Local Empirical and U–Statistic Processes and Applications

Lectures Notes

David M. Mason
University of Delaware

A good source for any result in these notes for which a reference is not provided is the monograph: *From dependence to independence. Randomly stopped processes. U-statistics and processes. Martingales and beyond* by de la Peña and Giné (1999). The author thanks Julia Dony for numerous corrections and helpful suggestions.

Local empirical process

Kernel Density Estimator

Let X, X_i, $i \in \mathbb{N}$, be i.i.d. random variables taking values in $(S, \mathcal{S})$, a measure space. Let

$$g : S \mapsto \mathbb{R}^d, 1 \leq d < \infty,$$

be a measurable function. Assume that $g(X)$ has density f_g.

Let $K : \mathbb{R}^d \mapsto \mathbb{R}$ be a kernel, meaning that

$$\int_{\mathbb{R}^d} K(x)dx = 1 \text{ and } 0 < \int_{\mathbb{R}^d} K^2(x)dx = ||K||_2^2 < \infty.$$

For $t \in \mathbb{R}^d$ and $\lambda \in [a, b]$, $0 < a \leq b < \infty$ and for $0 < h < 1$, we define the *kernel density estimator*

$$f_n(t, \lambda h) = \frac{1}{n\lambda h} \sum_{i=1}^{n} K_{\lambda h}\left(t - g\left(X_i\right)\right),$$

where

$$K_h\left(\cdot\right) = h^{-1} K\left(\cdot h^{-1/d}\right).$$

The conditions $h_n > 0$, $h_n \to 0$ and $nh_n \to \infty$ make $f_n(t, \lambda h_n)$ a pointwise consistent estimator of $f_g\left(t\right)$ at any continuity point t of f_g for any $a \leq \lambda \leq b$.

Local Empirical Process

We define the *local empirical process*

$$u_n\left(t, \lambda\right) := \sqrt{n}\left\{f_n(t, \lambda h_n) - \mathbf{E} f_n(t, \lambda h_n)\right\}.$$

It is easy to prove that, subject to smoothness conditions on f, for each $t \in \mathbb{R}^d$,

$$\sqrt{h_n} u_n\left(t\right) := \sqrt{h_n} u_n\left(t, 1\right) = \sqrt{n h_n}\left\{f_n(t, h_n) - \mathbf{E} f_n(t, h_n)\right\}$$

$$\to_d N\left(0, ||K||_2^2 f\left(t\right)\right),$$

whereas for any choice of $t_1 \neq t_2$ the random variables

$$\sqrt{h_n} u_n\left(t_1\right) \text{ and } \sqrt{h_n} u_n\left(t_2\right)$$

are asymptotically independent. This means that $\sqrt{h_n}u_n$ cannot converge weakly to a continuous bounded process on any non-trivial subset of $\mathbb{R}^d$.

A Law of the Logarithm

Let

$$\mathcal{K} = \left\{K\left(t - \cdot\gamma\right) : \gamma \geq 1, t \in \mathbb{R}^d\right\}.$$

Assume that

(F.i) $\mathcal{K}$ is a bounded point–wise measurable class.

(F.ii) $\mathcal{K}$ is VC for some $A \geq 3$ and $v \geq 1$.

Also assume that K has support in $[-1/2, 1/2]^d$;
f_g is uniformly continuous on $\mathbb{R}^d$;
$\{h_n\}_{n\geq 1}$ converges to zero at the rate:

(H.i) $h_n \searrow 0$, $nh_n \nearrow \infty$; (H.ii) $nh_n/\log\left(1/h_n\right) \to \infty$;

(H.iii) $\log(1/h_n)/\log\log n \to \infty$.

The conditions (H.i), (H.ii) and (H.iii) are sometimes called the Csörgő and Révész (1979) and Stute (1982) conditions.

For definitions of (F.i) and (F.ii) refer to the Appendix.

Theorem 1 of Mason (2004) implies that a.s.

$$\lim_{n\to\infty} \sup_{a\leq\lambda\leq b} \sup_{t\in\mathbb{R}^d} \frac{\sqrt{\lambda h_n}\left|u_n\left(t,\lambda\right)\right|}{\sqrt{2\log\left(1/h_n\right)}} = ||K||_2 \sup_{t\in\mathbb{R}^d} \sqrt{f_g\left(t\right)}. \tag{T}$$

This is a *uniform in bandwidth* version of a result of Giné and Guillou (2002).

Empirical Processes

Let X, X_i, $i \in \mathbb{N}$, be i.i.d. random variables taking values in $(S, \mathcal{S})$, a measure space. Consider a class $\mathcal{F}$ of bounded functions from S to $\mathbb{R}$. The *empirical process* $(\alpha_n(f))_{f\in\mathcal{F}}$ indexed by $\mathcal{F}$ is defined to be

$$\alpha_n(f) = \sum_{i=1}^{n}(f(X_i) - \mathbf{E}f(X_1))/\sqrt{n},\ f \in \mathcal{F}.$$

Introduce the class of functions on S,

$$\mathcal{F}_n = \left\{z \in S \to \frac{1}{\sqrt{\lambda}}K\left(\frac{t - g\left(z\right)}{\left(\lambda h_n\right)^{1/d}}\right) : t \in \mathbb{R}^d,\ a \leq \lambda \leq b\right\}.$$

We see that

$$\sup_{a\leq\lambda\leq b} \sup_{t\in\mathbb{R}^d} \frac{\sqrt{\lambda h_n}\left|u_n\left(t,\lambda\right)\right|}{\sqrt{2\log\left(1/h_n\right)}} = \frac{\sup_{\varphi\in\mathcal{F}_n}\left|\alpha_n(\varphi)\right|}{\sqrt{2h_n\log\left(1/h_n\right)}}.$$

This places us in the realm of empirical process theory, which enables us to bring two powerful tools into play.

Armed with these tools we shall sketch why (T) holds.

TOOL 1: Inequality (Talagrand, (1996))

Let $\mathcal{F}$ be a (countable) class of functions $f : \mathcal{X} \to \mathbb{R}$ and assume that $\exists M > 0 : |f| \leq M,\ f \in \mathcal{F}$. Then there exist absolute constants $A_1,\ A_2 > 0$ so that we have for $t \geq 0$:

$$P\{||\alpha_n||_{\mathcal{F}} \geq A_1(\mathbf{E}||\alpha_n||_{\mathcal{F}} + t)\} \leq 2\{\exp(-A_2 t^2/\sigma_{\mathcal{F}}^2) + \exp(-A_2\sqrt{n}t/M)\},$$

where $\sigma_{\mathcal{F}}^2 = \sup_{f\in\mathcal{F}} \mathrm{Var}(f(X_1))$.

