Processos empírics i aplicacions: visió general amb biaix

Evarist Giné-Masdeu*
The University of Connecticut

Introducció

La teoria de processos empírics gira entorn de la vella pregunta: com és de ben aproximada la probabilitat per la freqüència, o l'esperança per la mitjana? I tant els seus resultats com els seus mètodes tenen aplicació en estadística asimptòtica.

El *període clàssic*, 1920-1960, considera funcions de distribució de $\mathbb{R}$ i també de $\mathbb{R}^d$ (Glivenko i Cantelli, Kolmogorov i Smirnov, Cramér, Kac, Doob i Donsker, Kiefer, etc.). Aquest aspecte de la teoria encara es manté actiu, de fet amb nou ímpetu a causa de resultats importants sobre aproximacions fortes. Això no és tractat en aquestes notes.

Vapnik i Červonenkis, 1971, i Dudley, 1978, estengueren considerablement el camp d'estudi i, de camí, rejovenirem la teoria *de* processos empírics. Tractaré de descriure a) alguns dels resultats i mètodes principals en la direcció que ells establiren (obtinguts per diversos autors en el curs dels darrers quinze anys), *bootstrap* inclòs, i b) en forniré una aplicació, a tall d'il·lustració. La meva intenció és donar una vista panoràmica —necessàriament superficial, necessàriament esbiaixada— del tema, que permeti al lector fer-se una idea del contigut de la teoria i de les possibles aplicacions a l'estadística asimptòtica.

Plantejament d'alguns problemes

Sigui $(S, \mathcal{S}, \mathrm{P})$ un espai de probabilitats, i siguin X_i variables aleatòries i.i.d. (P) (una mica més: de fet volem que $X_i : S^{\mathbb{N}} \to S$, siguin les funcions coordenades). La *mesura empírica* P_n dóna massa $\frac{1}{n}$ a cada observació X_i, $i = 1,\ldots, n$:

$$\mathrm{P}_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i}, \quad n = 1, 2, \ldots$$

$$\text{(i.e.,}\quad \mathrm{P}_n(C) = \frac{\#\text{ de } X_i \text{ dins de } C}{n}, \quad \mathrm{P}_n f = \frac{1}{n}\sum_{i=1}^{n} f(X_i)).$$

---

* Aquest treball, basat en les Forum Lectures donades per l'autor al 20th European Meeting of Statisticians, Bath, Anglaterra, setembre 1992, ha estat subvencionat en part pel Centre de Recerca Matemàtica de l'IEC i per la National Science Foundation d'Estats Units

La pregunta que es feren Vapnik i Červonenkis és la següent: *per a quines famílies de conjunts mesurables, $\mathcal{C} \subset \mathcal{S}$, es verifica*

$$\sup_{C \in \mathcal{C}} |P_n(C) - P(C)| \to 0 \quad \text{en pr. o g.s.?}$$

O, més en general, *per a quines famílíes de funcions $\mathcal{F}$ es té que*

$$\sup_{f \in \mathcal{F}} |P_n(f) - P(f)| \to 0 \quad \text{en pr. o g.s.?}$$

Si S = IR i $\mathcal{C}$ és el conjunt de semirectes $(-\infty, x]$, la resposta a aquesta pregunta és el clàssic teorema de Glivenko-Cantelli. Vapnik i Červonenkis substituïren, doncs, en llur pregunta, (IR, B) per un espai mesurable general (S, $\mathcal{S}$) i la família de semirectes per famílies generals de conjunts i de funcions.

Podem fer-nos preguntes semblants sobre el teorema central del límit, la llei del logaritme iterat, desigualtats exponencials, etc.

[Les respostes a aquestes preguntes haurien de ser força útils ja que sovint, en estadística, es té necessitat de controlar quantitats com $\sum_{i=1}^n f_{\theta_n}(X_i)$, on θ_n és un estadístic: el control d'aquestes variables fóra conseqüència del control de $\sup | \sum_{i=1}^n f_\theta(X_i) |$, on el suprem és pres sobre el conjunt dels valors de la variable θ_n. Òbviament cal considerar famílies $\{f_\theta\}$ més generals que no només indicadors de semirectes —com veurem més avall.]

Ara considerarem breument com plantejar el teorema de Kolmogorov-Smirnov en aquest context general. Cal recordar que normalment el teorema de Kolmogorov-Smirnov s'obté com a conseqüència del teorema de Donsker sobre convergència feble de la funció de distribució (cumulativa) empírica, considerada com un element aleatori de (o com un procés amb trajectòries a) l'espai $D(-\infty, \infty)$, o $D(0, 1)$, de funcions càdlàg. No es pot definir un tal espai D sobre una classe general de funcions $\mathcal{F}$. Però, si

$$\sup_{f \in \mathcal{F}} |f(s)| := F(s) < \infty \quad \text{per a tot } s \in S$$

aleshores, l'aplicació

$$f \to P_n f$$

és una funció fitada sobre $\mathcal{F}$ i, si $PF < \infty$, l'aplicació $f \to Pf$ també és fitada. Per tant considerarem, en lloc de l'espai D, *l'espai de les funcions fitades sobre $\mathcal{F}$, amb la norma del suprem,*

$$\|x\|_{\mathcal{F}} = \sup_{f \in \mathcal{F}} |x(f)|,$$

que denotarem $l^\infty(\mathcal{F})$. El *pont brownià* ja no serà el procés límit, ara. Però si $Pf^2 < \infty$ aleshores, pel teorema central del límit a IR, $n^{\frac{1}{2}}(P_n - P)(f) \to_d N(0, \mathrm{Var}_P(f))$ i de fet la convergència és conjunta en qualsevol nombre finit de funcions f. Aleshores, caldrà canviar el pont brownià pel *procés gaussià* $\{G_P(f) : f \in \mathcal{F}\}$ centrat, de covariància

$$\mathrm{E}G_P(f)G_P(g) = \mathrm{P}[(f - \mathrm{P}f)(g - \mathrm{P}g)], \quad f, g \in \mathcal{F}.$$

I, per analogia amb una de les diferents definicions equivalents de convergència en llei a $\mathbb{R}$ o a $\mathbb{R}^d$, direm que

$$\mathcal{F} \in TCL(\mathrm{P}) \text{ o que } \mathcal{F} \text{ és } \mathrm{P} - Donsker$$

si

(1) *G_P és un 'bon' procés*, i si

(2) *per a tota $H : l^\infty(\mathcal{F}) \to \mathbb{R}$ fitada i contínua*,

$$\mathrm{E}^* H(n^{\frac{1}{2}}(\mathrm{P}_n - \mathrm{P})) \to \mathrm{E}H(G_P).$$

[Aclariment sobre (1): 'bon' procés significa que la llei de G_P és *ajustada* o *Radon* en $l^\infty(\mathcal{F})$, condició que és equivalent al fet que G_P tingui una versió de trajectòries *fitades i uniformement contínues en la seva pròpia L_2 - pseudomètrica* o, el el nostre cas, en la (pseudo) mètrica $e_P(f, g) = \sqrt{P(f-g)^2}$, lleugerament més tractable. Sobre (2): E* denota integral exterior ja que $H\,(n^{\frac{1}{2}}(\mathrm{P}_n - \mathrm{P}))$ no té per què ser mesurable —si bé en aquest treball tendirem a oblidar-nos de les condicions de mesurabilitat.]

Per exemple, si $\mathcal{F}$ és P-Donsker aleshores, sota condicions de mesurabilitat que a la pràctica es compleixen gairebé sempre, es dedueix el següent resultat, anàleg al teorema de Kolmogorov-Smirnov:

$$\sup_{f \in \mathcal{F}} \left| n^{\frac{1}{2}}(\mathrm{P}_n(f) - \mathrm{P}(f)) \right| \to_d \sup_{f \in \mathcal{F}} |G_P(f)|.$$

I també es verifica això mateix per a qualsevol funcional continu per la norma del suprem sobre $\mathcal{F}$.

La definició anterior és deguda a Dudley (1985) i a Hoffmann-Jørgensen (1991), conjuntament.

Ara, doncs, podem preguntar-nos: *per a quines classes de conjunts* (= funcions indicatrius) *o de funcions $\mathcal{F}$ es té que $\mathcal{F} \in TCL(\mathrm{P})$*?

I aquí cal dir que la resposta a aquesta pregunta produeix resultats intermedis que són sovint més útils que els mateixos teoremes de límit, concretament, *desigualtats maximals*: normalment, hom demostra que $\mathcal{F} \in TCL(\mathrm{P})$ verificant una espècie de criteri generalitzat de Prokhorov (Dudley, 1984, i una demostració més senzilla a Giné i Zinn, 1986): una classe mesurable de funcions $\mathcal{F}$ tal que $\|\mathrm{P}\|_{\mathcal{F}} < \infty$ és P-Donsker si i només si F compleix les dues condicions següents:

(i) *$(\mathcal{F}, e_P)$ és un espai totalment fitat* i

(ii) $\lim_{\delta \to 0} \limsup_{n \to \infty} \mathrm{E} \sup_{\substack{e_P(f,g) \leq \delta \\ f, g \in \mathcal{F}}} n^{\frac{1}{2}} \left| (\mathrm{P}_n - \mathrm{P})(f - g) \right| = 0.$

Anomenarem aquesta segona condició *equicontinuïtat asimptòtica* del procés empíric. Tot sovint les aplicacions requereixen simplement el control de la qualitat dins del límit de sobre, i atènyer aquest control és la part principal de la demostració de gairebé tots els TCL per a processos empírics.

Sobre les 'solucions' als problemes anteriors

Hi ha disponibles solucions gairebé completes als problemes plantejats a sobre (LGN, TCL; i també LLI), encara que en termes no gaire útils, almenys a primera vista. Tot seguit descriuré les solucions a l'LGN i al TCL, però només per a *classes de conjunts*.

A fi d'introduir el concepte adequat, primer descriuré un procediment d'*aleatorització* útil i senzill.

Siguin

◊ $\{X_i'\}$ una copia independent de $\{X_i\}$,

◊ $P_n' = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i'}$, i

◊ $\{\varepsilon_i\}$ una successió de Rademacher ($\varepsilon_i = +1$ *o* -1 *amb probabilitats* $\frac{1}{2}$, *independents*) independent de totes les altres variables.

Aleshores, suprimint el subíndex $\mathcal{F}$, es té, per simetria, Jensen conditional, i Fubini,

$$\begin{aligned}
E\|P_n - P\| &= E\|(P_n - P) - E'(P_n' - P)\| \\
&\le E\|P_n - P_n'\| = E\|\frac{1}{n}\sum_{i=1}^{n}(\delta_{X_i} - \delta_{X_i'})\| \\
&= E\|\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i(\delta_{X_i} - \delta_{X_i'})\| \\
&\le 2E\|\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i\delta_{X_i}\|
\end{aligned}$$

i

$$\begin{aligned}
E\|\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i\delta_{X_i}\| &\le E\|\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i(\delta_{X_i} - P)\| + (\sup_{f\in\mathcal{F}}|Pf|)E|\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i| \\
&\le 2E_\varepsilon E_X\|\frac{1}{n}\sum_{i\le n:\ \varepsilon_i=1}(\delta_{X_i} - P)\| \\
&\qquad + (\sup_{f\in\mathcal{F}}|Pf|)E|\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i| \\
&\le 2E\|P_n - P\| + (\sup_{f\in\mathcal{F}}|Pf|)E|\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i|.
\end{aligned}$$

Per tant, suposant $\sup_{f \in F} |Pf| < \infty$,

$$\|\mathrm{P}_n - \mathrm{P}\| \to 0 \quad si\ i\ només\ si \quad \|\frac{1}{n}\sum_{i=1}^{n} \varepsilon_i \delta_{X_i}\| \to 0$$

(g.s. o en pr.: aquestes successions són submartingales invertides).

