Gaussian and bootstrap approximations to suprema of empirical processes

Kengo Kato (U. Tokyo) Joint with V. Chernozhukov (MIT) and D. Chetverikov (UCLA)

May, 2015

This talk is based upon the following paper

- Chernozhukov, V., Cheteverikov, D. and K. (2014a). Gaussian approximation of suprema of empirical processes. AoS.
- Chernozhukov, V., Cheteverikov, D. and K. (2015+). Multiplier and empirical bootstraps for suprema of empirical processes of increasing complexity, and related Gaussian couplings. SPA, to appear.

Some other related papers.

- Chernozhukov, V., Chetverikov, D. and K. (2015+). Comparison and anti-concentration bounds for maxima of Gaussian random vectors. PTRF, to appear.
- Chernozhukov, V., Chetverikov, D., and K. (2014b). Anti-concentration and honest, adaptive confidence bands. AoS.

Problem

- Let $X_1, \dots, X_n \sim P$ be i.i.d. r.v.'s taking values in $(S, \mathcal{S})$.
- Suppose we have a class $\mathcal{F}_n$ (possibly depending on n) of m'ble functions $S \to \mathbb{R}$, and consider the *empirical process* indexed by $\mathcal{F}_n$:

$$f \mapsto \mathbb{G}_n f = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (f(X_i) - \mathbb{E}[f(X_i)]), \ f \in \mathcal{F}_n.$$

  Forget some technical details.
- Problem: want to approximate (in some way) the distribution of

$$\sup_{f \in \mathcal{F}_n} \mathbb{G}_n f.$$

Motivations

- Classical example: Kolmogorov-Smirnov statistic.
- Let $U_1, \ldots, U_n \sim_{i.i.d.} U(0,1)$, and $F_n(t) = n^{-1} \sum_{i=1}^n 1(U_i \le t)$. Then

$$\sup_{t \in [0,1]} \sqrt{n} |F_n(t) - t| \xrightarrow{d} \sup_{t \in [0,1]} |B_t^0|,$$

where $B_t^0, t \in [0,1]$ is a Brownian bridge on $[0,1]$. Used as a goodness-of-fit test via probability integral transformation.

Construction of confidence bands

- Want to construct uniform confidence bands for an unknown function (density, conditional mean, conditional quantile etc.) f on a subset $\mathcal{X}$ of the domain of f.
- Suppose that a suitable estimate $\hat{f}$ of f is given, and let

$$Z_n := \sup_{x \in \mathcal{X}} \frac{|\hat{f}(x) - f(x)|}{\sigma_n(x)},$$

where $\sigma_n(x) > 0$ is a scaling constant. Roughly think of $\sigma_n(x)$ as

$$\text{Avar}(\hat{f}(x)) \approx \sigma_n^2(x).$$

- Then by setting

$$\hat{I}(x) = [\hat{f}(x) - \sigma_n(x)c_n(1-\alpha), \hat{f}(x) + \sigma_n(x)c_n(1-\alpha)],\ x \in \mathcal{X},$$

where

$$c_n(1-\alpha) := (1-\alpha)\text{-quantile of } Z_n,$$

we have

$$\mathbb{P}(f(x) \in \hat{I}(x), \forall x \in \mathcal{X}) = \mathbb{P}(Z_n \le c_n(1-\alpha)) \ge 1-\alpha.$$

- For most standard estimators (such as kernel and series estimators), Z_n can be approximately written as the supremum of an empirical process. Caution: corresp. classes of functions depend on n.
- For kernel estimators, under some (strong) conditions, it is typically possible to show that a rescaled version of Z_n converges in dist. to a Gumbel distribution. See Smirnov (1950, Doklady) and Bickel and Rosenblatt (73, AoS).
- For wavelet projection estimators, see Giné and Nickl (2010, AoS).

