Lecture 13

Sub-exponential Random Variables (cont.)

Recall from our last lecture that sub-exponential random variables are defined as:

Definition 1. *A random variable X with mean $\mu = \mathbf{E}[X]$ is sub-exponential if there are non-negative parameters (ν^2, α) such that*

$$\mathbf{E}\left[e^{\lambda(X-\mu)}\right] \leq e^{\nu^2\lambda^2/2} \qquad \text{for all } \lambda \text{ for which } |\lambda| \leq \frac{1}{\alpha}\,.$$

We write $X \in \mathsf{subE}\,(\nu^2, \alpha)$.

Here, we start off by focusing on some of the properties of sub-exponentials.

Scaling: Sub-exponentiality is closed under scaling: for any scalar c, if a random variable X is in $\mathsf{subE}\,(\nu^2, \alpha)$, then $c\,X$ is in $\mathsf{subE}\,((c\,\nu)^2, |c| \cdot \alpha)$.

Summation: Moreover, the sum of two independent sub-exponential random variables is a sub-exponential random variable. In particular, if X_1 and X_2 are two independent sub-exponential random variables in $\mathsf{subE}\,(\nu_1^2, \alpha_1)$ and $\mathsf{subE}\,(\nu_2^2, \alpha_2)$ respectively, then we have:

$$X_1 + X_2 \in \mathsf{subE}\left(\nu_1^2 + \nu_2^2,\ \max\{\alpha_1, \alpha_2\}\right)\,.$$

In general, we have the following lemma for the sum of n sub-exponential random variables.

Lemma 2. *Let $X_1, X_2, ..., X_n$ be n independent random variables with $\mathbf{E}[X_i] = \mu_i$ and $X_i \in \mathsf{subE}\,(\nu_i^2, \alpha_i)$ for $i = 1, 2, ..., n$. Then, we have:*

$$\sum_{i=1}^{n} X_i \in \mathsf{subE}\left(\sum_{i=1}^{n} \nu_i^2,\ \max_{i\in[n]} \alpha_i\right)\,.$$

Proof. To prove the lemma, we will bound the moment generating function of the sum. We show that for any λ such that $|\lambda| < \frac{1}{\max_i \alpha_i}$, we have:

$$\mathbf{E}\left[\exp\left(\lambda\sum_{i=1}^{n}(X_i-\mu_i)\right)\right]\leq\exp\left(\left(\sum_{i=1}^{n}\nu_i^2\right)\lambda^2/2\right).$$

Since the random variables $X_1, X_2, ..., X_n$ are independent, the expectation of the product is equal to the product of the expectations. Hence we have:

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

Now, we can use the fact that each X_i is sub-exponential. Since $|\lambda|<\frac{1}{\max_i\alpha_i}\leq\frac{1}{\alpha_i}$ for all i, we can apply the MGF bound to each term in the product:

$$\prod_{i=1}^{n}\mathbf{E}[\exp\left(\lambda(X_i-\mu_i)\right)]\leq\prod_{i=1}^{n}\exp\left(\nu_i^2\lambda^2/2\right)=\exp\left(\left(\sum_{i=1}^{n}\nu_i^2\right)\lambda^2/2\right).$$

Therefore, we have shown that for any $|\lambda|<\frac{1}{\max_i\alpha_i}$:

$$\mathbf{E}\left[\exp\left(\lambda\sum_{i=1}^{n}(X_i-\mu_i)\right)\right]\leq\exp\left(\left(\sum_{i=1}^{n}\nu_i^2\right)\lambda^2/2\right),$$

implying the statement of the lemma. □

From this theorem and the scaling property, the following property is immediately implied:

Corollary 3. *Suppose $X_1, ..., X_n$ are independent zero-mean random variables with $X_i \in \mathsf{subE}(1,1)$. Let $a = (a_1, ..., a_n) \in \mathbb{R}^n$. Then,*

$$\sum_{i=1}^{n}a_iX_i\in\mathsf{subE}\left(\|a\|_2^2,\|a\|_\infty\right).$$

Bernstein's Inequality

Theorem 1. *Suppose $X_1, ..., X_n$ are independent zero-mean random variables with $X_i \in \mathsf{subE}(1,1)$. Let $a = (a_1, ..., a_n) \in \mathbb{R}^n$. Then, we have:*

$$\mathbf{Pr}\left[\left|\sum_{i=1}^{n}a_iX_i\right|\geq t\right]\leq 2\exp\left(-\min\left(\frac{t^2}{2\|a\|_2^2},\frac{t}{2\|a\|_\infty}\right)\right).$$

In particular, when $a_i = \frac{1}{n}$, we get:

$$\mathbf{Pr}\left[\left|\frac{1}{n}\sum_{i=1}^{n} X_i\right| \geq t\right] \leq 2\exp\left(-n\min\left(\frac{t^2}{2}, \frac{t}{2}\right)\right) .$$

Proof. By Corollary 3, $\sum_{i=1}^{n} a_i X_i \in \mathsf{subE}\left(\|a\|_2^2, \|a\|_\infty\right)$. The tail bound follows directly from the lemma we proved in our previous lecture. □