Posem per cas, ara, que $\mathcal{C}$ sigui una família tan complexa que, per a tot n, amb probabilitat $\alpha_n \not\to 0$, cada subconjunt de la mostra enèsima, $\{X_1(\omega),\ldots,X_n(\omega)\}$, sigui la intersecció (de la mostra) amb algun subconjunt $C \in \mathcal{C}$. En particular, els subconjunts $\{X_i(\omega) : \varepsilon_i = 1, i \leq n\}$ són interseccions d'aquestes i per tant es té que, amb probabilitat almenys α_n,

$$\|\frac{1}{n}\sum_{i=1}^{n} \varepsilon_i \delta_{X_i}\| \geq \frac{1}{n}\#\{i \leq n : \varepsilon_i = 1\}.$$

Com que aquestes variables aleatòries tendeixen g.s. a $\frac{1}{2}$ resulta que és impossible que $\| \frac{1}{n}\sum_{i=1}^{n} \varepsilon_i \delta_{X_i} \| \to 0$ en probabilitat, això és, per l'argument de simetrització donat a sobre, $\| \mathrm{P}_n - \mathrm{P} \| \not\to 0$ en probabilitat.

Per tant, cal examinar les traces dels conjunts de la col·lecció C sobre les mostres. La solució de VC a la LGN uniforme és: sota mesurabilitat, l'LGN uniforme sobre $\mathcal{C}$ es pot caracteritzar per la cardinalitat de les traces dels subconjunts de $\mathcal{C}$ sobre les mostres. Explícitament: Si, per a $\mathcal{C} \subset \mathcal{S}$ i $x_1,\ldots,x_n \in S$, definim

$$\Delta^{\mathcal{C}}(x_1,\ldots,x_n) = \#\{C \cap \{x_1,\ldots,x_n\} : C \in \mathcal{C}\},$$

es té:

TEOREMA 1. (Vapnik i Červonenkis, 1971). *Sota condicions de mesurabilitat,*

$$\|\mathrm{P}_n - \mathrm{P}\| \to 0 \quad g.s.\ (pr.)$$

si i només si

$$\frac{\log \Delta^{\mathcal{C}}(X_1,\ldots,X_n)}{n} \to 0 \quad pr.$$

De fet, Vapnik i Červonenkis obtingueren la llei feble i Steele observà que la llei forta es complia sota les mateixes hipòtesis.

[Una condició de mesurabilitat sota la qual tots els teoremes d'aquest article són certs és la següent: $(S,\mathcal{S})$ és un espai de Suslin (p.e. la imatge d'un espai mètric, separable i complet per una funció mesurable), i $\mathcal{C}$ es pot parametritzar pels elements d'un altre espai de Suslin $(\Theta, \mathcal{A})$ de manera que la funció d'avaluació $(\theta, s) \to I_{C_\theta}(s)$ és mesurable conjuntament. (Dudley, 1984)].

Exemple: si $C = \{(-\infty, x] : x \in \mathbb{R}\}$, aleshores $\Delta^C(x_1,\ldots,x_n) \leq n + 1$, i el teorema de Glivenko-Cantelli es dedueix del teorema 1, amb molt d'escreix.

Un fet remarcable, degut a Sauer (1972) i a Vapnik-Červonenkis (1971) independentment, és que, *o bé*

$$m^{\mathcal{C}}(n) := \sup_{\substack{T \subset S \\ \#T \leq n}} \Delta^{\mathcal{C}}(T) = 2^n \quad per\ a\ tot\ n$$

o bé

$$m^{\mathcal{C}}(n) \ creix\ polinomialment$$

En el segon cas direm que $\mathcal{C}$ és una *classe VC de conjunts*. I naturalment, pel teorema 1, la llei dels grans nombres es verifica uniformement sobre qualsevol classe VC.

Exemples de classes VC de conjunts:

◊ Els quadrants inferiors esquerres de $\mathbb{R}^d$ (VC, 1971).
◊ Els semiespais tancats de $\mathbb{R}^d$ (VC, 1971).
◊ Les esferes tancades de $\mathbb{R}^d$ (Dudley, 1979).
◊ Si G és un espai vectorial de dimensió finita de funcions reals, aleshores $\{g > 0 : g \in G\}$ és VC (Dudley, 1978). I si aquestes funcions són polinomis de grau fitat, les projeccions d'aquests conjunts sobre menys variables també formen una classe VC (Stengle i Yukich, 1989).
◊ Les classes obtingudes de classes VC per un nombre finit d'operacions booleanes també són VC (Dudley, 1978).

La solució al problema del teorema central del límit uniforme sobre classes de conjunts és descrita pel teorema següent. És convenient primer d'introduir la notació següent: direm que la classe de conjunts $\mathcal{C}$ és P- *pregaussiana* si el procés $\{G_P(C) : C \in \mathcal{C}\}$ és bo.

Aleshores:

TEOREMA 2. (Giné i Zinn, 1984, i Talagrand, 1988). *Una classe mesurable de conjunts, $\mathcal{C}$, és P-Donsker si i només si:*

(1) $\mathcal{C}$ *és P-pregaussiana i*
(2)

$$\frac{\log \Delta^{\mathcal{C}}(X_1, \ldots, X_n)}{n^{\frac{1}{2}}} \to 0 \quad pr.$$

Les classes VC de conjunts són P-pregaussianes i, per tant, si són mesurables, el teorema 2 s'aplica i es dedueix que són P-Donsker (el teorema central del límit per a classes VC mesurables fou demostrat originalment per Dudley, 1978).

Hi ha teoremes semblants al teorema 2 per a classes de funcions: vegeu Giné i Zinn, 1984 i 1986 i/o Ledoux i Talagrand, 1989.

La demostració del teorema 2 és massa llarga i complicada per a ser donada en una conferència general; donarem només les demostracions de la part de suficiència del teorema 1 i de la versió del teorema 2 per a classes VC. Les demostracions consistiran bàsicament en la tècnica de *transferència de propietats dels processos gaussians a processos empírics per mitjà d'aleatorització i condicionament*, que ha demostrat ser de molta utilitat. Aquesta tècnica fou introduïda en els processos empírics per Giné i Zinn, 1984, però no és nova: per exemple Marcus i Pisier la feren servir abans en llur estudi exemplar sobre sèries de Fourier aleatòries. L'aleatorització mitjançant variables de Rademacher, també feta servir abans en situacions similars, fou introduïda per Pollard, 1981, en els processos empírics, i és tot el que necessitaríem en les demostracions que segueixen, però l'aleatorització per variables normals, que és bàsica en les demostracions de les parts de necessitat del teoremes anteriors, és més còmoda.

És clàssic i prou conegut que si g_i, $i \leqslant N$, tenen distribució $N(0, \sigma_i^2)$ aleshores (encara que siguin independents)

$$(*) \qquad \mathrm{E} \max_{i \le N} |g_i| \le 2^{\frac{3}{2}} \sqrt{\log N} \max_{i \le N} \sigma_i .$$

[Vegeu-ne però la demostració de Pisier, àdhuc més general: sigui ϕ una funció convexa, creixent i no negativa (com ara $\phi(x) = \exp x^2$) i siguin c_i constants tals que

$$\mathrm{E}\phi\left(\frac{|\xi_i|}{c_i}\right) \le c;$$

aleshores

$$\begin{aligned}\phi\left(\mathrm{E}\frac{\max_{i \le N} |\xi_i|}{\max_{i \le N} c_i}\right) &\le \mathrm{E}\phi\left(\max_{i \le N} \frac{|\xi_i|}{c_i}\right) \\ &= E \max_{i \le N} \phi\left(\frac{|\xi_i|}{c_i}\right) \le cN\end{aligned}$$

i per tant,

$$\mathrm{E} \max_{i \le N} |\xi_i| \le \phi^{-1}(cN) \max_{i \le N} c_i;$$

(*) se'n dedueix ja que $\mathrm{E} \exp \dfrac{g_1^2}{4\sigma_i^2} = \sqrt{2}$.]

Sigui ara g una variable N (0, 1) genèrica i siguin g_i N (0, 1) i.i.d., independents de la mostra i de les ε_i. Continuant una de les desigualtats de simetrització de més amunt, tenim

$$\begin{aligned}\mathrm{E}\|\mathrm{P}_n - \mathrm{P}\|_{\mathcal{C}} &\le 2\mathrm{E}\Big\|\frac{1}{n}\sum_{i=1}^{n} \varepsilon_i \delta_{X_i}(C)\Big\|_{\mathcal{C}} \\ &= 2\mathrm{E}\Big\|\frac{1}{n}\sum_{i=1}^{n} \varepsilon_i (\mathrm{E}|g_i|) \delta_{X_i}(C)\Big\|_{\mathcal{C}} \frac{1}{\mathrm{E}|g|}\end{aligned}$$

$$\leq \sqrt{\frac{\pi}{2}}\mathrm{E}\|\frac{1}{n}\sum_{i=1}^{n} g_i\delta_{X_i}(C)\|_{\mathcal{C}}$$

$$= \sqrt{\frac{\pi}{2}}\mathrm{E}_X\mathrm{E}_g\|\frac{1}{n}\sum_{i=1}^{n} g_i\delta_{X_i}(C)\|_{\mathcal{C}}.$$

Condicionalment sobre la mostra, l'expressió $\|\frac{1}{n}\sum_{i=1}^{n} g_i\delta_{X_i}(C)\|_{\mathcal{C}}$ és un suprem sobre una col·lecció de normals. La variància de cada una d'elles és dominada per $\frac{1}{n^2}\sum_{i=1}^{n} 1 = \frac{1}{n}$. I ara observem que

$$C \cap \{X_1, \ldots, X_n\} = D \cap \{X_1, \ldots, X_n\}$$

$$\Rightarrow \sum_{i=1}^{n} g_i\delta_{X_i}(C) = \sum_{i=1}^{n} g_i\delta_{X_i}(D),$$

és a dir, que el suprem $\|\cdot\|_{\mathcal{C}}$ és realment només un màxim sobre

$$N = \Delta^{\mathcal{C}}(X_1, \ldots ; X_n)$$

termes. Per tant, (*) ens dona

$$\mathrm{E}\|\mathrm{P}_n - \mathrm{P}\|_{\mathcal{C}} \leq 2\sqrt{\pi}\mathrm{E}\sqrt{\frac{\log \Delta^{\mathcal{C}}(X_1, \ldots, X_n)}{n}},$$

que és essencialment la part directa del teorema 1.

