COMP 585 **Spring 2025**

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

February 18, 2025 *Scriber: Maryam Aliakbarpour*

Lecture 9

Sub-Gaussian Random Variables

In probability theory, we often encounter random variables that exhibit *light-weight tails*, where the probability of extremely large or small values (outliers) is relatively low. This property makes them particularly useful in analyzing concentration inequalities and establishing bounds on probabilities of rare events and finds applications in various areas such as statistics, machine learning, and high-dimensional probability.

Among these, sub-Gaussian random variables are characterized by their tail behavior, which closely resembles that of Gaussian random variables.

Definition 1. *A random variable X is sub-Gaussian with variance proxy K^2 (also known as variance factor or sub-Gaussianity parameter) if*

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

We write $X \in \mathsf{subG}\,(K^2)$.

In this definition, the tail probability drops at the rate e^{-t^2/K^2}, which resembles a Gaussian random variable with mean zero and variance of $\Theta(K^2)$. It turns out there are other ways to define this kind of random variables.

Lemma 2. *[Equivalent Properties of Sub-Gaussian Random Variables] The following properties are equivalent (up to constant factors, with the K_i's differing by at most an absolute constant factor) for a random variable X:*

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

$$\mathbf{Pr}[|X| \geq t] \leq 2\exp(-t^2/K_1^2)$$

for all $t \geq 0$.

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

$$\|X\|_{L_p} := (\mathbf{E}[|X|^p])^{1/p} \leq K_2\sqrt{p}$$

for all $p \geq 1$.

3. ***MGF of X^2:*** *The moment generating function (MGF) of X^2 satisfies the following bound: There exists $K_3 > 0$ such that, for all λ with $|\lambda| \leq \frac{1}{K_3}$, we have*

$$\mathbf{E}\left[e^{\lambda^2 X^2}\right] \leq \exp\left(K_3^2 \lambda^2\right).$$

4. ***MGF of X^2:*** *The moment generating function of X^2 is bounded at some point:*

$$\mathbf{E}\left[e^{X^2/K_4^2}\right] \leq 2,$$

for some K_4.

5. ***MGF of X:*** *If X is centered ($\mathbf{E}[X] = 0$), then the moment generating function of X satisfies*

$$\mathbf{E}\left[e^{\lambda X}\right] \leq \exp\left(K_5^2 \lambda^2\right),$$

for all $\lambda \in \mathbb{R}$.

We use the Θ notation in $X \in \mathsf{subG}\left(\Theta\left(K^2\right)\right)$ to emphasize that the sub-Gaussianity parameter can vary by constant factors.

Scaling: Sub-Gaussianity is closed under scaling: if a random variable X is sub-Gaussian with parameter K, then $c\,X$ is sub-Gaussian with parameter $c\,K$.

Summation: Moreover, the sum of two sub-Gaussian random variables is a sub-Gaussian random variable (you will prove this in the problem set). In particular, if X_1 and X_2 are two independent sub-Gaussian random variables in $\mathsf{subG}\left(K_1^2\right)$ and $\mathsf{subG}\left(K_2^2\right)$ respectively, then we have:

$$X_1 + X_2 \in \mathsf{subG}\left(K_1^2 + K_2^2\right).$$

Examples

We provide a few examples of sub-Gaussian random variables.

Gaussian: Gaussian random variables are indeed sub-Gaussian too. In particular, we have the following lemma about the tail bound of a standard normal random variable:

Lemma 3 (Tails of the Normal Distribution, Proposition 2.1.2 in [Ver18])**.** *Suppose $Z \sim \mathcal{N}(0,1)$ is a standard normal random variable. Then for all $t > 0$, we have:*

$$\left(\frac{1}{t} - \frac{1}{t^3}\right)\frac{1}{\sqrt{2\pi}}e^{-t^2/2} \leq \mathbf{Pr}[Z \geq t] \leq \frac{1}{t\sqrt{2\pi}}e^{-t^2/2}.$$

