Statistics 210B Lecture 3 Notes

Daniel Raban

January 25, 2022

1 Sub-Gaussian and Sub-Exponential Random Variables

1.1 Sub-Gaussian random variables

Last time, we used Chernoff's inequality to get an upper bound on the tail probability of $\frac{1}{n}\sum_{i=1}^n Z_i - \mu$, where Z_i are iid and supported in $[0,1]$. We made a claim about the moment generating function of such random variables:

$$\mathbb{E}[e^{\lambda(Z-\mathbb{E}[Z])}] \leq e^{\lambda^2/2}.$$

We can abstract this into a definition:

Definition 1.1. A random variable with $\mu = \mathbb{E}[X]$ is **σ-sub-Gaussian**[1] if there is a positive number $\sigma 0$ such that

$$\mathbb{E}[e^{\lambda(X-\mu)}] \leq e^{\lambda^2\sigma^2/2} \qquad \forall \lambda \in \mathbb{R}.$$

Combining with Chernoff's inequality, we have that if X is σ-sub-Gaussian, then

$$\begin{aligned}\mathbb{P}(X-\mu \geq t) &\leq \inf_\lambda \frac{\mathbb{E}[e^{\lambda(X-\mu)}]}{e^{\lambda t}} \\ &\leq \inf e^{\lambda\sigma^2/2-\lambda t}\end{aligned}$$

This quadratic function in the exponent is minimized at $\lambda = t/\sigma^2$:

$$\begin{aligned}&= e^{(t/\sigma^2)^2\cdot\sigma^2/2-t^2/\sigma^2} \\ &= e^{-t^2/(2\sigma^2)}.\end{aligned}$$

Why is this called "sub-Gaussian"?

[1] Some textbooks call this σ^2-sub-Gaussian, and you should think of σ as a surrogate for variance.

(a) If $G \sim N(\mu, \sigma^2)$, then

$$\mathbb{E}[e^{\lambda(G-\mu)}] = \int_{-\infty}^{\infty} e^{\lambda(x-\mu)} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \, dx$$

We can combine the exponentials and complete the square in the exponent to solve this integral.

$$= e^{\lambda^2\sigma^2/2}.$$

(b) If $G \sim N(0,1)$, then

$$\lim_{t\to\infty} \frac{\mathbb{P}(G \geq t)}{\underbrace{\frac{1}{t}\frac{1}{\sqrt{2\pi}}\exp(-t^2/2)}_{\phi(t)}} = 1.$$

In addition, if ϕ is the standard Gaussian probability density function, then

$$\frac{1}{t}\phi(t) \leq \mathbb{P}(G \geq t) \leq \left(\frac{1}{t} - \frac{1}{t^3} + \frac{3}{t^5}\right)\phi(t).$$

This is exercise 2.2 in Wainwright's textbook. To prove this, first show that $\phi(z) = -\frac{\phi'(z)}{z}$. Next, calculate $\int_t^\infty \phi(z)\,dz = \int_t^\infty -\frac{\phi'(z)}{z}\,dz$ by using integration by parts.

1.2 Hoeffding's inequality

Proposition 1.1 (Hoeffding's inequality)**.** *Suppose X_i, $i = 1, \ldots, n$ are independent, where X_i has mean μ and is σ_i-sub-Gaussian. Then*

1. *$\sum_{i=1}^n X_i$ has mean $\sum_{i=1}^n \mu_i$ and is sub-Gaussian with parameter $\sqrt{\sum_{i=1}^n \sigma_i^2}$.*
2. $$\mathbb{P}\left(\sum_{i=1}^n (X_i - \mu_i) \geq t\right) \leq \exp\left(-\frac{t^2}{2\sum_{i=1}^n \sigma_i^2}\right).$$

Proof.

