Tutorial Part 1b: Inference with many moment inequalities

Kengo Kato (U. of Tokyo)

May 15, 2014@NUS

This part is based upon the paper:

- Chernozhukov, V., Chetverikov, D. and K. (2013). Testing many moment inequalities. arXiv:1312.7614.

- In the previous lecture, we've seen that the following multivariate one-sided testing problem is closely connected to the inference problem on parameters defined by moment inequalities.
- Let $X_1, ..., X_n$ be an i.i.d. sequence of random vectors in $\mathbb{R}^p$ with $\mu = \mathrm{E}[X_1]$, and consider testing

$$H_0 : \mu_j \leq 0, 1 \leq \forall j \leq p,$$

against

$$H_1 : \mu_j > 0, 1 \leq \exists j \leq p.$$

Here

$$\mu = (\mu_1, \ldots, \mu_p)^T, X_i = (X_{i1}, \ldots, X_{ip})^T.$$

- Focus on the case where p *is large, and possibly much larger than* n.

Test statistic

- Define
$$\widehat{\mu}_j = \frac{1}{n}\sum_{i=1}^{n} X_{ij},\ \widehat{\sigma}_j^2 = \frac{1}{n}\sum_{i=1}^{n}(X_{ij} - \widehat{\mu}_j)^2.$$
- We consider here the max-type test statistic
$$T = \max_{1\le j\le p} \frac{\sqrt{n}\widehat{\mu}_j}{\widehat{\sigma}_j}.$$
Reject H_0 if $T > c$.

Critical values

- We want to choose c such that,

$$\sup_{\mu_j \le 0, \forall j} \mathrm{P}(T > c) \le \alpha + o(1),$$

which we also want to hold uniformly over a wide class of distributions.

- Interpret "$T > c$" as

$$\sqrt{n}\widehat{\mu}_j > \widehat{\sigma}_j c, \exists j,$$

whenever $\widehat{\sigma}_j = 0, \exists j$.

Basic idea for calculating critical values

- Under H_0, $\mu_j \le 0, \forall j$, so that

$$T \le \max_j \sqrt{n}(\widehat{\mu}_j - \mu_j)/\hat{\sigma}_j,$$

where the equality takes place when $\mu_j = 0, \forall j$. Enough to approximate or bound quantiles of the r.v. on the right side.

Methods

We consider two methods to calculate critical values:

- Self-Normalized (SN) method; fast, works under very weak conditions but conservative.
- Multiplier Bootstrap (MB) method; slower (requires simulations), requires stronger conditions but nonconservative.

For each method, we consider

- One-step method – no selection procedure.
- Two-step method – selection procedure is used to get rid of inequalities that are clearly non-binding.

Literature

- Inference with unconditional moment inequalities: Chernozhukov, Hong, and Tamer (2007, ECMT), Romano and Shaikh (2008, JSPI), Menzel (2009, thesis), Andrews and Guggenberger (2009, ET), Andrews and Soares (2010, ECMT), Canay (2010, JoE), Bugni (2011), Andrews and Jia Barwick (2012, ECMT), Romano, Shaikh, and Wolf (2014, ECMT).
- Except for Menzel (2009), the number of moment inequalities p is assumed to be *fixed* in the analysis of these papers.

- Andrews and Jia Barwick (2012) proposed computationally intensive methods using a novel moment selection (*recommended moment selection*) for inference on parameters defined by moment inequalities, which leads to CRs with good coverage properties.
- Romano, Shaikh, and Wolf (2014) proposed computationally less intensive alternatives; but still assume p is fixed in their analysis.
- Menzel (2009) studied the case where the number of inequalities $p = p_n$ is growing *slowly* with n; p should be $o(n^{2/7})$.
- This paper covers the case where p is possibly much larger than n.

- Inference with *conditional* moment inequalities: Andrews and Shi (2013, ECMT), Chernozhukov, Lee, and Rosen (2013, ECMT), Armstrong (2011), Chetverikov (2011), Armstrong and Chan (2012).
- A small number of conditional inequalities gives rise to a large number of unconditional inequalities, but these have certain continuity and tightness structure, which the literature on conditional moment inequalities heavily exploits/relies upon.
- Our approach does not exploit/rely upon such structure and can handle both many unstructured moment inequalities as well as many structured moment inequalities arising from conversion of a small number of conditional inequalities.

