ℓ_p Norm Concentration

Alireza Naderi

October 2021

Let $X = (X_1, \cdots, X_N)^\top$, where X_i are iid sub-gaussian random variables with $\|X_i\|_{\psi_2} \leq K$. Then, for every $p > 0$, the random variable $Y_i \coloneqq |X_i|^p$ satisfies

$$\mathbb{E}[\exp(Y_i^\beta/\sigma^\beta)] \leq 2,$$

for $\beta = 2/p$ and $\sigma = K^p$, i.e. Y_i is sub-weibull with $\|Y_i\|_{\psi_\beta} \leq K^p$. (Check!)

We want to bound the tail probability $\mathbb{P}\big(\big|\, \|X\|_p - (\mathbb{E}\,\|X\|_p^p)^{1/p}\big| \geq t\big)$ for $p \geq 2$. Let $\mu \coloneqq \mathbb{E}Y_i = \mathbb{E}|X_i|^p$, thus $(\mathbb{E}\,\|X\|_p^p)^{1/p} = (N\mu)^{1/p}$.

Lemma 1. *For any $p \geq 1$ and $z, \delta \geq 0$, if $|z-1| \geq \delta$, then $|z^p - 1| \geq \max\{\delta, \delta^p\}$.*

Theorem 1 (Theorem 3.1 in [1])**.** *Suppose $\{Y_i\}_{i=1}^N$ are independent sub-weibull random variables with $\|Y_i\|_{\psi_\beta} \leq \sigma$. Then for any $\mathbf{a} = (a_1, \cdots, a_N)^\top \in \mathbb{R}^N$ and all $s > 2$, we have*

$$\mathbb{P}\Big(\Big|\sum_{i=1}^N a_i Y_i - \mathbb{E}(\sum_{i=1}^N a_i Y_i)\Big| \geq C_\beta \sigma\big(\,\|\mathbf{a}\|_2 \sqrt{s} + \|\mathbf{a}\|_\infty s^{1/\beta}\big)\Big) \leq e^{-s}, \tag{1}$$

where C_β is an absolute constant only depending on β.

Remark. In our setting, $1/\beta = p/2 \geq 1$. Check that the following is equivalent to (1):

$$\mathbb{P}\Big(\Big|\sum_{i=1}^N a_i Y_i - \mathbb{E}(\sum_{i=1}^N a_i Y_i)\Big| \geq u\Big) \leq \exp\big(-c\min\{\frac{u^2}{C_\beta^2\sigma^2\|\mathbf{a}\|_2^2}, \frac{u^\beta}{C_\beta^\beta\sigma^\beta\|\mathbf{a}\|_\infty^\beta}\}\big). \tag{2}$$

Theorem 2. *Let $X = (X_1, \cdots, X_N)^\top$, where X_i are iid sub-gaussian random variables with $\|X_i\|_{\psi_2} \leq K$. Then, for every $p \geq 2$,*

$$\Big\| \|X\|_p - (\mathbb{E}\,\|X\|_p^p)^{1/p} \Big\|_{\psi_2} \lesssim C_p K^p,$$

where C_p is a constant that only depends on p.

Proof. Throughout, let $\beta = 2/p \leq 1$, $\sigma = K^p$, $Y_i = |X_i|^p$, and $\mu = \mathbb{E}Y_i = \mathbb{E}|X_i|^p$. Note that the sub-gaussianity of X_i implies that all the p-norms are finite, i.e. $\mu < \infty$. By Lemma 1, the event $\big\{ \big| \frac{\|X\|_p}{(N\mu)^{1/p}} - 1 \big| \geq \delta \big\}$ is included in the event $\big\{ \big| \frac{\|X\|_p^p}{N\mu} - 1 \big| \geq \max\{\delta, \delta^p\} \big\}$. Thus,

$$\begin{aligned} \mathbb{P}\Big(\Big| \frac{\|X\|_p}{(N\mu)^{1/p}} - 1 \Big| \geq \delta \Big) &\leq \mathbb{P}\Big(\Big| \frac{\|X\|_p^p}{N\mu} - 1 \Big| \geq \max\{\delta, \delta^p\} \Big) \\ &= \mathbb{P}\Big(\Big| \sum_{i=1}^N \frac{1}{N\mu} Y_i - \mathbb{E}(\sum_{i=1}^N \frac{1}{N\mu} Y_i) \Big| \geq \max\{\delta, \delta^p\} \Big). \end{aligned}$$

For $\delta \geq 1$, $\max\{\delta, \delta^p\} = \delta^p$. Then by Theorem 1, the tail probability can be bounded by

$$\begin{aligned} \exp\big(-c \min\{ \frac{\delta^{2p} \cdot N\mu^2}{C_\beta^2 \sigma^2}, \frac{\delta^2 \cdot (N\mu)^\beta}{C_\beta^\beta \sigma^\beta} \} \big) &\leq \exp\big(-\frac{c}{C_\beta^2 \sigma^2} \min\{\delta^{2p} N\mu^2, \delta^2 N^\beta \mu^\beta\} \big) \\ &\leq \exp\big(-c \frac{\delta^2 N^\beta \mu^\beta}{C_\beta^2 \sigma^2} \big). \end{aligned}$$

Hence, by letting $t \coloneqq \delta (N\mu)^{1/p}$, we have

$$\begin{aligned} \mathbb{P}\Big(\Big| \|X\|_p - (\mathbb{E}\,\|X\|_p^p)^{1/p} \Big| \geq t \Big) &= \mathbb{P}\Big(\Big| \|X\|_p - (N\mu)^{1/p} \Big| \geq t \Big) \\ &\leq \exp\big(-c \frac{t^2}{C_\beta^2 \sigma^2} \big). \end{aligned} \tag{3}$$

Therefore, $\Big\| \|X\|_p - (\mathbb{E}\,\|X\|_p^p)^{1/p} \Big\|_{\psi_2} \lesssim C_\beta \sigma = C_\beta K^p$. □

References

[1] Hao, B., Abbasi Yadkori, Y., Wen, Z., & Cheng, G. (2019). Bootstrapping Upper Confidence Bound. Advances in Neural Information Processing Systems, 32, 12123-12133.