MaPhySto
The Danish National Research Foundation:
Network in Mathematical Physics and Stochastics

Lecture Notes
no. 14 May 2004

Peter Gaenssler
and Daniel Rost:

Empirical and Partial-sum Processes; Revisited as Random Measure Processes

ISSN 1398-2699

MaPhySto
Centre for Mathematical
Physics and Stochastics

Summer School on
Empirical Processes

University of Aarhus, 9-20 August 1999

Prinicipal Course on

EMPIRICAL AND PARTIAL - SUM PROCESSES REVISITED AS RANDOM MEASURE PROCESSES

by Peter Gaenssler (Munich)

BASED ON THE PRESENT
Lecture Notes

by
Peter Gaenssler and Daniel Rost

Mathematical Institute
University of Munich
Theresienstrasse 39
D-80333 Munich
Germany

Second edition

$\mathcal{P}ref ace$

The interest of the first author of the present Lecture Notes in empirical process theory arose after having studied Ron Pyke's beautiful survey [Py72] on Empirical Processes where Ron underlines his view that "the development of empirical processes provides an excellent illustration of the interplay between statistics and probability and of increased sophistication of mathematical techniques which have been introduced into these disciplines in recent years." Since then the theory of empirical processes has grown in an enormous way initiated by Dudley's [Du78] fundamental paper and culminating in his book [Du99]. Also the books of Shorack-Wellner [Sh86] and van der Vaart-Wellner [Va96] together with Pollard [Po84],[Po90] and the overview given by Giné [Gi96] confirm Pyke's early view in a very impressive way.

In view of the large literature on empirical processes which have appeared in recent years, the present Lecture Notes will only cover a small amount of the subject. Our approach in revisiting Empirical and Partial-Sum Processes as so-called Random Measure Processes had its origin in the papers by Pyke [Py84] and Ossiander-Pyke [Os85].

We hope to raise with our presentation further interest in empirical process theory.

Munich, July 1999 *Peter Gaenssler* and *Daniel Rost*

$\mathcal{P}ref ace$ *to the second edition*

The present extended version of our Lecture Notes, first published as MaPhySto Lecture Notes no. 5 in August 1999, is based on Lectures by the first author given at the Ludwig-Maximilians-University in Munich during the Summer-term 2001. Compared with the first edition the additional sections are marked by an asterik; besides of the opening Section 0 and the closing Section 8 the additional sections are slight modifications of results as presented in Duembgen [Due00].

Munich, September 2003 *Peter Gaenssler* and *Daniel Rost*

Contents

0 To open the door by examples

This section is to present some typical examples in order to demonstrate how general empirical process theory (to be presented in the following sections) works w.r.t. applications in (nonparametric) statistics (cf. [Al85] and [We92]).

Let $\xi_i, i \in \mathbb{N}$, be independent identically distributed (i i d) random elements (re's) in an arbitrary measurable space $X = (X, \mathcal{X})$ with law $\nu \equiv \mathcal{L}\{\xi\}$ on $\mathcal{X}$, the ξ_i's being defined as coordinate projections on the probability space ($\equiv$ p-space)

$$(\Omega, \mathcal{A}, \mathbb{P}) := \Big(X^{\mathbb{N}}, \mathcal{X}^{\mathbb{N}} \equiv \bigotimes_{\mathbb{N}} \mathcal{X}, \nu^{\mathbb{N}} \equiv \bigotimes_{\mathbb{N}} \nu\Big),$$

i.e. in the i i d - case we always impose this so-called *canonical model* as underlying p-space.
In general, ξ is called a *re in* $X = (X, \mathcal{X}) :\Longleftrightarrow \quad \exists$ p-space $(\Omega, \mathcal{A}, \mathbb{P})$ such that (s.t.) $\xi : \Omega \longrightarrow X$ is $\mathcal{A}, \mathcal{X}$-measurable, i.e. $\xi^{-1}(B) \in \mathcal{A} \;\; \forall B \in \mathcal{X}$, where

$$\xi^{-1}(B) := \{\omega \in \Omega \; : \; \xi(\omega) \in B\} \equiv \{\xi \in B\}.$$

Let ν_n be the empirical measure based on $\xi_1, ..., \xi_n$, i.e.

$$\nu_n(B) := n^{-1} \sum_{j \leq n} \delta_{\xi_j}(B)\,, \quad B \in \mathcal{X}. \tag{0.1}$$

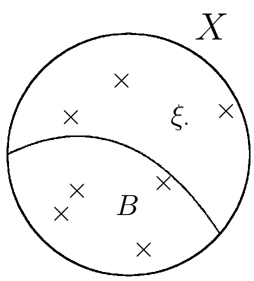
where δ_x denotes the Dirac measure in $x \in X$, i.e.

$$\delta_x(B) := \begin{cases} 1 & , \text{ if } x \in B \\ 0 & , \text{ if } x \in \complement B \equiv X \backslash B \end{cases} .$$

The n^{th} empirical process $\beta_n = (\beta_n(B))_{B \in \mathcal{X}}$ is defined by

$$\beta_n(B) := n^{1/2}\big(\nu_n(B) - \nu(B)\big);$$

thus $\beta_n(B)$ is the normalized deviation from its expected value of the fraction of the random points $\xi_1, ..., \xi_n$ which fall into B



$\mathbb{E}(\nu_n(B)) =$
$n^{-1} \sum_{j \leq n} \mathbb{E}(\delta_{\xi_j}(B)) =$
$n^{-1} \sum_{j \leq n} \mathbb{P}(\xi_j \in B) = \nu(B)$

More generally, given a measurable function $f : X \longrightarrow \mathbb{R}$ and a signed measure Q on $\mathcal{X}$, let $Q(f) :=$

$\int\limits_X f\,dQ$; in this way, given a class $\mathcal{F}$ of measurable functions $f : X \longrightarrow \mathbb{R}$, β_n may be viewed as a stochastic process $\beta_n = (\beta_n(f))_{f\in\mathcal{F}}$ indexed by $\mathcal{F}$, with

$$\beta_n(f) := \int\limits_X f\,d\beta_n = n^{-1/2} \sum_{j\le n} \Big(f(\xi_j) - \mathbb{E}(f(\xi_j))\Big).$$

(Note that for $Q := \nu_n - \nu$ we have

$$\begin{aligned} Q(f) = (\nu_n - \nu)(f) \ &= \ \nu_n(f) - \nu(f) \\ = \int\limits_X f\,d\nu_n - \nu(f) \ &= \ n^{-1} \sum_{j\le n} \big(f(\xi_j) - \nu(f)\big) \\ = n^{-1} \sum_{j\le n} \Big(f(\xi_j) - \mathbb{E}(f(\xi_j))\Big), \end{aligned}$$

since, by the transformation theorem ([Gae77], 1.10.4)

$$\mathbb{E}(f(\xi_j)) := \int\limits_\Omega f(\xi_j(\omega))\,\mathbb{P}(d\omega) = \int\limits_X f(x)\,\nu(dx) \ = \ \nu(f) \ ).$$

β_n may also be viewed as a process indexed by some class $\mathcal{C} \subset \mathcal{X}$; taking $(X, \mathcal{X}) = (\mathbb{R}, \mathcal{B})$ and

$$\mathcal{C} := \{(-\infty, t] \ : \ t \in \mathbb{R}\}$$

makes β_n equivalent to the *normalized empirical distribution function (edf)*

$$\Big(n^{1/2}(F_n(t) - F(t))\Big)_{t\in\mathbb{R}},$$

where $F_n(t) := n^{-1} \ \sum_{j\le n} \delta_{\xi_j}((-\infty, t])$ and $F(t) := \mathbb{P}(\xi_j \le t) = \nu((-\infty, t]), t \in \mathbb{R}$. By identifying sets with indicator functions, however, we may consider, when desired, also indexing by functions (see 4.3. below).

Part of what makes function-indexed empirical processes of interest to statisticians is that many statistics of interest can be expressed as functionals of such processes.

For example, if, as before, β_n is indexed by $\mathcal{C} = \{(-\infty, t] \ : \ t \in \mathbb{R}\}$, then

$$h(\beta_n) := \sup_{C\in\mathcal{C}} |\beta_n(C)| = \sup_{t\in\mathbb{R}} n^{1/2}|F_n(t) - F(t)|,$$

with F_n being the empirical distribution function based on the random variables (rv's) $\xi_1, ..., \xi_n$ with df F, is the *Kolmogorov-Smirnov statistic.*

If $\beta_n = (\beta_n(f))_{f\in\mathcal{F}}$, indexed by some class $\mathcal{F}$, converges weakly ($\mathcal{L}$ - convergence), in a sense to be described below in Section 2.3, to a Gaussian process $\mathbb{G}_\nu = (G_\nu(f))_{f\in\mathcal{F}}$ $\quad (\beta_n \xrightarrow[\text{sep}]{\mathcal{L}} \mathbb{G}_\nu)$, then the limit distribution of any statistic $h(\beta_n)$ will be immediately identified as the distribution of $h(\mathbb{G}_\nu)$ at least for "nice" h; see the *Continuous Mapping Theorem (CMT)* 2.3.16 below).

For example, Donsker's (1952) *Functional Central Limit Theorem (FCLT)* states (see Theorem 1.1.6 below) that the uniform empirical process $\alpha_n = (\alpha_n(t))_{t\in[0,1]}$ converges weakly (in law) to a Brownian bridge $B^\circ = (B^\circ(t))_{t\in[0,1]}$, which shows the limit distribution of the Kolmogorov-Smirnov statistic to be that of $\sup\limits_{t\in\mathbb{R}} |B^\circ(t)|$.
Thus for $\nu \equiv U[0,1]$ (uniform distribution on $[0,1]$) and $\mathcal{C} := \{[0,t] : t \in \mathbb{R}\}$ the process $\mathbb{G}_\nu = (G_\nu(C))_{C\in\mathcal{C}}$ is given by $G_\nu(C) = B^\circ(t)$ for $C = [0,t]$.
In general, $\mathbb{G}_\nu = (G_\nu(f))_{f\in\mathcal{F}}$ is an mean-zero Gaussian process with the same covariance structure as $\beta_n = (\beta_n(f))_{f\in\mathcal{F}}$, i.e. $cov(G_\nu(f), G_\nu(g)) = \nu(f\cdot g) - \nu(f)\cdot\nu(g)$ for $f, g \in \mathcal{F}$.

All of this motivates the study of empirical processes. This will be done in the more general context of the so-called *Random Measure Processes (RMP)* to cope at the same time with another class of important processes in probability theory, namely the *partial-sum processes* with either fixed or random locations (see Section 3 below).

Now, we are going to present some examples towards applications in Nonparametric Statistics where general empirical process theory proves to be an efficient and useful tool:

0.2. Example.
In dimension $d = 1$, the Kolmogorov-Smirnov statistic is a natural way to measure the distance between ν_n and ν. In dimension $d > 1$ however, the analogue

$$\sup_{\underline{t}\in\mathbb{R}^d} n^{1/2}\big|F_n(\underline{t}) - F(\underline{t})\big|,$$

is less natural. It corresponds to the class

$$\mathcal{C} := \{(-\infty, \underline{t}] \ : \ \underline{t} \in \mathbb{R}^d\}$$

of "lower left orthants" having few symmetries and giving special preference, for no good reason, to the "lower left" direction in which all coordinates approach $-\infty$.

The statistic $\sup\limits_{C\in\mathcal{C}} n^{1/2}|\nu_n(C) - \nu(C)|$ for a class $\mathcal{C}$ with more symmetries, e.g. the class of all half spaces, or all ellipsoids if ν is normal, seems more natural (cf. the Remarks after Theorem 2.1.6 below).

0.3. Example.
An M-estimator $\hat\vartheta_n$ for an unknown parameter $\vartheta \in \Theta \subset \mathbb{R}^d, d \geq 1$, is obtained by choosing $\hat\vartheta_n$ to minimize an expression of the form

$$\sum_{j\leq n} \gamma(\xi_j, \vartheta) \tag{0.4}$$

for some function $\gamma : X \times \Theta \longrightarrow \mathbb{R}$ and given i i d re's ξ_j in $(X, \mathcal{X}), j \in \mathbb{N}$, with law $\nu \equiv \mathcal{L}\{\xi_j\} = Q_{1,\vartheta_0}, \vartheta_0 \in \Theta$, being the true but unknown parameter; again the ξ_j's are considered as coordinate *projections on the p-space* $(\Omega, \mathcal{A}, \mathbb{P}) := \Big(X^{\mathbb{N}}, \mathcal{X}^{\mathbb{N}}, \bigotimes_{\mathbb{N}} Q_{1,\vartheta_0}\Big)$. *The true parameter* ϑ_0 *is considered to be that which minimizes*

$$\int_X \gamma(x,\vartheta) Q_{1,\vartheta_0}(dx); \tag{0.5}$$

γ *might, for example, be the negative of the log likelihood in case of Maximum Likelihood Estimators (MLE's).*

In many situations estimators that minimize a certain expression like (0.4) also solve a system of equations:
In particular, in the i i d case as before, let $\hat{\vartheta}_n$ satisfy the equation

$$\sum_{j\le n} \psi(\xi_j, \vartheta) = 0; \tag{0.6}$$

ψ might, for example, be the derivative of the log likelihood in case of MLE's $\hat{\vartheta}_n$.
Solutions of (0.6) are called Z-estimators (from "zero") for ϑ_0 being a solution of

$$H(\vartheta) := \int_X \psi(x,\vartheta) Q_{1,\vartheta_0}(dx) = 0. \tag{0.7}$$

(in (0.5) and (0.7) it is tacitly assumed that the integrals exist.)

We note that in the literature sometimes the name M-estimator is (also) used for what van der Vaart and Wellner ([Va96] call Z-estimator and the distinction between the different types of estimators is not always made.

Note also that (0.6) is the empirical analogue of (0.7) replacing in (0.7) the true underlying p-measure by the empirical measure ν_n based on the observations $\xi_1, ..., \xi_n$ (known as "plug-in-method"), i.e. (0.6) is equivalent to

$$\int_X \psi(x,\vartheta) \nu_n(dx) = 0. \tag{0.6$'$}$$

We are going to prove ASYMPTOTIC NORMALITY OF THE SEQUENCE $(\hat{\vartheta}_n)_{n\in\mathbb{N}}$ of Z- (M-) ESTIMATORS via general empirical process theory:
For this, let $\psi_\vartheta(x) := \psi(x,\vartheta)$ and $\mathcal{F} := \{\psi_\vartheta : \vartheta \in \Theta_0\}$, where $\Theta_0 \subset \mathbb{R}$ is some compact neighborhood of ϑ_0, and consider the empirical $\mathcal{F}$-process $\beta_n = (\beta_n(f))_{f\in\mathcal{F}}$, where

$$\beta_n(f) := n^{-1/2} \sum_{j\le n} \Big(f(\xi_j) - \mathbb{E}(f(\xi_j)) \Big).$$

Assume the following three conditions to hold (where $\xrightarrow{\mathbb{P}}$ denotes convergence in probability):

$$\hat{\vartheta}_n \xrightarrow{\mathbb{P}} \vartheta_0 \tag{0.8}$$

(i.e. weak consistency of $(\hat{\vartheta}_n)_{n\in\mathbb{N}}$)

(0.9) $$\beta_n \xrightarrow[\text{sep}]{\mathcal{L}} \beta_0$$

where $\beta_0 = (\beta_0(f))_{f\in\mathcal{F}}$ is a stochastic process with sample paths in $U^b(\mathcal{F},d) := \{x : \mathcal{F} \longrightarrow \mathbb{R} \ : \ x$ bounded and uniformly d-continuous $\}$ with $d(\psi_\vartheta,\psi_{\vartheta'}) := ||\vartheta - \vartheta'||$ $\quad(||\cdot||$ denoting the Euclidian norm in $\mathbb{R}^d)$

(0.10) $\quad H$ is differentiable on Θ_0 with continuous derivative H' and $H'(\vartheta_0) \neq 0$.

Then

(0.11) $$n^{1/2}(\hat{\vartheta}_n - \vartheta_0) \xrightarrow{\mathcal{L}} \mathcal{N}\big(0, \frac{\sigma^2}{[H'(\vartheta_0)]^2}\big),$$

where $\sigma^2 := \int\limits_X \psi^2(x,\vartheta_0)Q_{1,\vartheta_0}(dx)$ $\quad$ (tacitely assuming that $\sigma^2 > 0$).

(0.9) constitutes a *Functional Central Limit Theorem (FCLT)* for the empirical process β_n indexed by $\mathcal{F}$. Such a theorem holds e.g., if the "size of $\mathcal{F}$" is not too large, e.g., if $\mathcal{F}$ is a so-called *Vapnik-Chervonenkis graphclass (VCGC)* and if the metric d is equivalent to the pseudometric

$$d_\nu^{(2)}(\psi_\vartheta,\psi_{\vartheta'}) := \Big[\int\limits_X \big(\psi(x,\vartheta) - \psi(x,\vartheta')\big)^2 Q_{1,\vartheta_0}(dx)\Big]^{1/2}$$

being the case under weak smoothness conditions on $\psi(x,\,\cdot\,)$.
(cf. Sections 4.3 and 7.3 below).

Now, according to a CHARACTERIZATION THEOREM OF $\mathcal{L}$-CONVERGENCE (see Theorem 2.3.9 below) condition (0.9) implies the so-called ASYMPTOTIC EQUICONTINUITY CONDITION (AEC)

(0.12) $\quad \forall\, \varepsilon,\eta > 0 \ \exists\, \delta = \delta(\varepsilon,\eta) > 0$ and $\exists\, n_0 = n_0(\varepsilon,\eta) \in \mathbb{N}$ such that $\forall\, n \geq n_0$
$$\mathbb{P}^*\Big(\sup_{\vartheta,\vartheta'\in\Theta_0,||\vartheta-\vartheta'||\leq\delta} \big|\beta_n(\psi_\vartheta) - \beta_n(\psi_{\vartheta'})\big| \geq \varepsilon\Big) \leq \eta$$

($\mathbb{P}^*$ denotes outer probability, defined for any $A \subset \Omega$ by $\mathbb{P}^*(A) := \inf\{\mathbb{P}(B) \ : \ A \subset B, B \in \mathcal{A}\}$.)

As we will see, (0.12) will be crucial in proving (0.11).

