Empirical Process Theory for Statistics

Jon A. Wellner

University of Washington, Seattle, visiting Heidelberg

Short Course to be given at
Institut de Statistique, Biostatistique, et Sciences Actuarielles
Louvain-la-Neuve

29-30 May 2012

Short Course, Louvain-la-Neuve

- Day 1 (Tuesday):
  - ▷ Lecture 1: Introduction, history, selected examples.
  - ▷ Lecture 2: Some basic inequalities and Glivenko-Cantelli theorems.
  - ▷ Lecture 3: Using the Glivenko-Cantelli theorems: first applications.

Based on Courses given at Torgnon, Cortona,
and Delft (2003-2005). Notes available at:

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

- Day 2 (Wednesday):
  - ▷ Lecture 4: Donsker theorems and some inequalities
  - ▷ Lecture 5: Peeling methods and rates of convergence
  - ▷ Lecture 6: Some useful preservation theorems.

Lecture 4: Donsker theorems and some inequalities

- 1. Donsker theorems
  - ▷ Donsker theorem equivalences
  - ▷ Uniform entropy Donsker theorem
  - ▷ Bracketing entropy Donsker theorem
- 2. Bracketing Inequalities for expectations of suprema
- 3. Uniform entropy inequalities for expectations of suprema

1. (a) Donsker theorem equivalences

Reminder: the setting

- $X_1, \ldots, X_n$ are i.i.d. with probability measure P on $(\mathcal{X}, \mathcal{A})$.
- $\mathbb{P}_n = n^{-1} \sum_{i=1}^n \delta_{X_i}$, the empirical measure; here

$$\delta_x(A) = 1_A(x) = \begin{cases} 1, & x \in A, \\ 0, & x \in A^c \end{cases} \quad \text{for} \quad A \in \mathcal{A}.$$

  Hence we have

$$\mathbb{P}_n(A) = n^{-1} \sum_{i=1}^n 1_A(X_i), \quad \text{and} \quad \mathbb{P}_n(f) = n^{-1} \sum_{i=1}^n f(X_i).$$

- $\{\mathbb{G}_n(f) \equiv \sqrt{n}(\mathbb{P}_n(f) - P(f)) : \; f \in \mathcal{F} \subset L_2(P)\}$, the empirical process indexed by $\mathcal{F}$

Recall that: if $\mathcal{F}$ is a class of functions satisfying

$$\mathbb{G}_n \Rightarrow \mathbb{G} \quad \text{in} \quad \ell^{\infty}(\mathcal{F}), \tag{1}$$

where $\mathbb{G}$ is a tight, mean 0 Gaussian process with uniformly continuous sample paths with respect to ρ_P, then we say that $\mathcal{F}$ is a $P-$Donsker class of functions. Here $\mathbb{G}$ is a 0−mean $P-$Brownian bridge process with uniformly-continuous sample paths with respect to the semi-metric $\rho_P(f,g)$ defined by

$$\rho_P^2(f,g) = Var_P(f(X) - g(X)),$$

$\ell^{\infty}(\mathcal{F})$ is the space of all bounded, real-valued functions from $\mathcal{F}$ to $\mathbb{R}$:

$$\ell^{\infty}(\mathcal{F}) = \left\{ x : \mathcal{F} \mapsto \mathbb{R} \middle| \; \|x\|_{\mathcal{F}} \equiv \sup_{f \in \mathcal{F}} |x(f)| < \infty \right\},$$

and

$$E\{\mathbb{G}(f)\mathbb{G}(g)\} = P(fg) - P(f)P(g)$$

Here $\mathbb{G}_n \Rightarrow \mathbb{G}$ means that

$$E^* h(\mathbb{G}_n) \rightarrow E h(\mathbb{G})$$

for all bounded, continuous functions $h : \ell^\infty(\mathcal{F}) \rightarrow \mathbb{R}$, where

$$\begin{aligned} E^* h(\mathbb{G}_n) &\equiv \text{inf}\{EU : \; U \text{ measurable } \; U \geq h(\mathbb{G}_n)\} \\ &= \text{the "outer expectation"} \\ &\quad\; \text{(Hoffmann-Jørgensen theory)} \end{aligned}$$

Consequence:

$$\mathbb{G}_n \Rightarrow \mathbb{G} \;\; \text{implies that} \;\; h(\mathbb{G}_n) \rightarrow_d h(\mathbb{G})$$

for continuous functions h.

In general, a sequence $\{\mathbb{X}_n(t) : t \in T\}$ of sample-bounded stochastic processes (i.e. with values in $\ell^\infty(T)$) satisfies

$$\mathbb{X}_n \Rightarrow \mathbb{X} \quad \text{in} \quad \ell^\infty(T)$$

where $\mathbb{X}$ has a tight Borel probability distribution in $\ell^\infty(T)$ if and only the following two conditions hold:

- All the finite-dimensional distributions of $\mathbb{X}_n$ converge in distribution to those of $\mathbb{X}$.
- There exists a pseudo-metric ρ on T such that:
  - ▷ (T, ρ) is totally bounded.
  - ▷ The processes $\mathbb{X}_n$ are asymptotically $\rho-$equicontinuous in probability: i.e. for every $\epsilon > 0$

$$\lim_{\delta \to 0} \limsup_{n \to \infty} Pr\Big(\sup_{\rho(s,t) < \delta} |\mathbb{X}_n(s) - \mathbb{X}_n(t)| > \epsilon \Big) = 0.$$

For the empirical processes $\mathbb{G}_n$ indexed by $\mathcal{F}$ this (together with the Hoffmann-Jørgensen inequality) yields:

Corollary: Let $\mathcal{F}$ be a class of measurable functions. Then the following are equivalent:

(i) $\mathcal{F}$ is $P-$Donsker.