TOOL 2: A good bound for $\mathbf{E}||\alpha_n||_{\mathcal{F}}$

Proposition 1 (Moment Bound). (Einmahl and Mason (2000))

Let $\mathcal{F}$ be a class of functions $f : \mathcal{X} \to \mathbb{R}$ and let F be defined by

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

Assume that there are constants $C,\ \nu \geq 1$ and $0 < \sigma \leq \beta$ so that

(i) $N(\epsilon, \mathcal{F}) \leq C\epsilon^{-\nu}, 0 < \epsilon < 1$.

(ii) $\sup_{f\in\mathcal{F}} \mathbf{E}f^2(X_1) \leq \sigma^2$.

(iii) $\mathbf{E}F^2(X_1) \leq \beta^2$.

(iv) $\sup_{f\in\mathcal{F}} \|f\|_\infty \leq 1/(4\sqrt{\nu})\sqrt{n\sigma^2/\log(C_1\beta/\sigma)}$,

where $C_1 = C^{1/\nu} \vee e$.

Then we have for an absolute constant $A_3 > 0$

$$\mathbf{E}||\alpha_n||_{\mathcal{F}} \leq A_3\sqrt{\nu\sigma^2\log(C_1\beta/\sigma)}.$$

Combining these Two Results

We set $\mathcal{F} = \mathcal{F}_n$. Then it follows from Talagrand's inequality and a blocking argument that with probability one,

$$\|\alpha_n\|_{\mathcal{F}_n} = O\left(\sigma_n\sqrt{\log\log n} \vee \mathbf{E}\|\alpha_n\|_{\mathcal{F}_n}\right),$$

where $\sigma_n^2 = O(h_n)$ if f_g is bounded.

Our moment inequality implies that

$$\mathbf{E}\|\alpha_n\|_{\mathcal{F}_n} = O\left(\sqrt{h_n\log(1/h_n)}\right),$$

provided that $h_n \geq c\log n/n$. Thus when (H.ii) holds, that is

$$nh_n/\log(1/h_n) \to \infty,$$

we get that

$$\|\alpha_n\|_{\mathcal{F}_n} = O\left(\sqrt{h_n\log(1/h_n)}\right).$$

These tools also lead to the following uniform in bandwidth consistency result for the kernel density estimator.

Theorem (Einmahl-Mason 2005) Assume that the density f_g of $g(X)$ is bounded and let K be a bounded kernel function. Then we have with probability 1:

$$\limsup_{n\to\infty} \sup_{\frac{c\log n}{n}\le h\le 1} \frac{\sqrt{nh}\|f_n(t,h)-\mathbf{E}f_n(t,h)\|_\infty}{\sqrt{\log(1/h)\vee\log\log n}} < \infty.$$

(Here and elsewhere the supremum $\|\cdot\|_\infty$ is in the t variable.)

Corollary Let $\hat{h}_n = H_n(X_1,\ldots,X_n)$ be chosen so that with prob. 1:

$$\liminf_n \frac{\hat{h}_n}{a_n} > 0,$$

where

$$na_n/\log n \to \infty.$$

Then we have with probability 1,

$$\|f_n(t,\hat{h}_n)-\mathbf{E}f_n(t,\hat{h}_n)\|_\infty = O\left(\sqrt{\frac{\log(1/a_n)\vee\log\log n}{na_n}}\right) = o(1).$$

More precise results on the behavior of $\|f_n(t,\hat{h}_n)-\mathbf{E}f_n(t,\hat{h}_n)\|_\infty$ using certain plug–in–type data dependent bandwidth selectors are to be found in Deheuvels and Mason (2004).

Local U-Statistics

Frees (1994) made the interesting observation that, under natural, relatively weak hypotheses, if f_g is the density of a function of two or more sample variables, $g(X_1,\ldots,X_m)$, then the *local U-statistic*

$$\frac{(n-m)!}{n!h_n}\sum_{I_n^m} K\left(\frac{t-g(X_{i_1},\ldots,X_{i_m})}{h_n^{1/d}}\right),$$

with $t\in\mathbb{R}^d$ and

$$I_n^m = \{\mathbf{i}=(i_1,\ldots,i_m): 1\le i_j\le n, i_j\ne i_k \ \text{ if } \ j\ne k\},$$

estimates $f_g(t)$ for each fixed $t\in\mathbb{R}^d$, at the rate $O_P(1/\sqrt{n})$.

He motivates his study by considering as examples of g, the inter-point distance between spatial observations, and sums of independent insurance claims (in this last case, f_g is a convolution).

Schick and Wefelmeyer (2004) obtain the $O_P(1/\sqrt{n})$ rate for the sup norm and the L_1 norm in the case

$$g(X_1,\ldots,X_m) = u_1(X_1)+\cdots+u_m(X_m),$$

where the u_i are real functions. They obtain functional central limit theorems when the kernels are themselves convolutions.

Example 1: **Linear combinations**. (Frees (1994), Schick and Wefelmeyer (2004))

Suppose that

$$g(x_1,\ldots,x_m)=\sum_{i=1}^{m}u_i(x_i),\quad x_i\in S,$$

for measurable functions $u_1,\ldots,u_m$ from S to $\mathbb{R}^d$ such that the random variable $u_i(X)$ has a density f_i for each $i=1,\ldots,m$. In this case we consider the local U–statistic estimator of the density of $g(X_1,\ldots,X_m)$ given by

$$\frac{(n-m)!}{n!}\sum_{\mathbf{i}\in\mathbf{I_n^m}}K_{\lambda h_n}\left(t-\sum_{r=1}^{m}u_r(X_{i_r})\right).$$

Example 2: Local inter–point distance processes (Jammalamadaka and Janson (1986), Eastwood and Horváth (1999))