Goals

- Approximate $Z_n = \sup_{f \in \mathcal{F}_n} \mathbb{G}_n f$ by

$$\widetilde{Z}_n \stackrel{d}{=} \sup_{f \in \mathcal{F}_n} B_n f,$$

  where B_n is a Gaussian process indexed by $\mathcal{F}_n$ with mean zero and covariance function

$$\mathbb{E}[B_n(f) B_n(g)] = \mathrm{Cov}(f(X_1), g(X_1)), \ f, g \in \mathcal{F}_n.$$

- More precisely, we look for conditions under which such $\widetilde{Z}_n$ exist with

$$|Z_n - \widetilde{Z}_n| = O_{\mathbb{P}}(r_n), \ r_n \to 0. \tag{*}$$

  We develop a new technique toward this problem.
- Z_n itself needs not be weakly convergent (even after normalization).

- If (*) is verified, by using an *anti-concentration inequality* for the supremum of a Gaussian process, we have

$$r_n \mathbb{E}[\widetilde{Z}_n] = o(1) \Rightarrow \sup_{t \in \mathbb{R}} |\mathbb{P}(Z_n \le t) - \mathbb{P}(\widetilde{Z}_n \le t)| = o(1),$$

  as long as the variance of the empirical process is bounded and bounded away from zero.
- Note: $\mathbb{E}[\widetilde{Z}_n] \to \infty$ is allowed.

Theorem (CCK, 15+, PTRF)

Let $X(t), t \in T$ be a separable Gaussian process s.t. $\mathbb{E}[X(t)] = 0, \mathbb{E}[X^2(t)] = 1, \forall t \in T$, and $\sup_{t \in T} X(t) < \infty$ a.s. Then for every $\epsilon > 0$,

$$\sup_{x \in \mathbb{R}} \mathbb{P}(|\sup_{t \in T} X(t) - x| \le \epsilon) \le 4\epsilon(\mathbb{E}[\sup_{t \in T} X(t)] + 1).$$

Literature review

- The problem of approximating the distribution of the supremum of an empirical process is old.
- A (standard) approach is to approximate *whole* empirical processes in the sup-norm: construct (suitable) Gaussian processes B_n such that

$$\sup_{f \in \mathcal{F}_n} |(\mathbb{G}_n - B_n) f| = O_{\mathbb{P}}(r_n), \ r_n \to 0.$$

- A powerful result due to Komlós, Major and Tusnády (1975, PTRF).

- There are extensions of KMT approximations to general empirical processes. Rio (94, PTRF), Koltchinskii (94, J. Theoret. Probab.).
- These are powerful and have been used in the statistics literature.
- There are approaches different from KMT. Dudley and Philipp (83, PTRF), Berthet and Mason (06, IMS Lecture Notes), Setatti (09, SPA). These results are of limited use for statistical applications.
- *Directly* approximating the supremum of an empirical process (which we do) has not been well explored before.

Theorem (Rio, 94, PTRF, Theorem 1)

Let $X_1, \dots, X_n \sim_{i.i.d.} P$ on $[0,1]^d,\ d \geq 2$, and assume that P has a conti. and positive density on $[0,1]^d$. Moreover, $\mathcal{F}$ is a VC type class of m'ble functions $[0,1]^d \to [-1,1]$, and the total variation of functions in $\mathcal{F}$ is $\leq K(\mathcal{F})$ (+some m. condition). Then there exists a (suitable) Brownian bridge B_n indexed by $\mathcal{F}$ such that

$$\mathbb{P}\left(\sup_{f \in \mathcal{F}} |(\mathbb{G}_n - B_n)f| > Cn^{-1/(2d)}\sqrt{K(\mathcal{F})t} + n^{-1/2}Ct\sqrt{\log n}\right) \leq e^{-t},\ t > 0.$$

Main theorem

- Let $X_1, \dots, X_n \sim_{i.i.d.} P$ in $(S, \mathcal{S})$.
- $\mathcal{F}$: a class of m'ble functions $S \to \mathbb{R}$ such that

$$\int f dP = 0, \ \forall f \in \mathcal{F}.$$

Denote by F a m'ble envelope of $\mathcal{F}$.

$$F(x) \geq \sup_{f \in \mathcal{F}} |f(x)|, \ \forall x \in S.$$

- For a p.m. Q on $(S, \mathcal{S})$, write the $\mathcal{L}^2(Q)$-semimetric as e_Q:

$$e_Q^2(f, g) = \int (f - g)^2 dQ.$$

Assumptions

(A1) (PM class) $\exists \mathcal{G} \subset \mathcal{F}$: countable s.t. $\forall f \in \mathcal{F}, \exists g_m \in \mathcal{G}$ with $g_m \to f$ pointwise.