Ara demostrarem parcialment la part de suficiència del teorema 2 per a classes VC de conjunts mitjançant la transferència a processos empírics d'una propietat bàsica dels processos gaussians: *la desigualtat maximal en termes d'entropia mètrica.* Primer descriurem la versió més simple d'aquesta desigualtat. Sigui $\{G(t) : t \in T\}$ un procés gaussià centrat i definim $e_G^2(s, t) = \mathrm{E}(G(t) - G(s))^2$. Definim també el *nombre de cobriment* $N(T, e_G, \varepsilon)$ de l'espai (pseudo) mètric (T, e_G) com el mínim nombre de e_G-esferes de radi no més gran que ε necessari per a cobrir T. Aleshores hi ha una constant universal K tal que, per a tota versió separable de G, i $t_0 \in T$

$$(**) \qquad \mathrm{E}\sup_{t\in T}|G(t)| \leq K\int_0^\infty \sqrt{\log N(T, G, \varepsilon)}d\varepsilon + E|G(t_0)|.$$

Aquesta és una forma simplificada del teorema de Dudley, versió de Piser. Per a demostrar-la prenem, per a cada $k \in \mathbb{N}$, un conjunt de $N(2^{-k})$ punts de T, 2^{-k}-dens en (T, e_G), i n'assignem un a cada punt $t \in T$, posem per cas $\pi_k t$,amb la propietat que $e_G(\pi_k t, t) \leq 2^{-k}$. Aleshores suposant, sense pèrdua de generalitat, que T és finit i té diàmetre 1 i que $G(\pi_0 t) = 0$, tenim

$$\mathrm{E}\sup_T|G| \leq \sum \mathrm{E}\sup_T|G(\pi_{k-1}t) - G(\pi_k t)|$$

i podem aplicar la fita (*), tot observant que, per a cada k, el suprem és sobre no més de $N(2^{-k})^2$ termes, cada un amb desviació estàndard no més gran que $3 \cdot 2^{-k}$ (per la desigualtat triangular ja que $\pi_k t$ i $\pi_{k-1} t$ són tots dos prop de t). La sèrie numèrica que en resulta és equivalent a la condició integral.

Per aplicar (**) a classes VC necessitem una altra propietat remarcable d'aquestes classes de conjunts (Dudley, 1978): si $\mathcal{C}$ és una classe VC aleshores existeixen constants positives i finites c_1 i c_2 tals que, per a tota mesura de probabilitat Q sobre S,

$$N(\mathcal{C}, L_2(Q), \varepsilon) \leq c_1 \varepsilon^{-c_2}, \ \varepsilon < \frac{1}{2}.$$

Afegint aquestes dues desigualtats en un dels càlculs anteriors, tenim:

$$\begin{aligned} \mathrm{E} n^{\frac{1}{2}} \|\mathrm{P}_n - \mathrm{P}\|_{\mathcal{C}} &\leq K \mathrm{E}_X \mathrm{E}_g \|\frac{1}{n^{\frac{1}{2}}} \sum_{i=1}^{n} g_i \delta_{X_i}\|_{\mathcal{C}} \\ &\leq K \mathrm{E}_X \int_0^1 \sqrt{\log N(T, L_2(\mathrm{P}_n), \varepsilon)} \mathrm{d}\varepsilon + 1 \\ &\leq K \int_0^1 \sqrt{\log \frac{1}{\varepsilon}} \mathrm{d}\varepsilon + 1 < \infty \ \ \textit{uniformement en } n \end{aligned}$$

(on les K són constants universals no necessàriament iguals cada vegada que apareixen). Això ens dóna, doncs, la fitació estocàstica dels processos empírics, i no cal treballar gaire més per obtenir el TCL complet. La desigualtat de sobre és un exemple de desigualtat maximal per a processos empírics.

Tant el teorema 2 en tota la seva generalitat com la part de necessitat del teorema 1 es poden demostrar fent servir teoria de processos gaussians més refinada, càlculs més refinats, i una molt interessant desigualtat d'aleatorització gaussiana inversa que donarem més avall. De fet, com hem fet notar més amunt, l'aleatorització gaussiana no és estrictament necessària en les demostracions acabades de fer però és essencial en les parts inverses d'aquests teoremes.

Els teoremes descrits a sobre solucionen el problema (de l'LGN i/o del TCL) només en part: $\Delta^{\mathcal{C}}(X_1, \ldots, X_n)$ depèn de la mostra i de vegades és difícil de calcular. La situació encara és pitjor per a classes de funcions, on el lloc de $\Delta^{\mathcal{C}}$ el prenen els nombres de cobriment de $\mathcal{F}$ relatius a les distàncies $l_p(p_n)$ (els seus logaritmes s'anomemen *entropies empíriques*). Per tant, els resultats descrits a sobre, si bé descriuen admirablement la cancel·lació que fa treballar les coses, difícilment es pot dir que constitueixin *receptes* fàcils d'aplicar. De fet, són per a fer servir només quan els teoremes que donen condicions suficients de convergència més senzilles no funcionen. Donem ara, doncs, els dos criteris més importants d'aquest tipus (ordenats segons la freqüència amb què es fan servir).

Una classe de funcions $\mathcal{F}$ és VC-*subgràfica* si els subgràfics de les funcions de la classe formen una família VC de conjunts (subgràfic de f: $\{(x, t) \in S \times \mathbb{R} : 0 \leq t \leq f(x)$ o $f(x) \leq t \leq 0\}$). Els espais de funcions de dimensió finita són VC-subgràfics, i també ho és $\{q(C) I_C : C \in \mathcal{C}\}$ si $\mathcal{C}$ és VC.

TEOREMA 3. *Sota condicions de mesurabilitat, si $\mathcal{F}$ és VC-subgràfica aleshores:*

(*a*) (Giné i Zinn, 1984) $PF < \infty \Rightarrow \| P_n - P \|_{\mathcal{F}} \to 0$ g.s.
(*b*) (Alexander, 1987) *Les proporcions següents són equivalents, i són implicades per* $PF^2 < \infty$
(*i*) $\lim_{t\to\infty} t^2 P\{F < t\} = 0$ i *$\mathcal{F}$ és P-pregaussiana, i*
(*ii*) $\mathcal{F} \in TCL(P)$.

Tant (a) com la part de suficència de (b) també s'apliquen a altres classes de funcions relacionades amb la propietat VC, com les classes 'VC hull', etc.

Aquest resultat inclou el següent molt útil resultat de Pollard (1982):

COROL·LARI. *Sigui $\mathcal{F}$ una classe mesurable de funcions f tals que $|f| \leq F \in L_2(P)$, sigui $\| F \|_{2,Q} = (\int F^2 dQ)^{\frac{1}{2}}$, i sigui $D_{2,F}(x, \mathcal{F}) = \sup N(x, \frac{1}{\|F\|_{2,Q}} \cdot \mathcal{F}, L_2(Q))$, on el sup és sobre totes les probabilitats Q de suport finit; aleshores, si*

$$\int_0^1 (\log D_{2,F}(x,\mathcal{F}))^{\frac{1}{2}} dx < \infty,$$

la classe de funcions $\mathcal{F}$ és P-Donsker (és a dir, $\mathcal{F} \in TCL$(P)).

Si aquests resultats no s'apliquen a un problema donat, aleshores la millor manera de procedir és mirar si s'aplica un resultat d'*entropia mètrica per L_p-intervals* ('*bracketing*').

Definim

$$N^p_{[\,]}(\mathcal{F}, \mathrm{P}, \varepsilon), \ \varepsilon > 0,$$

com el mínim nombre de parelles de funcions mesurables f_i^L, f_i^U tals que

(i) per a tota $f \in \mathcal{F}$ hi ha una parella tal que $f_i^L \leq f \leq f_i^U$, i
(ii) $\mathrm{P}(f_i^U - f_i^L)^p \leq \varepsilon^p$.

$\log N^p_{[\,]}(\mathcal{F}, \mathrm{P}, \varepsilon)$ s'anomena l'entropia mètrica per L_p-intervals de $\mathcal{F}$.

Per exemple, si P és la mesura de Lebesgue del quadrat unitat de $\mathbb{R}^2$ i $\mathcal{C}$ és la família dels conjunts convexos tancats, aleshores $\log N^P_{[\,]}(\mathcal{F}, \mathrm{P}, \varepsilon)$ és de l'ordre de ε^{-1} (Bronštein, 1976). Dudley (1984) calcula (o estima) l'entropia per L_p-intervals de classes de funcions diferenciables i de classes de conjunts amb fronteres diferenciables.

TEOREMA 4
(i) (Blum, 1955; Dehardt, 1971) Si $N^1_{[\,]}(\mathcal{F}, \mathrm{P}, \varepsilon) < \infty$ *per a tot* $\varepsilon > 0$ *aleshores* $\| P_n - P \|_{\mathcal{F}} \to 0$ g.s.
(ii) (Ossiander, 1987) *Si* $\int_0^\infty \sqrt{\log N^2_{[\,]}(\mathcal{F}, \mathrm{P}, \varepsilon)}\, d\varepsilon < \infty$, *aleshores* $\mathcal{F} \in TCL$(P).

Andersen et al., 1988, conté un refinament immillorable del teorema d'Ossiander on la magnitud dels 'segments' és mesurada per la distància L_2 feble, i la seva cardinalitat controlada per 'mesures majorants'. L'estimació de l'entropia per L_2-segments es basa en el control de

$$\mathrm{E} \sup_{f: \mathrm{E}(f-g)^2 \leq \delta} (f-g)^2 (X),$$

que no té per què ser fàcil, però que és molt més senzill que el problema original. Més avall considerem un exemple on es fan servir els teoremes 3 i 4.

Encara hi ha una altra 'recepta' que potser és la que cal provar primer: si el suport de P està contingut dins l'esfera unitat d'un espai de Banach de tipus 2 o de cotipus 2 i si $\mathcal{F}$ és un conjunt fitat del dual, aleshores es pot aplicar la teoria del teorema central del límit sobre espais de Banach separables. Així s'obtenen, per exemple, els dos resultats següents:

TEOREMA 5

(i) (Borisov, 1981; Dudley i Durst, 1980) *Si* P *és discreta aleshores la família de tots els subconjunts de S sempre satisfà la llei dels grans nombres, i satisfà el teorema central del límit si i només si*

$$\sum p_i^{\frac{1}{2}} < \infty$$

(ii) (Giné i Zinn, 1986b) *La classe* $\mathcal{F}$ *de totes les funcions reals* f *tals que* $\|f\|_\infty \leq 1$ *i* $\|f\|_{\mathrm{Lip}} \leq 1$ *satisfà la llei dels grans nombres per a tota P, i satisfà el TCL per a* P *si i només si* $\sum \mathrm{P}\{i \leq |x| < i+1\}^{\frac{1}{2}} < \infty$. *Suprimint la condició de fitació, la condició per al TCL esdevé* $\sum \mathrm{P}\{|x| > i\}^{\frac{1}{2}} < \infty$.

El que hem fet fins ara constitueix una visió molt parcial del tipus de resultats que hom pot trobar dintre de la teoria de processos empírics. Hi ha molts temes que no s'han ni esmentat, en particular, desigualtats exponencials, classes universalment i uniformement Donsker, U-processos, etc. però ens aturem aquí a fi de comentar, en el que resta, la utilitat d'aquesta teoria i el *bootstrap*.

