A Local Maximal Inequality under Uniform Entropy

Jon A. Wellner

University of Washington, Seattle

XiAn, China

July 9, 2011

IMS-China International Conference on Statistics and Probability

Based on joint work with: Aad van der Vaart

- Last day of this conference: July 11 (7/11, both primes).
- 2011 is the sum of 11 consecutive primes.
- 3 of the 11 primes are (consecutive!) twin primes; e.g. 3 & 5 or 11 & 13.
- Prove the twin prime conjecture! (There are infinitely many twin primes.)

Outline

- 1. The setting and basic problem
- 2. Available bounds: bracketing and uniform entropy
- 3. The new bound: uniform entropy
- 4. The perspective of a convex (or concave) function
- 5. Proof, part 1: concavity of the entropy integral
- 6. Proof, part 2: inversion
- 7. Generalizations
- 8. An application

1. The setting and basic problem

Suppose that:

- $X_1, \ldots, X_n$ are i.i.d. P on a measurable space $(\mathcal{X}, \mathcal{A})$.
- $\mathbb{P}_n = n^{-1} \sum_{i=1}^n \delta_{X_i}$ = the empirical measure.
- $\mathbb{G}_n \equiv \sqrt{n}(\mathbb{P}_n - P)$ = the empirical process.
- If $f : \mathcal{X} \to R$ is measurable,

$$\mathbb{P}_n(f) = n^{-1} \sum_{i=1}^n f(X_i), \qquad \mathbb{G}_n(f) = n^{-1/2} \sum_{i=1}^n (f(X_i) - Pf).$$

- When $\mathcal{F}$ is a given class of measurable functions f, it is useful to consider

$$\|\mathbb{G}_n\|_{\mathcal{F}} \equiv \sup_{f \in \mathcal{F}} |\mathbb{G}_n(f)|.$$

1. The setting and basic problem

Problem: Find useful bounds for the mean value

$$E_P^* \|\mathbb{G}_n\|_{\mathcal{F}}.$$

Entropy and two entropy integrals:

Uniform entropy: For $r \geq 1$

$$N(\epsilon, \mathcal{F}, L_r(Q)) = \left\{ \begin{array}{l} \text{minimal number of balls of radius } \epsilon \\ \text{needed to cover } \mathcal{F} \end{array} \right\},$$

F an envelope function for $\mathcal{F}$:

i.e. $|f(x)| \leq F(x)$ for all $f \in \mathcal{F}$, $x \in \mathcal{X}$;

$\|f\|_{Q,r} \equiv Q(|f|^r)^{1/r}$;

$$J(\delta, \mathcal{F}, L_r) \equiv \sup_Q \int_0^\delta \sqrt{1 + \log N(\epsilon \|F\|_{Q,r}, \mathcal{F}, L_r(Q))}\, d\epsilon.$$

1. The setting and basic problem

Bracketing entropy: For $r \geq 1$

$$N_{[\,]}(\epsilon, \mathcal{F}, L_r(P)) = \left\{ \begin{array}{l} \text{minimal number of brackets } [l,u] \\ \text{of } L_r(P)\text{-size } \epsilon \ \text{ needed to cover } \mathcal{F} \end{array} \right\};$$

$$[l,u] \equiv \{f: \ l(x) \leq f(x) \leq u(x) \ \text{ for all } x \in \mathcal{X}\};$$

$$\|u - l\|_{r,P} < \epsilon;$$

$$J_{[\,]}(\delta, \mathcal{F}, L_r(P)) \equiv \int_0^\delta \sqrt{1 + \log N_{[\,]}(\epsilon \|F\|_{r,P}, \mathcal{F}, L_r(P))}\, d\epsilon.$$

2. Available bounds: bracketing and uniform entropy

Basic bound, uniform entropy: (Pollard, 1990) Under some measurability assumptions,

$$E_P^* \|\mathbb{G}_n\|_{\mathcal{F}} \lesssim J(1, \mathcal{F}, L_2) \|F\|_{P,2}. \tag{1}$$