Notation

- Assume in what follows

$$\mathbf{E}[X_{1j}^2] < \infty, \ \sigma_j^2 = \mathrm{Var}(X_{1j}) > 0, \ \forall j.$$

- Nominal size: $\alpha \in (0, 1/2)$.
- Define

$$Z_{ij} = \frac{X_{ij} - \mu_j}{\sigma_j}, \ Z_i = (Z_{i1}, \ldots, Z_{ip})^T,$$
$$M_{n,k} = \max_{1 \le j \le p} (\mathbf{E}[|Z_{ij}|^k)^{1/k}, \ B_n = (\mathbf{E}[\max_{1 \le j \le p} Z_{ij}^4])^{1/4}.$$

Note: $B_n \ge M_{n,4} \ge M_{n,3} \ge 1$.

SN methods

- Based on union (Bonferroni) bound combined with a moderate deviation inequality for self-normalized sums.
- Observe that under H_0,

$$\mathrm{P}(T > c) \le \sum_{j=1}^{p} \mathrm{P}(\sqrt{n}(\widehat{\mu}_j - \mu_j)/\widehat{\sigma}_j > c).$$

- Given that p is large, the union bound might look too conservative but in fact it is not.

- Define
$$U_j = \sqrt{n}\mathbb{E}_n[Z_{ij}]/\sqrt{\mathbb{E}_n[Z_{ij}^2]}.$$
Then
$$\sqrt{n}(\widehat{\mu}_j - \mu_j)/\widehat{\sigma}_j = U_j/\sqrt{1 - U_j^2/n},$$
so that
$$\mathrm{P}(T > c) \leq \sum_{j=1}^{p} \mathrm{P}(U_j > c/\sqrt{1 + c^2/n}),$$
since $u \mapsto u/\sqrt{1 - u^2/n}$ is increasing in u.
- The last quantity can be controlled well because U_j is a *self-normalized sum*, and behaves like $N(0,1)$ even if X_{ij} has only $2 + \delta$ finite moments

- Take c in such a way that

$$\mathbf{P}(N(0,1) > c/\sqrt{1+c^2/n}) = \alpha/p,$$

which leads to

$$c = c^{SN}(\alpha) = \frac{\Phi^{-1}(1-\alpha/p)}{\sqrt{1-\Phi^{-1}(1-\alpha/p)^2/n}}.$$

- When $p = p_n \to \infty$ but $\log p = o(n)$,

$$c^{SN}(\alpha) \sim \sqrt{\log(p/\alpha)},$$

so that the the SN critical value depends on p only through $\log p$.

Validity of one-step SN method

Theorem

Suppose that $\Phi^{-1}(1-\alpha/p) \le n^{1/6}/M_{n,3}$. *Then under* H_0,

$$\mathrm{P}(T > c^{SN}(\alpha)) \le \alpha\left[1 + Kn^{-1/2}M_{n,3}^3\{1+\Phi^{-1}(1-\alpha/p)\}^3\right],$$

where K *is universal. In particular, if there exist constants* $c_1 \in (0,1/2)$ *and* $C_1 > 0$ *such that*

$$M_{n,3}^3 \log^{3/2}(p/\alpha) \le C_1 n^{1/2-c_1},$$

then under H_0,

$$\mathrm{P}(T > c^{SN}(\alpha)) \le \alpha + Cn^{-c_1},$$

where $C = C(c_1, C_1) > 0$.

Moderate deviation inequality for self-normalized sums

Lemma

Let $\xi_1, \dots, \xi_n$ *be independent centered random variables with* $\mathrm{E}[\xi_i^2] = 1$ *and* $\mathrm{E}[|\xi_i|^{2+\nu}] < \infty$ *for all* $1 \le i \le n$ *where* $0 < \nu \le 1$. *Let* $S_n = \sum_{i=1}^n \xi_i, V_n^2 = \sum_{i=1}^n \xi_i^2$, *and* $D_{n,\nu} = (n^{-1} \sum_{i=1}^n \mathrm{E}[|\xi_i|^{2+\nu}])^{1/(2+\nu)}$. *Then uniformly in* $0 \le x \le n^{\frac{\nu}{2(2+\nu)}} / D_{n,\nu}$,

$$\left| \frac{\mathrm{P}(S_n/V_n \ge x)}{\bar{\Phi}(x)} - 1 \right| \le K n^{-\nu/2} D_{n,\nu}^{2+\nu} (1+x)^{2+\nu},$$

where K *is a universal constant.*

Proof.

