Sub-Gaussian Random Variable and its Tail Bound

Beier Zhu

Definition 1. ***(Sub-Gaussian R.V.)*** *A r.v. X with finite mean μ is sub-Gaussian with parameter σ if*

$$\mathbb{E}[\exp(\lambda(X-\mu))] \leq \exp(\frac{\sigma^2\lambda^2}{2}), \forall \lambda \in \mathbb{R}. \tag{1}$$

We say that X is σ-sub-Gaussian and say it has variance proxy σ^2.

We define the class of sub-Gaussian r.v.s by way of a bound on their moment-generating functions. It turns out to imply exactly the type of exponential tail bound we want:

Theorem 1. ***(Tail bound for sub-Gaussian r.v.).*** *If a r.v. X with finite mean μ is σ-sub-Gaussian, then*

$$\mathbb{P}(|X-\mu| > t) \leq 2\exp(-\frac{t^2}{2\sigma^2}),\ \forall t \in \mathbb{R}. \tag{2}$$

Proof. Fix $t > 0$. For any $\lambda > 0$,

$$\begin{aligned}
\mathbb{P}(X-\mu \geq t) &= \mathbb{P}(\exp(\lambda(X-\mu)) \geq \exp(\lambda t)) && (3)\\
&\leq \exp(-\lambda t)\mathbb{E}[\exp(\lambda(X-\mu)] && \text{[Markov's inequality]} \quad (4)\\
&\leq \exp(-\lambda t)\exp(\frac{\sigma^2\lambda^2}{2}) && \text{[Eq. (1)]} \quad (5)\\
&= \exp(-\lambda t + \sigma^2\lambda^2/2). && (6)
\end{aligned}$$

Because the bound 6 holds for any choose of $\lambda > 0$ and $\exp(\cdot)$ is monotonically increasing. We can optimize the bound by finding λ which minimizes the exponent $-\lambda t + \sigma^2\lambda^2/2$. Differentiating and setting the derivative equal to zero, we find that the optimal choice is $\lambda = t/\sigma^2$, yielding the one-sided tail bound

$$\mathbb{P}(X-\mu \geq t) \leq \exp(-\frac{t^2}{2\sigma^2}) \tag{7}$$

By the same line of reasoning, it is easy to show that for all $t > 0$:

$$\mathbb{P}(X-\mu \leq -t) \leq \exp(-\frac{t^2}{2\sigma^2}), \tag{8}$$

By applying the union bound:

$$\mathbb{P}(|X-\mu| \geq t) = \mathbb{P}(X-\mu \geq t) + \mathbb{P}(X-\mu \leq -t) \leq 2\exp(-\frac{t^2}{2\sigma^2}) \tag{9}$$

□

Discussion with Chebyshev's Inequality. The proof of Theorem 1 follows a similar reasoning to that of Chebyshev's Inequality, with both relying on **Markov's inequality** and **the bound on the central moments** (Chebyshev requires the k-th central moment exists and sub-Gaussian requires infinitely many moments exist). Recall that:

Theorem 2. ***(Chebyshev's Inequality)*** *If X is a r.v. with mean μ and finite k-th central moment $\mathbb{E}[|X-\mu|^k]$, then $\forall t > 0$,*

$$\mathbb{P}[|X-\mu| > t] \leq \frac{\mathbb{E}[|X-\mu|^k]}{t^k} \tag{10}$$

Specifically, when $k = 2$ and denote the variance as σ^2, we have:

$$\mathbb{P}[|X-\mu| > t] \leq \frac{\sigma^2}{t^2} \tag{11}$$

Proof.

$$\mathbb{P}(|X-\mu| \geq t) = \mathbb{P}(|X-\mu|^k) \geq t^k \tag{12}$$

$$\leq t^{-k}\mathbb{E}[|X-\mu|^k] \qquad \text{[Markov's inequality]} \tag{13}$$

□

Chebyshev's inequality provides a polynomial tail bound, while the sub-Gaussian inequality offers a tighter exponential tail bound. However, the assumption of sub-Gaussian is more strict, as it requires infinite central moments exist. If infinite moments also exist for Chebyshev, it turns out that

$$\inf_{k\in\mathbb{N}} \mathbb{P}[|X-\mu|^k] \leq \inf_{\lambda>0} \exp(-\lambda t)\mathbb{E}[\exp(\lambda(X-\mu)] \tag{14}$$

The optimal polynomial tail bound is tighter than the optimal exponential tail bound (see Tengyu Ma's notes). As we will see shortly though, using exponential functions of random variables allows us to prove results about sums of random variables more conveniently. This "tensorization" property is why most researchers use exponential tail bounds in practice.

