Subvector Inference in Partially Identified Models with Many Moment Inequalities

Alexandre Belloni
Duke

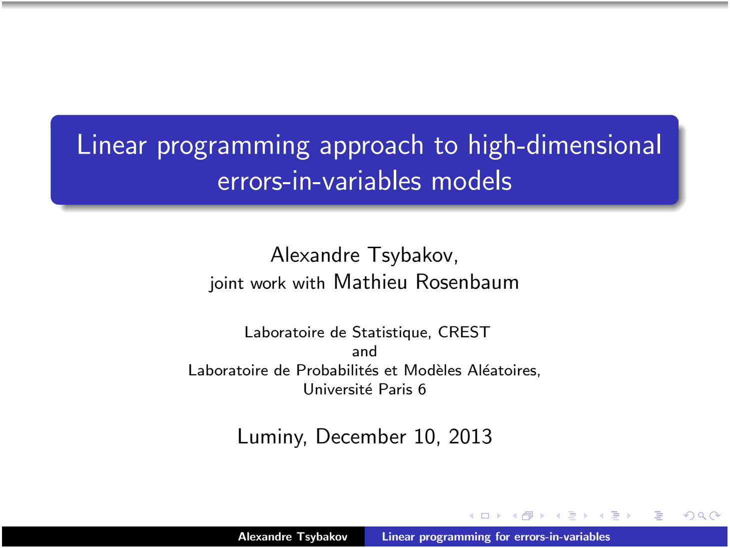
joint work with Federico Bugni (Duke) and Victor Chernozhukov (MIT)

Meeting in Mathematical Statistics
CIRM, December 20th, 2017

Happy Birthday

Luminy, December 12, 2013

Happy Birthday



(Mathieu: "Did you get the slides I sent? ... [send] my best to Sacha, I am feeling bad not having been able to make it to Luminy!")

Introduction

- There is a large literature on inference in partially identified (PI) models defined by moment (in)equalities.

- We consider a model characterized by (θ^*, F)
  - $\theta^* \in \mathbb{R}^{d_\theta}$ is a finite dimensional parameter of interest,
  - F is the distribution of data, i.e., $W \sim F$ (W_i, $i = 1, \ldots, n$, i.i.d.)

Introduction

- There is a large literature on inference in partially identified (PI) models defined by moment (in)equalities.

- We consider a model characterized by (θ^*, F)
  - $\theta^* \in \mathbb{R}^{d_\theta}$ is a finite dimensional parameter of interest,
  - F is the distribution of data, i.e., $W \sim F$ (W_i, $i = 1, \ldots, n$, i.i.d.)

- The *main prediction* of the model is that the true parameter θ^* satisfies

$$\begin{aligned} \mathbb{E}[m_j(W,\theta)] &\leqslant 0, \quad \text{for } j = 1, \ldots, p_I, \\ \mathbb{E}[m_j(W,\theta)] &= 0, \quad \text{for } j = p_I + 1, \ldots, p_I + p_E. \end{aligned} \tag{1}$$

- key issue: θ^* is *not* assumed to be point identified, i.e., given F, there might be a **set** of θ that satisfy (1).

$$\Theta_I \equiv \left\{ \theta \in \Theta \text{ s.t. } \begin{array}{l} \mathbb{E}[m_j(W,\theta)] \leqslant 0 \text{ for } j = 1, \ldots, p_I \\ \mathbb{E}[m_j(W,\theta)] = 0 \text{ for } j = p_I + 1, \ldots, p_I + p_E \end{array} \right\}.$$

Motivating Examples

Interval-Outcome Linear Regression (e.g., Manski and Tamer 2002)

- let Y_i^* denote a latent dependent variable

$$Y_i^* = X_i'\theta^* + \varepsilon_i, \quad \mathbb{E}[\varepsilon_i \mid X_i] = 0 \text{ a.s.}$$

- We only observe an interval s.t. $Y_i \in [Y_i^l, Y_i^u]$

Motivating Examples

Interval-Outcome Linear Regression (e.g., Manski and Tamer 2002)

- let Y_i^* denote a latent dependent variable

$$Y_i^* = X_i'\theta^* + \varepsilon_i, \quad \mathbb{E}[\varepsilon_i \mid X_i] = 0 \text{ a.s.}$$

- We only observe an interval s.t. $Y_i \in [Y_i^l, Y_i^u]$ which leads to

$$\mathbb{E}[X_i'\theta^* - Y_i^u \mid X_i] \leqslant 0$$
$$\mathbb{E}[Y_i^l - X_i'\theta^* \mid X_i] \leqslant 0$$

- We could use

$$\mathbb{E}[X_i'\theta^* - Y_i^u] \leqslant 0$$
$$\mathbb{E}[Y_i^l - X_i'\theta^*] \leqslant 0$$
$$\mathbb{E}[(X_i'\theta^* - Y_i^u)X_{ij}1\{X_{ij} \geqslant 0\}] \leqslant 0$$
$$\mathbb{E}[(Y_i^l - X_i'\theta^*)X_{ij}1\{X_{ij} \geqslant 0\}] \leqslant 0$$
$$\mathbb{E}[-(X_i'\theta^* - Y_i^u)X_{ij}1\{X_{ij} \leqslant 0\}] \leqslant 0$$
$$\mathbb{E}[-(Y_i^l - X_i'\theta^*)X_{ij}1\{X_{ij} \leqslant 0\}] \leqslant 0$$

Motivating Examples

Discrete Choice Model with Multiple Equilibria (Ciliberto and Tamer, 2009)

- m firms play an entry game (Nash Equilibrium) on n independent markets
- On each market, a firm makes an entry decision $d_j \in \{0, 1\}$

$$\text{profit function} \quad \pi_j(d_j, d_{-j}, X, \varepsilon, \theta^*)$$

Motivating Examples

Discrete Choice Model with Multiple Equilibria (Ciliberto and Tamer, 2009)

- m firms play an entry game (Nash Equilibrium) on n independent markets
- On each market, a firm makes an entry decision $d_j \in \{0, 1\}$

$$\text{profit function} \quad \pi_j(d_j, d_{-j}, X, \varepsilon, \theta^*) \geqslant \pi_j(1 - d_j, d_{-j}, X, \varepsilon, \theta^*)$$

where X firm/market characteristics, and we would like to infer θ^*

Motivating Examples

Discrete Choice Model with Multiple Equilibria (Ciliberto and Tamer, 2009)

- m firms play an entry game (Nash Equilibrium) on n independent markets
- On each market, a firm makes an entry decision $d_j \in \{0, 1\}$

$$\text{profit function} \quad \pi_j(d_j, d_{-j}, X, \varepsilon, \theta^*) \geqslant \pi_j(1 - d_j, d_{-j}, X, \varepsilon, \theta^*)$$

  where X firm/market characteristics, and we would like to infer θ^*
- There are set-valued functions R_1, R_2 such that
  - d is the unique equilibrium if $\varepsilon \in R_1(d, X, \theta)$
  - d is a possible equilibrium if $\varepsilon \in R_2(d, X, \theta)$

Motivating Examples

Discrete Choice Model with Multiple Equilibria (Ciliberto and Tamer, 2009)

- m firms play an entry game (Nash Equilibrium) on n independent markets
- On each market, a firm makes an entry decision $d_j \in \{0, 1\}$

$$\text{profit function} \quad \pi_j(d_j, d_{-j}, X, \varepsilon, \theta^*) \geqslant \pi_j(1 - d_j, d_{-j}, X, \varepsilon, \theta^*)$$

  where X firm/market characteristics, and we would like to infer θ^*
- There are set-valued functions R_1, R_2 such that
  - d is the unique equilibrium if $\varepsilon \in R_1(d, X, \theta)$
  - d is a possible equilibrium if $\varepsilon \in R_2(d, X, \theta)$
- Conclude that

$$\begin{aligned} \mathbb{E}[1\{d = d'\} \mid X] &\geqslant \mathbb{E}[1\{\varepsilon \in R_1(d, X, \theta_0)\} \mid X] \\ \mathbb{E}[1\{d = d'\} \mid X] &\leqslant \mathbb{E}[1\{\varepsilon \in R_1(d, X, \theta_0) \cup R_2(d, X, \theta_0)\} \mid X] \end{aligned}$$

Motivating Examples

Discrete Choice Model with Multiple Equilibria (Ciliberto and Tamer, 2009)

- m firms play an entry game (Nash Equilibrium) on n independent markets
- On each market, a firm makes an entry decision $d_j \in \{0, 1\}$

$$\text{profit function} \quad \pi_j(d_j, d_{-j}, X, \varepsilon, \theta^*) \geqslant \pi_j(1 - d_j, d_{-j}, X, \varepsilon, \theta^*)$$

  where X firm/market characteristics, and we would like to infer θ^*
- There are set-valued functions R_1, R_2 such that
  - d is the unique equilibrium if $\varepsilon \in R_1(d, X, \theta)$
  - d is a possible equilibrium if $\varepsilon \in R_2(d, X, \theta)$
- Conclude that

$$\begin{aligned} \mathbb{E}[1\{d = d'\} \mid X] &\geqslant \mathbb{E}[1\{\varepsilon \in R_1(d, X, \theta_0)\} \mid X] \\ \mathbb{E}[1\{d = d'\} \mid X] &\leqslant \mathbb{E}[1\{\varepsilon \in R_1(d, X, \theta_0) \cup R_2(d, X, \theta_0)\} \mid X] \end{aligned}$$

- If conditional distribution of ε given X is known (up to a subvector of θ_0), we can calculate numerically right-hand sides of both inequalities
- we have 2^{m+1} moment inequalities for each value of $X \in \mathcal{X}$ (a discrete set).

Introduction

- ▷ We consider a model characterized by (θ^*, F)
  - $\theta \in \mathbb{R}^{d_\theta}$ is a finite dimensional parameter of interest,
  - F is the distribution of data, i.e., $W \sim F$ (W_i, $i = 1, \ldots, n$, i.i.d.)
- ▷ The *main prediction* of the model is that the true parameter θ^* satisfies

$$\begin{aligned} \mathbb{E}[m_j(W,\theta)] &\leqslant 0, \quad \text{for } j = 1, \ldots, p_I, \\ \mathbb{E}[m_j(W,\theta)] &= 0, \quad \text{for } j = p_I + 1, \ldots, p_I + p_E. \end{aligned} \tag{2}$$

- ▷ key issue: θ^* is *not* assumed to be point identified,
  i.e., given F, there might be a **set** of θ that satisfy (2).

$$\Theta_I \equiv \left\{\theta \in \Theta \text{ s.t. } \mathbb{E}[m_j(W,\theta)] \leqslant 0 \text{ for } j = 1, \ldots, p \ \right\}.$$

where $p = p_I + 2p_E$.

Introduction

- ▷ We consider a model characterized by (θ^*, F)
  - $\theta \in \mathbb{R}^{d_\theta}$ is a finite dimensional parameter of interest,
  - F is the distribution of data, i.e., $W \sim F$ (W_i, $i = 1, \ldots, n$, i.i.d.)
- ▷ The *main prediction* of the model is that the true parameter θ^* satisfies

$$\begin{aligned} \mathbb{E}[m_j(W,\theta)] &\leqslant 0, \quad \text{for } \; j = 1, \ldots, p_I, \\ \mathbb{E}[m_j(W,\theta)] &= 0, \quad \text{for } \; j = p_I + 1, \ldots, p_I + p_E. \end{aligned} \tag{2}$$

- ▷ key issue: θ^* is *not* assumed to be point identified,
  i.e., given F, there might be a **set** of θ that satisfy (2).

$$\Theta_I \equiv \left\{\theta \in \Theta \text{ s.t. } \mathbb{E}[m_j(W,\theta)] \leqslant 0 \text{ for } j = 1, \ldots, p \;\right\}.$$

  where $p = p_I + 2p_E$.

- ▷ We formally deal with unconditional moments

Introduction

- ▷ We consider a model characterized by (θ^*, F)
  - $\theta \in \mathbb{R}^{d_\theta}$ is a finite dimensional parameter of interest,
  - F is the distribution of data, i.e., $W \sim F$ (W_i, $i = 1, \ldots, n$, i.i.d.)
- ▷ The *main prediction* of the model is that the true parameter θ^* satisfies

$$\begin{aligned} \mathbb{E}[m_j(W,\theta)] &\leqslant 0, \quad \text{for } j = 1, \ldots, p_I, \\ \mathbb{E}[m_j(W,\theta)] &= 0, \quad \text{for } j = p_I + 1, \ldots, p_I + p_E. \end{aligned} \tag{2}$$

- ▷ key issue: θ^* is *not* assumed to be point identified,
  i.e., given F, there might be a **set** of θ that satisfy (2).

$$\Theta_I \equiv \left\{\theta \in \Theta \text{ s.t. } \mathbb{E}[m_j(W,\theta)] \leqslant 0 \text{ for } j = 1, \ldots, p \ \right\}.$$

  where $p = p_I + 2p_E$.

- ▷ We formally deal with unconditional moments
  but conditional moments can be approximated via

$$\mathbb{E}[m_j(W,\theta^*) \mid z_i] \leqslant 0 \Rightarrow \mathbb{E}[m_j(W,\theta^*)1\{z_i \in [a,b]\}] \leqslant 0 \ \text{ for all } [a,b]$$

Introduction (ctd.)

- The literature focuses on inference on PI parameter vector $\theta^* \in \Theta$

Introduction (ctd.)

- The literature focuses on inference on PI parameter vector $\theta^* \in \Theta$, i.e.,

$$H_0 : \theta^* = \theta_0 \quad \text{vs.} \quad H_1 : \theta^* \neq \theta_0.$$

Introduction (ctd.)

- The literature focuses on inference on PI parameter vector $\theta^* \in \Theta$, i.e.,

$$H_0 : \theta^* = \theta_0 \quad \text{vs.} \quad H_1 : \theta^* \neq \theta_0.$$

- We are not interested on θ^* but on $h(\theta^*)$ for a *known* fn. $h : \Theta \to \Lambda$. This is the problem addressed in this paper:

  Hypothesis test (HT): For fixed h_0, we want to test:

$$H_0 : h(\theta^*) = h_0 \quad \text{vs.} \quad H_1 : h(\theta^*) \neq h_0 \tag{3}$$

Introduction (ctd.)

