Concentration inequality and high dimensional probability

Du Junye, Li Zhenghao
Department of Statistics and Actuarial Science
The University of Hong Kong (HKU)

February 20, 2025

Abstract

Concentration inequalities and high-dimensional probability are foundational tools in modern probability theory and statistical learning. These concepts provide rigorous frameworks to quantify how a random variable deviates from its expected value. In this note, we are going to present the definitions of Sub-Gaussian and Sub-Exponential variables, together with common concentration inequalities. **Note:** All the contents come from Vershynin's High Dimensional Probability[Ver18] and HKU reading group of 2025.02.13.

1 Classical Inequalities

Lemma 1 (Jensen's inequality). *For any random variable X and a convex function $\psi : \mathbb{R} \to \mathbb{R}$, we have:*

$$\psi(\mathbb{E}X) \leq \mathbb{E}\psi(X)$$

The simple consequence of the Jensen's inequality is that, $\|X\|_{L_p}$ is an increasing function in p, which is:

$$\|X\|_{L_p} \leq \|X\|_{L_q} \text{ for any } 0 \leq p \leq q \leq \infty$$

Lemma 2 (Minkowski's Inequality). *For any $p \in [1, \infty]$ and any random variables $X, Y \in L^p$, we have:*

$$\|X + Y\|_{L^p} \leq \|X\|_{L^p} + \|X\|_{L^p}$$

This can be viewed as the triangle inequality, which implies that the $\|\cdot\|_{L^p}$ is a norm when $p \in [1, \infty]$.

Lemma 3 (Hölder's Inequality). *If $p, q \in [1, \infty]$ are conjugate exponents, that is, $1/p + 1/q = 1$, then the random variable $X \in L^p$ and $Y \in L^q$ satisfy:*

$$|\mathbb{E}XY| \leq \|X\|_{L^p}\|Y\|_{L^q}$$

This inequality is the generalized form of the Cauchy-Schwarz inequality, which is:

$$|\mathbb{E}XY| \leq \|X\|_{L^2}\|Y\|_{L^2}$$

Lemma 4 (Markov Inequality). *For any non-negative random variable X and $t > 0$, we have:*

$$\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}X}{t}$$

Lemma 5 (Chebyshev's Inequality). *Let X be a random variable with mean μ and variance σ^2. Then, for any $t > 0$, we have:*

$$\mathbb{P}\{|X - \mu| \geq t\} \leq \frac{\sigma^2}{t^2}$$

2 Concentration of sums of independent random variables

Concentration inequalities quantify how a random variable X deviates around its mean μ. They usually take the form of two-sided bounds for the tails of $X - \mu$, such as:

$$\mathbb{P}\{|X - \mu| > t\} \leq \text{something small}$$

Proposition 1 (Tails of the normal distribution)**.** *Let $g \sim N(0,1)$. Then for all $t > 0$, we have*

$$\left(\frac{1}{t} - \frac{1}{t^3}\right) \cdot \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \leq \mathbb{P}\left\{g \geq t\right\} \leq \frac{1}{t} \cdot \frac{1}{\sqrt{2\pi}} e^{-t^2/2}$$

In particular, for $t \geq 1$ the tail is bounded by the density:

$$\mathbb{P}\left\{g \geq t\right\} \leq \frac{1}{\sqrt{2\pi}} e^{-t^2/2}.$$

Lemma 6 (Hoeffding's inequality)**.** *Let $X_1, \dots, X_N$ be independent symmetric Bernoulli random variables, and $a = (a_1, \dots, a_N) \in \mathbb{R}^N$. Then, for any $t \geq 0$, we have:*

$$\mathbb{P}\left\{\sum_{i=1}^{N} a_i X_i \geq t\right\} \leq \exp\left(-\frac{t^2}{2||a||_2^2}\right).$$

Lemma 7 (Hoeffding's inequality for general bounded random variables)**.** *Let $X_1, \dots, X_N$ be independent random variables. Assume that $X_i \in [m_i, M_i]$ for every i. Then, for any $t > 0$, we have:*

$$\mathbb{P}\left\{\sum_{i=1}^{N}(X_i - \mathbb{E}X_i) \geq t\right\} \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^{N}(M_i - m_i)^2}\right).$$