Proof. To obtain an upper bound on the tail probability $\mathbf{Pr}[Z \geq t]$, we start by integrating the PDF:

$$\mathbf{Pr}[Z \geq t] = \frac{1}{\sqrt{2\pi}} \int_t^\infty e^{-x^2/2}\, dx.$$

We change variables by setting $x = t + y$. This gives:

$$\begin{aligned}
\mathbf{Pr}[Z \geq t] &= \frac{1}{\sqrt{2\pi}} \int_0^\infty e^{-(t+y)^2/2}\, dy \\
&= \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \int_0^\infty e^{-ty} e^{-y^2/2}\, dy \\
&\leq \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \int_0^\infty e^{-ty}\, dy,
\end{aligned}$$

where we used the fact that $e^{-y^2/2} \leq 1$. The last integral evaluates to $1/t$, so we obtain the desired upper bound:

$$\mathbf{Pr}[Z \geq t] \leq \frac{1}{t\sqrt{2\pi}} e^{-t^2/2}.$$

To obtain the lower bound, we use the following identity (which can be verified by integration by parts):

$$\int_t^\infty (1 - 3x^{-4}) e^{-x^2/2}\, dx = \left(\frac{1}{t} - \frac{1}{t^3}\right) e^{-t^2/2}.$$

Since, $1 - 3x^{-4}$ is at most one, we have:

$$\begin{aligned}
\left(\frac{1}{t} - \frac{1}{t^3}\right) \frac{1}{\sqrt{2\pi}} e^{-t^2/2} &= \frac{1}{\sqrt{2\pi}} \int_t^\infty (1 - 3x^{-4}) e^{-x^2/2}\, dx \\
&\leq \frac{1}{\sqrt{2\pi}} \int_t^\infty e^{-x^2/2}\, dx = \mathbf{Pr}[Z \geq t]\,.
\end{aligned}$$

Hence, the proof is complete. □

Given this bound, one can show that Z is a sub-Gaussian random variable as well. Using the symmetry of the Gaussian distribution, i.e., $\mathbf{Pr}[Z \leq -t] = \mathbf{Pr}[Z \geq t]$, we have for all $t \geq 0$:

$$\mathbf{Pr}[|Z| \geq t] \leq \min\left(1, 2\,\frac{1}{t\,\sqrt{2\pi}} e^{-t^2/2}\right) \leq 2\, e^{-t^2/2}$$

where the first inequality is due to the fact that the probability is always at most one, or it is bounded by the tail bound we have shown earlier. For the second inequality, note that for $t \geq 1/\sqrt{2\pi}$, the inequality is trivial. For $t < 1/\sqrt{2\pi}$, it is easy to see that the right-hand side is greater than one. In addition to this proof, you may see the plots of the two sides of the inequality here.

In general, if $Z \sim \mathcal{N}(0, \sigma^2)$, then $Z \in \mathsf{subG}\left(\Theta\left(\sigma^2\right)\right)$. This follows from the fact that $Z/\sigma \sim \mathcal{N}(0, 1)$, and applying the previous result shows that Z is sub-Gaussian with a variance proxy scaled by $\Theta(\sigma^2)$. To see the effect of scaling, see Definition 1.

Rademacher random variables: If X is a random variable taking values ± 1 with equal probability, then $X \in \mathsf{subG}\left(\Theta\left(1\right)\right)$. To see this, we can bound the moment generating function of X:

$$\begin{aligned}
\mathbf{E}\left[e^{\lambda X}\right] &= \frac{1}{2}e^{\lambda} + \frac{1}{2}e^{-\lambda} = \frac{1}{2}\sum_{k=0}^{\infty}\frac{\lambda^k}{k!} + \frac{(-\lambda)^k}{k!} && \text{(Since } \textstyle\sum_{n=0}^{\infty}\frac{x^n}{n!} = e^x \text{ for all } x \in \mathbb{R}\text{)} \\
&= \frac{1}{2}\sum_{k=0}^{\infty}\frac{\lambda^{2k}}{(2k)!} && \text{(Only even terms are non-zero.)} \\
&\le \sum_{k=0}^{\infty}\frac{(\lambda^2)^k}{2^k k!} = e^{\lambda^2/2}.
\end{aligned}$$