Basic bound, bracketing entropy: (Pollard)

$$E_P^* \|\mathbb{G}_n\|_{\mathcal{F}} \lesssim J_{[\,]}(1, \mathcal{F}, L_2(P)) \|F\|_{P,2}.$$

Small f bound, bracketing entropy: vdV & W (1996)
If $\|f\|_\infty \le 1$ and $Pf^2 \le \delta^2 PF^2$ for all $f \in \mathcal{F}$ and some $\delta \in (0,1)$, then

$$E_P^* \|\mathbb{G}_n\|_{\mathcal{F}} \lesssim J_{[\,]}(\delta, \mathcal{F}, L_2(P)) \|F\|_{P,2} \left(1 + \frac{J_{[\,]}(\delta, \mathcal{F}, L_2(P))}{\delta^2 \sqrt{n} \|F\|_{P,2}}\right).$$

3. The new bound: uniform entropy

Small f bound, uniform entropy?

Goal here:
provide a bound analogous to the "small f bound, bracketing entropy", but for uniform entropy.

Definition: The class of functions $\mathcal{F}$ is $P-$measurable if the map

$$(X_1, \ldots, X_n) \mapsto \sup_{f \in \mathcal{F}} \Big| \sum_{i=1}^{n} e_i f(X_i) \Big|$$

on the completion of the probability space $(\mathcal{X}^n, \mathcal{A}^n, P^n)$ is measurable, for every sequence $e_1, e_2, \ldots, e_n \in \{-1, 1\}$.

3. The new bound: uniform entropy

Theorem 1. Suppose that $\mathcal{F}$ is a $P-$measurable class of measurable functions with envelope function $F \leq 1$ and such that $\mathcal{F}^2$ is $P-$measurable. If $Pf^2 < \delta^2 P(F^2)$ for every f and some $\delta \in (0,1)$, then

$$E_P^* \|\mathbb{G}_n\|_{\mathcal{F}} \lesssim J(\delta, \mathcal{F}, L_2)\|F\|_{P,2}\left(1 + \frac{J(\delta, \mathcal{F}, L_2)}{\delta^2\sqrt{n}\|F\|_{P,2}}\right).$$

4. The perspective of a convex or concave function

Suppose that $f : \mathbb{R}^d \to \mathbb{R}$. Then the perspective of f is the function $g = g_f : \mathbb{R}^{d+1} \to \mathbb{R}$ defined by

