COMP 585 **Spring 2025**

Probabilistic Toolkit for Learning and Computing *Instructor: Maryam Aliakbarpour*

February 27, 2025 *Scriber: Maryam Aliakbarpour*

Lecture 11

Sub-Exponential Random Variables

In the previous lecture, we explored sub-Gaussian random variables and observed their limitations. Specifically, we encountered simple random variables that did not fall under the sub-Gaussian category. Additionally, we saw that Hoeffding's Lemma, while useful, has limitations due to its invariance to variance. To address these limitations, we now introduce the concept of sub-exponential random variables. These variables offer a more nuanced approach to characterizing random variables and their concentration properties, particularly in scenarios where sub-Gaussianity proves insufficient.

Recall from the last lecture that for $|\lambda| \leq 1/4$, we have:

$$\mathbf{E}\left[e^{Z^2-1}\right] \leq e^{2\lambda^2},$$

where $Z \sim \mathcal{N}(0,1)$ was a standard normal distribution.

The definition of sub-exponential random variables is inspired by the behavior of Z^2. Like Z^2, sub-exponential random variables have moment generating functions (MGFs) that are bounded for a range of λ values. In fact, we define a sub-exponential random variable with this property:

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*

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

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

These variables also exhibit heavier tails than sub-Gaussian random variables, decaying at a rate e^{-t}, similar to that of Z^2. This slower tail decay distinguishes sub-exponential random variables from sub-Gaussian ones and makes them suitable for modeling distributions with heavier tails.

Lemma 2. *If $X \in \mathsf{subE}(\nu^2, \alpha)$, then we have:*

$$\mathbf{Pr}[X - \mathbf{E}[X] \geq t] \leq \begin{cases} e^{\frac{-t^2}{2\nu^2}} & 0 \leq t \leq \frac{\nu^2}{\alpha} \\ e^{\frac{-t}{2\alpha}} & t \geq \frac{\nu^2}{\alpha} \end{cases} \tag{1}$$

The same bound holds for the tail of $\mathbf{E}[X] - X$*. Alternatively, we can write:*

$$\begin{aligned}\mathbf{Pr}[X - \mathbf{E}[X] \geq t] &\leq \exp\left(-\min\left\{\frac{t^2}{2\nu^2}, \frac{t}{2\alpha}\right\}\right) \\ &\leq \exp\left(-\frac{t^2/2}{\nu^2 + t\alpha}\right)\end{aligned}$$

Proof. To establish this tail bound, we employ the Cramér-Chernoff method. This method leverages the fact that for any strictly increasing function f and any $x, y \in \mathbb{R}$, $x \geq y$ if and only if $f(x) \geq f(y)$. This implies that $\mathbf{Pr}[X \geq Y] = \mathbf{Pr}[f(X) \geq f(Y)]$. For any $\lambda > 0$, the function $f(x) = e^{\lambda x}$ is strictly increasing. Therefore, we get:

$$\begin{aligned}\mathbf{Pr}[X - \mathbf{E}[X] \geq t] &= \mathbf{Pr}\left[e^{\lambda(X - \mathbf{E}[X])} \geq e^{\lambda t}\right] \\ &\leq \frac{\mathbf{E}\left[e^{\lambda(X - \mathbf{E}[X])}\right]}{e^{\lambda t}} . \qquad \text{(by Markov's inequality)}\end{aligned}$$

The expected value in the numerator is the moment generating function (MGF) of the centered random variable $X - \mathbf{E}[X]$. Since X is sub-exponential, we can bound this MGF as follows:

$$\mathbf{Pr}[X - \mathbf{E}[X] \geq t] \leq \frac{\mathbf{E}\left[e^{\lambda(X - \mathbf{E}[X])}\right]}{e^{\lambda t}} \leq \frac{e^{\frac{\nu^2\lambda^2}{2}}}{e^{\lambda t}} = \exp\left(\frac{\nu^2\lambda^2}{2} - \lambda t\right) .$$