Here we get the local $U-$statistic

$$\frac{1}{\lambda h_n n(n-1)}\sum_{i\neq j}^{n}I\left\{X_i-X_j\in(\lambda h_n)^{1/d}D\right\},$$

where $X_1,\ldots,X_n$ are i.i.d. in $\mathbb{R}^d$ and D is the unit ball. In this case

$$m=2,\ S=\mathbb{R}^d \text{ and } g\left(X_1,X_2\right)=X_1-X_2.$$

Closely related to this is the kernel density estimator of the density f of the *interpoint distribution*

$$F\left(t\right)=P\left\{\left|X_1-X_2\right|\leq t\right\}$$

proposed by Frees (1994)

$$\frac{1}{\lambda h_n n(n-1)}\sum_{i\neq j}^{n}K\left(\frac{t-|X_i-X_j|}{\lambda h_n}\right).$$

Example 3: Integral of the Square of the Density

Assuming X has cdf F with density f, a frequently used estimator of

$$T\left(F\right):=\int_{\mathbb{R}}f^2(x)\,dx=\mathbf{E}f\left(X\right)$$

is the *Local U-statistic*

$$T_n(F,\lambda h_n)=\frac{1}{n\left(n-1\right)}\frac{1}{\lambda h_n}\sum_{i\neq j}^{n}K\left(\frac{X_i-X_j}{\lambda h_n}\right),$$

which when

$$g\left(x,y\right)=x-y \text{ and } t=0$$

gives

$$T_n(F,\lambda h_n)=U_n(0,\lambda),$$

where for any t

$$U_n(t,\lambda) := \frac{(n-2)!}{n!}\frac{1}{\lambda h_n}\sum_{\mathbf{i}\in I_n^2} K\left(\frac{t-g(X_{i_1},X_{i_2})}{\lambda h_n}\right).$$

Local U-Statistic Process

Nearly all of the results in the remainder of these notes are taken from Giné and Mason (2006a,b,c).

Let X, X_i, $i \in \mathbb{N}$, be i.i.d. random variables taking values in $(S, \mathcal{S})$, a measure space. Further for $m \geq 1$ let

$$g : S^m \mapsto \mathbb{R}^d, 1 \leq d < \infty,$$

be a measurable function and $K : \mathbb{R}^d \mapsto \mathbb{R}$ be a kernel with *not necessary compact support.* For $t \in \mathbb{R}^d$ and $\lambda \in [a,b]$, $0 < a \leq b < \infty$, we define the *local U–statistic*

$$U_n(t,\lambda) := \frac{(n-m)!}{n!}\sum_{\mathbf{i}\in I_n^m} K_{\lambda h_n}(t - g(X_{i_1},\ldots,X_{i_m})),$$

with

$$I_n^m = \{\mathbf{i} = (i_1,\ldots,i_m) : 1 \leq i_j \leq n, i_j \neq i_k \ \text{ if } \ j \neq k\}.$$

Define the *local U–statistic process* for $t \in \mathbb{R}^d$, $\lambda \in [a,b]$, $0 < a \leq b < \infty$, to be

$$u_n(t,\lambda) := \sqrt{n}\left\{U_n(t,\lambda) - \mathbf{E}K_{\lambda h_n}(t - g(X_1,\ldots,X_m))\right\}.$$

Uniform CLT

We always assume $g(X_1,\ldots,X_m)$ has density f_g and that X, X_i, $i \in \mathbb{N}$, are i.i.d. with common law P such that for each $i = 1,\ldots,m$, the random variable

$$g(X_1,\ldots,X_m), \text{ conditionally on } X_i = x,$$

has a density $f_i(t,x)$, jointly measurable in t and x, satisfying for each $x \in IR^d$, $f_i(\cdot,x) \in L_\infty\left(\mathbb{R}^d\right)$.

This condition is not always satisflied. For instance, it does not hold for the function

$$g(X_1,\ldots,X_m) = (X_1,\ldots,X_m).$$

Definition: CLT for $u_n(t,\lambda)$ uniformly in $a \leq \lambda \leq b$.

For $m \geq 2$, we say that the processes $u_n(t,\lambda)$ satisfy the CLT uniformly in $a \leq \lambda \leq b$, if

$$\sup_{\lambda\in[a,b]}\left\| u_n(t,\lambda) - \alpha_n\left(\sum_{i=1}^m f_i(t,\cdot)\right)\right\|_\infty \to 0$$

and

$$\alpha_n\left(\sum_{i=1}^m f_i(t,\cdot)\right)$$

converges weakly in the sense of an empirical process indexed by the class of functions

$$\mathcal{G} = \left\{ \sum_{i=1}^{m} f_i(t, \cdot) : t \in \mathbb{R}^d \right\}$$

to a Gaussian process $\mathcal{Z}$ indexed by the class $\mathcal{G}$.

Definition: Here is what we mean by weak convergence of α_n.

Let $l^\infty(\mathcal{F})$ denote the space of bounded functions on $\mathcal{F}$. We equip $l^\infty(\mathcal{F})$ with the supremum norm. Clearly $\alpha_n \in l^\infty(\mathcal{F})$. We say that α_n converges weakly uniformly in $f \in \mathcal{F}$ to a Gaussian process $\mathcal{Z}$ indexed by the class $\mathcal{F}$ taking values in $l^\infty(\mathcal{F})$ if for all functions

$$H : l^\infty(\mathcal{F}) \to \mathbb{R}, \text{bounded and continuous,}$$

we have

$$\mathbf{E}^* H(\alpha_n) \to \mathbf{E}(H(\mathcal{Z})).$$

(E^* denotes the outer expectation. For more details refer to pages 209-210 of de la Peña and Giné (1999).)

Definition: Compact LIL for $u_n(t, \lambda)$ uniformly in $a \leq \lambda \leq b$

For $m \geq 2$, we say that the processes $, \ldots,$ satisfy the compact LIL uniformly in $a \leq \lambda \leq b$, if a.s.

$$\frac{\sup_{\lambda \in [a,b]} \left\| u_n(t, \lambda) - \alpha_n \left(\sum_{i=1}^m f_i(t, \cdot) \right) \right\|_\infty}{\sqrt{\log \log n}} \to 0$$

and the class of functions

$$\mathcal{G} = \left\{ \sum_{i=1}^{m} f_i(t, \cdot) : t \in \mathbb{R}^d \right\}$$

is P-separable and satisfies the compact LIL for the class $\mathcal{G}$. (We shall define this soon.) In particular this implies that

$$\limsup_{n \to \infty} \sup_{\lambda \in [a,b]} \frac{\| u_n(t, \lambda) \|_\infty}{\sqrt{2 \log \log n}} = \sup_{t \in \mathbb{R}^d} \sigma_g(t), \quad \text{a.s.} \tag{LIL}$$

where

$$\sigma^2(t) = Var\left(\sum_{i=1}^{m} f_i(t, X). \right).$$

(For the definition of the compact LIL refer to Chapter 8 of Ledoux and Talagrand (1991).) In these notes we are only interested in the conclusion that (LIL) holds.