(ii) $(\mathcal{F}, \rho_P)$ is totally bounded and $\mathbb{G}_n$ is asymptotically equicontinuous in probability: for every $\epsilon > 0$

$$\lim_{\delta \to 0} \limsup_{n \to \infty} Pr^*(\sup_{\rho_P(f,g)<\delta} |\mathbb{G}_n(f) - \mathbb{G}_n(g)| > \epsilon) = 0.$$

(iii) $(\mathcal{F}, \rho_P)$ is totally bounded and $\mathbb{G}_n$ is asymptotically equicontinuous in mean:

$$\lim_{\delta \to 0} \limsup_{n \to \infty} E^* \left(\sup_{\rho_P(f,g)<\delta} |\mathbb{G}_n(f) - \mathbb{G}_n(g)| \right) = 0.$$

1. (b) Uniform entropy Donsker theorem

Theorem. (Pollard, Koltchinskii) Let $\mathcal{F}$ be a class of measurable functions with square-integrable envelope function F. Let the classes $\mathcal{F}_\delta = \{f - g : \ f, g \in \mathcal{F}, \ \|f - g\|_{P,2} < \delta\}$ and $\mathcal{F}_\infty^2$ be $P-$measurable for every $\delta > 0$. Then $\mathcal{F}$ is $P-$Donsker provided that the following uniform entropy condition holds:

$$\int_0^\infty \sup_{Q \in \mathcal{Q}} \sqrt{\log N(\epsilon \|F\|_{Q,2}, \mathcal{F}, L_2(Q))} d\epsilon < \infty;$$

here $\mathcal{Q}$ denotes the class of all finitely discrete probability measures on $(\mathcal{X}, \mathcal{A})$, and

$N(\epsilon, \mathcal{F}, \|\cdot\|) \equiv$ minimal number of $\epsilon -$ balls needed to cover $\mathcal{F}$.

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

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

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

1. (c) Bracketing entropy Donsker theorem

Theorem. (Ossiander)
Suppose that $\mathcal{F}$ is a class of measurable functions satisfying

$$\int_0^\infty \sqrt{\log N_{[\,]}(\epsilon, \mathcal{F}, L_2(P))}\, d\epsilon < \infty.$$

Then $\mathcal{F}$ is $P-$Donsker: $\mathbb{G}_n \Rightarrow \mathbb{G}$ in $\ell^\infty(\mathcal{F})$.

For a refinement of this theorem using a "weak $L_2(P)$ norm", see van der Vaart & W (1996) Theorem 2.5.6, page 130. For further refinements involving majorizing measures, see Andersen, Giné, Ossiander, and Zinn (1988).

Note that in both of these theorems, the entropy integral condition holds if the entropy (i.e. the logarithm of the covering or bracketing number) is bounded by

$$K\epsilon^{-r} \quad \text{with} \quad r < 2.$$

Proofs of the Donsker theorems?

- Uniform entropy CLT: (VdV & W, page 128)
  - ▷ Symmetrization with Rademacher rv's: Let $\mathbb{G}_n^0(f) \equiv n^{-1/2}\sum_{i=1}^n \epsilon_i f(X_i)$. Then
$$E^*\|\mathbb{G}_n\|_{\mathcal{F}_{\delta_n}} \le 2E^*\Big\|\mathbb{G}_n^0\|_{\mathcal{F}_{\delta_n}}.$$
  - ▷ Hoeffding's inequality:
$$P(|\sum_1^n a_i\epsilon_i| > x) \le 2\exp(-\frac{x^2}{2\|a\|^2})$$
  - ▷ Maximal inequality for Ψ_2- Orlicz norm (conditionally on X_i's) and chaining:
$$E_\epsilon\Big\|\frac{1}{\sqrt{n}}\sum_{i=1}^n \epsilon_i f(X_i)\Big\|_{\mathcal{F}_{\delta_n}} \lesssim \int_0^\infty \sqrt{\log N(\epsilon, \mathcal{F}_{\delta_n}, L_2(\mathbb{P}_n))}d\epsilon$$
  - ▷ Relate $\mathcal{F}_{\delta_n}$ to $\mathcal{F}$ and integrate over X_i's: measurability needed for Fubini to apply!

Proofs of the Donsker theorems?

- Bracketing entropy CLT: (VdV & W, pages 129 - 133)
  - ▷ Truncation at multiples of $\sqrt{n}$.
  - ▷ Bernstein's inequality for a fixed bounded f:

$$P(|\mathbb{G}_n(f)| > x) \le 2\exp\left(-\frac{1}{2}\frac{x^2}{Pf^2 + (1/3)\|f\|_\infty x/\sqrt{n}}\right).$$

  - ▷ Orlicz norm type maximal inequality related to Bernstein's inequality: for a finite set $\mathcal{F}$ with $|\mathcal{F}| > 2$,

$$E\|\mathbb{G}_n\|_{\mathcal{F}} \lesssim \max_f \frac{\|f\|_\infty}{\sqrt{n}} \log|\mathcal{F}| + \max_f \|f\|_{P,2}\sqrt{\log|\mathcal{F}|}.$$

  - ▷ Set truncation levels so that $\|f\|_\infty \approx \sqrt{n}\|f\|_{P,2}/\sqrt{\log|\mathcal{F}|}$.
  - ▷ Chaining argument!