$$g(x, t) = t f(x/t),$$

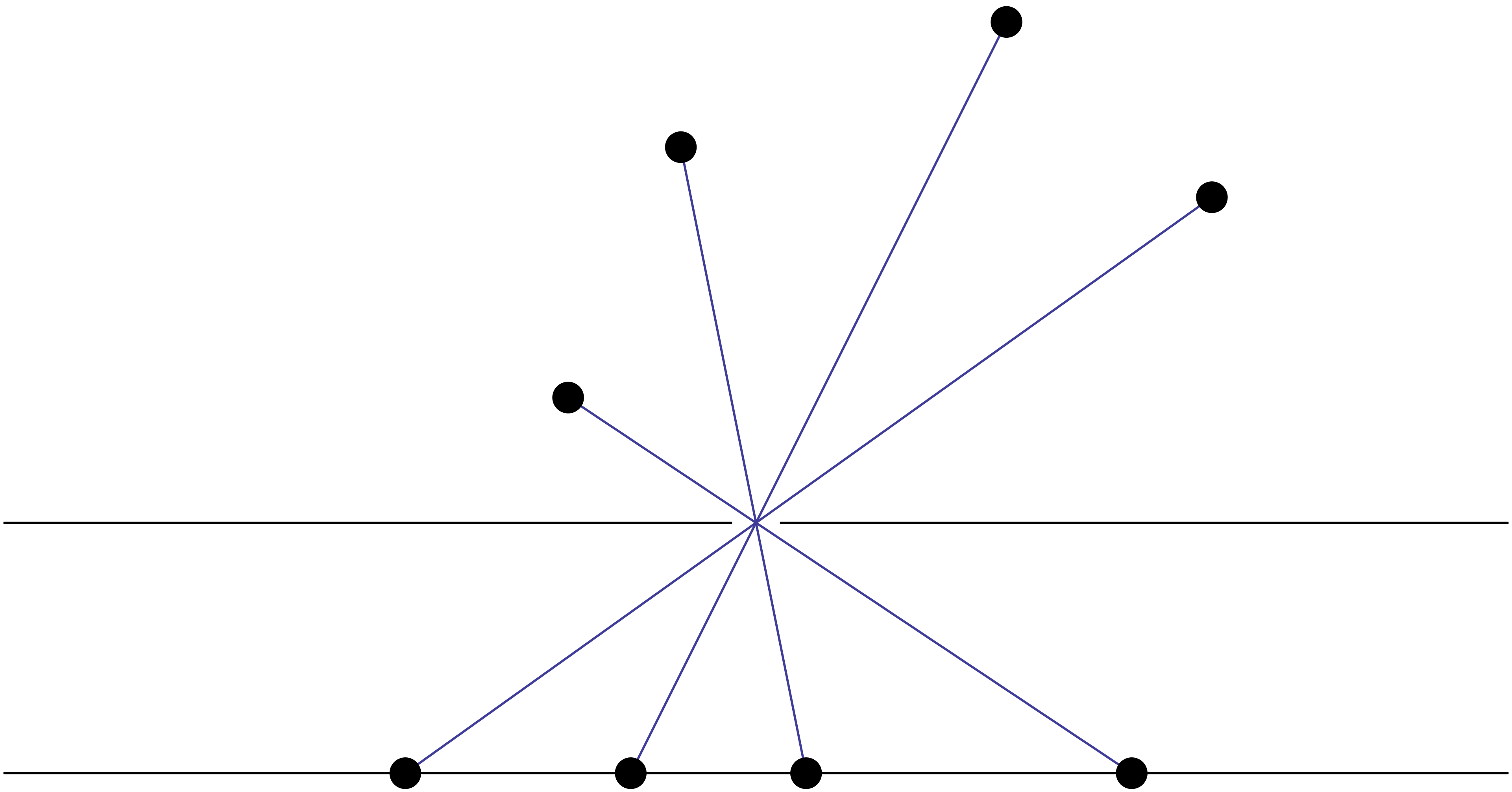
for $(x, t) \in \text{dom}(g) = \{(x, t) : x/t \in \text{dom}(f),\ t > 0\}$.
Then:

- If f is convex, then g is also convex.
- If f is concave, then g is also concave.