PROOF of (0.11). $\quad$ Let $\varepsilon,\eta > 0$ be arbitrary and choose $\delta > 0$ such that $\{\vartheta : ||\vartheta - \vartheta_0|| \leq \delta\} \subset \Theta_0$ and (0.12) holds; then

$$\mathbb{P}\Big(\big|\beta_n(\psi_{\hat{\vartheta}_n}) - \beta_n(\psi_{\vartheta_0})\big| \geq \varepsilon\Big) \leq \mathbb{P}\Big(\Big\{\big|\beta_n(\psi_{\hat{\vartheta}_n}) - \beta_n(\psi_{\vartheta_0})\big| \geq \varepsilon\Big\} \cap \{||\hat{\vartheta}_n - \vartheta_0|| \leq \delta\}\Big) + \mathbb{P}(||\hat{\vartheta}_n - \vartheta_0|| > \delta),$$

where $\mathbb{P}(||\hat{\vartheta}_n - \vartheta_0|| > \delta) \underset{n\to\infty}{\longrightarrow} 0$ $\quad$ by (0.8).
Thus

$$\limsup_{n\to\infty} \mathbb{P}\Big(\big|\beta_n(\psi_{\hat{\vartheta}_n}) - \beta_n(\psi_{\vartheta_0})\big| \geq \varepsilon\Big) \leq \limsup_{n\to\infty} \mathbb{P}^*\Big(\sup_{\vartheta,\vartheta'\in\Theta_0,||\vartheta-\vartheta'||\leq\delta} \big|\beta_n(\psi_\vartheta) - \beta_n(\psi_{\vartheta'})\big| \geq \varepsilon\Big) \underset{(0.12)}{\leq} \eta$$

for all $\eta > 0$, whence

$$(+) \qquad \beta_n(\psi_{\hat\vartheta_n}) - \beta_n(\psi_{\vartheta_0}) \xrightarrow{\mathbb{P}} 0.$$

Now, noticing that $H(\vartheta_0) = 0$ by (0.7), we get by the definition of $H(\vartheta)$:

$$n^{1/2}(H(\hat\vartheta_n) - H(\vartheta_0)) = n^{1/2} \int\limits_X \psi(x, \hat\vartheta_n) Q_{1,\vartheta_0}(dx).$$

Since, by (0.6)$'$ $\int\limits_X \psi(x, \hat\vartheta_n)\nu_n(dx) = 0$ we get

$$n^{1/2}(H(\hat\vartheta_n) - H(\vartheta_0)) = -\Big[\underbrace{n^{1/2} \int\limits_X \psi(x, \hat\vartheta_n)\nu_n(dx)}_{=n^{-1}\sum_{j\le n} \psi_{\hat\vartheta_n}(\xi_j)} - \underbrace{n^{1/2} \int\limits_X \psi(x, \hat\vartheta_n) Q_{1,\vartheta_0}(dx)}_{\substack{=\mathbb{E}(\psi_{\hat\vartheta_n}(\xi_1)) = n^{-1}\sum_{j\le n} \mathbb{E}(\psi_{\hat\vartheta_n}(\xi_j)), \\ \text{where the expectation } \mathbb{E} \text{ is only} \\ \text{taken w.r.t. the } \xi_i\text{'s}}}\Big],$$

i.e.

$$\begin{aligned} n^{1/2}(H(\hat\vartheta_n) - H(\vartheta_0)) &= -n^{-1/2} \sum_{j\le n} (\psi_{\hat\vartheta_n}(\xi_j) - \mathbb{E}(\psi_{\hat\vartheta_n}(\xi_j)) \\ &= -\beta_n(\psi_{\hat\vartheta_n}) \\ &= -\underbrace{[\beta_n(\psi_{\hat\vartheta_n}) - \beta_n(\psi_{\vartheta_0})]}_{\xrightarrow{\mathbb{P}} 0 \ \text{ by } (+)} - \beta_n(\psi_{\vartheta_0}) \\ &= -\beta_n(\psi_{\vartheta_0}) + o_{\mathbb{P}}(1). \end{aligned}$$

(In this context, for a sequence $(\eta_n)_{n\in\mathbb{N}}$ of random variables (rv's) defined on a common p-space $(\Omega, \mathcal{A}, \mathbb{P})$

$$\eta_n = o_{\mathbb{P}}(1) \quad :\Longleftrightarrow \quad \eta_n \xrightarrow{\mathbb{P}} 0\ .\ )$$

Therefore, we have shown

$$n^{1/2}(H(\hat\vartheta_n) - H(\vartheta_0)) = -\beta_n(\psi_{\vartheta_0}) + o_{\mathbb{P}}(1)$$

and, since by the classical CLT

$$-\beta_n(\psi_{\vartheta_0}) = -n^{-1/2} \sum_{j\le n} \left[\psi(\xi_j, \vartheta_0) - \mathbb{E}(\psi(\xi_j, \vartheta_0))\right] \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2)$$

with $\sigma^2 := \int\limits_X \psi^2(x, \vartheta_0)\, Q_{1,\vartheta_0}(dx)$ (tacitly assuming $\sigma^2 > 0$), we have

$$(++) \qquad n^{1/2}(H(\hat\vartheta_n) - H(\vartheta_0)) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2).$$

Now (0.11) follows easily from

$$n^{1/2}(H(\hat{\vartheta}_n) - H(\vartheta_0)) = n^{1/2}H'(\vartheta_n^*)(\hat{\vartheta}_n - \vartheta_0),$$

where ϑ_n^* is between $\hat{\vartheta}_n$ and ϑ_0, using (++) and a Cramér-Slutzky theorem (note that by (0.8) and the continuity of H' we have

$$H'(\vartheta_n^*) \xrightarrow{\mathbb{P}} H'(\vartheta_0) \neq 0 \text{).}$$

□

0.3 Example - continued.

Concerning Huber's paper [Hu67], the crucial step in establishing Asymptotic Normality of Maximum Likelihood Estimators $\hat{\vartheta}_n$ *is proving that for some* $d_0 > 0$ *(with* $\vartheta_0 \in \Theta \subset \mathbb{R}^d$, $d \geq 1$, *being the true but unknown parameter and again with* $|| \cdot ||$ *denoting the Euclidian norm)*

$$\sup_{||\tau-\vartheta_0|| \leq d_0} Z_n(\tau, \vartheta_0) \xrightarrow{\mathbb{P}} 0, \tag{0.13}$$

where $Z_n(\tau, \vartheta) := \frac{\beta_n(\psi_\tau - \psi_\vartheta)}{1+n^{1/2}|\nu(\psi_\tau - \psi_\vartheta)|}$ *(see Huber's Lemma 3). As Pollard has observed, under Huber's assumptions, the crucial condition (0.13) holds, whenever*

$$\beta_n \xrightarrow[sep]{\mathcal{L}} \mathbb{G}_\nu \tag{0.14}$$

where $\beta_n = (\beta_n(f))_{f \in \mathcal{F}}$ *and* $\mathbb{G}_\nu = (G_\nu(f))_{f \in \mathcal{F}}$ *are indexed by*

$$\mathcal{F} := \{\psi_\tau \ : \ ||\tau - \vartheta_0|| \leq d_0\}.$$

Let us postpone here again what $\xrightarrow[sep]{\mathcal{L}}$ *means (cf. Section 2.3).*

One PROOF is as follows:
Huber's assumptions readily imply that for some positive constants a, b and all τ with $||\tau - \vartheta_0|| \leq d_0$

$$|\nu(\psi_\tau - \psi_{\vartheta_0})| \geq a||\tau - \vartheta_0||, \quad \text{and} \tag{*}$$

$$\nu(\psi_\tau - \psi_{\vartheta_0})^2 \leq b||\tau - \vartheta_0||. \tag{**}$$

Therefore, for any $\varepsilon > 0$ and $M < \infty$

$$\begin{aligned}
\nu(\sup_{||\tau-\vartheta_0|| \leq d_0} |Z_n(\tau, \vartheta_0)| > \varepsilon) &\leq \nu\Big(\sup_{||\tau-\vartheta_0|| \leq d_0} \Big\{|\beta_n(\psi_\tau - \psi_{\vartheta_0})| \ : \ |\nu(\psi_\tau - \psi_{\vartheta_0})| \leq \frac{M}{\varepsilon n^{1/2}}\Big\} > \varepsilon\Big) \\
&\quad + \nu\Big(\sup_{||\tau-\vartheta_0|| \leq d_0} \Big\{|\beta_n(\psi_\tau - \psi_{\vartheta_0})| \ : \ |\nu(\psi_\tau - \psi_{\vartheta_0})| > \frac{M}{\varepsilon n^{1/2}}\Big\} > M\Big) \\
&\leq \nu\Big(\sup_{||\tau-\vartheta_0|| \leq d_0} \Big\{|\beta_n(\psi_\tau) - \beta_n(\psi_{\vartheta_0})| \ : \ \nu(\psi_\tau - \psi_{\vartheta_0})^2 \leq \frac{b\,M}{a\,\varepsilon n^{1/2}}\Big\} > \varepsilon\Big) \\
&\quad + \nu\Big(\sup_{||\tau-\vartheta_0|| \leq d_0} |\beta_n(\psi_\tau)| > \frac{M}{2}\Big).
\end{aligned}$$

(Note: $|\nu(\psi_\tau - \psi_{\vartheta_0})| \le \frac{M}{\varepsilon n^{1/2}} \quad \Longrightarrow \quad \nu(\psi_\tau - \psi_{\vartheta_0})^2 \underset{(**)}{\le} b||\tau - \vartheta_0|| \underset{(*)}{\le} \frac{b\,|\nu(\psi_\tau - \psi_{\vartheta_0})|}{a} \le \frac{b}{a}\frac{M}{\varepsilon\, n^{1/2}} \underset{n\to\infty}{\longrightarrow} 0$.)

But weak convergence of β_n (i.e. $\beta_n \underset{\text{sep}}{\xrightarrow{\mathcal{L}}} \mathbb{G}_\nu$) implies that $(\beta_n)_{n\in\mathbb{N}}$ is stochastically bounded (cf. [St94], 1.6) whence, since M is arbitrary, the term $\nu(\sup_{||\tau-\vartheta_0||\le d_0} |\beta_n(\psi_\tau)| > \frac{M}{2})$ can be made arbitrarily small as $n \to \infty$.
Concerning the other term, the same holds as $n \to \infty$ due to the (AEC) (with d replaced by $d_\nu^{(2)}$; cf. 2.3.12 below). Thus (0.13) is shown. □

In the next example we will see that by a result of Ossiander ([Os87]) $\beta_n \underset{\text{sep}}{\xrightarrow{\mathcal{L}}} \mathbb{G}_\nu$ (indexed by $\mathcal{F}$) (i.e. (0.14) holds true under Huber's assumptions). In this way Huber's result follows fairly directly from a FCLT for function-indexed empirical processes.

0.15. Example.
We are going to prove (0.14), i.e. $\beta_n \underset{sep}{\xrightarrow{\mathcal{L}}} \mathbb{G}_\nu$*, where the processes* $\beta_n, n \in \mathbb{N}$*, and* $\mathbb{G}_\nu$ *are indexed by*

$$\mathcal{F} := \{\psi_\tau \ : \ ||\tau - \vartheta_0|| \le d_0\}$$

$(\psi_\tau := \psi(\cdot\,, \tau))$.

Suppose that the function

$$u(x, \vartheta, r) := \sup_{\tau:\, ||\tau-\vartheta||\le r} |\psi_\tau(x) - \psi_\vartheta(x)|$$

(assumed to be measurable as a function in x) satisfies Huber's critical assumption

$$(+) \qquad \nu(u(\cdot\,, \vartheta, r)^2) \le cr \qquad \forall\, r > 0 \ \text{ and } \forall\, ||\vartheta - \vartheta_0|| \le d_0 \text{ for some } c > 0.$$

Fix $\varepsilon > 0$ and let $\tau_1, ..., \tau_{N(\varepsilon)}$ be points in $\Theta_0 := \{\tau : \ ||\tau - \vartheta_0|| \le d_0\}$ such that $\forall \tau \in \Theta_0 \ \exists\, \tau_i, 1 \le i \le N(\varepsilon)$ s.t. $||\tau - \tau_i|| \le \frac{\varepsilon^2}{4c}$.
The number $N(\varepsilon)$ of points needed is of order $O(\varepsilon^{-2d})$ as $\varepsilon \to 0$ in case $\Theta \subset \mathbb{R}^d$. Now, if $||\tau - \tau_i|| \le \frac{\varepsilon^2}{4c}$, then, by the definition of u (with $r := \frac{\varepsilon^2}{4c}$)

$$f_l(x) := \psi_{\tau_i}(x) - u(x, \tau_i, \frac{\varepsilon^2}{4c}) \ \le \ \psi_\tau(x) \ \le \ \psi_{\tau_i}(x) + u(x, \tau_i, \frac{\varepsilon^2}{4c}) =: f_u(x)$$

where

$$d_\nu^{(2)}(f_l, f_u)^2 := \nu(|f_u - f_l|^2) = \nu\Big([2\, u(\cdot\,, \tau_i, \frac{\varepsilon^2}{4c})]^2\Big) \underset{(+)}{\le} 4c\frac{\varepsilon^2}{4c} = \varepsilon^2,$$

whence $d_\nu^{(2)}(f_l, f_u) \le \varepsilon$. Therefore

$$N^{[]}(\varepsilon, \mathcal{F}, d_\nu^{(2)}) \le N(\varepsilon) = O(\varepsilon^{-2d}) \quad \text{ as } \quad \varepsilon \to 0,$$

whence

$$(++) \qquad \int_0^1 \big(\log N^{[]}(\varepsilon, \mathcal{F}, d_\nu^{(2)})\big)^{1/2} d\varepsilon < \infty.$$

Here, given a class $\mathcal{F}$ of measurable functions $f : X \longrightarrow \mathbb{R}$, suppose that for each $\varepsilon > 0$ there exists a finite collection $\mathcal{F}(\varepsilon)$ of measurable functions on X s.t. for each $f \in \mathcal{F}$ there are functions f_l, f_u in $\mathcal{F}(\varepsilon)$ with $f_l \leq f \leq f_u$ s.t. $d_\nu^{(2)}(f_l, f_u) \leq \varepsilon$, where $d_\nu^{(2)}(f_l, f_u) = \nu(|f_u - f_l|^2)^{1/2}$.
The minimal cardinality of such a collection $\mathcal{F}(\varepsilon)$ is denoted by

$$N^{[\,]}(\varepsilon, \mathcal{F}, d_\nu^{(2)})$$

and the function $\log N^{[\,]}(\cdot\,, \mathcal{F}, d_\nu^{(2)})$ is called METRIC ENTROPY WITH BRACKETING OF $\mathcal{F}$ w.r.t. the metric $d_\nu^{(2)}$.

Now, by a fundamental result of Ossiander ([Os87]) (++) implies that the empirical process $\beta_n = (\beta_n(f))_{f \in \mathcal{F}}$ indexed by $\mathcal{F}$ converges weakly (in law) to $\mathbb{G}_\nu$ $(\beta_n \underset{\text{sep}}{\xrightarrow{\mathcal{L}}} \mathbb{G}_\nu)$ (cf. [Gee00], Theorem 6.3), where $\mathbb{G}_\nu = (G_\nu(f))_{f \in \mathcal{F}}$ is a mean-zero Gaussian processes with sample paths in the space $U^b(\mathcal{F}, d_\nu^{(2)}) := \{x : \mathcal{F} \longrightarrow \mathbb{R} \;:\; x$ bounded and uniformly $d_\nu^{(2)}$-continuous$\}$
and the covariance structure given by $cov(G_\nu(f_1)\,, G_\nu(f_2)) = \nu(f_1 \cdot f_2) - \nu(f_1) \cdot \nu(f_2),\;\; f_i \in \mathcal{F}, i = 1, 2;$ (cf. also Theorem 7.3.5 below).
This, as said before, provides the crucial step in [Hu67] to proving asymptotic normality of Maximum Likelihood Estimators.

0.16. Example *(Pollard's k-means clustering procedure, to be considered in more detail in Section 4.3 A below).*
Given data points $x_1, ..., x_n \in X = \mathbb{R}^d$ *viewed as realizations of i i d re's* ξ_j *in* $(\mathbb{R}^d, \mathcal{B}^d)$*, the k-means (with k being arbitrary but fixed and given in advance) empirical cluster centers* $a_{n1}^*, ..., a_{nk}^* \in \mathbb{R}^d$ *are the k points which best approximate* $\xi_1, ..., \xi_n$ *in the sense that* $\sum\limits_{j \leq n} \min\limits_{1 \leq i \leq k} |\xi_j - a_i|^2$ *is minimized by* $(a_{n1}^*, ..., a_{nk}^*)$ *over all* $(a_1, ..., a_k)$ *with* $a_i \in \mathbb{R}^d$.

Pollard [Po82b] applied empirical process theory (as examined in [Po84], Chapter VII) with (cf. 0.3 Example)

$$\gamma(x, (a_1, ..., a_k)) := \min_{1 \leq i \leq k} |x - a_i|^2$$

to obtain asymptotic normality of $(a_{n1}^*, ..., a_{nk}^*)$ as $n \to \infty$ whenever $\mathbb{E}(|\xi_1|^2) < \infty$.

0.17. Example *(DENSITY ESTIMATION).*
(considered within the frame of so-called smoothed empirical processes in Sections 6.4 and 7.4).
Let $\eta_j, j \in \mathbb{N}$, *be i i d re's in* $(\mathbb{R}^d, \mathcal{B}^d)$ *with* $\nu := \mathcal{L}\{\eta_j\}$ *having an unknown density g w.r.t. Lebesgue measure. Let* $K \geq 0$ *be a kernel function on* $\mathbb{R}^d$ *with* $\int\limits_{\mathbb{R}^d} K(v)\,dv = 1$*, and let* (h_n) *be a given sequence of bandwidths. Then the density g can be estimated by the so-called kernel density estimator* $\hat{g}_n$*, defined by*

$$\hat{g}_n(t) := h_n^{-d} \frac{1}{n} \sum_{j \leq n} K(\frac{t - \eta_j}{h_n}) \;=\; h_n^{-d} \int\limits_{\mathbb{R}^d} K(\frac{t - y}{h_n})\, \nu_n(dy)\;, \qquad t \in \mathbb{R}^d.$$

Now, we have with $K_n(t)(\cdot) := h_n^{-d} K(\frac{t-\cdot}{h_n})$

$$\hat{g}_n(t) = \nu_n(K_n(t)),$$

whence (with $\beta_n(f) := n^{1/2}(\nu_n(f) - \nu(f))$ *)* $\beta_n(K_n(t))$ *is the random part*

$$n^{1/2}(\hat{g}_n(t) - \mathbb{E}(\hat{g}_n(t)))$$

in the decomposition

$$\hat{g}_n(t) - g(t) = \hat{g}_n(t) - \mathbb{E}(\hat{g}_n(t)) + \mathbb{E}(\hat{g}_n(t)) - g(t)$$

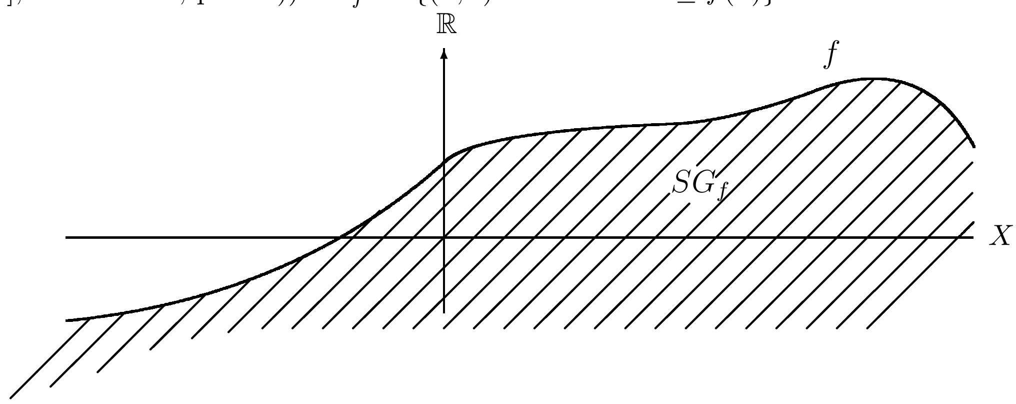
with $\mathbb{E}(\hat{g}_n(t)) - g(t)$ *being the "BIAS".*
(Note also that under regularity conditions

$$\begin{aligned} \mathbb{E}(\hat{g}_n(t)) &= h_n^{-d}\mathbb{E}\Big(K(\frac{t-\eta_1}{h_n})\Big) \ = \ \nu(K_n(t)) \\ &= h_n^{-d} \int\limits_{\mathbb{R}^d} K(\frac{t-y}{h_n}) g(y)\, dy \underset{n\to\infty}{\longrightarrow} g(t).) \end{aligned}$$

Using this point of view, empirical process techniques are in order with indexing sets given by

$$\mathcal{F} := \{K\Big(\frac{t-\cdot}{h_n}\Big) \ : \ t \in \mathbb{R}^d\}.$$

Now, results from general empirical process theory will be available if, as already remarked in connection with (0.9), $\mathcal{F}$ is not too large, e.g. if $\mathcal{F}$ is a VCGC (see Definition 4.3.16 below) or, equivalently (cf. [Va96], Problem 11, p. 152), a *Vapnik-Chervonenkis Subgraph Class (VCSGC)*, where the *subgraph* of a function $f : X \longrightarrow \mathbb{R}$ is defined by $SG_f := \{(x, r) \in X \times \mathbb{R} \ : \ r < f(x)\}$ or (equivalently (cf. [Va96], Problem 10, p. 152)) $SG_f := \{(x, r) \in X \times \mathbb{R} \ : \ r \leq f(x)\}$.



Now, $\mathcal{F}$ is a VCSGC $:\Longleftrightarrow$ $\{SG_f \ : \ f \in \mathcal{F}\}$ is a Vapnik-Chervonenkis Class in $(X \times \mathbb{R}, \mathcal{X} \bigotimes \mathcal{B})$. For the definition of a *Vapnik-Chervonenkis Class (VCC) of sets* see Section 4.2 below. As to the

VCC-property, let us mention here the following result (Lemma 4.2.5):

(0.18) Let $\mathcal{G}$ be an arbitrary m-dimensional vector space of real-valued functions g being defined on an arbitrary set X equipped with the σ-field $\mathcal{X} = \mathcal{P}(X)$ (whence each g is measurable). Then the class

$$\mathcal{C} := \{\{g \geq 0\} \ : \ g \in \mathcal{G}\} \text{ is a VCC.}$$

Based on this result we are going to show next

(0.19) Let $H : \mathbb{R}_+ \longrightarrow \mathbb{R}$ be monotone increasing (or decreasing), $X := \mathbb{R}^d, d \geq 1$, and

$$\mathcal{F} := \{H(b||t - \cdot||^2) \ : \ b > 0, t \in \mathbb{R}^d\}.$$

Then $\mathcal{F}$ is a VCSGC.