See Theorem 7.4 in Lai, T.L., de la Pen a, V., and Shao, Q.-M. (2009), *Self-Normalized Processes: Limit Theory and Statistical Applications*, Springer; or the original reference: Jing, B.-Y., Shao, Q.-M., and Wang, Q. (2003, AoP). □

Two-step SN method

- Let $0 < \beta_n < \alpha/2$ be some constant.
- Let $c^{SN}(\beta_n)$ be the SN critical value with size β_n, and define

$$\hat{J}_{SN} = \{j \in \{1,\dots,p\} : \sqrt{n}\hat{\mu}_j > -2\hat{\sigma}_j c^{SN}(\beta_n)\}.$$

- Then the two-step SN critical value is defined by

$$c^{SN,MS}(\alpha) = \begin{cases} \frac{\Phi^{-1}(1-(\alpha-2\beta_n)/\hat{k})}{\sqrt{1-\Phi^{-1}(1-(\alpha-2\beta_n)/\hat{k})^2/n}}, & \text{if } \hat{k} \geq 1, \\ 0, & \text{if } \hat{k} = 0. \end{cases}$$

Here $\hat{k} = |\hat{J}_{SN}|$.

Validity of two-step SN method

Theorem

Suppose that $\sup_n \beta_n \leq \alpha/3$ and there exist constants $0 < c_1 < 1/2$ and $C_1 > 0$ such that

$$M_{n,3}^3 \log^{3/2}(p/\beta_n) \leq C_1 n^{1/2-c_1}, \ B_n^2 \log^2(p/\beta_n) \leq C_1 n^{1/2-c_1}.$$

Then $\exists c, C$ depending only on c_1, C_1 such that under H_0,

$$\mathrm{P}(T > c^{SN,MS}(\alpha)) \leq \alpha + Cn^{-c}.$$

- Intuition:

$$\max_j \sqrt{n}(\widehat{\mu}_j - \mu_j)/\widehat{\sigma}_j \approx \max_j \sqrt{n}(\widehat{\mu}_j - \mu_j)/\sigma_j$$
$$\approx \max_j \sqrt{n}\mathbb{E}_n[Z_{ij}],$$

so that when p *is fixed*, by CLT and continuous mapping theorem,

$$\max_j \sqrt{n}\mathbb{E}_n[Z_{ij}] \xrightarrow{d} \max_j Y_j,$$

where

$$Y = (Y_1, \dots, Y_p)^T \sim N(0, \mathrm{E}[Z_1 Z_1^T]).$$

- When $p = p_n \to \infty$? The previous argument does not apply as it stands: what does "$\stackrel{d}{\to}$" mean when $p = p_n \to \infty$?
- We apply a high-dimensional CLT from Chernozhukov, Chetverikov, K. (AoS, 2013).
- It proved that under mild regularity conditions, the distribution of $\max_j \sqrt{n}\mathbb{E}_n[Z_{ij}]$ can be approximated by the that of $\max_j Y_j$ in the sense

$$\sup_t |\mathrm{P}(\max_j \sqrt{n}\mathbb{E}_n[Z_{ij}] \le t) - \mathrm{P}(\max_j Y_j \le t)| \to \text{small},$$

even when $p \gg n$. The main idea is to directly compare the distributions of $\max_j \sqrt{n}\mathbb{E}_n[Z_{ij}]$ and $\max_j Y_j$ instead of trying first to compare the whole vectors $\sqrt{n}\mathbb{E}_n[Z_i]$ and Y.