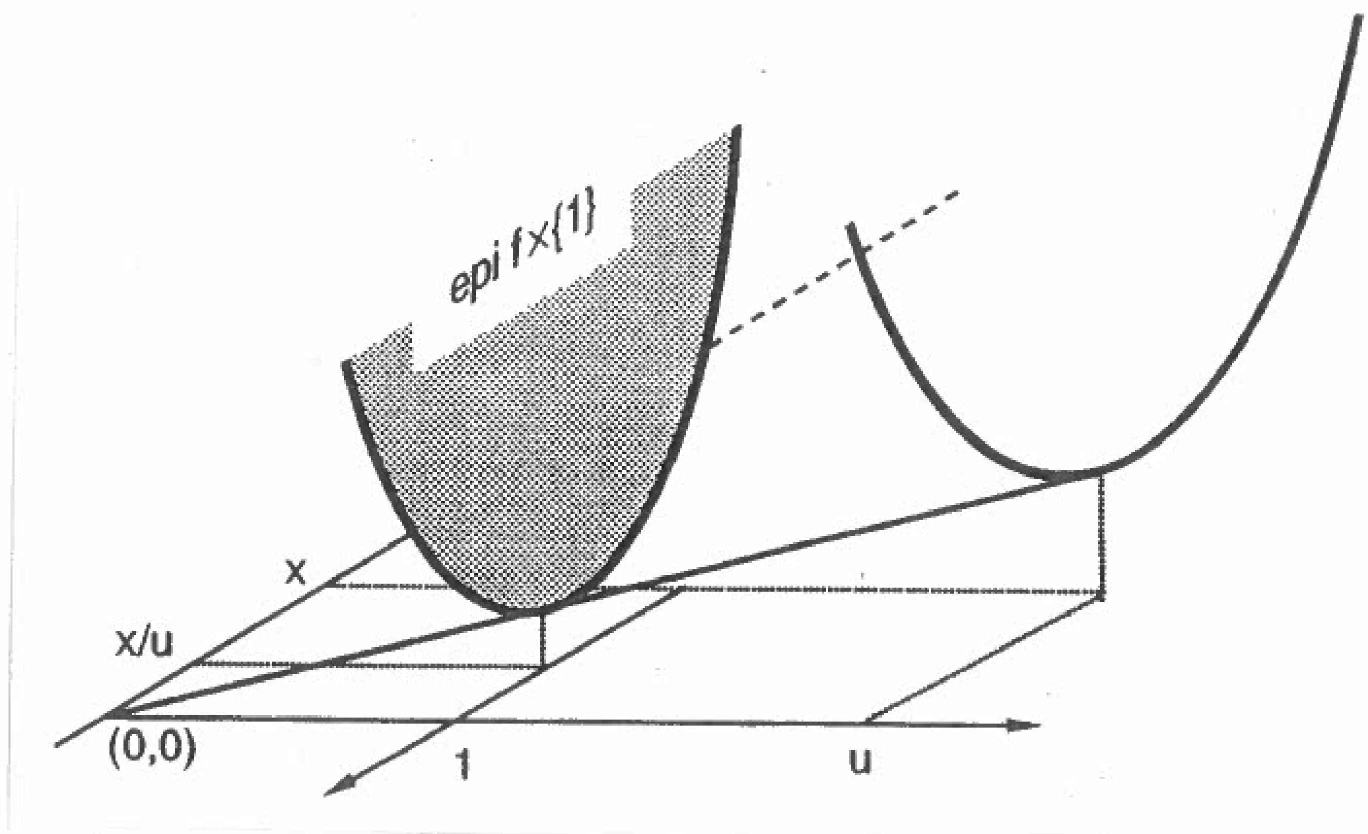
This seems to be due to Hiriart-Urruty and and Lemaréchal (1990), vol. 1, page 100; see also Boyd and Vandenberghe (2004), page 89.

Example: $f(x) = x^2$; then $g(x, t) = t(x/t)^2 = x^2/t$.

4. The perspective of a convex or concave function



4. The perspective of a convex or concave function



4. The perspective of a convex or concave function

Suppose that $h : \mathbb{R}^p \to \mathbb{R}$ and $g_i : \mathbb{R}^d \to \mathbb{R}$ for $i = 1, \ldots, p$. Then consider

$$f(x) = h(g_1(x), \ldots, g_p(x))$$

as a map from $\mathbb{R}^d$ to $\mathbb{R}$.

A preservation result:

- If h is concave and nondecreasing in each argument and $g_1, \ldots, g_d$ are all concave, then f is concave. See e.g. Boyd and Vandenberghe (2004), page 86.

5. Proof, part 1: concavity of the entropy integral

The proof begins much as in the proof of the easy bound (1); see e.g. van der Vaart and Wellner (1996), sections 2.5.1 and 2.14.1 and especially the fourth display on page 128, section 2.5.1: this argument yields

$$E_P^* \|\mathbb{G}_n\|_{\mathcal{F}} \lesssim E_P^* J\left(\frac{\sup_f (\mathbb{P}_n f^2)^{1/2}}{(\mathbb{P}_n F^2)^{1/2}}, \mathcal{F}, L_2\right)(\mathbb{P}_n F^2)^{1/2}. \qquad (2)$$

Since $\delta \mapsto J(\delta, \mathcal{F}, L_2)$ is the integral of a non-increasing nonnegative function, it is a concave function. Hence its perspective function

$$(x,t) \mapsto tJ(x/t, \mathcal{F}, L_2)$$

is a concave function of its two arguments. Furthermore, by the composition rule with $p = 2$, the function

$$(x,y) \mapsto \sqrt{y} J(\sqrt{x}/\sqrt{y}, \mathcal{F}, L_2)$$

is concave.