PROOF (Taken from [Due00]): This is easily seen if H can be extended to an isotone bijection on $\mathbb{R}$; in this case one has for any $f(x) = H(b||t-x||^2), x \in \mathbb{R}^d$:

$$SG_f = \{(x,r) \in \mathbb{R}^d \times \mathbb{R} \ : \ b||t||^2 - 2bt^T x + bx^T x - H^{-1}(r) \geq 0\} = \{g \geq 0\}$$

(where x^T denotes the transpose of $x = \begin{pmatrix} x_1 \\ \vdots \\ x_d \end{pmatrix}$) with $g \in \mathcal{G}$, $\mathcal{G}$ being the $(d+3)$- dimensional vector space generated by the functions $g_0(x,r) := 1$, $g_i(x,r) := x_i, 1 \leq i \leq d$, $g_{d+1}(x,r) := x^T x$, and $g_{d+2}(x,r) := H^{-1}(r)$, whence the assertion follows from (0.18).

In the general case, one must find a proper substitute for H^{-1} to argue in a similar way: For this, let F be an arbitrary finite subset of $\mathbb{R}^d \times \mathbb{R}$ and suppose that for any $F' \subset F$ there exists an $f' \in \mathcal{F}$, i.e. $f' = H(b_{F'}||t_{F'} - \cdot||^2)$ with $b_{F'} > 0, t_{F'} \in \mathbb{R}^d$, such that $F' = F \cap SG_{f'}$. Then, one shows that this necessarily implies $F' = \{(x,r) \in F \ : \ g(x,r) \geq 0\} = F \cap \{g \geq 0\}$ for some $g \in \mathcal{G} := span\{g_0, g_1, ..., g_{d+1}, g_{d+2}\}$ where now $g_{d+2}(x,r) := \bar{H}(r) := \min\{t \in T \ : \ H(t) \geq r\}$ with $T := \{b_{F'}||t_{F'} - x||^2 \ : \ (x,r) \in F, F' \subset F\}$. Thus again the assertion follows from (0.18). □

0.20. Example *(from classical probability theory: A strong law of large numbers).*
Let $\xi_i, i \in \mathbb{N}$, be i i d rv's with law $\nu = \mathcal{L}\{\xi_i\}$, the ξ_i's being defined as coordinate projections on the p-space $(\Omega, \mathcal{A}, \mathbb{P}) = (\mathbb{R}^{\mathbb{N}}, \mathcal{B}^{\mathbb{N}}, \nu^{\mathbb{N}})$. Assume $\mathbb{E}(|\xi_1|) < \infty$ and let $\mu := \mathbb{E}(\xi_1)$ and $\bar{\xi}_n := n^{-1} \sum_{i=1}^n \xi_i$. Then

(1) $$A_n := n^{-1} \sum_{i=1}^n |\xi_i - \bar{\xi}_n| \longrightarrow a := \mathbb{E}(|\xi_1 - \mu|) \quad \mathbb{P}-a.s.$$

PROOF (by Empirical Process Methods).
Let ν_n be the empirical measure based on $\xi_1, ..., \xi_n$. For $t \in \mathbb{R}$, let $f_t(x) := |x - t|, x \in \mathbb{R}$, $H_n(t) := \nu_n(f_t)$ and $H(t) := \nu(f_t)$. Then $A_n = H_n(\bar{\xi}_n)$ and $a = H(\mu)$. Next, let $\mathcal{F} := \{f_t \ : \ t \in \mathbb{R}\}$, then

the ENVELOPE F of $\mathcal{F}$, defined by $F(x) := \sup_{t\in\mathbb{R}} |f_t(x)|, x \in \mathbb{R}$, is the function $F \equiv \infty$. Considering instead for a fixed $\delta > 0$ the class

$$\mathcal{F}_\delta := \{f_t \ : \ |t-\mu| \leq \delta\},$$

the envelope F_δ of $\mathcal{F}_\delta$ is given by $F_\delta(x) = |x-\mu| + \delta, x \in \mathbb{R}$, i.e. F_δ is real-valued and ν-integrable:

$$\nu(F_d) = \mathbb{E}(|\xi_1 - \mu| + \delta) = \mathbb{E}(|\xi_1 - \mu|) + \delta < \infty,$$

since $\mathbb{E}(|\xi_1|) < \infty$ by assumption.
(Note that $f_t(x) \leq |x-\mu| + |\mu - t| \leq |x-\mu| + \delta \quad \text{if} \quad |\mu - t| \leq \delta$.)
According to the strong law of large numbers, $\bar{\xi}_n \longrightarrow \mu \quad \mathbb{P}-a.s$, i.e. $\forall\, \varepsilon > 0 \quad \forall\, \delta > 0$

$$\lim_{m\to\infty} \mathbb{P}(\sup_{n\geq m} |\bar{\xi}_n - \mu| > \delta) < \varepsilon. \tag{2}$$

Concerning the class $\mathcal{F}_\delta$ one shows (cf. (0.19)) that $\mathcal{F}_\delta$ is a VCSGC (equivalently a VCGC) and thus it follows from 6.3.3 (cf. (6.3.5) below) that

$$||\nu_n - \nu||_{\mathcal{F}_\delta} := \sup_{|t-\mu|\leq\delta} |(\nu_n - \nu)(f_t)| \longrightarrow 0 \quad \mathbb{P}-a.s. \tag{3}$$

But now,

$$\begin{aligned} |A_n - a| &= |H_n(\bar{\xi}_n) - H(\mu)| \\ &\leq |H_n(\bar{\xi}_n) - H(\bar{\xi}_n)| \ + \ \underbrace{|H(\bar{\xi}_n) - H(\mu)|.}_{\longrightarrow 0\ \mathbb{P}-a.s.,\text{ since } \bar{\xi}_n \longrightarrow \mu\ \mathbb{P}-a.s. \text{ and } H \text{ being continuous.}} \end{aligned}$$

So, it remains to show that $\forall\, \rho > 0$

$$\lim_{m\to\infty} \mathbb{P}(\sup_{n\geq m} |H_n(\bar{\xi}_n) - H(\bar{\xi}_n)| > \rho) = 0 \tag{4}$$

For this, let $\rho > 0$ be arbitrary; then

$$\begin{aligned} \mathbb{P}(\sup_{n\geq m} |H_n(\bar{\xi}_n) - H(\bar{\xi}_n)| > \rho) &\leq \mathbb{P}\Big(\{\sup_{n\geq m} |H_n(\bar{\xi}_n) - H(\bar{\xi}_n)| > \rho\} \\ &\qquad \cap \{\sup_{n\geq m} |\bar{\xi}_n - \mu| \leq \delta\}\Big) + \mathbb{P}(\sup_{n\geq m} |\bar{\xi}_n - \mu| > \delta) \\ &\underset{(2)}{\leq} \mathbb{P}(\sup_{n\geq m} \sup_{|t-\mu|\leq\delta} |H_n(t) - H(t)| > \rho) + \varepsilon \\ &= \mathbb{P}(\sup_{n\geq m} ||\nu_n - \nu||_{\mathcal{F}_\delta} > \rho) + \varepsilon. \end{aligned}$$

Since $\varepsilon > 0$ was arbitrary, the assertion (4) follows from (3). $\square$

Of course, (1) follows more or less immediately from the strong law of large numbers, nevertheless the proof here nicely reveals the elegant operation principle how a (functional) uniform law of large numbers as given in (3) may be applied.

Further applications of empirical process theory together with the concepts and results already mentioned will be considered in the following chapters presenting also the theoretical background to cope with future applications beyond our present knowledge.

In fact, as already emphasized in [We92], "modern empirical process theory deals with empirical measures and processes based on data with values in *completely arbitrary*, perhaps *infinite-dimensional* sample spaces. This aspect of the theory will undoubtly become more important in future applications as statisticians develop methods for dealing with 'function'- and 'picture'-valued data such as seismographs, noise level tracings, electrodiagrams, and high-dimensional biomedical data (survival times together with hundreds of covariates)"

"I (Wellner) believe that one important consequence of the rapid developments in modern empirical process tools and techniques is a shortening of the lag time between the introduction of a new method (e.g. a new estimator or test statistic) in statistics and the development of an understanding of the properties and performance of the method."

In an announcement for a workshop on 'Statistical Modelling – Nonparametric Models' it was said:
"As in all areas of science, models serve as a portrayal of the reality. Their quality and usefulness thereby heavily depends on the complexity of the model itself. A simple model can only mirror simple things. Moreover, as in all quantitive sciences, classical quantitative models suffer from being parametric.
Nonparametric models go beyond this scope by modelling relations and effects nonparametrically.
Various applications have been developed in the last years. Numerical and theoretical results allow a wide range of applications of Nonparametric models.
Further research is required to make more profit of this powerful modelling technique.
Investigation of theoretical aspects and applicability of the available routines in more depth is demanded to access and describe their impact.
Finally, a fair comparison of different methods is still rudimentary but desired."

Hopefully, the present COURSE ON SELECTED TOPICS IN MATHEMATICAL STOCHASTICS will serve as a solid basis to cope with applications in nonparametric statistics based on empirical process theory.

Summary

In a general framework of so-called Random Measure Processes (RMP's) we present uniform laws of large numbers (ULLN) and functional central limit theorems (FCLT) for RMP's yielding known and also new results for empirical processes and for so-called smoothed empirical processes based on data in general sample spaces. At the same time one obtains results for Partial-sum processes with either fixed or random locations. Proofs are based on tools from modern empirical process theory as presented e.g. in [Va96].

Our presentation will be also guided by showing up some aspects of the development of empirical process theory from its classical origin up to its present generality which now offers a wide variety of applications in statistics as demonstrated e.g. in Part 3 of [Va96].

1 Introduction to the theory of empirical processes (and partial-sum processes)

1.1 The uniform empirical process α_n

Two important processes in probability and statistics are the *empirical* and *partial-sum process*.

Let $\eta_j, j \in \mathbb{N}$, be independent identically distributed (i i d) random variables (rv's) with law $\mathcal{L}\{\eta_j\} = U[0,1]$ (the uniform distribution on $I = [0,1]$), defined on a basic probability space (p-space) $(\Omega, \mathcal{A}, \mathbb{P})$, i.e. $\eta_j : \Omega \longrightarrow I$ with $\mathbb{P}(\eta_j \leq t) = F(t) := t \;\; \forall t \in I$.

Let F_n be the empirical distribution function (edf) based on $\eta_1, ..., \eta_n$, i.e.

$$F_n(t) := n^{-1} \sum_{j \leq n} 1_{[0,t]}(\eta_j). \quad t \in I;$$

to indicate that F_n is random, i.e. depending on $\omega \in \Omega$, we also write instead of $F_n(t)$

$$F_n(t, \omega) = n^{-1} \sum_{j \leq n} 1_{[0,t]}(\eta_j(\omega)).$$

(1_A denotes the indicator function of a set A.)

THEN:

$\forall t \in I \;\; \mathbb{E}(F_n(t)) = F(t)$ (i.e. $F_n(t)$ is an unbiased estimator for $F(t)$)

$\forall t \in I$ by the classical central limit theorem (CLT)

$$\alpha_n(t) := n^{1/2}\big(F_n(t) - F(t)\big) \xrightarrow{\mathcal{L}} \mathcal{N}(0, F(t)(1 - F(t))$$

(where $\xrightarrow{\mathcal{L}}$ denotes convergence in law), and by the strong law of large numbers (LLN)

$\forall t \in I \;\; F_n(t) \longrightarrow F(t)$ $\mathbb{P}$-almost surely (a.s.)

(i.e. $\big(F_n(t)\big)_{n\in\mathbb{N}}$ is a strongly consistent sequence of estimators for $F(t)$);
moreover, by the *GLIVENKO-CANTELLI THEOREM*,

$$\sup_{t\in I}|F_n(t)-F(t)| \longrightarrow 0 \quad \mathbb{P}-a.s.$$

(Note that $\sup_{t\in I}|F_n(t)-F(t)|$ is measurable since it remains unchanged when replacing I by the countable index set $I\cap\mathbb{Q}$.)

FUNCTIONAL VIEWPOINT

The stochastic process $\alpha_n=(\alpha_n(t))_{t\in I}$ is called UNIFORM EMPIRICAL PROCESS (of sample size n). Its sample paths

$$\alpha_n(\omega)=\Big(n^{1/2}\big(F_n(t,\omega)-F(t)\big)\Big)_{t\in I}$$

are contained in the space $D:=\{x\in\mathbb{R}^I : x \text{ satisfies } (i)-(iii)\}$:

(i) $\forall t\in[0,1)\ \exists x(t+0):=\lim_{s\searrow t}x(s)$

(ii) $\forall t\in(0,1]\ \exists x(t-0):=\lim_{s\nearrow t}x(s)$

(iii) $\forall t\in[0,1)\ x(t)=x(t+0)$.

Since $\sup_{t\in I}|x(t)|<\infty\ \ \forall x\in D$, it is tempting to endow the space D with the sup-metric ρ, i.e. with

$$\rho(x_1,x_2):=\sup_{t\in I}|x_1(t)-x_2(t)|,\quad x_1,x_2\in D,$$

which is usually considered in the space $C\equiv C(I)$ of continuous functions on I.
Note that (C,ρ) is a closed separable subspace of (D,ρ) being also complete (cf. [Bi68], p.220).
In contrast, (D,ρ) is **not** separable and $\alpha_n:\Omega\longrightarrow D$ is **not** $\mathcal{A},\mathcal{B}(\rho)$-measurable if D is equipped with the σ-field $\mathcal{B}(\rho)$ of Borel sets w.r.t. the ρ-topology; cf. [Bi68],p.152.

At this place there were two ways to overcome this difficulty (cf. [Bi68]):

(i) Skorokhod's metric s being weaker than ρ which makes $\alpha_n:\Omega\longrightarrow D\quad \mathcal{A},\mathcal{B}(s)$-measurable ($\mathcal{B}(s):=$ Borel σ-field in (D,s))

(ii) [Du66] (cf. also [Wi68] and [Gae83]):
Consider instead of $\mathcal{B}(\rho)$ the smaller σ-field $\mathcal{B}_b(\rho)$ generated by the open ρ-balls in (D,ρ); then again $\alpha_n:\Omega\longrightarrow D$ becomes $\mathcal{A},\mathcal{B}_b(\rho)$-measurable, since $\mathcal{B}_b(\rho)=\sigma(\{\pi_t : t\in I\})$ ($\equiv$ σ-field generated by the projections $\pi_t:D\longrightarrow\mathbb{R}$, $\pi_t(x):=x(t)$).

1.1.1. Remarks.

Let $\mathcal{B}(C,\rho)$ be the Borel σ-field in (C,ρ) and $\mathcal{B}_b(C,\rho)$ be the σ-field generated by the open ρ-balls in (C,ρ); then $\mathcal{B}(C,\rho)=\mathcal{B}_b(C,\rho)=C\cap\mathcal{B}_b(\rho)$, whence

(1.1.2) $$\mathcal{B}(C,\rho)=\sigma(\{rest_C\pi_t : t\in I\});$$

furthermore $C \in \mathcal{B}_b(\rho)$ *and (cf. [Bi68], Th. 14.5)* $\mathcal{B}(s) = \sigma(\{\pi_t : t \in I\})$, *whence*

$$\mathcal{B}(s) = \mathcal{B}_b(\rho). \tag{1.1.3}$$

In the following let $B = (B(t))_{t \in I}$ be the Wiener process (Brownian Motion) with parameter set $T = I$, and let $B^\circ = (B^\circ(t))_{t \in I}$ be the Brownian Bridge ($B^\circ(t) := B(t) - tB(1)$); both processes are mean-zero Gaussian processes with sample paths in C, whose covariance structure is given by

$$cov(B(t_1), B(t_2)) = t_1 \wedge t_2, \quad t_1, t_2 \in I, \text{ and} \tag{1.1.4}$$

$$cov(B^\circ(t_1), B^\circ(t_2)) = t_1 \wedge t_2 - t_1 \cdot t_2, \quad t_1, t_2 \in I, \tag{1.1.5}$$

respectively. Both processes can be viewed as random elements (re) in $(C, \mathcal{B}(C, \rho))$ or as random elements in $(D, \mathcal{B}_b(\rho))$ with $\mathcal{L}\{B\}(C) = 1$ and $\mathcal{L}\{B^\circ\}(C) = 1$ respectively.

HERE: Given a measurable space $(X, \mathcal{X})$, we say that η is a re in $(X, \mathcal{X}) :\Longleftrightarrow \exists$ p-space $(\Omega, \mathcal{A}, \mathbb{P})$ s.t. $\eta : \Omega \longrightarrow X$ is $\mathcal{A}, \mathcal{X}$-measurable.

The following prospect is taken from [Do49]:

"Noticing that, by the multivariate CLT, the finite-dimensional distributions (fidis) of α_n are asymptotically (as $n \to \infty$) the same as those of B°, we may assume– until a contradiction frustrates our devotion to heuristic reasoning – that in calculating asymptotic distributional results for the α_n-process one may simply replace the α_n's by B°."

This prospect was justified by the following Functional Central Limit Theorem (FCLT):

1.1.6. THEOREM *([Don51],[Don52],[Pro56]).*

$$\alpha_n \xrightarrow{\mathcal{L}} B^\circ \quad \textit{in } (D, s),$$

i.e. $\lim_{n\to\infty} \mathbb{E}(f(\alpha_n)) = \mathbb{E}(f(B^\circ)) \quad \forall f \in C^b(D)$,
where $C^b(D) := \{f : D \longrightarrow \mathbb{R} : f$ *s-continuous and bounded* $\}$.

Note also that $\mathcal{L}\{B^\circ\}(C) = 1$ in view of 1.1.7 below.
Taking instead of s the sup-metric ρ one gets

1.1.6′ THEOREM.

$$\alpha_n \xrightarrow{\mathcal{L}_b} B^\circ \quad \textit{in } (D, \rho),$$

i.e. $\lim_{n\to\infty} \mathbb{E}(f(\alpha_n)) = \mathbb{E}(f(B^\circ))) \ \forall f \in C_b^b(D)$,
where $C_b^b(D) := \{f : D \longrightarrow \mathbb{R} : f \ \ \rho - \textit{continuous}, \mathcal{B}_b(\rho) - \textit{measurable and bounded}\}$.

In fact 1.1.6 and 1.1.6′ are equivalent according to the following lemma. For this, let $\eta_n, n \geq 0$, be a sequence of re's in $(D, \mathcal{B}_b(\rho)) \underset{(1.1.3)}{=} (D, \mathcal{B}(s))$, and

$$\eta_n \xrightarrow{\mathcal{L}} \eta_0 \ :\Longleftrightarrow \ \lim_{n\to\infty} \mathbb{E}(f(\alpha_n)) = \mathbb{E}(f(B^\circ)) \ \forall f \in C^b(D), \text{ and}$$

$$\eta_n \xrightarrow{\mathcal{L}_b} \eta_0 \ :\Longleftrightarrow \ \lim_{n\to\infty} \mathbb{E}(f(\alpha_n)) = \mathbb{E}(f(B^\circ)) \ \forall f \in C_b^b(D),$$

respectively.

(Note that $\mathcal{L}\{\eta_n\}, n \geq 0$, is well defined on $\mathcal{B}_b(\rho) = \mathcal{B}(s)$ and that $\lim_{n\to\infty} \mathbb{E}(f(\alpha_n)) = \mathbb{E}(f(B^\circ)) \iff lim_{n\to\infty} \int_D f d\mathcal{L}\{\eta_n\} = \int_D f d\mathcal{L}\{\eta_0\}$.)

1.1.7. Lemma *([Gae83],Lemma 18.p.93).*

$$\eta_n \xrightarrow{\mathcal{L}} \eta_0 \quad and \quad \mathcal{L}\{\eta_0\}(C) = 1 \implies \eta_n \xrightarrow{\mathcal{L}_b} \eta_0;$$

conversely, $\eta_n \xrightarrow{\mathcal{L}_b} \eta_0 \implies \eta_n \xrightarrow{\mathcal{L}} \eta_0$.

The same situation is met in Section 1.2 in connection with the classical partial-sum process.

As we shall see in Section 2.3 the concept of weak convergence ($\mathcal{L}$ - convergence) can be generalized in such a way that the approximating sequence $(\eta_n)_{n\in\mathbb{N}}$ is not assumed to consist of re's, i.e. arbitrary η_n's will be allowed; measurability is solely assumed for η_0 to which η_n converges weakly.

1.2 The classical partial-sum process ζ_n

Let $\xi_j, j \in \mathbb{N}$, be i i d rv's defined on a basic p-space $(\Omega, \mathcal{A}, \mathbb{P})$ with $\mathbb{E}(\xi_j) = 0$ and $\mathbb{E}(\xi_j^2) = 1$. Let

$$\zeta_n(t) := n^{-1/2} \sum_{\{j:j/n\leq t\}} \xi_j \quad , t \in I;$$

THEN:

$$\begin{aligned} &\mathbb{E}(\zeta_n(t)) = 0 \quad \text{and} \quad \zeta_n(t) \xrightarrow{\mathcal{L}} \mathcal{N}(0,t) \quad \forall t \in I \\ \text{and} \quad &cov(\zeta_n(t_1), \zeta_n(t_2)) = t_1 \wedge t_2 \quad , t_1, t_2 \in I. \end{aligned}$$

FUNCTIONAL VIEWPOINT

The stochastic process $\zeta_n = (\zeta_n(t))_{t\in I}$ is the CLASSICAL (standardized) PARTIAL-SUM PROCESS (of sample size n).