Lemma 8 (Chernoff's inequality)**.** *Let X_i be independent Bernoulli random variables with parameters p_i. Consider their sum $S_N = \sum_{i=1}^{N} X_i$ and denote its mean by $\mu = \mathbb{E}S_N$. Then, for any $t > \mu$, we have:*

$$\mathbb{P}\left\{S_N \geq t\right\} \leq e^{-\mu}\left(\frac{e\mu}{t}\right)^t.$$

3 Sub-Gaussian Properties

Let X be a random variable. Then the following properties are equivalent; the parameters $K_i > 0$ appearing in these properties differ from each other by at most an absolute constant factor.

1. There exists $K_1 > 0$ such that the tails of X satisfy

$$P\{|X| \geq t\} \leq 2\exp(-t^2/K_1^2) \quad \text{for all } t \geq 0.$$

2. There exists $K_2 > 0$ such that the moments of X satisfy

$$\|X\|_{L^p} = (\mathbb{E}|X|^p)^{1/p} \leq K_2\sqrt{p} \quad \text{for all } p \geq 1.$$

3. There exists $K_3 > 0$ such that the MGF of X^2 satisfies

$$\mathbb{E}\exp(\lambda^2 X^2) \leq \exp(K_3^2\lambda^2) \quad \text{for all } \lambda \text{ such that } |\lambda| \leq \frac{1}{K_3}.$$

4. There exists $K_4 > 0$ such that the MGF of X^2 is bounded at some point, namely

$$\mathbb{E}\exp(X^2/K_4^2) \leq 2.$$

Moreover, if $\mathbb{E}X = 0$ then properties 1–4 are also equivalent to the following one.

5. There exists $K_5 > 0$ such that the MGF of X satisfies

$$\mathbb{E}\exp(\lambda X) \leq \exp(K_5^2\lambda^2) \quad \text{for all } \lambda \in \mathbb{R}.$$

Definition 1 (Sub-Gaussian Norm). *The Sub-Gaussian norm of a random variable X, denoted $\|X\|_{\psi_2}$, is defined by:*

$$\|X\|_{\psi_2} := \sup_{p\geq 1} \frac{1}{\sqrt{p}}(\mathbb{E}|X|^p)^{1/p}$$

Lemma 9 (Centering). *If X is a Sub-Gaussian random variable then $X - \mathbb{E}X$ is Sub-Gaussian too and:*

$$\|X - \mathbb{E}X\|_{\psi_2} \leq 2\|X\|_{\psi_2}$$

Note: Due to the different definitions of the Sub-Gaussian norm, the 2 here in the lemma could be a constant C.

Lemma 10 (General Hoeffding's inequality for Sub-Gaussian Random Variable). *Let $X_1, \ldots, X_N$ be independent, mean zero, sub-gaussian random variables, and $a = (a_1, \ldots, a_N) \in \mathbb{R}^N$. Then, for every $t \geq 0$, we have*

$$\mathbb{P}\left\{\left|\sum_{i=1}^{N} a_i X_i\right| \geq t\right\} \leq 2\exp\left(-\frac{ct^2}{K^2\|a\|_2^2}\right)$$

where $K = \max_i \|X_i\|_{\psi_2}$.

3.1 Sub-Gaussian Vectors's Properties

Lemma 11 (Sub-Gaussian Vector). *$\boldsymbol{X} \in \mathbb{R}^p$ is Sub-Gaussian random vector if and only if for any $\boldsymbol{a} \in \mathbb{R}^p$ s.t $\|\boldsymbol{a}\|_2 = 1$ such that $\sum_i a_i X_i$ is Sub-Gaussian random variable and $\|\boldsymbol{X}\|_{\psi_2} = \max_{\|\boldsymbol{a}\|_2=1} \|\sum_i a_i X_i\|_{\psi_2}$.*