2. Bracketing Inequalities for expectations of suprema

For a given norm $\|\cdot\|$, define a bracketing integral of a class of functions $\mathcal{F}$ by

$$J_{[\,]}(\delta, \mathcal{F}, \|\cdot\|) = \int_0^\delta \sqrt{1 + \log N_{[\,]}(\epsilon\|F\|, \mathcal{F}, \|\cdot\|)}\, d\epsilon.$$

A basic bracketing maximal inequality uses the $L_2(P)-$norm:

Theorem: Let $\mathcal{F}$ be a class of measurable functions with measurable envelope function F. For $\eta > 0$ set

$$a(\eta) \equiv \frac{\eta\|F\|_{P,2}}{\sqrt{1 + \log N_{[\,]}(\eta\|F\|_{P,2}, \mathcal{F}, L_2(P))}}.$$

Then:

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

- Furthermore

$$E^*\|\mathbb{G}_n\|_{\mathcal{F}} \lesssim J_{[\,]}(\eta,\mathcal{F},L_2(P))\|F\|_{P,2}+\sqrt{n}PF1\{F>\sqrt{n}a(\eta)\}$$
$$+\ \|\|f\|_{P,2}\|_{\mathcal{F}}\sqrt{1+\log N_{[\,]}(\eta\|F\|_{P,2},\mathcal{F},L_2(P))}.$$

- In particular, if $\|f\|_{P,2}<\delta\|F\|_{P,2}$ for every $f\in\mathcal{F}$, then

$$E^*\|\mathbb{G}_n\|_{\mathcal{F}} \lesssim J_{[\,]}(\delta,\mathcal{F},L_2(P))\|F\|_{P,2}+\sqrt{n}P(F1\{F>\sqrt{n}a(\delta)\}).$$

- If $\|f\|_\infty\leq 1$ and $Pf^2<\delta^2PF^2$ for every $f\in\mathcal{F}$, then

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

3. Uniform entropy inequalities for expectations of suprema

Define a uniform entropy integral of a class of functions $\mathcal{F}$ by

$$J(\delta, \mathcal{F}) = \sup_Q \int_0^\delta \sqrt{1 + \log N(\epsilon \|F\|_{Q,2}, \mathcal{F}, L_2(Q))} d\epsilon.$$

A basic uniform entropy maximal inequality is given by:

Theorem: Let $\mathcal{F}$ be a $P-$measurable class of measurable functions with measurable envelope function F. Then with $\theta_n \equiv \sup_{f \in \mathcal{F}} \{\|f\|_n / \|F\|_n\}$, and $\|f\|_n \equiv \|f\|_{\mathbb{P}_n,2} \equiv \|f\|_{L_2(\mathbb{P}_n)}$,

$$E^* \|\mathbb{G}_n\|_{\mathcal{F}} \lesssim E\left\{J(\theta_n, \mathcal{F}) \|F\|_n\right\} \lesssim J(1, \mathcal{F}) \|F\|_{P,2}. \tag{3}$$

Proof. See vdV & W (1996), page 239.

Problem: Note that θ_n and $\|F\|_n$ are both random!

Question: Is there a bound analogous to the "small f bound, bracketing entropy", but for uniform entropy?

Answer: Yes! (VdV & W, 2011, *Electronic Journal of Statistics*)

Theorem: Suppose that $\mathcal{F}$ is a $P-$measurable class of measurable functions with envelope function $F \le 1$ 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

