Lecture 10

Sub-Gaussian Random Variables (cont.)

Recall that in our previous lecture we introduced the equivalent definitions of sub-Gaussian random variables. In this lecture, we focus on proving some of these equivalences. The purpose of presenting this proof is to illustrate the connections between the tail bound, the moments bound, and the moment generating function (MGF). This approach shows how these properties can be manipulated to derive one from another.

Lemma 1. *[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:*
$$\mathbf{E}\left[e^{\lambda^2 X^2}\right] \leq \exp\left(K_3^2\,\lambda^2\right)\,,$$
for some K_3.

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

(Partial) Proof of Lemma 1

Deriving Moment Bound from Tail Bound: Proof of 1 ⇒ 2

Proof. We start by the integral identity of expectation. For a non-negative random variable $Y \geq 0$, we have the following integral identity:
$$\mathbf{E}[Y] = \int_0^\infty \mathbf{Pr}[Y > t]\,dt\,.$$

This is a useful tool here, since it connects the expected value to the tail bound of that random variable. To show that the tail bound implies the moment growth condition, we can use the integral identity and a change of variables. Starting with the definition of the p-th moment, we have:
$$\begin{aligned}\mathbf{E}[|X|^p] &= \int_0^\infty \mathbf{Pr}[|X|^p > t]\,dt\\ &= \int_0^\infty \mathbf{Pr}\left[|X| \geq \sqrt[p]{t}\right]dt.\end{aligned}$$

Now, we make the change of variables $u = \sqrt[p]{t}$, which implies $t = u^p$ and $dt = pu^{p-1}du$. This gives:
$$\begin{aligned}\mathbf{E}[|X|^p] &= \int_0^\infty \mathbf{Pr}[|X| \geq u] \cdot p\,u^{p-1}\,du\\ &\leq \int_0^\infty 2\,e^{-u^2/K_1^2} \cdot p\,u^{p-1}\,du,\end{aligned}$$

where we used the tail bound in the last inequality. Another change of variables $z = u^2/K_1^2$ leads to:
$$\begin{aligned}\mathbf{E}[|X|^p] &\leq \int_0^\infty 2e^{-z} \cdot p(K_1\sqrt{z})^{p-1} \cdot \frac{K_1}{2\sqrt{z}}\,dz\\ &= p\,K_1^p \int_0^\infty e^{-z} z^{p/2-1}\,dz\\ &= p\,K_1^p\,\Gamma\left(\frac{p}{2}\right)\\ &\leq 3\,p\,K_1^p \left(\frac{p}{2}\right)^{p/2},\end{aligned}$$

where $\Gamma(\cdot)$ is the Gamma function. Here, without a proof, we are using this fact that $\Gamma(x) \leq 3\,x^x$ for all $x \geq 0.5$. Taking the p-th root of both sides, we get:

$$(\mathbf{E}[|X|^p])^{1/p} \leq \frac{(3\,p)^{1/p}}{\sqrt{2}}\,K_1\,\sqrt{p} \leq 2.13\,K_1\,\sqrt{p}.$$

To see the last inequality note that $(3x)^{1/x}$ has a negative derivative for $x \geq 1$ implying it takes it maximum at $x = 1$. This shows that the moment growth condition Definition 2 holds for $K_2 \geq 2.13\,K_1$. □

Deriving tail bound from MGF bound: Proof of 5 ⇒ 1

Proof. Suppose X is a zero mean random variable with bounded MGF. For all $\lambda \in \mathbb{R}$, we have:

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

Our goal is to bound the tail probability. In particular, for any $\lambda > 0$, we have:

$$\mathbf{Pr}[X \geq t] = \mathbf{Pr}\left[e^{\lambda X} \geq e^{t\lambda}\right]$$

This identity holds because the function $e^{\lambda x}$ is strictly increasing in x for $\lambda > 0$. This means that $X \geq t$ if and only if $e^{\lambda X} \geq e^{\lambda t}$. Since the events are equivalent, their probabilities are equal.

Next we use the Markov's inequality and the MGF bound we found earlier:

$$\mathbf{Pr}[X \geq t] = \mathbf{Pr}\left[e^{\lambda X} \geq e^{t\lambda}\right] \leq \frac{\mathbf{E}\left[e^{\lambda X}\right]}{e^{t\lambda}} \leq \exp\left(K_5^2\lambda^2 - t\lambda\right)\,.$$

Note that the above bound holds for any $\lambda > 0$. Hence, to obtain the strongest upper bounds, we pick λ that minimizes the right hand side:

$$\mathbf{Pr}[X \geq t] \leq \inf_{\lambda > 0} \exp\left(K_5^2\lambda^2 - t\lambda\right)$$

It suffices to minimize the exponent, which is a quadratic function with roots at $\lambda = 0$ and $\lambda = t/K_5^2$. Hence the minimum occurs at $\lambda = t/(2K_5^2)$. By substituting λ in here, we obtain:

$$\mathbf{Pr}[X \geq t] \leq \exp\left(-\frac{t^2}{2K_5^2}\right)\,.$$

Note that the same bound can be proved for $-X$. Hence, we get:

$$\mathbf{Pr}[X \geq t] \leq 2\exp\left(-\frac{t^2}{2K_5^2}\right)\,.$$

Note that this is the desired bound for Definition 1, if we set $K_1 \geq \sqrt{2}\,K_5$. Hence, the proof is complete. □

For the rest of the proofs, see Vershynin's book [Ver18].