(A2) (VC type)

$$\sup_{Q:\text{finitely discrete}} N(\mathcal{F}, e_Q, \varepsilon \|F\|_{Q,2}) \leq (A/\varepsilon)^v, \ 0 < \forall \varepsilon \leq 1.$$

(A3) There exist $b \geq \sigma > 0$ and $q \in [4, +\infty]$ s.t.

$$\sup_{f \in \mathcal{F}} P|f|^k \leq \sigma^2 b^{k-2} \text{ for } k = 2, 3, 4, \quad \text{and} \quad \|F\|_{P,q} \leq b.$$

Theorem (CCK, 14a, AoS)

Let $Z = \sup_{f \in \mathcal{F}} \mathbb{G}_n f$. *Then there exists a tight, Borel m'ble Gaussian random element* G_P *in* $\ell^\infty(\mathcal{F})$ *with mean zero and covariance function* $\mathbb{E}[G_P(f) G_P(g)] = P(fg)$, *and for every* $\gamma \in (0,1)$, *there exists a random variable* $\widetilde{Z} \stackrel{d}{=} \sup_{f \in \mathcal{F}} G_P f$ *such that*

$$\mathbb{P}\left\{ |Z - \widetilde{Z}| > C_1 \left(\frac{bK}{\gamma^{1/2} n^{1/2 - 1/q}} + \frac{(b\sigma)^{1/2} K^{3/4}}{\gamma^{1/2} n^{1/4}} + \frac{(b\sigma^2 K^2)^{1/3}}{\gamma^{1/3} n^{1/6}} \right) \right\}$$
$$\leq C_2 \{\gamma + (\log n)/n\},$$

where C_1, C_2 *depend only on* q, *and*

$$K = K(n, \sigma, b, A, v) = v(\log n \vee \log(Ab/\sigma)).$$

Further results: Bootstrap

- Focus here on the Gaussian multiplier bootstrap (essentially similar results hold for empirical bootstrap).
- Let $\xi_1, \dots, \xi_n \sim N(0,1)$ i.i.d. indep. of $X_1^n := \{X_1, \dots, X_n\}$.

$$\mathbb{G}_n^{\xi} f = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_i (f(X_i) - P_n f), \ f \in \mathcal{F}.$$

Theorem (CCK, 15, arXiv)

Keep the setting of the previous theorem (but assume $q < +\infty$ *and* S *to be a separable metric space). Let* $Z^\xi = \sup_{f \in \mathcal{F}} \mathbb{G}_n^\xi f$*. Then for every* $\gamma, \varphi \in (0,1)$*, there exists a random variable* $\tilde{Z}^\xi$ *such that* $\tilde{Z}^\xi \mid X_1^n \stackrel{d}{=} \sup_{f \in \mathcal{F}} G_P f$ *and*

$$\mathbb{P}\left\{ |Z^\xi - \tilde{Z}^\xi| > C_3 \left(\frac{bK}{\varphi \gamma^{1/q} n^{1/2-1/q}} + \frac{(b\sigma K^{3/2})^{1/2}}{\varphi \gamma^{1/q} n^{1/4}} \right) \right\}$$
$$\leq C_4(\gamma + \varphi + n^{-1}),$$

provided that $K \leq n$*, where* C_3, C_4 *depend only on* q*.*

Application: local empirical processes

- $(Y_1, X_1), \dots, (Y_n, X_n)$: i.i.d. r.v.'s taking values in $\mathcal{Y} \times \mathbb{R}^d$.
- $\mathcal{G}$: a class of m'ble functions $\mathcal{Y} \to \mathbb{R}$.
- $k(\cdot) : \mathbb{R}^d \to \mathbb{R}$: a (possibly higher order) kernel function.
- $h_n \to 0$: a bandwidth.
- $\mathcal{I}$: a Borel m'ble subset of $\mathbb{R}^d$.
- Consider the kernel type statistic of the form:

$$S_n(x, g) = \frac{1}{n h_n^d} \sum_{i=1}^{n} g(Y_i) k(h_n^{-1}(X_i - x)), \ (x, g) \in \mathcal{I} \times \mathcal{G}.$$

Various applications.