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

Key notion: 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

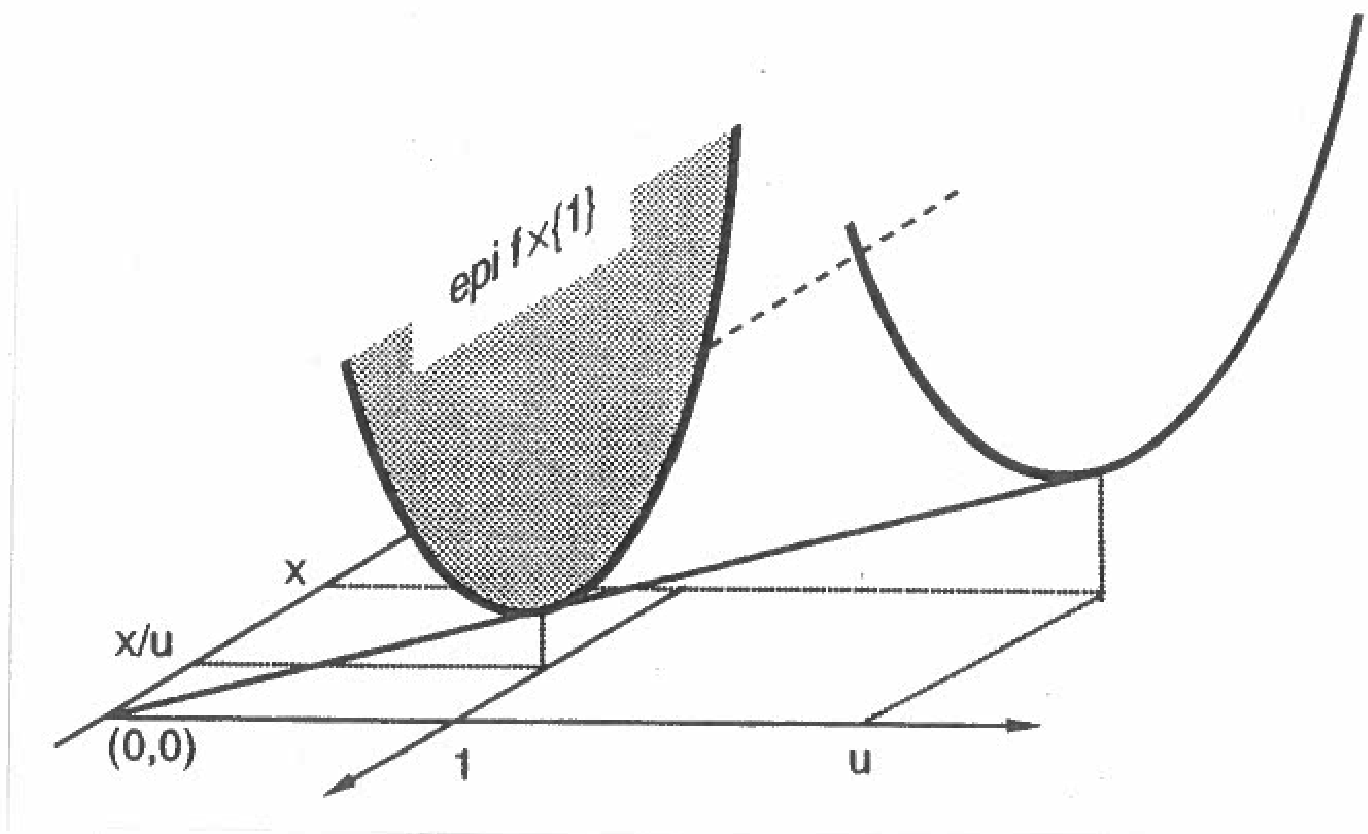
$$g(x,t) = tf(x/t),$$

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 $g(x,t) = t(x/t)^2 = x^2/t$.


epi f×{1}
x
x/u
(0,0)
1
u

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.

Proof of the new bound: This begins much as in the proof of the easy bound (3); 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

$$E_P^* \|\mathbb{G}_n\|_{\mathcal{F}} \lesssim E_P^* J\left(\frac{\sup_f (\mathbb{P}_n f^2)^{1/2}}{(\mathbb{P}_n F^2)^{1/2}}, \mathcal{F}\right) (\mathbb{P}_n F^2)^{1/2}. \tag{4}$$

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

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

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

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

is concave.

Note that $E_P \mathbb{P}_n F^2 = \|F\|_{P,2}^2$. Therefore, by Jensen's inequality applied to the right side of (4) it follows that

$$E_P^* \|\mathbb{G}_n\|_{\mathcal{F}} \lesssim J\left(\frac{\{E_P^*(\sup_f \mathbb{P}_n f^2)\}^{1/2}}{\|F\|_{P,2}}, \mathcal{F}\right) \|F\|_{P,2}. \tag{5}$$

Now since $\mathbb{P}_n(f^2) = Pf^2 + n^{-1/2}\mathbb{G}_n f^2$ and $Pf^2 \leq \delta^2 PF^2$ for all f, it follows, by using symmetrization, the contraction inequality for Rademacher random variables, de-symmetrization, and then (5), that

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

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

$$
z^2 \lesssim \delta^2 + \frac{J(z, \mathcal{F}, L_2)}{\sqrt{n}\|F\|_{P,2}}. \tag{6}
$$

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

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

Combining this with (5) completes the proof:

$$\begin{aligned} E_P^*\|\mathbb{G}_n\|_{\mathcal{F}} &\lesssim J\left(\frac{\{E_P^*(\sup_f \mathbb{P}_n f^2)\}^{1/2}}{\|F\|_{P,2}}, \mathcal{F}\right)\|F\|_{P,2} \\ &\lesssim J(\delta,\mathcal{F})\left(1+\frac{J(\delta,\mathcal{F})}{\delta^2\sqrt{n}\|F\|_{P,2}}\right)\|F\|_{P,2}. \qquad (7) \end{aligned}$$

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.

Lecture 5: Peeling methods and rates of convergence

- 1. The peeling method: first rate theorem
  - ▷ General theorem 1.
  - ▷ Application 1: Chernoff's "mode" estimator.
- 2. The peeling method: second rate theorem
  - ▷ General theorem 2.
  - ▷ Application 2: Grenander's estimator of a monotone decreasing density
  - ▷ Application 3: ??

The peeling method: first rate theorem:
Suppose that $\widehat{\theta}_n$ maximizes

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

for given measurable functions $m_\theta : \mathcal{X} \rightarrow R$ 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,

$$P m_\theta - P m_{\theta_0} \lesssim -d^2(\theta, \theta_0). \tag{8}$$

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 $\mathbb{G}_n m_\theta$ indexed by the functions m_θ.

Theorem 1. Suppose that (8) holds. If ϕ_n is a function such that $\delta \mapsto \phi_n(\delta)/\delta^\alpha$ is decreasing for some $\alpha < 2$ and

$$E \sup_{\theta: d(\theta,\theta_0)<\delta} |\mathbb{G}_n(m_\theta - m_{\theta_0})| \lesssim \phi_n(\delta), \tag{9}$$

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 (9) 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.

Proof. For simplicity, assume that $\widehat{\theta}_n = \text{argmax}\ \mathbb{M}_n(\theta)$. For each n, partition $\Theta \setminus \{\theta_0\}$ into "shells" $\{S_{n,j} : \ j \in \mathbb{Z}\}$ defined as follows:

$$S_{n,j} = \{\theta \in \Theta : 2^{j-1} < r_n d(\theta, \theta_0) \le 2^j\}, \qquad j \in \mathbb{Z}\,.$$

If $r_n d(\widehat{\theta}_n, \theta_0) > 2^M$ for some M, then $\widehat{\theta}_n \in S_{n,j}$ for some $j > M$, and hence $\sup_{\theta \in S_{n,j}}(\mathbb{M}_n(\theta) - \mathbb{M}_n(\theta_0)) \ge 0$. We want to show that

$$\limsup_{n\to\infty} P^*(r_n d(\widehat{\theta}_n, \theta_0) > 2^M) \to 0 \qquad \text{as} \qquad M \to \infty\,.$$