Els teoremes anteriors s'apliquen d'una manera molt útil i senzilla a la construcció de regions de confiança: si $\mathcal{F}$ captura propietats d'interès d'una distribució, i si és P-*Donsker* per a un conjunt prou gran de P (potser tota P, o tota P tal que $\mathrm{P}F^2 < \infty$), pot ser interessant construir regions de confiança per a P de la forma $\|\mathrm{P}_n - \mathrm{P}\|_{\mathcal{F}} \leq \delta n^{-\frac{1}{2}}$. Per exemple, Beran i Millar, 1986, prenen per $\mathcal{F}$ la classe dels semiespais de $\mathbb{R}^d$. Els cal afrontar dos problemes: un és la computabilitat de la norma $\|\mathrm{P}_n - \mathrm{P}\|_{\mathcal{F}}$ i l'altre és que la llei límit de $n^{\frac{1}{2}} \|\mathrm{P}_n - \mathrm{P}\|_{\mathcal{F}}$, és a dir la llei de $\|G_P\|$, depèn (en aquest cas, i molt sovint) de P, que és desconeguda. La primera dificultat es pot atacar cercant el màxim per procediments aleatoris (*random search*), cosa que Millar està estudiant. I hom pot eludir la segona dificultat mitjançant el *bootstrap*, el qual serà considerat a la darrera secció.

Els teoremes de límit anteriors també són útils en conjunció amb el «*delta method*»: $\theta(\mathrm{P})$ podria ser una funció sobre mesures de probabilitat Fréchet (o Hadamard) diferenciable en $\ell^\infty(\mathcal{F})$; e.g.

$$\theta(\mathrm{Q}) - \theta(\mathrm{P}) = \int f_P \mathrm{d}(\mathrm{Q}-\mathrm{P}) + o(\|\mathrm{Q}-\mathrm{P}\|_F).$$

Aleshores, si f_P és de quadrat P integrable i si $\mathcal{F}$ és P-Donsker (amb una mica menys n'hi ha prou), es té

$$n^{\frac{1}{2}}(\theta(\mathrm{P}_n) - \theta(\mathrm{P})) \to_d N(0, \mathrm{Var}_P f_P.$$

El *bootstrap* de processos empírics commuta amb la diferenciació (com va ser observat per Bickel i Freedman, 1981, i com és fàcil de veure).

I, certament, hom té les desigualtats maximals, que s'estan aplicant amb molt d'èxit tant en l'estimació de funcions com en la teoria asimptòtica d'estadístics complicats.

Wellner, 1992, descriu unes quantes aplicacions dels processos empírics en estadística i, en un congrés recent, D. Nolan passà una llista de més de 90 referències dels últims deu anys on es fa servir alguna part de la teoria acabada d'exposar.

A les dues seccions que resten veurem un exemple d'aplicació dels resultats acabats de descriure (el límit de la mediana simplicial empírica) i també considerarem el *bootstrap* de processos empírics.

Acabarem però aquesta secció amb la desigualtat d'aleatorització normal esmentada a sobre. Aquest resultat és una petita modificació d'un resultat de Piser (Giné i Zinn, 1984), i la demostració que en segueix és nova. Com hem vist a sobre, per Jensen, és ben clar que $\|\sum g_i \delta_{X_i}\|$ és estocàsticament més gran que $\|\sum \varepsilon_i \delta_{X_i}\|$. El que és sorprenent és que aquestes variables són pràcticament de la mateixa mida. En lloc de normals, prendrem multiplicadors bastant més generals.

PROPOSICIÓ 6. (Pisier, comunicació personal: Giné i Zinn, 1984). *Siguin Y, Y_i, $i \in \mathbb{N}$, variables aleatòries i.i.d. prenent valors en un espai de Banach i siguin ξ, ξ_i, $i \in \mathbb{N}$, variables aleatòries reals simètriques, i.i.d., independents de les Y_i. Suposem $E\|Y_i\| < \infty$ i $\Lambda_{2,1} := \int_0^\infty (\Pr\{|\xi| > u\})^{\frac{1}{2}} \mathrm{d}u < \infty$. Aleshores es té que per a tot $n \in \mathbb{N}$ i tot $n_0 \leq n$,*

$$\mathrm{E}\|\frac{1}{n^{\frac{1}{2}}}\sum_{i=1}^{n}\xi_i Y_i\| \leq n_0 E\|Y\|\mathrm{E}\Big(\frac{\max_{k\leq n}|\xi_i|}{n^{\frac{1}{2}}}\Big) + \Lambda_{2,1}(\xi)\max_{n_0<k\leq n}\mathrm{E}\|\frac{1}{k^{\frac{1}{2}}}\sum_{n_0<i\leq k}\varepsilon_i Y_i\|.$$

(Canviant E per E, aquesta desigualtat també es verifica encara que Y no sigui mesurable.)*

DEMOSTRACIÓ. A fi que es pugui apreciar la simplicitat de la demostració, primer considerarem el cas $n_0 = 0$, que és la desigualtat que Pisier ens comunicà. Vet aquí la demostració a base de desigualtats que, per al lector que ha seguit aquest article fins aquí, s'expliquen soles:

$$
\begin{aligned}
\mathrm{E}\|\frac{1}{n^{\frac{1}{2}}}\sum_{i=1}^{n}\xi_i Y_i\| &= \mathrm{E}\|\frac{1}{n^{\frac{1}{2}}}\sum_{i=1}^{n}\varepsilon_i|\xi_i|Y_i\| \\
&= \mathrm{E}\|\frac{1}{n^{\frac{1}{2}}}\sum_{i=1}^{n}\Big(\int_0^\infty I(t \le |\xi_i|)\mathrm{d}t\Big)\varepsilon_i Y_i\| \\
&= \mathrm{E}\|\frac{1}{n^{\frac{1}{2}}}\int_0^\infty \Big(\sum_{i=1}^{n} I(t \le |\xi_i|)\varepsilon_i Y_i\Big)\mathrm{d}t\| \\
&\le \int_0^\infty \mathrm{E}\|\frac{1}{n^{\frac{1}{2}}}\sum_{i=1}^{n} I(t \le |\xi_i|)\varepsilon_i Y_i\|\mathrm{d}t \\
&= \int_0^\infty \mathrm{E}\|\frac{1}{n^{\frac{1}{2}}}\sum_{i=1}^{\#\{i\le n:\ |\xi_i|\ge t\}} \varepsilon_i Y_i\|\mathrm{d}t \\
&= \int_0^\infty \Big(\sum_{k=1}^{n}\Pr\{\sum_{i=1}^{n} I(|\xi_i| \ge t) = k\}\mathrm{E}\|\frac{1}{n^{\frac{1}{2}}}\sum_{i=1}^{k}\varepsilon_i Y_i\|\Big)\mathrm{d}t \\
&\le \Big(\frac{1}{n^{\frac{1}{2}}}\int_0^\infty \sum_{k=1}^{\infty}\sqrt{k}\Pr\{\sum_{i=1}^{n} I(|\xi_i| \ge t) = k\}\mathrm{d}t\Big)\max_{k\le n}\mathrm{E}\|\frac{1}{\sqrt{k}}\sum_{i=1}^{k}\varepsilon_i Y_i\| \\
&= \Big(\frac{1}{n^{\frac{1}{2}}}\int_0^\infty \mathrm{E}\sqrt{\sum_{i=1}^{n} I(|\xi_i| \ge t)\ \mathrm{d}t}\Big)\max_{k\le n}\mathrm{E}\|\frac{1}{\sqrt{k}}\sum_{i=1}^{k}\varepsilon_i Y_i\| \\
&\le \Big(\int_0^\infty \sqrt{\Pr\{|\xi| \ge t\}}\mathrm{d}t\Big)\max_{k\le n}\mathrm{E}\|\frac{1}{\sqrt{k}}\sum_{i=1}^{k}\varepsilon_i Y_i\|.
\end{aligned}
$$

En el cas $n_0 > 0$, continuem després de la sisena línea de la manera següent:

$$
\begin{aligned}
&\le \Big(\int_0^\infty \Pr\{\sum_{i=1}^{n} I(|\xi_i| \ge t) > 0\}\mathrm{d}t\Big)\max_{k\le n_0}\mathrm{E}\|\frac{1}{n^{\frac{1}{2}}}\sum_{i=1}^{k}\varepsilon_i Y_i\| \\
&+ \Big(\frac{1}{n^{\frac{1}{2}}}\int_0^\infty \sum_{k=n_0+1}^{\infty}\sqrt{k}\Pr\{\sum_{i=1}^{n} I(|\xi_i| \ge t) = k\}\mathrm{d}t\Big)\max_{n_0<k\le n}\mathrm{E}\|\frac{1}{\sqrt{k}}\sum_{i=1}^{k}\varepsilon_i Y_i\| \\
&\le \Big(\int_0^\infty \Pr\{\max_{i\le n}|\xi_i| \ge t\}\mathrm{d}t\Big)\frac{n_0}{n^{\frac{1}{2}}}\mathrm{E}\|Y\| \quad + \Lambda_{2,1}(\xi)\max_{n_0<k\le n}\mathrm{E}\|\frac{1}{\sqrt{k}}\sum_{i=1}^{k}\varepsilon_i Y_i\| \\
&= n_0\mathrm{E}\|Y\|\mathrm{E}\Big(\frac{\max_{k\le n}|\xi_i|}{n^{\frac{1}{2}}}\Big) + \Lambda_{2,1}(\xi)\max_{n_0<k\le n}\mathrm{E}\|\frac{1}{k^{\frac{1}{2}}}\sum_{n_0<i\le k}\varepsilon_i Y_i\|.
\end{aligned}
$$

Una aplicació: distribució asimptònica de la mediana simplicial empírica

Sigui P una probabilitat en el pla. La mediana simplicial de P és el punt (o conjunt de punts) amb més probabilitat de pertànyer al triangle determinat per tres observacions independents de P. Fórmules: donat $\theta \in \mathbb{R}^2$ (o $\mathbb{R}^d$, però prenem d = 2 per simplicitat d'exposició), sigui $C_\theta \subset \mathbb{R}^2 \times \mathbb{R}^2 \times \mathbb{R}^2$ el conjunt de les ternes de punts $(x_1, x_2, x_3) \in (\mathbb{R}^2)^3$ tals que $\theta \in \mathcal{S}(x_1, x_2, x_3)$, el triangle (obert) determinat pels punts x_1, x_2, x_3; aleshores la *mediana simplicial* de P es defineix com

$$\theta_0 (= \theta(\mathrm{P})) = \arg \text{màx}\ \mathrm{P}^3 C_\theta.$$

Aquesta definició és deguda a R. Liu, 1989, que fa servir símplexs tancats (no importa que els triangles siguin oberts o tancats si P dóna massa zero a qualsevol línea recta, com suposarem; però sí que hi ha diferència a la definició següent). Si X_i, $i \leq n$, són n observacions independents de P, la corresponent *mediana simplicial empírica* és qualsevol dels punts que pertanyen al màxim nombre de triangles simplicials $\mathcal{S}(X_i, X_j, X_k)$. És a dir, la mediana simplicial empírica és

$$\theta_n \in \{\arg \text{màx}\ D_n(\theta)\}$$

on

$$D_n(\theta) = U_n^{(3)}(C_\theta) = \frac{1}{\binom{n}{3}} \sum_{i<j<k\leq n} I_{C_\theta}(X_i, X_j, X_k).$$

(De fet, puix que considerem triangles oberts, $D_n(\theta) = \mathrm{P}_n^3 C_\theta$, però jo prefereixo la notació d'U-estadístiques.) Es pot escollir $\theta_n \in \{\arg \text{màx}\ D_n(\theta)\}$ de manera que θ_n sigui una variable aleatòria per a tot n, i així ho fem d'aquí endavant. El procés $\{D_n(\theta) : \theta \in \mathbb{R}^2\}$ s'anomena *procés de profunditat empírica*. Si P és angularment simètrica al voltant de θ_0 i té densitat no nul·la a θ_0, aleshores θ_0 és la seva mediana simplicial *única*; farem aquestes hipòtesis, i més, sobre P. Puix que θ és un paràmetre interessant de P (n'és un paràmetre de posició bastant robust), la pregunta a fer-se és: *com és de ben aproximada θ per θ_n?*

$D_n(\theta)$ no és un procés empíric sinó quelcom que es pot estudiar amb mètodes semblants: $D_n(\theta)$ és un U-*procés*. Per a cada θ, no és una suma de variables independents, sinó una U-estadística. Nolan i Pollard, 1987, 1988, i Arcones i Giné, 1991, han desenvolupat la teoria d'U-processos; en particular tenim resultats completament anàlegs als teoremes 1, 3 i 4. El corol·lari d'aquests resultats que és rellevant per a l'estudi de la mediana simplicial empírica és el següent:

PROPOSICIÓ 7. (Arcones i Giné, 1991). *Sota mesurabilitat, si $\mathcal{C}$ és una classe VC de conjunts mesurables de $\mathcal{S}^m$ aleshores*

(1) $\| U_n^{(m)}(s(C)) - \mathrm{P}^m(C) \|_{\mathcal{C}} \to 0$ *g.s.* *(LGN uniforme)*

(2) $n^{\frac{1}{2}}(U_n^{(m)}(s(C)) - \mathrm{P}^m(C)) \rightarrow_{\mathcal{C}}$ *un procés gaussià, en* $l^\infty(\mathcal{C})$

(3) $n^{\frac{k}{2}}(U_n^{(k)}(\pi_k s(C)) \rightarrow_{\mathcal{C}}$ *un procés de caos d'ordre k, en* $l^\infty(\mathcal{C})$ *per a* $1 \leq k \leq m$.

