Bernstein's inequality for unbounded random variables

High-dimensional Probability and Statistics

HSE University, spring 2023

Let us recall the standard Bernstein inequality.

Theorem 1 (Bernstein's inequality). *Let $\xi_1, \dots, \xi_n$ be i.i.d. centered random variables supported on $[-B, B]$. Then, for any $t \geqslant 0$, it holds that*

$$\mathbb{P}\left(\sum_{i=1}^n \xi_i \geqslant t\right) \leqslant \exp\left\{-\frac{t^2/2}{n\sigma^2 + Bt/3}\right\},$$

where $\sigma^2 = \mathrm{Var}(\xi_1)$.

Bernstein's concentration inequality provides a powerful tool for analysis of sums of random variables. However, it holds only for bounded random variables. The goal of this note is to prove a counterpart of the Bernstein inequality for sums of unbounded random variables. First, let us introduce an auxiliary definition.

Definition 2 (Orlicz norm). For any $p \geqslant 1$, the ψ_p-Orlicz norm of a random variable ξ is defined as

$$\|\xi\|_{\psi_p} = \inf\left\{t > 0 : \mathbb{E} e^{|\xi|^p / t^p} \leqslant 2\right\}.$$

Remark 3. For any $p \geqslant 1$, the $\|\cdot\|_{\psi_p}$-norm is indeed a norm. In particular, it satisfies the triangle inequality.

We are ready to present the main result of this note.

Theorem 4 (Bernstein's inequality for unbounded random variables). *Let $\xi_1, \dots, \xi_n$ be centered i.i.d. random variables, such that $\mathrm{Var}(\xi_1) = \sigma$ and $\|\xi_1\|_{\psi_1} < \infty$. Then, for any $t > 0$ and any $\rho > \|\xi_1\|_{\psi_1} \log(4n/t)$, it holds that*

$$\mathbb{P}\left(\sum_{i=1}^n \xi_i \geqslant t\right) \leqslant \exp\left\{-\frac{t^2/8}{n\sigma^2 + \rho t/6}\right\} + 2\exp\left\{-\frac{\rho}{\left\|\max\limits_{1 \leqslant i \leqslant n} |\xi_i|\right\|_{\psi_1}}\right\}.$$

Proof. Fix $t > 0$ and ρ, satisfying the inequality $\rho > \|\xi_1\|_{\psi_1} \log(4n/t)$ and introduce random variables

$$\eta_i = \xi_i \mathbb{1}(|\xi_i| \leqslant \rho), \quad \zeta_i = \xi_i \mathbb{1}(|\xi_i| > \rho), \quad \text{where } 1 \leqslant i \leqslant n.$$

Since $\mathbb{E}\xi_i = 0$ for all $i \in \{1, \dots, n\}$, it is straightforward to observe that $\mathbb{E}\eta_i = -\mathbb{E}\zeta_i$. Then it holds that

$$\mathbb{P}\left(\sum_{i=1}^n \xi_i \geqslant t\right) \leqslant \mathbb{P}\left(\sum_{i=1}^n (\eta_i - \mathbb{E}\eta_i) \geqslant t/2\right) + \mathbb{P}\left(\sum_{i=1}^n (\zeta_i - \mathbb{E}\zeta_i) \geqslant t/2\right). \tag{1}$$

Applying Bernstein's inequality to bounded random variables $\eta_1, \dots, \eta_n$, we obtain that

$$\mathbb{P}\left(\sum_{i=1}^{n}(\eta_i - \mathbb{E}\eta_i) \geqslant t/2\right) \leqslant \exp\left\{-\frac{t^2/8}{n\sigma^2 + \rho t/6}\right\}. \tag{2}$$

It only remains to bound the second term in the right-hand side of (1).

$$\mathbb{P}\left(\sum_{i=1}^{n}(\zeta_i - \mathbb{E}\zeta_i) \geqslant t/2\right) \leqslant \mathbb{P}\left(\sum_{i=1}^{n}(\zeta_i - \mathbb{E}\zeta_i) \geqslant t/2 \text{ and } \max_{1\leqslant i\leqslant n}|\xi_i| \leqslant \rho\right) + \mathbb{P}\left(\max_{1\leqslant i\leqslant n}|\xi_i| > \rho\right).$$

On the event $\{\max\{|\xi_1|, \dots, |\xi_n|\} \leqslant \rho\}$ we have $\zeta_i = 0$ for all $i \in \{1, \dots, n\}$. Hence,

$$\mathbb{P}\left(\sum_{i=1}^{n}(\zeta_i - \mathbb{E}\zeta_i) \geqslant t/2 \text{ and } \max_{1\leqslant i\leqslant n}|\xi_i| \leqslant \rho\right) = \mathbb{P}\left(-\sum_{i=1}^{n}\mathbb{E}\zeta_i \geqslant t/2 \text{ and } \max_{1\leqslant i\leqslant n}|\xi_i| \leqslant \rho\right).$$