This bound holds for all $\lambda \in \left(0, \frac{1}{\alpha}\right]$. Thus, we have:

$$\mathbf{Pr}[X - \mathbf{E}[X] \geq t] \leq \inf_{\lambda \in (0, \alpha^{-1}]} \exp\left(\frac{\nu^2\lambda^2}{2} - \lambda t\right) .$$

To obtain the tightest upper bound, we minimize the exponent over this range. The exponent is a quadratic function in λ:

$$g(\lambda) \coloneqq \frac{\nu^2\lambda^2}{2} - \lambda t = \lambda\left(\frac{\nu^2}{2}\lambda - t\right) .$$

This quadratic has roots at $\lambda = 0$ and $\lambda = \frac{2t}{\nu^2}$, and it attains its minimum at the midpoint, $\lambda_{\min} = \frac{t}{\nu^2}$. Now, we consider two cases based on whether $\lambda_{\min}$ is in $(0, \alpha^{-1}]$ or not:

Case 1: $\lambda_{\min} \leq \frac{1}{\alpha}$. In this case, the minimum is within the allowed range for λ, and we have:

$$\mathbf{Pr}[X - \mathbf{E}[X] \geq t] \leq \exp\left(\frac{\nu^2\lambda_{\min}^2}{2} - \lambda_{\min} t\right) = \exp\left(-\frac{t^2}{2\nu^2}\right) .$$

Case 2: $\lambda_{\min} > \frac{1}{\alpha}$. Here, the minimum falls outside the allowed range. Since $g(\lambda)$ is decreasing for $\lambda < \lambda_{\min}$, the minimum value within the allowed range is achieved at $\lambda = \frac{1}{\alpha}$:

$$\begin{aligned}\mathbf{Pr}[X - \mathbf{E}[X] \geq t] &\leq \exp\left(\frac{\nu^2}{2\alpha^2} - \frac{t}{\alpha}\right) \leq \exp\left(\frac{t}{2\alpha} - \frac{t}{\alpha}\right) \quad (\text{using } t > \frac{\nu^2}{\alpha}) \\ &= \exp\left(-\frac{t}{2\alpha}\right).\end{aligned}$$

The above bounds together imply Equation (1). To see that the same bound applies to the other tail, $\mathbf{Pr}[\mathbf{E}[X] - X \geq t]$, we simply note that if X is sub-exponential, then so is $-X$ with the same parameters. Applying the above argument to $-X$ yields the desired bound.

The alternative form of the bound can be easily derived from the above cases, and we leave its proof as an exercise. □

This lemma reveals that sub-exponential random variables exhibit tail behavior similar to sub-Gaussian random variables near their mean. However, as we move further away from the mean, their tail behavior transitions to an e^{-t} decay. The parameter ν acts as a variance parameter, while α serves as an inverse width parameter, controlling the range over which the sub-Gaussian-like tail behavior holds.

Alternative Definitions

It is worth noting that alternative definitions of sub-exponential random variables exist (e.g., Section 2.7 in [Ver18]), using roughly the same parameters for both ν and α. For these alternative definitions, we have an analogous lemma to sub-Gaussians establishing equivalent properties.

Lemma 3 (Equivalent Properties of Sub-Exponential Random Variables)**.** *The following properties are equivalent (up to constant factors, with the C_i's differing by at most an absolute constant factor) for a random variable X:*

1. ***Tail Bound:*** *The tail probability of X satisfies*

$$\mathbf{Pr}[|X| \geq t] \leq 2\exp(-t/C_1) \qquad \textit{for all } t \geq 0\,.$$

2. ***Moment bound:*** *The moments of X satisfy*

$$\|X\|_{L_p} := (\mathbf{E}[|X|^p])^{1/p} \leq C_2\, p \qquad \textit{for all } p \geq 1\,.$$