5. Proof, part 1: concavity of the entropy integral

Note that $E_P \mathbb{P}_n F^2 = \|F\|_{P,2}^2$. Therefore, by Jensen's inequality applied to the right side of (2) it follows that

$$E_P^* \|\mathbb{G}_n\|_{\mathcal{F}} \lesssim J\left(\frac{\{E_P^*(\sup_f \mathbb{P}_n f^2)\}^{1/2}}{\|F\|_{P,2}}, \mathcal{F}, L_2\right) \|F\|_{P,2}. \tag{3}$$

Now since $\mathbb{P}_n(f^2) = Pf^2 + n^{-1/2}\mathbb{G}_n f^2$ and $Pf^2 \le \delta^2 PF^2$ for all f, it follows, by using symmetrization, the contraction inequality for Rademacher random variables, de-symmetrization, and then (3), that

5. Proof, part 1: concavity of the entropy integral

$$
\begin{aligned}
E_P^*(\sup_f \mathbb{P}_n f^2) &\leq \delta^2 \|F\|_{P,2}^2 + \frac{1}{\sqrt{n}} E_P * \|\mathbb{G}_n\|_{\mathcal{F}^2} \\
&\leq \delta^2 \|F\|_{P,2}^2 + \frac{2}{\sqrt{n}} E_P * \|\mathbb{G}_n^0\|_{\mathcal{F}^2} \\
&\leq \delta^2 \|F\|_{P,2}^2 + \frac{4}{\sqrt{n}} E_P^* \|\mathbb{G}_n^0\|_{\mathcal{F}} \\
&\leq \delta^2 \|F\|_{P,2}^2 + \frac{8}{\sqrt{n}} E_P^* \|\mathbb{G}_n\|_{\mathcal{F}} \\
&\lesssim \delta^2 \|F\|_{P,2}^2 + \frac{8}{\sqrt{n}} J\left(\frac{\{E_P^*(\sup_f \mathbb{P}_n f^2)\}^{1/2}}{\|F\|_{P,2}}, \mathcal{F}, L_2 \right) \|F\|_{P,2}.
\end{aligned}
$$

Dividing through by $\|F\|_{P,2}^2$ we see that $z^2 \equiv E_P^*(\sup_f \mathbb{P}_n f^2)\}/\|F\|_{P,2}^2$ satisfies

$$
z^2 \lesssim \delta^2 + \frac{J(z, \mathcal{F}, L_2)}{\sqrt{n}\|F\|_{P,2}}. \qquad (4)
$$

2. Proof, part 2: inversion

Lemma. (Inversion) Let $J : (0,\infty) \to \mathbb{R}$ be a concave, nondecreasing function with $J(0) = 0$. If $z^2 \leq A^2 + B^2 J(z^r)$ for some $r \in (0,2)$ and $A, B > 0$, then

$$J(z) \lesssim J(A) \left\{ 1 + J(A^r) \left(\frac{B}{A} \right)^2 \right\}^{1/(2-r)}.$$

Applying this Lemma with $r = 1$, $A = \delta$ and $B^2 = 1/(\sqrt{n}\|F\|_{P,2})$ yields

$$J(z, \mathcal{F}, L_2) \lesssim J(\delta, \mathcal{F}, L_2) \left(1 + \frac{J(\delta, \mathcal{F}, L_2)}{\delta^2 \sqrt{n} \|F\|_{P,2}} \right).$$

Combining this with (3) completes the proof:

$$\begin{aligned} E_P^* \|\mathbb{G}_n\|_{\mathcal{F}} &\lesssim J\left(\frac{\{E_P^*(\sup_f \mathbb{P}_n f^2)\}^{1/2}}{\|F\|_{P,2}}, \mathcal{F}, L_2 \right) \|F\|_{P,2} \\ &\lesssim J(\delta, \mathcal{F}, L_2) \left(1 + \frac{J(\delta, \mathcal{F}, L_2)}{\delta^2 \sqrt{n} \|F\|_{P,2}} \right) \|F\|_{P,2}. \qquad (5) \end{aligned}$$