By the definition of ζ_i and Markov's inequality, we have

$$-\mathbb{E}\zeta_i = -\mathbb{E}\xi_i \mathbb{1}(|\xi_i| > \rho) \leqslant \mathbb{E}|\xi_i|\mathbb{1}(|\xi_i| > \rho) = \int_{\rho}^{+\infty}\mathbb{P}\big(|\xi_i| > t\big)\mathrm{d}t$$
$$\leqslant \int_{\rho}^{+\infty} e^{-t\|\xi_i\|_{\psi_1}}\mathbb{E}e^{|\xi_i|/\|\xi_i\|_{\psi_1}}\mathrm{d}t \leqslant 2\int_{\rho}^{+\infty} e^{-t/\|\xi_i\|_{\psi_1}}\mathrm{d}t = 2e^{-\rho/\|\xi_i\|_{\psi_1}}.$$

This implies that

$$-\sum_{i=1}^{n}\mathbb{E}\zeta_i = 2ne^{-\rho/\|\xi_1\|_{\psi_1}},$$

and, therefore, the event $\left\{-\sum_{i=1}^{n}\mathbb{E}\zeta_i \geqslant t/2\right\}$ never happens, because $\rho > \|\xi_1\|_{\psi_1}\log(4n/t)$ by the condition of the theorem. Hence, it holds that

$$\mathbb{P}\left(\sum_{i=1}^{n}(\zeta_i - \mathbb{E}\zeta_i) \geqslant t/2\right) \leqslant \mathbb{P}\left(\max_{1\leqslant i\leqslant n}|\xi_i| > \rho\right),$$

and, applying Markov's inequality, we obtain that

$$\mathbb{P}\left(\sum_{i=1}^{n}(\zeta_i - \mathbb{E}\zeta_i) \geqslant t/2\right) \leqslant e^{-\rho/\|\max\limits_{1\leqslant i\leqslant n}|\xi_i|\|_{\psi_1}}\mathbb{E}e^{\max\limits_{1\leqslant i\leqslant n}|\xi_i|/\|\max\limits_{1\leqslant i\leqslant n}|\xi_i|\|_{\psi_1}} \leqslant 2e^{-\rho/\|\max\limits_{1\leqslant i\leqslant n}|\xi_i|\|_{\psi_1}}. \tag{3}$$

The claim of the theorem follows from the inequalities (1), (2), and (3).

□

There are several remarks we would like to make.

Remark 5. It holds that
$$\Big\| \max_{1\leqslant i\leqslant n} |\xi_i| \Big\|_{\psi_1} \leqslant \|\xi_1\|_{\psi_1} \log_2 n.$$
Indeed, due to the Hölder inequality, we have
$$\mathbb{E} e^{\max_{1\leqslant i\leqslant n} |\xi_i|/(\|\xi_1\|_{\psi_1} \log_2 n)} \leqslant \left(\mathbb{E} e^{\max_{1\leqslant i\leqslant n} |\xi_i|/\|\xi_1\|_{\psi_1}} \right)^{1/\log_2 n}.$$
Since the function $f(u) = e^u$ is monotone and positive, it holds that
$$e^{\max_{1\leqslant i\leqslant n} |\xi_i|/\|\xi_1\|_{\psi_1}} = \max_{1\leqslant i\leqslant n} e^{|\xi_i|/\|\xi_1\|_{\psi_1}} \leqslant \sum_{1\leqslant i\leqslant n} e^{|\xi_i|/\|\xi_1\|_{\psi_1}}.$$
Hence, we obtain that
$$\mathbb{E} e^{\max_{1\leqslant i\leqslant n} |\xi_i|/(\|\xi_1\|_{\psi_1} \log_2 n)} \leqslant \left(\sum_{i=1}^n \mathbb{E} e^{|\xi_i|/\|\xi_1\|_{\psi_1}} \right)^{1/\log_2 n} \leqslant (2n)^{1/\log_2 n} = 2.$$
Remark 6. In [1], the author proved a more general result. Let $X_1, \dots, X_n$ be independent random variables and let $\mathcal{F}$ be a separable family of univariate zero-mean functions. Denote
$$Z = \sup_{f\in\mathcal{F}} \sum_{i=1}^n f(X_i).$$
Then, for any $p \geqslant 1$ and any $\delta \in (0,1)$, with probability at least $1-\delta$, it holds that
$$Z \lesssim \mathbb{E} Z + \sigma_{\mathcal{F}} \sqrt{\log(1/\delta)} + \left(\left\| \max_{1\leqslant i\leqslant n} \sup_{f\in\mathcal{F}} |f(X_i)| \right\|_{\psi_p} \log(1/\delta) \right)^{1/p},$$
where
$$\sigma_{\mathcal{F}}^2 = \sup_{f\in\mathcal{F}} \mathrm{Var} \left(\sum_{i=1}^n f(X_i) \right).$$

References

[1] R. Adamczak. A tail inequality for suprema of unbounded empirical processes with applications to Markov chains. *Electronic Journal of Probability*, 13:no. 34, 1000–1034, 2008.