1. $$\mathbb{E}[e^{\lambda\sum_{i=1}^n (X_i-\mu_i)}] = \mathbb{E}\left[\prod_{i=1}^n e^{\lambda(X_i-\mu_i)}\right]$$
$$= \prod_{i=1}^n \mathbb{E}[e^{\lambda(X_i-\mu_i)}]$$

$$
\begin{aligned}
&\leq \prod_{i=1}^{n} e^{\lambda\sigma_i^2/2} \\
&= e^{\lambda^2(\sum_{i=1}^{n}\sigma_i^2)/2}.
\end{aligned}
$$

2. The second statement is by Chernoff's inequality, as above. □

Let $(X_i)_{i\in[n]} \overset{oniid}{\sim} X$ be σ-sub-Gaussian. Then

$$
\begin{aligned}
\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mu \geq t\right) &= \mathbb{P}\left(\sum_{i=1}^{n}(X_i - \mu) \geq nt\right) \\
&\leq \exp\left(-\frac{(nt)^2}{2n\sigma^2}\right) \\
&= \exp\left(-\frac{nt^2}{2\sigma^2}\right).
\end{aligned}
$$

(a) How do we extract the order of $\frac{1}{n}\sum_{i=1}^{n} X_I - \mu$? Let $\delta = \exp(-\frac{nt^2}{2\sigma^2}$ and solve for t to get $t = \sigma\sqrt{\frac{2\log(1/\delta)}{n}}$. Thus,

$$
\frac{1}{n}\sum_{i=1}^{n} X_i \leq \mu + \sigma\sqrt{\frac{2\log(1/\delta)}{n}} \qquad \text{with probability at least } 1-\delta.
$$

To check for mistakes, look at the units: X_i, μ, and σ have the same units, while δ and n are unitless. Here, we can see that the units match up.

(b) How many samples are needed to that $\frac{1}{n}\sum_{i=1}^{n} X_i - \mu \leq t$ with probability $1-\delta$? Let $\delta = \exp(-\frac{nt^2}{2\sigma^2})$, and solve for n to get $n = \frac{2\sigma^2}{t^2}\log(1/\delta)$.

1.3 Examples of sub-Gaussian random variables

Example 1.1 (Rademacher random variables)**.** Consider a **Rademacher random variable** $\varepsilon \sim \mathrm{Unif}(\{\pm 1\})$. ε is 1-sub-Gaussian.

Proof.

$$
\mathbb{E}[e^{\lambda\varepsilon}] = \frac{1}{2}e^{\lambda} + \frac{1}{2}e^{-\lambda}
$$

We want to upper bound this by $e^{\lambda^2/2}$. One way is to use the Taylor expansion:

$$
= \frac{1}{2}\sum_{k=1}^{\infty}\frac{\lambda^k}{k!} + \frac{(-\lambda)^k}{k!}
$$

$$= \sum_{k=0}^{\infty} \frac{\lambda^{2k}}{(2k!)}.$$

If we take the Taylor expansion of $e^{\lambda^2/2}$, we get $1 + \sum_{k=1}^{\infty} \lambda^{2k} 2^k k!$. To compare the Taylor expansions, we only need to show that $(2k)! \geq 2^k k!$. $\square$

Example 1.2 (Bounded random variable)**.** Let $X \in \mathcal{P}([a,b])$. We claim that X is $(b-a)$-sub-Gaussian.[2]

Proof. Instead of a direct calculation, we use a series of tricks.

Trick 1: Let $X' \overset{d}{=} X$ with X, X' independent. Then

$$\mathbb{E}_X[e^{\lambda(X-\mu)}] = \mathbb{E}_X[e^{\lambda X - \mathbb{E}_X[X']}]$$

Trick 2: Use Jensen's inequality to get $e^{-\lambda \mathbb{E}[X']} \leq \mathbb{E}[e^{-\lambda X'}]$. This gives

$$\leq \mathbb{E}_{X,X'} \mathbb{E}[e^{\lambda(X-X')}]$$

Trick 3: Introduce $\varepsilon \sim \mathrm{Unif}(\{\pm 1\})$ with ε independent of (X, X'). Then $\varepsilon(X - X') \overset{d}{=} X - X'$.