- Einmahl and Mason (97, PTRF) called the process

$$g \mapsto S_n(x, g)$$

a local empirical process at x with fixed x. We would call the process

$$(x, g) \mapsto S_n(x, g)$$

a local empirical process.

- Consider

$$W_n = \sup_{(x,g) \in \mathcal{I} \times \mathcal{G}} c_n(x, g) \sqrt{n h_n^d} (S_n(x, g) - \mathbb{E}[S_n(x, g)]).$$

Assumptions

(B1) $\mathcal{G}$ is a PM class of functions $\mathcal{Y} \to \mathbb{R}$ uniformly bounded by a constant $b > 0$, and is VC type with envelope $\equiv b$.

(B1)' $\mathcal{G}$ is a PM class of functions $\mathcal{Y} \to \mathbb{R}$ with m'ble envelope G such that $\mathbb{E}[G^q(Y_1)] < \infty$ for some $q \geq 4$ and $\sup_{x \in \mathbb{R}^d} \mathbb{E}[G^4(Y_1) \mid X_1 = x] < \infty$. Moreover, $\mathcal{G}$ is VC type with envelope G.

(B2)-(B5) are standard and skipped.

Proposition

Assume (B1)-(B5) (or (B1)',(B2)-(B5)). Then there exist tight Gaussian random elements B_n *in* $\ell^\infty(\mathcal{I} \times \mathcal{G})$ *with mean zero and covariance function*

$$\mathbb{E}[B_n(x,g)B_n(\check{x},\check{g})] = h_n^{-d} c_n(x,g) c_n(\check{x},\check{g}) \times \mathrm{Cov}[g(Y_1)k(h_n^{-1}(X_1 - x)), \check{g}(Y_1)k(h_n^{-1}(X_1 - \check{x}))],$$

and r.v.'s $\widetilde{W}_n \stackrel{d}{=} \sup_{(x,g)\in\mathcal{I}\times\mathcal{G}} B_n(x,g)$ *s.t.*

$$|W_n - \widetilde{W}_n| = \begin{cases} O_{\mathbb{P}}\{(nh_n^d)^{-1/6}\log n + (nh_n^d)^{-1/4}\log^{5/4} n + (nh_n^d)^{-1/2}\log^{3/2} n\}, & \text{(B1)}, \\ O_{\mathbb{P}}\{(nh_n^d)^{-1/6}\log n + (nh_n^d)^{-1/4}\log^{5/4} n + (n^{1-2/q}h_n^d)^{-1/2}\log^{3/2} n\}, & \text{(B1)'}. \end{cases}$$

Moreover, if $\mathbf{Var}(c_n(x,g)\sqrt{nh_n^d}S_n(x,g)) \geq \underline{\sigma}^2 > 0$ for all $(x,g) \in \mathcal{I} \times \mathcal{G}$, we have

$$|W_n - \widetilde{W}_n| = o_{\mathbb{P}}(\log^{-1/2} n)$$
$$\Rightarrow \sup_{t \in \mathbb{R}} |\mathbb{P}(W_n \leq t) - \mathbb{P}(\widetilde{W}_n \leq t)| = o(1).$$

Note that $|W_n - \widetilde{W}_n| = o_{\mathbb{P}}(\log^{-1/2} n)$ (i) if $nh_n^d / \log^c n \to \infty$ under (B1), and (ii) if $n^{(1-2/q)}h_n^d / \log^c n \to \infty$ under (B1)'.

