A Local Maximal Inequality under Uniform Entropy

Jon A. Wellner

University of Washington, Seattle, visiting Heidelberg

Seminaire, Institut de Mathématiques de Toulouse

8 March 2012

Seminaire 4, Toulouse

Thursday, 8 March

Based on joint work with: Aad van der Vaart

Seminaire 1: Tuesday, 28 February
arXiv:1203.0828v1
Seminaire 2: Thursday, 1 March
Seminaire 3: Monday, 5 March

Available at: *http://www.stat.washington.edu/jaw/RESEARCH/TALKS/talks.html*

Outline

- 1. The setting and basic problem
- 2. Available bounds: bracketing and uniform entropy
- 3. Applications to convergence rates
  - ▷ Example 1: Chernoff's mode estimator.
  - ▷ Example 2: Grenander's estimator.
- 4. The new bound: uniform entropy
- 5. The perspective of a convex (or concave) function
- 6. Proof, part 1: concavity of the entropy integral
- 7. Proof, part 2: inversion
- 8. Generalizations
- 9. More applications

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 $f \in \mathcal{F}$, $x \in \mathcal{X}$;

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

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

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)

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

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

$$E_P^* \|\mathbb{G}_n\|_{\mathcal{F}} \lesssim J_{[\,]}(\delta, \mathcal{F}, L_2(P)) \|F\|_{P,2} \left(1 + \frac{J_{[\,]}(\delta, \mathcal{F}, L_2(P))}{\delta^2 \sqrt{n} \|F\|_{P,2}} \right). \tag{3}$$

3. Applications to convergence rates

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,

$$P m_\theta - P m_{\theta_0} \gtrsim d^2(\theta, \theta_0). \tag{4}$$

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.

3. Applications to convergence rates

Theorem 1. Suppose that (4) holds. If ϕ_n is a function such that $\delta \mapsto \phi_n(\delta)/\delta^\alpha$ is decreasing for some $\alpha < 2$ and

$$E \sup_{\theta: d(\theta,\theta_0)<\delta} |\mathbb{G}_n(m_\theta - m_{\theta_0})| \lesssim \phi_n(\delta), \tag{5}$$

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 (5) 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.

Example 1. Suppose that $X_1, \ldots, X_n$ are i.i.d. P on $\mathbb{R}$ with density p with respect to Lebesgue measure λ. Fix $a > 0$ and let

$$\mathbb{M}_n(\theta) = \mathbb{P}_n 1_{[\theta - a, \theta + a]} = \mathbb{P}_n m_\theta \,,$$

the proportion of the sample in the interval $[\theta - a, \theta + a]$. Correspondingly,

$$\mathbb{M}(\theta) = P m_\theta = P(|X - \theta| \le a) = F_X(\theta + a) - F_X((\theta - a)-)$$

where $F_X(x) = P(X \le x)$ is the distribution function of X. Is this maximized uniquely by some θ_0? Since P has Lebesgue density p, it follows that $\mathbb{M}$ is differentiable and

$$\mathbb{M}'(\theta) = p(\theta + a) - p(\theta - a) = 0$$

if $p(\theta+a) = p(\theta-a)$ which clearly holds for the point of symmetry θ_0 if p is symmetric and unimodal about θ_0. If p is just unimodal, with $p'(x) > 0$ for $x < \theta_0$ and $p'(x) < 0$ for $x > \theta_0$, then $\theta_0 \equiv$ argmax $\mathbb{M}(\theta)$ might not agree with the mode, but it is "nearby".

Does it hold that

$$\widehat{\theta}_n = \text{argmax}\, \mathbb{M}_n(\theta) \to_p \text{argmax}\, \mathbb{M}(\theta) = \theta_0\,?$$

If this holds, do we have

$$r_n(\widehat{\theta}_n - \theta_0) \;\begin{cases} = O_p(1) & \text{for some } \; r_n \to \infty \\ \to_d \mathbb{Z} & \text{for some limiting random variable } \; \mathbb{Z}? \end{cases}$$