- The literature focuses on inference on PI parameter vector $\theta^* \in \Theta$, i.e.,

$$H_0 : \theta^* = \theta_0 \quad \text{vs.} \quad H_1 : \theta^* \neq \theta_0.$$

- We are not interested on θ^* but on $h(\theta^*)$ for a *known* fn. $h : \Theta \to \Lambda$. This is the problem addressed in this paper:

  Hypothesis test (HT): For fixed h_0, we want to test:

$$H_0 : h(\theta^*) = h_0 \quad \text{vs.} \quad H_1 : h(\theta^*) \neq h_0 \tag{3}$$

  Confidence set (CS) for $h(\theta^*)$: based on HT inversion of a test for (3).

Introduction (ctd.)

- The literature focuses on inference on PI parameter vector $\theta^* \in \Theta$, i.e.,

$$H_0 : \theta^* = \theta_0 \quad \text{vs.} \quad H_1 : \theta^* \neq \theta_0.$$

- We are not interested on θ^* but on $h(\theta^*)$ for a *known* fn. $h : \Theta \to \Lambda$. This is the problem addressed in this paper:

  Hypothesis test (HT): For fixed h_0, we want to test:

$$H_0 : h(\theta^*) = h_0 \quad \text{vs.} \quad H_1 : h(\theta^*) \neq h_0 \tag{3}$$

  Confidence set (CS) for $h(\theta^*)$: based on HT inversion of a test for (3).

- Main application: *Subvector inference*: For $\theta^* \in \Theta \subset \mathbb{R}^{d_\theta}$, $d_\theta > 1$,

$$H_0 \ : \ \theta_1^* = h_0 \quad \text{vs.} \quad H_1 \ : \ \theta_1^* \neq h_0.$$

  $\Rightarrow$ Special case of Eq. (3) with $h(\theta) = \theta_1$ and $h_0 \in \Lambda \subseteq \mathbb{R}$.

Literature review

Most of literature on inference in PI moment (in)eq. is on "vector inference"

$$H_0 : \theta^* = \theta_0 \text{ vs. } H_1 : \theta^* \neq \theta_0$$

Literature review

Most of literature on inference in PI moment (in)eq. is on "vector inference"

$$H_0 : \theta^* = \theta_0 \text{ vs. } H_1 : \theta^* \neq \theta_0$$

- **Testing unconditional moment inequalities:** Chernozhukov, Hong, and Tamer (2007), Romano and Shaikh (2008), Andrews and Guggenberger (2009), Andrews and Soares (2010), Canay (2010), Bugni (2011), Andrews and Jia Barwick (2012), Romano, Shaikh, and Wolf (2012)
- **Testing conditional moment inequalities:** Andrews and Shi (2013), Chernozhukov, Lee, and Rosen (2013), Armstrong (2011), Chetverikov (2011), Armstrong and Chan (2012)

Literature review

Most of literature on inference in PI moment (in)eq. is on "vector inference"

$$H_0 : \theta^* = \theta_0 \text{ vs. } H_1 : \theta^* \neq \theta_0$$

- **Testing unconditional moment inequalities:** Chernozhukov, Hong, and Tamer (2007), Romano and Shaikh (2008), Andrews and Guggenberger (2009), Andrews and Soares (2010), Canay (2010), Bugni (2011), Andrews and Jia Barwick (2012), Romano, Shaikh, and Wolf (2012)
- **Testing conditional moment inequalities:** Andrews and Shi (2013), Chernozhukov, Lee, and Rosen (2013), Armstrong (2011), Chetverikov (2011), Armstrong and Chan (2012)

In both cases, the number of moments p is fixed (explicitly or due to the structure)

Testing unconditional moment inequalities with $p \to \infty$

- Menzel (2014), where $p \ll n$
- Chernozhukov, Chetverikov and Kato (WP 2013), where $p \gg n$

Literature review: subvector

Asymptotically uniformly valid inference for

$$H_0 : h(\theta^*) = h_0 \ \text{ vs. } \ H_1 : h(\theta^*) \neq h_0$$

- **Projections of CS:** Project usual CS for θ onto space of H_0. Considered by Andrews et. al. (2009, 10).
  - *Related work improving projections:*
    Kaido, Molinari & Stoye (WP, 2015), Gafarov (affine models, WP 2017)
- **Subsampling:** Profile the criterion function and approximate critical value with subsampling. Proposed by Romano & Shaikh (2008, 10).
- **Project the Criterion Function**: Bugni, Canay and Shi (2017)

Literature review: subvector

Asymptotically uniformly valid inference for

$$H_0 : h(\theta^*) = h_0 \ \text{ vs. } \ H_1 : h(\theta^*) \neq h_0$$

▷ **Projections of CS:** Project usual CS for θ onto space of H_0. Considered by Andrews et. al. (2009, 10).

- *Related work improving projections:*
  Kaido, Molinari & Stoye (WP, 2015), Gafarov (affine models, WP 2017)

▷ **Subsampling:** Profile the criterion function and approximate critical value with subsampling. Proposed by Romano & Shaikh (2008, 10).

▷ **Project the Criterion Function**: Bugni, Canay and Shi (2017)

In all cases, the number of moments inequalities p **is fixed** and asymptotic analysis (i.e., based on the limiting distribution of the process)

Positioning in the literature

	Donsker (e.g. p fixed)	non-Donsker (e.g. p growing)
Vector Inference	Chernozhukov et al (2007) Romano and Shaikh (2008) Andrews and Guggenberger (2009) Andrews and Soares (2010) ...	Menzel (2014, $p \ll n$) Chernozhukov et al (WP 2013, $p \gg n$)
Subvector Inference	Andrews et. al. (2009, 10) Romano and Shaikh (2008, 10) Bugni, Canay and Shi (2017) Kaido et al (WP, 2015) Gafarov (WP 2017)	

Positioning in the literature

	Donsker (e.g. p fixed)	non-Donsker (e.g. p growing)
Vector Inference	Chernozhukov et al (2007) Romano and Shaikh (2008) Andrews and Guggenberger (2009) Andrews and Soares (2010) ...	Menzel (2014, $p \ll n$) Chernozhukov et al (WP 2013, $p \gg n$)
Subvector Inference	Andrews et. al. (2009, 10) Romano and Shaikh (2008, 10) Bugni, Canay and Shi (2017) Kaido et al (WP, 2015) Gafarov (WP 2017)	

Starting point:

- ▷ the minimum resampling critical value in Bugni, Canay and Shi (2017); and
- ▷ CLTs for the max of high-dim vectors used in Chernozhukov et al (WP 2013)

Setting and Contributions

Profiled test statistics for $H_0 : h(\theta^*) = h_0$ vs. $H_1 : h(\theta^*) \neq h_0$

$$T_n(h_0) = \inf_{\theta \in \Theta(h_0)} \max_{j \in [p]} \sqrt{n} \bar{m}_{\theta,j} / \widehat{\sigma}_{\theta,j}$$

where $\Theta(h_0) = h^{-1}(h_0) = \{\theta \in \Theta : h(\theta) = h_0\}$ and

$$\bar{m}_{\theta,j} = \frac{1}{n} \sum_{i=1}^{n} m_j(W_i, \theta) \quad \text{and} \quad \widehat{\sigma}^2_{\theta,j} = \frac{1}{n} \sum_{i=1}^{n} \{m_j(W_i, \theta) - \bar{m}_{\theta,j}\}^2$$

The test: reject if $T_n(h_0) > c_n(h_0, 1-\alpha)$ where $c_n(h_0, 1-\alpha)$ is a critical value.

Setting and Contributions

Profiled test statistics for $H_0 : h(\theta^*) = h_0$ vs. $H_1 : h(\theta^*) \neq h_0$

$$T_n(h_0) = \inf_{\theta \in \Theta(h_0)} \max_{j \in [p]} \sqrt{n} \bar{m}_{\theta,j} / \widehat{\sigma}_{\theta,j}$$

where $\Theta(h_0) = h^{-1}(h_0) = \{\theta \in \Theta : h(\theta) = h_0\}$ and

$$\bar{m}_{\theta,j} = \frac{1}{n} \sum_{i=1}^{n} m_j(W_i, \theta) \quad \text{and} \quad \widehat{\sigma}^2_{\theta,j} = \frac{1}{n} \sum_{i=1}^{n} \{m_j(W_i, \theta) - \bar{m}_{\theta,j}\}^2$$

The test: reject if $T_n(h_0) > c_n(h_0, 1-\alpha)$ where $c_n(h_0, 1-\alpha)$ is a critical value.

Under H_0: $\qquad P(T_n(h_0) > c_n(h_0, 1-\alpha)) \leqslant \alpha + o(1)$

Setting and Contributions

Profiled test statistics for $H_0 : h(\theta^*) = h_0$ vs. $H_1 : h(\theta^*) \neq h_0$

$$T_n(h_0) = \inf_{\theta \in \Theta(h_0)} \max_{j \in [p]} \sqrt{n}\bar{m}_{\theta,j}/\widehat{\sigma}_{\theta,j}$$

where $\Theta(h_0) = h^{-1}(h_0) = \{\theta \in \Theta : h(\theta) = h_0\}$ and

$$\bar{m}_{\theta,j} = \frac{1}{n}\sum_{i=1}^{n} m_j(W_i, \theta) \quad \text{and} \quad \widehat{\sigma}^2_{\theta,j} = \frac{1}{n}\sum_{i=1}^{n}\{m_j(W_i,\theta) - \bar{m}_{\theta,j}\}^2$$

The test: reject if $T_n(h_0) > c_n(h_0, 1-\alpha)$ where $c_n(h_0, 1-\alpha)$ is a critical value.

Under H_0: $\qquad P(T_n(h_0) > c_n(h_0, 1-\alpha)) \leqslant \alpha + o(1)$

Our contribution is to construct critical values $c_n(h_0, 1-\alpha)$ that

- ▷ uniformly controls asymptotic size over a large class of dgps ($F \in \mathcal{P}_n$)
- ▷ in the presence of many moment inequalities ($p \to \infty$ as $n \to \infty$)

Setting and Contributions

Profiled test statistics for $H_0 : h(\theta^*) = h_0$ vs. $H_1 : h(\theta^*) \neq h_0$

$$T_n(h_0) = \inf_{\theta \in \Theta(h_0)} \max_{j \in [p]} \sqrt{n}\bar{m}_{\theta,j}/\widehat{\sigma}_{\theta,j}$$

where $\Theta(h_0) = h^{-1}(h_0) = \{\theta \in \Theta : h(\theta) = h_0\}$ and

$$\bar{m}_{\theta,j} = \frac{1}{n}\sum_{i=1}^{n} m_j(W_i,\theta) \quad \text{and} \quad \widehat{\sigma}^2_{\theta,j} = \frac{1}{n}\sum_{i=1}^{n}\{m_j(W_i,\theta) - \bar{m}_{\theta,j}\}^2$$

The test: reject if $T_n(h_0) > c_n(h_0, 1-\alpha)$ where $c_n(h_0, 1-\alpha)$ is a critical value.

Under H_0: $$P(T_n(h_0) > c_n(h_0, 1-\alpha)) \leqslant \alpha + o(1)$$

Our contribution is to construct critical values $c_n(h_0, 1-\alpha)$ that

- uniformly controls asymptotic size over a large class of dgps ($F \in \mathcal{P}_n$)
- in the presence of many moment inequalities ($p \to \infty$ as $n \to \infty$)
- allow for $p \gg n$ (also $d_\theta \to \infty$ but not clear if empirically relevant)
- finite sample analysis, and rate for size error (e.g. polynomially in n)
- towards data-driven choice of parameters

Overview of Proposals

Profiled test statistics for $H_0 : h(\theta^*) = h_0$ vs. $H_1 : h(\theta^*) \neq h_0$

$$T_n(h_0) = \inf_{\theta \in \Theta(h_0)} \max_{j \in [p]} \sqrt{n} \bar{m}_{\theta,j} / \widehat{\sigma}_{\theta,j}$$

We consider different methods to calculate the critical value $c_n(h_0, 1-\alpha)$:

▷ Self-Normalized method (not covering today)

- fast
- works under very weak conditions
- potentially conservative

▷ Bootstrap-based methods

- slower (requires simulations)
- requires stronger conditions
- but less conservative

▷ Hybrids are possible (not covering today)

- potentially useful to speed up bootstrap-based methods

Proposal via Bootstrap-based methods

Profiled test statistics for $H_0 : h(\theta^*) = h_0$ vs. $H_1 : h(\theta^*) \neq h_0$

$$T_n(h_0) = \inf_{\theta \in \Theta(h_0)} \max_{j \in [p]} \sqrt{n} \bar{m}_{\theta,j} / \widehat{\sigma}_{\theta,j}$$

Proposal via Bootstrap-based methods

Profiled test statistics for $H_0 : h(\theta^*) = h_0$ vs. $H_1 : h(\theta^*) \neq h_0$

$$T_n(h_0) = \inf_{\theta \in \Theta(h_0)} \max_{j \in [p]} \sqrt{n}\bar{m}_{\theta,j}/\widehat{\sigma}_{\theta,j}$$

Letting $\widehat{v}_{\theta,j} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \{m_j(W_i, \theta) - \mathbb{E}[m_j(W_i, \theta)]\}/\widehat{\sigma}_{\theta,j}$

Proposal via Bootstrap-based methods

Profiled test statistics for $H_0 : h(\theta^*) = h_0$ vs. $H_1 : h(\theta^*) \neq h_0$

$$T_n(h_0) = \inf_{\theta \in \Theta(h_0)} \max_{j \in [p]} \sqrt{n}\bar{m}_{\theta,j}/\widehat{\sigma}_{\theta,j}$$

Letting $\widehat{v}_{\theta,j} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \{m_j(W_i,\theta) - \mathbb{E}[m_j(W_i,\theta)]\}/\widehat{\sigma}_{\theta,j}$ we can rewrite $T_n(h_0)$ as

$$T_n(h_0) = \inf_{\theta \in \Theta(h_0)} \max_{j \in [p]} \quad \widehat{v}_{\theta,j} + \sqrt{n}\mathbb{E}[m_j(W,\theta)]/\widehat{\sigma}_{\theta,j}$$

bootstrap: $\widehat{v}^*_{\theta,j} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_i \frac{m_j(W_i,\theta) - \bar{m}_{\theta,j}}{\widehat{\sigma}_{\theta,j}}$ where ξ_i's are i.i.d. $N(0,1)$.