Theorem 3. ***(Sum of sub-Gaussian r.v.s is sub-Gaussian).*** *If $X_1, ..., X_n$ are independent sub-Gaussian r.v.s with variance proxies $\sigma_1^2, ..., \sigma_n^2$, then $Z = \sum_{i=1}^n X_i$ is sub-Gaussian with variance proxy $\sum_{i=1}^n \sigma_i^2$. As a consequence, we have the tail bound*

$$\mathbb{P}(|Z-\mathbb{E}[Z]|) \leq 2\exp(-\frac{t^2}{2\sum_{i=1}^n \sigma_i^2}) \tag{15}$$

Proof. Using the independence of $X_1, ..., X_n$, we have that for any $\lambda \in \mathbb{R}$:

$$\mathbb{E}[\exp\{\lambda(Z-\mathbb{E}[Z])\}] = \mathbb{E}[\prod_{i=1}^n \exp\{\lambda(X_i - \mathbb{E}[X_i])\}] \tag{16}$$

$$= \prod_{i=1}^n \mathbb{E}[\exp\{\lambda(X_i - \mathbb{E}[X_i])\}] \tag{17}$$

$$\leq \prod_{i=1}^n \exp(\frac{\lambda^2\sigma_i^2}{2}) \tag{18}$$

$$= \exp(\frac{\lambda^2\sum_{i=1}^n \sigma^2}{2}), \tag{19}$$

So Z is sub-Gaussian with variance proxy $\sum_{i=1}^n \sigma_i^2$. The tail bound then follows immediately from Theorem 1. □

1 Examples of Sub-Gaussian R.V.s

Example 1. (Rademacher R.V.) A *Rademacher r.v.* ϵ takes a value of 1 with probability $1/2$ and a value of -1 with probability $1/2$. Rademacher r.v. is 1-sub-Gaussian.

Proof.

$$\mathbb{E}[\exp(\lambda\epsilon)] = \frac{1}{2}\{\exp(-\lambda) + \exp(\lambda)\} \tag{20}$$

$$= \frac{1}{2}\{\sum_{k=0}^\infty \frac{(-\lambda)^k}{k!} + \sum_{k=0}^\infty \frac{\lambda^k}{k!}\} \tag{21}$$

$$= \sum_{k=0}^\infty \frac{\lambda^{2k}}{(2k)!} \qquad \text{[for odd } k, (-\lambda)^k + (\lambda)^k] \tag{22}$$

$$\leq 1 + \sum_{k=1}^\infty \frac{(\lambda^2)^k}{2^k k!} \qquad [2^k! \text{ is every other term of } (2k)!] \tag{23}$$

$$= \exp(\lambda^2/2) \tag{24}$$

□

Example 2. (R.V. with bounded distance to mean.) Suppose a random variable X satisfies $|X - \mathbb{E}[X]| \leq M$ almost surely for some constant M. Then X is $O(M)$-sub-Gaussian.

Example 3. (Bounded R.V.) If X is a r.v. such that $a \leq X \leq b$ almost surely for some constant $a, b \in \mathbb{R}$, then

$$\mathbb{E}[\exp\{\lambda(X - \mathbb{E}[X])\}] \leq \exp\{\frac{\lambda^2(b-a)^2}{8}\}, \tag{25}$$

i.e., X is sub-Gaussian with variance proxy $(\frac{b-a}{2})^2$ (**Hoeffding's Lemma**).

Using Theorem 1, we have the tail bound for bounded r.v.:

$$\mathbb{P}[|X - \mu| \geq t] \leq 2\exp(-\frac{2t^2}{(b-a)^2}) \tag{26}$$

Let $X_1, X_2, ..., X_n$ be independent real-valued r.v.s, such that $a_i \leq X_i \leq b_i$ almost surely. Define $\bar{X} = \frac{1}{n}\sum_{i=1}^{N} X_i$, and let $\mu = \mathbb{E}[\bar{X}]$, Then for any $t > 0$, using Thereom 3, we have the tail bound:

$$\mathbb{P}[|\bar{X} - \mu| \geq t] = \mathbb{P}[|n\bar{X} - n\mu| \geq nt] \tag{27}$$

$$\leq 2\exp(-\frac{n^2t^2}{2\sum_{i=1}^{n}\sigma_i^2}) \tag{28}$$

$$= 2\exp(-\frac{2n^2t^2}{\sum_{i=1}^{n}(b_i - a_i)^2}) \tag{29}$$

Example 4. (Gaussian R.V.). The variance and the variance proxy are the same.