Gine and Mason (2006) also derive CLT and LIL for $u_n(\cdot, \lambda)$ in $L_p(\mathbb{R}^d)$ for $1 \leq p < \infty$. For the sake of brevity we shall only discuss sup norm results.

Returning to Example 3

Under suitable regularity conditions we get

$$\sup_{\lambda \in [a,b]} \left| T_n(F, \lambda h_n) - T(F) - \frac{2}{n} \sum_{i=1}^{n} (f(X_i) - T(F)) \right| = o_p\left(1/\sqrt{n}\right)$$

and with probability 1

$$\sup_{\lambda\in[a,b]}\left|T_n(F,\lambda h_n)-T(F)-\frac{2}{n}\sum_{i=1}^{n}\left(f(X_i)-T\left(F\right)\right)\right|=o\left(\sqrt{\log\log n}/\sqrt{n}\right).$$

The last statement implies that

$$\limsup_{n\to\infty}\pm\frac{\sqrt{n}}{\sqrt{2\log\log n}}\sup_{\lambda\in[a,b]}\left\{T_n(F,\lambda h_n)-T\left(F\right)\right\}$$

$$=\sigma\left(f\right)=2\sqrt{\mathbf{E}f^2\left(X\right)-\left(\mathbf{E}f(X)\right)^2}.$$

Tools and Methods of Proof

Hoeffding Decomposition

Let L be a function of m variables, symmetric in its entries. Then, for $1\le k\le m$, the Hoeffding projections with respect to P are defined as

$$\pi_kL(x_1,\ldots,x_k)=(\delta_{x_1}-P)\times\cdots\times(\delta_{x_k}-P)\times P^{m-k}(L)$$

with $\pi_0L=\mathbf{E}L(X_1,\ldots,X_m)$.

The *Hoeffding decomposition* states the following:

$$\frac{(n-m)!}{n!}\sum_{\mathbf{i}\in I_n^m}L(X_{i_1},\ldots,X_{i_m})-\mathbf{E}L=:U_n^{(m)}(L)-\mathbf{E}L$$

$$=\sum_{k=1}^{m}\binom{m}{k}U_n^{(k)}(\pi_kL).$$

Assuming L is in $L_2(P^m)$, this is an orthogonal decomposition and

$$\mathbf{E}\left(\pi_kL\,|X_2,\ldots,X_k\right)=0\text{ for }k\ge 1,$$

that is, the kernels π_kL are canonical for P. Also, π_k, $k\ge 1$, are nested projections, that is $\pi_k\circ\pi_\ell=\pi_k$ if $k\le\ell$.

The Resulting Expansion

The function $K_h(t-g(X_1,\ldots,X_m))$ is not necessarily symmetric in its entries, but we can symmetrize it as

$$\overline{K}_h(t,x_1,\ldots,x_m):=\frac{1}{m!}\sum_{\sigma\in\rho_m}K_h(t-g(x_{\sigma(1)},\ldots,x_{\sigma(m)})),$$

where ρ_m are the permutations of $1,\ldots,m$. Then, clearly, for each $t\in\mathbb{R}^d$,

$$U_n(t,\lambda)-\mathbf{E}K_{\lambda h_n}(t-g(X_1,\ldots,X_m))$$

$$= U_n^{(m)}(\overline{K}_{\lambda h_n}(t,\cdot,\ldots,\cdot)) - \mathbf{E}\overline{K}_{\lambda h_n}(t,X_1,\ldots,X_m).$$

Moreover, we get

$$u_n(t,\lambda) = \sqrt{n}\sum_{k=1}^{m}\binom{m}{k}U_n^{(k)}(\pi_k\overline{K}_{\lambda h_n}(t,\cdot)).$$

Smoothed Empirical Process

It turns out that the first term of this expansion,

$$\sqrt{n}mU_n^{(1)}(\pi_1\overline{K}_{\lambda h_n}(t,\cdot)) = \frac{m}{\sqrt{n}}\sum_{i=1}^{n}\pi_1\overline{K}_{\lambda h_n}(t,X_i), \tag{1}$$

is a *smoothed empirical process* (as studied by Yukich (1992), van der Vaart (1994) and Rost (1999)).

This term controlls both the CLT and LIL for the local U-statistic process $u_n(t,\lambda)$.

We shall confine our discussion to describing the tools that lead to the LIL.

These tools should have many uses in other contexts. The CLT follows similarly.

A General Proposition for LIL

Assume that X, X_k, $k \in \mathbb{N}$, are i.i.d. with common law P such that for each $i = 1,\ldots,m$, the random variable $g(X_1,\ldots,X_m)$, conditionally on $X_i = x$, has a density $f_i(t,x)$, jointly measurable in t and x, satisfying for each $x \in \mathbb{R}^d$, $f_i(\cdot,x) \in L_\infty\left(\mathbb{R}^d\right)$. Also assume for each $i = 1,\ldots,m$, a.s.

$$\lim_{\delta\searrow 0}\limsup_{n}\sup_{|u-v|\le\delta}\frac{|\alpha_n(f_i(u,\cdot)-f_i(v,\cdot))|}{\sqrt{\log\log n}} = 0; \tag{2}$$

$$\limsup_{n}\sup_{t\in\mathbb{R}^d}\frac{|\alpha_n\left(f_i(t,\cdot)\right)|}{\sqrt{\log\log n}} < \infty. \tag{3}$$

Proposition 2. *Whenever $h_n \searrow 0$, under the above assumptions, we have uniformly in $a \le \lambda \le b$,* a.s.

$$\frac{\left\|\sqrt{n}\left(mU_n^{(1)}(\pi_1\overline{K}_{\lambda h_n}(t,\cdot))\right) - \alpha_n\left(\sum_{i=1}^m f_i(t,\cdot)\right)\right\|_\infty}{\sqrt{\log\log n}} \to 0.$$

**The analogous General Proposition holds for the CLT with $\sqrt{\log\log n}$ removed and convergence almost surely replaced by convergence in probability.*