En aquest teorema s denota simetrització i les π_k són les projeccions de Hoeffding. En el nostre cas, C_θ és simètric i, per tant, s no s'aplica; projeccions de Hoeffding:

$$(\pi_k C)(x_1,\ldots,x_k) = (\delta_{x_1} - \mathrm{P}) \ldots (\delta_{x_k} - \mathrm{P})\, \mathrm{P}^{m-k} C,$$

e.g.

$$\begin{aligned}
\pi_1 C_\theta(x) &= (\delta_x - \mathrm{P})\mathrm{P}^2 C_\theta \\
&= (\mathrm{P}^2 C_\theta)(x) - \mathrm{P}^3 C_\theta \\
&= \int\int I_{C_\theta}(x_1, x_2, x)\mathrm{dP}(x_1)\mathrm{dP}(x_2) - \mathrm{P}^3 C_\theta.
\end{aligned}$$

La classe $\mathcal{C} = \{ C_\theta : \theta \in \mathbb{R}^2\}$ és VC: donats n ternes de punts $S_1,\ldots,S_n \in (\mathbb{R}^2)^3$ tenim, e.g.

$$C_\theta \cap \{S_1,\ldots,S_n\} = \{S_1, S_2\}$$
$$\textit{si i només si } \theta \in S_1 \cap S_2 \cap S_3^c \cap \ldots \cap S_n^c.$$

Per tant, $\Delta^{\mathcal{C}}(S_1,\ldots,S_n) \leq$ *el nombre més gran possible de regions que* $3n$ *línies determinen en el pla*, nombre que, per recurrència, es pot veure fàcilment que és

$$2 + 2 + 3 + \ldots + 3n = 1 + \frac{3n(3n+1)}{2}$$

(cada sumand correspon a l'addició d'una línia nova en el pla). Així doncs, $\mathcal{C}$ és VC i la proposició 7 s'aplica al procés de profunditat simplicial.

Comencem per descriure com emprar la llei dels grans nombres uniforme per a obtenir la consistència de la mediana simplicial empírica. Suposem que P té una mediana simplicial úica θ_0 i que dóna massa zero a les línies; aleshores,

◊ l'ajustament de P^3 implica $\lim_{|\theta| \to \infty} \mathrm{P}^3 C_\theta = 0$
◊ $\tau_n \to \theta \Rightarrow \limsup \mathrm{P}^3 C_{\tau_n} \leq \mathrm{P}^3 C_\theta$, és a dir, la funció $\theta \to \mathrm{P}^3 C_\theta$ és semicontínua superiorment.

(Noten $\mathrm{P}^3 \partial C_\theta = 0$ per a demostrar, com un execici, la segona afirmació). Aquestes dues observacions elementals, juntament amb la unicitat, impliquen *identificabilitat*, és a dir,

$$\delta\varepsilon := \mathrm{P}^3 C_{\theta_0} - \sup_{|\theta - \theta_0| > \varepsilon} \mathrm{P}^3 C_\theta > 0 \textit{ per a tot } \varepsilon > 0.$$

Això, juntament amb l'LGN per a la classe $\mathcal{C}$, dóna

$$\Pr\{\sup_{k\geq n}|\theta_k-\theta_0|>\varepsilon\}\leq \Pr\{\sup_{k\geq n}(\mathrm{P}^3C_{\theta_0}-\mathrm{P}^3C_{\theta_k})>\frac{\delta}{2}\}$$
$$=\Pr\Big\{\sup_{k\geq n}(\mathrm{P}^3C_{\theta_0}-D_k(\theta_0)+\underbrace{D_k(\theta_0)-D_k(\theta_k)}_{\leq 0}$$
$$+D_k(\theta_k)-\mathrm{P}^3(\theta_k))>\frac{\delta}{2}\Big\}$$
$$\leq \Pr\{2\sup_{k\geq n}\|\mathrm{P}^3C_\theta-D_k(\theta)\|>\frac{\delta}{2}\}\to 0.$$

Ara obtindrem la *velocitat de convergència* (i la normalitat asimptòtica) de θ_n mitjançant un procediment de Pollard, 1985, per a M-estimadors. Van de Geer, 1992, i Birgé i Massart, 1991, tenen altres maneres de fer servir la teoria de processos empírics per a obtenir velocitat de convergència d'estimadors.

Suposem $\theta_0=0$ i definim $U(\theta)=\mathrm{P}^3C_\theta$. Sota regularitat de $f:=\mathrm{dP}/\mathrm{d}x$ tenim

$$U(\theta)=U(0)-\frac{1}{2}\theta\cdot A\cdot\theta'+o(|\theta|^2)$$

per a θ petit, on $\mathcal{A}$ és simètrica i positiva definida. (En el nostre cas, sota suficient regularitat,

$$A=-\int_{C_0}[\Pi f]''(x_1,x_2,x_3)\mathrm{d}x_1\,\mathrm{d}x_2\,\mathrm{d}x_3$$

on $[\Pi f]''$ és la matriu de segones derivades de la funció $\Pi_{i=1}^3 f(x_i+\theta)$ respecte de θ a $\theta=0$; si $\mathrm{P}=N(0,I)$ aleshores $A=\frac{3}{2\pi}Id.$) Com que $\theta_n\to 0$ en pr. (de fet g.s.) existeix $c>0$ tal que

$$cn|\theta_n|^2\leq n(U(0)-U(\theta_n))+o_P(1)$$
$$\leq n(U_n^{(3)}-\mathrm{P}^3)(C_{\theta_n}-C_0)+\underbrace{n(D_n(0)-D_n(\theta_n))}_{\leq 0}+o_P(1).$$

La descomposició de Hoeffding ens dóna

$$n(U_n^{(3)}-\mathrm{P}^3)(C_{\theta_n}-C_0)=n(\mathrm{P}_n-\mathrm{P})(\mathrm{P}^2C_{\theta_n}-\mathrm{P}^2C_0)$$
$$+n\sum_{k=2}^{3}\binom{3}{k}U_n^{(k)}(\pi_k(C_{\theta_n}-C_0)),$$

i com que $\mathrm{P}^3(C_\theta-C_0)^2\to 0$ quan $\theta\to 0$, la proposició 7 implica (per la condició d'equicontinuïtat asimptòtica associada a aquests teoremes de límit)

$$\lim_{\delta\to 0}\limsup_{n\to\infty}\mathrm{E}\sup_{|\theta|<\delta}n^{\frac{k}{2}}|U_n^{(k)}(\pi_k(C_\theta-C_0))|=0$$

Amb això ens desempallegeum doncs dels termes d'ordre més alt i obtenim

$$cn|\theta_n|^2\leq n(\mathrm{P}n-\mathrm{P})(\mathrm{P}^2C_{\theta_n}-\mathrm{P}^2C_0)+o_P(1).$$

Suposem que P^2C_θ és suau (les dues integracions fan versemblant que ho sigui) en el sentit que existeixen «bones» Δ i r tals que

$$(*) \qquad (P^2C_\theta)(x) = (P^2C_0)(x) + \theta\cdot\Delta(x) + |\theta|\, r(x,\theta).$$

Aleshores,

$$cn\,|\theta_n|2 \le n^{\frac{1}{2}}|\theta_n|\left(|n^{\frac{1}{2}}(P_n-P)(\Delta)| + |n^{\frac{1}{2}}(P_n-P)(r(\cdot,\theta_n))|\right) + o_P(1).$$

De fet, si es verifiqués que

$$E|\Delta|^2<\infty \quad \text{i} \quad \sup_n \; E \sup_{|\theta|<\delta} |n^{\frac{1}{2}}(P_n-P)(r(\cdot,\theta))| < \infty$$

tindríem que

$$c(n^{\frac{1}{2}}|\theta_n|)^2 \le ((O_P(1))(n^{\frac{1}{2}}|\theta_n|) + o_P(1),$$

i això ens donaria

$$n^{\frac{1}{2}}|\theta_n| = O_P(1).$$

Si la classe $\{r(x,\theta) : |\theta| < \delta\}$ és P – *Donsker* i $Er^2(X,\theta) \to 0$ quan $\theta \to 0$, aquest argument es pot refinar i obtenir

$$n^{\frac{1}{2}}\theta_n \to_d N(0, A^{-1}(\mathrm{Cov}_P\Delta)A^{-1})$$

que és el que passa en aquest cas. De fet,

$$\Delta(x) = \int_{(R^2)^2} I_{C_0}(x_1,x_2,x)[\Pi f]'(x_1,x_2)\mathrm{d}x_1\mathrm{d}x_2$$

on $[\Pi f]'(x_1,x_2)$ és el vector de derivades parcials de la funció $f(x_1+\theta)f(x_2+\theta)$ respecte de θ a $\theta=0$. Si $P=N(0,I)$, $\Delta(x) = (\frac{2}{\pi^3})^{\frac{1}{2}}\frac{x}{|x|}$, $x\neq 0$, i el límit és $N(0,\frac{4}{\pi}I)$.

