Gaussian Approximations and Multiplier Bootstrap for Maxima of Sums of High-Dimensional Random Vectors

Victor Chernozhukov (MIT),
Denis Chetverikov (MIT),
and Kengo Kato (University of Tokyo)

June 2013

Introduction

Let $x_1, ..., x_n$ be an i.i.d. sequence of random vectors in $\mathbb{R}^p$

Let

$$\mu = \mathrm{E}[x_i]$$

Denote

$$x_i = (x_{i1}, ...x_{ip})'$$

and

$$\mu = (\mu_1, ..., \mu_p)'$$

This paper is about approximating the distribution of

$$T_0 = \max_{1 \leq j \leq p} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (x_{ij} - \mu_j)$$

Introduction

Let $y_1, ..., y_n$ be independent Gaussian random vectors such that

$$\mathrm{E}[y_i] = \mathrm{E}[x_i] = \mu \text{ and } \mathrm{E}[y_i y_i'] = \mathrm{E}[x_i x_i']$$

Define

$$Z_0 = \max_{1 \leq j \leq p} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (y_{ij} - \mu_j)$$

Classical CLT:

- when p is fixed, the distribution of T_0 can be approximated by the distribution of Z_0

Introduction

Recall that

$$T_0 = \max_{1 \le j \le p} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (x_{ij} - \mu_j) \text{ and } Z_0 = \max_{1 \le j \le p} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (y_{ij} - \mu_j)$$

Define

$$\rho = \sup_{t \in \mathbb{R}} |\mathrm{P}(T_0 \le t) - \mathrm{P}(Z_0 \le t)|$$

Question:

How large p can be in relation with n so that $\rho \to 0$

Probably the *best* previous result (due to Bentkus (2003)):

$$p^{7/2} = o(n)$$

So, p has to be much smaller than n.

Introduction

We show that

$$(\log p)^7 = o(n)$$

suffices (under some conditions).

So, p can be much larger than n

One can approximate the distribution of T_0 by Z_0 but this method is *infeasible* because covariance structure is unknown

- We also show how to approximate the distribution of T_0 using the Gaussian multiplier bootstrap

Applications

There is a huge number of applications:

- selecting a regularization parameter for Dantzig and Lasso estimators
- multiple hypothesis testing with the number of hypothesis much larger than the sample size
- adaptive specification testing
- testing many moment inequalities
- selecting truncation parameter for adaptive estimation and testing based on Lepski's method

Graphical Illustration

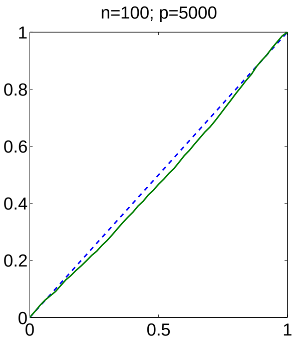

Figure : P-P plots comparing distributions of T_0 and Z_0. Here x_{ij} are generated as $x_{ij} = z_{ij}\varepsilon_i$ with $\varepsilon_i \sim t(4)$, and z_{ij} are non-stochastic (simulated once using $U[0,1]$ distribution independently across i and j). The dashed line is 45°.

Literature

Classical CLTs:

- Senatov (1980), Asriev and Rotar (1985), Portnoy (1986), Gotze (1991), Bentkus (2003), among many others

Modern invariance principles:

- Chatterjee (2005), Korada and Montanari (2011)

We use the following techniques

- Slepian's interpolation
- Stein's leave-one-out approach

Main GAR: Conditions

Let c_1 and C_1 be some strictly positive constants

Let $(B_n)_{n=1}^{\infty}$ be a sequence of positive constants

We will assume that one of the following conditions holds:

$$\text{(E.1)} \quad \max_{1 \le j \le p} \mathrm{E}[\exp(|x_{ij} - \mu_j|/B_n)] \le 2$$

$$\text{(E.2)} \quad \mathrm{E}[(\max_{1 \le j \le p} |x_{ij} - \mu_j|/B_n)^4] \le 2$$

