Notes of concentration inequality

September 13, 2023

In probability theory, concentration inequalities provide bounds on how a random variable deviates from some value (typically, its expected value).
Suppose $x_1, \ldots, x_n$ are i.i.d. random variables, by Law of Large Number, we have

$$\frac{1}{n}\sum_{i=1}^{n} X_i = \mathbb{E}X_1$$

By central limit theorem, we arriave at

$$\frac{1}{\sqrt{n}} z \xrightarrow{d} \mathcal{N}(0,1)$$

where $z = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i - EX_1)$. By the definition of Gaussian, we have

$$\begin{aligned} \mathbb{P}(|z| > t) \quad &= \int_t^{\infty} \frac{2}{2\pi} \exp(-\frac{1}{2}z^2) dz \\ &\leqslant 2\exp(-\frac{1}{2}t^2) \end{aligned}$$

Hence,

$$\begin{aligned} \mathbb{P}(\frac{1}{\sqrt{n}}|z| > t) \quad &\leqslant 2\exp(-\frac{1}{2}nt^2) \\ &:= \delta. \end{aligned}$$

With probability at least $1-\delta$,

$$|\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}X_i| \leqslant \sqrt{\frac{\log(2/\delta)}{2n}}$$

1 Sub-Gaussian

Lemma 1. *Let X be a random varible, the following properties are equivalent with $K_i > 0$ differing from each other by at most an absolute constant factor.*

- *Tails* $\mathbb{P}(|x| > t) \leqslant 2\exp(t^2/K_1^2)$
- *Moment* $(\mathbb{E}|X|^p)^{1/p} \leqslant K_2\sqrt{p}$
- *Super-exponential moment* $\mathbb{E}\exp(x^2/K_3) \leqslant 2$ *(bounded)*
- *Moment generating function* $\quad$ *if* $EX = 0, \mathbb{E}\exp(tx) \leqslant \exp(t^2K_4^2), \forall t \in \mathbb{R}$

Definition 1. *A random variable X is called a sub-Gaussian random variable if either one of the equivalent conditions above holds.*

The sub-Gaussian norm of X, denoted $\|X\|_{\psi_2}$, is defined by

$$\|X\|_{\psi_2} := \sup_{p\geqslant 1} p^{-1/2}(\mathbb{E}|X|^p)^{1/p}$$

Proposition 1. $\|\cdot\|_{\psi_2}$ *is indeed a norm.*

By this proposition, we can derive that $Y = X - EX$ is a sub-Gaussian when X is a sub-Gaussian.

Some examples

- Gaussian is sub-Gaussian.
- Bounded random variable is sub-Gaussian. For example, Bernoulli $\mathbb{P}(x = 1) = p, \mathbb{P}(x = -1) = 1 - p$.

1.1 Hoeffding-type inequality

Definition 2. *Let $X_1, \dots, X_n$ be zero-mean independent sub-Gaussian random variables, denote $K := \max_i \|X\|_{\psi_2}$. Then for every $a \in \mathbb{R}^n$ and for all $t > 0$, we have*

$$\mathbb{P}(|\sum_{i=1}^{n} a_i X_i| > t) \leqslant 2\exp(-\frac{ct^2}{K^2\|a\|_2^2})$$

For instance, let $a_i = 1/n$,

$$\mathbb{P}(|\frac{1}{n}\sum_{i=1}^{n} X_i| > t) \leqslant 2\exp(-\frac{cnt^2}{K^2})$$

Moreover, if X_i is bounded, that is $X_i \leqslant M$, the Hoeffding's theorem states that

$$\mathbb{P}(|\frac{1}{n}\sum_{i=1}^{n} X_i| > t) \leqslant 2\exp(-\frac{cnt^2}{M^2})$$