But

$$\begin{aligned} &P^*(r_n d(\widehat{\theta}_n, \theta_0) > 2^M) \\ &\quad\le\ P^*(r_n d(\widehat{\theta}_n, \theta_0) > 2^M,\ r_n d(\widehat{\theta}_n, \theta_0) \le 2^J) \\ &\qquad\quad + P^*(\ r_n d(\widehat{\theta}_n, \theta_0) > 2^J > \eta r_n/2) \\ &\quad\le\ \sum_{M<j\le J} P^*\left(\sup_{\theta\in S_{n,j}} (\mathbb{M}_n(\theta) - \mathbb{M}_n(\theta_0)) \ge 0\right) \\ &\qquad\quad + P^*(2d(\widehat{\theta}_n, \theta_0) \ge \eta)\,. \end{aligned}$$

Suppose that we choose η so small that the condition given by (8) holds for $d(\theta, \theta_0) \le \eta$, and the second condition (9) holds for all $\delta \le \eta$. Then, for every j in the sum, and all $\theta \in S_{n,j}$

$$\mathbb{M}(\theta) - \mathbb{M}(\theta_0) \lesssim -d^2(\theta, \theta_0) \lesssim -\frac{2^{2(j-1)}}{r_n^2}.$$

Thus in terms of the centered process $W_n(\theta) = (\mathbb{M}_n(\theta) - \mathbb{M}(\theta))$, the sum is bounded by

$$\begin{aligned}
&\sum_{M<j\le J} P^* \left(\|W_n(\theta) - W_n(\theta_0)\|_{S_{n,j}} \ge \frac{2^{2j-2}}{r_n^2} \right) \\
&\le \sum_{M<j\le J} \frac{\phi_n(2^j/r_n)}{\sqrt{n}2^{2j-2}} r_n^2 \\
&\le \sum_{M<j\le J} \frac{2^{j\alpha}\phi_n(1/r_n) r_n^2}{\sqrt{n}} 2^{-(2j-2)} \\
&\le 4 \sum_{j>M} 2^{j\alpha - 2j} \to 0 \qquad \text{as} \quad M \nearrow \infty;
\end{aligned}$$

where we used the definition of r_n together with

$$\frac{\phi_n(c\delta)}{(c\delta)^\alpha} \le \frac{\phi_n(\delta)}{\delta^\alpha}$$

for $c > 1$ to conclude that $\phi_n(c\delta) \le c^\alpha \phi_n(\delta)$. $\square$

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 (3) 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 (8). This will typically hold for unimodal densities since

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

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,

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

Remark: Chernoff (1964) shows that

- $n^{1/3}(\widehat{\theta}_n - \theta_0) \to_d \left(\frac{8p(\theta_0)}{c}\right)^{1/3} Z$ where

$$c \equiv p'(\theta_0 - a) - p'(\theta_0 + a).$$

- $Z = \text{argmax}\{W(t) - t^2\}$ for a two-sided Brownian motion process W; see also VdV&W, page 295.

Here is a simple result that handles many parametric examples. This formulation is from Van der Vaart (1998).

Corollary. Suppose that $x \mapsto m_\theta(x)$ is a measurable function for each $\theta \in \Theta \subset \mathbb{R}^d$ where Θ is open, and suppose that for all θ_1, θ_2 in some neighborhood of $\theta_0 \in \Theta$ there is a measurable function $\dot{m} \in L_2(P)$ such that

$$|m_{\theta_1}(x) - m_{\theta_2}(x)| \leq \dot{m}(x) \|\theta_1 - \theta_2\| \,. \tag{10}$$

Furthermore, suppose that the function $\theta \mapsto M(\theta) = P m_\theta$ has a second-order Taylor expansion at the point of maximum θ_0 with nonsingular second derivative. If $\mathbb{M}_n(\widehat{\theta}_n) \geq \mathbb{M}_n(\theta_0) - O_p(n^{-1})$, then $\sqrt{n}(\widehat{\theta}_n - \theta_0) = O_p(1)$ provided that $\widehat{\theta}_n \to_p \theta_0$.

Proof. The hypothesis (8) holds with the metric d replaced by the Euclidean distance. To verify (9), we apply the bracketing moment bound (2) to the class of functions $\mathcal{M}_\delta = \{m_\theta - m_{\theta_0} : \|\theta - \theta_0\| < \delta\}$. This class has envelope $M_\delta = \dot{m}\delta$, so that

$(PM_\delta^2)^{1/2} = \delta\|\dot{m}\|_{P,2}$. By Lemma 1.6.2 and Exercise 1.3.18 it follows that

$$\begin{aligned} N_{[\,]}(2\epsilon\|\dot{m}\|_{P,2}, \mathcal{M}_\delta, L_2(P)) &\le N(\epsilon, B(\theta_0,\delta), \|\cdot\|) \\ &\le \left(\frac{6\delta}{\epsilon}\right)^d, \end{aligned}$$

or

$$\begin{aligned} N_{[\,]}(\epsilon\delta\|\dot{m}\|_{P,2}, \mathcal{M}_\delta, L_2(P)) &= N_{[\,]}(\epsilon\|M_\delta\|_{P,2}, \mathcal{M}_\delta, L_2(P)) \\ &\le \left(\frac{12}{\epsilon}\right)^d, \end{aligned}$$

and hence,

$$\begin{aligned} J_{[\,]}(1, \mathcal{M}_\delta, L_2(P)) &\lesssim \sqrt{d}\int_0^1 \sqrt{\log\left(\frac{12}{\epsilon}\right)}d\epsilon \\ &= 12\sqrt{d}\int_{\log(12)}^\infty v^{1/2}e^{-v}dv < \infty\,. \end{aligned}$$

Thus we conclude that (9) holds with $\phi_n(\delta) \lesssim \delta$. Thus Theorem 1 yields the rate of convergence $r_n = \sqrt{n}$ if $\widehat{\theta}_n$ is consistent. $\square$