Limitations of Sub-Gaussians

Issue with variance-insensitive bounds: One limitation of sub-Gaussian tail bounds, such as Hoeffding's inequality, is that they rely on the range of the random variables but not their actual concentration behavior, such as their variance. This can lead to loose bounds for distributions with small variance but large range.

For example, consider a random variable X that takes value 0 with high probability and large values $\pm k$ with small probability. In particular, for some large k, we have:

$$X = \begin{cases} -k & \text{with probability } \dfrac{1}{2\,k^2}\,, \\ 0 & \text{with probability } 1 - \dfrac{1}{k^2}\,, \\ k & \text{with probability } \dfrac{1}{2\,k^2}\,. \end{cases}$$

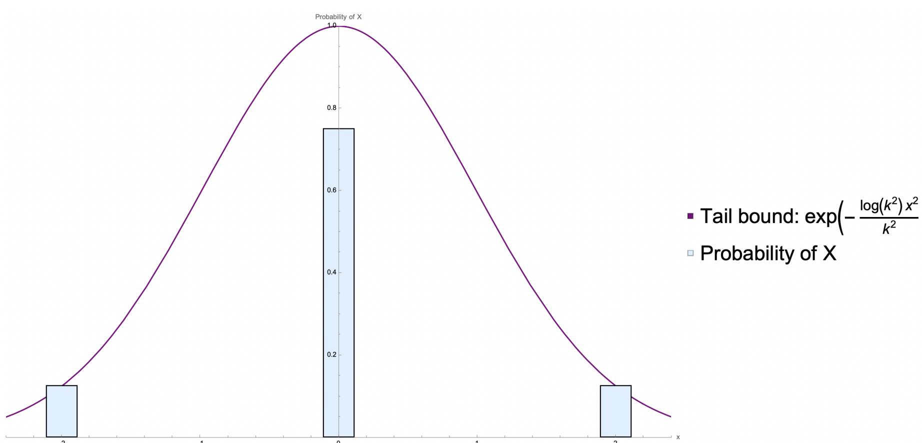

Figure 1: Distribution of X with low variance, but high subGaussianity parameter

The mean of X is 0, and its variance is small ($\mathbf{Var}[X] = 1/k$), but its range is large ($2k$). If we draw k i.i.d. copies of X, $X_1, \ldots, X_k$, the probability that all of them are 0 is $(1 - 1/k^2)^k \approx e^{-1/k} \approx 1$, which is close to 1 for large k. This means that the sum $\sum_{i=1}^{k} X_i$ is very likely to be 0.

However, Hoeffding's inequality, which only depends on the range, cannot distinguish this small-variance distribution from one that is uniform on $[-k, k]$ or one that has $1/2$ probability mass on k and $-k$. This is because Hoeffding's inequality only considers the worst-case scenario, where the random variables are concentrated at the endpoints of their range. As a result, it gives a very loose upper bound on the tail probability $\mathbf{Pr}\Big[|\sum_{i=1}^k X_i| \geq k\epsilon\Big]$, especially for small ϵ:

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

For ϵ that is $o(\sqrt{k})$, this bound is roughly $e^{-\text{tiny}} \approx 1$. Hence, the upper bound provided by Hoeffding's inequality is very loose: for a probability that is almost 0, we provide an upper bound of almost 1.

This limitation highlights the importance of considering the actual concentration behavior of random variables, not just their range, when applying sub-Gaussian tail bounds. In practice, it may be necessary to use more refined tail bounds that take into account the variance or other concentration properties of the random variables to obtain tighter bounds.

Not all random variables are sub-Gaussian: One might conjecture that every random variable is a sub-Gaussian random variable for a sufficiently large parameter K. This is not true for some unbounded random variables. For instance, the square of a standard Gaussian random variable Z^2 is not sub-Gaussian. To see this, consider the moment generating function of the centered variable $Z - \mathbf{E}[Z] = Z^2 - 1$:

$$\begin{aligned}\mathbf{E}\left[e^{\lambda(Z^2-1)}\right] &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{\lambda(z^2-1)}e^{-z^2/2}\,dz = \frac{e^{-\lambda}}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-z^2(1-2\lambda)/2}\,dz \\ &= \begin{cases} \dfrac{e^{-\lambda}}{\sqrt{1-2\lambda}} & \lambda < 1/2\,, \\ \text{unbounded} & \lambda \geq 1/2\,, \end{cases}\end{aligned}$$

which is unbounded for $\lambda \geq 1/2$. However, for $|\lambda| \leq 1/4$, Z^2 behaves like a sub-Gaussian random variable. One can verify that for $\lambda \leq 1/4$:

$$\frac{e^{-\lambda}}{\sqrt{1-2\,\lambda}} \leq e^{2\lambda^2} .$$

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.

References

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