$$= \mathbb{E}_{\varepsilon,X,X'} \mathbb{E}[e^{\lambda\varepsilon(X-X')}]$$

Using the tower property of conditional expectation,

$$= \mathbb{E}_{X,X'}[\mathbb{E}_{\varepsilon}[e^{\lambda\varepsilon(X-X')} \mid X, X']]$$

By the 1-sub-Gaussianity of ε,

$$\leq \mathbb{E}_{X,X'}[e^{\lambda^2(X-X')^2/2}]$$

Since $(X - X') \leq (b-a)^2$ by the boundedness of X, X',

$$\leq e^{\lambda^2(b-a)^2/2}.$$ $\square$

Remark 1.1. These tricks will be useful in later lectures and in statistics research. This technique is known as **symmetrization**.

1.4 Equivalent characterizations of sub-Gaussianity

Here are some

Theorem 1.1 (HDP 2.6 or RV 2.5.1)**.** *Let X be a random variable. Then the following are equivalent:*

(i) The tails of X satisfy

$$\mathbb{P}(|X| \geq t) \leq 2\exp\left(-\frac{t^2}{\kappa_1^2}\right) \qquad \forall t \geq 0.$$

[2] It is actually possible to show this with parameter $(b-a)/2$, but we will not show this fact in this lecture.

(ii) The moments of X satisfy

$$\|X\|_{L^p} = (\mathbb{E}[|X^p|])^{1/p} \leq \kappa_2 \sqrt{p}, \qquad \forall p \geq 1.$$

(iii) The moment generating function of X^2 satisfies

$$\mathbb{E}[\exp(\lambda^2 X^2)] \leq \exp(\kappa_3^2 \lambda^2) \qquad \forall \lambda \textit{ such that } |\lambda \leq \frac{1}{\kappa_3}.$$

(iv) The moment generating function of X^2 is bounded at some point:

$$\mathbb{E}[\exp(X^2/\kappa_4^2)] \leq 2.$$

Moreover, if $\mathbb{E}[X] = 0$, then properties (i)-(iv) are also equivalent to

5. *The moment generating function of X satisfies*

$$\mathbb{E}[\exp(\lambda X)] \leq \exp(\kappa_5^2 \lambda^2 / 2) \qquad \forall \lambda \in \mathbb{R}.$$

Here, $\kappa_1, \ldots, \kappa_5$ are universal constants.

Proof. Proof is an exercise. □

Remark 1.2. Some people define sub-Gaussian through property (i) instead of (v). It can also be defined in terms of **Orlicz norms**, which are covered in an exercise in Wainwright's book. We use the moment generating function definition because a *tensorization* property will be important to us later.

Proposition 1.2. *There is a universal constant κ such that if X is σ-sub-Gaussian and Z is a random variable bounded by 1, then ZX is $\kappa\sigma$-sub-Gaussian.*

Remark 1.3. Z and X can be dependent!

Proof. We can use any of the characterizations (i), (ii), (ii), (iv) to prove this. (v) doesn't work as easily. □

1.5 Sub-exponential random variables

Let $G \sim N(0, 1)$. Then G^2 is not sub-Gaussian. This is because $\mathbb{E}[G^2] = 1$, and

$$\begin{aligned}
\mathbb{E}[e^{\lambda(G^2-1)}] &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\lambda(z^2-1)} e^{-z^2/2} \, dz \\
&= \begin{cases} \frac{e^{-\lambda}}{\sqrt{1-2\lambda}} & \lambda < 1/2 \\ \infty & \lambda \geq 1/2. \end{cases}
\end{aligned}$$

We can still derive a good but weaker tail bound for this kind of random variable.

Definition 1.2. A random variable X is (ν, α)**-sub-exponential** if

$$\mathbb{E}[e^{\lambda(X-\mu)}] \leq e^{\lambda^2\nu^2/2} \qquad \forall|\lambda| \leq \frac{1}{\alpha}.$$

We can see from this definition that sub-Gaussian variables are sub-exponential with any $\alpha > 0$.