The peeling method: second rate theorem:

Theorem 2. (Birgé and Massart, 1993).
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

$$E^*_{P_0}\|\mathbb{G}_n\|_{\mathcal{M}_\delta} \lesssim \tilde{J}_{[\,]}(\delta, \mathcal{P}, h)\left(1 + \frac{\tilde{J}_{[\,]}(\delta, \mathcal{P}, h)}{\delta^2\sqrt{n}}\right) \equiv \phi_n(\delta)\,. \qquad (11)$$

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

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

Proof of the last part of the Theorem statement: Note that the Hellinger equivalence noted in Proposition 1.3 (Lecture 2) followed by Proposition 1.3 yields

$$
\begin{aligned}
h^2(\widehat{p}_n, p_0) &\leq 24 h^2((\widehat{p}_n + p_0)/2, p_0) \\
&\leq 24(\mathbb{P}_n - P_0)\left(\frac{1}{2}\log\left(\frac{\widehat{p}_n + p_0}{2p_0}\right) 1_{[p_0>0]}\right) \\
&= 12(\mathbb{P}_n - P_0)(m_{\widehat{p}_n} - m_{p_0}) \\
&\leq 12 n^{-1/2} \|\mathbb{G}_n\|_{\mathcal{M}_\delta}
\end{aligned}
$$

on the event $\{h(\widehat{p}_n, p_0) \leq \delta\}$. Thus for $x > 0$ and $\delta > 0$

$$
\begin{aligned}
&P(r_n h(\widehat{p}_n, p_0) > x) \\
&\quad = P(r_n h(\widehat{p}_n, p_0) > x, h(\widehat{p}_n, p_0) \le \delta) \\
&\qquad\qquad + P(r_n h(\widehat{p}_n, p_0) > x, h(\widehat{p}_n, p_0) > \delta) \\
&\quad \le P(r_n h(\widehat{p}_n, p_0) > x, h(\widehat{p}_n, p_0) \le \delta) + P(h(\widehat{p}_n, p_0) > \delta) \\
&\quad \le \frac{E^*[r_n^2 h^2(\widehat{p}_n, p_0) 1\{h(\widehat{p}_n, p_0) \le \delta\}]}{x^2} + P(h(\widehat{p}_n, p_0) > \delta) \\
&\quad \le \frac{12 n^{-1/2} r_n^2 E^* \|\mathbb{G}_n\|_{\mathcal{M}_\delta}}{x^2} + P(h(\widehat{p}_n, p_0) > \delta)
\end{aligned}
$$

Choosing $x = 2^j$, $\delta = 2^{j+1}/r_n$ and assuming that $E^*\|\mathbb{G}_n\|_{\mathcal{M}_\delta} \le \phi_n(\delta)$, this yields

$$
\begin{aligned}
&P(r_n h(\widehat{p}_n, p_0) > 2^j) \\
&\quad \le \frac{12 n^{-1/2} r_n^2 \phi_n(2^{j+1}/r_n)}{2^{2j}} + P(h(\widehat{p}_n, p_0) > 2^{j+1}/r_n)
\end{aligned}
$$

for all j. But then by recursion, the bound $\phi_n(c\delta) \le c^\alpha \phi_n(\delta)$ used in the proof of Theorem 3.1, and choosing

J so large that $2^{J+1}/r_n > \eta$, we find that

$$
\begin{aligned}
& P(r_n h(\widehat{p}_n, p_0) > 2^M) \\
& \le \ 12 \sum_{j=M}^{J} \frac{n^{-1/2} r_n^2 \phi_n(2^{j+1}/r_n)}{2^{2j}} + \ P(h(\widehat{p}_n, p_0) > 2^{J+1}/r_n) \\
& \le \ 12 \sum_{j=M}^{J} \frac{r_n^2 2^{j\alpha} \phi_n(1/r_n)}{\sqrt{n}} 2^{-2j} + \ P(h(\widehat{p}_n, p_0) > \eta) \\
& \le \ 12 \sum_{j=M}^{\infty} 2^{j\alpha - 2j} + P(h(\widehat{p}_n, p_0) > \eta) \, .
\end{aligned}
$$

(Note that we could have summed out to J so large that $2^{J+1}/r_n > 1$, and then the second term on the ride side is zero since $h(p,q) \le 1$; thus consistency of $\widehat{p}_n$ is not needed in this case.) By consistency of $\widehat{p}_n$ this yields

$$
\lim_{M\to\infty} \limsup_{n\to\infty} P(r_n h(\widehat{p}_n, p_0) > 2^M) = 0 \, .
$$

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

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

Let Q denote the uniform distribution on $[0, M]$. Then

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

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

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

Lecture 6: Some Useful Preservation Theorems

- Preservation of the VC property.
- Preservation of Euclidean classes.
- Preservation of Bracketing properties.
- Preservation of the Glivenko-Cantelli property.
- Preservation of the Donsker property.
- Preservation of the Donsker property under ... ??? .