(with $< a >:= \max\{z \in \mathbb{Z} : z \leq a\}, a \in \mathbb{R}$, $\zeta_n(t)$ can also be written as $\zeta_n(t) = n^{-1/2} \sum_{j=1}^{<nt>} \xi_j$.)

Its sample paths $\zeta_n(\omega) = \left(n^{-1/2} \sum_{j\leq<nt>} \xi_j(\omega)\right)_{t\in I}$ are contained in D. ζ_n can be viewed as re in $(D, \mathcal{B}_b(\rho)) = (D, \mathcal{B}(s))$ and the FCLT for ζ_n is also due to Donsker:

1.2.1. THEOREM.

$$\zeta_n \xrightarrow{\mathcal{L}} B \quad in\ (D,s) \quad or,\ equivalently,$$
$$\zeta_n \xrightarrow{\mathcal{L}_b} B \quad in\ (D,\rho).$$

Theorem 1.2.1 and 1.1.6 are special cases of FCLT's to be considered in Section 7.

Nevertheless we want to present here proofs in a form due to Franz Strobl [St90] yielding some indications when dealing later with more general processes.

The following proofs are based on the characterization theorem of $\mathcal{L}$-convergence (CT$\mathcal{L}$-C) presented in Section 2.3 (here with parameter space $T = I$ and metric $d(t_1, t_2) := |t_1 - t_2|, t_1, t_2 \in I$).

PROOF OF THEOREM 1.1.6. According to CT$\mathcal{L}$-C we have to show (i) and (ii), where

(i) $\alpha_n \xrightarrow[\text{fidi}]{\mathbb{P}} B^\circ$, i.e. weak convergence of the finite-dimensional distributions (fidis) of α_n to the corresponding fidis of B°.

(ii) $\lim_{\delta\to 0} \limsup_{n\to\infty} \mathbb{P}^*(w_{\alpha_n}(\delta) > \varepsilon) = 0 \quad \forall \varepsilon > 0$,
where $\mathbb{P}^*(A) := \inf\{\mathbb{P}(B) : A \subset B, B \in \mathcal{A}\} \quad \forall A \subset \Omega$, and where
$w_{\alpha_n}(\delta) := \sup_{t,t' \in I_{|t-t'|\le\delta}} |\alpha_n(t) - \alpha_n(t')| \quad \forall \delta > 0$.

The proof of (i) follows by the multivariate CLT and can be found in standard textbooks (c.f. e.g. [Gae77], 12.2.1).
PROOF OF (ii). Let $\varepsilon > 0$ be arbitrary; w.l.o.g. let

(1) $$\delta \in \mathbb{Q},\ \ 0 < \delta < \frac{1}{4} \quad \text{and } n \ge 9216 \cdot \delta^{-3}.$$

STEP 1: "We are going back to a grid of span δ" in the parameter-space $T = I$; then

$$\{w_{\alpha_n}(\delta) > \varepsilon\} = \{\exists t_i \in I, i = 1,2, \ s.t.\ 0 < t_2 - t_1 \le \delta \ \text{ and } \ |\alpha_n(t_1) - \alpha_n(t_2)| > \varepsilon\}.$$

Now, to each gridpoint t_i we associate a $k_i \in \mathbb{Z}_+$ s.t. $k_i\delta < t_i \le (k_i + 1)\delta$, where $k_2 - k_1 \le 1$ if $t_2 - t_1 \le \delta$. Then

$$\{w_{\alpha_n}(\delta) > \varepsilon\} = \bigcup_{\substack{k \in \mathbb{Z}_+ \\ k < \frac{1}{\delta}}} \{\exists t \in (k\delta, (k+1)\delta] \cap I \ : \ |\alpha_n(t) - \alpha_n(k\delta)| > \varepsilon/3\},$$

whence

$$\begin{aligned}
\mathbb{P}^*(w_{\alpha_n}(\delta) > \varepsilon) &\le \sum_{\substack{k \in \mathbb{Z}_+ \\ k < \frac{1}{\delta}}} \mathbb{P}^*\Big(\sup_{t \in (k\delta,(k+1)\delta]\cap I} |\alpha_n(t) - \alpha_n(k\delta)| > \varepsilon/3\Big) \\
&= \sum_{\substack{k \in \mathbb{Z}_+ \\ k < \frac{1}{\delta}}} \mathbb{P}\Big(\sup_{t \in (k\delta,(k+1)\delta]\cap I\cap\mathbb{Q}} \big|n^{-1/2} \sum_{i \le n} \big(1_{(k\delta,t]}(\eta_i) - (t - k\delta)\big)\big| > \varepsilon/3\Big) \\
&\le (\frac{1}{\delta} + 1)\, \mathbb{P}\Big(\sup_{t \in (0,\delta]\cap\mathbb{Q}} |\alpha_n(t)| > \varepsilon/3\Big),
\end{aligned}$$

i.e. we have

(2) $$\mathbb{P}^*(w_{\alpha_n}(\delta) > \varepsilon) \le (\frac{1}{\delta} + 1)\,\mathbb{P}\Big(\sup_{t \in (0,\delta]\cap\mathbb{Q}} |\alpha_n(t)| > \varepsilon/3\Big).$$

Now, let $T_m \subset (0, \delta] \cap \mathbb{Q}$ be s.t. $|T_m| = m \ \forall m \in \mathbb{N}$, and $T_m \nearrow (0, \delta] \cap \mathbb{Q}$ as $m \to \infty$; then we have

(3) $$\mathbb{P}\Big(\sup_{t \in (0,\delta]\cap\mathbb{Q}} |\alpha_n(t)| > \varepsilon/3\Big) = \lim_{m\to\infty} \mathbb{P}\Big(\sup_{t \in T_m} |\alpha_n(t)| > \varepsilon/3\Big).$$

STEP 2: Let $m \in \mathbb{N}$ be arbitrary but fixed, $T_m = \{t_1, ..., t_m\}$, $0 =: t_0 < t_1 < ... < t_m \leq \delta$, and $A := \{\sup_{t \in T_m} |\alpha_n(t)| > \varepsilon\}$; then $A \subset B + \sum_{k \leq m} A_k^+ + \sum_{k \leq m} A_k^-$ with

$$B := \{ \sum_{i \leq n} 1_{[0,\delta]}(\eta_i) > n/2 \},$$

$$A_k^+ := \{\alpha_n(t_j) \leq \varepsilon/3, j = 1, ..., k-1, \alpha_n(t_k) > \varepsilon/3\,, \sum_{i \leq n} 1_{[0,t_k]}(\eta_i) \leq n/2\},$$

$$A_k^- := \{\alpha_n(t_j) \geq -\varepsilon/3, j = 1, ..., k-1, \alpha_n(t_k) < -\varepsilon/3\,, \sum_{i \leq n} 1_{[0,t_k]}(\eta_i) \leq n/2\}, \quad 1 \leq k \leq m;$$

now, we are going to show

$$\text{(4)} \qquad \mathbb{P}(A_k^{+/-}) \leq 4\, \mathbb{P}\Big(A_k^{+/-} \cap \big\{\alpha_n(2\delta) \geq\!/\!\leq \alpha_n(t_k) \frac{1-2\delta}{1-t_k}\big\}\Big) \quad \forall\, 1 \leq k \leq m:$$

Let $k \in \{1, ..., m\}$ be arbitrary but fixed and

$$R := \{\underline{r} = (r_1, ..., r_k) : r_i \in \mathbb{Z}_+, n^{-1/2}(r_1 + ... + r_j - nt_j) \leq \varepsilon/3 \;\; \forall 1 \leq j \leq k-1,$$
$$n^{-1/2}(r_1 + ... + r_k - nt_k) > \varepsilon/3\,,\; r_1 + ... + r_k \leq n/2\};$$

then

$$\begin{aligned}
&\mathbb{P}\Big(A_k^+ \cap \big\{\alpha_n(2\delta) \geq \alpha_n(t_k) \frac{1-2\delta}{1-t_k}\big\}\Big) \\
&= \mathbb{P}\Big(n^{-1/2}\big(\sum_{i \leq n} 1_{[0,t_j]}(\eta_i) - nt_j\big) \leq \varepsilon/3, j = 1, ..., k-1, n^{-1/2}\big(\sum_{i \leq n} 1_{[0,t_k]}(\eta_i) - nt_k\big) > \varepsilon/3, \\
&\qquad \sum_{i \leq n} 1_{[0,t_k]}(\eta_i) \leq n/2, \sum_{i \leq n} 1_{(t_k,2\delta]}(\eta_i) \geq \big(\sum_{i \leq n} 1_{[0,t_k]}(\eta_i) - nt_k\big) \frac{1-2\delta}{1-t_k} + 2n\delta - \sum_{i \leq n} 1_{[0,t_k]}(\eta_i)\Big) \\
&= \sum_{\substack{\underline{r} \in R \\ r := r_1 + ... + r_k}} \mathbb{P}\Big(\sum_{i \leq n} 1_{(t_{j-1},t_j]}(\eta_i) = r_j, j = 1, ..., k,\; \sum_{i \leq n} 1_{(t_k,2\delta]}(\eta_i) \geq (n-r) \frac{2\delta - t_k}{1-t_k}\Big) \\
&= \sum_{\substack{\underline{r} \in R \\ r := r_1 + ... + r_k}} \;\sum_{s \in [(n-r)\frac{2\delta - t_k}{1-t_k}, n-r) \cap \mathbb{Z}_+} \binom{n}{r_1, ..., r_k, s, n-r-s} \cdot (t_1 - t_0)^{r_1} \ldots (t_k - t_{k-1})^{r_k} \cdot (2\delta - t_k)^s (1-2\delta)^{n-r-s} \\
&= \sum_{\underline{r}} \sum_s \binom{n}{r_1, ..., r_k, n-r} \frac{(n-r)!}{s!(n-r-s)!} \cdot (t_1 - t_0)^{r_1} \ldots (t_k - t_{k-1})^{r_k} (1-t_k)^{n-r} \frac{(2\delta - t_k)^s (1-2\delta)^{n-r-s}}{(1-t_k)^{n-r}} \\
&= \sum_{\underline{r}} \mathbb{P}\Big(\sum_{i \leq n} 1_{(t_{j-1},t_j]}(\eta_i) = r_j, j = 1, ..., k\Big) \cdot \;\sum_s \mathbb{P}\Big(\sum_{i \leq n-r} 1_{[0, \frac{2\delta - t_k}{1-t_k}]}(\eta_i) = s\Big);
\end{aligned}$$

since $\sigma^2 := Var(1_{[0,\frac{2\delta-t_k}{1-t_k}]}(\eta_1)) = \frac{2\delta-t_k}{1-t_k} \cdot \frac{1-2\delta}{1-t_k} \cdot \underset{(1)}{\geq} \delta/2$, it follows that

$$\sum_s \mathbb{P}\Big(\sum_{i\leq n-r} 1_{[0,\frac{2\delta-t_k}{1-t_k}]}(\eta_i) = s\Big) = \mathbb{P}\Big(\sum_{i\leq n-r} 1_{[0,\frac{2\delta-t_k}{1-t_k}]}(\eta_i) \geq (n-r)\frac{2\delta-t_k}{1-t_k}\Big)$$
$$= \mathbb{P}\Big(\sum_{i\leq n-r} 1_{[0,\frac{2\delta-t_k}{1-t_k}]}(\eta_i) - (n-r)\frac{2\delta-t_k}{1-t_k} \geq 0\Big)$$
$$\underset{\text{(BERRY-ESSÉEN)}}{\geq} \frac{1}{2} - \frac{6}{\sqrt{n-r}\sigma^3}\,\mathbb{E}\Big(\big|1_{[0,\frac{2\delta-t_k}{1-t_k}]}(\eta_1) - \frac{2\delta-t_k}{1-t_k}\big|^3\Big)$$
$$\underset{(r\leq n/2)}{\geq} \frac{1}{2} - \frac{6}{\sqrt{n/2}(\delta/2)^{3/2}} \underset{(1)}{\geq} \frac{1}{4};$$

therefore

$$\mathbb{P}\Big(A_k^+ \cap \big\{\alpha_n(2\delta) \geq \alpha_n(t_k)\frac{1-2\delta}{1-t_k}\big\}\Big) \geq$$
$$\frac{1}{4}\,\mathbb{P}\Big(\alpha_n(t_j) \leq \varepsilon/3,\, j=1,...,k-1,\ \alpha_n(t_k) > \varepsilon/3\,,\sum_{i\leq n} 1_{[0,t_k]}(\eta_i) \leq n/2\Big),$$

which proves (4) for A_k^+. Analogously one shows that

$$\mathbb{P}(A_k^-) \leq 4\,\mathbb{P}\Big(A_k^- \cap \big\{\alpha_n(2\delta) \leq \alpha_n(t_k)\frac{1-2\delta}{1-t_k}\big\}\Big).$$

STEP 3: According to STEP 2 we have for any fixed $m \in \mathbb{N}$ with $T_m = \{t_1,...,t_m\}$ and $0 =: t_0 < t_1 < ... < t_m \leq \delta$ that

$$\mathbb{P}\big(\sup_{t\in T_m} |\alpha_n(t)| > \varepsilon\big) \leq$$
$$\mathbb{P}\big(\sum_{i\leq n} 1_{[0,\delta]}(\eta_i) > \frac{n}{2}\big) + 4\sum_{k\leq m} \mathbb{P}\Big(A_k^+ \cap \big\{\alpha_n(2\delta) \geq \alpha_n(t_k)\frac{1-2\delta}{1-t_k}\big\}\Big) +$$
$$4\sum_{k\leq m} \mathbb{P}\Big(A_k^- \cap \big\{\alpha_n(2\delta) \leq \alpha_n(t_k)\frac{1-2\delta}{1-t_k}\big\}\Big).$$

Now, $\frac{1-2\delta}{1-t_k} \underset{(1)}{\geq} \frac{1}{2}$ implies that $\forall 1 \leq k \leq m$

$$A_k^+ \cap \big\{\alpha_n(2\delta) \geq \alpha_n(t_k)\frac{1-2\delta}{1-t_k}\big\} \subset A_k^+ \cap \{\alpha_n(2\delta) \geq \varepsilon/6\} \text{ and}$$
$$A_k^- \cap \big\{\alpha_n(2\delta) \leq \alpha_n(t_k)\frac{1-2\delta}{1-t_k}\big\} \subset A_k^- \cap \{\alpha_n(2\delta) \leq -\varepsilon/6\},$$

whence (noticing that the A_k^+'s as well as the A_k^-'s are pairwise disjoint (p.d.))

$$\mathbb{P}\big(\sup_{t\in T_m} |\alpha_n(t)| > \varepsilon\big) \leq$$
$$\mathbb{P}\big(\sum_{i\leq n} 1_{[0,\delta]}(\eta_i) > \frac{n}{2}\big) + 4\,\mathbb{P}\big(|\alpha_n(2\delta)| \geq \varepsilon/6\big).$$

STEP 4: We are now in the position to verify that

$$\lim_{\delta\to 0}\limsup_{n\to\infty}\mathbb{P}^*(w_{\alpha_n}(\delta)>\varepsilon)=0 \quad \forall\varepsilon>0:$$

$$\begin{aligned}
\mathbb{P}^*(w_{\alpha_n}(\delta)>\varepsilon) &\underset{(2)+(3)}{\leq} (\frac{1}{\delta}+1)\lim_{m\to\infty}\mathbb{P}\Big(\sup_{t\in T_m}|\alpha_n(t)|>\varepsilon/3\Big)\\
&\underset{(\text{STEP 3})}{\leq} (\frac{1}{\delta}+1)\ \Big(\mathbb{P}\big(\sum_{i\leq n}1_{[0,\delta]}(\eta_i)>\frac{n}{2}\big)+4\,\mathbb{P}\big(|\alpha_n(2\delta)|\geq\varepsilon/6\big)\Big)\\
&\underset{(1)}{\leq} \frac{2}{\delta}\ \Big[\mathbb{P}\Big(\frac{\sum_{i\leq n}1_{[0,\delta]}(\eta_i)-n\delta}{\sqrt{n\delta(1-\delta)}}>\frac{n/2-n\delta}{\sqrt{n\delta(1-\delta)}}\Big)+4\mathbb{P}\big(|\alpha_n(s\delta)|\geq\varepsilon/6\big)\Big]\\
&@>CLT>n\to\infty> \frac{2}{\delta}\Big[0+4\cdot 2\big(1-\Phi(\frac{\varepsilon}{6\sqrt{2\delta(1-2\delta)}})\big)\Big],
\end{aligned}$$

where Φ denotes the standard normal df.

Therefore, $\forall\varepsilon>0$

$$\begin{aligned}
\limsup_{n\to\infty}\mathbb{P}^*(w_{\alpha_n}(\delta)>\varepsilon) &\leq \frac{2}{\delta}\cdot 8\frac{6\sqrt{2\delta(1-2\delta)}}{\varepsilon}\frac{1}{\sqrt{2\pi}}e^{-\frac{\varepsilon^2}{2\cdot 36\cdot 2\delta(1-2\delta)}}\\
&\leq \frac{2}{\delta}\cdot 8\ \ \frac{6\sqrt{2\delta}}{\varepsilon\sqrt{2\pi}}\cdot\frac{144\cdot\delta(1-2\delta)}{\varepsilon^2}@>>\delta\to 0>0.
\end{aligned}$$

□

PROOF OF THEOREM 1.2.1. According to CT$\mathcal{L}$-C we have to show (i) and (ii), where

(i) $\zeta_n \xrightarrow[\text{fidi}]{\mathbb{P}} B$, and

(ii) $\lim_{\delta\to 0}\limsup_{n\to\infty}\mathbb{P}^*(w_{\zeta_n}(\delta)>\varepsilon)=0 \quad \forall\varepsilon>0.$

As before, we skip the standard proof of (i).

PROOF OF (ii). Let $\varepsilon>0$ be arbitrary; w.l.o.g. let

$$0<\delta<\frac{1}{2}\,(\frac{\varepsilon}{12})^2,\quad \delta\leq 1,\quad n\delta\geq 1. \tag{5}$$

As in STEP 1 of the proof above we get

$$\begin{aligned}
\mathbb{P}^*(w_{\zeta_n}(\delta)>\varepsilon) &\leq \sum_{\substack{k\in\mathbb{Z}_+\\ k<\frac{1}{\delta}}}\mathbb{P}^*\Big(\sup_{t\in(k\delta,(k+1)\delta]\cap I}|\zeta_n(t)-\zeta_n(k\delta)|>\varepsilon/3\Big)\\
&\leq \sum_{\substack{k\in\mathbb{Z}_+\\ k<\frac{1}{\delta}}}\mathbb{P}\Big(\sup_{t\in(k\delta,(k+1)\delta]\cap I\cap\mathbb{Q}}\Big|n^{-1/2}\sum_{i=<nk\delta>+1}^{<nt>}\xi_i\Big|>\varepsilon/3\Big)\\
&\leq (\frac{1}{\delta}+1)\,\mathbb{P}\Big(\sup_{1\leq k\leq<2n\delta>}\Big|n^{-1/2}\sum_{i\leq k}\xi_i\Big|>\varepsilon/3\Big).
\end{aligned}$$

To obtain an upper estimate for the last expression we make use of the first Lévy-inequality as follows:

Let $m :=< 2n\delta >, S_k := \sum_{i\leq k} \xi_i/n^{\frac{1}{2}},\ \varepsilon' := \varepsilon/3, s_m^2 := \sum_{i\leq m} Var(\xi_i/n^{\frac{1}{2}}) = m/n = \frac{<2n\delta>}{n} \leq 2\delta$, i.e. $s_m \underset{(5)}{\leq} \varepsilon/12$; then

$$\mathbb{P}(\sup_{1\leq k\leq <2n\delta>} \left| n^{-1/2} \sum_{i\leq k} \xi_i \right| > \varepsilon/3) =$$

$$\mathbb{P}(\sup_{1\leq k\leq m} |S_k| > \varepsilon') \leq \mathbb{P}(\sup_{1\leq k\leq m} S_k > \varepsilon') + \mathbb{P}(\sup_{1\leq k\leq m} (-S_k) \geq \varepsilon')$$

$$\underset{(1.\ \text{Lévy-Ineq.})}{\leq} \frac{a^2}{a^2-1} \big[\mathbb{P}(S_m \geq \varepsilon' - a s_m) + \mathbb{P}(-S_m \geq \varepsilon' - a s_m)\big]$$

for all $a > 1$. Taking $a = 2$ and noticing that $\varepsilon' - 2s_m \geq \varepsilon/3 - 2\varepsilon/12 = \varepsilon/6 > 0$, the last expression is

$$\leq\ \frac{4}{3}\,\mathbb{P}(|S_m| \geq \varepsilon/6) = \frac{4}{3}\,\mathbb{P}(|n^{-1/2} \sum_{i\leq <2n\delta>} \xi_i| \geq \varepsilon/6).$$

Therefore (note that $1 \underset{(5)}{\leq} \frac{1}{\delta}$) we get

$$\mathbb{P}^*(w_{\zeta_n}(\delta) > \varepsilon) \leq$$

$$\frac{2}{\delta}\frac{4}{3}\,\mathbb{P}\Big(|n^{-1/2} \sum_{i\leq <2n\delta>} \xi_i| \geq \varepsilon/6\Big) = \frac{8}{3\delta}\,\mathbb{P}\Big(| < 2n\delta >^{-1/2} \sum_{i\leq <2n\delta>} \xi_i| \geq \Big(\frac{n}{< 2n\delta >}\Big)^{1/2} \frac{\varepsilon}{6}\Big)$$

$$\leq \frac{8}{3\delta}\,\mathbb{P}\Big(| < 2n\delta >^{-1/2} \sum_{i\leq <2n\delta>} \xi_i| \geq \frac{\varepsilon}{6\sqrt{2\delta}}\Big)$$

$$@ > CLT > n\to\infty > \quad \frac{8}{3\delta}\, 2\,\big(1 - \Phi(\frac{\varepsilon}{6\sqrt{2\delta}})\big) @ >> \delta\to 0 > 0 \qquad \text{(as in Step 4 above).}$$

□

1.3 The multivariate case

Let $d \geq 1$ and $\eta_j, j \in \mathbb{N}$, be i i d random vectors uniformly distributed on $I^d = [0,1]^d$, defined on a basic p-space $(\Omega, \mathcal{A}, \mathbb{P})$, i.e. $\eta_j : \Omega \longrightarrow I^d$ with $\mathbb{P}(\eta_j \leq \underline{t}) = F(\underline{t}) := \prod_{i\leq d} t_i\ \forall \underline{t} := (t_1, ..., t_d) \in I^d$. Let $\alpha_n = \alpha_n(\underline{t})_{\underline{t}\in I^d}$ be defined by

$$\alpha_n(\underline{t}) := n^{1/2}\big(F_n(\underline{t}) - F(\underline{t})\big), \quad \underline{t} \in I^d,$$

where $F_n(\underline{t}) := n^{-1} \sum_{j\leq n} 1_{[0,\underline{t}]}(\eta_j)$ and $[0, \underline{t}] = [0, t_1] \times \cdots \times [0, t_d]$.