- The distribution of $\max_j Y_j$ is still infeasible since the covariance structure of Z is unknown.
- To deal with this problem, CCK13 suggested to use the Multiplier Bootstrap method:
  1. Generate independent $N(0,1)$ random variables $\epsilon_1, \ldots, \epsilon_n$ independent of the data $X_1^n = \{X_1, \ldots, X_n\}$.
  2. Construct the Multiplier Bootstrap test statistic:

$$W = \max_{1 \leq j \leq p} \frac{\sqrt{n}\mathbb{E}_n[\epsilon_i(X_{ij} - \widehat{\mu}_j)]}{\widehat{\sigma}_j}.$$

  3. Calculate $c = c^{MB}(\alpha)$ as conditional $(1-\alpha)$ of $W \mid X_1^n$.

Validity of one-step MB method

Theorem

Suppose that there exist constants $0 < c_1 < 1/2$ and $C_1 > 0$ such that

$$(M_{n,3}^3 \vee M_{n,4}^2 \vee B_n)^2 \log^{7/2}(pn) \le C_1 n^{1/2-c_1}.$$

Then $\exists c, C$ depending only on c_1, C_1 such that under H_0,

$$\mathrm{P}(T > c^{MB}(\alpha)) \le \alpha + Cn^{-c}.$$

If $\mu_j = 0, \forall j$, then

$$\left|\mathrm{P}(T > c^{MB}(\alpha)) - \alpha\right| \le Cn^{-c}.$$

Two-step MB method

- Let $0 < \beta_n < \alpha/2$ be some constant.
- Let $c^{MB}(\beta_n)$ be the (one-step) MB critical value with size β_n.
- Define the set $\hat{J}_{MB}$ by

$$\hat{J}_{MB} = \{j \in \{1, \ldots, p\} : \sqrt{n}\hat{\mu}_j > -2\hat{\sigma}_j c^{MB}(\beta_n)\}.$$

Then the two-step MB critical value is defined by:

1. Generate independent standard normal random variables $\epsilon_1, \ldots, \epsilon_n$ independent of the data X_1^n.
2. Construct the bootstrap test statistic

$$W_{\hat{J}_{MB}} = \begin{cases} \max_{j \in \hat{J}_{MB}} \frac{\sqrt{n}\mathbb{E}_n[\epsilon_i(X_{ij} - \hat{\mu}_j)]}{\hat{\sigma}_j}, & \text{if } \hat{J}_{MB} \neq \emptyset \\ 0 & \text{if } \hat{J}_{MB} = \emptyset. \end{cases}$$

3. Calculate $c^{MB,MS}(\alpha)$ as

$$c^{MB,MS}(\alpha) = \text{conditional } (1 - \alpha + 2\beta_n)\text{-quantile of } W_{\hat{J}_{MB}} \mid X_1^n.$$

Theorem

Suppose that the assumption of the previous theorem is satisfied. Moreover, suppose that $\sup_n \beta_n < \alpha/2$ *and* $\log(1/\beta_n) \leq C_1 \log n$*. Then all the conclusions of the previous theorem hold with* $c^{MB}(\alpha)$ *replaced by* $c^{MB,MS}(\alpha)$*.*

Hybrid method

- Using the SN method for moment selection and applying the MB method to the selected moments.
- Recall

$$\widehat{J}_{SN} = \{j \in \{1, \ldots, p\} : \sqrt{n}\widehat{\mu}_j > -2\widehat{\sigma}_j c^{SN}(\beta_n)\}.$$

- Hybrid method:
  1. Generate independent $N(0,1)$ random variables $\epsilon_1, \dots, \epsilon_n$ independent of the data X_1^n
  2. Construct the bootstrap test statistic

$$W_{\hat{J}_{SN}} = \begin{cases} \max_{j \in \hat{J}_{SN}} \frac{\sqrt{n}\mathbb{E}_n[\epsilon_i(X_{ij} - \hat{\mu}_j)]}{\hat{\sigma}_j}, & \text{if } \hat{J}_{SN} \neq \emptyset \\ 0 & \text{if } \hat{J}_{SN} = \emptyset. \end{cases}$$

  3. Calculate $c = c^{HB}(\alpha)$ as the conditional $(1 - \alpha + 2\beta_n)$-quantile of $W_{\hat{J}_{SN}} \mid X_1^n$
- Under the same assumption as in the theorem for the two-step MB method, all the conclusions of the theorem holds for the hybrid method.