Preservation of the VC property:

Proposition 6.1. (Operations preserving the VC property for sets). Suppose that $\mathcal{C}$ and $\mathcal{D}$ are VC-classes of subsets of a set $\mathcal{X}$, and that $\phi : \mathcal{X} \mapsto \mathcal{Y}$ and $\psi : \mathcal{Z} \mapsto \mathcal{X}$ are fixed functions. Then:

- $\mathcal{C}^c = \{C^c : C \in \mathcal{C}\}$ is VC and $S(\mathcal{C}^c) = S(\mathcal{C})$.
- $\mathcal{C} \sqcap \mathcal{D} = \{C \cap D : C \in \mathcal{C}, D \in \mathcal{D}\}$ is VC.
- $\mathcal{C} \sqcup \mathcal{D} = \{C \cup D : C \in \mathcal{C}, D \in \mathcal{D}\}$ is VC.
- $\mathcal{C} \times \mathcal{D} = \{C \times D : C \in \mathcal{C}, D \in \mathcal{D}\}$ is VC for VC-classes $\mathcal{C}$ and $\mathcal{D}$ in sets $\mathcal{X}$ and $\mathcal{Y}$.
- $\phi(\mathcal{C})$ is VC if ϕ is one-to-one.
- $\psi^{-1}(\mathcal{C})$ is VC and $S(\psi^{-1}(\mathcal{C})) \le S(\mathcal{C})$ with equality if ψ is onto.
- The sequential closure of $\mathcal{C}$ for pointwise convergence of indicator functions is VC.

Can we quantify or bound $S(\mathcal{C} \sqcap \mathcal{D})$, $S(\mathcal{C} \sqcap \mathcal{D})$, $S(\mathcal{C} \times \mathcal{D})$ in terms of $S(\mathcal{C})$ and $S(\mathcal{D})$? Yes! Even more generally:

Proposition. Let $S \equiv \sum_{j=1}^m S_j$. Then the following bounds hold:

$$\left\{ \begin{array}{l} S(\sqcup_{j=1}^m \mathcal{C}_j) \\ S(\sqcap_{j=1}^m \mathcal{C}_j) \\ S(\boxtimes_1^m \mathcal{D}_j) \end{array} \right\} \le c_1 S \log\left(\frac{c_2 m}{e^{Ent(\underline{S})/\overline{S}}} \right) \le c_1 S \log(c_2 m). \qquad (12)$$

where $\underline{V} \equiv (S_1, \ldots, S_m)$, $c_1 \equiv \frac{e}{(e-1)\log(2)} \doteq 2.28231...$, $c_2 \equiv \frac{e}{\log 2} \doteq 3.92165...$,

$$Ent(\underline{S}) \equiv m^{-1} \sum_{j=1}^m S_j \log S_j - \overline{S} \log \overline{S}$$

is the "entropy" of the S_j's under the discrete uniform distribution with weights $1/m$ and $\overline{S} = m^{-1} \sum_{j=1}^m S_j$.

Proof: VdV&W, 2009, HD Prob V.

Proposition 6.2. (Operations preserving the VC-subgraph property for functions). Suppose that $\mathcal{F}$ and $\mathcal{G}$ are VC-subgraph classes of functions on a set $\mathcal{X}$, and $g : \mathcal{X} \mapsto R$, $\phi : R \mapsto R$, and $\psi : \mathcal{Z} \mapsto \mathcal{X}$ fixed functions. Then:

- $\mathcal{F} \wedge \mathcal{G} = \{f \wedge g : f \in \mathcal{F}, g \in \mathcal{G}\}$ is VC subgraph;
- $\mathcal{F} \vee \mathcal{G} = \{f \vee g : f \in \mathcal{F}, g \in \mathcal{G}\}$ is VC subgraph;
- $\{\mathcal{F} > 0\} = \{\{f > 0\} : f \in \mathcal{F}\}$ is VC;
- $-\mathcal{F}$ is VC-subgraph;
- $g + \mathcal{F} = \{g + f : f \in \mathcal{F}\}$ is VC subgraph;
- $g \cdot \mathcal{F} = \{g \cdot f : f \in \mathcal{F}\}$ is VC subgraph;
- $\mathcal{F} \circ \psi = \{f(\psi) : f \in \mathcal{F}\}$ is VC subgraph;
- $\phi \circ \mathcal{F} = \{\phi(f) : f \in \mathcal{F}\}$ is VC subgraph for monotone ϕ.

Euclidean classes: It is sometimes easier to work with the following notion: a class of real functions on a set $\mathcal{X}$ is said to be *a Euclidean class* for the envelope function F if there exist constants A and V such that

$$N(\epsilon\|F\|_{Q,1}, \mathcal{F}, L_1(Q)) \leq A\epsilon^{-V}, \qquad 0 < \epsilon \leq 1$$

whenever $0 < \|F\|_{Q,1} = QF < \infty$. Note that the constants A and V may not depend on Q.

If $\mathcal{F}$ is Euclidean, then for each $r > 1$

$$N(\epsilon\|F\|_{Q,r}, \mathcal{F}, L_r(Q)) \leq A2^{rV}\epsilon^{-rV}, \qquad 0 < \epsilon \leq 1$$

whenever $0 < QF^r < \infty$, as follows from the definition of $N(2(\epsilon/2)^r\|F\|_{\mu,1}, \mathcal{F}, L_1(\mu))$ for the measure $\mu(\cdot) = Q(\cdot(2F)^{r-1})$.

Here is an example of a preservation or stability result for Euclidean classes:

Proposition 6.3. Suppose that $\mathcal{F}$ and $\mathcal{G}$ are Euclidean classes of functions with envelopes F and G respectively, and suppose that Q is a measure with $QF^r < \infty$ and $QG^r < \infty$ for some $r \geq 1$. Then the class of functions

$$\mathcal{F} + \mathcal{G} = \{f + g : f \in \mathcal{F},\, g \in \mathcal{G}\}$$

is Euclidean for the envelope $F + G$; moreover,

$$\begin{aligned} & N((2\epsilon + 2\delta)\|F + G\|_{Q,r}, \mathcal{F} + \mathcal{G}, L_2(Q)) \\ & \quad \leq \; N(\epsilon\|F\|_{Q,r}, \mathcal{F}, L_r(Q))N(\delta\|G\|_{Q,r}, \mathcal{G}, L_r(Q))\,. \end{aligned}$$

Here are two specific results concerning affine transformations of $\mathbb{R}^d$ and then composition with a fixed function of bounded variation.

