Qianqian Lei[1] Soham Bonnerjee[1]
Yuefeng Han[2] Wei Biao Wu[1]

[1]University of Chicago [2]University of Notre Dame

June 1, 2026

Stability analysis under Finite L_p Moments

ICML 2026

Why revisit stability?

Problem. In modern regimes, often risks may be unbounded or covariates/response be heavy-tailed.

Prior art

- Uniform stability + bounded loss/ sub-weibull loss gives high-probability generalization bounds via McDiarmid-type bounded differences. Bousquet and Elisseeff (2002); Maurer (2005); Bousquet et al. (2020).
- Distributional relaxations allow sub-Gaussian, sub-exponential, or sub-Weibull stability increments Kontorovich (2014); Maurer and Pontil (2021); Li et al. (2024).
- All moments must exist $\implies$ exponentially decaying tail.

Goal: Stability analysis when only p moments exist.

Example: what goes wrong in previous theory?

Ridge regression with heavy-tailed noise

- DGP: $(x_i, y_i) \in \mathbb{R}^5 \times \mathbb{R}$ from linear model with heavy-tailed errors with exactly $\nu/2$ finite moments.
- $R = \mathbb{E}_{x,y}[(y - x^\top \hat{\beta})^2]$, R_{emp} empirical risk. Here $\hat{\beta}$ ridge estimate with $\lambda = 1$.
- Look at $p(y) = \frac{\mathbb{P}(|R - R_{\text{emp}}| > y)}{\mathbb{P}(|R - R_{\text{emp}}| > C_0 y)}$ versus y for $C_0 = 1.5$.

Example (Ctd)

$$p(y) = \frac{\mathbb{P}(|R - R_{\mathrm{emp}}| > y)}{\mathbb{P}(|R - R_{\mathrm{emp}}| > C_0 y)}, \; C_0 = 1.5.$$

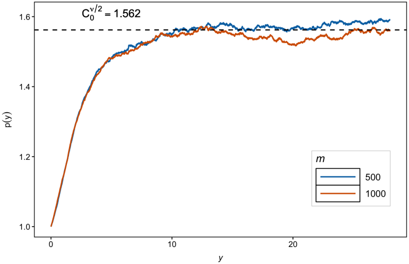

Figure: $\nu = 2.2$

Takeaway. The curves first show Gaussian-type behavior, then stabilize towards $C_0^p \implies$ a polynomial tail like behavior.

Key tool: A Fuk-Nagaev type inequality

Notation: $\|X\|_p = (\mathbb{E}[|X|^p])^{1/p}$, $|\cdot| \leftarrow$ Absolute value/Euclidean norm/Frobenius norm.

Let $x_1, \ldots, x_n$ be independent and let x_i' be an independent copy of x_i. For a statistic $f(x_1, \ldots, x_n)$, assume

$$\left| f(x_1, \ldots, x_i, \ldots, x_n) - f(x_1, \ldots, x_i', \ldots, x_n) \right| \leq H_i(x_i, x_i'),$$

where

$$\|H_i(x_i, x_i')\|_p < \infty.$$

H_i **is a random, data-dependent stability envelope.**

Key tool: A Fuk-Nagaev type inequality

Theorem (Simplified)

For $p \geq 2$, it follows

$$\mathbb{P}\left(|f - \mathbb{E}f| > y\right) \lesssim \frac{\sum_{i=1}^{n} \mathbb{E}|H_i|^p}{y^p} + \exp\left\{-\frac{y^2}{\sum_{i=1}^{n} \mathbb{E}|H_i|^2}\right\}.$$

Message.

- Moderate deviations behave like a sub-Gaussian bound.
- Large deviations have the polynomial tail from finite p-th moments.
- This result is a general version of Nagaev (1970)'s result.

Application: ERM generalization

Training sample $S = (z_1, \ldots, z_m)$, algorithm A, loss $\ell(A_S, z)$. Define

$$R = \mathbb{E}_z \ell(A_S, z), \qquad R_{\mathrm{emp}} = \frac{1}{m} \sum_{i=1}^{m} \ell(A_S, z_i).$$

Let: $S^i = \{z_1, \ldots, z_{i-1}, z_i', z_{i+1}, \ldots, z_m\}$.

Twin L_p stability assumption:

$$\underbrace{|\ell(A_S, z) - \ell(A_{S^i}, z)| \leq H(z_i, z_i')}_{\text{training-data stability}}, \ \underbrace{|\ell(A_S, z) - \ell(A_S, z')| \leq G(z, z')}_{\text{test-data stability}}$$

where $\|H\|_p \vee \|G\|_p < \infty$.

Application: ERM generalization

Theorem (Informal)

With probability $\geq 1-\delta$,

$$|R - R_{\mathrm{emp}}| \lesssim \mathbb{E}H + \underbrace{\left(\|H\|_2\sqrt{m} + \|G\|_2 m^{-1/2}\right)\sqrt{\log(1/\delta)}}_{\text{sub-Gaussian tail}} + \underbrace{\left(\|H\|_p m^{1/p} + \|G\|_p m^{-(1-1/p)}\right)\delta^{-1/p}}_{\text{polynomial tail}}.$$

- Weaker assumptions $\rightarrow$ slightly weaker rate. For comparison, similar rates also with sub-Weibull assumptions Li et. al. (2024).
- Stability increment H can be data-dependent.
- More examples on meta learning, transductive regression in the paper.

Takeaway

Finite-moment stability is enough for high-probability generalization.

- Replace bounded differences by L_p replace-one envelopes.
- Obtain sharp two-regime large deviation:

$$\text{sub-Gaussian} + \text{polynomial tails.}$$

- Simulations support the necessity of the polynomial correction.

Questions: mail to sohambonnerjee@uchicago.edu.