Thank you for your attention!

Supplement

A sketch of the proof of the main theorem

The proof of the main theorem has the following four steps.

1. the reduction of the problem vis Strassen's theorem.
2. discretization of the empirical and Gaussian processes (here the covering numbers appear).
3. coupling to the maximum of the discretized empirical process (here Stein's method is used).
4. control of the discretization errors.

A version of Strassen's theorem

Lemma

Let μ and ν be Borel probability measures on $\mathbb{R}$, and let $V \sim \mu$. Let $\varepsilon > 0$ and $\delta > 0$ be two positive constants. Then there exists a $W \sim \nu$ s.t. $\mathbb{P}(|V - W| > \delta) \leq \varepsilon$ iff $\mu(A) \leq \nu(A^{\delta}) + \varepsilon$ for every Borel subset A of $\mathbb{R}$.

Coupling to the maximum of the discretized empirical process

Theorem (CCK, 14a, AoS)

Let $X_1, \dots, X_n$ *be independent random vectors in* $\mathbb{R}^p$ *with mean zero and finite absolute third moment. Let* $Z = \max_{1 \le j \le p} \sum_{i=1}^n X_{ij}$*, and take*

$$Y_i \overset{\text{indep.}}{\sim} N(0, \mathbb{E}[X_i X_i^T]), 1 \le i \le n.$$

Then $\forall \beta > 0, \forall \delta > 1/\beta$, $\exists \widetilde{Z} \overset{d}{=} \max_{1 \le j \le p} \sum_{i=1}^n Y_{ij}$ *s.t.*

$$\mathbb{P}(|Z - \widetilde{Z}| > 2\beta^{-1} \log p + 3\delta) \le \frac{\varepsilon + C\beta\delta^{-1}\{B_1 + \beta(B_2 + B_3)\}}{1 - \varepsilon},$$

where $\varepsilon = \varepsilon_{\beta,\delta}$ *is given by*

$$\varepsilon = \sqrt{e^{-\alpha}(1+\alpha)} < 1, \ \alpha = \beta^2\delta^2 - 1 > 0.$$

Moreover, B_1, B_2, B_3 are given by

$$B_1 = \mathbb{E}\left[\max_{1\le j,k\le p} |\sum_{i=1}^n (X_{ij}X_{ik} - \mathbb{E}[X_{ij}X_{ik}])|\right],$$

$$B_2 = \mathbb{E}\left[\max_{1\le j\le p}\sum_{i=1}^n |X_{ij}|^3\right],$$

$$B_3 = \sum_{i=1}^n \mathbb{E}\left[\max_{1\le j\le p}|X_{ij}|^3 \cdot 1\left(\max_{1\le j\le p}|X_{ij}| > \beta^{-1}/2\right)\right].$$

Corollary

$\forall \delta > 0$, $\exists \widetilde{Z} \stackrel{d}{=} \max_{1 \le j \le p} \sum_{i=1}^{n} Y_{ij}$ *s.t.*

$$\mathbb{P}(|Z - \widetilde{Z}| > 16\delta) \lesssim \delta^{-2}\{B_1 + \delta^{-1}(B_2 + B_4)\log(p \vee n)\}\log(p \vee n) + n^{-1}\log n.$$

Here

$$B_4 = \sum_{i=1}^{n} \mathbb{E}\left[\max_{1 \le j \le p} |X_{ij}|^3 \cdot 1\left(\max_{1 \le j \le p} |X_{ij}| > \delta / \log(p \vee n)\right)\right].$$

- If $p = p_n \to \infty$ as $n \to \infty$, $X_{ij} = x_{ij}/\sqrt{n}, |x_{ij}| \le C$, then

$$B_1 = O(n^{-1/2} \log^{1/2} p_n), \ B_2 + B_4 = O(n^{-1/2}).$$

Hence to guarantee $|Z - \widetilde{Z}| = o_{\mathbb{P}}(1)$, we just need

$$\log p_n = o(n^{1/4})!$$

- Important is the fact that the bound depends on p only through $\log p$.