This shows that the moment generating function of X is bounded by $e^{(1/\sqrt{2})^2\lambda^2}$. Combined with the fact that $\mathbf{E}[X]$ is zero, Definition 5 implies that $X \in \mathsf{subG}\left(\Theta\left(1\right)\right)$.

Bounded random variables Suppose Y is a random variable that only takes two possible values: a and b, each with probability $1/2$. The centered version of Y is $Y - \mathbf{E}[Y]$ and is defined as:

$$Y' = \begin{cases} \dfrac{b-a}{2} & \text{with probability } \dfrac{1}{2} \\ \dfrac{a-b}{2} & \text{with probability } \dfrac{1}{2} \end{cases}$$

Note that Y' has the same distribution as $(b-a)\cdot X/2$, where X is the Rademacher random variable we defined earlier. Using the fact that X is a sub-Gaussian random variable, and by the scaling property of sub-Gaussians, we have $Y' \in \mathsf{subG}\left((b-a)^2/4\right)$.

More generally, we can show the same result for any bounded random variable Z in $[a, b]$ with $\mathbf{E}[Z] = 0$:

Lemma 4 (Hoeffding's Lemma)**.** *Suppose X is a zero-mean random variable in $[a, b]$. Then,*

$$\mathbf{E}\left[e^{\lambda X}\right] \le \exp\left(\frac{\lambda^2(b-a)^2}{8}\right),$$

which implies $X \in \mathsf{subG}\left((b-a)^2/8\right)$.

The proof of this lemma can be found in [Rak22], Lemma 2.

This result shows that bounded random variables with zero mean are sub-Gaussian, and their sub-Gaussianity parameter depends on the range of the variable. This is a useful property because many random variables encountered in practice are bounded.

Hoeffding's Bound

The sub-Gaussianity of bounded random variables has an interesting implication. We can show our first sophisticated concentration lemma here: Hoeffding's bound!

Theorem 1 (Hoeffding's Bound)**.** *Let $X_1, \ldots, X_n$ be i.i.d. random variables in the range $[a, b]$. Then,*

$$\mathbf{Pr}\left[\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbf{E}[X_i]\right| \geq \epsilon\right] \leq 2\exp\left(-\frac{2\epsilon^2 n}{(b-a)^2}\right).$$

Proof. Here we prove this theorem ignoring the constant factor in the exponent. Let's first focus on centering these random variables. Define $Y_i = X_i - \mathbf{E}[X_i]$. Then, Y_i is a zero-mean random variable in the interval $[a - \mathbf{E}[X_i], b - \mathbf{E}[X_i]]$. By Hoeffding's Lemma (Lemma 4), Y_i's are in $\mathsf{subG}\left((b-a)^2/8\right)$. Since $Y_1, \ldots, Y_n$ are independent, we have

$$\sum_{i=1}^{n} Y_i \in \mathsf{subG}\left(n(b-a)^2/8\right) .$$

Therefore, using the tail bound property of sub-Gaussians, we have:

$$\mathbf{Pr}\left[\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbf{E}[X_i]\right| \geq \epsilon\right] = \mathbf{Pr}\left[\left|\sum_{i=1}^{n} Y_i\right| \geq \epsilon n\right] \leq 2\exp\left(-\Theta\left(\frac{\epsilon^2 n^2}{n(b-a)^2}\right)\right) .$$

Ignoring the constant in the exponent, this bound implies the statement of the lemma. □

Bibliographic Note

The content of this lecture was derived from Section 2.5 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.