LIL for the Local U-Statistic Process

It is clear by Proposition 2 and the Hoeffding expansion, that to determine conditions under which $u_n(\cdot,\lambda)$ considered as a process taking values on $\ell^\infty(\mathbb{R}^d)$ indexed by $\lambda \in [a,b]$ obeys a **compact LIL uniformly in** $a \le \lambda \le b$, as defined above, it suffices to impose conditions so that simultaneously those needed for Proposition 2 and the compact LIL are in effect and so that a.s., for $k = 2,\ldots,m$,

$$\sup_{a\le\lambda\le b}\frac{\sqrt{n}\left\|U_n^{(k)}(\pi_k\overline{K}_{\lambda h_n}(t,\cdot,\ldots,\cdot))\right\|_\infty}{\sqrt{\log\log n}} \to 0.$$

**Again an analogous CLT statement holds.*

Compact LIL for the Empirical Process

Recall that a class $\mathcal{F}$ satisfies the compact LIL for P whenever the sequence

$$\left\{ \frac{\alpha_n(f)}{\sqrt{2\log\log n}} : f \in \mathcal{F} \right\}_{n=1}^{\infty}$$

is a.s. relatively compact in $\ell^{\infty}(\mathcal{F})$ with set of limit points

$$\mathcal{H} = \left\{ f \mapsto \mathbf{E}\left\{ (f(X) - Pf)h(X) \right\} : \mathbf{E}h^2(X) \le 1 \right\}.$$

In particular, assuming separability, if $\mathbf{E}F^2 < \infty$, where F is the envelope function of the class $\mathcal{F}$, and $\mathcal{F}$ is P-Donsker then $\mathcal{F}$ satisfies the compact LIL. (Donsker just means that the empirical process α_n indexed by $\mathcal{F}$ converges weakly.)

Applying the compact LIL for empirical processes of Ledoux and Talagrand (1988, 89), we see that for both our general proposition and the compact LIL to hold it is enough for the classes of functions

$$\mathcal{F}_i := \left\{ f_i(t, \cdot) : t \in \mathbb{R}^d \right\}$$

to be P-Donsker and separable for the law P of X,

$$\mathbf{E}F_i^2(X) < \infty, \text{ where } F_i(x) = \sup_{t \in \mathbb{R}^d} f_i(t, x),$$

and $\mathrm{Var}(f_i(u, X) - f_i(v, X))$ is uniformly continuous.

A Useful Class of Functions

Let $X, X_1, X_2, \ldots$ be a sequence of i.i.d. random variables in $\mathbb{R}^d$. Let $\mathcal{H}_k$ denote a countable class of measurable functions defined on $\mathbb{R}^{dk}$ such that each $H \in \mathcal{H}_k$ is symmetric in its arguments and for any $1 \le i \le k$

$$\mathbf{E}H\left(x_1, \ldots, x_{i-1}, X, x_{i+1}, \ldots, x_k\right) = 0.$$

For any $n \ge k$ and $H \in \mathcal{H}_k$ set

$$U_n(H) = \sum_{\mathbf{i} \in \mathbf{I}_n^k} H\left(X_{i_1}, \ldots, X_{i_k}\right)$$

and

$$S_n := \sup_{H \in \mathcal{H}_k} |U_n(H)|.$$

Just like we required the two ngredients of an exponential inequality and a moment bound to study the local empirical process, we shall need two such ingredients to investigate the local U–statistic process.

Ingredient 1: Major's Inequality

For any functional Ψ defined on a class of functions $\mathcal{F}$ set

$$\|\Psi(f)\|_{\mathcal{F}} = \sup_{f \in \mathcal{F}} |\Psi(f)|.$$

Assume $U_n(H)$ is as above and that $\mathcal{H}_k$ is VC and bounded. Choose

$$\sigma^2 \geq \sup_{H \in \mathcal{H}_k} \mathbf{E} H^2(X_1, \ldots, X_k).$$

Then, there exist constants C_1, C_2, C_3 depending only on k and the characteristics of the class $\mathcal{H}_k$ such that

$$P\left\{ \left\| \frac{\sum_{\mathbf{i} \in I_n^k} H(X_{i_1}, \ldots, X_{i_k})}{n^{k/2}} \right\|_{\mathcal{H}_k} > y \right\} \leq C_1 \exp\left(-C_2 \left(\frac{y}{\sigma} \right)^{2/k} \right),$$

whenever y satisfies

$$n\sigma^2 \geq \left(\frac{y}{\sigma} \right)^{2/k} \geq C_3 \log\left(\frac{2}{\sigma} \right).$$

Ingredient 2: A Moment Bound

Consider a class of measurable functions $\mathcal{F}$ defined on $(S^m, \mathcal{S}^m)$ taking values in $[-1, 1]$, and assume that $0 \in \mathcal{F}$. Our object is to obtain a bound for

$$\mathbf{E} \| U_n^{(k)}(\pi_k f) \|_{\mathcal{F}}$$

where $\mathcal{F}$ is of VC type and is suitably measurable.

Theorem *Let $\mathcal{F}$ be a collection of measurable functions $S^m \mapsto \mathbb{R}$ symmetric in their entries with absolute values bounded by 1 and let P be any probability measure on $(S, \mathcal{S})$ (with X_i i.i.d. P). Assume $\mathcal{F}$ is VC with respect to the envelope function $F = 1$ with characteristics A and v, then for every $m \in \mathbb{N}$, $A \geq e^m$, $v \geq 1$, there exist constants $C_1 := C_1(m, A, v)$ and $C_2 = C_2(m, A, v)$ such that, for $k = 1, \ldots, m$,*

$$n^k \mathbf{E} \left\| U_n^{(k)}(\pi_k f) \right\|_{\mathcal{F}}^2 \leq C_1^2 2^k \sigma^2 \left(\log \frac{A}{\sigma} \right)^k,$$

assuming

$$n\sigma^2 \geq C_2 \log\left(\frac{A}{\sigma} \right),$$

where σ^2 is any number satisfying

$$\| P^m f^2 \|_{\mathcal{F}} \leq \sigma^2 \leq 1.$$

Application of Ingredients 1 and 2

Assume now that $\mathcal{H}_k$ is VC and bounded and let

$$\sigma^2 \geq \sup_{H \in \mathcal{H}_k} \mathbf{E} H^2(X_1, \ldots, X_k).$$

Then, Major's inequality and a martingale inequality due to Brown (1971) can be adapted to show that there exist constants C_1, C_2, C_3 depending only on k and the VC characteristics of the class $\mathcal{H}_k$ such that for any $0 < c < 1$ and all $y = x/2$ as above

$$\Pr\left\{ \frac{\max_{k \leq m \leq n} S_m}{n^{k/2}} > x \right\} \leq \frac{C_1^{1/2} \exp\left(-\frac{C_2}{2} \left(\frac{cx}{\sigma} \right)^{2/k} \right) \left(\mathbf{E} \left(S_n / n^{k/2} \right)^2 \right)^{1/2}}{x(1-c)},$$

where

$$S_n := \sup_{H \in \mathcal{H}_k} |U_n(H)|,$$

and for a suitable class of functions $\mathcal{F}$

$$\mathcal{H}_k = \{\pi_k f : f \in \mathcal{F}\}.$$

Refer to the Appendix to see how this maximal inequality is derived.