In addition, we will assume that the following moment conditions hold:

$$\text{(M.1)} \quad c_1 \le \mathrm{E}[(x_{ij} - \mu_j)^2] \le C_1$$

$$\text{(M.2)} \quad \max_{k=1,2} \mathrm{E}[|x_{ij} - \mu_j|^{2+k}/B_n^k] \le 2$$

Main GAR: Theorem

Theorem (Gaussian Approximation Result)

Suppose that one of the following conditions is satistfied:

(i) *(E.1) and* $B_n^2(\log(pn))^7/n \leq C_1 n^{-c_1}$

(ii) *(E.2) and* $B_n^4(\log(pn))^7/n \leq C_1 n^{-c_1}$

In addition, suppose that the conditions (M.1) and (M.2) are satisfied. Then

$$\rho = \sup_{t \in \mathbb{R}} |\mathrm{P}(T_0 \leq t) - \mathrm{P}(Z_0 \leq t)| \leq Cn^{-c}$$

for some strictly positive constants c *and* C *that depend only on* c_1 *and* C_1*.*

Remark: the theorem in the paper is more general

GAR: In old English, a gar means a spear

Multiplier Bootstrap I

Suppose that $x_1, ..., x_n$ is our data and

$$T_0 = \max_{1 \leq j \leq p} \frac{1}{\sqrt{n}} \sum_{1 \leq i \leq n} (x_{ij} - \mu_j)$$

is our test statistic. To derive a critical value, we would like to use quantiles of

$$Z_0 = \max_{1 \leq j \leq p} \frac{1}{\sqrt{n}} \sum_{1 \leq i \leq n} (y_{ij} - \mu_j)$$

However, this critical value is *infeasible* because $\mathrm{E}[x_i x_i']$ is unknown.

Instead, we derive a Gaussian multiplier bootstrap critical value Let $e_1, ..., e_n$ be an i.i.d. sequence of $N(0, 1)$ random variables that are independent of $x_1, ..., x_n$, and let

$$W_0 = \max_{1 \leq j \leq p} \frac{1}{\sqrt{n}} \sum_{1 \leq i \leq n} (x_{ij} - \mu_j) e_i$$

Multiplier Bootstrap II

Bootstrap test statistic:

$$W_0 = \max_{1 \le j \le p} \frac{1}{\sqrt{n}} \sum_{1 \le i \le n} (x_{ij} - \mu_j) e_i$$

Bootstrap critical value:

$$c_{W_0}(\alpha) = \inf\{t \in \mathbb{R} : \mathrm{P}_e(W_0 \le t) \ge \alpha\}$$

where $\mathrm{P}_e(\cdot)$ denotes the probability measure induced by the multiplier variables $e_1, ..., e_n$ holding $x_1, ..., x_n$ fixed.

Multiplier Bootstrap Theorem

Theorem (Multiplier Bootstrap Theorem)

Suppose that one of the following conditions is satistfied:

(i) *(E.1) and* $B_n^2(\log(pn))^7/n \leq C_1 n^{-c_1}$

(ii) *(E.2) and* $B_n^4(\log(pn))^7/n \leq C_1 n^{-c_1}$

In addition, suppose that the conditions (M.1) and (M.2) are satisfied. Then

$$\sup_{\alpha \in (0,1)} \left| \mathrm{P}(T_0 \leq c_{W_0}(\alpha)) - \alpha \right| \leq C n^{-c}$$

for some strictly positive constants c *and* C *that depend only on* c_1 *and* C_1.

Testing Many Moment Inequalities

Let $x_1, \dots, x_n$ be a sequence of i.i.d. random vectors in $\mathbb{R}^p$

- p can be much larger than n

Denote $\mu = \mathrm{E}[x_1]$ where $\mu = (\mu_1, \dots, \mu_p)'$.
We are interested in testing the null hypothesis, H_0, that

$$\mu_j \leq 0 \text{ for all } j = 1, \dots, p$$

against the alternative, H_a, that

$$\mu_j > 0 \text{ for some } j = 1, \dots, p.$$

We allow for triangular array asymptotics, so that p is allowed to depend on n.