Example 1.3. If $G \sim N(0,1)$, then G^2 is $(2,4)$-sub-exponential.

Proof. We want to show that

$$\mathbb{E}[e^{\lambda(G^2-1)}] = \frac{e^{-\lambda}}{\sqrt{1-2\lambda}} \leq e^{2\lambda^2} \qquad \forall|\lambda| \leq \frac{1}{4}.$$

we can do this by comparing Taylor series. □

Combining this definition with Chernoff's inequalitiy, we have that if X is (ν, α)-sub-exponential, then

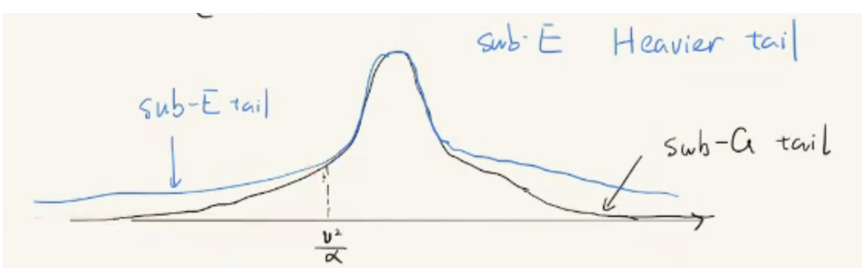
$$\begin{aligned}
\mathbb{P}(X - \mu \geq t) &\leq \inf_{\lambda} \frac{\mathbb{E}[e^{\lambda(X-\mu)}]}{e^{\lambda t}} \\
&\leq \inf_{|\lambda|\leq 1/\alpha} \frac{e^{\nu^2\lambda^2/2}}{e^{\lambda t}} \\
&= \exp\left(\inf_{\lambda \leq 1/\alpha} \nu^2\lambda^2/2 - \lambda t\right)
\end{aligned}$$

If this interval contains $\lambda = t/\nu^2$, then this is the minimum. Otherwise, the minimum will be on the boundary.

$$= \begin{cases} \exp(-\frac{t^2}{2\nu^2}) & \text{if } \frac{t}{\nu^2} \leq \frac{1}{\alpha} \\ \exp(\frac{\nu^2}{2\alpha^2} - \frac{t}{\alpha}) & \text{if } \frac{t}{\nu^2} > \frac{1}{\alpha} \end{cases}$$

The second expression is $\leq \exp(-\frac{t}{\nu^2}\frac{\nu^2}{2\alpha} - \frac{t}{\alpha}) = \exp(-\frac{t}{2\alpha})$. So we can write this as

$$\leq \exp\left(-\min\left\{\frac{t^2}{2\nu^2}, \frac{t}{2\alpha}\right\}\right).$$



Why is this called "sub-exponential?

(a) If $Z \sim \text{Exp}(1/\alpha)$, then
$$\mathbb{P}(Z \geq t) = \exp\left(-\frac{t}{\alpha}\right).$$

(b) Exp(1) is $(\sqrt{2}, 2)$-sub-exponential: If $Z \sim \text{Exp}(1)$, then
$$Z \overset{d}{=} \frac{1}{2}(G_1^2 + G_2^2), \qquad G_1, G_2 \overset{\text{iid}}{\sim} N(0,1).$$

Then
$$\begin{aligned}\mathbb{E}[e^{\lambda(Z-1)}] &= \mathbb{E}[e^{\frac{\lambda}{2}(G_1^2+G_2^2-2)}] \\ &= \mathbb{E}[e^{\frac{\lambda}{2}(G_1^2-1)}]\,\mathbb{E}[e^{\frac{\lambda}{2}(G_2^2-1)}]\end{aligned}$$

for $|\lambda| \leq 1/2$,
$$\begin{aligned}&\leq \frac{e^{-\lambda}}{1-\lambda} \\ &\leq e^{\lambda^2}.\end{aligned}$$