The stochastic process $\alpha_n = (\alpha_n(\underline{t}))_{\underline{t}\in I^d}$ is called MULTIVARIATE UNIFORM EMPIRICAL PROCESS (of sample size n).

The MULTIVARIATE (standardized) PARTIAL-SUM PROCESS (of sample size n) $\zeta_n = (\zeta_n(\underline{t}))_{\underline{t}\in I^d}$ is defined by

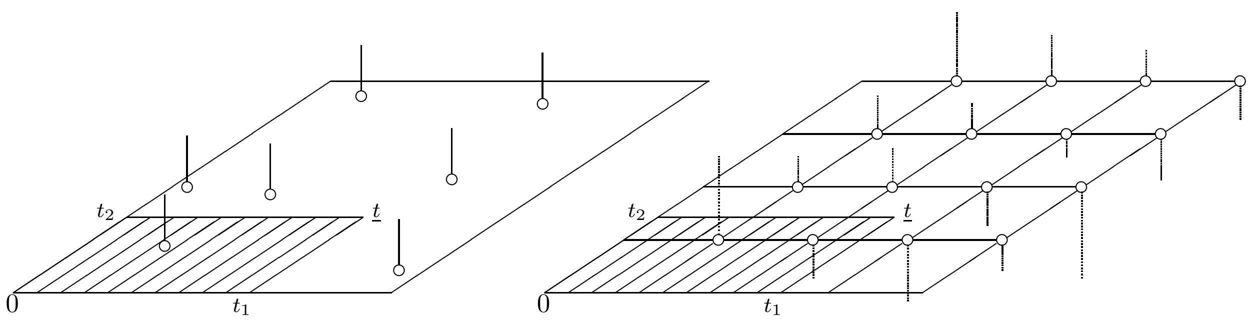
$$\zeta_n(\underline{t}) := n^{-d/2} \sum_{\underline{j}\in J_n : \underline{j}/n \in [0,\underline{t}]} \xi_{\underline{j}} \quad , \underline{t} \in I^d,$$

where $J_n := \{1, ..., n\}^d$, $\underline{j} = (j_1, ..., j_d)$, and where the rv's $\xi_{\underline{j}}, \underline{j} \in \mathbb{N}^d$, are assumed to be iid with $\mathbb{E}(\xi_{\underline{j}}) = 0$ and $\mathbb{E}(\xi_{\underline{j}}^2) = 1$.

The EMPIRICAL MEASURE ν_n pertaining to $F_n = (F_n(\underline{t}))_{\underline{t} \in I^d}$ is given by $\nu_n = n^{-1} \sum_{j \le n} \delta_{\eta_j}$, where $\delta_{\underline{y}}$ denotes the Dirac measure in $\underline{y} \in I^d$.

The following picture illustrates ν_n in comparison with ζ_n:

$d = 2$



$\circ \ = \eta_j(\omega), 1 \le j \le n,$ RANDOM LOCATIONS

$\equiv n^{-1}$ FIXED MASSES

$\circ \ = \underline{j}/n, \underline{j} \in J_n,$ FIXED LOCATIONS

(+) (-) RANDOM MASSES (positive (+) or negative (-))

Generalizations of the FCLT's 1.1.6 and 1.2.1 to the multivariate case were obtained by Bickel and Wichura [Bi71], Neuhaus [Ne71] and Straf [Str71] after having extended the Skorokhod-metric to $D(I^d), d > 1$, to ensure the necessary measurability of the processes considered. In contrast, based on the concept of weak convergence ($\mathcal{L}$-convergence) of Hoffmann-Jørgensen [Ho84], [Ho91] in Section 2.3 below, the corresponding FCLT's for α_n and ζ_n in the multivariate case can also be obtained in a much simpler way by choosing a proper metric space, endowed with its natural sup-metric, as sample space of the processes, where the α_n's adn ζ_n's need not be measurable as we shall see (cf. Section 7).

1.4 α_n and ζ_n as set-indexed processes

Identifying each $\underline{t} \in I^d$ with the quadrant $C := [0, \underline{t}] \subset I^d, d \ge 1$, one gets the representations

(1.4.1) $$\alpha_n(C) = n^{1/2}(\nu_n(C) - \nu(C)),\, C \in \mathcal{C},$$

$$\text{where } \ \nu_n(C) = n^{-1} \sum_{j \le n} 1_C(\eta_j) \quad \text{and}$$

$$\nu := \text{ Lebesgue measure on } I^d, \text{ and}$$

(1.4.2) $$\zeta_n(C) = n^{-d/2} \sum_{\underline{j} \in J_n} 1_C(\underline{j}/n)\xi_{\underline{j}} \quad, C \in \mathcal{C},$$

(1.4.3) $$\text{with } \ \mathcal{C} := \{[0, \underline{t}] : \underline{t} \in I^d\}.$$

Both processes can be considered as set-indexed Partial-sum processes with random or fixed locations. Their sample paths are contained in the Banachspace

$$l^\infty(\mathcal{C}) := \{x : \mathcal{C} \longrightarrow \mathbb{R} \, : ||x||_{\mathcal{C}} := \sup_{C \in \mathcal{C}} |x(C)| < \infty\},$$

endowed with the sup-norm $|| \cdot ||_{\mathcal{C}}$.

Moreover, both processes will occur as special cases of so-called RANDOM MEASURE PROCESSES to be considered in Section 3.

1.5 A first glance at Glivenko-Cantelli convergence and Vapnik-Chervonenkis classes of sets

(Cf. also [Gae79] at this place.)

Let $\xi_j, j \in \mathbb{N}$, be i i d rv's with $\nu := \mathcal{L}\{\xi_j\}$, defined on a basic p-space $(\Omega, \mathcal{A}, \mathbb{P})$; then the classical GLIVENKO-CANTELLI THEOREM ([Gl33], [Ca33]) states:

(1.5.1)
$$\forall \nu \quad \sup_{C \in \mathcal{C}} |\nu_n(C) - \nu(C)| \longrightarrow 0 \quad \mathbb{P} - a.s.,$$
$$\text{where } \nu_n(C) = n^{-1} \sum_{j \leq n} 1_C(\xi_j) \quad \text{and}$$
$$\mathcal{C} := \{(-\infty, t] : t \in \mathbb{R}\}.$$

There are a lot of generalizations of (1.5.1) in the literature. Let us mention here only a few of them:

(1.5.2) Wolfowitz[Wo60], Dehardt [De71]

(1.5.1) is also valid for re's ξ_j in $(\mathbb{R}^d, \mathcal{B}^d), d \geq 1$,
being i i d with law ν and with $\mathcal{C} := \{(-\infty, \underline{t}] : \underline{t} \in \mathbb{R}^d\}$.
($\mathcal{B}^d :=$ Borel $\sigma -$ field in $\mathbb{R}^d$.)

(1.5.3) Ranga Rao [Ra62]

(1.5.1) is also valid for re's ξ_j in $(\mathbb{R}^d, \mathcal{B}^d), d \geq 1$,
being i i d with law ν and with $\mathcal{C} := \{C = \bigcap_{i \leq m} H_i : H_i$ halfspace in $\mathbb{R}^d\}$,
where $m \in \mathbb{N}$ is arbitrary but fixed.

(1.5.4) Elker-Pollard-Stute [El79]

(1.5.1) is also valid for re's ξ_j in $(\mathbb{R}^d, \mathcal{B}^d), d \geq 1$,
being i i d with law ν and with $\mathcal{C} := \{C \subset \mathbb{R}^d : C$ closed Euclidian ball$\}$.

The proofs of (1.5.2) - (1.5.4) are mainly based on *geometric arguments*.

That Glivenko-Cantelli convergence fails to hold *for any* ν when choosing larger classes of sets can be seen from the following

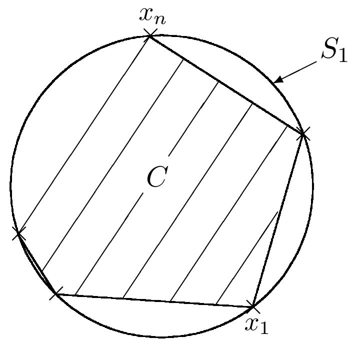
1.5.5. Example.

Let $d \geq 2$ and $\mathcal{C} := \{C \subset \mathbb{R}^d : C$ convex Borel set$\}$; let ν be the (normalized) uniform distribution on the unit sphere S_1 in $\mathbb{R}^d$ and $\xi_j, j \in \mathbb{N}$, be identically distributed with $\mathcal{L}\{\xi_j\} = \nu$ (defined on $(\Omega, \mathcal{A}, \mathbb{P})$). Let ν_n be the empirical measure based on $\xi_1, ..., \xi_n$; then $\forall n \in \mathbb{N}$

$$\sup_{C \in \mathcal{C}} |\nu_n(C) - \nu(C)| = 1 \qquad \mathbb{P} - a.s.$$

In fact, given any $x_j = \xi_j(\omega), 1 \leq j \leq n$, where $\xi_j(\omega) \in S_1$ for $\mathbb{P}$-almost all $\omega \in \Omega$, there exists a $C \in \mathcal{C}$ with $C \subset \{z \in \mathbb{R}^d : |z| \leq 1\}$ s.t. $C \cap S_1 = \{x_1, ..., x_n\}$: Choose $C := co(\{x_1, ..., x_n\})$ where $co(A)$ denotes the convex hull of $A \subset \mathbb{R}^d$. But then $\nu_n(C) = 1$, wheras $\nu(C) = 0$.

$d = 2:$



Choosing even $\mathcal{C} = \mathcal{B}^d$ one gets for i i d re's ξ_j in $(\mathbb{R}^d, \mathcal{B}^d)$ with law ν:

1.5.6. Lemma.

The following assertions are equivalent:

(i) $\exists \Omega_0 \in \mathcal{A}$ with $\mathbb{P}(\Omega_0) = 1$ s.t. $\quad \sup_{C \in \mathcal{B}^d} |\nu_n(C, \omega) - \nu(C)| \longrightarrow 0 \quad \forall \omega \in \Omega_0$.

(ii) ν is discrete, i.e. $\nu = \sum_{i \in N} m_i \delta_{x_i}, x_i \in \mathbb{R}^d, m_i > 0, \sum_{i \in N} m_i = 1$, $N \subset \mathbb{N}$.

PROOF. (i) $\Longrightarrow$ (ii): By assumption there exists $\Omega_0 \in \mathcal{A}$ with $\mathbb{P}(\Omega_0) = 1$ s.t. $\quad \sup_{C \in \mathcal{B}^d} |\nu_n(C, \omega) - \nu(C)| \longrightarrow 0 \quad \forall \omega \in \Omega_0$; thus $\Omega_0 \neq \varnothing$ and for $\omega_0 \in \Omega_0$ we have $\lim_{n\to\infty} \nu_n(C, \omega_0) = \nu(C)$ $\forall C \in \mathcal{B}^d$, whence for $C_0 := \{\xi_j(\omega_0) : j \in \mathbb{N}\} \in \mathcal{B}^d \quad \nu(C_0) = \lim_{n\to\infty} \nu_n(C_0, \omega_0) = 1$, since $\nu_n(C_0, \omega_0) = 1$ $\forall n \in \mathbb{N}$ by definition of C_0. But $\nu(C_0) = 1$ implies that ν is discrete.

(ii) $\Longrightarrow$ (i): Let ν be discrete, i.e. $\nu = \sum_{i \in N} m_i \delta_{x_i}, x_i \in \mathbb{R}^d, m_i > 0, \sum_{i \in N} m_i = 1, N \subset \mathbb{N}$; then, by the strong law of large numbers, there exists $N_1 \in \mathcal{A}$ with $\mathbb{P}(N_1) = 0$ s.t. $\forall \omega \in \complement N_1$ and $\forall i \in N \quad \lim_{n\to\infty} \nu_n(\{x_i\}, \omega) = \nu(\{x_i\})$. Furthermore, since ν concentrates on $D := \{x_i : i \in N\}$, there exists $N_2 \in \mathcal{A}$ with $\mathbb{P}(N_2) = 0$ s.t. $\forall \omega \in \complement N_2$ and $\forall n \in \mathbb{N} \quad \nu_n(A, \omega) = 0$ $\forall A \subset \mathbb{R}^d \backslash D$. Therefore, $\forall \omega \in \complement(N_1 \cup N_2)$ $(\nu_n(\,\cdot\,, \omega))_{n \in \mathbb{N}}$ is a sequence of p-measures on $(D, \mathcal{P}(D))$ (where $\mathcal{P}(D)$ denotes the power set of D) which converges pointwise (i.e. $\forall x_i : i \in N$)

towards ν. Applying Scheffé's lemma yields

$$\sum_{i \in N} |\nu_n(\{x_i\},\omega) - \nu(\{x_i\})| \longrightarrow 0 \quad \text{ as } n \to \infty,$$

and therefore $\lim_{n\to\infty} \left(\sup_{\Delta \in \mathcal{P}(D)} |\nu_n(\Delta,\omega) - \nu(\Delta)| \right) = 0$, yielding (i) with $\Omega_0 := \complement(N_1 \cup N_2)$. $\square$

In the following let $X = (X, \mathcal{X})$ be an arbitrary measurable space, $\xi_j, j \in \mathbb{N}$, be i i d re's in X with $\nu := \mathcal{L}\{\xi_j\}$ defined as coordinate projections on the p-space

$$(\Omega, \mathcal{A}, \mathbb{P}) := (X^{\mathbb{N}}, \mathcal{X}^{\mathbb{N}} \equiv \bigotimes_{\mathbb{N}} \mathcal{X}, \nu^{\mathbb{N}} \equiv \times_{\mathbb{N}} \nu);$$

this is what we call *CANONICAL MODEL* which will always be imposed as our basic p-space when dealing with i i d re's in X.

Let ν_n be the empirical measure based on $\xi_1, ..., \xi_n$, i.e.

(1.5.7) $$\nu_n(B) := n^{-1} \sum_{j \le n} 1_B(\xi_j) \equiv n^{-1} \sum_{j \le n} \delta_{\xi_j}(B)\,,\ B \in \mathcal{X}.$$

Now, especially from the statistical point of view (i.e. when ν is *unknown*), it is of interest to know whether

(1.5.8) $$\forall \nu \quad \sup_{C \in \mathcal{C}} |\nu_n(C) - \nu(C)| \longrightarrow 0 \ \mathbb{P} - a.s.,$$

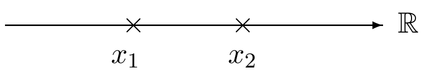
for suitable classes $\mathcal{C} \subset \mathcal{X}$ (taking over the role of the classes considered in the special cases (1.5.1) - (1.5.4)).
According to 1.5.5 and 1.5.6 the classes $\mathcal{C} \subset \mathcal{X}$ for which (1.5.8) holds true are not allowed to be too "rich". As we shall see later in Section 6.3, up to measurability, (1.5.8) will hold true in case of i i d re's ξ_j in X, if $\mathcal{C} \subset \mathcal{X}$ is a so-called *VAPNIK-CHERVONENKIS CLASS* (VCC), i.e. if $\mathcal{C}$ fulfills

(1.5.9) $$\exists s \in \mathbb{N} \text{ s.t. } \forall F \subset X \text{ with } |F| = s \quad \Delta^{\mathcal{C}}(F) < 2^s,$$

where $\Delta^{\mathcal{C}}(F) := |\{F \cap C : C \in \mathcal{C}\}|$. (1.5.9) means that a VCC is not too rich in a combinatorial sense, namely that from a certain s on "no s-element subset of X can be shattered by $\mathcal{C}$" (i.e. $\forall F \subset X$ with $|F| = s$ there is at least one $F' \subset F$ for which $F' \neq F \cap C \ \forall C \in \mathcal{C}$).

Note that for $F \subset X$ with $|F| = n$ $\Delta^{\mathcal{C}}(F) \le 2^n \equiv$ number of all subsets of F including the empty set, i.e. the case were $F \cap C = \varnothing$ is also counted here and in the following.)

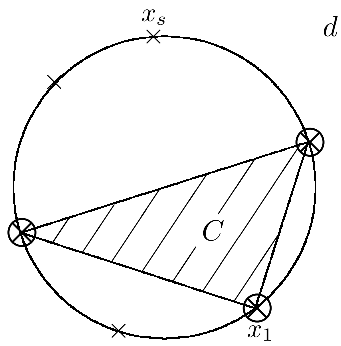
In the special case $X = \mathbb{R}$ and $\mathcal{C} = \{(-\infty, t] : t \in \mathbb{R}\}$ (1.5.9) holds true with $s = 2$:



$$\forall F = \{x_1, x_2\}, x_1 < x_2 \quad \Longrightarrow$$
$$\{x_2\} \neq F \cap (-\infty, t] \ \ \forall t \in \mathbb{R}.$$

In contrast, considering again 1.5.5 and choosing for any $s \in \mathbb{N}$ $F := \{x_1, ..., x_s\}$ with pairwise different $x_i \in S_1$, it follows that every subset $F' = \{x_{i_1}, ..., x_{i_k}\}$ of F can be represented as $F' = F \cap C$

with a convex Borel set C:
Choose $C := co(\{x_{i_1}, ..., x_{i_k}\})$.



$d = 2$

$F = \{x_1, ..., x_s\} \subset S_1$
$F' = \{\bigotimes\}$
$C = co(F')$.