3. ***MGF of $|X|$:*** *The moment generating function of $|X|$ satisfies the following bound:*

$$\mathbf{E}\left[e^{\lambda|X|}\right] \leq \exp(C_3\,\lambda) \qquad \textit{for all } \lambda \textit{ such that } 0 \leq \lambda \leq \frac{1}{C_3}\,.$$

4. ***MGF of $|X|$:*** *The moment generating function of $|X|$ is bounded at some point. For some C_4, we have:*
$$\mathbf{E}\big[e^{|X|/C_4}\big] \le 2\,.$$

5. ***MGF of X:*** *If X is centered ($\mathbf{E}[X] = 0$), then the moment generating function of X satisfies:*
$$\mathbf{E}\big[e^{\lambda X}\big] \le \exp\left(C_5^2\,\lambda^2\right) \qquad \text{for all } \lambda \text{ such that } |\lambda| \le \frac{1}{C_5}\,.$$

Deriving MGF Bound from Moment Bound: Proof of 2 ⇒ 5

Proof. Suppose X is a zero mean random variable with bounded moments. For all $p \ge 1$, we have:
$$\|X\|_{L_p} := (\mathbf{E}[|X|^p])^{1/p} \le C_2\,p\,.$$
Our goal is to bound the moment generating function of X. We start by the Taylor expansion of the exponential function:
$$e^x = \sum_{p=0}^{\infty} \frac{x^p}{p!}\,.$$
Using this, we can write the MGF of X as:
$$\begin{aligned}
\mathbf{E}\big[e^{\lambda X}\big] &= \mathbf{E}\left[\sum_{p=0}^{\infty} \frac{(\lambda X)^p}{p!}\right] \le \mathbf{E}\left[1 + \lambda X + \sum_{p=2}^{\infty} \frac{(\lambda X)^p}{p!}\right] \\
&= 1 + \lambda \mathbf{E}[X] + \sum_{p=2}^{\infty} \frac{\lambda^p \mathbf{E}[X^p]}{p!} = 1 + \sum_{p=2}^{\infty} \frac{\lambda^p \mathbf{E}[X^p]}{p!}\,,
\end{aligned}$$
where we used the fact that $\mathbf{E}[X] = 0$ in the last equality. Now, we can use the moment bound to bound the terms in the summation:
$$\begin{aligned}
\mathbf{E}\big[e^{\lambda X}\big] &\le 1 + \sum_{p=2}^{\infty} \frac{\lambda^p \mathbf{E}[X^p]}{p!} \le 1 + \sum_{p=2}^{\infty} \frac{\lambda^p (C_2\,p)^p}{p!} \\
&\le 1 + \sum_{p=2}^{\infty} \frac{(\lambda\,C_2\,p)^p}{(p/e)^p} && \text{(using Stirling's approximation: } p! \ge (p/e)^p) \\
&= 1 + \sum_{p=2}^{\infty} (\,C_2\,\lambda\,e)^p = 1 + \frac{(C_2\,\lambda\,e)^2}{1 - C_2\,\lambda\,e} && \text{(for } |C_2\,\lambda\,e| < 1)
\end{aligned}$$

Note that the series in the last line converges only when $|C_2 \lambda e|$ is bounded away from one. The denominator in the last term can get arbitrarily close to zero, which makes the bound useless. Here, we set λ in a way that the denominator is at most a constant. (Remember that we have control in determining the range of λ.) Let's assume λ is in a range for which we have $|C_2 \lambda e|$ at most $1/2$. Then, we obtain:

$$\begin{aligned}\mathbf{E}\left[e^{\lambda X}\right] &\leq 1 + 2(C_2\,\lambda\,e)^2 \leq \exp\left(2(C_2\,\lambda\,e)^2\right) && \left(\text{for all } |\lambda| < \tfrac{1}{2C_2 e}\right)\\ &\leq \exp\left((2\,e\,C_2)^2\ \lambda^2\right)\ .\end{aligned}$$

This shows that the MGF of X is bounded as described in Definition 5 for $C_5 = 2\,C_2\,e$. □

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.