Lemma 12 (DNN trasnformation of Sub-Gaussian vector). *Suppose that $\boldsymbol{X} \in \mathbb{R}^p$ is a sub-gaussian random vector, let $g(\boldsymbol{x}) = \boldsymbol{W}_L\sigma_L(\boldsymbol{W}_{L-1}\cdots\sigma_2(\boldsymbol{W}_2\sigma_1(\boldsymbol{W}_1\boldsymbol{x})))$ be a ReLU deep neural network. Then $g(\boldsymbol{X})$ is a sub-gaussian random variable.*

Lemma 13 (β-hölder smooth trasformation of Sub-Gaussian vector). *Suppose that $\boldsymbol{X} \in \mathbb{R}^p$ is a sub-gaussian random vector, let f be a β-hölder smooth function with $0 < \beta \leq 1$. Then $f(\boldsymbol{X})$ is a sub-gaussian random variable.*

4 Sub-Exponential Properties

Let X be a random variable. Then the following properties are equivalent; the parameters $K_i > 0$ appearing in these properties differ from each other by at most an absolute constant factor.

1. The tails of X satisfy
$$\mathbb{P}\{|X| \geq t\} \leq 2\exp(-t/K_1) \quad \text{for all } t \geq 0.$$

2. The moments of X satisfy
$$\|X\|_{L^p} = (\mathbb{E}|X|^p)^{1/p} \leq K_2 p \quad \text{for all } p \geq 1.$$

3. The MGF of $|X|$ satisfies
$$\mathbb{E}\exp(\lambda|X|) \leq \exp(K_3\lambda) \quad \text{for all } \lambda \text{ such that } 0 \leq \lambda \leq \frac{1}{K_3}.$$

4. The MGF of $|X|$ is bounded at some point, namely
$$\mathbb{E}\exp(|X|/K_4) \leq 2.$$

Moreover, if $\mathbb{E}X = 0$ then properties 1-4 are also equivalent to the following one.

5. The MGF of X satisfies

$$\mathbb{E}\exp(\lambda X) \le \exp(K_5^2\lambda^2) \quad \text{for all } \lambda \text{ such that } |\lambda| \le \frac{1}{K_5}.$$

Definition 2 (Sub-Exponential Norm)**.** *The Sub-Exponential norm of a random variable X, denoted $\|X\|_{\psi_1}$, is defined by:*

$$\|X\|_{\psi_2} := \sup_{p\ge 1} \frac{1}{p}(\mathbb{E}|X|^p)^{1/p}$$

Lemma 14. *Suppose that x is a sub-gaussian random variable, then x^2 is a sub-exponential random variable and:*

$$\|x\|_{\psi_2}^2 \le \left\|x^2\right\|_{\psi_1} \le 2\,\|x\|_{\psi_2}^2$$

Lemma 15 (Product of sub-gaussians is sub-exponential)**.** *Let X and Y be sub-gaussian random variables. Then XY is sub-exponential. Moreover,*

$$\|XY\|_{\psi_1} \le 2\|X\|_{\psi_2}\|Y\|_{\psi_2}.$$

5 Berstein's Inequality

Lemma 16 (Bernstein's inequality)**.** *Let $X_1, \dots, X_N$ be independent, mean zero, sub-exponential random variables, and $\boldsymbol{a} = (a_1, \dots, a_N) \in \mathbb{R}^N$. Then, for every $t \ge 0$, we have*

$$\mathbb{P}\left\{\left|\sum_{i=1}^N a_i X_i\right| \ge t\right\} \le 2\exp\left[-c\min\left(\frac{t^2}{K^2\|\boldsymbol{a}\|_2^2}, \frac{t}{K\|\boldsymbol{a}\|_\infty}\right)\right]$$

where $K = \max_i \|X_i\|_{\psi_1}$.

A special case is that when $a_i = 1/N$, we could obtain a form of Berstein's inequality for averages:

Corollary 1 (Bernstein's inequality)**.** *Let $X_1, \dots, X_N$ be independent, mean zero, sub-exponential random variables. Then, for every $t \ge 0$, we have*

$$\mathbb{P}\left\{\left|\frac{1}{N}\sum_{i=1}^N X_i\right| \ge t\right\} \le 2\exp\left[-c\min\left(\frac{t^2}{K^2}, \frac{t}{K}\right)N\right]$$

where $K = \max_i \|X_i\|_{\psi_1}$.