Further Calculation

Assume for some $c > 1$

$$c^{-1} h_n \le h_{2n} \le c h_n$$

and

$$n h_n \log\log n / \left(\log\log n \vee \log(1/h_n)\right)^2 \to \infty.$$

We get after some calculation using our moment bound and the maximal inequality on a S_n defined via a suitable $\mathcal{H}_k$ on the blocks $2^{r-1} \le n \le 2^r$, $r \ge 1$, in combination with the Borel–Cantelli lemma that for $k = 2, \ldots, m$,

$$\frac{\sqrt{n}}{\sqrt{\log\log n}} \sup_{a \le \lambda \le b} \sup_{t \in \mathbb{R}^d} \left| U_n^{(k)}(\pi_k \overline{K}_{\lambda h_n}(t, \cdot)) \right| = O\left(\frac{\left(\log\log n \vee \log(1/h_n)\right)^{k/2}}{n^{\frac{k-1}{2}} h_n^{1/2} \sqrt{\log\log n}} \right), \quad \text{a.s.}$$

which gives for $k = 2, \ldots, m$,

$$\lim_{n \to \infty} \frac{\sqrt{n}}{\sqrt{\log\log n}} \sup_{a \le \lambda \le b} \sup_{t \in \mathbb{R}^d} \left| U_n^{(k)}(\pi_k \overline{K}_{\lambda h_n}(t, \cdot)) \right| = 0, \quad \text{a.s.}$$

This implies by Proposition 2 that the LIL holds for $u_n(t, \lambda)$ in the sup norm.

Recapping Everything

1. Assume for each $x \in S$;

$$\mathcal{K} = \left\{ K(t - \cdot \gamma) : \gamma \ge 1,\ t \in \mathbb{R}^d \right\}$$

is a bounded point–wise measurable class and is VC for some $A \ge 3$ and $v \ge 1$;

2. the random variable $g(X_1, \ldots, X_m)$ has a bounded density f_g with respect to Lebesgue measure on $\mathbb{R}^d$;

3. for each $i = 1, \ldots, m$, the class of functions $\mathcal{F}_i := \{f_i(t, \cdot) : t \in \mathbb{R}^d\}$ is P-Donsker and separable, where the conditional densities $f_i(t, x)$ are jointly measurable in t and x;.

4. for each $i = 1, \ldots, m$,

$$\mathbf{E} F_i^2(X) < \infty, \text{ where } F_i(x) = \sup_{t \in \mathbb{R}^d} f_i(t, x);$$

5. for each $i = 1, \ldots, m$, $(\mathbb{R}^d, |\cdot|) \mapsto (\mathbb{R}^d, \rho_i)$ is uniformly continuous, where

$$\rho_i^2(u, v) = \mathrm{Var}(f_i(u, X) - f_i(v, X));$$

6. $h_n \to 0$ and for some $c > 1$

$$c^{-1}h_n \leq h_{2n} \leq ch_n;$$

7. as $n \to \infty$,

$$nh_n \log\log n / \left(\log\log n \vee \log\left(1/h_n\right)\right)^2 \to \infty.$$

Then, the processes $u_n(t, \lambda)$ satisfy the compact LIL in $L_\infty(\mathrm{I\!R}^d)$ uniformly in $a \leq \lambda \leq b$. The uniform CLT also holds under these conditions.

Summary

To quickly summarize we get dramatically different behavior according as $m = 1$ or $m \geq 2$. Subject to regularity, when $m = 1$,

$$\lim_{n\to\infty} \sup_{a\leq\lambda\leq b} \sup_{t\in\mathrm{I\!R}^d} \frac{\sqrt{\lambda h_n}\left|u_n\left(t,\lambda\right)\right|}{\sqrt{2\log\left(1/h_n\right)}} = ||K||_2 \sup_{z\in\mathrm{I\!R}^d} \sqrt{f_g\left(z\right)}, \quad \text{a.s.}$$

where f_g is the density of $g\left(X\right)$. Whereas, for $m \geq 2$

$$\limsup_{n\to\infty} \sup_{a\leq\lambda\leq b} \sup_{t\in\mathrm{I\!R}^d} \frac{\left|u_n\left(t,\lambda\right)\right|}{\sqrt{2\log\log n}} = \sup_{t\in\mathrm{I\!R}^d} \sigma_g\left(t\right), \quad \text{a.s.}$$

where

$$\sigma_g^2\left(t\right) = Var\left(\sum_{i=1}^{m} f_i(t,X)\right).$$

Likewise the uniform CLT holds for $u_n\left(t,\lambda\right)$ for $m \geq 2$ but only pointwise for $\sqrt{h_n}u_n\left(t,\lambda\right)$ when $m = 1$.

Further Application of our Tools

Giné and Mason (2006c) have used these tools to study uniform in bandwidth consistency of a class of local U-statistic type estimators introduced by Levit (1978) of the following general class of integral functionals of the cumulative distribution:

$$T(F) = \int_{\mathrm{I\!R}} \varphi(x, F(x), F^{(1)}(x), ..., F^{(r)}(x))\, dF\left(x\right),$$

where F is a cumulative distribution function on $\mathrm{I\!R}$ with $r \geq 1$ derivatives $F^{(m)}$, $1 \leq m \leq r$,.

Example 3 given above is a special case. Here

$$T(F) = \int_{\mathrm{I\!R}} f\left(x\right) dF\left(x\right) = \int_{\mathrm{I\!R}} f^2\left(x\right) dx,$$

and the Levit estimator of this $T(F)$ becomes the local U-statistic $T_n(F, h_n)$ introduced above.