Let $\mathcal{F} = \{m_\theta : \; \theta \in \mathbb{R}\}$. This is a VC -subgraph class of functions of dimension $S(\mathcal{F}) = 2$. Now it is easily seen that with $\mathcal{M}_\delta(\theta_0) = \{m_\theta - m_{\theta_0} : \; d(\theta, \theta_0) < \delta\}$ we have

$$\begin{aligned} N(\epsilon, \mathcal{M}_\delta(\theta_0), L_2(Q)) &\leq N(\epsilon, \mathcal{F}_\infty, L_2(Q)) \\ &\leq N^2(\epsilon/2, \mathcal{F}, L_2(Q)) \leq \left(\frac{K}{\epsilon}\right)^8, \end{aligned}$$

and hence the entropy integral

$$J(1, \mathcal{M}_\delta) \lesssim \int_0^1 \sqrt{8 \log(K/\epsilon)}\, d\epsilon < \infty\,.$$

Furthermore, $\mathcal{M}_\delta(\theta_0)$ has envelope function

$$\begin{aligned} M_\delta(x) &= \sup\{|m_\theta(x) - m_{\theta_0}(x)| : |\theta - \theta_0| < \delta\} \\ &= 1_{[\theta_0+a-\delta,\theta_0+a+\delta]}(x) + 1_{[\theta_0-a-\delta,\theta_0-a+\delta]}(x) \end{aligned}$$

for $\delta < a$, and we compute

$$\begin{aligned} P(M_\delta^2) &= P(\theta_0 + a - \delta \le X \le \theta_0 + a + \delta) \\ &\qquad + P(\theta_0 - a - \delta \le X \le \theta_0 - a + \delta) \\ &\le 4\|p\|_\infty \delta\,, \end{aligned}$$

so $\|M_\delta\|_{P,2} \le 2\|p\|_\infty^{1/2}\delta^{1/2}$. Combining these calculations with Pollard's bound (1) yields

$$E^*\|\mathbb{G}_n\|_{\mathcal{M}_\delta} \lesssim J(1, \mathcal{M}_\delta)\|M_\delta\|_{P,2} \lesssim \delta^{1/2} \equiv \phi(\delta)\,.$$

The only remaining ingredient to apply the rate Theorem 1 is to verify (4). This will typically hold for unimodal densities since

$$\mathbb{M}(\theta) - \mathbb{M}(\theta_0) = \frac{1}{2}\Big(p'(\theta_0 + a) - p'(\theta_0 - a)\Big)(\theta - \theta_0)^2 + o(\|\theta - \theta_0\|^2)$$

where $p'(\theta_0 - a) > 0$ and $p'(\theta_0 + a) < 0$.

Now with $\phi_n(\delta) \equiv \phi(\delta) \equiv C\delta^{1/2}$ we have

$$C\delta_n^{1/2} = \phi_n(\delta_n) \lesssim \sqrt{n}\delta_n^2$$

if $\delta_n = n^{-1/3}$:

$$Cn^{-1/6} \lesssim n^{1/2}n^{-2/3} = n^{-1/6}.$$

Thus we find that $r_n = 1/\delta_n = n^{1/3}$, and hence, by Theorem 1,

$$n^{1/3}(\widehat{\theta}_n - \theta_0) = O_p(1).$$

Theorem 2. Suppose that $X_1, \ldots, X_n$ are i.i.d. P_0 with density $p_0 \in \mathcal{P}$. Let H be the Hellinger distance between densities, and let m_p be defined, for $p \in \mathcal{P}$, by

$$m_p(x) = \log\left(\frac{p(x) + p_0(x)}{2p_0(x)}\right).$$

Then $\mathbb{M}(p) - \mathbb{M}(p_0) = P_0(m_p - m_{p_0}) \lesssim -H^2(p, p_0)$. Furthermore, with $\mathcal{M}_\delta = \{m_p - m_{p_0} : \ H(p, p_0) \le \delta\}$, we also have

$$E^*_{P_0} \|\mathbb{G}_n\|_{\mathcal{M}_\delta} \lesssim \tilde{J}_{[\,]}(\delta, \mathcal{P}, H)\left(1 + \frac{\tilde{J}_{[\,]}(\delta, \mathcal{P}, H)}{\delta^2 \sqrt{n}}\right) \equiv \phi_n(\delta). \qquad (6)$$

Thus if $\delta_n = 1/r_n$ satisfies $\phi_n(\delta_n) \lesssim \sqrt{n}\delta_n^2$, then

$$r_n H(\widehat{p}_n, p_0) = O_p(1).$$

Example 2. The Grenander estimator of a monotone decreasing density. Let

$$\mathcal{P} \equiv \{p : [0, B] \to [0, M] \Big| p \text{ is nonincreasing}\}.$$