2. Proof, part 2: inversion

Proof of the inversion lemma: For $0 < s < t$ we can write $s = (s/t)t + (1 - s/t)0$, so by concavity of J and $J(0) = 0$ we have

$$J(s) \geq \frac{s}{t} J(t),$$

and hence $J(t)/t$ is decreasing. Thus for $C \geq 1$ and $t > 0$ it follows that

$$J(Ct) \leq C J(t). \tag{6}$$

Now since J is $\nearrow$ it follows from the hypothesis on z that a

$$\begin{aligned}
J(z^r) &\leq J((A^2 + B^2 J(z^r))^{r/2}) \\
&= J(A^r (1 + (B/A)^2 J(z^r))^{r/2}) \equiv J(tC) \quad \text{with } C \geq 1 \\
&\leq J(A^r) \left(1 + (B/A)^2 J(z^r)\right)^{r/2} \\
&\leq 2 \max\{J(A^r),\ J(A^r)(B/A)^r J(z^r)^{r/2}\}.
\end{aligned}$$

2. Proof, part 2: inversion

If $J(z^r) \le J(A^r)(B/A)^r J(z^r)^{r/2}$, then $J(z^r)^{1-r/2} \le J(A^r)(B/A)^r$, so

$$J(z^r) \le \{J(A^r)(B/A)^r\}^{2/(2-r)}.$$

Hence we conclude that

$$J(z^r) \lesssim J(A^r) + J(A^r)^{2/(2-r)}(B/A)^{2r/(2-r)}.$$

Repeating the argument above, but starting with $J(z)$ and then using the above bound for $J(z^r)$ yields

$$\begin{aligned}
J(z) &\le J((A^2 + B^2 J(z^r))^{1/2}) \\
&= J(A(1 + (B/A)^2 J(z^r))^{1/2}) \equiv J(tC) \quad \text{with } C \ge 1 \\
&\le J(A)\left(1 + (B/A)^2 J(z^r)\right)^{1/2} \\
&\le J(A)\left(1 + (B/A)^2\left(J(A^r) + J(A^r)^{2/(2-r)}(B/A)^{2r/(2-r)}\right)\right)^{1/2} \\
&\le J(A)\left(1 + J(A^r)^{1/2}(B/A) + J(A^r)^{1/(2-r)}(B/A)^{2/(2-r)}\right).
\end{aligned}$$

2. Proof, part 2: inversion

But by Young's inequality the second term $x \equiv J(A^r)^{1/2}(B/A)$ is bounded above by $1^p + x^q$ for any conjugate exponents p and q (ie for $a, b > 0$, $ab \leq a^p + b^q$). Choosing $p = 2/r$ and $q = 2/(2-r)$ yields

$$J(A^r)^{1/2}(B/A) \leq 1 + J(A^r)^{1/(2-r)}(B/A)^{2/(2-r)}.$$

Thus the preceding argument yields the conclusion:

$$\begin{aligned} J(z) &\leq 2J(A)\left(1 + J(A^r)^{1/(2-r)}(B/A)^{2/(2-r)}\right) \\ &\lesssim J(A)\left(1 + J(A^r)(B/A)^2\right)^{1/(2-r)}. \end{aligned}$$

7. Generalizations to unbounded classes $\mathcal{F}$

Theorem 2. Let $\mathcal{F}$ be a $P-$measurable class of measurable functions with envelope function F such that $PF^{(4p-2)/(p-1)} < \infty$ for some $p > 1$ and such that $\mathcal{F}^2$ and $\mathcal{F}^4$ are $P-$measurable. If $Pf^2 < \delta^2 PF^2$ for every $f \in \mathcal{F}$ and some $\delta \in (0,1)$, then

$$E_P^* \|\mathbb{G}_n\|_{\mathcal{F}} \lesssim J(\delta, \mathcal{F}, L_2)\|F\|_{P,2}\left(1 + \frac{J(\delta^{1/p}, \mathcal{F}, L_2)}{\delta^2\sqrt{n}} \frac{\|F\|_{P,(4p-2)/(p-1)}^{2-1/p}}{\|F\|_{P,2}^{2-1/p}}\right)^{p/(2p-1)}.$$

