Big|_{s=0} = \mathbb{E}[X^k]$$

If we can control the MGF, we can produce tail bounds using Chernoff bounds.

**Definition (subGaussian random variables):** A random variable  $X$  is subGaussian with variance proxy  $\sigma^2$  if  $\mathbb{E}X = 0$  and  $\mathbb{E}[e^{sx}] \leq e^{\frac{s^2\sigma^2}{2}}$  for all  $s \in \mathbb{R}$ . This is written as  $X \sim \text{subG}(\sigma^2)$ .

Note that  $X \sim \text{subG}(\sigma^2)$  is an abuse of notation as this is a class of distributions, not a particular one. For convenience we will assume subGaussians are centred, i.e., have mean 0.

**Proposition:** Let  $X$  be a random variable with mean 0 and variance 1. Then the following are equivalent:

- (i)  $\mathbb{E}[e^{sX}] \leq e^{c_1 s^2}$  for all  $s \in \mathbb{R}$ ,
- (ii)  $\mathbb{P}[|X| \geq t] \leq 2e^{-c_2 t^2}$  for all  $t \geq 0$ ,
- (iii)  $\|X\|_p = (\mathbb{E}|X|^p)^{\frac{1}{p}} \leq c_3 \sqrt{p}$  for all  $p = 1, 2, \dots$ ,
- (iv)  $\mathbb{E}[e^{sX^2}] \leq e^{c_4 s}$  for all  $s \in (0, c'_4)$ ,

Note that (ii) is the tail bound we wanted, and (iv) says that  $X^2$  is sub-exponential.

*Proof.* (i)  $\Rightarrow$  (ii). We will apply a *Chernoff bound*. We note that for any positive  $s$ ,

$$\mathbb{P}[X \geq t] = \mathbb{P}[e^{sX} \geq e^{st}] \leq \frac{\mathbb{E}[e^{sX}]}{e^{st}} \leq e^{c_1 s^2 - st}$$

where the first inequality is the application of Markov's inequality and the second inequality is the application of (i). This is true for any positive  $s$ , and we can minimize with respect to  $s$ :

$$\frac{\partial}{\partial s} c_1 s^2 - st = 0 \Leftrightarrow s = \frac{t}{2c_1} > 0$$

Plugging this value into the exponent yields that  $\mathbb{P}[X \geq t] \leq e^{-\frac{t^2}{4c_1}}$ . Similarly, we get that  $\mathbb{P}[X \leq -t] \leq e^{-\frac{t^2}{4c_1}}$  and we conclude the proof of (ii) using a union bound.

(ii)  $\Rightarrow$  (iii). Here we apply integration of tails which is just using Fubini's theorem to switch integrals. If  $x \geq 0$  then

$$x = \int_0^x dt = \int_0^\infty \mathbb{I}[t \leq x] dt$$

In particular, if  $X \geq 0$  a.s., then

$$\mathbb{E}X = \int_0^\infty \mathbb{P}[X \geq t] dt$$

If we apply this to  $|X|^p$ , we get using Fubini's theorem to switch the expectation and the integral sign, we get

$$\mathbb{E}|X|^p = \int_0^\infty \mathbb{P}[|X|^p > t] dt \leq 2 \int_0^\infty e^{-c_2 t^{2/p}} dt$$

We apply the change of variable  $u = c_2 t^{2/p}$  to get

$$\frac{p}{c_2^{p/2}} \int_0^\infty e^{-u} u^{\frac{p}{2}-1} du$$

We recognize the integral as  $\Gamma(\frac{p}{2})$ , which we can upper-bound by  $(\frac{p}{2})^{p/2}$ . We can then produce the bound

$$\|X\|_p \leq Cp^{1/p} \sqrt{\frac{p}{2}}$$

We can however bound  $p^{1/p}$  by a constant, so we get our desired result.

(iii)  $\Rightarrow$  (iv).

We first employ the Taylor Expansion of the exponential function and linearity of expectation to observe that

$$\mathbb{E}[e^{sX^2}] = 1 + \sum_{p=1}^\infty \frac{s^p}{p!} \mathbb{E}X^{2p}$$

Our bound from (iii) states that for all  $p > 0$ ,

$$\mathbb{E}X^{2p} \leq (2c_3^2 p)^p = (2c_3^2 e)^p \left(\frac{p}{e}\right)^p \leq (2c_3^2 e)^p p!$$

The last inequality follows from Stirling's approximation. We substitute this to produce the bound

$$1 + \sum_{p=1}^\infty (2sc_3^2 e)^p = 1 + 2sc_3^2 e \sum_{p=0}^\infty (2sc_3^2 e)^p$$

Let us take  $s$  small enough such that  $2sc_3^2 e < \frac{1}{2}$ . Then this sum is bounded above by

$$1 + 2sc_3^2 e \leq e^{2sc_3^2 e}$$

as desired.