Then with Q denoting the uniform distribution on $[0, M]$,

$$\log N_{[\,]}(\epsilon, \mathcal{P}, L_2(Q)) \lesssim \epsilon^{-1},$$

and

$$\log N_{[\,]}(\epsilon, \mathcal{P}, H) \lesssim \epsilon^{-1}.$$

Thus

$$J_{[\,]}(\delta, \mathcal{P}, H) \lesssim \int_0^\delta \epsilon^{-1/2} d\epsilon = 2\delta^{1/2}.$$

Then we have

$$\phi_n(\delta) \lesssim \delta^{1/2}\left(1 + \frac{\delta^{1/2}}{\delta^2\sqrt{n}}\right) = \delta^{1/2} + \frac{1}{\delta\sqrt{n}}$$

So we have $\phi_n(\delta_n) \lesssim \sqrt{n}\delta_n^2$ if $\delta_n = n^{-1/3}$:

$$\begin{aligned}\phi_n(\delta_n) &= n^{-1/6} + \frac{1}{n^{-1/3} n^{1/2}} = 2n^{-1/6} \\ &\lesssim n^{1/2} n^{-2/3} = n^{-1/6}.\end{aligned}$$

From Theorem 2 we conclude that

$$n^{1/3} H(\widehat{p}_n, p_0) = O_p(1).$$

4. 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 $e_1, e_2, \ldots, e_n \in \{-1, 1\}$.

4. 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

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

5. 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

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

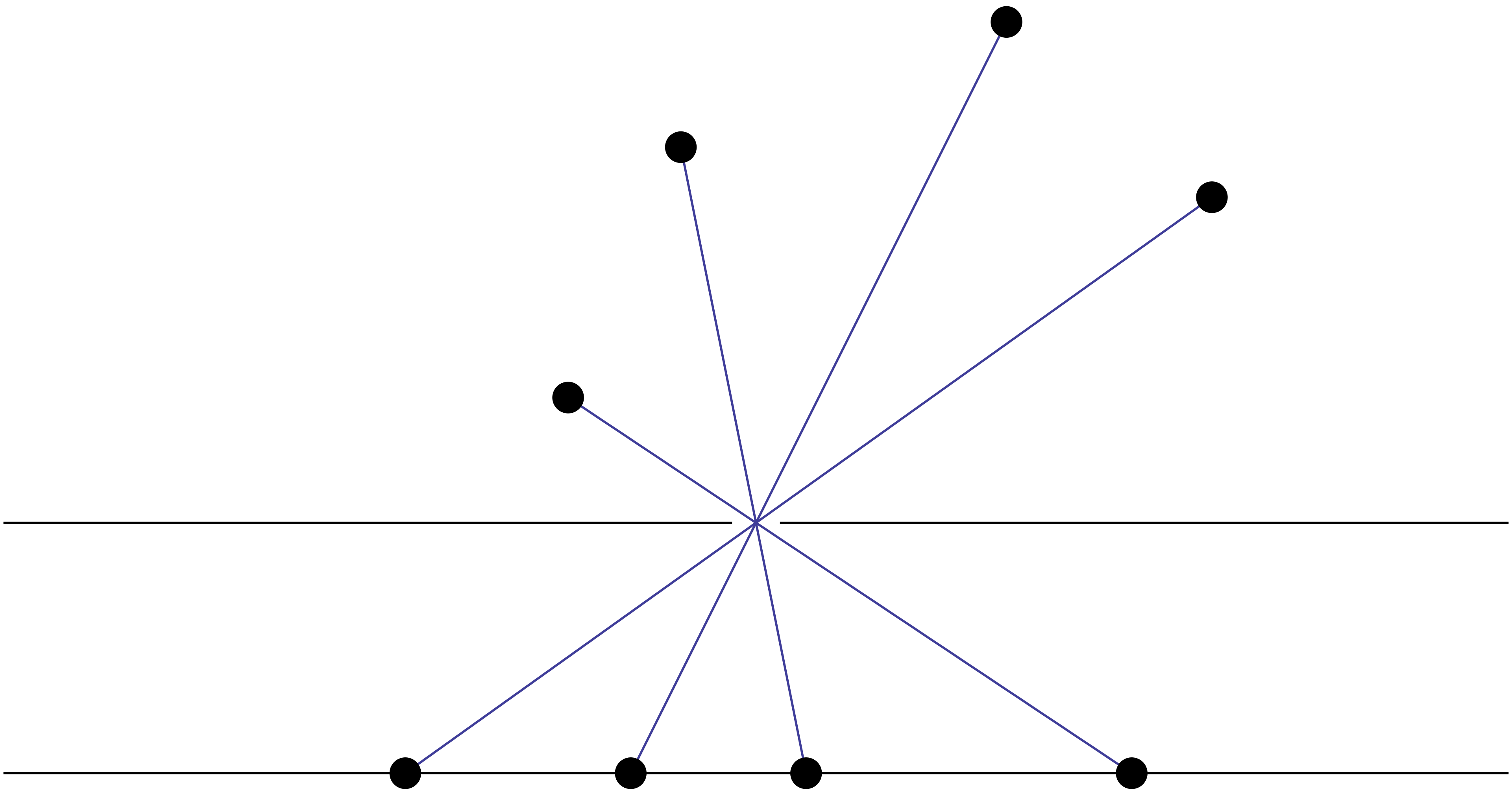
for $(x,t) \in \text{dom}(g) = \{(x,t) : x/t \in \text{dom}(f),\ t > 0\}$.
Then:

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

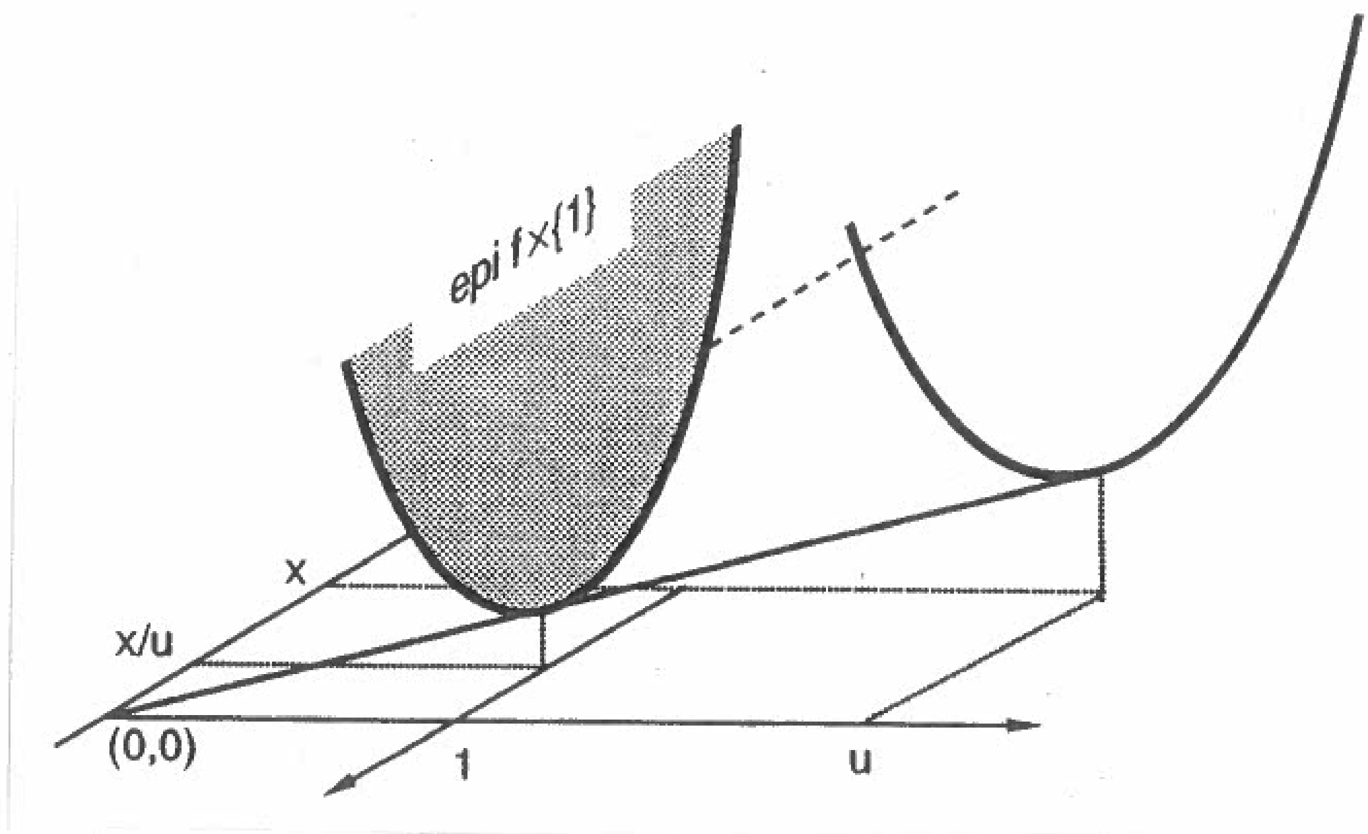
This seems to be due to Hiriart-Urruty and Lemaréchal (1990), vol. 1, page 100; see also Boyd and Vandenberghe (2004), page 89.