Proposal via Bootstrap-based methods

Profiled test statistics for $H_0 : h(\theta^*) = h_0$ vs. $H_1 : h(\theta^*) \neq h_0$

$$T_n(h_0) = \inf_{\theta \in \Theta(h_0)} \max_{j \in [p]} \sqrt{n}\bar{m}_{\theta,j}/\widehat{\sigma}_{\theta,j}$$

Letting $\widehat{v}_{\theta,j} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \{m_j(W_i,\theta) - \mathbb{E}[m_j(W_i,\theta)]\}/\widehat{\sigma}_{\theta,j}$ we can rewrite $T_n(h_0)$ as

$$T_n(h_0) = \inf_{\theta \in \Theta(h_0)} \max_{j \in [p]} \quad \widehat{v}_{\theta,j} + \sqrt{n}\mathbb{E}[m_j(W,\theta)]/\widehat{\sigma}_{\theta,j}$$

bootstrap: $\widehat{v}^*_{\theta,j} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_i \frac{m_j(W_i,\theta) - \bar{m}_{\theta,j}}{\widehat{\sigma}_{\theta,j}}$ where ξ_i's are i.i.d. $N(0,1)$.

Remark: It has been shown that although we can suitably approximate

$$\widehat{v}_{\theta,j} \text{ by } \widehat{v}^*_{\theta,j}$$

Proposal via Bootstrap-based methods

Profiled test statistics for $H_0 : h(\theta^*) = h_0$ vs. $H_1 : h(\theta^*) \neq h_0$

$$T_n(h_0) = \inf_{\theta \in \Theta(h_0)} \max_{j \in [p]} \sqrt{n}\bar{m}_{\theta,j}/\widehat{\sigma}_{\theta,j}$$

Letting $\widehat{v}_{\theta,j} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \{m_j(W_i,\theta) - \mathbb{E}[m_j(W_i,\theta)]\}/\widehat{\sigma}_{\theta,j}$ we can rewrite $T_n(h_0)$ as

$$T_n(h_0) = \inf_{\theta \in \Theta(h_0)} \max_{j \in [p]} \quad \widehat{v}_{\theta,j} + \sqrt{n}\mathbb{E}[m_j(W,\theta)]/\widehat{\sigma}_{\theta,j}$$

bootstrap: $\widehat{v}^*_{\theta,j} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_i \frac{m_j(W_i,\theta) - \bar{m}_{\theta,j}}{\widehat{\sigma}_{\theta,j}}$ where ξ_i's are i.i.d. $N(0,1)$.

Remark: It has been shown that although we can suitably approximate

$$\widehat{v}_{\theta,j} \text{ by } \widehat{v}^*_{\theta,j}$$

Andrews and Soares (2010) show it is more delicate to approximate

$$\sqrt{n}\mathbb{E}[m_j(W,\theta)]/\widehat{\sigma}_{\theta,j} \quad \text{by} \quad \sqrt{n}\bar{m}_{\theta,j}/\widehat{\sigma}_{\theta,j}$$

as there is a non-vanishing noise due to the scaling by $\sqrt{n}$.

Standard Bootstrap-based methods via GMS

Profiled test statistics for $H_0 : h(\theta^*) = h_0$ vs. $H_1 : h(\theta^*) \neq h_0$

$$T_n(h_0) = \inf_{\theta \in \Theta(h_0)} \max_{j \in [p]} \quad \widehat{v}_{\theta,j} + \sqrt{n}\mathbb{E}[m_j(W,\theta)]/\widehat{\sigma}_{\theta,j}$$

Standard Bootstrap-based methods via GMS

Profiled test statistics for $H_0 : h(\theta^*) = h_0$ vs. $H_1 : h(\theta^*) \neq h_0$

$$T_n(h_0) = \inf_{\theta \in \Theta(h_0)} \max_{j \in [p]} \quad \widehat{v}_{\theta,j} + \sqrt{n}\mathbb{E}[m_j(W,\theta)]/\widehat{\sigma}_{\theta,j}$$

A standard way to proceed is to use Generalized Moment Selection (GMS)

$$\varphi_{\theta,j} = \begin{cases} 0, \quad \text{if } \sqrt{n}\bar{m}_{\theta,j}/\widehat{\sigma}_{\theta,j} \geqslant -\kappa_n, \\ \\ -\infty, \quad \text{otherwise} \quad \text{(i.e., inequality will not be used)} \end{cases}$$

for a tuning parameter $\kappa_n \to \infty$

Standard Bootstrap-based methods via GMS

Profiled test statistics for $H_0 : h(\theta^*) = h_0$ vs. $H_1 : h(\theta^*) \neq h_0$

$$T_n(h_0) = \inf_{\theta \in \Theta(h_0)} \max_{j \in [p]} \quad \widehat{v}_{\theta,j} + \sqrt{n}\mathbb{E}[m_j(W,\theta)]/\widehat{\sigma}_{\theta,j}$$

A standard way to proceed is to use Generalized Moment Selection (GMS)

$$\varphi_{\theta,j} = \begin{cases} 0, & \text{if } \sqrt{n}\bar{m}_{\theta,j}/\widehat{\sigma}_{\theta,j} \geqslant -\kappa_n, \\ -\infty, & \text{otherwise} \quad \text{(i.e., inequality will not be used)} \end{cases}$$

for a tuning parameter $\kappa_n \to \infty$ (recommendation $\sim \{\log n\}^{1/2}$ when p is fixed)

Then set

$$T_n^{GMS*}(h_0) := \inf_{\theta \in \Theta(h_0)} \max_{j \in [p]} \quad \widehat{v}^*_{\theta,j} + \varphi_{\theta,j}$$

and compute the critical values based on the quantile of $T_n^{GMS*}(h_0)$.

Standard Bootstrap-based methods via GMS

Example: $d_\theta = 2$, and $\Theta = [-1, 1]^2$. Let $p = 2$, and consider

$$\begin{aligned} \mathbb{E}[m_1(W_i, \theta)] &= \mathbb{E}[\theta_1 + \theta_2 - W_{i,1}] \leqslant 0 \\ \mathbb{E}[m_2(W_i, \theta)] &= \mathbb{E}[W_{i,2} - \theta_1 - \theta_2] \leqslant 0 \end{aligned}$$

where $W_i \in \mathbb{R}^p$, $W_i \sim N(0, I)$ and we are interest on testing

$$H_0 : \theta_1 = 0 \quad vs. \quad H_1 : \theta_1 \neq 0$$

Standard Bootstrap-based methods via GMS

Example: $d_\theta = 2$, and $\Theta = [-1, 1]^2$. Let $p = 2$, and consider

$$\begin{aligned} \mathbb{E}[m_1(W_i, \theta)] &= \mathbb{E}[\theta_1 + \theta_2 - W_{i,1}] \leqslant 0 \\ \mathbb{E}[m_2(W_i, \theta)] &= \mathbb{E}[W_{i,2} - \theta_1 - \theta_2] \leqslant 0 \end{aligned}$$

where $W_i \in \mathbb{R}^p$, $W_i \sim N(0, I)$ and we are interest on testing

$$H_0 : \theta_1 = 0 \quad vs. \quad H_1 : \theta_1 \neq 0 \qquad \text{so that}$$

$$\Theta(h_0) = \{\theta \in \Theta : \theta_1 = 0\} \quad \text{and} \quad \Theta_I = \{\theta \in \Theta : \theta_1 + \theta_2 = 0\}$$

Standard Bootstrap-based methods via GMS

Example: $d_\theta = 2$, and $\Theta = [-1, 1]^2$. Let $p = 2$, and consider

$$\begin{aligned}\mathbb{E}[m_1(W_i, \theta)] &= \mathbb{E}[\theta_1 + \theta_2 - W_{i,1}] \leqslant 0 \\ \mathbb{E}[m_2(W_i, \theta)] &= \mathbb{E}[W_{i,2} - \theta_1 - \theta_2] \leqslant 0\end{aligned}$$

where $W_i \in \mathbb{R}^p$, $W_i \sim N(0, I)$ and we are interest on testing

$$H_0 : \theta_1 = 0 \quad vs. \quad H_1 : \theta_1 \neq 0 \qquad \text{so that}$$
$$\Theta(h_0) = \{\theta \in \Theta : \theta_1 = 0\} \quad \text{and} \quad \Theta_I = \{\theta \in \Theta : \theta_1 + \theta_2 = 0\}$$

It follows that for $(Z_1, Z_2) \sim N(0, I)$ we have

$$T_n(0) = \inf_{-1 \leqslant \theta_2 \leqslant 1} \max\left\{ \frac{\sqrt{n}\theta_2 - \bar{W}_1}{\widehat{\sigma}_1}, \frac{\bar{W}_2 - \sqrt{n}\theta_2}{\widehat{\sigma}_2} \right\}$$

Standard Bootstrap-based methods via GMS

Example: $d_\theta = 2$, and $\Theta = [-1,1]^2$. Let $p = 2$, and consider

$$\begin{aligned} \mathbb{E}[m_1(W_i,\theta)] &= \mathbb{E}[\theta_1 + \theta_2 - W_{i,1}] \leqslant 0 \\ \mathbb{E}[m_2(W_i,\theta)] &= \mathbb{E}[W_{i,2} - \theta_1 - \theta_2] \leqslant 0 \end{aligned}$$

where $W_i \in \mathbb{R}^p$, $W_i \sim N(0,I)$ and we are interest on testing

$$H_0 : \theta_1 = 0 \quad vs. \quad H_1 : \theta_1 \neq 0 \qquad \text{so that}$$

$$\Theta(h_0) = \{\theta \in \Theta : \theta_1 = 0\} \quad \text{and} \quad \Theta_I = \{\theta \in \Theta : \theta_1 + \theta_2 = 0\}$$

It follows that for $(Z_1, Z_2) \sim N(0,I)$ we have

$$T_n(0) = \inf_{-1 \leqslant \theta_2 \leqslant 1} \max\left\{ \frac{\sqrt{n}\theta_2 - \bar{W}_1}{\widehat{\sigma}_1}, \frac{\bar{W}_2 - \sqrt{n}\theta_2}{\widehat{\sigma}_2} \right\} \to_d \frac{Z_2 - Z_1}{2} \sim N(0, 1/2)$$

Standard Bootstrap-based methods via GMS

Example: $d_\theta = 2$, and $\Theta = [-1,1]^2$. Let $p = 2$, and consider

$$\begin{aligned} \mathbb{E}[m_1(W_i,\theta)] &= \mathbb{E}[\theta_1 + \theta_2 - W_{i,1}] \leqslant 0 \\ \mathbb{E}[m_2(W_i,\theta)] &= \mathbb{E}[W_{i,2} - \theta_1 - \theta_2] \leqslant 0 \end{aligned}$$

where $W_i \in \mathbb{R}^p$, $W_i \sim N(0,I)$ and we are interest on testing

$$H_0 : \theta_1 = 0 \;\; vs. \;\; H_1 : \theta_1 \neq 0 \quad \text{so that}$$
$$\Theta(h_0) = \{\theta \in \Theta : \theta_1 = 0\} \;\text{ and }\; \Theta_I = \{\theta \in \Theta : \theta_1 + \theta_2 = 0\}$$

It follows that for $(Z_1, Z_2) \sim N(0,I)$ we have

$$T_n(0) = \inf_{-1 \leqslant \theta_2 \leqslant 1} \max \left\{ \frac{\sqrt{n}\theta_2 - \bar{W}_1}{\widehat{\sigma}_1}, \frac{\bar{W}_2 - \sqrt{n}\theta_2}{\widehat{\sigma}_2} \right\} \to_d \frac{Z_2 - Z_1}{2} \sim N(0, 1/2)$$

In turn, for GMS, using $\kappa_n = \sqrt{\log n}$ we select both inequalities whp and

$$T_n^{GMS*}(0) \mid (W_i)_{i=1}^n \approx \inf_{-1 \leqslant \theta_2 \leqslant 1} \max \{-Z_1 + \varphi_{\theta_2,1}, Z_2 + \varphi_{\theta_2,2}\}$$

Standard Bootstrap-based methods via GMS

Example: $d_\theta = 2$, and $\Theta = [-1,1]^2$. Let $p = 2$, and consider

$$\begin{aligned} \mathbb{E}[m_1(W_i,\theta)] &= \mathbb{E}[\theta_1 + \theta_2 - W_{i,1}] \leqslant 0 \\ \mathbb{E}[m_2(W_i,\theta)] &= \mathbb{E}[W_{i,2} - \theta_1 - \theta_2] \leqslant 0 \end{aligned}$$

where $W_i \in \mathbb{R}^p$, $W_i \sim N(0,I)$ and we are interest on testing

$$H_0 : \theta_1 = 0 \quad vs. \quad H_1 : \theta_1 \neq 0 \qquad \text{so that}$$
$$\Theta(h_0) = \{\theta \in \Theta : \theta_1 = 0\} \quad \text{and} \quad \Theta_I = \{\theta \in \Theta : \theta_1 + \theta_2 = 0\}$$

It follows that for $(Z_1, Z_2) \sim N(0,I)$ we have

$$T_n(0) = \inf_{-1 \leqslant \theta_2 \leqslant 1} \max\left\{ \frac{\sqrt{n}\theta_2 - \bar{W}_1}{\widehat{\sigma}_1}, \frac{\bar{W}_2 - \sqrt{n}\theta_2}{\widehat{\sigma}_2} \right\} \to_d \frac{Z_2 - Z_1}{2} \sim N(0, 1/2)$$

In turn, for GMS, using $\kappa_n = \sqrt{\log n}$ we select both inequalities whp and

$$T_n^{GMS*}(0) \mid (W_i)_{i=1}^n \to_d \min\{-Z_1, Z_2\} \quad whp$$

Standard Bootstrap-based methods via GMS

Example: $d_\theta = 2$, and $\Theta = [-1, 1]^2$. Let $p = 2$, and consider

$$\begin{aligned} \mathbb{E}[m_1(W_i, \theta)] &= \mathbb{E}[\theta_1 + \theta_2 - W_{i,1}] \leqslant 0 \\ \mathbb{E}[m_2(W_i, \theta)] &= \mathbb{E}[W_{i,2} - \theta_1 - \theta_2] \leqslant 0 \end{aligned}$$

where $W_i \in \mathbb{R}^p$, $W_i \sim N(0, I)$ and we are interest on testing

$$H_0 : \theta_1 = 0 \;\; vs. \;\; H_1 : \theta_1 \neq 0 \quad \text{so that}$$
$$\Theta(h_0) = \{\theta \in \Theta : \theta_1 = 0\} \;\text{ and }\; \Theta_I = \{\theta \in \Theta : \theta_1 + \theta_2 = 0\}$$

It follows that for $(Z_1, Z_2) \sim N(0, I)$ we have

$$T_n(0) = \inf_{-1 \leqslant \theta_2 \leqslant 1} \max\left\{ \frac{\sqrt{n}\theta_2 - \bar{W}_1}{\widehat{\sigma}_1}, \frac{\bar{W}_2 - \sqrt{n}\theta_2}{\widehat{\sigma}_2} \right\} \to_d \frac{Z_2 - Z_1}{2} \sim N(0, 1/2)$$

In turn, for GMS, using $\kappa_n = \sqrt{\log n}$ we select both inequalities whp and

$$T_n^{GMS*}(0) \mid (W_i)_{i=1}^n \to_d \min\{-Z_1, Z_2\} \;\; whp$$

Critical values based on $T_n^{GMS*}(0)$ fail to control size.