Resulta doncs que el problema es redueix a demostrar que la classe de funcions $\{r(x,\theta) : |\theta| < \delta\}$ és P – *Donsker*. Per simetria angular $P^2C_0(x) = \frac{1}{4}$ si $x\neq 0$ i, com que els triangles són oberts, $\theta_n \neq X_i$; aquestes dues observacions i la definició (*) de r donen, per a $x\neq 0$,

$$r(x,\theta) = \frac{1}{|\theta|}[P^2C_\theta(x) - P^2C_0(x-\theta) - \theta\cdot(\Delta(x-\theta) + (\Delta(x-\theta)-\Delta(x)))]$$

$$= \frac{1}{|\theta|}\int\!\!\int I_{C_0}(x_1,x_2,x-\theta)\big(\Pi_{i=1}^2 f(x_i+\theta) - \Pi_{i=1}^2 f(x_i) - \theta\cdot[\Pi f]'(x_1,x_2)\big)\mathrm{d}x_1\mathrm{d}x_2$$

$$+ \frac{\theta}{|\theta|}\cdot(\Delta(x-\theta)-\Delta(x)).$$

El segon sumand és una suma d'integrals de la forma

$$\frac{\theta_i}{|\theta|}\int\int [I_{C_0}(x_1,x_2,x-\theta)-I_{C_0}(x_1,x_2,x)]f_i(x_1)f(x_2)\mathrm{d}x_1\mathrm{d}x_2.$$

Amb x_1, x_2, θ fixes, el conjunt de x tals que $(x_1, x_2, x) \in C_0$ és el con amb vèrtex a θ i fronteres les semirectes $\{\lambda x_1 : \lambda \leq 0\}$, $\{\lambda x_2 : \lambda \leq 0\}$, i aquest cons formen una família VC de subconjunts de $\mathbb{R}^2$. Si suposem que f, $|f_i|$ i $\sup_{|\eta| \leq \delta} |f_{ij}(x+\eta)|$ (per a alguna $\delta > 0$) són integrables Riemann, aleshores resulta que, per a algun $M > 0$, les funcions $\frac{1}{M} r(\cdot, \theta)$ són dins la clausura puntual de l'embolcall convexa, simètrica del conjunt d'indicadors d'aquests cons. És sabut que això implica que la classe $\{r(x, \theta) : |\theta| < \delta\}$ és P-Donsker (Dudley, 1985).

Sota condicions no tan restrictives sobre f, f_i, f_{ij}, àdhuc sense l'existència de f_{ij}, encara es pot demostrar, aplicant el teorema d'Ossiander, que aquesta classe és P-Donsker. Això implica estimar

$$\mathrm{E}\sup_{|\theta|\leq\varepsilon} r^2(X,\theta)$$

i

$$\mathrm{E}\sup_{\theta':|\theta-\theta'|\leq\varepsilon^3} |r(x,\theta)-r(X,\theta')|^2,\ |\theta|>\varepsilon.$$

Resulta que aquestes quantitats són dominades per potències positives de ε i, per tant, que els nombres de cobriment per L_2-intervals són petits.

Aquesta secció és una ressenya d'Arcones, Chen i Giné, 1992.

El bootstrap

El test de Kolmogorov-Smirnov en dimensió més gran que u és difícil d'aplicar perquè la llei de $n^{\frac{1}{2}} \| F_n - F \|_\infty$ depèn de la distribució F. Això no és una excepció: normalment la distribució límit de $n^{\frac{1}{2}} \| P_n - P \|_{\mathcal{F}}$ depèn de P (que, normalment, és desconeguda). É el fet que, en dimensió u i per a $\mathcal{C} = \{(-\infty, x] : x \in \mathbb{R}\}$ i P contínua, no en depengui, que és excepcional. Per tant, el *bootstrap* hauria d'estendre molt considerablement l'aplicabilitat del TCL per a processos empírics. Bickel i Freedman, 1981, demostraren, fent servir mètodes que només s'apliquen satisfactòriament en $\mathbb{R}$ (representacions gairebé segures), que l'estadística de Kolmogorov-Smirnov en $\mathbb{R}$ es por «bootstrapejar». Gaenssler, 1985, estengué aquest resultat a classes VC de conjunts fent servir les excel·lents propietats de l'entropia mètrica d'aquestes classes. Giné i Zinn, 1990, demostraren que el TCL per a processos empírics *sempre* es pot bootstrapejar, almenys sota les condicions de mesurabilitat habituals. Els nostres resultats han estat estesos fa poc al *bootstrap* amb pesos intercanviables («*exchangeable*») i amb magnitud arbitrària (però tendent a infinit) de la mostra *bootstrap* per Præstgaard i Wellner, 1992, fent servir en part, els nostres mètodes. En aquesta secció comentaré els nostres resultats i els mètodes comuns, que poden ser d'interès en altres situacions. També consideraré breument el *bootstrap* paramètric i, més en general, el *bootstrap* basat en el model estadístic.

Com en seccions anteriors, tenim $(S, \mathcal{S}, \mathrm{P})$, un espai de probabilitat general, la mostra $\{X_i\}$, i, sempre que els necessitem, multiplicadors independents de la mostra, Rademacher $\{\varepsilon_i\}$, normal estàndard $\{g_i\}$ i Poisson amb paràmetre $\frac{1}{2}$, $\{N_i\}$, o Poisson simetritzats, $\{N_i - N_i' = \tilde{N}\}$, (successions i.i.d. totes elles)

Siguin $X^*_{n_i}$, $i = 1, \ldots, n$, condicionalment i.i.d. donada la mostra, amb llei condicional

$$\Pr{}^*\{X^*_{ni} = X_j\} = \frac{1}{n}.$$

(El superindex * sobre Pr o E denotarà Pr o E condicionals donada la mostra –i no probabilitat o integral exteriors com a les seccions anteriors.) Siguin P^*_n les mesures empíriques *bootstrap*

$$\mathrm{P}^*_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{X^*_{ni}}, \quad n \in I\!N.$$

Direm que *el bootstrap funciona g.s. per a $\mathcal{F}$ i* P si

$$n^{\frac{1}{2}}(\mathrm{P}^*_n - \mathrm{P}_n) \to_{\mathcal{L}^*} G_P \quad en\ \ell^\infty(\mathcal{F}), \quad g.s.$$

Si això es verifica aleshores es té que, per a tot funcional mesurable i continu H sobre $l^\infty(\mathcal{F})$,

$$H(n^{\frac{1}{2}}(\mathrm{P}^*_n - \mathrm{P}_n)) \to_{d^*} H(G_P) \quad g.s.$$

Sovint només es necessita el *bootstrap* en probabilitat. La convergència en llei en $l^\infty(\mathcal{F})$ es pot metritzar per e.g. la distància dual-fitada-Lipschitz («*dual-bounded-Lipschitz*»), d_{BL^*}, i, per tant, direm que *el bootstrap funciona en probabilitat* si

$$d^*_{BL^*}(n^{\frac{1}{2}}(\mathrm{P}^*_n - \mathrm{P}_n), G_P) \to 0 \quad en\ pr.$$

($d^*_{BL^*}(n^{\frac{1}{2}}(\mathrm{P}^*_n - \mathrm{P}_n), G_P) := \sup\{\, |E^* f(n^{\frac{1}{2}}(\mathrm{P}^*_n - \mathrm{P}_n)) - Ef(G_P)| : \|f\|_\infty \leq 1,\ \|f\|_{\mathrm{Lip}} \leq 1\}$, on $f\colon \mathcal{F} \to \mathbb{R}$.) Si això es verifica aleshores es té que per a tot funcional mesurable i continu H sobre $l^\infty(\mathcal{F})$,

$$\|F_{H(n^{\frac{1}{2}}(\mathrm{P}^*_n - \mathrm{P}_n))}(x) - F_{H(G_P)}(x)\|_\infty \to 0 \quad en\ pr.$$

(suposant $F_{H(G_P)}(x)$ contínua, altrament només es té aquest límit per a qualsevol distància que metritzi la convergència en distribució en $\mathbb{R}$). I també es té el *bootstrap* per a qualsevol funcional diferenciable (F o H) per $\|\cdot\|_{\mathcal{F}}$ a P. Per tant, *el bootstrap del procés empíric dóna automàticament el bootstrap d'una gran munió d'estadístiques*, i això el fa molt útil i atractiu.

La nostra demostració es basa en el *bootstrap* de les desigualtats maximals necessàries i suficients per als teoremes de límit i, per tant, els nostre mètodes també donen el *bootstrap* d'estimadors el comportament asimptòtic dels quals és conseqüéncia d'aquest tipus de desigualtats, com per exemple, d'M-estimadors força generals (Romo, inèdit; Arcones i Giné, 1992). Tanmateix, almenys ara, no sé si el teorema de límit de la mediana simplicial empírica es bootstrapeja (si bé sospito que sí) ja que encara resten alguns problemes a resoldre sobre el *bootstrap* d'U-processos degenerats –la mediana d'Oja, més suau, es bootstrapeja.

Formalment, aquest és el teorema:

TEOREMA 7. (Giné i Zinn, 1990). *Sota les hipòtesis de mesurabilitat habituals,*

(a) *el* bootstrap *per a $\mathcal{F}$ i P funciona en probabilitat si i només si $\mathcal{F}$ és* P-*Donsker, i*
(b) *el* bootstrap *per a $\mathcal{F}$ i P funciona g.s. si i només si $\mathcal{F}$ és* P-*Donsker i* $\mathrm{P}F^2 < \infty$.

(Contràriament al que s'esdevé en dimensió finita, es pot tenir el TCL en dimensió infinita i al mateix temps $\mathrm{P}F^2 = \infty$; el que és necessari, sobre integrabilitat, és $t^2\,\mathrm{Pr}\{\mathcal{F} > t\} \to 0$.)

Si el procés empíric *bootstrap* convergeix, ha de convergir per força a G_P a causa de TCL *bootstrap* en dimensió finita. L'únic problema, doncs, és l'ajustament uniforme condicional (p.e. les condicions d'equicontinuïtat asimptòtica condicionals). En altres paraules, cal que relacionen les quantitats

$$\mathrm{E}^* \sup_{\substack{f,g\in\mathcal{F}\\ \mathrm{P}(f-g)^2\le\delta}} |n^{\frac{1}{2}}(\mathrm{P}_n^* - \mathrm{P}_n)(f-g)|$$

i

$$\mathrm{E} \sup_{\substack{f,g\in\mathcal{F}\\ \mathrm{P}(f-g)^2\le\delta}} |n^{\frac{1}{2}}(\mathrm{P}_n - \mathrm{P})(f-g)|.$$

La segona hauria de controlar la primera en pr. o g.s. i viceversa. Simplifiquem la notació denotant aquests sups per símbols de norma genèrics.

El *bootstrap* és una suma amb pesos multinominals $(n; \frac{1}{n}, \ldots, \frac{1}{n})$, o sigui, $n^{-\frac{1}{2}}\sum_{i=1}^n M_{ni}\delta x_i$, i asimptòticament s'han de poder substituir les multinomials per variables de Pisson independents equidistribuïdes. Per tant, hauríem de poder comparar

$$\mathrm{E}^*\|\frac{1}{n^{\frac{1}{2}}}\sum_{i=1}^n \varepsilon_i\delta_{X^*_{ni}}\|$$

i

$$\mathrm{E}_N\|\frac{1}{n^{\frac{1}{2}}}\sum_{i=1}^n \tilde{N}_i\delta_{X_i}\|.$$

Per les desigualtats de simetrització estàndard, les primeres quantitats controlen i són controlades pel TCL *bootstrap*. Per la proposició 7, la segona quantitat és controlada, en esperança, pel TCL (encara que les variables de Poisson siguin estocàsticament més grans que les de Rademacher), i la seva esperança òbviament controla el TCL. La comparació d'aquestes dues quantitats ens donarà, donc, el *bootstrap* en probabilitat. En canvi, per al *bootstrap* g.s., necessitarem el control g.s. (i no en esperança) de les segones variables aleatòries. En termes més concrets, el teorema es dedueix de les tres proposicions següents:

A. *Desigualtats bàsiques*:

$$\frac{e-1}{\sqrt{2}e}\mathrm{E}_\epsilon\|\sum_{i=1}^{n}\varepsilon_i\delta_{X_i}\| \le \mathrm{E}^*\|\sum_{i=1}^{n}\varepsilon_i\delta_{X^*_{ni}}\| \le \frac{e}{e-1}\mathrm{E}_N\|\sum_{i=1}^{n}\tilde{N}_i\delta_{X_i}\|.$$

B. *La desigualtat dels multiplicadors: Per a tot* $n \in \mathbb{N}$ *i tot* $n_0 \le n$,

$$\mathrm{E}\|\frac{1}{n^{\frac{1}{2}}}\sum_{i=1}^{n}\tilde{N}_i\delta_{X_i}\| \le n_0(\mathrm{P}F)\mathrm{E}\Big(\frac{\max_{k\le n}|\tilde{N}_k|}{n^{\frac{1}{2}}}\Big) + M\max_{n_0<k\le n}\mathrm{E}\|\frac{1}{k^{\frac{1}{2}}}\sum_{n_0\le i\le k}\varepsilon_i\delta_{X_i}\|$$

on $M = \int_0^\infty(\Pr\{|\tilde{N}|>u\})^{\frac{1}{2}}du$.