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

5. The perspective of a convex or concave function



5. The perspective of a convex or concave function



5. 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.

6. 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

$$E_P^* \|\mathbb{G}_n\|_{\mathcal{F}} \lesssim E_P^* J \left(\frac{\sup_f (\mathbb{P}_n f^2)^{1/2}}{(\mathbb{P}_n F^2)^{1/2}}, \mathcal{F}, L_2 \right) (\mathbb{P}_n F^2)^{1/2}. \qquad (7)$$

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.

6. 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 (7) it follows that

$$E_P^* \|\mathbb{G}_n\|_{\mathcal{F}} \lesssim J\left(\frac{\{E_P^*(\sup_f \mathbb{P}_n f^2)\}^{1/2}}{\|F\|_{P,2}}, \mathcal{F}, L_2\right) \|F\|_{P,2}. \tag{8}$$

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 (8), that

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

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

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

$$z^2 \lesssim \delta^2 + \frac{J(z, \mathcal{F}, L_2)}{\sqrt{n}\|F\|_{P,2}}. \qquad (9)$$

7. 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 (8) completes the proof:

$$\begin{aligned} E_P^* \|\mathbb{G}_n\|_{\mathcal{F}} &\lesssim J\left(\frac{\{E_P^*(\sup_f \mathbb{P}_n f^2)\}^{1/2}}{\|F\|_{P,2}}, \mathcal{F}, L_2 \right) \|F\|_{P,2} \\ &\lesssim J(\delta, \mathcal{F}, L_2) \left(1 + \frac{J(\delta, \mathcal{F}, L_2)}{\delta^2 \sqrt{n} \|F\|_{P,2}} \right) \|F\|_{P,2}. \qquad (10) \end{aligned}$$

7. 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

$$J(Ct) \leq CJ(t). \tag{11}$$

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

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

7. Proof, part 2: inversion

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

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

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

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

7. 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:

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

8. 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

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

Proof: Replace the contraction inequality with an argument involving Hölder's inequality together with preservation properties of uniform entropy and use concavity of the perspective function again with powers other than $\sqrt{x}$.

8. 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 $P\exp(F^{p+\rho})$,

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

9. More applications: minimum contrast estimators

- Global rate of convergence for estimating a $k-$monotone density; Gao and Wellner (2009).
- Global rate of convergence for case 2 interval censoring:
  - ▷ van de Geer (1993), (1996)
  - ▷ Geskus and Groeneboom (1997, 1999).
- $\cdots$!

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.

- van der Vaart, A. W. and Wellner, J. A. (2011). A local maximal inequality under uniform entropy. *Electronic Journal of Statistics* **5**, 192-203.

- van de Geer, S. (1993). Hellinger consistency of certain nonparametric maximum likelihood estimators. *Ann. Statist.* **21**, 14 - 44.

- van de Geer, S. (1996). Rates of convergence for the maximum likelihood estimator in mixture models. *J. Nonpar. Statist.* **6**, 293310.

- Gao, F. and Wellner, J. A. (2009). On the rate of convergence of the maximum likelihood estimator of a $k-$monotone density. *Science in China Series A: Mathematics* **52**, 1-14.

- Geskus, R. B. and Groeneboom, P. (1997). Asymptotically optimal estimation of smooth functionals for interval censoring. II. *Statist. Neerlandica* **51**, 201219.

- Wellner, Jon A. (2005). *Empirical Processes: Theory and Application.* Notes for a course given at Delft University of Technology, 2005. On-line at: *http://www.stat.washington.edu/jaw/RESEARCH/TALKS/talks.html*

Merci beaucoup!