An interesting observation arises when comparing the tail bound of a single sub-exponential random variable to the above tail bound for the average. For a single sub-exponential random variable $X \in \mathsf{subE}(1,1)$, the tail bound for $t \geq 1$ is given by:

$$\mathbf{Pr}\left[\left|\frac{1}{n}X\right| \geq t\right] = \mathbf{Pr}\big[|X| \geq n \cdot t\big] \leq 2\exp\left(-\min\left(n^2t^2/2, nt/2\right)\right) = 2\exp\left(-nt/2\right) .$$

This bound is identical to the tail bound for the average. Intuitively, these bounds suggest that the probability of the average of n sub-exponential random variables exceeding t (for $t \geq 1$) is comparable to the probability of a single sub-exponential random variable exceeding $n \cdot t$. This highlights a key characteristic of sub-exponential random variables: that their large deviations are typically driven by a single, exceptionally large observation rather than by the cumulative effect of many moderate ones.

Comparison with Central Limit Theorem (CLT): It is natural to wonder if the CLT could provide a tighter bound on the tail probabilities of sums of independent sub-exponential random variables. The CLT states that for large n, the distribution of the normalized sum $\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i$ approaches a standard Gaussian distribution:

$$Y_n \coloneqq \sqrt{n}\left(\frac{\sum_{i=1}^{n} X_i/n}{\sigma}\right) \xrightarrow{d} \mathcal{N}(0,1) ,$$

where $\xrightarrow{d}$ denotes convergence in distribution.

If we assume $\sigma = 1$ for simplicity, we might expect a tail bound of the form:

$$\mathbf{Pr}\left[\left|\frac{1}{n}\sum_{i=1}^{n} X_i\right| \geq t\right] = \mathbf{Pr}\left[\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i\right| \geq t\sqrt{n}\right] \approx \mathbf{Pr}\big[Y_n > \sqrt{n}\,t\big] \overset{?}{\leq} \exp\left(-n\,t^2/2\right) . \quad (1)$$

However, this line of reasoning is not entirely correct. The CLT describes the limiting behavior of the distribution of the normalized sum for a fixed t, but it does not provide precise tail bounds when t itself grows with n. More formally, for any fixed t, we have:

$$\lim_{n\to\infty} \mathbf{Pr}\big[Y_n > t\big] = \mathbf{Pr}_{Z\sim\mathcal{N}(0,1)}\big[Z > t\big] .$$

But we cannot directly use this limit to obtain a tail bound when t depends on n, as in Equation (1). Therefore, while the CLT offers valuable insights into the asymptotic behavior of sums of independent random variables (and for a very large category of them), it does not necessarily yield tighter tail bounds than Bernstein's inequality for sub-exponentials.

Bernstein's Condition

In this section, we explore a more general condition that implies sub-exponentiality, known as the Bernstein condition. This condition is particularly useful when dealing with random variables that have low variance but potentially heavy tails or wide range (as we saw in an example before).

Definition 4 (Bernstein Condition)**.** *A random variable X with mean μ is said to satisfy the Bernstein condition with parameter b if for all $i \geq 3$:*

$$\mathbf{E}\left[(X-\mu)^i\right] \leq \frac{1}{2}\, i!\, \sigma^2\, b^{i-2}$$

where $\sigma^2 = Var(X)$.

This condition essentially bounds the centered moments of the random variable. The following lemma establishes the connection between the Bernstein condition and sub-exponentiality:

Lemma 5. *If X satisfies the Bernstein condition with parameter b, then $X \in \mathsf{subE}\,(2\,(\sigma)^2, 2\,b)$.*

Proof. For a random variable X satisfying the Bernstein condition with parameter b, we want to bound the MGF of $X-\mu$, where $\mu = \mathbf{E}[X]$. Using the Taylor expansion of the exponential function, we have:

$$\mathbf{E}\left[e^{\lambda(X-\mu)}\right] = \mathbf{E}\left[\sum_{i=0}^{\infty} \frac{(\lambda(X-\mu))^i}{i!}\right].$$

Expanding the first three terms of the sum and using linearity of expectation, we get:

$$\mathbf{E}\left[e^{\lambda(X-\mu)}\right] = 1 + \lambda \mathbf{E}[X-\mu] + \frac{\lambda^2}{2}\mathbf{E}\left[(X-\mu)^2\right] + \sum_{i=3}^{\infty} \frac{\lambda^i}{i!}\mathbf{E}\left[(X-\mu)^i\right].$$

Since $\mathbf{E}[X-\mu] = 0$ and $\mathbf{E}[(X-\mu)^2] =$ $VarX = \sigma^2$, we can simplify this to:

$$\mathbf{E}\big[e^{\lambda(X-\mu)}\big] = 1 + \frac{\lambda^2\sigma^2}{2} + \sum_{i=3}^{\infty} \frac{\lambda^i}{i!}\mathbf{E}\big[(X-\mu)^i\big] .$$