(iv)  $\Rightarrow$  (i). For all  $x \in \mathbb{R}$ ,

$$e^x \leq x + e^{x^2} \tag{2.1}$$

To verify (2.1), define the function  $\psi(x) := e^{x^2} + x - e^x$ . We have

$$\psi'(x) = 2xe^{x^2} + 1 - e^x, \quad \psi''(x) = 2e^{x^2}(1 + 2x^2) - e^x,$$

Next, observe that

$$\psi''(x) \geq 2e^{x^2} - e^{|x|} = e^{|x|}(2e^{x^2-|x|} - 1) \geq \frac{e^{|x|}}{2} \geq \frac{1}{2},$$

where, in the penultimate inequality, we used the fact that  $e^{x^2-|x|} \geq 1 + x^2 - |x| \geq 3/4$ .

In particular  $\psi$  is convex and since  $\psi'(0) = 0$ , it is minimized at 0 where it takes the value 0. Therefore,  $\psi(x) \geq 0$  for all  $x \in \mathbb{R}$ . This concludes the proof of (2.1).

We now resume the proof of (iv)  $\Rightarrow$  (i). Using (2.1), we get

$$\mathbb{E}[e^{sX}] \leq \mathbb{E}[sX + e^{s^2X^2}] \leq e^{c_4s^2}$$

for  $s^2 \in (0, c'_4)$ . Note that the expectation of  $sX$  disappears as  $X$  is centered. If  $s^2 \geq c'_4$ , we use a different technique. We use the fact that

$$2\lambda x \leq \delta\lambda^2 + \frac{x^2}{\delta}$$

for all  $\delta > 0$ . This inequality follows from  $(\sqrt{\delta}\lambda - x/\sqrt{\delta})^2 \geq 0$ .

Choosing of  $\lambda = s/2$  and  $x = X$ , we get

$$\mathbb{E}[e^{sX}] \leq e^{\frac{\delta s^2}{4}} \mathbb{E}[e^{\frac{X^2}{\delta}}]$$

Taking  $\delta = 1/c'_4$ , we get from (iv) that

$$\mathbb{E}[e^{sX}] \leq e^{\frac{s^2}{4c'_4}} e^{c_4c'_4} \leq \exp\left(\frac{s^2}{4c'_4} + c_4s^2\right) = e^{c_1s^2}, \quad c_1 = \frac{1}{4c'_4} + c_4$$

□

We now describe a fifth, alternative definition of subGaussianity. It relies on the general notion of Orlicz norms.

**Definition (Orlicz norm):** The Orlicz  $\psi_2$ -norm of a random variable  $X$  is

$$\|X\|_{\psi_2} = \inf \{t > 0, \quad \mathbb{E}[e^{\frac{X^2}{t^2}}] \leq 2\}$$

Note that not all random variables have a finite  $\psi_2$ -norm. In fact, the following proposition shows that only subGaussian ones do.

**Proposition:**

$$X \sim \text{subG}(\sigma^2) \quad \text{iff} \quad \|X\|_{\psi_2} \leq c\sigma^2$$

Keep in mind that  $X$  is centered.

*Proof.* We assume that  $\sigma = 1$  without loss of generality.

Assume first that  $X \sim \text{subG}(1)$  so that

$$\mathbb{E}[e^{sX}] \leq e^{\frac{s^2}{2}}, \quad \forall s \in \mathbb{R}$$

Let  $S \sim \mathcal{N}(0, 3/4)$  and note that the above inequality implies that

$$\mathbb{E}[e^{\frac{3X^2}{8}}] = \mathbb{E}[\mathbb{E}[e^{SX} | X]] = \mathbb{E}[\mathbb{E}[e^{SX} | S]] \leq \mathbb{E}[e^{\frac{S^2}{2}}] = 2.$$

Therefore, we have  $\|X\|_{\psi_2}^2 \leq 4/3$ .