Standard Bootstrap-based methods via GMS

Example: $d_\theta = 2$, and $\Theta = [-1,1]^2$. Let $p = 2$, and consider

$$\begin{aligned} \mathbb{E}[m_1(W_i,\theta)] &= \mathbb{E}[\theta_1 + \theta_2 - W_{i,1}] \leqslant 0 \\ \mathbb{E}[m_2(W_i,\theta)] &= \mathbb{E}[W_{i,2} - \theta_1 - \theta_2] \leqslant 0 \end{aligned}$$

where $W_i \in \mathbb{R}^p$, $W_i \sim N(0,I)$ and we are interest on testing

$$H_0 : \theta_1 = 0 \;\; vs. \;\; H_1 : \theta_1 \neq 0 \quad \text{so that}$$
$$\Theta(h_0) = \{\theta \in \Theta : \theta_1 = 0\} \;\; \text{and} \;\; \Theta_I = \{\theta \in \Theta : \theta_1 + \theta_2 = 0\}$$

It follows that for $(Z_1, Z_2) \sim N(0,I)$ we have

$$T_n(0) = \inf_{-1 \leqslant \theta_2 \leqslant 1} \max \left\{ \frac{\sqrt{n}\theta_2 - \bar{W}_1}{\widehat{\sigma}_1}, \frac{\bar{W}_2 - \sqrt{n}\theta_2}{\widehat{\sigma}_2} \right\} \to_d \frac{Z_2 - Z_1}{2} \sim N(0, 1/2)$$

In turn, for GMS, using $\kappa_n = \sqrt{\log n}$ we select both inequalities whp and

$$T_n^{GMS*}(0) \mid (W_i)_{i=1}^n \to_d \min\{-Z_1, Z_2\} \;\; whp$$

Critical values based on $T_n^{GMS*}(0)$ fail to control size. Indeed, for $\alpha = 0.1$

- $c_n^{GMS}(0, 1-\alpha) \approx 0.5$ and $P(T_n(0) > c_n^{GMS}(0, 1-\alpha)) \approx 0.24$.
- $c_n(0, 1-\alpha) \approx 0.86$
- GMS quantiles are "too" small

Bootstrap-based methods for subvector inference

Bootstrap-based methods for subvector inference

1) **"Discard Resampling" (DR):**

$$T_n^{DR*}(h_0) \equiv \inf_{\theta \in \widehat{\Theta}_I(h_0)} \max_{j \in [p]} \widehat{v}^*_{\theta,j} + \varphi_{\theta,j}$$

where $\widehat{\Theta}_I(h_0) \subseteq$ "$\arg\min$"$_{\theta \in \Theta(h_0)} \max_{j \in [p]} \sqrt{n}\bar{m}_{\theta,j}/\widehat{\sigma}_{\theta,j}$

$$\varphi_{\theta,j} = \begin{cases} 0, & \text{if } \sqrt{n}\bar{m}_{\theta,j}/\widehat{\sigma}_{\theta,j} \geqslant \max_{\ell \in [p]} \sqrt{n}\bar{m}_{\theta,\ell}/\widehat{\sigma}_{\theta,\ell} - \kappa_n, \\ -\infty, & \text{otherwise} \quad \text{(i.e., inequality will not be used)} \end{cases}$$

Bootstrap-based methods for subvector inference

1) **"Discard Resampling" (DR):**

$$T_n^{DR*}(h_0) \equiv \inf_{\theta \in \widehat{\Theta}_I(h_0)} \max_{j \in [p]} \widehat{v}^*_{\theta,j} + \varphi_{\theta,j}$$

where $\widehat{\Theta}_I(h_0) \subseteq$ "$\arg\min$"$_{\theta \in \Theta(h_0)} \max_{j \in [p]} \sqrt{n}\bar{m}_{\theta,j}/\widehat{\sigma}_{\theta,j}$

$$\varphi_{\theta,j} = \begin{cases} 0, & \text{if } \sqrt{n}\bar{m}_{\theta,j}/\widehat{\sigma}_{\theta,j} \geqslant \max_{\ell \in [p]} \sqrt{n}\bar{m}_{\theta,\ell}/\widehat{\sigma}_{\theta,\ell} - \kappa_n, \\ -\infty, & \text{otherwise} \quad \text{(i.e., inequality will not be used)} \end{cases}$$

2) **"Penalized Resampling" (PR):**

$$T_n^{PR*}(h_0) \equiv \inf_{\theta \in \Theta(h_0)} \max_{j \in [p]} \quad \widehat{v}^*_{\theta,j} + \kappa_n^{-1}\sqrt{n}\bar{m}_{\theta,j}/\widehat{\sigma}_{\theta,j},$$

where $\kappa_n \geqslant 1$ is a penalty parameter

Bootstrap-based methods for subvector inference

1) **"Discard Resampling" (DR):**

$$T_n^{DR*}(h_0) \equiv \inf_{\theta \in \widehat{\Theta}_I(h_0)} \max_{j \in [p]} \widehat{v}^*_{\theta,j} + \varphi_{\theta,j}$$

where $\widehat{\Theta}_I(h_0) \subseteq$ "$\arg\min$"$_{\theta \in \Theta(h_0)} \max_{j \in [p]} \sqrt{n}\bar{m}_{\theta,j}/\widehat{\sigma}_{\theta,j}$

$$\varphi_{\theta,j} = \begin{cases} 0, & \text{if } \sqrt{n}\bar{m}_{\theta,j}/\widehat{\sigma}_{\theta,j} \geqslant \max_{\ell \in [p]} \sqrt{n}\bar{m}_{\theta,\ell}/\widehat{\sigma}_{\theta,\ell} - \kappa_n, \\ -\infty, & \text{otherwise} \quad \text{(i.e., inequality will not be used)} \end{cases}$$

2) **"Penalized Resampling" (PR):**

$$T_n^{PR*}(h_0) \equiv \inf_{\theta \in \Theta(h_0)} \max_{j \in [p]} \quad \widehat{v}^*_{\theta,j} + \kappa_n^{-1}\sqrt{n}\bar{m}_{\theta,j}/\widehat{\sigma}_{\theta,j},$$

where $\kappa_n \geqslant 1$ is a penalty parameter

3) **"Minimum Resampling" (MR):**

$$T_n^{MR*}(h_0) \equiv \min\{T_n^{DR*}(h_0), T_n^{PR*}(h_0)\}$$

The impact of many moment inequalities, $p \gg n$

- ▷ lack of a Donsker property for the whole process $\{v_{\theta,j} : \theta \in \Theta(h_0), j \in [p]\}$
  - no limiting distributions guaranteed to exist
  - cannot invoke Donsker's functional CLT to establish the convergence in distribution of $T_n(h_0)$

- ▷ restriction on the criterion functions
  - we use $Q(\theta) = \max_{j\in[p]} \sqrt{n}\bar{m}_{\theta,j}/\widehat{\sigma}_{\theta,j}$
  - do not use (MMM): $Q(\theta) = \sum_{j=1}^{p}\{\sqrt{n}\bar{m}_{\theta,j}/\widehat{\sigma}_{\theta,j}\}_+^2$
  - do not use (AQLR): $Q(\theta) = \min_{t\in\mathbb{R}^p}(\sqrt{n}\bar{m}_{\theta,j}/\widehat{\sigma}_{\theta,j} - t)'\widetilde{\Sigma}_\theta^{-1}(\sqrt{n}\bar{m}_{\theta,j}/\widehat{\sigma}_{\theta,j} - t)$

- ▷ tuning parameters need to account for growing entropy

Assumptions for Hypothesis Testing

Condition M.

Assumptions for Hypothesis Testing

Condition M.

(i) $\Theta(h_0) = \{\theta \in \Theta : h(\theta) = h_0\}$ *is well behaved*

Assumptions for Hypothesis Testing

Condition M.

(i) $\Theta(h_0) = \{\theta \in \Theta : h(\theta) = h_0\}$ *is well behaved*

(ii) $\{\widetilde{m}_j(\cdot, \theta) := \sigma_{\theta,j}^{-1} m_j(\cdot, \theta) : \theta \in \Theta(h_0), j \in [p]\}$ *is well behaved*

Assumptions for Hypothesis Testing

Condition M.

(i) $\Theta(h_0) = \{\theta \in \Theta : h(\theta) = h_0\}$ *is well behaved*

(ii) $\{\widetilde{m}_j(\cdot, \theta) := \sigma_{\theta,j}^{-1} m_j(\cdot, \theta) : \theta \in \Theta(h_0), j \in [p]\}$ *is well behaved*

(iii) Polynomial Minorant condition away from the identified set

Assumptions for Hypothesis Testing

Condition M. *The following conditions hold:*
(i) set $\Theta(h_0)$ *is convex and* $\sup_{\theta \in \Theta(h_0)} \|\theta\|_\infty \leqslant C\sqrt{n}$

Assumptions for Hypothesis Testing

Condition M. *The following conditions hold:*
(i) set $\Theta(h_0)$ *is convex and* $\sup_{\theta\in\Theta(h_0)} \|\theta\|_\infty \leqslant C\sqrt{n}$
(ii) $\{\widetilde{m}_j(\cdot,\theta) : \theta \in \Theta(h_0), j \in [p]\}$ *is VC type class of functions*

- *with constants* $\bar{A}$ *and* $v \geqslant 1$ *and envelope* F *(i.e.* $F(W) \geqslant |\widetilde{m}_j(W,\theta)|$*)*
- *for some* $b > 0$, $q \geqslant 4$, *we have*

$$Ep[F^q]^{1/q} \leqslant b \ \ \textit{and} \ \ \mathbb{E}[|\widetilde{m}_j(W,\theta)|^k] \leqslant b^{k-2}, \quad k = 3, 4$$

Assumptions for Hypothesis Testing

Condition M. *The following conditions hold:*
(i) set $\Theta(h_0)$ *is convex and* $\sup_{\theta\in\Theta(h_0)}\|\theta\|_\infty \leqslant C\sqrt{n}$
(ii) $\{\widetilde{m}_j(\cdot,\theta):\theta\in\Theta(h_0), j\in[p]\}$ *is VC type class of functions*

- *with constants* $\bar{A}$ *and* $v\geqslant 1$ *and envelope* F *(i.e.* $F(W)\geqslant|\widetilde{m}_j(W,\theta)|$*)*
- *for some* $b>0$, $q\geqslant 4$, *we have*

$$Ep[F^q]^{1/q}\leqslant b \ \ \text{and}\ \ \mathbb{E}[|\widetilde{m}_j(W,\theta)|^k]\leqslant b^{k-2}, \quad k=3,4$$

- $\mathbb{E}[\{\widetilde{m}_j(W,\theta)-\widetilde{m}_j(W,\tilde{\theta})\}^2]\leqslant L_C\|\theta-\tilde{\theta}\|^\chi$ *for some* $\chi\geqslant 1$.
- $\max_{j\in[p]}\|\nabla_\theta\mathbb{E}[\widetilde{m}_j(W,\theta)]\|\leqslant L_G$ *for every* $\theta\in\Theta(h_0)$

Assumptions for Hypothesis Testing

Condition M. *The following conditions hold:*
(i) set $\Theta(h_0)$ *is convex and* $\sup_{\theta \in \Theta(h_0)} \|\theta\|_\infty \leqslant C\sqrt{n}$
(ii) $\{\widetilde{m}_j(\cdot, \theta) : \theta \in \Theta(h_0), j \in [p]\}$ *is VC type class of functions*

- *with constants* $\bar{A}$ *and* $v \geqslant 1$ *and envelope* F *(i.e.* $F(W) \geqslant |\widetilde{m}_j(W, \theta)|$*)*
- *for some* $b > 0$, $q \geqslant 4$, *we have*

$$Ep[F^q]^{1/q} \leqslant b \;\; \text{and} \;\; \mathbb{E}[|\widetilde{m}_j(W, \theta)|^k] \leqslant b^{k-2}, \quad k = 3, 4$$

- $\mathbb{E}[\{\widetilde{m}_j(W, \theta) - \widetilde{m}_j(W, \tilde{\theta})\}^2] \leqslant L_C \|\theta - \tilde{\theta}\|^\chi$ *for some* $\chi \geqslant 1$.
- $\max_{j \in [p]} \|\nabla_\theta \mathbb{E}[\widetilde{m}_j(W, \theta)]\| \leqslant L_G$ *for every* $\theta \in \Theta(h_0)$

(iii) For every $\theta \in \Theta(h_0) \setminus \Theta_I$ *we have*

$$\max_{j \in [p]} \mathbb{E}\left[\widetilde{m}_j\left(W, \theta\right)\right] \geqslant \vartheta_n \min\left\{\delta, \inf_{\tilde{\theta} \in \Theta(h_0) \cap \Theta_I} \|\theta - \tilde{\theta}\|\right\}$$

Assumptions for Hypothesis Testing

Condition M. *The following conditions hold:*
(i) set $\Theta(h_0)$ *is convex and* $\sup_{\theta \in \Theta(h_0)} \|\theta\|_\infty \leqslant C\sqrt{n}$
(ii) $\{\widetilde{m}_j(\cdot, \theta) : \theta \in \Theta(h_0), j \in [p]\}$ *is VC type class of functions*