7. Generalizations to unbounded classes $\mathcal{F}$

Theorem 3. Let $\mathcal{F}$ be a $P-$measurable class of measurable functions with envelope function F such that $P\exp(F^{p+\rho}) < \infty$ for some $p, \rho > 0$ and such that $\mathcal{F}^2$ and $\mathcal{F}^4$ are $P-$measurable. If $Pf^2 < \delta^2 PF^2$ for every $f \in \mathcal{F}$ and some $\delta \in (0, 1/2)$, then for a constant c depending on p, PF^2, PF^4 and $P\exp(F^{p+\rho})$,

$$E_P^* \|\mathbb{G}_n\|_{\mathcal{F}} \lesssim c J(\delta, \mathcal{F}, L_2) \left(1 + \frac{J(\delta(\log(1/\delta))^{1/p}, \mathcal{F}, L_2)}{\delta^2 \sqrt{n}}\right).$$

8. An application: minimum contrast estimators

Suppose that $\widehat{\theta}_n$ minimizes

$$\theta \mapsto \mathbb{M}_n(\theta) \equiv \mathbb{P}_n m_\theta$$

for given measurable functions $m_\theta : \mathcal{X} \to R$ indexed by a parameter θ, and that the population contrast

$$\theta \mapsto \mathbb{M}(\theta) = P m_\theta$$

satisfies, for $\theta_0 \in \Theta$ and some metric d on Θ,

$$P m_\theta - P m_{\theta_0} \gtrsim d^2(\theta, \theta_0).$$

A bound on the rate of convergence of $\widehat{\theta}_n$ to θ_0 can then be derived from the modulus of continuity of the empirical process $\mathbb{G}_n m_\theta$ index by the functions m_θ.

8. An application: minimum contrast estimators

If ϕ_n is a function such that $\delta \mapsto \phi_n(\delta)/\delta^\alpha$ is decreasing for some $\alpha < 2$ and

$$E \sup_{\theta:\delta(\theta,\theta_0)<\delta} |\mathbb{G}_n(m_\theta - m_{\theta_0})| \lesssim \phi_n(\delta), \tag{7}$$

then $d(\widehat{\theta}_n, \theta_0) = O_p(\delta_n)$ for δ_n any solution to

$$\phi_n(\delta_n) \leq \sqrt{n}\delta_n^2.$$

The inequality (7) involves the empirical process indexed by the class of functions $\mathcal{M}_\delta = \{m_\theta - m_{\theta_0} : \ d(\theta, \theta_0) < \delta\}$. If d dominates the $L_2(P)-$norm, or another norm $\| \, \|$ (such as the Bernstein norm) and the norms of the envelopes M_δ of the classes $\mathcal{M}_\delta$ are bounded in δ, then we can choose

$$\phi_n(\delta) = J(\delta, \mathcal{M}_\delta, \| \cdot \|) \left(1 + \frac{J(\delta, \mathcal{M}_\delta, \| \cdot \|)}{\delta^2 \sqrt{n}} \right).$$

where J is an appropriate entropy integral.

Selected References:

- Boyd, S. and Vandenberghe, L. (2004). *Convex Optimization*. Cambridge University Press.

- Hiriart-Urruty, J-B. and Lemaréchal, C. (2004). *Fundamentals of Convex Analysis.* Corrected Second Printing. Springer, Berlin.

- Pollard, D. (1990). *Empirical processes: theory and applications.* NSF-CBMS Regional Conference Series in Probability and Statistics, 2. Institute of Mathematical Statistics, Hayward, CA.

- Van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes.* Springer-Verlag, New York.

Thank You!

$$2011 = \sum\{157, 163, 167, 173, \underline{179, 181}, \underline{191, 193}, \underline{197, 199}, 211\}$$