We now prove the converse by showing part (ii) from the proposition. Set  $s := 1/(2\|X\|_{\psi_2}^2)$

$$\mathbb{P}(|X| > t) \leq \mathbb{E}[e^{sX^2}]e^{-st^2} \leq 2e^{-\frac{t^2}{2\|X\|_{\psi_2}^2}}$$

□

## 2.1 Examples of subGaussian random variables

1. Gaussian random variables are clearly subGaussian (see, e.g., (i))
2.  $X \sim \text{Rad}(\frac{1}{2})$ , where a Rademacher variable takes value  $+1$  with probability  $\frac{1}{2}$  and value  $-1$  with probability  $\frac{1}{2}$ . This can be seen by checking (iv) for example:

$$\mathbb{E}[e^{sX^2}] = e^s.$$

3. Bounded support  $|X| \leq a$  a.s. We can use (iv) here as well. In fact, sharp constants may be obtained in this case. This is the purpose of Hoeffding's inequality.

## 3. Hoeffding's Inequality

We begin with a lemma that gives sharp bounds for the variance proxy of random variables with bounded support.

**Lemma (Hoeffding's lemma):** Let  $X$  be a centered random variable such that  $x \in [a, b]$  a.s. Then

$$\mathbb{E}[e^{sX}] \leq e^{\frac{s^2(b-a)^2}{8}} \quad \forall s \in \mathbb{R}$$

*Proof.* Define the log-MGF (a.k.a. *cumulant generating function*):

$$\psi(s) = \log \mathbb{E}[e^{sX}].$$

Let us define a new probability measure  $\tilde{\mathbb{P}}$  on the real line by

$$\tilde{\mathbb{P}}(A) = \frac{\int_A e^{sx} \mathbb{P}(dx)}{\int e^{sy} \mathbb{P}(dy)}.$$

We can compute the derivatives of the log-MGF:

$$\psi'(s) = \frac{\mathbb{E}[Xe^{sX}]}{\mathbb{E}[e^{sX}]} = \tilde{\mathbb{E}}(x).$$

Moreover,

$$\begin{aligned}
\psi''(s) &= \frac{\mathbb{E}[X^2 e^{sX}] \mathbb{E}[e^{sX}] - \mathbb{E}[X e^{sX}] \mathbb{E}[X e^{sX}]}{\mathbb{E}[e^{sX}]^2} \\
&= \frac{\mathbb{E}[X^2 e^{sX}]}{\mathbb{E}[e^{sX}]} - \left( \frac{\mathbb{E}[X e^{sX}]}{\mathbb{E}[e^{sX}]} \right)^2 \\
&= \mathbb{E}_{\tilde{\mathbb{P}}}[X^2] - \mathbb{E}_{\tilde{\mathbb{P}}}[X]^2 \\
&= \widetilde{\text{var}}(X),
\end{aligned}$$

where  $\widetilde{\text{var}}(X)$  denotes variance of  $X$  under the probability measure  $\tilde{\mathbb{P}}$ .

Note that

$$\widetilde{\text{var}}(X) = \widetilde{\text{var}}\left(X - \frac{a-b}{2}\right) \leq \mathbb{E}\left[\left(X - \frac{a+b}{2}\right)^2\right] \leq \frac{(b-a)^2}{4}$$

Moreover, since  $\mathbb{E}[X] = 0$ , we have  $\psi'(0) = 0$  and it is not hard to check that  $\psi(0) = 0$ . By the fundamental theorem of calculus, we have

$$\psi(s) = \int_0^s \int_0^r \psi''(t) dt dr \leq \int_0^s \int_0^r \frac{(b-a)^2}{4} dt dr = \frac{s^2(b-a)^2}{8}$$

□

We are now in a position to state the main result of this section.

**Theorem (Hoeffding's inequality):** Let  $X_1, \dots, X_n$  be independent random variables such that  $X_i \sim \text{subG}(\sigma_i^2)$ . Then for any  $t > 0$ , we have

$$\mathbb{P}[\bar{X}_n > t] \vee \mathbb{P}[\bar{X}_n < -t] \leq \exp\left(-\frac{n^2 t^2}{2 \sum_{i=1}^n \sigma_i^2}\right)$$

In particular, if  $X_i \in [a_i, b_i]$  a.s., we have

$$\mathbb{P}[|\bar{X}_n| > t] \leq 2 \exp\left(-\frac{2n^2 t^2}{\sum_{i=1}^n (b_i - a_i)^2}\right)$$