Representation of Estimators of $T(F)$

Giné and Mason (2006c) have considered estimation of $T\left(F\right)$ by the Levit (1978) type estimator

$$\widehat{T}_n(F) = \frac{1}{n}\sum_{i=1}^{n} \varphi(X_i, \widehat{F}_n(X_i), \widehat{F_n}^{(1)}(X_i), ..., \widehat{F_n}^{(r)}(X_i)),$$

for appropriate U-statistic-type estimators

$$\widehat{F}_n, \widehat{F_n}^{(1)}, ..., \widehat{F_n}^{(r)} \text{of } F, F^{(1)}, ..., F^{(r)}.$$

We show that for suitable i.i.d. $Y_1, Y_2, \ldots,$

$$\widehat{T}_n(F) - T(F) = \frac{1}{n}\sum_{i=1}^{n} Y_i + R_n \longleftarrow \text{[Remainder Term]}$$

For Example 3

$$T_n(F, h_n) = \widehat{T}_n(F) = \frac{1}{n(n-1)h_n}\sum_{i\neq j} K\left(\frac{X_i - X_j}{h_n}\right)$$

and

$$\widehat{T}_n(F) - T(F) = \frac{2}{n}\sum_{i=1}^{n}\{f(X_i) - T(F)\} + R_n.$$

We prove that R_n converges to zero in various senses and rates using the techniques that we have just described.

We also establish uniform in bandwidth versions of these results.

Conditional U-Statistics Estimation

Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be i.i.d. with common joint density $f_{X,Y}$ and let $\varphi(Y_1, \ldots, Y_m)$ be a function of $Y_1, \ldots, Y_m$. Consider the regression function

$$H\left(\overrightarrow{t}\right) = \mathbf{E}\left(\varphi(Y_1, \ldots, Y_m) \,|X_1, \ldots, X_m) = \overrightarrow{t}\right),$$

where $\overrightarrow{t} = (t_1, \ldots, t_m)$.

Stute (1991) studied pointwise consistency of the *conditional U–statistic* estimator of $H\left(\overrightarrow{t}\right)$ given by

$$\widehat{H}_n\left(\overrightarrow{t}; h_n\right) = \frac{\sum_{(i_1,\ldots,i_m)\in I_n^m}\varphi(Y_{i_1}, \ldots, Y_{i_m})\, K\left(\frac{t_1 - X_{i_1}}{h_n}\right)\ldots K\left(\frac{t_m - X_{i_m}}{h_n}\right)}{\sum_{(i_1,\ldots,i_m)\in I_n^m} K\left(\frac{t_1 - X_{i_1}}{h_n}\right)\ldots K\left(\frac{t_m - X_{i_m}}{h_n}\right)}.$$

Uniform in Bandwidth Consistency of the Estimator

Recently Dony and Mason have applied the methods developed in Einmahl and Mason (2005) and Giné and Mason (2006a,b) to show that under appropriate regularity conditions, with probability 1,

$$\limsup_{n\to\infty}\sup_{a_n\le h\le b_n}\sqrt{\frac{h}{\log(1/h)}}\sup_{\overrightarrow{t}\in[a,b]^m}\left|\widehat{H}_n\left(\overrightarrow{t}; h\right) - H\left(\overrightarrow{t}\right)\right| < \infty,$$

for $-\infty < a < b < \infty$, $a_n < b_n$, $a_n \to 0$, $b_n \to 0$ and $b_n/a_n \to \infty$ at rates depending upon the moments of $\varphi(Y_1, \ldots, Y_m)$.

This is a generalization of a result of Einmahl and Mason (2005), who prove it for the case $m = 1$.

Appendix

VC Classes of Functions

Recall that we say that a class of measurable P-square integrable functions $\mathcal{F}$ defined on a measurable space $(S, \mathcal{S})$ is VC–type or VC (VC for Vapnik and Červonenkis) with respect to an envelope F (meaning a measurable function F such that $|f| \le F$ for all $f \in \mathcal{F}$) if the covering number

$$N(\mathcal{F}, L_2(Q), \varepsilon),$$

defined as the smallest number of $L_2(Q)$ open balls of radius ε required to cover $\mathcal{F}$, satisfies

$$N(\mathcal{F}, L_2(Q), \varepsilon) \le \left(\frac{A\|F\|_{L_2(Q)}}{\varepsilon}\right)^v, \quad 0 < \varepsilon \le 2\|F\|_{L_2(Q)},$$

for some $A \ge 3$ and $v \ge 1$, for every probability measure Q on $\mathcal{S}$. If this holds for $\mathcal{F}$, then we say that the *VC class* $\mathcal{F}$ admits the characteristics A and v.

There are a lot of VC classes. Here is a way to generate them. Let $\Psi : \mathbb{R} \mapsto \mathbb{R}$ be a function of bounded variation on $\mathbb{R}$ (Ψ is the difference of two bounded non-decreasing functions). The proof of Lemma 22 in Nolan and Pollard (1987) shows that if

$$K(x) = \Psi(p(x)), \quad x \in \mathbb{R}^d, \tag{4}$$

where p is either a real polynomial on $\mathbb{R}^d$ or the α–th power of the absolute value of a real polynomial on $\mathbb{R}^d$, $\alpha > 0$, then the class of functions

$$\mathcal{K} = \left\{K\left(\gamma^{-1}(t - \cdot)\right) : t \in \mathbb{R}^d, \gamma > 0\right\}$$

is VC–type.

Point–wise Measurable

A class of functions $\mathcal{F}$ is a point–wise measurable class whenever there exists a countable subclass $\mathcal{F}_0$ of $\mathcal{F}$ such that we can find for any function $g \in \mathcal{F}$ a sequence of functions $\{g_m\}$ in $\mathcal{F}_0$ for which

$$g_m(z) \to g(z), \ z \in \mathbb{R}^d.$$

This condition is discussed in van der Vaart and Wellner (1996). It is satisfied for the class

$$\mathcal{K} = \left\{K\left(\gamma^{-1}(t - \cdot)\right) : t \in \mathbb{R}^d, \gamma > 0\right\}$$

whenever K is right continuous.