2 Sub-exponential

Similar to sub-Gaussian, the following properties are equivalent

- Tails $\mathbb{P}(|x| > t) \leqslant 2\exp(-t/K_1)$, for $\forall t \geqslant 0$.
- Moment $(\mathbb{E}|X|^p)^{1/p} \leqslant K_2 p$, for $\forall p \geqslant 1$.
- Super-exponential moment $\mathbb{E}\exp(x/K_3) \leqslant 2$ (bounded)

The sub-exponential norm defined by

$$\|X\|_{\psi_1} := \sup_{p\geqslant 1} p^{-1}(\mathbb{E}|X|^p)^{1/p}$$

2.1 Bernstein inequality

Definition 3. *Let $X_1, \dots, X_n$ be zero-mean independent sub-exponential random variables, denote $K := \max_i \|X\|_{\psi_1}$. Then for every $a \in \mathbb{R}^n$ and for all $t > 0$, we* *have*

$$\mathbb{P}(|\sum_{i=1}^{n} a_i X_i| > t) \leqslant 2\exp\left(-c\min\{\frac{t^2}{K^2\|a\|_2^2}, \frac{t}{K\|a\|_\infty}\}\right)$$

Specifically, take $a_i = 1/n$,

$$\mathbb{P}(|\frac{1}{n}\sum_{i=1}^{n} X_i| > t) \leqslant 2\exp\left(-cn\min\{\frac{t^2}{K^2}, \frac{t}{K}\}\right)$$

We always consider $t/K < 1$. For small t this is sub-Gaussian in nature.
Pick $t = K\sqrt{\frac{\log(2/\delta)}{cn}}$, then when $n > \frac{1}{c}\log(2/\delta)$, with probability at least $1-\delta$, we have

$$|\frac{1}{n}\sum_{i=1}^{n} X_i| \leqslant K\sqrt{\frac{\log(2/\delta)}{cn}}$$

Lemma 2. *(Sub-exponential is sub-gaussian squared) If X is sub-Gaussian, if and only if X^2 is sub-exponential when*

$$\|X\|_{\psi_2}^2 \leqslant \|X^2\|_{\psi_1} \leqslant 2\|X\|_{\psi_2}^2$$

Lemma 3. *(The product of sub-gaussians is sub-exponential) Let X and Y be sub-gaussian random variables. Then XY is sub-exponential. Moreover,*

$$\|XY\|_{\psi_1} \leqslant \|X\|_{\psi_2}\|Y\|_{\psi_2}$$

Examples:

1. $x = (x_1, \ldots, x_p) \in \mathbb{R}^p$, each $x_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$. Thus $\mathbb{E}\|x\|_2^2 = \sum_{i=1}^{n}\mathbb{E}x_i^2 = p$.
   Let $z_i = x_i^2 - \mathbb{E}x_i^2$, that is z_i is a centered sub-exponential random variable. Denote $\|z_i\|_{\psi_1} \leqslant M$ for some constant M. By Berstein inequality,

$$\mathbb{P}(|\sum_{i=1}^{p} z_i| > t) \leqslant 2\exp\left(-c\min\{\frac{t^2}{M^2 p}, \frac{t}{K}\}\right)$$

Pick $t = M\sqrt{p}\sqrt{\frac{\log(2/\delta)}{c}}$, then when $p > \frac{\log(2/\delta)}{c}$, with probability at least $1-\delta$, we have

$$\begin{aligned} |\|x\|_2^2 - p| &= |\frac{1}{p}\sum_{i=1}^{p} z_i| \\ &\leqslant M\sqrt{p}\sqrt{\frac{\log(2/\delta)}{c}} \end{aligned}$$

That means $\|x\|_2^2 = p \pm \mathcal{O}(\sqrt{p})$. This helps us to explain the points tend to concentrate near the surface of the unit ball in the high-dimension space.