The next example of Durst and Dudley [Dur80] shows that (1.5.8) may fail to hold for a VCC without imposing additional measurability assumptions (cf. [Gae83], p.37-38):

1.5.10. Example.
Let $X = (X, <)$ be an uncountable well-ordered set such that all its initial segments $\{x \in X : x < y\}, y \in X$, are countable (cf. [Ke61], p.29-). Then $\mathcal{C} := \{\{x \in X : x < y\}, y \in X\}$ does not shatter any $F \subset X$ with $|F| = 2$ (in fact: $\forall F = \{x_1, x_2\} \subset X$ with $x_1 < x_2$ we have $\{x_2\} \neq F \cap C \, \forall C \in \mathcal{C}$, since $x_2 \in \mathcal{C}$ would necessarily imply that $x_1 \in C \; \forall C \in \mathcal{C}$).
Note that $\mathcal{C}$ is linearly ordered by inclusion.
Now, by choosing ν properly, we will see that

$$\sup_{C \in \mathcal{C}} |\nu_n(C) - \nu(C)| \equiv 1 :$$

For this, let $\mathcal{X} := \{B \subset X : B$ countable or $\complement B$ countable$\}$, and let ν on $\mathcal{X}$ be defined by

$$\nu(B) := \begin{cases} 0 & , \text{ if } B \text{ is countable} \\ 1 & , \text{ if } \complement B \text{ is countable} \end{cases} \quad , B \in \mathcal{X}.$$

Then $\mathcal{C} \subset \mathcal{X}$ and $\nu(C) = 0 \; \forall C \in \mathcal{C}$.
On the other hand, given any observations $x_i, 1 \leq i \leq n, n \in \mathbb{N}$, of i i d re's $\xi_1, ..., \xi_n$ in $X = (X, \mathcal{X})$ with $\mathcal{L}\{\xi_j\} = \nu$, there exists a $C \in \mathcal{C}$ s.t. $x_i \in C \;\; \forall 1 \leq i \leq n$, whence

$$\sup_{C \in \mathcal{C}} |\nu_n(C) - \nu(C)| \equiv 1.$$

To avoid discussions about measurability assumptions we shall usually assume for simplicity that the index sets like $\mathcal{C} \subset \mathcal{X}$ are *countable.*

Even so, note that in cases where

$$\sup_{C\in\mathcal{C}} |\nu_n(C)-\nu(C)| = \sup_{C\in\mathcal{C}_0} |\nu_n(C)-\nu(C)|$$

with a countable $\mathcal{C}_0 \subset \mathcal{C}$, this is no restriction.
Also in case of empirical processes and partial-sum processes considered in 1.1 and 1.2, respectively, their sample paths are completely determined through the behaviour on a countable index set.

When considering later classes $\mathcal{F}$ of measurable functions $f : X \longrightarrow \mathbb{R}$ (instead of $1_{\mathcal{C}} \equiv \{1_C : C \in \mathcal{C}\}$) there will be instances where $\forall \delta_j \in \{-1,1\}, 1 \le j \le n$,

$$(+) \qquad ||\sum_{j\le n} \delta_j f(x_j)||_{\mathcal{F}} := \sup_{f\in\mathcal{F}} |\sum_{j\le n} \delta_j f(x_j)| = \sup_{f\in\mathcal{F}_0} |\sum_{j\le n} \delta_j f(x_j)|$$

for a countable subclass $\mathcal{F}_0$ of $\mathcal{F}$, implying measurability of $(x_1,...,x_n) \longmapsto ||\sum_{j\le n} \delta_j f(x_j)||_{\mathcal{F}}$. The underlying measurability concept can be found in [Va96], Example 2.3.4, called there "Pointwise measurability of $\mathcal{F}$" which means that there exists a countable $\mathcal{F}_0 \subset \mathcal{F}$ s.t. $\forall f \in \mathcal{F}$ there exists a sequence $(f_n) \subset \mathcal{F}_0$ with $f_n(x) \longrightarrow f(x)\, \forall x \in X$. In fact, this property implies (+): It is enough to show that for any $\varepsilon > 0$ there exists $f_{n_0} \in \mathcal{F}_0$ s.t. $|\sum_{j\le n} \delta_j f_{n_0}(x_j)| > ||\sum_{j\le n} \delta_j f(x_j)||_{\mathcal{F}} - \varepsilon$. For this, choose $f \in \mathcal{F}$ with $|\sum_{j\le n} \delta_j f(x_j)| > ||\sum_{j\le n} \delta_j f(x_j)||_{\mathcal{F}} - \varepsilon/2$ and $(f_m) \subset \mathcal{F}_0$ s.t. $f_m(x_j) \longrightarrow f(x_j)$ as $m \longrightarrow \infty \quad \forall\, 1 \le j \le n$ which implies $|\sum_{j\le n} \delta_j f_m(x_j)| \longrightarrow |\sum_{j\le n} \delta_j f(x_j)|$ whence there exists an n_0 s.t. $|\sum_{j\le n} \delta_j f_{n_0}(x_j)| > ||\sum_{j\le n} \delta_j f(x_j)||_{\mathcal{F}} - \varepsilon$.

For more about measurability concepts we refer to [Du99].

When restricting to *countable* $\mathcal{C} \subset \mathcal{X}$ one may wonder if one ends up with a VCC; this is not the case as seen by the following example:

Let $(X,\mathcal{X}) := (\mathbb{R},\mathcal{B}), J_1 := \{[a,b] : a < b, a, b \in \mathbb{Q}\}$ and $\forall n \in \mathbb{N} \quad J(n) := \bigcup_{i\le n} J_i$ with $J_i \equiv J_1$. Then $\mathcal{C} := \bigcup_{n\in\mathbb{N}} J(n)$ is a countable subclass of $\mathcal{B}$ with with the following property:

$$\forall n \in \mathbb{N}\ \exists F \subset \mathbb{R} \text{ with } |F| = n\ s.t.\ |\{F \cap C : C \in \mathcal{C}\}| = 2^n,$$

i.e. $\mathcal{C}$ is not a VCC.

More about VCC's in arbitrary sample spaces $X = (X,\mathcal{X})$ and so-called Vapnik-Chervonenkis graph classes (VCGC) of $\mathcal{X}$-measurable functions $f : X \longrightarrow \mathbb{R}$ will be contained in Sections 4.2 and 4.3 below.

2 Empirical measures in general sample spaces

2.1 Empirical discrepancies, Glivenko-Cantelli convergence and some consequences in statistics

Let $X = (X, \mathcal{X})$ be an arbitrary measurable space serving as sample space of i i d re's $\xi_j, j \in \mathbb{N}$, with $\mathcal{L}\{\xi_j\} = \nu$, defined as coordinate projections on $(\Omega, \mathcal{A}, \mathbb{P}) := (X^{\mathbb{N}}, \mathcal{X}^{\mathbb{N}}, \nu^{\mathbb{N}})$, i.e. our basic model will be the canonical one as introduced in Section 1.5.

Let ν_n be the empirical measure based on $\xi_1, ..., \xi_n$ (cf. (1.5.7)) and let $\mathcal{C} \subset \mathcal{X}$ be arbitrary but countable for simplicity. The so-called EMPIRICAL DISCREPANCY is defined by

$$||\nu_n - \nu||_{\mathcal{C}} := \sup_{C \in \mathcal{C}} |\nu_n(C) - \nu(C)| \tag{2.1.1}$$

(Since $\mathcal{C}$ is supposed to be countable, $||\nu_n - \nu||_{\mathcal{C}}$ is a rv, defined on $(\Omega, \mathcal{A}, \mathbb{P})$.)

The empirical discrepancies have the following property; in case of arbitrary (i.e. not necessarily countable) index sets we refer to [St95]:

2.1.2. Lemma.
$(||\nu_n - \nu||_{\mathcal{C}})_{n\in\mathbb{N}}$ is a reversed sub-martingale w.r.t. the sequence $(\mathcal{G}_n)_{n\in\mathbb{N}}$ of σ-fields

$$\mathcal{G}_n := \sigma\big(\{\nu_k(B) : k \geq n, B \in \mathcal{X}\}\big),$$

i.e. $||\nu_n - \nu||_{\mathcal{C}}$ is $\mathcal{G}_n$-measurable and $\mathbb{P}$-integrable $\forall n \in \mathbb{N}$, and $\forall n, m \in \mathbb{N}$ with $m \leq n$ one has

$$||\nu_n - \nu||_{\mathcal{C}} \leq \mathbb{E}(||\nu_m - \nu||_{\mathcal{C}}|\mathcal{G}_n) \quad \mathbb{P} - a.s. \tag{2.1.3}$$

PROOF. As shown in [Gae77], 6.5.5(c), the following holds:
$\forall C \in \mathcal{C}$ the sequence $(\nu_n(C) - \nu(C))_{n\in\mathbb{N}}$ is a reversed martingale w.r.t. $(\mathcal{G}_n)_{n\in\mathbb{N}}$, i.e. $\forall n, m \in \mathbb{N}$ with $m \leq n$ one has

$$\nu_n(C) - \nu(C) = \mathbb{E}((\nu_m(C) - \nu(C))|\mathcal{G}_n) \quad \mathbb{P} - a.s.;$$

therefore, since $\mathcal{C}$ is countable, it follows that $\mathbb{P} - a.s.$

$$\sup_{C\in\mathcal{C}} |(\nu_n(C) - \nu(C)| =$$
$$\sup_{C\in\mathcal{C}} |\mathbb{E}((\nu_m(C) - \nu(C))|\mathcal{G}_n)| \leq \mathbb{E}(\sup_{C\in\mathcal{C}} |\nu_m(C) - \nu(C)|\,|\mathcal{G}_n),$$

i.e. (2.1.3). □

Now, as in the case of sub-martingales, there holds an analogous CONVERGENCE THEOREM FOR REVERSED SUB-MARTINGALES (cf. e.g. [Gae77], 6.5.10) stating that for any reversed sub-martingale $(T_n)_{n\in\mathbb{N}}$ (on some p-space $(\Omega, \mathcal{A}, \mathbb{P})$) w.r.t. a monotone decseasing sequence $(\mathcal{G}_n)_{n\in\mathbb{N}}$ of

sub-σ-fields of $\mathcal{A}$ satisfying the condition that $\inf_{n\in\mathbb{N}} \mathbb{E}(T_n) > -\infty$ there exists an $\mathbb{P}$-integrable rv. T_∞ s.t. $T_n \longrightarrow T_\infty$ $\mathbb{P}$-a.s. and in the mean.
From this and Lemma 2.1.2 one obtains a rather simple proof of the following result (cf. [Po81]) which, in a similar form, was one of the main results in [Ste78] proved there with different methods based on ergodic theory of subadditive processes.

2.1.4. Lemma.
Let $(\tau_n)_{n\in\mathbb{N}}$ be an arbitrary sequence of non-negative integer-valued rv's on $(\Omega, \mathcal{A}, \mathbb{P})$ such that $\tau_n \xrightarrow{\mathbb{P}} \infty$ (where $\xrightarrow{\mathbb{P}}$ denotes convergence in probability); then

$$||\nu_n - \nu||_{\mathcal{C}} \longrightarrow 0 \quad \mathbb{P}-a.s. \iff ||\nu_{\tau_n} - \nu||_{\mathcal{C}} \xrightarrow{\mathbb{P}} 0;$$

in particular, $||\nu_n - \nu||_{\mathcal{C}} \xrightarrow{\mathbb{P}} 0 \implies ||\nu_n - \nu||_{\mathcal{C}} \longrightarrow 0 \quad \mathbb{P}-a.s.$*, whence*

$$\mathbb{E}(||\nu_n - \nu||_{\mathcal{C}}) \longrightarrow 0 \quad \implies \quad ||\nu_n - \nu||_{\mathcal{C}} \longrightarrow 0 \quad \mathbb{P}-a.s. \tag{2.1.5}$$

Note that (2.1.5)will lead later to an essential simplification in proving Glivenko-Cantelli convergence of $\nu_n = (\nu_n(C))_{C\in\mathcal{C}}$. Especially we will obtain along this way (cf. Section 6.3)) the following fundamental result of Vapnik-Chervonenkis ([Vap71]):

2.1.6. THEOREM.
Let $\mathcal{C} \subset \mathcal{X}$ be a VCC; then – under appropriate measurability conditions – it is true that $\forall \nu$ one has

$$||\nu_n - \nu||_{\mathcal{C}} \longrightarrow 0 \quad \mathbb{P}-a.s.$$

PROOF OF 2.1.4 $\implies$: $\tau_n \xrightarrow{\mathbb{P}} \infty$ implies that for any subsequence $(\tau_{n'})$ of (τ_n) there exists a further subsequence $(\tau_{n''})$ s.t. $\tau_{n''} \longrightarrow \infty$ $\mathbb{P}$-a.s., whence $||\nu_{\tau_{n''}} - \nu|| \longrightarrow 0$ $\mathbb{P}$-a.s. as n'' tends to infinity, and therefore $||\nu_{\tau_n} - \nu||_{\mathcal{C}} \xrightarrow{\mathbb{P}} 0$.
$\Longleftarrow$: According to 2.1.2 $(||\nu_n - \nu||_{\mathcal{C}})_{n\in\mathbb{N}}$ is a reversed sub-martingale. It is uniformly bounded; therefore, by the convergence theorem for reversed sub-martingales mentioned before, there exists an $\mathbb{P}$-integrable rv T_∞ s.t. $||\nu_n - \nu||_{\mathcal{C}} \longrightarrow T_\infty$ $\mathbb{P}$-a.s. From this it follows as in the first part of our proof that $||\nu_{\tau_n} - \nu||_{\mathcal{C}} \xrightarrow{\mathbb{P}} T_\infty$ whence, by assumption, it follows that $T_\infty = 0$ $\mathbb{P}$-a.s. □

Some consequences in statistics

In his book on Probability Theory Alfred Rényi considers the (classical) Glivenko-Cantelli theorem to be the “Fundamental Theorem of Mathematical Statistics” ([Re70], Chap.VII, §8). Given data $x_1, x_2, ...$ viewed as realizations of re's $\xi_1, \xi_2, ...$ in $(X, \mathcal{X})$ with $\mathcal{L}\{\xi_i\} = \nu$, Theorem 2.1.6 yields information about an unknown ν through its “statistical pictures” in form of the empirical measures

ν_n, e.g. in connection with a test for the null-hypothesis $H^0 : \nu = \nu_0$, ν_0 being a given hypothetical distribution on $\mathcal{X}$, versus the alternative $H^1 : \nu \neq \nu_0$. In the classical case the corresponding Kolmogorov-test is based on the test-statistic

$$D_n(\mathcal{C}, \nu_0) \equiv ||\nu_n - \nu_0||_{\mathcal{C}}$$

with (cf. (1.5.2)) $\mathcal{C} := \{(-\infty, \underline{t}] : \underline{t} \in \mathbb{R}^d\}$:

$$Reject\ \ H^0 \quad if \quad D_n(\mathcal{C}, \nu_0) > c, \quad c > 0.$$

Another possibility would be to use a Kolmogorov-test based on

$$D_n(\mathcal{C}_0, \nu_0) \quad \text{with } \mathcal{C}_0 := \{x + C_0 : x \in \mathbb{R}^d\},$$

where C_0 is a given closed Euclidian ball. Also in this case one has for any ν that for the so-called "scan-statistic"

$$\lim_{n\to\infty} D_n(\mathcal{C}_0, \nu) = 0 \qquad \mathbb{P} - a.s. \ \ (cf.(1.5.4)),$$

and under $H^1 : \nu \neq \nu_0$ one has

$$\lim_{n\to\infty} D_n(\mathcal{C}_0, \nu_0) = d \quad \mathbb{P} - a.s.,$$

where $d := ||\nu - \nu_0||_{\mathcal{C}} > 0$ (cf. [Py84], Theorem 6.1), i.e. Kolmogorov-tests based on $D_n(\mathcal{C}_0, \nu_0)$ are also consistent against all alternatives.

Furthermore, simulation results in [Py84], Section 6, indicated a considerable improvement in power that is possible when using the scan-statistic $D_n(\mathcal{C}_0, \nu_0)$ instead of $D_n(\mathcal{C}, \nu_0)$; cf. also the very interesting Monte-Carlo study of Pyke and Wilbour ([Py88]) concerning the power of such tests; as mentioned in [We92] it would be of some interest to have available sufficient theory in order to theoretically compute (or at least approximate) the power of their tests.

2.2 Functional Central Limit Theorems for set-indexed empirical and partial-sum processes, respectively

Let $X = (X, \mathcal{X})$ be an arbitrary sample space, $\eta_j, j \in \mathbb{N}$, be i i d re's in X, $\nu = \mathcal{L}\{\eta_j\}$, and ν_n be the empirical measure based on $\eta_1, ..., \eta_n$. Let $\mathcal{C} \subset \mathcal{X}$ be a VCC and $\beta_n = (\beta_n(C))_{C\in\mathcal{C}}$ be the *empirical* $\mathcal{C}$*-process* (of sample size n), defined by

$$\beta_n(C) := n^{1/2}(\nu_n(C) - \nu(C)), \quad C \in \mathcal{C}.$$

Then, under appropriate measurability conditions, the following generalization of Theorem 1.1.6 has been obtained by Dudley:

2.2.1. THEOREM *([Du78]).*

$$\beta_n \xrightarrow{\mathcal{L}} \mathbb{G}_\nu \quad in \ \ (l^\infty(\mathcal{C}), ||\cdot||_{\mathcal{C}}),$$

where $\mathbb{G}_\nu = (G_\nu(C))_{C\in\mathcal{C}}$ *is a mean-zero Gaussian process whose covariance structure is given by*

$$cov(G_\nu(C_1), G_\nu(C_2)) = \nu(C_1 \cap C_2) - \nu(C_1)\,\nu(C_2), \quad C_1, C_2 \in \mathcal{C},$$

and $\xrightarrow{\mathcal{L}}$ *-convergence is defined as in Section 2.3 below.*

The sample paths of $\mathbb{G}_\nu$ are contained in the space

$$U^b(\mathcal{C}, d_\nu) := \{x \in l^\infty(\mathcal{C}) : x \ \text{ uniformly } d_\nu\text{-continuous}\},$$

d_ν being the pseudo-metric in $\mathcal{C}$, defined by $d_\nu(C_1, C_2) := \nu(C_1 \Delta C_2)$, $C_1, C_1 \in \mathcal{C}$.

($C_1 \Delta C_2$ denotes the symmetric difference between C_1 and C_2; note that a pseudo-metric has all properties of a metric besides that $d_\nu(C_1, C_2) = 0$ does not imply $C_1 = C_2$.)

Compare Theorem 2.2.1 with Theorem 1.1.6 in case of the uniform empirical process α_n, where $X = I \equiv [0,1], \mathcal{X} = I \cap \mathcal{B}, \nu =$ Lebesgue measure on $\mathcal{X}$, $\mathcal{C} = \{[0,t] : t \in I\}$ being a VCC, $d_\nu(C_1, C_2) = |t_1 - t_2|$ for $C_i = [0, t_i]$ and where $U^b(\mathcal{C}, d_\nu) = C \equiv C(I)$.

Functional Central Limit Theorems (FCLT's) for set-indexed partial-sum processes have been obtained by the SEATTLE-SCHOOL around Ron Pyke: cf. [Py84], [Os84], [Ba85], [Os85], [Al86], and Section 7.2 below.

At this place here we want to mention only the following two results. The first is concerned with the multivariate (standardized) partial-sum process $\zeta_n = (\zeta_n(C))_{n\in\mathbb{N}}$ of Section 1.4 (cf. (1.4.2) and (1.4.3)) generalizing Theorem 1.2.1:

2.2.2. THEOREM *([Al86] and [Gae94], Remark 2.16).*

$$\zeta_n \xrightarrow{\mathcal{L}} \mathbb{G}_\nu \quad in \ \ (l^\infty(\mathcal{C}), ||\cdot||_{\mathcal{C}}), \quad \nu \equiv \ Lebesgue \ measure \ on \ I^d,$$

where $\mathbb{G}_\nu = (G_\nu(C))_{C\in\mathcal{C}}$ *is a mean-zero Gaussian process whose covariance structure is given by*

$$cov(G_\nu(C_1), G_\nu(C_2)) = \nu(C_1 \cap C_2), \quad C_1, C_2 \in \mathcal{C},$$

and again $\xrightarrow{\mathcal{L}}$ *-convergence is defined as in Section 2.3 below.*
Also here the sample paths of $\mathbb{G}_\nu$ *are contained in* $U^b(\mathcal{C}, d_\nu)$.

The second result is concerned with set-indexed partial-sum processes with random locations (cf. 3.2.1 and 7.2 below):

2.2.3. THEOREM *(cf. [Gae94], Cor. 2.15).*
Let $\mathcal{C} \subset \mathcal{X}$ be a VCC in an arbitrary sample space $X = (X, \mathcal{X})$ and $\xi_{nj} := j(n)^{-1/2}\xi_j$ for each $1 \le j \le j(n)$ and $n \in \mathbb{N}$ with $j(n) \longrightarrow \infty$ as $n \longrightarrow \infty$, the ξ_j's being i i d rv's with $\mathbb{E}(\xi_j) = 0$ and $\mathbb{E}(\xi_j^2) = 1$. Let $(\eta_{nj})_{1 \le j \le j(n), n \in \mathbb{N}}$ be a triangular array of rowwise independent (but not necessarily identically destributed) re's in X which is independent of $(\xi_j)_{j \in \mathbb{N}}$.
Assume that there is a p-measure ν on $\mathcal{X}$ s.t. with $\nu_{nj} := \mathcal{L}\{\eta_{nj}\}$ the following two conditions are fulfilled:

(i) $\quad \lim_{n\to\infty} j(n)^{-1} \sum_{j \le j(n)} \nu_{nj}(C \cap D) = \nu(C \cap D) \quad \forall C, D \in \mathcal{C}$

(ii) $\quad \lim_{\delta\to 0} \limsup_{n\to\infty} \sup_{\{C \in \mathcal{C}: \nu(C) \le \delta\}} j(n)^{-1} \sum_{j \le j(n)} \nu_{nj}(C) = 0.$

Then

$$\Big(j(n)^{-1/2} \sum_{j \le j(n)} 1_C(\eta_{nj}) \cdot \xi_j\Big)_{C \in \mathcal{C}} \xrightarrow{\mathcal{L}} \mathbb{G}_\nu \quad \text{in } (l^\infty(\mathcal{C}), ||\cdot||_{\mathcal{C}}),$$

where $\mathbb{G}_\nu = (G_\nu(C))_{C \in \mathcal{C}}$ is a mean-zero Gaussian process whose covariance structure is given by

$$cov(G_\nu(C_1), G_\nu(C_2)) = \nu(C_1 \cap C_2), \quad C_1, C_2 \in \mathcal{C},$$

and again $\xrightarrow{\mathcal{L}}$ -convergence is defined as in Section 2.3 below. Also here the sample paths of $\mathbb{G}_\nu$ are contained in $U^b(\mathcal{C}, d_\nu)$.