Now, we can use the Bernstein condition to bound the higher-order moments $\mathbf{E}\big[(X-\mu)^i\big]$ for $i \geq 3$:

$$\begin{aligned}\mathbf{E}\big[e^{\lambda(X-\mu)}\big] &\leq 1 + \frac{\lambda^2\sigma^2}{2} + \sum_{i=3}^{\infty} \frac{\lambda^i}{i!} \cdot \frac{1}{2}\, i!\, \sigma^2\, b^{i-2} = 1 + \frac{\lambda^2\sigma^2}{2} + \frac{\lambda^2\sigma^2}{2}\sum_{i=3}^{\infty} |\lambda|^{i-2} b^{i-2} \\ &= 1 + \frac{\lambda^2\sigma^2}{2}\left(1 + \sum_{i=1}^{\infty} |\lambda\, b|^i\right) = 1 + \frac{\lambda^2\sigma^2}{2}\left(\sum_{i=0}^{\infty} |\lambda\, b|^i\right) .\end{aligned}$$

For $|\lambda\, b| < 1$, the geometric series converges, and we have:

$$\mathbf{E}\big[e^{\lambda(X-\mu)}\big] = 1 + \frac{\lambda^2\sigma^2}{2} \cdot \frac{1}{1 - |\lambda b|}$$

If we further assume $|\lambda\, b| < \frac{1}{2}$, we can simplify this to:

$$\mathbf{E}\big[e^{\lambda(X-\mu)}\big] \leq 1 + \lambda^2\sigma^2 \leq \exp(\lambda^2\sigma^2)$$

This shows that X satisfies the sub-exponential condition with parameters $(2\,\sigma^2, 2\,b)$, completing the proof. □

Bernstein's Inequality for Bounded Variables

The Bernstein condition is particularly useful for bounded random variables.

Theorem 2. *Let X be a random variable with mean μ such that $|X - \mu| < B$. Then, X satisfies the Bernstein condition with parameter $b = \frac{B}{3}$.*

Proof. The proof involves showing that the centered moments of X can be bounded by B. Specifically, for any $i \geq 3$ we have:

$$\begin{aligned}\mathbf{E}\big[(X-\mu)^i\big] &\leq \mathbf{E}\Big[|X-\mu|^{i-2} \cdot (X-\mu)^2\Big] \leq B^{i-2} \cdot \mathbf{E}\Big[|X-\mu|^{i-2}\Big] \\ &\leq B^{i-2} \cdot \sigma^2 \leq \sigma^2 B^{i-2}\left(\frac{i!}{2 \cdot 3^{i-2}}\right)\end{aligned}$$

To see why the last inequality holds, we can show that $\left(\frac{i!}{2\cdot 3^{i-2}}\right)$ is at least one by an inductive argument. For $i = 3$, $i!/(2 \cdot 3^{i-1}) = 3!/6 \geq 1$. And for any $i > 3$, we have:

$$\frac{i!}{2 \cdot 3^{i-2}} = \frac{i}{3} \cdot \frac{(i-1)!}{2 \cdot 3^{(i-1)-2}} \geq 1\,.$$

Thus, we get:

$$\mathbf{E}\big[(X-\mu)^i\big] \leq \sigma^2 B^{i-2} \left(\frac{i!}{2 \cdot 3^{i-2}}\right) \leq \frac{1}{2}\, i!\, \sigma^2 \left(\frac{B}{3}\right)^{i-2}\,.$$

Hence, the proof is complete. ☐

Using this result, we can derive a version of Bernstein's inequality specifically tailored for bounded random variables:

Theorem 3. *Let $X_1, X_2, ..., X_n$ be independent random variables with $\mathbf{E}[X_i] = \mu$, $\mathbf{Var}[X_i] = \sigma^2$, and range $|X_i - \mu| \leq B$. Then,*

$$\mathbf{Pr}\left[\left|\sum_{i=1}^{n}(X_i-\mu)\right| \geq t\right] \leq 2\exp\left(-\frac{t^2/2}{n\sigma^2 + Bt/3}\right)$$

or, in the normalized version,

$$\mathbf{Pr}\left[\left|\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)\right| \geq t\right] \leq 2\exp\left(-\frac{nt^2/2}{\sigma^2 + Bt/3}\right)$$

Proof. Ignoring the constant factors, this follows directly from Bernstein's inequality and the fact that bounded variables satisfy the Bernstein condition, and hence they are sub-exponential. ☐

This version of Bernstein's inequality is particularly useful in scenarios where we have bounded random variables with variance much smaller than their range. As we saw previously, this bound implies that we have a fast-dropping tail that resembles the behavior of a Gaussian around the mean, where $t \ll \sigma^2/B$. The heavy tail kicks in later when we get further from the mean.

Bibliographic Note

The content of this lecture was derived from Section 2.7 of [Ver18] and the lecture notes of Prof. Sasha Rakhlin for "Mathematical Statistics: A Non-Asymptotic Approach", which can be found here [Rak22].

References

[Rak22] Alexander Rakhlin. Mathematical statistics: A non-asymptotic approach, 2022. Lecture notes for MIT course IDS.160, Spring 2022.

[Ver18] Roman Vershynin. *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge University Press, 2018.