C. *Fita g.s. per a sumes amb multiplicadors: Si* $\mathrm{P}F^2 < \infty$

$$\limsup_{n\to\infty}\mathrm{E}_N\|\frac{1}{n^{\frac{1}{2}}}\sum_{i=1}^{n}\tilde{N}_i\delta_{X_i}\| \le 4\limsup_{n\to\infty}\mathrm{E}\|\frac{1}{n^{\frac{1}{2}}}\sum_{i=1}^{n}\tilde{N}_i\delta_{X_i}\| \quad a.s.$$

A i B demostren que el *bootstrap* en pr. es verifica per a $\mathcal{F}$ i P si i només si $\mathcal{F}$ és P-Donsker. Amb C, obtenim el resultat sobre el *bootstrap* g.s. (llevat de la necessitat de $\mathrm{P}F^2 < \infty$ que és fàcil d'establir i que es verifica àdhuc a $\mathbb{R}$).

Aquestes desigualtats es poden enunciar amb més generalitat per a elements aleatoris en espais de Banach i, la segona i la tercera, per a multiplicadors no necessàriament Poisson. La segona, generalitzada, és simplement la proposició 7, que Zinn i jo hem fet servir repetidament, per a multiplicadors gaussians, des de 1984: és crucial per a demostrar la part de necessitat dels teoremes de límit descrits a la Secció 3.

La desigualtat C és trivial en $\mathbb{R}$: si Y_i són variables reals i.i.d., centrades i amb segon moment finit, i si les ξ_i són reals i.i.d. simètriques (només que siguin centrades n'hi ha prou), amb segon moment 1, i independents de les Y_i, aleshores

$$\mathrm{E}_\xi\Big[\frac{1}{n^{\frac{1}{2}}}\sum_{i=1}^{n}\xi_iY_i\Big]^2 = \frac{1}{n}\sum_{i=1}^{n}Y_i^2 \to \mathrm{E}Y_1^2.$$

En canvi, en dimensió infinita, C és una desigualtat força més interessant i difícil. Curiosament, els seus autors, Ledoux, Talagrand i Zinn, no sabien res del *bootstrap* quan l'obtingueren i de fet, abans que Zinn i jo la féssim servir, Zinn preguntà a més d'un estadístic si aquesta desigualtat podia ser útil per a alguna cosa (amb respostes sempre negatives!). Ref.: Ledoux i Talagrand, 1988.

La desigualtat B, més precisament la més general de la proposició 7 (per a $n_0 = 0$), té a veure amb un problema teòric de probabilitats en espais de Banach que també és trivial en $\mathbb{R}$: quines ξ reals i simètriques verifiquen que si un vector aleatori X a valors en un Banach satisfà el TCL i és independent de ξ, aleshores també ξX satisfà el

TCL? En $\mathbb{R}$ la resposta és clara: ξ satisfà aquesta propietat per a l'espai $\mathbb{R}$ si i només si $E\xi^2 < \infty$. Pisier n'obtingué una condició suficient (que la integral de la rel quadrada de la cua de la distribució de ξ sigui finita) probablement el 1976, inèdit, que m'explicà el 1977, i que fins al 1984 no férem servir -amb ξ normals; Zinn i jo estàvem tan segurs que aquesta desigualtat només s'utilitzaria amb multiplicadors normals que ni se'ns acudí de publicar-la amb multiplicadors generals, com a la proposició 7. Ledoux i Talagrand, 1986, demostraren que la condició de Pisier és, en general, la millor possible.

La desigualtat A (Giné i Zinn, 1990) també adapta tècniques inventades en el context d'espais de Banach. És interessant perquè dóna una relació quantitativa entre l'estadística original i la *bootstrap* (simetritzades, però aquest és només un punt tècnic irrellevant). Indicarem com obtenir la desigualtat de la dreta. Els conceptes resultaran més clars si la reenunciem en termes més abstractes.

Considerem

◊ punts fixos $x_1,., x_n$ d'un espai de Banach B (en el nostre cas, $B = l^\infty(\mathcal{F})$ i $x_i = \delta X_{i(\omega)}$);

◊ variables aleatòries $x_1^*,\ldots, x_n^*$ a valors en B, i.i.d. i tals que

$$\Pr\{x_j^* = x_i\} = \frac{1}{n}, \quad i = 1, \ldots, n$$

(en el nostre cas, $x_j^* = \delta X_{nj}^*$).

Demostrarem

$$E\| \sum \epsilon_j x_j^* \| \leq \frac{e}{e-1} E\| \sum \tilde{N}_k x_k \|.$$

L'argument principal s'inspira en un argument de poissonització de LeCam, 1970, que emprà per a demostrar que, en espais generals, l'ajustament uniforme de les lleis acompanyants de Poisson implica l'ajustament uniforme de les sumes (de vectors aleatoris independents).

Recordem el següent sobre *lleis de Poisson compostes*:

◊ Si μ és una mesura finita de massa total $|\mu|$, i $\in \mathbb{N}$, són i.i.d. amb llei $\frac{\mu}{|\mu|}$ i si $N(|\mu|)$ és Poisson amb paràmetre $|\mu|$, aleshores la llei de la variable aleatòria

$$\sum_{i=0}^{N(|\mu|)} Y_i$$

(amb $Y_0 = 0$) és

$$Pois\ \mu = e^{-|\mu|} \sum_{k=0}^{\infty} \frac{\mu^k}{k!}$$

(on les potències són preses en el sentit de convolució).

◊ Per les propietats de les funcions exponencials en àlgebres de Banach, $Pois\ (\mu + \nu) = (Pois\ \mu) * (Pois\ \nu)$, en particular, si Y_i són independents i per a cada i, Y_{ij}

són còpies independents de Yi, amb la convenció que $Y_{i0} = 0$, i si $N_i(1)$ són i.i.d. Poisson amb paràmetre 1, independents de les Y_{ij}, aleshores la llei de

$$\sum_{i=1}^{\infty} \sum_{j=0}^{N_i(1)} Y_{ij}$$

és

$$Pois \sum \mathcal{L}(Y_i).$$

La nostra modificació de l'argument de LeCam consisteix en el següent: (ometent l'argument (1) de N_i) Jensen i Fubini donen

$$\begin{aligned}(1 - e^{-1})\mathrm{E}\|\sum Y_i\| &= \mathrm{E}_Y\mathrm{E}_N\|\sum (N_i \wedge 1)Y_{i1}\| \\ &= \mathrm{E}_N\mathrm{E}_Y\|\sum (N_i \wedge 1)Y_{i1}\| \\ &\leq \mathrm{E}_N\mathrm{E}_Y\|\sum_i \sum_{j=0}^{N_i} Y_{ij}\| \\ &= \int \|x\| \mathrm{d}(Pois \sum \mathcal{L}(Y_i))(x).\end{aligned}$$

Tractem ara d'aplicar aquesta desigualtat al nostre problema:

$$Y_i = \varepsilon_i x_i^* \quad \text{i} \quad \mathcal{L}(Y_i) = \frac{1}{2n} \sum_{k=1}^{n} (\delta_{x_k} + \delta_{-x_k}),$$

d'on

$$\sum_{i=1}^{n} \mathcal{L}(Y_i) = \frac{1}{2} \sum_{k=1}^{n} (\delta_{x_k} + \delta_{-x_k}).$$

Per tant,

$$\begin{aligned}&Pois(\sum_{i=1}^{n} \mathcal{L}(Y_i)) = Pois[\frac{1}{2} \sum_{k=1}^{n} (\delta_{x_k} + \delta_{-x_k})] \\ &= Pois(\frac{1}{2}\delta_{x_1}) \star Pois(\frac{1}{2}\delta_{-x_1}) \star \ldots \star Pois(\frac{1}{2}\delta_{x_n}) \star Pois(\frac{1}{2}\delta_{-x_n}).\end{aligned}$$

I ara notem que

$$Pois(\frac{1}{2}\delta_x) = e^{-\frac{1}{2}} \sum_{k=0}^{\infty} \frac{(\frac{1}{2}\delta_x)^k}{k!} = e^{-\frac{1}{2}} \sum_{k=0}^{\infty} \frac{\delta_{kx}}{2^k k!},$$

que dóna massa $e^{-\frac{1}{2}} \frac{1}{2^k k!}$ a kx i, per tant, és la llei de la variable

$$N(\frac{1}{2})x.$$

És a dir,

$$Pois(\sum_{i=1}^{n}\mathcal{L}(Y_i)) = \mathcal{L}\Big[\sum_{k=1}^{n}(N_k(\frac{1}{2}) - N'_k(\frac{1}{2}))x_k\Big],$$

on les N i N' són independents i equidistribuïdes. Aleshores la desigualtat de poissonització es tradueix en

$$\mathrm{E}\|\sum \varepsilon_j x_j^*\| \leq \frac{e}{e-1}\mathrm{E}\|\sum \tilde{N}_k x_k\|,$$

que és el que volíem. [Vet aquí els dos elements clau en aquesta demostració: la desigualtat de poissonització i el fet obvi —però molt específic de les esponencials— que si $\sum \mu i = \sum \nu i$ aleshores $\Pi\, Pois(\mu_i) = \Pi\, Pois(\mu_i)$.]

Aquí m'agradaria cridar l'atenció sobre el fet curiós (però no tan excepcional) que la solució definitiva al problema del *bootstrap* per a processos empírics, un problema eminentment «aplicat», ha estat conseqüència de resultats de teoria de probabilitats de la més «abstracta», obtinguts no pas amb vista a les possibles aplicacions, sinó per imperatius de la pròpia teoria

Tornant al *bootstrap*: existeix un resultat igualment general per al *bootstrap paramètric o el bootstrap basat en el model estadístic?* Suposem que, donada la mostra, prenem mostres (*resampling*), no segons la distribuició empírica, P_n, sinó segons una «funció» seva, $\tau_n(\mathrm{P}_n)$. Sigui P_n^B la mesura empírica de la pseudo-mostra. Sota quines condicions encara tindrem, g.s. o en probabilitat, que

$$n^{\frac{1}{2}}(\mathrm{P}_n^B - \tau_n(\mathrm{P}_n)) \to_{\mathcal{L}^*} G_P \quad en\ \ell^\infty(\mathcal{F})?$$

Per exemple, $\tau_n(\mathrm{P}_n)$ podria ser una simetrització o una suavització de P_n o, si tenim un model paramètric i θ_n és un estimador de θ, $\tau_n(\mathrm{P}_n)$ podria ser P_{θ_n}.