2.3 Weak convergence ($\mathcal{L}$-convergence) in the sense of Hoffmann-Jørgensen

The classical concept of weak convergence (convergence in law) for random elements (re's) $\eta_n, n \ge 0$, in a metric space $S = (S, \mathcal{B}(S))$, endowed with its Borel σ-field $\mathcal{B}(S)$, is defined by (cf. [Bi68])

$$\eta_n \xrightarrow{\mathcal{L}} \eta_0 \; :\Longleftrightarrow \; \lim_{n\to\infty} \mathbb{E}(f \circ \eta_n) = \mathbb{E}(f \circ \eta_0) \;\; \forall f \in C^b(S) \tag{2.3.1}$$

where $C^b(S) := \{f : S \longrightarrow \mathbb{R} : f \text{ continuous and bounded}\}$.

For such η_n's, being re's in $(S, \mathcal{B}(S))$, their laws $\mathcal{L}\{\eta_n\}$ are well defined on $\mathcal{B}(s)$, whence (2.3.1) is equivalent to

$$\eta_n \xrightarrow{\mathcal{L}} \eta_0 \iff \lim_{n\to\infty} \int_S f d\mathcal{L}\{\eta_n\} = \int_S f d\mathcal{L}\{\eta_0\} \quad \forall f \in C^b(S). \tag{2.3.1$'$}$$

But, as we have learned from the uniform empirical process, the approximating sequence $(\eta_n)_{n\in\mathbb{N}}$ of a limitting re η_0 may **not** be ad hoc measurable and this leads to the concept of weak convergence ($\mathcal{L}$-convergence) in the sense of Hoffmann-Jørgensen ([Ho84], [Ho91]). In this context, i.e. where the η_n's, $n \in \mathbb{N}$, are allowed to be completely arbitrary maps, we will speak of RANDOM QUANTITIES (rq's) instead of RANDOM ELEMENTS (re's). So, given a basic p-space $(\Omega, \mathcal{A}, \mathbb{P})$, let $\eta_n : \Omega \longrightarrow S$ be rq's and $\eta_0 : \Omega \longrightarrow S$ be $\mathcal{A}, \mathcal{B}(S)$-measurable (i.e. only η_0 is assumed to be a re in $(S, \mathcal{B}(S))$); then:

(2.3.2) $$\eta_n \xrightarrow{\mathcal{L}} \eta_0 \ :\Longleftrightarrow \ \lim_{n\to\infty} \mathbb{E}^*(f \circ \eta_n) = \mathbb{E}(f \circ \eta_0) \ \ \forall f \in C^b(S).$$

Here, given an arbitrary $g : \Omega \longrightarrow \overline{\mathbb{R}}$ (defined on a p-space $(\Omega, \mathcal{A}, \mathbb{P})$), the so-called "outer expectation" ("outer integral") of g w.r.t. $\mathbb{P}$ is defined by

(2.3.3) $$\mathbb{E}^*(g) := \inf\{\mathbb{E}(h) : h \geq g, \ h : \Omega \longrightarrow \overline{\mathbb{R}} \text{ measurable and } \mathbb{E}(h) \text{exists}\}.$$

In view of (2.3.2) one should note that $\mathbb{E}(f \circ \eta_0)$ is well defined, since $f \in C^b(S) \Longrightarrow f \ \mathcal{B}(S), \mathcal{B}$-measurable and bounded $\Longrightarrow f \circ \eta_0$ $\mathbb{P}$-integrable (i.e. $f \circ \eta_0 \in \mathcal{L}(\Omega, \mathcal{A}, \mathbb{P})$). If, in addition, also the η_n's, $n \in \mathbb{N}$, are $\mathcal{B}(S), \mathcal{B}$-measurable, then $\mathbb{E}^*(f \circ \eta_n) = \mathbb{E}(f \circ \eta_n)$, i.e. in this case (2.3.2) coinces with the classical definition (2.3.1).

In connection with (2.3.2) the following definitions and formulas are in order:
Let the so-called "inner expectation" ("inner integral") of $g : \Omega \longrightarrow \overline{\mathbb{R}}$ w.r.t. $\mathbb{P}$ be defined by

$$\mathbb{E}_*(g) := \sup\{\mathbb{E}(h) : h \leq g, \ h : \Omega \longrightarrow \overline{\mathbb{R}} \text{ measurable and } \mathbb{E}(h) \text{ exists}\},$$

then, for any $A \subset \Omega$, $\mathbb{E}_*(1_A) = \mathbb{P}_*(A) := \sup\{\mathbb{P}(B) : B \subset A, B \in \mathcal{A}\}$, whereas $\mathbb{E}^*(1_A) = \mathbb{P}^*(A) := \inf\{\mathbb{P}(B) : B \supset A, B \in \mathcal{A}\}$; furthermore

(2.3.4) $$\begin{aligned} &\mathbb{E}_*(g) = -\mathbb{E}^*(-g) \ , \ \mathbb{E}_*(g) \leq \mathbb{E}^*(g); \\ &g_1 \leq g_2 \quad \Longrightarrow \quad \mathbb{E}_*(g_1) \leq \mathbb{E}_*(g_2) \text{ and } \mathbb{E}^*(g_1) \leq \mathbb{E}^*(g_2) \\ &\mathbb{E}^*(g_1 + g_2) \leq \mathbb{E}^*(g_1) + \mathbb{E}^*(g_2) \\ &|\mathbb{E}^*(g_1) - \mathbb{E}^*(g_2)| \leq \mathbb{E}^*(|g_1 + g_2|) \quad \text{ if } |\mathbb{E}^*(g_i)| < \infty, i = 1, 2; \\ &\mathbb{E}_*(g) = \mathbb{E}^*(g) = \mathbb{E}(g), \quad \text{ if } g \in \mathcal{L}(\Omega, \mathcal{A}, \mathbb{P}); \\ &\mathbb{P}^*(A) + \mathbb{P}_*(\complement A) = 1 \quad \forall A \subset \Omega. \end{aligned}$$

For some applications it might be useful to allow also rq's $\eta_n, n \in \mathbb{N}$, with values in a larger space $E \supset S$; one may think (in case of stochastic processes with parameter set T) of $E = \mathbb{R}^T \supset S := l^\infty(T)$, where S is endowed with the sup-metric $||x||_T := \sup_{t \in T} |x(t)|$; this leads to the following *more general model of weak convergence ($\mathcal{L}$-convergence)* considered in [St94]:

(2.3.5) Let $S = (S, s)$ be a metric space (with metric s) and $E \supset S$ be arbitrary; let $\eta_n : \Omega \longrightarrow E$ be rq's, $n \in \mathbb{N}$, and $\eta_0 : \Omega \longrightarrow S$ be $\mathcal{A}, \mathcal{B}(S)$-measurable; then

(2.3.6) $$\eta_n \xrightarrow{\mathcal{L}} \eta_0 \ :\Longleftrightarrow \ \lim_{n\to\infty} \mathbb{E}^*(f \circ \eta_n) = \mathbb{E}(f \circ \eta_0)$$
$\forall f : E \longrightarrow \mathbb{R}$, f bounded and $rest_S(f) \in C^b(S)$, where $rest_S(f)$ denotes the restriction of f onto S.

If, in addition (compare with the classical situation of Section 1), for a *separable* subspace S_0 of S with $S_0 \in \mathcal{B}(S)$, $\mathbb{P}(\eta_0 \in S_0) = 1$, then *the limiting re η_0 is said to be separable* and in this case we write $\eta_n \xrightarrow[\text{sep}]{\mathcal{L}} \eta_0$.

Within this general model of $\mathcal{L}$-convergence the known results from the classical theory of weak convergence, like the Portmanteau-Theorem, Cramér-Slutzky-type result, Continuous Mapping Theorem, etc. remain valid as we shall see below.

In passing we mention the following two facts:

(2.3.7) (cf. [Va96], 1.3.7 and 1.3.8(i)): $\eta_n \xrightarrow{\mathcal{L}} \eta_0 \implies \mathbb{E}^*(f \circ \eta_n) - \mathbb{E}_*(f \circ \eta_n) \longrightarrow 0$ $\forall f : E \longrightarrow \mathbb{R}$, f bounded and $rest_S(f) \in C^b(S)$, i.e. the η_n's are "asymptotically measurable".

(2.3.8) $\eta_n \xrightarrow{\mathcal{L}} \eta_0 \quad \implies \quad \mathbb{P}_*(\eta_n \in S) \longrightarrow 1.$

In case of stochastic processes $\eta_n = (\eta_n(t))_{t \in T}$, indexed by a pseudo-metric parameter space $T = (T, d)$, being all defined on some basic p-space $(\Omega, \mathcal{A}, \mathbb{P})$, the following theorem characterizes weak convergence ($\mathcal{L}$-convergence), i.e. $\eta_n \xrightarrow[\text{sep}]{\mathcal{L}} \eta_0$ based on the situation (2.3.5) with $S = (l^\infty(T), ||\cdot||_T) \subset E = \mathbb{R}^T$.

2.3.9. CHARACTERIZATION THEOREM OF $\mathcal{L}$-CONVERGENCE (CT$\mathcal{L}$-C).
Let $\eta_n = (\eta_n(t))_{t \in T}, n \in \mathbb{N}$, be a sequence of stochastic processes, indexed by a pseudo-metric parameter space $T = (T, d)$, being all defined on some basic p-space $(\Omega, \mathcal{A}, \mathbb{P})$ and let $\overline{\eta}_0 = (\overline{\eta}_0(t))_{t \in T}$ be a stochastic process viewed as coordinate process on $(\mathbb{R}^T, \mathcal{B}^T, \mathcal{L}\{\overline{\eta}_0\})$ (where the law $\mathcal{L}\{\overline{\eta}_0\}$ of $\overline{\eta}_0$ is well defined on the product σ-field $\mathcal{B}^T \equiv \bigotimes_T \mathcal{B}$ according to Kolmogorov's theorem) such that

(2.3.10) $\eta_n \xrightarrow[fidi]{\mathbb{P}} \overline{\eta}_0,$ *i.e. weak convergence of the finite-dimensional distributions (fidis) of η_n to the corresponding fidis of $\overline{\eta}_0$.*

Then, if

(2.3.11) (T, d) *is totally bounded,*

and if the so-called "Asymptotic Equicontinuity Condition" (AEC) is fulfilled, i.e. if

(2.3.12) $$\lim_{\delta \to 0} \limsup_{n \to \infty} \mathbb{P}^*(w_{\eta_n}(\delta) > \varepsilon) = 0 \quad \forall \varepsilon > 0,$$

there exists a stochastic process $\eta_0 = (\eta_0(t))_{t \in T}$ with sample paths in $S_0 \equiv U^b(T, d)$ $((U^b(T, d), ||\cdot||_T)$ being a separable subspace of $S = (l^\infty(T), ||\cdot||_T))$ such that

(2.3.13) $$\eta_n \xrightarrow[sep]{\mathcal{L}} \eta_0,$$

where $\eta_0 \overset{\mathcal{L}}{\underset{fidi}{=}} \overline{\eta}_0$, i.e. η_0 and $\overline{\eta}_0$ have the same fidis.

Conversely, (2.3.13) (with $S_0 = (U^b(T, d), ||\cdot||_T)$ as separable subspace of $S = (l^\infty(T), ||\cdot||_T)$) implies (2.3.11) and (2.3.12).

Here, $U^b(T,d) := \{x \in l^\infty(T) : x \text{ uniformly } d\text{-continuous}\}$, and for any $x \in \mathbb{R}^T$ and $\delta > 0$

$$w_x(\delta) := \sup_{t,t' \in T, d(t,t') \leq \delta} |x(t) - x(t')|$$

is the oszillation-modulus of x. Note that (2.3.10) will be fulfilled in most of the later applications according to the classical multivariate CLT's.

There are several possibilities presented in the literature for proving the CT$\mathcal{L}$-C; cf. e.g. [Gi86], [An87], Theorem 5.5, [Po90], Theorem 10.2, [Du92], Theorem 3.7.2, [Gae92], Theorem 3.10, and [Va96], Section 1.5. Independently, we want to give here a different (and as we think rather lucid) proof of 2.3.9 based on the following auxiliary lemma and partially on ideas of [Po90] (cf. STEP 2 below).

AUXILIARY LEMMA *(Cf. [Bi68], Cor. 1, p. 14, and [Gae83], Thm. 8).*
Given the situation as in (2.3.5), let $S_0 \subset S$ be separable and $\mathbb{P}(\eta_0 \in S_0) = 1$. Suppose that the class

$$\mathcal{C} \subset \{B \in \mathcal{B}(S) : \mathbb{P}(\eta_0 \in \partial B) = 0\}$$

(where ∂B denotes the boundary of B) satisfies

($\star$) *$\forall G$ open $\subset S$ and $\forall x \in G \cap S_0^c$ (where S_0^c denotes the closure of S_0 in S)*
$\exists C_x \in \mathcal{C}$ s.t. $x \in C_x^0 \subset C_x \subset G$ (where C_x^0 denotes the interior of C_x).

Then the following two statements are equivalent:

(i) $\eta_n \xrightarrow[sep]{\mathcal{L}} \eta_0$

(ii) $\limsup \mathbb{P}^*(\eta_n \in C) \leq \mathbb{P}(\eta_0 \in C)$ *and*
$\liminf \mathbb{P}_*(\eta_n \in C) \geq \mathbb{P}(\eta_0 \in C) \quad \forall C \in \mathcal{C}^{\cap f}$,

where $\mathcal{C}^{\cap f}$ denotes the class of all subsets of S which are finite intersections of sets in $\mathcal{C}$.

PROOF OF THE CT$\mathcal{L}$-C 2.3.9 (carried out in three steps).

Assume (2.3.11) and (2.3.12).

STEP 1: According to (2.3.11) there exists a countable and d-dense subset D of T. We are going to show:

There exists a stochastic process $\eta_0 = (\eta_0(t))_{t\in T}$ with sample paths in $S_0 \equiv U^b(T,d)$ such that

$$\eta_{0,D} \overset{\mathcal{L}}{\underset{\text{fidi}}{=}} \overline{\eta}_{0,D}, \qquad \text{where } \eta_{0,D} := (\eta_0(t))_{t\in D} \text{ and } \overline{\eta}_{0,D} = (\overline{\eta}_0(t))_{t\in D}.$$

For this, let $U(D,d) := \{x : D \longrightarrow \mathbb{R} : x \text{ uniformly } d\text{-continuous}\}$. Then it suffices to show that

(a) there exists a stochastic process $\eta_{0,D} := (\eta_0(t))_{t\in D}$ on some proper p-space $(\Omega_0, \mathcal{A}_0, \mathbb{P}_0)$ with sample paths in $U(D,d)$ such that $\eta_{0,D} \overset{\mathcal{L}}{\underset{\text{fidi}}{=}} \overline{\eta}_{0,D}$.

In fact, once (a) is shown, we can define for each $\omega \in \Omega_0$ $\eta_0(\omega)$ as the uniquely determined uniformly d-continuous extension on T of $\eta_{0,D}(\omega)$ being also bounded since T is totally bounded, whence $\eta_0(\omega) \in U^b(T,d)$ for each $\omega \in \Omega_0$.

Now, verifying (a) is equivalent (cf. [Gae77], 7.2.31 and 7.1.18) with proving

(b) $\quad P_{\overline{\eta}_0}(\overline{\eta}_{0,D} \in U(D,d)) = 1, \quad$ where $\quad P_{\overline{\eta}_0} := \mathcal{L}\{\overline{\eta}_0\}$.

For this, let $D = \{t_1, t_2, ...\}$; then

$$\begin{aligned}
&P_{\overline{\eta}_0}(\overline{\eta}_{0,D} \in U(D,d)) \\
&= P_{\overline{\eta}_0}\Big(\forall \varepsilon > 0\, \exists \delta > 0\, \forall t, t' \in D: \quad d(t,t') \le \delta \implies |\overline{\eta}_0(t) - \overline{\eta}_0(t')| \le \varepsilon\Big) \\
&= P_{\overline{\eta}_0}\Big(\forall \varepsilon > 0\, \exists \delta > 0\, \forall m \in \mathbb{N}\, \forall 1 \le i,j \le m: \quad d(t_i,t_j) \le \delta \implies |\overline{\eta}_0(t_i) - \overline{\eta}_0(t_j)| \le \varepsilon\Big) \\
&= P_{\overline{\eta}_0}\Big(\forall \varepsilon > 0\, \exists \delta > 0\, \forall m \in \mathbb{N}\, \forall 1 \le i,j \le m: \quad (\overline{\eta}_0(t_1), ..., \overline{\eta}_0(t_m)) \in F_{ij}(\varepsilon,\delta,m)\Big),
\end{aligned}$$

where

$$F_{ij}(\varepsilon,\delta,m) := \begin{cases} \mathbb{R}^m, & \text{if } d(t_i,t_j) > \delta \\ \{(r_1,...,r_m) \in \mathbb{R}^m : |r_i - r_j| \le \varepsilon\}, & \text{if } d(t_i,t_j) \le \delta. \end{cases}$$

By the way, since the F_{ij}'s are closed and since we may restrict ourselves to rational ε's and δ's, this shows that $\{\overline{\eta}_{0,D} \in U(D,d)\}$ is measurable.

Furthermore, by σ-continuity of $P_{\overline{\eta}_0}$

$$\begin{aligned}
&P_{\overline{\eta}_0}(\overline{\eta}_{0,D} \in U(D,d)) = \\
&\lim_{\varepsilon\to 0}\lim_{\delta\to 0}\lim_{m\to\infty} P_{\overline{\eta}_0}\Big((\overline{\eta}_0(t_1), ..., \overline{\eta}_0(t_m)) \in \bigcap_{1\le i,j\le m} F_{ij}(\varepsilon,\delta,m)\Big) \ge \\
(+) \qquad &\lim_{\varepsilon\to 0}\lim_{\delta\to 0}\lim_{m\to\infty}\limsup_{n\to\infty} \mathbb{P}\Big((\eta_n(t_1), ..., \eta_n(t_m)) \in \bigcap_{1\le i,j\le m} F_{ij}(\varepsilon,\delta,m)\Big),
\end{aligned}$$

where the inequality follows by (2.3.10) and the classical Portmanteau-Theorem ([Bi68], Theorem 2.1 with $S = \mathbb{R}^m$); furthermore,

$$\begin{aligned}
(+) \quad &\ge \lim_{\varepsilon\to 0}\lim_{\delta\to 0}\limsup_{n\to\infty} \mathbb{P}^*\Big(\forall t, t' \in T: \quad d(t,t') \le \delta \implies |\eta_n(t) - \eta_n(t')| \le \varepsilon\Big) \\
&= \lim_{\varepsilon\to 0}\lim_{\delta\to 0}\limsup_{n\to\infty} \mathbb{P}^*\big(w_{\eta_n}(\delta) \le \varepsilon\big) \underset{(2.3.12)}{=} 1.
\end{aligned}$$

The proof of the following step is due to Franz Strobl ([St94], Thm. 2.1).

STEP 2: Using the auxiliary lemma from above, we are going to show now

(2.3.13) $\qquad \eta_n \xrightarrow[\text{sep}]{\mathcal{L}} \eta_0, \quad$ where w.l.o.g η_0 is assumed to be also defined on our basic p-space $(\Omega, \mathcal{A}, \mathbb{P})$ properly enlarged.