Relation to Multiple Hypothesis Testing

Testing many moment inequalities (MMI) is closely related to testing multiple hypothesis tesing (MHT)

- MHT is a very popular research topic in statistics

But the emphasis is different:

- MMI is about increasing power given that some inequalities are not binding
- MHT is about increasing power given that some inequalities are not satisfied

Literature on Testing Moment Inequalities

Testing unconditional moment inequalities (among others):

- Chernozhukov, Hong, and Tamer (2007), Romano and Shaikh (2008), Rosen (2008), Chiburis (2009), Andrews and Guggenberger (2009), Andrews and Soares (2010), Canay (2010), Bugni (2010), Bugni (2011), Andrews and Jia Barwick (2012), Bugni, Canay, and Guggenberger (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 is *fixed*

When Many Moment Inequalities?

Behavioral choice models (Pakes, 2010):

- Consumer's decision what bundle to buy

Entry games (Ciliberto and Tamer, 2009):

- The number of moment inequalities corresponds to the number of different combinations of firms in the market

Conditional moment inequalities:

- Andrews and Shi (2013): represent conditional moments as an infinite number of unconditional moments

Test Statistic

Our test statistic is

$$T = \max_{1 \le j \le p} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} x_{ij}$$

We can also work with the studentized test statistic:

$$\hat{T} = \max_{1 \le j \le p} \frac{\sum_{i=1}^{n} x_{ij}}{\left(\sum_{i=1}^{n} (x_{ij} - \hat{\mu}_j)^2 \right)^{1/2}}$$

where $\hat{\mu}_j = \sum_{i=1}^{n} x_{ij}/n$.

Plug-in Critical Value

Let $e_1, \dots, e_n$ be an i.i.d. sequence of $N(0,1)$ random variables that are independent of the data.

Define

$$W = \max_{1 \le j \le p} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (x_{ij} - \hat{\mu}_j) e_i$$

where $\hat{\mu}_j = \sum_{i=1}^{n} x_{ij}/n$.

Then the multiplier bootstrap critical value $c_W(1-\alpha)$ is the conditional $(1-\alpha)$-quantile of W given $x_1, \dots, x_n$.

Refined Critical value

Let γ be some number such that $\gamma < \alpha/2$.

Let

$$J = \{j = 1, \dots, p : \hat{\mu}_j \geq -2c_W(1-\gamma)/\sqrt{n}\}.$$

Define

$$W_R = \max_{j \in J} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (x_{ij} - \hat{\mu}_j) e_i$$

Then the refined critical value $c_{W_R}(1-\alpha)$ is the conditional $(1 - \alpha + 2\gamma)$ quantile of W_R given $x_1, \dots, x_n$.

Size Control for the Test of Many Moment Inequalities

Theorem

Suppose that one of the following conditions is satisfied:

(i) *(E.1) and* $B_n^2(\log(pn))^7/n \leq C_1 n^{-c_1}$

(ii) *(E.2) and* $B_n^4(\log(pn))^7/n \leq C_1 n^{-c_1}$

In addition, suppose that the conditions (M.1) and (M.2) are satisfied. Then there exist strictly positive constants c *and* C*, depending only on* c_1 *and* C_1 *such that under* H_0*,*

$$\mathrm{P}(T \leq c_W(1-\alpha)) \geq 1 - \alpha - Cn^{-c},$$
$$\mathrm{P}(T \leq c_{W_R}(1-\alpha)) \geq 1 - \alpha - Cn^{-c}.$$

In addition, if all inequalities are binding and $\gamma \leq C_1 n^{-c_1}$*, then*

$$\mathrm{P}(T \leq c_W(1-\alpha)) \leq 1 - \alpha + Cn^{-c},$$
$$\mathrm{P}(T \leq c_{W_R}(1-\alpha)) \leq 1 - \alpha + Cn^{-c}.$$

Conclusion

We derived a new GAR

- dimension p of the data can be much larger than the sample size n
- we work formally with maxima but results can be extended to cover the class of all rectangles

We proved validity of the Gaussian multiplier method

- results are extended to the case of approximate maxima

We demonstrated usefulness of the results for

- testing many moment inequalities

Our results have many other applications as well.