- *with constants* $\bar{A}$ *and* $v \geqslant 1$ *and envelope* F *(i.e.* $F(W) \geqslant |\widetilde{m}_j(W, \theta)|$*)*
- *for some* $b > 0$, $q \geqslant 4$, *we have*

$$Ep[F^q]^{1/q} \leqslant b \;\; \textit{and} \;\; \mathbb{E}[|\widetilde{m}_j(W, \theta)|^k] \leqslant b^{k-2}, \quad k = 3, 4$$

- $\mathbb{E}[\{\widetilde{m}_j(W, \theta) - \widetilde{m}_j(W, \tilde{\theta})\}^2] \leqslant L_C \|\theta - \tilde{\theta}\|^\chi$ *for some* $\chi \geqslant 1$.
- $\max_{j \in [p]} \|\nabla_\theta \mathbb{E}[\widetilde{m}_j(W, \theta)]\| \leqslant L_G$ *for every* $\theta \in \Theta(h_0)$

(iii) For every $\theta \in \Theta(h_0) \setminus \Theta_I$ *we have*

$$\max_{j \in [p]} \mathbb{E}\left[\widetilde{m}_j\left(W, \theta\right)\right] \geqslant \vartheta_n \min\left\{\delta, \inf_{\tilde{\theta} \in \Theta(h_0) \cap \Theta_I} \|\theta - \tilde{\theta}\|\right\}$$

Define $\gamma = o(1)$, in particular $\gamma \ll \alpha$

- ▷ $\bar{w}_n = (1 - \gamma)$-quantile of $\sup_{\theta \in \Theta(h_0), j \in [p]} |\widehat{v}^*_{\theta, j}|$
- ▷ $K_n = v \log(n\bar{A}b) + d_\theta \log(nb) + \log p$

Assumptions for Hypothesis Testing

Condition M. *The following conditions hold:*
(i) set $\Theta(h_0)$ *is convex and* $\sup_{\theta\in\Theta(h_0)}\|\theta\|_\infty \leqslant C\sqrt{n}$
(ii) $\{\widetilde{m}_j(\cdot,\theta) : \theta \in \Theta(h_0), j \in [p]\}$ *is VC type class of functions*

- *with constants* $\bar{A}$ *and* $v \geqslant 1$ *and envelope* F *(i.e.* $F(W) \geqslant |\widetilde{m}_j(W,\theta)|$*)*
- *for some* $b > 0$*,* $q \geqslant 4$*, we have*

$$Ep[F^q]^{1/q} \leqslant b \;\; \textit{and} \;\; \mathbb{E}[|\widetilde{m}_j(W,\theta)|^k] \leqslant b^{k-2}, \quad k = 3, 4$$

- $\mathbb{E}[\{\widetilde{m}_j(W,\theta) - \widetilde{m}_j(W,\tilde{\theta})\}^2] \leqslant L_C\|\theta - \tilde{\theta}\|^\chi$ *for some* $\chi \geqslant 1$.
- $\max_{j\in[p]}\|\nabla_\theta \mathbb{E}[\widetilde{m}_j(W,\theta)]\| \leqslant L_G$ *for every* $\theta \in \Theta(h_0)$

(iii) For every $\theta \in \Theta(h_0) \setminus \Theta_I$ *we have*

$$\max_{j\in[p]} \mathbb{E}\left[\widetilde{m}_j\left(W,\theta\right)\right] \geqslant \vartheta_n \min\left\{\delta, \inf_{\tilde{\theta}\in\Theta(h_0)\cap\Theta_I} \|\theta - \tilde{\theta}\|\right\}$$

Define $\gamma = o(1)$, in particular $\gamma \ll \alpha$ (e.g., $\gamma = n^{-c}$ for some $c > 0$)

▷ $\bar{w}_n = (1-\gamma)$-quantile of $\sup_{\theta\in\Theta(h_0),j\in[p]}|\widehat{v}^*_{\theta,j}| \lesssim \sqrt{d_\theta \log(pn)}$

▷ $K_n = v\log(n\bar{A}b) + d_\theta \log(nb) + \log p \quad \lesssim d_\theta \log(pn)$

Rates for Size Control

Rates for Size Control for Discard Resampling

$$T_n^{DR*}(h_0) \equiv \inf_{\theta \in \widehat{\Theta}_I(h_0)} \max_{j \in [p]} \widehat{v}^*_{\theta,j} + \varphi_{\theta,j}$$

where $\widehat{\Theta}_I(h_0) \subseteq$ "arg min"$_{\theta \in \Theta(h_0)} \max_{j \in [p]} \sqrt{n}\bar{m}_{\theta,j}/\widehat{\sigma}_{\theta,j}$

$$\varphi_{\theta,j} = \begin{cases} 0, & \text{if } \sqrt{n}\bar{m}_{\theta,j}/\widehat{\sigma}_{\theta,j} \geqslant \max_{\ell \in [p]} \sqrt{n}\bar{m}_{\theta,\ell}/\widehat{\sigma}_{\theta,\ell} - \kappa_n, \\ -\infty, & \text{otherwise} \quad \text{(i.e., inequality will not be used)} \end{cases}$$

Rates for Size Control for Discard Resampling

$$T_n^{DR*}(h_0) \equiv \inf_{\theta \in \widehat{\Theta}_I(h_0)} \max_{j \in [p]} \widehat{v}^*_{\theta,j} + \varphi_{\theta,j}$$

where $\widehat{\Theta}_I(h_0) \subseteq$ "$\arg\min$"$_{\theta \in \Theta(h_0)} \max_{j \in [p]} \sqrt{n}\bar{m}_{\theta,j}/\widehat{\sigma}_{\theta,j}$

$$\varphi_{\theta,j} = \begin{cases} 0, & \text{if } \sqrt{n}\bar{m}_{\theta,j}/\widehat{\sigma}_{\theta,j} \geqslant \max_{\ell \in [p]} \sqrt{n}\bar{m}_{\theta,\ell}/\widehat{\sigma}_{\theta,\ell} - \kappa_n, \\ -\infty, & \text{otherwise} \quad \text{(i.e., inequality will not be used)} \end{cases}$$

Issues to address:

- no functional min max CLT since $p \to \infty$ (and potentially $d_\theta \to \infty$)
- handle random set $\widehat{\Theta}_I(h_0)$
- handle random selection of inequalities
- penalty parameter κ_n

Rates for Size Control for Discard Resampling

$$T_n^{DR*}(h_0) \equiv \inf_{\theta \in \widehat{\Theta}_I(h_0)} \max_{j \in [p]} \widehat{v}^*_{\theta,j} + \varphi_{\theta,j}$$

where $\widehat{\Theta}_I(h_0) \subseteq$ "$\arg\min$"$_{\theta \in \Theta(h_0)} \max_{j \in [p]} \sqrt{n}\bar{m}_{\theta,j}/\widehat{\sigma}_{\theta,j}$

$$\varphi_{\theta,j} = \begin{cases} 0, & \text{if } \sqrt{n}\bar{m}_{\theta,j}/\widehat{\sigma}_{\theta,j} \geqslant \max_{\ell \in [p]} \sqrt{n}\bar{m}_{\theta,\ell}/\widehat{\sigma}_{\theta,\ell} - \kappa_n, \\ -\infty, & \text{otherwise} \quad \text{(i.e., inequality will not be used)} \end{cases}$$

Theorem (Simplified)

Suppose Condition M is satisfied with $d_\theta + L_G/\vartheta_n \leqslant C$ and that H_0 holds. Then

$$P(T_n(h_0) \geqslant t) \leqslant P(T_n^{DS*}(h_0) \geqslant t - C\delta'_{n,\gamma}) + C\{\gamma + n^{-1}\}$$

Rates for Size Control for Discard Resampling

$$T_n^{DR*}(h_0) \equiv \inf_{\theta \in \widehat{\Theta}_I(h_0)} \max_{j \in [p]} \widehat{v}^*_{\theta,j} + \varphi_{\theta,j}$$

where $\widehat{\Theta}_I(h_0) \subseteq$ "arg min"$_{\theta \in \Theta(h_0)} \max_{j \in [p]} \sqrt{n}\bar{m}_{\theta,j}/\widehat{\sigma}_{\theta,j}$

$$\varphi_{\theta,j} = \begin{cases} 0, & \text{if } \sqrt{n}\bar{m}_{\theta,j}/\widehat{\sigma}_{\theta,j} \geqslant \max_{\ell \in [p]} \sqrt{n}\bar{m}_{\theta,\ell}/\widehat{\sigma}_{\theta,\ell} - \kappa_n, \\ -\infty, & \text{otherwise} \quad \text{(i.e., inequality will not be used)} \end{cases}$$

Theorem (Simplified)

Suppose Condition M is satisfied with $d_\theta + L_G/\vartheta_n \leqslant C$ and that H_0 holds. Then

$$P(T_n(h_0) \geqslant t) \leqslant P(T_n^{DS*}(h_0) \geqslant t - C\delta'_{n,\gamma}) + C\{\gamma + n^{-1}\}$$

where we have

$$\delta'_{n,\gamma} \lesssim \frac{\log^{2/3}(np)}{\gamma^{1/3} n^{1/6}}$$

provided that $\kappa_n/\bar{w}_n \to \infty$.

Rates for Size Control for Discard Resampling

$$T_n^{DR*}(h_0) \equiv \inf_{\theta \in \widehat{\Theta}_I(h_0)} \max_{j \in [p]} \widehat{v}^*_{\theta,j} + \varphi_{\theta,j}$$

where $\widehat{\Theta}_I(h_0) \subseteq$ "arg min"$_{\theta \in \Theta(h_0)} \max_{j \in [p]} \sqrt{n}\bar{m}_{\theta,j}/\widehat{\sigma}_{\theta,j}$

$$\varphi_{\theta,j} = \begin{cases} 0, & \text{if } \sqrt{n}\bar{m}_{\theta,j}/\widehat{\sigma}_{\theta,j} \geqslant \max_{\ell \in [p]} \sqrt{n}\bar{m}_{\theta,\ell}/\widehat{\sigma}_{\theta,\ell} - \kappa_n, \\ -\infty, & \text{otherwise} \quad \text{(i.e., inequality will not be used)} \end{cases}$$

Theorem (Simplified)

Suppose Condition M is satisfied with $d_\theta + L_G/\vartheta_n \leqslant C$ *and that* H_0 *holds. Then*

$$P(T_n(h_0) \geqslant t) \leqslant P(T_n^{DS*}(h_0) \geqslant t - C\delta'_{n,\gamma}) + Cn^{-c}$$

for some $0 < c < 1/6$

$$\delta'_{n,\gamma} \lesssim \frac{\log^{2/3}(np)}{n^{1/6-c}}$$

provided that $\kappa_n/\sqrt{\log p} \to \infty$.

Rates for Size Control for Discard Resampling

$$T_n^{DR*}(h_0) \equiv \inf_{\theta \in \widehat{\Theta}_I(h_0)} \max_{j \in [p]} \widehat{v}^*_{\theta,j} + \varphi_{\theta,j}$$

where $\widehat{\Theta}_I(h_0) \subseteq$ "arg min"$_{\theta \in \Theta(h_0)} \max_{j \in [p]} \sqrt{n}\bar{m}_{\theta,j}/\widehat{\sigma}_{\theta,j}$

$$\varphi_{\theta,j} = \begin{cases} 0, & \text{if } \sqrt{n}\bar{m}_{\theta,j}/\widehat{\sigma}_{\theta,j} \geqslant \max_{\ell \in [p]} \sqrt{n}\bar{m}_{\theta,\ell}/\widehat{\sigma}_{\theta,\ell} - \kappa_n, \\ -\infty, & \text{otherwise} \quad \text{(i.e., inequality will not be used)} \end{cases}$$

Theorem

Assume that Condition M is satisfied and that H_0 holds. Then

$$P(T_n(h_0) \geqslant t) \leqslant P(T_n^{DS*}(h_0) \geqslant t - C\delta'_{n,\gamma}) + C\{\gamma + n^{-1}\}$$

where we have

$$\delta'_{n,\gamma} \quad := \frac{CbK_n}{\gamma^{3/q} n^{1/2}} + \frac{C(bK_n^2)^{1/3}}{\gamma^{1/3} n^{1/6}} + CL_C^{1/2} \left(\frac{CK_n^{1/2}}{\gamma^{1/q} n^{1/2} \vartheta_n} \right)^{\chi/2} \frac{K_n^{1/2}}{\gamma^{1/q}} + \frac{CbK_n}{\gamma^{1/q} n^{1/2 - 1/q}}$$

provided that $\kappa_n \geqslant \bar{w}_n\{6 + 2L_G/\vartheta_n\}$

Rates for Size Control for Penalized Resampling

$$T_n^{PR*}(h_0) = \inf_{\theta \in \Theta(h_0)} \max_{j \in [p]} \widehat{v}^*_{\theta,j} + \kappa_n^{-1} \sqrt{n} \bar{m}_{\theta,j} / \widehat{\sigma}_{\theta,j}$$

Rates for Size Control for Penalized Resampling

$$T_n^{PR*}(h_0) = \inf_{\theta \in \Theta(h_0)} \max_{j \in [p]} \widehat{v}^*_{\theta,j} + \kappa_n^{-1} \sqrt{n} \bar{m}_{\theta,j} / \widehat{\sigma}_{\theta,j}$$

Issues to address:

- no functional CLT for min max since $p \to \infty$ (and potentially $d_\theta \to \infty$)
- need to handle random centering $\kappa_n^{-1} \sqrt{n} \bar{m}_{\theta,j} / \widehat{\sigma}_{\theta,j}$
- penalty parameter κ_n

Rates for Size Control for Penalized Resampling

$$T_n^{PR*}(h_0) = \inf_{\theta \in \Theta(h_0)} \max_{j \in [p]} \widehat{v}^*_{\theta,j} + \kappa_n^{-1} \sqrt{n} \bar{m}_{\theta,j} / \widehat{\sigma}_{\theta,j}$$

Theorem (Simplified)

Suppose Condition M is satisfied with $d_\theta + L_G/\vartheta_n \leqslant C$, $\chi = 2$, and that H_0 holds. Then

$$P(T_n(h_0) \geqslant t) \leqslant P(T_n^{PR*}(h_0) \geqslant t - C\delta''_{n,\gamma}) + C\{\gamma + n^{-1}\}$$

where we have

$$\delta''_{n,\gamma} := \frac{\log^{2/3}(np)}{\gamma^{1/3} n^{1/6}} + \kappa_n \frac{\log^{3/2}(np)}{n^{1/2}} + \frac{\bar{w}_n}{\kappa_n}$$