En el teorema de *bootstrap* descrit més amunt la poissonització és molt important i està relacionada molt estretament amb l'estructura del procediment, és a dir, amb l'ús de multiplicadors multinomials. Però ara fem molt poques hipòtesis sobre l'estructura del procediment de remostreig. Tanmateix, existeix una resposta positiva al problema si la classe $\mathcal{F}$ no és gaire gran.

Suposem que la classe $\mathcal{F}$ satisfà les condicions següents:

(1) donades mesures de probabilitat R_n, R, sempre que

$$\mathrm{Cov}_{R_n}(f,g) \to \mathrm{Cov}_R(f,g)$$

uniformement en $f, g \in \mathcal{F}$, es té que

$$G_{R_n} \to_{\mathcal{L}} G_R$$

en $l^\infty(\mathcal{F})$;

(2) $n^{\frac{1}{2}}(\mathrm{P}_n - \mathrm{P}) \to_{\mathcal{L}} G_P$ en $l^\infty(\mathcal{F})$ *uniformement* en P.

Notació: denotem per $\mathcal{F}^2$ la classe de funcions $\{f, fg : f, g \in \mathcal{F}\}$. Aleshores, una simple desigualtat triangular dóna:

$$\|\tau_n(P_n) - P\|_{\mathcal{F}^2} \to 0 \;\; g.s. \; (o\ en\ pr.) \;\Rightarrow\; n^{\frac{1}{2}}(\mathrm{P}_n^B - \tau_n(\mathrm{P}_n)) \to_{\mathcal{L}} G_P.$$

[Si d metritza la convergència en llei, es té

$$d^*\left(n^{\frac{1}{2}}(\mathrm{P}_n^B - \tau_n(\mathrm{P}_n)), G_P\right) \leq d^*\left(n^{\frac{1}{2}}(\mathrm{P}_n^B - \tau_n(\mathrm{P}_n)), G_{\tau_n(\mathrm{P}_n)}\right) + d(G_P, G_{\tau_n(\mathrm{P}_n)});$$

com que P_n^B és la mesura empírica corresponent a P_n i el TCL és satisfet uniformement en P, per tant en P_n, el primer sumand tendeix a zero; i el segon sumand tendeix a zero g.s. —o en probabilitat— per (1) i la hipòtesi.]

Si τ és la identitat aquest argument dóna una demostració molt senzilla del TCL *bootstrap* (per a aquestes classes de funcions, no en general, és clar).

Com observen Sheehy i Wellner, 1990, si la distància de Hellinger entre $\tau_n(\mathrm{P}_n)$ i P tendeix a zero g.s. o en probabilitat) aleshores també $\| \tau_n(P_n) - P \|_{\mathcal{F}^2} \to 0$ g.s. o en probabilitat. Van Zwet anuncià l'agost de 1992, a les *Wald Lectures*, un resultat semblant, d'aplicació més ample, però molt més fluix.

Tornant al problema que ens ocupa, la pregunta pertinent és: quines classes de funcions satisfan (1) i (2)? La resposta és molt lluny del conjunt buit. Aquestes classes inclouen les classes VC-subgràfiques fitades i moltes més. De fet, són aquelles classes per a les quals tots els processos gaussians

$$\sum_{i=1}^{m} \alpha_i f(x_i) g_i, \quad f \in \mathcal{F},$$

amb $m < \infty$, g_i i.i.d. $N(0, 1)$, $x_i \in S$ i $\sum \alpha_i^2 = 1$, són uniformement «bons». Zinn i Giné, 1991, demostràrem que aquesta condició implica (1) i que és essencialment equivalent a (2). Anomenàrem aquestes classes *uniformement pregaussianes* o *uniformement Donsker*. També cal donar crèdit per a la introducció d'aquestes classes a Sheehy i Wellner, 1992; aquest autors n'han obtingut altres propietats i també tenen un treball sobre el TCL uniforme sobre famílies de probabilitats diferents del conjunt total. Prèviament, Dudley, 1987, introduí les *classes universalment Donsker*: Classes sobre les quals el procés empíric satisfà el TCL per a tota P. De fet, la majoria d'exemples en el seu article satisfan el TCL uniformement sobre P. Per acabar, és interessant de fer notar que, mentre hi ha una descripció gaussiana de les classes uniformement Donsker, no n'hi ha cap, per ara, de les classes universalment Donsker.

Referències

ALEXANDER, K. S. (1987). the central limit theorem for empirical processes on Vapnik-Cervonenkis classes. *Ann. Probability* 15, 178-203.

ANDERSEN, N. T., GINÉ, E.; OSSIANDER, M. I ZINN, J. (1988). The central limit theorem and the law of iterated logarithm for empirical processes under local conditions. *Prob. Th. Rel. Fields* 77, 271-305.

ARCONES, M. I GINÉ, E. (1991). Limit theorems for U-processes. *Ann. Probability,* pendent de publicació.

ARCONES, M. I GINÉ, E. (1992). On the bootstrap of M-estimators and other statistical functionals. A *Exploring the limits of bootstrap* (R. LePage i L. Billard eds.) pp. 13-48. Wiley, New York.

ARCONES, M.; CHEN, Z.; GINÉ, E. (1992). Estimators related to U-processes with applications to multivariate medians: asymptotic normality. Preprint.

BICKEL, P. J. I FREEDMAN, D. P. (1981). Some asymptotic theory for the bootstrap. *Ann. Statist.* 9, 1196-1216.

BIRGÉ, L. I MASSART, P. (1991). Rates of convergence for minimum contrast estimators. Report 140, Université Paris VI.

BORISOV, I. S. (1981). Some limit theorems for empirical distributions. Abstracts of Reports, Third Vilnius Conference on Probability and Math. Statist. I, 71-72. Vilnius.

BLUM, J. R. (1955). Ont the convergence of empiric distribution functions. *Ann. Math. statist.* 26, 527-729.

BRONSTEIN, E. M. (1976). ε-entropy of convex sets and functions. *Siberian Math. J.* 17, 393-398.

DEHARDT, J. (1971). Generalization of the Glivenko-Cantelli theorem. Ann. Math. Statist. 42, 2050-2055.

DUDLEY, R. M. (1978). Central limit theorem for empirical measures. *Ann. Probability.* 6, 899-929.

DUDLEY, R. M. (1979). Balls in $\mathbb{R}^k$ do not cut all subsets of $k+2$ points. *Adv. in Math.* 31, 306-308.

DUDLEY, R. M. (1984). A course on empirical processes. *Lecture Notes in Math.* 1097, 1-142. Springer, New York.

DUDLEY, R. M. (1985). An extended Wichura theorem, definitions of Donsker class, and weighted emprical functions. *Lect. Notes in Math.* 1153, 141-178. Springer, New York.

DUDLEY, R. M. (1987). Universal Donsker classes and metric entropy. *Ann. Probability.* 15, 1306-1326.

DUDLEY, R. M. I DURST, M. (1980). Empirical processes, Vapnik-Červonenkis classes and Poisson processes. *Prob. Math. Statist.* 1, 109-115.

GAENSSLER, P. (1987). Bootstrapping empirical measures indexed by Vapnik-Červonenkis classes of sets. A *Probability theory and Math. Statist.* 1, 467-481. VNU Science Press, Utrecht.

GINÉ, E. I ZINN, J. (1984). Some limit theorems for empirical processes. *Ann. Probability.* 12, 929--989.

GINÉ, E. I ZINN, J. (1986). Lectures on the central limit theorem for empirical processes. *Lecture Notes in Math.* 1221, 50-113. Springer, Berlín.

GINÉ, E. I ZINN, J. (1986b). Empirical processes indexed by Lipschitz functions. *Ann Probability*. 14, 1329-1338.

GINÉ, E. I ZINN, J. (1990). Bootstrapping general empirical measures. *Ann. Probability*. 18, 851-869.

GINÉ, E. I ZINN, J. (1991). Gaussian characterization of uniform Donsker clases of functions. *Ann. Probability*. 19, 758-782.

HOFFMANN-JØRGENSEN, J. (1984). *Stochastics processes on Polish spaces,* Aarhus Universitet, Matematisk Inst., Various Publication Series No. 39, 1991. Aarhus.

LE CAM, L. (1970). Remarques sur le théorème limite central dans les espaces localement convexes. A *Les Probabilités sur les Structures Algébriques*, Clermont-Ferrand 1969. Colloque CNRS, París, 233-249.

LEDOUX, M. I TALAGRAND, M. (1986). Conditions d'integrabilité pour les multiplicateurs dans le TLC banachique. *Ann. Probability*. 14, 916-921.

LEDOUX, M. I TALAGRAND, M. (1988). Un critère sur les petites boules dans le théorème limite central. *Prob. Theory Rel. Fields*. 77,29-47.

LEDOUX, M. I TALAGRAND, M. (1991). *Probability in Banach Spaces*. Springer-Verlag, New York.

LIU, R. V. (1990). On a notion of data depth based on random simplices. *Ann. Statistics*. 18, 404-414.

NOLAN, D. I POLLARD, D. (1987). U-Processes: rates of convergence. *Ann. Statist*. 15, 780-799.

NOLAN, D. I POLLARD, D. (1988). Functional limit theorems for U-processes. *Ann. Probability*. 16, 1291-1298.

OSSIANDER, M. (1987). A central limit theorem under metric entropy with L2 bracketing. *Ann. Probability*. 15, 897-919.

POLLARD, D. (1981). Limit theorems for empirical processes. *Zeits. Wahrsch. verw. Geb*. 57, 181-195.

POLLARD, D. (1982). A central limit theorem for empirical processes. *J. Australian Math. Soc., Series A*. 33, 235-248.

POLLARD, D. (1985). New ways to prove central limit theorems. *Econometric Theory*. 1, 295-314.

PRÆSTGAARD, J. I WELLNER, J. A. (1992). Exchangeably weighted bootstraps of the general empirical process. *Ann. Probability*, pendent de publicació.

ROMO, J. (1991). The bootstrap in probability of M-estimators. Preprint.

VAPNIK, V. N. I CERVONENKIS. A. JA. (1971). On the uniform convergence of relative frequencies of events to their probabilities. *Theory Prob. Appl*. 16, 164-280.

SAUER, N. (1972). On the density of families of sets. *J. Comb. Theory, A*. 13, 145-147.

SHEEHY, A. I WELLNER, J. A. (1992). Uniform Donsker classes of functions. *Ann. Probability*, pendent de publicació.

STENGLE, G. I YUKICH, J. E. (1989). Some new Vapnik-Cervonenkis classes. *Ann. Statist*. 17, 1441-1446.

TALAGRAND, M. (1988). Donsker classes of sets. *Prob. Theory Rel. Fields*. 78, 169-191.

VAN DE GEER, S. (1992). Hellinger consistency of certain non-parametric maximum likelihood estimators. *Ann. Statist.*, pendent de publicació.

VAPNIK, V. N. I CERVONENKIS. A. JA. (1981). Necessary and sufficient conditions for the convergence of means to theri expectation. *Theory Prob. Appl.* 26, 532-553.

WELLNER, J. A. (1992). Empirical processes in action: a review. *Statist. Review*, pendent de publicació.

University of Connecticut
Departaments of Mathematics
and Statistics
Storrs, CT 06269, USA