2. $x, y \in \mathbb{R}^p, x \perp\!\!\!\perp y$. $x_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1), y_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$. $\langle x, y\rangle = 0$. By lemma 3, we have

$$\|x_i y_i\|_{\psi_1} \leqslant \|x_i\|_{\psi_2}\|y_i\|_{\psi_2}$$

Denote $z_i := x_i y_i$, hence $\sum_{i=1}^{p} |z_i| = |\langle x, y\rangle|$. By Berstein inequality,

$$|\langle x, y\rangle| \leqslant M\sqrt{p}\sqrt{\frac{\log(2/\delta)}{c}}$$

It implies that

$$\cos\theta = \frac{|\langle x, y\rangle|}{\|x\|\|y\|} \leqslant \mathcal{O}(\frac{1}{\sqrt{p}})$$

This helps us to explain that vectors tend to be orthogonal in the high-dimension space.

3. Consider $\sum_{i=1}^{p} x_i x_{i+1}$ where $\mathbb{P}(x_i = 1) = 1/2$ and $\mathbb{P}(x_i = 1) = -1/2$. We may use the Azuma-Hoeffding inequality. The Azuma–Hoeffding inequality gives a concentration result for the values of martingales that have bounded differences.

4. Consider the order of $\frac{1}{n}\sum_{i=1}^{n} x_i^5$ close to $\frac{1}{n}\sum_{i=1}^{n} \mathbb{E}x_i^5$ where each x_i is sub-gaussian.

The proof idea is follows

$$\frac{1}{n}\sum_{i=1}^{n} x_i^5 \xrightarrow{1} \frac{1}{n}\sum_{i=1}^{n} x_i^5 \mathbf{1}\{x_i \leqslant M\}) \xrightarrow{2} \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}x_i^5 \mathbf{1}\{x_i \leqslant M\}) \xrightarrow{3} \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}x_i^5$$

To prove the $\xrightarrow{1}$,

$$\begin{aligned}
\mathbb{P}(\frac{1}{n}\sum_{i=1}^{n} x_i^5 \neq \frac{1}{n}\sum_{i=1}^{n} x_i^5 \mathbf{1}\{x_i \leqslant M\}) \quad & \leqslant \mathbb{P}(\exists i, s.t., |x_i| > M) \\
& \leqslant \sum_{i=1}^{n} \mathbb{P}(|x_i| > M) \\
& = n\mathbb{P}(|x_i| > M) \\
& \leqslant 2n\exp(-cnM^2) \quad \text{by the tail of sub-gaussian} \\
& = 2\exp(-cnM^2 + \log n)
\end{aligned}$$

We can prove the $\xrightarrow{2}$ directly in fact that the bounded random variable is sub-gaussian. Hence, with probability $1-\delta$, we have

$$|\frac{1}{n}\sum_{i=1}^{n} x_i^5 \mathbf{1}\{x_i \leqslant M \neq \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}x_i^5 \mathbf{1}\{x_i \leqslant M\}| \leqslant c\sqrt{\frac{\log(2/\delta)}{n}}.$$

For $\xrightarrow{3}$,

$$\begin{aligned}
|\frac{1}{n}\sum_{i=1}^{n} \mathbb{E}x_i^5 \mathbf{1}\{x_i \leqslant M\} - \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}x_i^5| \quad & = |\frac{1}{n}\sum_{i=1}^{n} \mathbb{E}x_i^5 \mathbf{1}\{x_i \geqslant M\}| \\
& \leqslant |\frac{1}{n}\sum_{i=1}^{n} \sqrt{\mathbb{E}x_i^{10}\mathbb{P}(x_i \geqslant M)}| \quad \text{by Cauchy-schwarz ineq} \\
& \leqslant c\sqrt{\mathbb{P}(x_i \geqslant M)} \quad \text{by } \mathbb{E}x_i^{10} \text{ is a constant} \\
& \leqslant c\sqrt{2\exp(-cM^2)}
\end{aligned}$$

In summary, pick $M = c\sqrt{\frac{\log n + \log(2/\delta)}{n}}$, with probability $1-\delta$, we have

$$|\frac{1}{n}\sum_{i=1}^{n} x_i^5 - \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}x_i^5| \leqslant c\sqrt{\frac{\log n + \log(2/\delta)}{n}}.$$