Rates for Size Control for Penalized Resampling

$$T_n^{PR*}(h_0) = \inf_{\theta \in \Theta(h_0)} \max_{j \in [p]} \widehat{v}^*_{\theta,j} + \kappa_n^{-1} \sqrt{n} \bar{m}_{\theta,j} / \widehat{\sigma}_{\theta,j}$$

Theorem

Assume that Condition M is satisfied and that H_0 holds. Then

$$P(T_n(h_0) \geqslant t) \leqslant P(T_n^{PR*}(h_0) \geqslant t - C\delta'_{n,\gamma}) + C\{\gamma + n^{-1}\}$$

where we have

$$\delta'_{n,\gamma} := \frac{L_G \kappa_n K_n}{\gamma^{2/q} n^{1/2} \vartheta_n^2} + \frac{(bK_n^2)^{1/3}}{\gamma^{1/3} n^{1/6}} + \frac{(b)^{1/2} K_n^{3/4}}{\gamma^{1/q} n^{1/4}} + \frac{bK_n}{\gamma^{1/q} n^{1/2-1/q}}$$

$$+ \frac{\bar{w}_n}{\kappa_n} + L_C^{1/2} \left(\frac{\kappa_n K_n^{1/2}}{n^{1/2} \vartheta_n \gamma^{1/q}} \right)^{\chi/2} \frac{K_n^{1/2}}{\gamma^{1/q}}$$

Rates for Size Control for Minimum Resampling

$$T_n^{MR*}(h_0) = \min\{T_n^{DR*}(h_0), T_n^{MR*}(h_0)\}$$

Rates for Size Control for Minimum Resampling

$$T_n^{MR*}(h_0) = \min\{T_n^{DR*}(h_0), T_n^{MR*}(h_0)\}$$

Issues:

- ▷ note that T_n^{MR*} is also a MinMax statistics
  - as it is the minimum of two MinMax statistics
  - need to handle random set in the minimization
  - need to handle random centering
- ▷ clearly need to couple the statistics (use the same $\xi_i \sim N(0,1)$ for both)
- ▷ no functional CLT for MinMax as $p \to \infty$ (and potentially $d_\theta \to \infty$)

Rates for Size Control for Minimum Resampling

$$T_n^{MR*}(h_0) = \min\{T_n^{DR*}(h_0), T_n^{MR*}(h_0)\}$$

Issues:

- ▷ note that T_n^{MR*} is also a MinMax statistics
  - as it is the minimum of two MinMax statistics
  - need to handle random set in the minimization
  - need to handle random centering
- ▷ clearly need to couple the statistics (use the same $\xi_i \sim N(0,1)$ for both)
- ▷ no functional CLT for MinMax as $p \to \infty$ (and potentially $d_\theta \to \infty$)

Theorem

Assume that Condition M is satisfied and that H_0 holds. Then

$$P(T_n(h_0) \geqslant t) \leqslant P(T_n^{MR*}(h_0) \geqslant t - C\delta_{n,\gamma}) + C\{\gamma + n^{-1}\}$$

where we have

$$\delta_{n,\gamma} := \delta'_{n,\gamma} + \delta''_{n,\gamma}$$

Key New Coupling Result

Theorem. Let $X_1, \ldots, X_n$ be independent random matrices in $\mathbb{R}^{N \times p}$ ($Np \geqslant 2$), $Y_1, \ldots, Y_n$ be independent random matrices in $\mathbb{R}^{N \times p}$ with $Y_i \sim N(\mathbb{E}[X_i], \operatorname{Var} X_i)$.

$$\text{Define} \quad T = \min_{k \in [N]} \max_{j \in [p]} \sum_{i=1}^{n} \frac{X_{ikj}}{\sqrt{n}}, \quad \text{and} \quad \widetilde{T} = \min_{k \in [N]} \max_{j \in [p]} \sum_{i=1}^{n} \frac{Y_{ikj}}{\sqrt{n}}$$

Key New Coupling Result

Theorem. Let $X_1, \ldots, X_n$ be independent random matrices in $\mathbb{R}^{N \times p}$ ($Np \geqslant 2$), $Y_1, \ldots, Y_n$ be independent random matrices in $\mathbb{R}^{N \times p}$ with $Y_i \sim N(\mathbb{E}[X_i], \operatorname{Var} X_i)$.

$$\text{Define} \quad T = \min_{k \in [N]} \max_{j \in [p]} \sum_{i=1}^{n} \frac{X_{ikj}}{\sqrt{n}}, \quad \text{and} \quad \widetilde{T} = \min_{k \in [N]} \max_{j \in [p]} \sum_{i=1}^{n} \frac{Y_{ikj}}{\sqrt{n}}$$

Then for every $\delta > 0$ and every Borel subset A of $\mathbb{R}$ we have

$$P(T \in A) \leqslant P(\widetilde{T} \in A^{C\delta}) + \frac{C \log^2(Np)}{\delta^3 n^{1/2}} \{L_n + M_{n,X}(\delta) + M_{n,Y}(\delta)\}$$

where C is a universal positive constant

Key New Coupling Result

Theorem. Let $X_1, \ldots, X_n$ be independent random matrices in $\mathbb{R}^{N \times p}$ ($Np \geqslant 2$), $Y_1, \ldots, Y_n$ be independent random matrices in $\mathbb{R}^{N \times p}$ with $Y_i \sim N(\mathbb{E}[X_i], \operatorname{Var} X_i)$.

$$\text{Define} \quad T = \min_{k \in [N]} \max_{j \in [p]} \sum_{i=1}^{n} \frac{X_{ikj}}{\sqrt{n}}, \quad \text{and} \quad \widetilde{T} = \min_{k \in [N]} \max_{j \in [p]} \sum_{i=1}^{n} \frac{Y_{ikj}}{\sqrt{n}}$$

Then for every $\delta > 0$ and every Borel subset A of $\mathbb{R}$ we have

$$P(T \in A) \leqslant P(\widetilde{T} \in A^{C\delta}) + \frac{C \log^2(Np)}{\delta^3 n^{1/2}} \{L_n + M_{n,X}(\delta) + M_{n,Y}(\delta)\}$$

where C is a universal positive constant, and

$$L_n = \max_{k \in N, j \in [p]} \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[|\widetilde{X}_{ikj}|^3],$$

$$M_{n,W}(\delta) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\left[\max_{k \in N, j \in [p]} |\widetilde{W}_{ikj}|^3 \cdot 1\left\{ \max_{k \in N, j \in [p]} |\widetilde{W}_{ikj}| > \delta\sqrt{n}/\log(Np) \right\} \right],$$

for $\widetilde{W}_i = W_i - \mathbb{E}[W_i]$.

Key New Coupling Result

Theorem. Let $X_1, \ldots, X_n$ be independent random matrices in $\mathbb{R}^{N \times p}$ ($Np \geqslant 2$), $Y_1, \ldots, Y_n$ be independent random matrices in $\mathbb{R}^{N \times p}$ with $Y_i \sim N(\mathbb{E}[X_i], \operatorname{Var} X_i)$.

$$\text{Define} \quad T = \min_{k \in [N]} \max_{j \in [p]} \sum_{i=1}^{n} \frac{X_{ikj}}{\sqrt{n}}, \quad \text{and} \quad \widetilde{T} = \min_{k \in [N]} \max_{j \in [p]} \sum_{i=1}^{n} \frac{Y_{ikj}}{\sqrt{n}}$$

Then for every $\delta > 0$ and every Borel subset A of $\mathbb{R}$ we have

$$P(T \in A) \leqslant P(\widetilde{T} \in A^{C\delta}) + \frac{C \log^2(Np)}{\delta^3 n^{1/2}} \{L_n + M_{n,X}(\delta) + M_{n,Y}(\delta)\}$$

where C is a universal positive constant. In many settings

$$L_n + M_{n,X}(\delta) + M_{n,Y}(\delta) \leqslant C$$

Key New Coupling Result

Theorem. Let $X_1, \ldots, X_n$ be independent random matrices in $\mathbb{R}^{N \times p}$ ($Np \geqslant 2$), $Y_1, \ldots, Y_n$ be independent random matrices in $\mathbb{R}^{N \times p}$ with $Y_i \sim N(\mathbb{E}[X_i], \operatorname{Var} X_i)$.

$$\text{Define} \quad T = \min_{k \in [N]} \max_{j \in [p]} \sum_{i=1}^{n} \frac{X_{ikj}}{\sqrt{n}}, \quad \text{and} \quad \widetilde{T} = \min_{k \in [N]} \max_{j \in [p]} \sum_{i=1}^{n} \frac{Y_{ikj}}{\sqrt{n}}$$

Then for every $\delta > 0$ and every Borel subset A of $\mathbb{R}$ we have

$$P(T \in A) \leqslant P(\widetilde{T} \in A^{C\delta}) + \frac{C \log^2(Np)}{\delta^3 n^{1/2}} \{L_n + M_{n,X}(\delta) + M_{n,Y}(\delta)\}$$

where C is a universal positive constant. In many settings

$$L_n + M_{n,X}(\delta) + M_{n,Y}(\delta) \leqslant C$$

We apply with $A = [t, \infty)$, so $A^{C\delta} = [t - C\delta, \infty)$

Key New Coupling Result

Theorem. Let $X_1, \ldots, X_n$ be independent random matrices in $\mathbb{R}^{N \times p}$ ($Np \geqslant 2$), $Y_1, \ldots, Y_n$ be independent random matrices in $\mathbb{R}^{N \times p}$ with $Y_i \sim N(\mathbb{E}[X_i], \operatorname{Var} X_i)$.

$$\text{Define} \quad T = \min_{k \in [N]} \max_{j \in [p]} \sum_{i=1}^{n} \frac{X_{ikj}}{\sqrt{n}}, \quad \text{and} \quad \widetilde{T} = \min_{k \in [N]} \max_{j \in [p]} \sum_{i=1}^{n} \frac{Y_{ikj}}{\sqrt{n}}$$

Then for every $\delta > 0$ and every Borel subset A of $\mathbb{R}$ we have

$$P(T \in A) \leqslant P(\widetilde{T} \in A^{C\delta}) + \frac{C \log^2(Np)}{\delta^3 n^{1/2}} \{L_n + M_{n,X}(\delta) + M_{n,Y}(\delta)\}$$

where C is a universal positive constant. In many settings

$$L_n + M_{n,X}(\delta) + M_{n,Y}(\delta) \leqslant C$$

We apply with $A = [t, \infty)$, so $A^{C\delta} = [t - C\delta, \infty)$, and for some $\gamma \to 0$

$$\frac{C \log^2(Np)}{\delta^3 n^{1/2}} \leqslant \gamma$$

Key New Coupling Result

Theorem. Let $X_1, \ldots, X_n$ be independent random matrices in $\mathbb{R}^{N\times p}$ ($Np \geqslant 2$), $Y_1, \ldots, Y_n$ be independent random matrices in $\mathbb{R}^{N\times p}$ with $Y_i \sim N(\mathbb{E}[X_i], \operatorname{Var} X_i)$.

$$\text{Define} \quad T = \min_{k\in[N]} \max_{j\in[p]} \sum_{i=1}^{n} \frac{X_{ikj}}{\sqrt{n}}, \quad \text{and} \quad \widetilde{T} = \min_{k\in[N]} \max_{j\in[p]} \sum_{i=1}^{n} \frac{Y_{ikj}}{\sqrt{n}}$$

Then for every $\delta > 0$ and every Borel subset A of $\mathbb{R}$ we have

$$P(T \in A) \leqslant P(\widetilde{T} \in A^{C\delta}) + \frac{C \log^2(Np)}{\delta^3 n^{1/2}} \{L_n + M_{n,X}(\delta) + M_{n,Y}(\delta)\}$$

where C is a universal positive constant. In many settings

$$L_n + M_{n,X}(\delta) + M_{n,Y}(\delta) \leqslant C$$

We apply with $A = [t, \infty)$, so $A^{C\delta} = [t - C\delta, \infty)$, and for some $\gamma \to 0$

$$\frac{C \log^2(Np)}{\delta^3 n^{1/2}} \leqslant \gamma$$

which makes the error

$$\delta = \frac{C \log^{2/3}(Np)}{\gamma^{1/3} n^{1/6}}$$

Key New Coupling Result

Theorem. Let $X_1, \dots, X_n$ be independent random matrices in $\mathbb{R}^{N\times p}$ ($Np \geqslant 2$), $Y_1, \dots, Y_n$ be independent random matrices in $\mathbb{R}^{N\times p}$ with $Y_i \sim N(\mathbb{E}[X_i], \operatorname{Var} X_i)$.

$$\text{Define} \quad T = \min_{k\in[N]} \max_{j\in[p]} \sum_{i=1}^{n} \frac{X_{ikj}}{\sqrt{n}}, \quad \text{and} \quad \widetilde{T} = \min_{k\in[N]} \max_{j\in[p]} \sum_{i=1}^{n} \frac{Y_{ikj}}{\sqrt{n}}$$

Then for every $\delta > 0$ and every Borel subset A of $\mathbb{R}$ we have

$$P(T \in A) \leqslant P(\widetilde{T} \in A^{C\delta}) + \frac{C \log^2(Np)}{\delta^3 n^{1/2}} \{L_n + M_{n,X}(\delta) + M_{n,Y}(\delta)\}$$

where C is a universal positive constant. In many settings

$$L_n + M_{n,X}(\delta) + M_{n,Y}(\delta) \leqslant C$$

We apply with $A = [t, \infty)$, so $A^{C\delta} = [t - C\delta, \infty)$, and for some $\gamma \to 0$

$$\frac{C \log^2(Np)}{\delta^3 n^{1/2}} \leqslant \gamma$$

which makes the error

$$\delta = \frac{C \log^{2/3}(Np)}{\gamma^{1/3} n^{1/6}} \qquad \text{in our case} \quad N \leqslant n^{C d_\theta} \quad (N = 1 \text{ recovers the case of max})$$

Key New Technical Result

Proof is based on Stein's method.
(Extends to processes, Empirical Bootstrap as in Deng and Zhang, 2017)
One key new step is a smooth approximation of the MinMax.