Lemma 6.1.

- Suppose that $\psi : \mathbb{R}^+ \mapsto \mathbb{R}$ is of bounded variation. For A an $m \times d$ matrix and $b \in \mathbb{R}^m$, let $f_{A,b} : \mathbb{R}^d \mapsto \mathbb{R}$ be defined by $f_{A,b}(x) = \psi(|Ax + b|)$. Then the collection

$$\mathcal{F} = \{f_{A,b} : A \text{ an } m \times d \, \text{matrix}, b \in \mathbb{R}^m\}$$

  is Euclidean for a constant envelope $F = \|\psi\|_\infty$.

- Suppose that $\psi : \mathbb{R} \mapsto \mathbb{R}$ is of bounded variation. For $a \in \mathbb{R}^d$, $b \in \mathbb{R}$, let $g_{a,b} : \mathbb{R}^d \mapsto \mathbb{R}$ be defined by $f_{a,b}(x) = \psi(a'x + b)$. Then the collection $\mathcal{G} = \{g_{a,b} : \; a \in \mathbb{R}^d, b \in \mathbb{R}\}$ is Euclidean for a constant envelope $F = \|\psi\|_\infty$.

Preservation of bracketing entropy:

Proposition 6.4. Suppose that for every θ a compact subset Θ of $\mathbb{R}^d$, the class $\mathcal{F}_\theta = \{f_{\theta,\gamma} : \gamma \in \Gamma\}$ satisfies

$$\log N_{[\,]}(\epsilon, \mathcal{F}_\theta, L_2(P)) \leq K \left(\frac{1}{\epsilon}\right)^W$$

for a constant $W < 2$ and K not depending on θ. Suppose, moreover, that

$$|f_{\theta_1,\gamma} - f_{\theta_2,\gamma}| \leq F|\theta_1 - \theta_2|$$

for a function F with $PF^2 < \infty$. Then $\mathcal{F} \equiv \cup_{\theta\in\Theta}\mathcal{F}_\theta$ satisfies

$$\log N_{[\,]}(\epsilon, \mathcal{F}, L_2(P)) \lesssim d\log(1/\epsilon) + K \left(\frac{1}{\epsilon}\right)^W .$$

Preservation of the Glivenko-Cantelli property:

Proposition 6.5. (Operations preserving the Glivenko-Cantelli-subgraph property).

- If $\mathcal{F}_1, \ldots, \mathcal{F}_k$ are $P-$Glivenko-Cantelli classes, and $\varphi : \mathbb{R}^k \to \mathbb{R}$ is continuous, then $\mathcal{H} \equiv \varphi(\mathcal{F}_1, \ldots, \mathcal{F}_k)$ is Glivenko-Cantelli provided it has a $P-$integrable envelope.
- If $\{\mathcal{X}_j\}_{j=1}^{\infty}$ is a partition of $\mathcal{X}$ and $\mathcal{F}_j \equiv \{f 1_{\mathcal{X}_j} : \ f \in \mathcal{F}\}$ is $P-$Glivenko-Cantelli for each j and $\mathcal{F}$ has an integrable envelope function F, then $\mathcal{F}$ is $P-$Glivenko-Cantelli.
- ??

Preservation of the Donsker property:

Proposition 6.6. (Operations preserving the Donsker property).

- If $\mathcal{F}$ is Donsker and $\mathcal{G} \subset \mathcal{F}$ is Donsker, then $\mathcal{G}$ is Donsker.
- If $\mathcal{F}$ is Donsker, then $\overline{\mathcal{F}}$ is Donsker where $\overline{\mathcal{F}}$ denotes the set of all $f : \mathcal{X} \to \mathbb{R}$ for which there exists a sequence $\{f_m\} \subset \mathcal{F}$ with $f_m \to f$ both pointwise and in $L_2(P)$.
- If $\mathcal{F}$ is Donsker then sconv($\mathcal{F}$) is Donsker where sconv($\mathcal{F}$) denotes the set of convex combinations $\sum_{i=1}^{\infty} \lambda_i f_i$ of functions f_i in $\mathcal{F}$ where $\sum |\lambda_i| \leq 1$ and the series converges both pointwise and in $L_2(P)$.

- Let $\mathcal{F}_1, \ldots, \mathcal{F}_k$ be Donsker classes with $\|P\|_{\mathcal{F}_i} < \infty$ for each i. Let $\phi : \mathbb{R}^k \to \mathbb{R}$ satisfy

$$|\phi(f(x)) - \phi(g(x))|^2 \leq \sum_{l=1}^{k} (f_i(x) - g_i(x))^2$$

for every $f, g \in \mathcal{F}_1, \ldots, \mathcal{F}_k$ and x. (This holds if ϕ is Lipschitz.) Then the class $\phi(\mathcal{F}_1, \ldots, \mathcal{F}_k)$ is Donsker provided that $\phi(f_1, \ldots, f_k)$ is square integrable for at least one $(f_1, \ldots, f_k)$.

- If $\mathcal{F}$ and $\mathcal{G}$ are Donsker classes with $\|P\|_{\mathcal{F}\cup\mathcal{G}} < \infty$, then $\mathcal{F} \wedge \mathcal{G}$, $\mathcal{F} \vee \mathcal{G}$, $\mathcal{F} + \mathcal{G}$, and $\mathcal{F} \cup \mathcal{G}$ are all Donsker.

Useful things not covered

- Multiplier central limit theorems.
- Bootstrap central limit theorems.
- $M-$ and $Z-$theorems.
- Concentration inequalities
  (with applications to model selection).

See van der Vaart and Wellner (1996) ...

and

VdV & W Second Edition (2013?)

Thank You!