*Proof.* We use a Chernoff bound.

$$\mathbb{P}[\bar{X}_n > t] \leq \mathbb{E}\left[e^{s \sum_{i=1}^n X_i}\right] e^{-nst} = \left[\prod_{i=1}^n \mathbb{E}\left[e^{s X_i}\right]\right] e^{-nst} \leq \left[\prod_{i=1}^n e^{\frac{\sigma_i^2 s^2}{2}}\right] e^{-nst} = e^{\frac{\sum_{i=1}^n \sigma_i^2 s^2}{2} - nst}.$$

We minimize the right-hand side with respect to  $s$ . To do this we want to minimize our exponent value of  $\frac{\sum_{i=1}^n \sigma_i^2 s^2}{2} - nst$ . Taking the derivative with respect to  $s$  and setting the derivative equal to zero gives us

$$\sum_{i=1}^n \sigma_i^2 s - nt = 0 \implies s = \frac{nt}{\sum_{i=1}^n \sigma_i^2}$$

Plugging in this value of  $s$  into our exponent gives us

$$\frac{\sum_{i=1}^n \sigma_i^2 \left( \frac{nt}{\sum_{i=1}^n \sigma_i^2} \right)^2}{2} - n \left( \frac{nt}{\sum_{i=1}^n \sigma_i^2} \right) t = \frac{n^2 t^2}{2 \sum_{i=1}^n \sigma_i^2} - \frac{n^2 t^2}{\sum_{i=1}^n \sigma_i^2} = -\frac{n^2 t^2}{2 \sum_{i=1}^n \sigma_i^2}$$

This yields

$$\mathbb{P}[\bar{X}_n > t] \leq \exp\left(-\frac{n^2 t^2}{2 \sum_{i=1}^n \sigma_i^2}\right).$$

The same bound on  $\mathbb{P}[\bar{X}_n < -t]$  may be obtained using the same method.

Finally, the second bound by observing that Hoeffding's lemma yields

$$X_i \sim \text{subG}\left(\frac{(b_i - a_i)^2}{4}\right)$$

and applying a union bound. □

**Summary: SubGaussian random variables.** A centered r.v.  $X$  is sub-Gaussian with variance proxy 1 if either of the following equivalent definitions hold

- (i)  $\mathbb{E}[e^{sX}] \leq e^{c_1 s^2}$  for all  $s \in \mathbb{R}$ ,
- (ii)  $\mathbb{P}[|X| \geq t] \leq 2e^{-c_2 t^2}$  for all  $t \geq 0$ ,
- (iii)  $\|X\|_p = (\mathbb{E}|X|^p)^{\frac{1}{p}} \leq c_3 \sqrt{p}$  for all  $p = 1, 2, \dots$ ,
- (iv)  $\mathbb{E}[e^{sX^2}] \leq e^{c_4 s}$  for all  $s \in (0, c'_4)$ ,
- (v)  $\|X\|_{\psi_2} \leq c_5$

where  $\|X\|_{\psi_2} = \inf \{t > 0, \mathbb{E}[e^{\frac{X^2}{2t^2}}] \leq 2\}$  is the  $\psi_2$  Orlicz norm of  $X$ .

**Sums of independent subGaussian random variables.** Let  $X_1, \dots, X_n$  be independent random variables such that  $X_i \sim \text{subG}(\sigma_i^2)$ . Then for any  $t > 0$ , we have

$$\mathbb{P}[|\bar{X}_n| > t] \leq 2 \exp\left(-\frac{n^2 t^2}{2 \sum_{i=1}^n \sigma_i^2}\right)$$

**Hoeffding's lemma.** If  $X \in [a, b]$  a.s., then  $X \sim \text{subG}\left(\frac{(b - a)^2}{4}\right)$

**Hoeffding's inequality.** Let  $X_1, \dots, X_n$  be independent copies of  $X_i \sim \text{subG}(\sigma^2)$ . Then for any  $t > 0$ , we have

$$\mathbb{P}[|\bar{X}_n| > t] \leq 2 \exp\left(-\frac{nt^2}{2\sigma^2}\right)$$

In particular, if  $X \in [a, b]$  a.s., then

$$\mathbb{P}[|\bar{X}_n| > t] \leq 2 \exp\left(-\frac{2nt^2}{(b-a)^2}\right)$$