Key New Technical Result

Proof is based on Stein's method.
(Extends to processes, Empirical Bootstrap as in Deng and Zhang, 2017)
One key new step is a smooth approximation of the MinMax.
Recall the LSE function that approximates the max. For $X_k \in \mathbb{R}^p$

$$F_\beta(X_k) = \beta^{-1} \log \left(\sum_{j=1}^{p} \exp(\beta X_{kj}) \right)$$

Key New Technical Result

Proof is based on Stein's method.
(Extends to processes, Empirical Bootstrap as in Deng and Zhang, 2017)
One key new step is a smooth approximation of the MinMax.
Recall the LSE function that approximates the max. For $X_k \in \mathbb{R}^p$

$$F_\beta(X_k) = \beta^{-1} \log \left(\sum_{j=1}^{p} \exp(\beta X_{kj}) \right)$$

In order to approximate $\min \max$

Key New Technical Result

Proof is based on Stein's method.
(Extends to processes, Empirical Bootstrap as in Deng and Zhang, 2017)
One key new step is a smooth approximation of the MinMax.
Recall the LSE function that approximates the max. For $X_k \in \mathbb{R}^p$

$$F_\beta(X_k) = \beta^{-1} \log \left(\sum_{j=1}^{p} \exp(\beta X_{kj}) \right)$$

In order to approximate $\min \max = -\max\{-\max\}$

Key New Technical Result

Proof is based on Stein's method.
(Extends to processes, Empirical Bootstrap as in Deng and Zhang, 2017)
One key new step is a smooth approximation of the MinMax.
Recall the LSE function that approximates the max. For $X_k \in \mathbb{R}^p$

$$F_\beta(X_k) = \beta^{-1} \log \left(\sum_{j=1}^{p} \exp(\beta X_{kj}) \right)$$

In order to approximate $\min \max = -\max\{-\max\}$ and proceed to use

$$G_\beta(X) = -F_\beta(-\{F_\beta(X_k)\}_{k=1}^{N})$$

Key New Technical Result

Proof is based on Stein's method.
(Extends to processes, Empirical Bootstrap as in Deng and Zhang, 2017)
One key new step is a smooth approximation of the MinMax.
Recall the LSE function that approximates the max. For $X_k \in \mathbb{R}^p$

$$F_\beta(X_k) = \beta^{-1} \log \left(\sum_{j=1}^{p} \exp(\beta X_{kj}) \right)$$

In order to approximate $\min \max = -\max\{-\max\}$ and proceed to use

$$G_\beta(X) = -F_\beta(-\{F_\beta(X_k)\}_{k=1}^N)$$

that satisfies:

$$-\beta^{-1} \log N \leqslant G_\beta(X) - \min_{k \in [N]} \max_{j \in [p]} X_{kj} \leqslant \beta^{-1} \log p$$

$$\|\nabla G_\beta(X)\|_1 \leqslant 1$$

$$\|\nabla^2 G_\beta(X)\|_1 \leqslant 4\beta$$

$$\|\nabla^3 G_\beta(X)\|_1 \leqslant 24\beta^2$$

Key New Coupling Result

LEMMA 11. *Consider* $m(X) = g \circ G_\beta(X)$. *Then we have*

(1) for $(k, j) \in [N] \times [p]$

$$m'_{kj}(X) = g'(G_\beta(X))\pi_k(-F_\beta(X))\pi_j^{\mu_k}(X_{k\cdot})$$

(2) for $(k, j) \in ([N] \times [p])^2$, *we have*

$$\begin{aligned}
m''_{(k,j)}(X) \quad &= g''(G_\beta(X))\pi_{k_2}(-F_\beta(X))\pi_{j_2}^{\mu_{k_2}}(X_{k_2\cdot})\pi_{k_1}(-F_\beta(X))\pi_{j_1}^{\mu_{k_1}}(X_{k_1\cdot}) \\
&-g'(G_\beta(X))\beta w_{k_1k_2}(-F_\beta(X))\pi_{j_2}^{\mu_{k_2}}(X_{k_2\cdot})\pi_{j_1}^{\mu_{k_1}}(X_{k_1\cdot}) \\
&+g'(G_\beta(X))\pi_{k_1}(-F_\beta(X))\delta_{k_1k_2}\beta w_{j_1j_2}^{\mu_{k_1}}(X_{k_1\cdot})
\end{aligned}$$

(3) for $(k, j) \in ([N] \times [p])^3$, *we have*

$$\begin{aligned}
m'''_{(k,j)}(X) \quad &= g'''(G_\beta(X)) \textstyle\prod_{\ell=1}^3 \pi_{k_\ell}(-F_\beta(X))\pi_{j_\ell}^{\bar{\mu}_{k_\ell}}(X_{k_\ell\cdot}) \\
&-g''(G_\beta(X))\beta w_{k_2k_3}(-F_\beta(X))\pi_{k_1}(-F_\beta(X)) \textstyle\prod_{\ell=1}^3 \pi_{j_\ell}^{\bar{\mu}_{k_\ell}}(X_{k_\ell\cdot}) \\
&+g''(G_\beta(X))\pi_{k_2}(-F_\beta(X))\delta_{k_2k_3}\beta w_{j_2j_3}^{\bar{\mu}_{k_2}}(X_{k_2\cdot})\pi_{k_1}(-F_\beta(X))\pi_{j_1}^{\bar{\mu}_{k_1}}(X_{k_1\cdot}) \\
&-g''(G_\beta(X))\pi_{k_2}(-F_\beta(X))\beta w_{k_1k_3}(-F_\beta(X)) \textstyle\prod_{\ell=1}^3 \pi_{j_\ell}^{\bar{\mu}_{k_\ell}}(X_{k_\ell\cdot}) \\
&+g''(G_\beta(X))\pi_{k_2}(-F_\beta(X))\pi_{j_2}^{\bar{\mu}_{k_2}}(X_{k_2\cdot})\pi_{k_1}(-F_\beta(X))\delta_{k_1k_3}\beta w_{j_1j_3}^{\bar{\mu}_{k_1}}(X_{k_1\cdot}) \\
&-g''(G_\beta(X))\pi_{k_3}(-F_\beta(X))\beta w_{k_1k_2}(-F_\beta(X)) \textstyle\prod_{\ell=1}^3 \pi_{j_\ell}^{\bar{\mu}_{k_\ell}}(X_{k_\ell\cdot}) \\
&+g'(G_\beta(X))\beta^2 q_{k_1k_2k_3}(-F_\beta(X)) \textstyle\prod_{\ell=1}^3 \pi_{j_\ell}^{\bar{\mu}_{k_\ell}}(X_{k_\ell\cdot}) \\
&-g'(G_\beta(X))\beta w_{k_1k_2}(-F_\beta(X))\delta_{k_2k_3}\beta w_{j_2j_3}^{\bar{\mu}_{k_2}}(X_{k_2\cdot})\pi_{j_1}^{\bar{\mu}_{k_1}}(X_{k_1\cdot}) \\
&-g'(G_\beta(X))\beta w_{k_1k_2}(-F_\beta(X))\pi_{j_2}^{\bar{\mu}_{k_2}}(X_{k_2\cdot})\delta_{k_1k_3}\beta w_{j_1j_3}^{\bar{\mu}_{k_1}}(X_{k_1\cdot}) \\
&+g''(G_\beta(X))\pi_{k_3}(-F_\beta(X))\pi_{j_3}^{\bar{\mu}_{k_3}}(X_{k_3\cdot})\pi_{k_1}(-F_\beta(X))\delta_{k_1k_2}\beta w_{j_1j_2}(X_{k_1\cdot}) \\
&-g'(G_\beta(X))\beta w_{k_1k_3}(-F_\beta(X))\pi_{j_3}^{\bar{\mu}_{k_3}}(X_{k_3\cdot})\delta_{k_1k_2}\beta w_{j_1j_2}^{\bar{\mu}_{k_1}}(X_{k_1\cdot}) \\
&+g'(G_\beta(X))\pi_{k_1}(-F_\beta(X))\delta_{k_1k_2k_3}\beta^2 q_{j_1j_2j_3}^{\bar{\mu}_{k_1}}(X_{k_1\cdot})
\end{aligned}$$

Back to the Size Control Bound

We obtained

$$P(T_n(h_0) \geqslant t) \leqslant P(T_n^{MR*}(h_0) \geqslant t - C\delta_{n,\gamma}) + C\{\gamma + n^{-1}\}$$

where we can take $\gamma \to 0$, and $\delta_{n,\gamma} = o(1)$.

Back to the Size Control Bound

We obtained

$$P(T_n(h_0) \geqslant t) \leqslant P(T_n^{MR*}(h_0) \geqslant t - C\delta_{n,\gamma}) + C\{\gamma + n^{-1}\}$$

where we can take $\gamma \to 0$, and $\delta_{n,\gamma} = o(1)$.

Use a critical value $c_{n,1-\alpha} := c_n^{MR}(h_0, 1-\alpha)$ based on $T_n^{MR*}(h_0)$ for HT

Back to the Size Control Bound

We obtained

$$P(T_n(h_0) \geqslant t) \leqslant P(T_n^{MR*}(h_0) \geqslant t - C\delta_{n,\gamma}) + C\{\gamma + n^{-1}\}$$

where we can take $\gamma \to 0$, and $\delta_{n,\gamma} = o(1)$.

Use a critical value $c_{n,1-\alpha} := c_n^{MR}(h_0, 1-\alpha)$ based on $T_n^{MR*}(h_0)$ for HT. Then

$$P(T_n(h_0) \geqslant c_{n,1-\alpha})$$

Back to the Size Control Bound

We obtained

$$P(T_n(h_0) \geqslant t) \leqslant P(T_n^{MR*}(h_0) \geqslant t - C\delta_{n,\gamma}) + C\{\gamma + n^{-1}\}$$

where we can take $\gamma \to 0$, and $\delta_{n,\gamma} = o(1)$.

Use a critical value $c_{n,1-\alpha} := c_n^{MR}(h_0, 1-\alpha)$ based on $T_n^{MR*}(h_0)$ for HT. Then

$$\begin{aligned} P(T_n(h_0) \geqslant c_{n,1-\alpha}) &\leqslant P(T_n^{MR*}(h_0) \geqslant c_{n,1-\alpha} - C\delta_{n,\gamma}) + o(1) \\ &\leqslant \alpha + P(|T_n^{MR*}(h_0) - c_{n,1-\alpha}| \leqslant C\delta_{n,\gamma}) + o(1) \end{aligned}$$

Back to the Size Control Bound

We obtained

$$P(T_n(h_0) \geqslant t) \leqslant P(T_n^{MR*}(h_0) \geqslant t - C\delta_{n,\gamma}) + C\{\gamma + n^{-1}\}$$

where we can take $\gamma \to 0$, and $\delta_{n,\gamma} = o(1)$.

Use a critical value $c_{n,1-\alpha} := c_n^{MR}(h_0, 1-\alpha)$ based on $T_n^{MR*}(h_0)$ for HT. Then

$$\begin{aligned} P(T_n(h_0) \geqslant c_{n,1-\alpha}) &\leqslant P(T_n^{MR*}(h_0) \geqslant c_{n,1-\alpha} - C\delta_{n,\gamma}) + o(1) \\ &\leqslant \alpha + P(|T_n^{MR*}(h_0) - c_{n,1-\alpha}| \leqslant C\delta_{n,\gamma}) + o(1) \end{aligned}$$

We need to ensure that

$$P(|T_n^{MR*}(h_0) - c_{n,1-\alpha}| \leqslant C\delta_{n,\gamma}) \quad \text{is small}$$

Back to the Size Control Bound

We obtained

$$P(T_n(h_0) \geqslant t) \leqslant P(T_n^{MR*}(h_0) \geqslant t - C\delta_{n,\gamma}) + C\{\gamma + n^{-1}\}$$

where we can take $\gamma \to 0$, and $\delta_{n,\gamma} = o(1)$.

Use a critical value $c_{n,1-\alpha} := c_n^{MR}(h_0, 1-\alpha)$ based on $T_n^{MR*}(h_0)$ for HT. Then

$$\begin{aligned} P(T_n(h_0) \geqslant c_{n,1-\alpha}) &\leqslant P(T_n^{MR*}(h_0) \geqslant c_{n,1-\alpha} - C\delta_{n,\gamma}) + o(1) \\ &\leqslant \alpha + P(|T_n^{MR*}(h_0) - c_{n,1-\alpha}| \leqslant C\delta_{n,\gamma}) + o(1) \end{aligned}$$

We need to ensure that

$$P(|T_n^{MR*}(h_0) - c_{n,1-\alpha}| \leqslant C\delta_{n,\gamma}) \quad \text{is small}$$

i.e., that $T_n^{MR*}(h_0)$ does not concentrate too fast around $c_{n,1-\alpha}$ as $p \to \infty$.