A Useful Submartingale and a Maximal Inequality

Let $X, X_1, X_2, \ldots$ be a sequence of i.i.d. random variables in $\mathbb{R}^d$. Let $\mathcal{H}_k$ denote a countable class of measurable functions defined on $\mathbb{R}^{dk}$ such that each $H \in \mathcal{H}_k$ is symmetric in its arguments and for any $1 \le i \le k$

$$\mathbf{E}H(x_1, \ldots, x_{i-1}, X, x_{i+1}, \ldots, x_k) = 0.$$

For any $n \ge k$ and $H \in \mathcal{H}_k$ set

$$U_n(H) = \sum_{\mathbf{i} \in \mathbf{I}_n^k} H(X_{i_1}, \ldots, X_{i_k})$$

and

$$S_n := \sup_{H \in \mathcal{H}_k} |U_n(H)|\,.$$

Let $\mathcal{F}_n$ denote the smallest sigma field generated by $X_1, \ldots, X_n$. Clearly

$$\mathbf{E}\left(U_{n+1}\left(H\right)|\mathcal{F}_n\right) = U_n(H)$$

and hence $(S_n|\mathcal{F}_n)_{n \geq k}$ is a submartingale since

$$\mathbf{E}\left(\sup_{H \in \mathcal{H}_k} |U_{n+1}\left(H\right)| \;\; |\mathcal{F}_n\right) \geq S_n.$$

Next by a result of Brown (1971), see Inequality 2 on page 870 of Shorack and Wellner (1986)), for any $\gamma > 0$ and $0 < c < 1$,

$$\Pr\left\{\max_{k \leq m \leq n} S_m > \gamma\right\} \leq \frac{\int_{\{S_n > c\gamma\}} S_n dP}{\gamma(1-c)},$$

which gives

$$\Pr\left\{\max_{k \leq m \leq n} S_n > \gamma\right\} \leq \frac{\Pr\left\{S_n > c\gamma\right\}^{1/2}\left(\mathbf{E}S_n^2\right)^{1/2}}{\gamma(1-c)}.$$

This is the maximal inequality that was used in conjunction with Ingredients 1 and 2 in the notes.

References

Brown, B. M. Martingale central limit theorems. Ann. Math. Statist. 42 1971 59–66.

Csörgő, M.; Révész, P. How big are the increments of a Wiener process? Ann. Probab. 7 (1979), no. 4, 731–737.

de la Peña, Víctor H.; Giné, Evarist Decoupling. From dependence to independence. Randomly stopped processes. U-statistics and processes. Martingales and beyond. Probability and its Applications (New York). Springer-Verlag, New York, 1999.

Deheuvels, Paul; Mason, David M. General asymptotic confidence bands based on kernel-type function estimators. Stat. Inference Stoch. Process. 7 (2004), no. 3, 225–277.

Eastwood, Vera R.; Horváth, Lajos Limit theorems for short distances in R_m. Statist. Probab. Lett. 45 (1999), no. 3, 261–268..

Einmahl, Uwe; Mason, David M. An empirical process approach to the uniform consistency of kernel-type function estimators. J. Theoret. Probab. 13 (2000), no. 1, 1–37.

Einmahl, Uwe; Mason, David M. Uniform in bandwidth consistency of kernel-type function estimators. Ann. Statist. 33 (2005), no. 3, 1380–1403

Frees, Edward W. Estimating densities of functions of observations. J. Amer. Statist. Assoc. 89 (1994), no. 426, 517–525.

Giné, Evarist; Guillou, Armelle Rates of strong uniform consistency for multivariate kernel density estimators. En l'honneur de J. Bretagnolle, D. Dacunha-Castelle, I. Ibragimov. Ann. Inst. H. Poincaré Probab. Statist. 38 (2002), no. 6, 907–921.

Giné, Evarist; Mason, David M. Laws of the iterated logarithm for the local U-statistic process. J. Theoret. Probab., (2006a) To appear.

Giné, Evarist; Mason, David M. On local U–statistic processes and the estimation of densities of functions of several sample variables (2006b), Ann. Statist., To appear.

Giné, Evarist; Mason, David M.Uniform in Bandwidth Estimation of Integral Functionals of the Density Function (2006c). Preprint.

Jammalamadaka, S. Rao; Janson, Svante Limit theorems for a triangular scheme of U-statistics with applications to inter-point distances. Ann. Probab. 14 (1986), no. 4, 1347–1358.

Ledoux, M.; Talagrand, M. Characterization of the law of the iterated logarithm in Banach spaces. Ann. Probab. 16 (1988), no. 3, 1242–1264.

Ledoux, M.; Talagrand, M. Comparison theorems, random geometry and some limit theorems for empirical processes. Ann. Probab. 17 (1989), no. 2, 596–631.

Ledoux, Michel; Talagrand, Michel Probability in Banach spaces. Isoperimetry and processes. Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)], 23. Springer-Verlag, Berlin, 1991.

Levit, B. Ya. Asymptotically efficient estimation of nonlinear functionals. (Russian) Problems Inform. Transmission 14 (1978), no. 3, 65–72.

Major, Peter. An estimate on the supremum of a nice class of stochastic integrals and U-statistics. *Probab. Th. Rel. Fields.* (2005), the online version.

Mason, David M. A uniform functional law of the logarithm for the local empirical process. Ann. Probab. 32 (2004), no. 2, 1391–1418.

Nolan, Deborah; Pollard, David U-processes: rates of convergence. Ann. Statist. 15 (1987), no. 2, 780–799.

Rost, Daniel Limit theorems for smoothed empirical processes. High dimensional probability, II (Seattle, WA, 1999), 107–113, Progr. Probab., 47, Birkhäuser Boston, Boston, MA, 2000.

Schick, Anton; Wefelmeyer, Wolfgang Root n consistent density estimators for sums of independent random variables. J. Nonparametr. Stat. 16 (2004), no. 6, 925–935.

Shorack, Galen R.; Wellner, Jon A. Empirical processes with applications to statistics. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, 1986.

Stute, Winfried The oscillation behavior of empirical processes. Ann. Probab. 10 (1982), no. 1, 86–107.

Stute, Winfried Conditional U-statistics. Ann. Probab.. 19 (1991), no. 2, 812–825.

Talagrand, Michel. New concentration inequalities in product spaces. Invent. Math. 126 (1996), no. 3, 505–563.

van der Vaart, Aad Weak convergence of smoothed empirical processes. Scand. J. Statist. 21 (1994), no. 4, 501–504.

van der Vaart, Aad W.; Wellner, Jon A. Weak convergence and empirical processes. With applications to statistics. Springer Series in Statistics. Springer-Verlag, New York, 1996.

Yukich, J. E. Weak convergence of smoothed empirical processes. Scand. J. Statist. 19 (1992), no. 3, 271–279.