Note: A common trick is to let $\frac{t}{K} < 1$ so that $\min\left(\frac{t^2}{K^2}, \frac{t}{K}\right) = \frac{t^2}{K^2}$.

For an integer $d > 0$, let P_d be the set of partitions of $[d]$ into nonempty, pairwise disjoint sets. For a partition $\mathcal{J} = \{J_1, \cdots, J_k\}$, and a d-indexed tensor $\boldsymbol{A} = (a_{\boldsymbol{i}})_{\boldsymbol{i}\in[n]^d}$, define:

$$\|\boldsymbol{A}\|_{\mathcal{J}} = \sup\left\{\sum_{\boldsymbol{i}\in[n]^d} a_{\boldsymbol{i}} \prod_{l=1}^k \boldsymbol{x}^{(l)}_{\boldsymbol{i}_{J_l}} : \left\|\boldsymbol{x}^{(l)}\right\|_2 \le 1, \boldsymbol{x}^{(l)} \in \mathbb{R}^{n^{|J_l|}}, 1 \le l \le k\right\} \tag{1}$$

Lemma 17 (Theorem 1.4 of [Ada15])**.** *Let $X = (X_1, \cdots, X_n)$ be a random vector with independent components, such that for all $i \in [n]$, $\|X_i\|_{\psi_2} \le \gamma$. Then for every polynomial $f : \mathbb{R}^n \to \mathbb{R}$ of degree D and any $t > 0$,*

$$\mathbb{P}\left(|f(X) - \mathbb{E}f(X)| \ge t\right) \le 2\exp\left\{-\frac{1}{C_D}\eta_f(t)\right\} \tag{2}$$

where $\eta_f(t) = \min_{1\le d\le D}\min_{\mathcal{J}\in\mathcal{P}_d}\left(\frac{t}{\gamma^d\|\mathbb{E}\mathcal{D}^d f(X)\|_{\mathcal{J}}}\right)^{2/|\mathcal{J}|}$, $\mathcal{D}^d f$ denotes the d-th derivative of f and C_D is some positive constant depending on D.

Let $f(X) = \sum\limits_{i=1}^{N} a_i X_i^d$, we have the following Bernstein-type inequality.

Lemma 18 (Bernstein-type inequality in high order case)**.** *Let $X_1, \cdots, X_N$ be independent subgaussian random variables, and $\boldsymbol{a} = (a_1, \cdots, a_N) \in \mathbb{R}^N$. Then, for every $d > 1$, there exists some constants $c, m_1, m_2 > 0$ such that for $0 < t < m_1 K^d \|\boldsymbol{a}\|_\infty$,*

$$\mathbb{P}\left(\left|\sum_{i=1}^{n} a_i X_i^d - a_i \mathbb{E} X_i^d\right| \geq t\right) \leq 2\exp\left\{-c \min\left[\left(\frac{t}{K^d \|\boldsymbol{a}\|_2}\right)^2, \frac{t}{K^d \|\boldsymbol{a}\|_\infty}\right]\right\} \tag{3}$$

and for $t \geq m_2 K^d \|\boldsymbol{a}\|_\infty$,

$$\mathbb{P}\left(\left|\sum_{i=1}^{n} a_i X_i^d - a_i \mathbb{E} X_i^d\right| \geq t\right) \leq 2\exp\left\{-c \min\left[\left(\frac{t}{K^d \|\boldsymbol{a}\|_2}\right)^2, \left(\frac{t}{K^d \|\boldsymbol{a}\|_\infty}\right)^{2/d}\right]\right\} \tag{4}$$

where $K = \max_i \|X_i\|_{\psi_2}$.

6 Summary

In this note, we have introduced the concepts of Sub-Gaussian and Sub-Exponential random variables, together with common concentration inequalities. These concepts are foundational tools in modern probability theory and statistical learning, which provide rigorous frameworks to quantify how a random variable deviates from its expected value.

References

[Ada15] Wolff P. Adamczak, R. Concentration inequalities for non-lipschitz functions with bounded derivatives of higher order. *Probability Theory and Related Fields*, 2015.

[Ver18] Roman Vershynin. High-dimensional probability. 2018.