Anti-Concentration

Anti-Concentration: Max

Anti-Concentration: Max

Anti-concentration essentially bounds the probability density function

Anti-Concentration: Max

Anti-concentration essentially bounds the probability density function

Theorem (Chernozhukov, Chetverikov and Kato (2011))

Let $X \in \mathbb{R}^p$ be a vector of Gaussian random variables such that $\mathrm{Var}(X_j) \geqslant 1$. Let $Z = \max_{j \in [p]} X_j$. Then for any $\epsilon > 0$ and $x \in \mathbb{R}$

$$\mathbb{P}\left(|Z - x| \leqslant \epsilon\right) \leqslant C\epsilon\sqrt{\log p}$$

In particular the probability density function of Z satisfies $\max_{t \in \mathbb{R}} f_Z(t) \leqslant C\sqrt{\log p}$

- allows for non-central and arbitrary correlation structure

Anti-Concentration: Max

Anti-concentration essentially bounds the probability density function

Theorem (Chernozhukov, Chetverikov and Kato (2011))

Let $X \in \mathbb{R}^p$ be a vector of Gaussian random variables such that $\mathrm{Var}(X_j) \geqslant 1$. Let $Z = \max_{j \in [p]} X_j$. Then for any $\epsilon > 0$ and $x \in \mathbb{R}$

$$\mathbb{P}\left(|Z - x| \leqslant \epsilon\right) \leqslant C\epsilon\sqrt{\log p}$$

In particular the probability density function of Z satisfies $\max_{t \in \mathbb{R}} f_Z(t) \leqslant C\sqrt{\log p}$

- allows for non-central and arbitrary correlation structure

For coupling between Max statistics ($N = 1$), say T and Z, we have

$$P(T \geqslant c_{1-\alpha}) \leqslant P(Z \geqslant c_{1-\alpha} - \delta_{n,\gamma}) + C\gamma$$

Anti-Concentration: Max

Anti-concentration essentially bounds the probability density function

Theorem (Chernozhukov, Chetverikov and Kato (2011))

Let $X \in \mathbb{R}^p$ be a vector of Gaussian random variables such that $\mathrm{Var}(X_j) \geqslant 1$. Let $Z = \max_{j \in [p]} X_j$. Then for any $\epsilon > 0$ and $x \in \mathbb{R}$

$$\mathbb{P}\left(|Z - x| \leqslant \epsilon\right) \leqslant C\epsilon\sqrt{\log p}$$

In particular the probability density function of Z satisfies $\max_{t \in \mathbb{R}} f_Z(t) \leqslant C\sqrt{\log p}$

- allows for non-central and arbitrary correlation structure

For coupling between Max statistics ($N = 1$), say T and Z, we have

$$P(T \geqslant c_{1-\alpha}) \leqslant P(Z \geqslant c_{1-\alpha} - \delta_{n,\gamma}) + C\gamma \leqslant \alpha + P(|Z - c_{1-\alpha}| \leqslant \delta_{n,\gamma}) + \gamma$$

Anti-Concentration: Max

Anti-concentration essentially bounds the probability density function

Theorem (CCK (2011))

Let $X \in \mathbb{R}^p$ be a vector of Gaussian random variables such that $\mathrm{Var}(X_j) \geqslant 1$. Let $Z = \max_{j \in [p]} X_j$. Then for any $\epsilon > 0$ and $x \in \mathbb{R}$

$$\mathbb{P}\left(|Z - x| \leqslant \epsilon\right) \leqslant C\epsilon\sqrt{\log p}$$

In particular the probability density function of Z satisfies $\max_{t \in \mathbb{R}} f_Z(t) \leqslant C\sqrt{\log p}$

- allows for non-central and arbitrary correlation structure

For coupling between Max statistics ($N = 1$), say T and Z,

we need $\delta_{n,\gamma}\sqrt{\log p} + \gamma \to 0$

Anti-Concentration: Max

Anti-concentration essentially bounds the probability density function

Theorem (CCK (2011))

Let $X \in \mathbb{R}^p$ be a vector of Gaussian random variables such that $\mathrm{Var}(X_j) \geqslant 1$. Let $Z = \max_{j \in [p]} X_j$. Then for any $\epsilon > 0$ and $x \in \mathbb{R}$

$$\mathbb{P}\left(|Z - x| \leqslant \epsilon\right) \leqslant C\epsilon\sqrt{\log p}$$

In particular the probability density function of Z satisfies $\max_{t \in \mathbb{R}} f_Z(t) \leqslant C\sqrt{\log p}$

- allows for non-central and arbitrary correlation structure

For coupling between Max statistics ($N = 1$), say T and Z,

$$\text{we need } \delta_{n,\gamma}\sqrt{\log p} + \gamma \to 0 \;\; \text{implied by} \;\; \delta_{n,\gamma} = \frac{\log^{2/3}(p)}{\gamma^{1/3} n^{1/6}} = o\left(\frac{1}{\sqrt{\log p}}\right)$$

That is, $\log^{7/6}(p) = o(n^{1/6})$.

Anti-Concentration: MinMax

Anti-Concentration: MinMax

Anti-concentration essentially bounds the probability density function

Lemma

For $X_{kj} \sim N(0,1)$, i.i.d., $k \in [N]$, $j \in [p]$, let

$$Z = \min_{k \in [N]} \max_{j \in [p]} X_{kj}.$$

Anti-Concentration: MinMax

Anti-concentration essentially bounds the probability density function

Lemma

For $X_{kj} \sim N(0,1)$, *i.i.d.*, $k \in [N]$, $j \in [p]$, *let*

$$Z = \min_{k \in [N]} \max_{j \in [p]} X_{kj}.$$

If $p/\sqrt{2\pi} > \log(Np) \geqslant 3$, *the probability density function* f_Z *satisfies*

$$\left\{ \sqrt{2} \log^{1/2} \left(\frac{p/\sqrt{2\pi}}{\log N} \right) - 2 \right\} \frac{\log(N)}{e} \leqslant \max_{t \in \mathbb{R}} f_Z(t) \leqslant 4\sqrt{2} \log^{3/2}(Np)$$

Anti-Concentration: MinMax

Anti-concentration essentially bounds the probability density function

Lemma

For $X_{kj} \sim N(0,1)$, i.i.d., $k \in [N]$, $j \in [p]$, let

$$Z = \min_{k \in [N]} \max_{j \in [p]} X_{kj}.$$

If $p/\sqrt{2\pi} > \log(Np) \geqslant 3$, the probability density function f_Z satisfies

$$\left\{ \sqrt{2} \log^{1/2} \left(\frac{p/\sqrt{2\pi}}{\log N} \right) - 2 \right\} \frac{\log(N)}{e} \leqslant \max_{t \in \mathbb{R}} f_Z(t) \leqslant 4\sqrt{2} \log^{3/2}(Np)$$

That is, if $p = N$, for some universal constants $0 < c < C$ we have

$$c \log^{3/2} p \leqslant \max_{t \in \mathbb{R}} f_Z(t) \leqslant C \log^{3/2}(p)$$

Anti-Concentration: MinMax

Anti-concentration essentially bounds the probability density function

Lemma

For $X_{kj} \sim N(0,1)$, i.i.d., $k \in [N]$, $j \in [p]$, let

$$Z = \min_{k \in [N]} \max_{j \in [p]} X_{kj}.$$

If $p/\sqrt{2\pi} > \log(Np) \geqslant 3$, the probability density function f_Z satisfies

$$\left\{ \sqrt{2} \log^{1/2} \left(\frac{p/\sqrt{2\pi}}{\log N} \right) - 2 \right\} \frac{\log(N)}{e} \leqslant \max_{t \in \mathbb{R}} f_Z(t) \leqslant 4\sqrt{2} \log^{3/2}(Np)$$

That is, if $p = N$, for some universal constants $0 < c < C$ we have

$$c \log^{3/2} p \leqslant \max_{t \in \mathbb{R}} f_Z(t) \leqslant C \log^{3/2}(p)$$

- ▷ suggests anti-concentration of MinMax is quite different from the Max
- ▷ currently only partial results for arbitrary correlation structures

Anti-Concentration

However note that our bounds are

$$P(T_n(h_0) \geqslant t) \leqslant P(T_n^{MR*}(h_0) \geqslant t - C\delta_{n,\gamma}) + C\{\gamma + n^{-1}\}$$

It suffices to control the concentration of the bootstrapped statistics

- not of the original statistics

Anti-Concentration

However note that our bounds are

$$P(T_n(h_0) \geqslant t) \leqslant P(T_n^{MR*}(h_0) \geqslant t - C\delta_{n,\gamma}) + C\{\gamma + n^{-1}\}$$

It suffices to control the concentration of the bootstrapped statistics

- not of the original statistics

We can estimate

$$P(|T_n^{MR*}(h_0) - t| \leqslant 2C\delta_{n,\gamma})$$

via bootstrap for $t = c_n(h_0, 1-\alpha)$ and bound the anti-concentration factor

▷ adaptive to the setting (in contrast to analytical bounds)

▷ can be estimated using the same bootstrap that computed $c_n(h_0, 1-\alpha)$

Let $\mathcal{A}_{1-\alpha}^* := \frac{P(|T_n^{MR*}(h_0)-t| \leqslant 2C\delta_{n,\gamma})}{2C\delta_{n,\gamma}}$ denote the anti-concentration rate.

Examples of Simple Conditions for "Penalized Resampling"

Suppose Condition M, m_j and its derivatives are uniformly bounded, $\sigma_{\theta,j} \geqslant c$, $L_G/\vartheta_n + L_C \leqslant C$. Then, letting $\mathcal{A}^*_{1-\alpha}$ denote the anti-concentration rate, provided

$$\frac{K_n^{2/3}}{n^{1/6}} + \kappa_n \frac{d_\theta^{1/2} K_n}{n^{1/2}} + \frac{\bar{w}}{\kappa_n} = o\left(\frac{1}{\mathcal{A}^*_{1-\alpha}}\right), \tag{4}$$

where $K_n = \log p + d_\theta \log n$, we have $P(T_n(h_0) \geqslant c_n(h_0, 1-\alpha)) \leqslant \alpha + o(1)$

Examples of Simple Conditions for "Penalized Resampling"

Suppose Condition M, m_j and its derivatives are uniformly bounded, $\sigma_{\theta,j} \geqslant c$, $L_G/\vartheta_n + L_C \leqslant C$. Then, letting $\mathcal{A}^*_{1-\alpha}$ denote the anti-concentration rate, provided

$$\frac{K_n^{2/3}}{n^{1/6}} + \kappa_n \frac{d_\theta^{1/2} K_n}{n^{1/2}} + \frac{\bar{w}}{\kappa_n} = o\left(\frac{1}{\mathcal{A}^*_{1-\alpha}}\right), \tag{4}$$

where $K_n = \log p + d_\theta \log n$, we have $P(T_n(h_0) \geqslant c_n(h_0, 1-\alpha)) \leqslant \alpha + o(1)$

Remark: we can simulate $\bar{w}$ and bound $\mathcal{A}^*_{1-\alpha}$

- ▷ yields a data-driven choice of κ_n

Examples of Simple Conditions for “Penalized Resampling”

Suppose Condition M, m_j and its derivatives are uniformly bounded, $\sigma_{\theta,j} \geqslant c$, $L_G/\vartheta_n + L_C \leqslant C$. Then, letting $\mathcal{A}^*_{1-\alpha}$ denote the anti-concentration rate, provided

$$\frac{K_n^{2/3}}{n^{1/6}} + \kappa_n \frac{d_\theta^{1/2} K_n}{n^{1/2}} + \frac{\bar{w}}{\kappa_n} = o\left(\frac{1}{\mathcal{A}^*_{1-\alpha}}\right), \tag{4}$$

where $K_n = \log p + d_\theta \log n$, we have $P(T_n(h_0) \geqslant c_n(h_0, 1-\alpha)) \leqslant \alpha + o(1)$

Remark: we can simulate $\bar{w}$ and bound $\mathcal{A}^*_{1-\alpha}$

- ▷ yields a data-driven choice of κ_n

For the traditional setting, e.g., fixed p and d_θ

- ▷ $\mathcal{A}^*_{1-\alpha} \leqslant C$
- ▷ $K_n \leqslant C$
- ▷ $\kappa_n \to \infty$ and $\kappa_n/n^{1/2} \to 0$

Examples of Simple Conditions for "Penalized Resampling"

For non-Donsker cases:

Example (Many inequalities and fixed d_θ)

Let $p = n^C$ for some fixed $C > 1$, $d_\theta \leqslant C$, and the anti-concentration $\mathcal{A}^*_{1-\alpha} \leqslant C \log^{3/2} n$. It suffices $\kappa_n \in [\log^{5/2} n, \quad n^{\frac{1}{2}} \log^{-3} n]$.

Examples of Simple Conditions for "Penalized Resampling"

For non-Donsker cases:

Example (Many inequalities and fixed d_θ)

Let $p = n^C$ for some fixed $C > 1$, $d_\theta \leqslant C$, and the anti-concentration $\mathcal{A}^*_{1-\alpha} \leqslant C \log^{3/2} n$. It suffices $\kappa_n \in [\log^{5/2} n, \quad n^{\frac{1}{2}} \log^{-3} n]$.

Example (Polynomially many inequalities and large d_θ)

Let $p = n^C$ for some fixed $C > 1$, $d_\theta = n^a$ for some $a < 1/4$, and the anti-concentration $\mathcal{A}^*_{1-\alpha} \leqslant C \log^{3/2} n$. It suffices $\kappa_n \in [n^{a/2} \log^{5/2} n, \quad n^{\frac{1}{2} - \frac{3}{2}a} \log^{-3} n]$.

Examples of Simple Conditions for "Penalized Resampling"

For non-Donsker cases:

Example (Many inequalities and fixed d_θ)

Let $p = n^C$ for some fixed $C > 1$, $d_\theta \leqslant C$, and the anti-concentration $\mathcal{A}^*_{1-\alpha} \leqslant C \log^{3/2} n$. It suffices $\kappa_n \in [\log^{5/2} n, \quad n^{\frac{1}{2}} \log^{-3} n]$.

Example (Polynomially many inequalities and large d_θ)

Let $p = n^C$ for some fixed $C > 1$, $d_\theta = n^a$ for some $a < 1/4$, and the anti-concentration $\mathcal{A}^*_{1-\alpha} \leqslant C \log^{3/2} n$. It suffices $\kappa_n \in [n^{a/2} \log^{5/2} n, \quad n^{\frac{1}{2} - \frac{3}{2}a} \log^{-3} n]$.

Example (Exponentially many inequalities)

Suppose that $d_\theta \leqslant C \log n$, $p \geqslant n^{\log n}$ and the anti-concentration $\mathcal{A}^*_{1-\alpha} \leqslant C \log^{3/2} p$. It suffices $\kappa_n \in [\log^2 p \, \log n, \quad n^{1/2} \log^{-5/2} p \log^{-1} n]$, provided that $n^{-1/6} \log^{13/6} p \, \log n = o(1)$.

Conclusion

- ▷ subvector inference in PI models with many moment restrictions
  - allow for non-Donsker classes
  - finite sample analysis
  - need more than $\kappa_n \to \infty$ and $\kappa_n/\sqrt{n} \to 0$ when $p \to \infty$
  - valid data-driven choice of penalty parameters (via additional bootstrap)
- ▷ new CLTs for $\min_{k\in[N]} \max_{j\in[p]} W_{kj}$
  - results parallel results for $\max_{j\in[p]} W_j$
  - approximation based on composition of smooth maximum (LSE)
- ▷ new anti-concentration pattern
  - does not parallel results for $\max_{j\in[p]} W_j$ (counter example)
  - estimate anti-concentration via bootstrap
- ▷ Future (ongoing) work
  - sharper constants
  - hybrid methods
  - power comparisons
  - analytical bounds for anti-concentration
  - orthogonal moment conditions