Optimality from minimax point of view

- Discuss optimality from minimax point of view in the sense of Ingster (1993, Math. Meth. Stat.) and Ingster and Suslina (2003, Lecture Notes in Stat.).
- Consider testing

$$H_0 : \max_j \mu_j \le 0, \text{ v.s. } H_{1,r} : \max_j (\mu_j / \sigma_j) \ge r,$$

  where $r > 0$.
- The constant r is thought of as "distance" between the null and alternative hypotheses; the smaller r is, the harder to detect $H_{1,r}$ is.

Minimax rate of testing

- Question: when $p = p_n \to \infty$, determine the *minimax rate of testing*, $r = r_n^* \to 0$, such that

$$\inf_{\phi_n} \left\{ \sup_{\mu_j \leq 0, \forall j} \mathrm{E}[\phi_n] + \sup_{\max_j (\mu_j/\sigma_j) \geq r_n} (1 - \mathrm{E}[\phi_n]) \right\} = 1 - o(1),$$

whenever $r_n = o(r_n^*)$, where $\inf_{\phi_n}$ is taken over all tests, i.e., all m'ble functions

$$\phi_n : (X_1, \ldots, X_n) \mapsto \phi(X_1, \ldots, X_n) \in [0, 1];$$

and moreover if for every $\alpha \in (0, 1)$ there exists a test ϕ_n^* with size $\alpha + o(1)$ such that

$$\sup_{\max_j (\mu_j/\sigma_j) \geq r_n} (1 - \mathrm{E}[\phi_n^*]) = o(1),$$

whenever $r_n / r_n^* \to \infty$.

Normal case

Lemma

Let $V_1, \ldots, V_n \sim N(\mu, \Sigma)$ *i.i.d. with* $\Sigma = \mathrm{diag}\{\sigma_1^2, \ldots, \sigma_p^2\}$ *and* $\sigma_j^2 > 0, \forall j$*. Consider testing* $H_0 : \max_j \mu_j \le 0$ *v.s.* $H_1 : \max_j(\mu_j/\sigma_j) \ge r$ *with* $r > 0$*. Then for any test* ϕ_n*,*

$$\inf_{\max_j(\mu_j/\sigma_j)\ge r} \mathrm{E}_\mu[\phi_n(V_1, \ldots, V_n)] \le \alpha + \mathrm{E}[|p^{-1}\textstyle\sum_{j=1}^p e^{\sqrt{n} r \xi_j - nr^2/2} - 1|],$$

where $\alpha = \sup_{\mu_j \le 0, \forall j} \mathrm{E}_\mu[\phi_n(V_1, \ldots, V_n)]$ *and* $\xi_1, \ldots, \xi_p \sim N(0,1)$ *i.i.d. If* $p = p_n \to \infty$ *as* $n \to \infty$ *and* $r = \underline{r}_n = (1-\epsilon_n)\sqrt{2(\log p_n)/n}$ *where* $\epsilon_n > 0$ *is such that* $\epsilon_n \to 0$ *and* $\epsilon_n \sqrt{\log p_n} \to \infty$*, the second term on the right side is* $o(1)$*.*

Back to general case

Lemma

Take $c(\alpha) \in \{c^{SN}(\alpha), c^{SN,MS}(\alpha), c^{MB}(\alpha), c^{MB,MS}, c^{HB}\}$. *Suppose that*

$$B_n^2 \log^{3/2} p = o(n^{1/2}), \ \sup_n \beta_n \leq \alpha/3.$$

Then

$$\inf_{\max_j (\mu_j/\sigma_j) \geq r} \mathrm{P}(T > c(\alpha)) \geq 1 - o(1),$$

if $p = p_n \to \infty$ *and* $r = \overline{r}_n = (1 + \epsilon_n)\sqrt{2(\log p_n)/n}$ *where* $\epsilon_n > 0$ *is such that* $\epsilon_n \to 0$ *and* $\epsilon_n \sqrt{\log p_n} \to \infty$.

- The minimax rate of testing is $\sqrt{(\log p_n)/n}$, and out tests are all rate-optimal (we've proved a bit more stronger assertion).
- Our tests can detect with probability approaching one any deviation from the null of which the size is

$$\bar{r}_n = (1 + \epsilon_n)\sqrt{2(\log p_n)/n},$$

and for *any* test that with correct size (at least asymptotically), there exists an alternative that is separated from the null by

$$\underline{r}_n = (1 - \epsilon_n)\sqrt{2(\log p_n)/n},$$

against which the test is trivial.