Since (by (2.3.11)), $S_0 = U^b(T,d)$ is a separable and closed subspace of $(S,s) = (l^\infty(T), ||\cdot||_T)$ (cf. [Gae90], Corollary 2), we can apply the auxiliary lemma. For this, let

$$\mathcal{C} := \{B(x,r) : x \in S_0, r > 0, \mathbb{P}(\eta_0 \in \partial B(x,r)) = 0\},$$

where $B(x,r) := \{y \in S : ||y - x||_T \le r\}$. Then one easily verifies

$(\star)$ $\quad \forall G$ open $\subset S$ and $\forall x \in G \cap S_0^c = G \cap S_0 \; \exists C_x \in \mathcal{C}$ s.t. $\; x \in C_x^0 \subset C_x \subset G$.

Therefore, to verify (2.3.13) it remains to show

(c) $\quad \limsup \mathbb{P}^*(\eta_n \in C) \le \mathbb{P}(\eta_0 \in C)$ and

(d) $\quad \liminf \mathbb{P}_*(\eta_n \in C) \ge \mathbb{P}(\eta_0 \in C)$ for all $C \in \mathcal{C}^{\cap f} = \{\bigcap_{i \le n} C_i : n \in \mathbb{N}, C_i \in \mathcal{C}, 1 \le i \le n\}$.

Now, given any $C \in \mathcal{C}^{\cap f}$ one can choose appropriate $g, h \in U^b(T,d)$ such that C can be represented as

$$C = \{y \in S : g(t) \le y(t) \le h(t) \;\; \forall t \in T\}.$$

(If $C = B(x,r)$, choose $g := x - r$ and $h := x + r$; in case of finite intersections of balls $B(x_i, r_i)$ one has to choose maxima and minima of such g_i's and h_i's, respectively.)

Next, given $C = \bigcap_{i \le n} B(x_i, r_i) = \{y \in S : g \le y \le h\}$ and an arbitrary $\varepsilon > 0$, choose $\lambda = \lambda(\varepsilon) > 0$ s.t. with $C_\lambda := \{y \in S : g + \lambda \le y \le h - \lambda\}$

(e) $\quad \mathbb{P}(\eta_o \in C) \le \mathbb{P}(\eta_0 \in C_{5\lambda/4}) + \varepsilon/2$.

Before making the next step rigorous, we argue at first informally:

By the AEC (2.3.12) one can choose $\delta > 0$ s.t. for n large enough up to probability $\varepsilon/2$ the oscillation of η_n within span δ is at most $\lambda/2$ and this is also true for g and h (due to their uniform continuity). Since T is totally bounded, we can choose a δ-net $\{t_1, ..., t_m\} \subset T$ (which means that for each $t \in T$ there is a t_i with $d(t, t_i) < \delta$); since the oscillations of g, h and η_n (up to probability $\varepsilon/2$) within $V(t_i) := \{t \in T : d(t, t_i) < \delta\}$ are at most $\lambda/2$, we get

(f) $\quad \mathbb{P}(\forall i \; : g(t_i) + \lambda < \eta_n(t_i) < h(t_i) - \lambda) \;\le\; \mathbb{P}_*(\eta_n \in C) + \varepsilon/2$,

whence by fidi-convergence we obtain (d):

$$\begin{aligned}
\mathbb{P}(\eta_0 \in C) &\underset{(e)}{\le} \mathbb{P}\big(\eta_0 \in C_{5\lambda/4}\big) + \varepsilon/2 \;\le\; \mathbb{P}\big(\forall i \; : g(t_i) + \lambda < \eta_0(t_i) < h(t_i) - \lambda\big) + \varepsilon/2 \\
&\underset{n\to\infty}{\longleftarrow} \mathbb{P}\big(\forall i \; : g(t_i) + \lambda < \eta_n(t_i) < h(t_i) - \lambda\big) + \varepsilon/2 \\
&\underset{(f)}{\le} \mathbb{P}_*(\eta_n \in C) + \varepsilon/2.
\end{aligned}$$

Now, making the above reasoning rigorous, note first that $C_\lambda \uparrow C^0$ as $\lambda \downarrow 0$ and $\mathbb{P}(\eta_0 \in \partial C) = 0$ implies that for each $\varepsilon > 0$ there exists a $\lambda = \lambda(\varepsilon) > 0$ s.t. (e) holds true.

Since $\mathbb{P}(\eta_0 \in U^b(T,d)) = 1$, there exists a $\delta = \delta(\varepsilon) > 0$ s.t. $\mathbb{P}(\eta_0 \in H) \ge 1 - \frac{\varepsilon}{2}$, where

$$H := \{y \in S : \sup_{t,t' \in T, d(t,t') < \delta} |y(t) - y(t')| \le \lambda/2\}.$$

By (2.3.12), choosing δ small enough, we have in addition that

$$\limsup_{n\to\infty} \mathbb{P}^*(\eta_n \in \complement H) < \varepsilon/2 \qquad (\complement H \equiv \mathbb{R}^T \backslash H).$$

Since $g, h \in U^b(T,d)$, we may also assume that $g, h \in H$ (again by choosing δ small enough).

Now, let $D = \{t_1, t_2, ...\}$ be as above and $m \in \mathbb{N}$ large enough s.t. $T = \bigcup_{i \le m} V(t_i, \delta)$ with $V(t_i, \delta) := \{t \in T : d(t, t_i) < \delta\}$ (such an m exists by (2.3.11) and since D is d-dense in T).

Then, for any $x \in H$ and $t \in T$ we have the following implications (choosing $i \in \{1, ..., m\}$ s.t. $d(t, t_i) < \delta$):

$$x(t_i) < h(t_i) - \lambda \implies x(t) \underset{(x \in H)}{\le} x(t_i) + \lambda/2 < h(t_i) - \lambda/2 \underset{(h \in H)}{\le} h(t), \quad \text{and}$$
$$x(t_i) > g(t_i) + \lambda \implies x(t) \underset{(x \in H)}{\ge} x(t_i) - \lambda/2 > g(t_i) + \lambda/2 \underset{(g \in H)}{\ge} g(t).$$

Thus,

$$\{\eta_n \in H \text{ and } g(t_i) + \lambda < \eta_n(t_i) < h(t_i) - \lambda \ \forall 1 \le i \le m\} \subset \{\eta_n \in C\},$$

and therefore

$$\begin{aligned}
\mathbb{P}_*(\eta_n \in C) &\ge \mathbb{P}_*\Big(\{\eta_n \in H\} \cap \{g(t_i) + \lambda < \eta_n(t_i) < h(t_i) - \lambda \ \forall 1 \le i \le m\}\Big) \\
&= 1 - \mathbb{P}^*\Big(\{\eta_n \in \complement H\} \cup \complement\{g(t_i) + \lambda < \eta_n(t_i) < h(t_i) - \lambda \ \forall 1 \le i \le m\}\Big) \\
&\ge 1 - \mathbb{P}^*(\eta_n \in \complement H) - \mathbb{P}^*\Big(\complement\{g(t_i) + \lambda < \eta_n(t_i) < h(t_i) - \lambda \ \forall 1 \le i \le m\}\Big) \\
&= \underbrace{\mathbb{P}_*\Big(g(t_i) + \lambda < \eta_n(t_i) < h(t_i) - \lambda \ \forall 1 \le i \le m\Big)}_{= \mathbb{P}((\eta_n(t_1), ..., \eta_n(t_m)) \in G), \text{ where}} - \mathbb{P}^*(\eta_n \in \complement H)
\end{aligned}$$

$G := \{(r_1, ..., r_m) \in \mathbb{R}^m : g(t_i) + \lambda < r_i < h(t_i) - \lambda \ \forall 1 \le i \le m\}$ is an open subset of $\mathbb{R}^m$.

Thus, (2.3.10) and the classical Portmanteau-Theorem together with STEP 1 (according to which $\eta_{0,D} \underset{\text{fidi}}{\overset{\mathcal{L}}{=}} \overline{\eta}_{0,D}$) imply

$$\begin{aligned}
\liminf_{n \to \infty} \mathbb{P}_*(\eta_n \in C) &\ge \liminf_{n \to \infty} \mathbb{P}((\eta_n(t_1), ..., \eta_n(t_m)) \in G) - \limsup_{n \to \infty} \mathbb{P}^*(\eta_n \in \complement H) \\
&> \mathbb{P}((\eta_0(t_1), ..., \eta_0(t_m)) \in G) - \varepsilon/2 \\
&= \mathbb{P}\Big(g(t_i) + \lambda < \eta_0(t_i) < h(t_i) - \lambda \ \forall 1 \le i \le m\Big) - \varepsilon/2 \\
&\ge \mathbb{P}\big(\eta_0 \in C_{5\lambda/4}\big) - \varepsilon/2 \underset{(e)}{\ge} \mathbb{P}(\eta_0 \in C) - \varepsilon.
\end{aligned}$$

Since $\varepsilon > 0$ was choosen arbitrary, (d) is shown.

As to (c), the proof runs quite similarly: For $\lambda > 0$ let

$$C^\lambda := \{y \in S : g(y) - \lambda \le y(t) \le h(t) + \lambda \ \forall t \in T\}.$$

Since $C^\lambda \downarrow C$ as $\lambda \downarrow 0$, we can choose $\forall \varepsilon > 0$ a $\lambda = \lambda(\varepsilon) > 0$ s.t.

$$\mathbb{P}(\eta_0 \in C^\lambda) \le \mathbb{P}(\eta_0 \in C) + \varepsilon/2.$$

Let H, δ and m be as before. Then, analogously, for $x \in H, t \in T$ and $i \in \{1, ..., m\}$ with $d(t, t_i) < \delta$:

$$x(t_i) \le h(t_i) \implies x(t) \le x(t_i) + \lambda/2 \le h(t_i) + \lambda/2 \le h(t) + \lambda, \quad \text{and}$$
$$x(t_i) \ge g(t_i) \implies x(t) \ge x(t_i) - \lambda/2 \ge g(t_i) - \lambda/2 \ge g(t) - \lambda.$$

Now, $\mathbb{P}^*(\eta_n \in C) \le \mathbb{P}^*(g(t_i) \le \eta_n(t_i) \le h(t_i)\ \forall 1 \le i \le m\} = \mathbb{P}((\eta_n(t_1), ..., \eta_n(t_m)) \in F)$, where $F := \{(r_1, ..., r_m) \in \mathbb{R}^m : g(t_i) \le r_i \le h(t_i) \ \ \forall 1 \le i \le m\}$ is a closed subset of $\mathbb{R}^m$. Thus, as before (2.3.10) and the classical Portmanteau-Theorem together with STEP 1 imply

$$\begin{aligned}\limsup_{n\to\infty} \mathbb{P}^*(\eta_n \in C) \le \mathbb{P}((\eta_0(t_1), ..., \eta_0(t_m)) \in F) \ &= \ \mathbb{P}\Big(g(t_i) \le \eta_0(t_i) \le h(t_i) \ \ \forall 1 \le i \le m\Big) \\ &\le \mathbb{P}(\eta_0 \in \complement H) + \mathbb{P}\Big(g(t) - \lambda \le \eta_0(t) \le h(t) + \lambda \ \ \forall t \in T\Big) \\ &\le \varepsilon/2 + \mathbb{P}(\eta_0 \in C^\lambda) \ \ \le \ \ \mathbb{P}(\eta_0 \in C) + \varepsilon.\end{aligned}$$

Since $\varepsilon > 0$ was choosen arbitrary, also (c) is shown. Thus, (2.3.13) is proved.

It is now easy to verify $\eta_0 \underset{\text{fidi}}{\overset{\mathcal{L}}{=}} \overline{\eta}_0$ using the continuous mapping theorem 2.3.16 below together with (2.3.10).

The proof of the converse part of 2.3.9 is as follows:

STEP 3: (2.3.13) (with $U^b(T,d)$ as a separable subspace of $l^\infty(T)$) implies (2.3.11) (according to [Gae90], Corollary 2). So it remains to show that (2.3.13) implies the AEC (2.3.12):

For this, let $\varepsilon > 0$ and $H(\delta) := \{x \in \mathbb{R}^T : w_x(\delta) \ge \varepsilon\}, \delta > 0$; then $H(\delta) \cap S$ is a closed subset of S, and therefore, by Theorem 2.3.14 (ii) below

$$\lim_{\delta\to 0} \limsup_{n\to\infty} \mathbb{P}^*(w_{\eta_n}(\delta) > \varepsilon) \le$$
$$\lim_{\delta\to 0} \limsup_{n\to\infty} \mathbb{P}^*(\eta_n \in H(\delta)) \ \le \ \lim_{\delta\to 0} \mathbb{P}(\eta_0 \in H(\delta)) =$$
$$\lim_{\delta\to 0} \mathbb{P}(\sup_{t,t' \in T, d(t,t') \le \delta} |\eta_0(t) - \eta_0(t')| \ge \varepsilon) \underset{(\sigma-\text{continuity of } \mathbb{P})}{=}$$
$$\mathbb{P}(\forall \delta > 0 \ : \sup_{t,t' \in T, d(t,t') \le \delta} |\eta_0(t) - \eta_0(t')| \ge \varepsilon) \le \mathbb{P}(\eta_0 \notin U^b(T,d)) = 0. \qquad \square$$

REMARK. The just given proof together with 2.3.14 below also shows:
If $\eta_n = (\eta_n(t))_{t\in T}$ with $T = (T,d)$ being totally bounded, $n \in \mathbb{N}$, is a sequence of RANDOM QUANTITIES $\eta_n : \Omega \longrightarrow \mathbb{R}^T$ (i.e. with $\eta_n(t), t \in T$, not being necessarily rv's on $(\Omega, \mathcal{A}, \mathbb{P})$, if η_0 is a re in $l^\infty(T)$ with sample paths in $U^b(T,d)$ s.t. $(\eta_n(t_1), ..., \eta_n(t_m)) \xrightarrow{\mathcal{L}} (\eta_0(t_1), ..., \eta_0(t_m))$ (in the sense of (2.3.6)) $\forall t_1, ..., t_m \in D, m \in \mathbb{N}$ (i.e. if the fidi-convergence on D holds true), then (2.3.12) implies (2.3.13) (in the sense of (2.3.6) with $S = l^\infty(T), E = \mathbb{R}^T, S_0 = U^b(T,d)$).

The following theorem is part of the Portmanteau-Theorem needed for our purposes. For a more comprehensive list of equivalent conditions for $\mathcal{L}$-convergence in our general model (2.3.5) we refer to [St94], Thm. 1.5; cf. also [Va96], Thm. 1.3.4.

2.3.14. THEOREM .
Given the general model (2.3.5), the following assertions are equivalent:

(i) $\eta_n \xrightarrow{\mathcal{L}} \eta_0$ *(in the sense of (2.3.6))*

(ii) $\limsup_{n\to\infty} \mathbb{P}^*(\eta_n \in H) \le \mathbb{P}(\eta_0 \in H) \quad \forall H \subset E, H \cap S$ *closed in* S

(ii')

$\limsup_{n\to\infty} \mathbb{P}^*(\eta_n \in F) \le \mathbb{P}(\eta_0 \in F) \quad \forall F$ *closed* $\subset S$, *and*
$\limsup_{n\to\infty} \mathbb{P}^*(\eta_n \in E\backslash S) = 0$

(iii) $\liminf_{n\to\infty} \mathbb{P}_*(\eta_n \in H) \ge \mathbb{P}(\eta_0 \in H) \quad \forall H \subset E, H \cap S$ *open in* S

(iii′) $\liminf_{n\to\infty} \mathbb{P}_*(\eta_n \in G) \ge \mathbb{P}(\eta_0 \in G) \quad \forall G$ *open* $\subset S$.

The proof of 2.3.14 is tailored along arguments used to prove the classical Portmanteau-Theorem as in [Bi68].

2.3.15. THEOREM *(Cramér-Slutzky-type result) ([St94], Thm. 1.16).*
Given a basic p-space $(\Omega, \mathcal{A}, \mathbb{P})$*, a metric space* $S = (S, s)$ *and an arbitrary* $E \supset S$*, let* $\eta_n, \zeta_n : \Omega \longrightarrow E, n \ge 1$*, be rq's and* $\eta_0 : \Omega \longrightarrow S$ *be a re in* $(S, \mathcal{B}(S))$ *such that*

$$(+) \qquad \lim_{n\to\infty} \mathbb{P}_*(\eta_n, \zeta_n \in S \ \textit{ and } s(\eta_n, \zeta_n) \le \varepsilon) = 1 \qquad \forall \varepsilon > 0$$

where $s(\eta_n, \zeta_n)(\omega) := s(\eta_n(\omega), \zeta_n(\omega)), \omega \in \Omega$*. Then*

$$\eta_n \xrightarrow{\mathcal{L}} \eta_0 \quad \Longleftrightarrow \quad \zeta_n \xrightarrow{\mathcal{L}} \eta_0.$$

PROOF. By symmetry, it suffices to show "$\Longrightarrow$": We are going to use the criterion (ii) from 2.3.14. For this, given any $H \subset E$ and $\varepsilon > 0$ s.t. $H \cap S$ is closed in S, the set

$$F := (H \cap S)^\varepsilon := \{x \in S : \inf_{y \in H \cap S} s(x, y) \le \varepsilon\}$$

is also closed in S, whence

$$\begin{aligned}
\limsup_{n\to\infty} \mathbb{P}^*(\zeta_n \in H) \ &\le \ \limsup_{n\to\infty} \Big[\mathbb{P}^*(\zeta_n \in H, \eta_n, \zeta_n \in S \text{ and } s(\eta_n, \zeta_n) \le \varepsilon) + \\
&\quad \mathbb{P}^*(\complement\{\eta_n, \zeta_n \in S \text{ and } s(\eta_n, \zeta_n) \le \varepsilon\})\Big] \ \le \\
&\limsup_{n\to\infty} \mathbb{P}^*(\eta_n \in F) \ + \ 1 \ - \liminf_{n\to\infty} \mathbb{P}_*(\eta_n, \zeta_n \in S \text{ and } s(\eta_n, \zeta_n) \le \varepsilon) \underset{(+)}{=} \\
&\limsup_{n\to\infty} \mathbb{P}^*(\eta_n \in F) \ \underset{2.3.14(ii)}{\le} \ \mathbb{P}(\eta_0 \in F).
\end{aligned}$$

Since $H \cap S$ is closed in S, we have $(H \cap S)^\varepsilon \downarrow H \cap S$ as $\varepsilon \downarrow 0$, whence for $\varepsilon \downarrow 0$

$$\mathbb{P}(\eta_0 \in F) \downarrow \mathbb{P}(\eta_0 \in H \cap S) = \mathbb{P}(\eta_0 \in H),$$

and therefore

$$\limsup_{n\to\infty} \mathbb{P}^*(\zeta_n \in H) \ \le \mathbb{P}(\eta_0 \in H)$$

from which $\zeta_n \xrightarrow{\mathcal{L}} \eta_0$ follows according to 2.3.14(ii). □

2.3.16. THEOREM *(Continuous Mapping Theorem (CMT)) ([St94], Thm. 1.8; cf. also [Va96], Thm. 1.3.6)).*
In addition to our general model (2.3.5), let $S' = (S', s')$ *be a further metric space and* E' *be arbitrary,* $E' \supset S'$*. Let* $g : E \longrightarrow E'$ *be a given map with* $g(S) \subset S'$ *and let* $S_g \in \mathcal{B}(S)$ *be such that* $rest_S g$ *is continuous at every point in* S_g*; then, assuming in addition that* η_0 *takes its values in* S_g*,* $\eta_n \xrightarrow{\mathcal{L}} \eta_0$ *implies* $g \circ \eta_n \xrightarrow{\mathcal{L}} g \circ \eta_0$*.*

In connection with RANDOM QUANTITIES (rq's) $\zeta : \Omega \longrightarrow \overline{\mathbb{R}} := \mathbb{R} \cup \{\pm\infty\}$ (being not necessarily rv's) the following concept and its consequences turn out to be useful at some places later on; cf. [Va96], Lemma 1.2.1 and Lemma 1.2.2; cf. also the Notes on p. 75 in [Va96] refering to early papers by Blumberg in 1935 and by Eames and May in 1967.

(2.3.17) Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a p-space and $\zeta : \Omega \longrightarrow \overline{\mathbb{R}}$ be a rq. Then there exists a measurable function $\zeta^* : \Omega \longrightarrow \overline{\mathbb{R}}$ with

(i) $\zeta^* \geq \zeta$, and

(ii) $\zeta^* \leq \eta \; \mathbb{P} - a.s \; \forall$ measurable $\eta : \Omega \longrightarrow \overline{\mathbb{R}}$ with $\eta \geq \zeta \;\; \mathbb{P} - a.s.$

ζ^* is $\mathbb{P} - a.s.$ uniquely determined and for any ζ^* fulfilling (i) and (ii), it holds that $\mathbb{E}^*(\zeta) = \mathbb{E}(\zeta^*)$, provided $\mathbb{E}(\zeta^*)$ exists; the latter is certainly true if $\mathbb{E}^*(\zeta) < \infty$; furthermore one has $\mathbb{P}^*(\zeta > t) = \mathbb{P}(\zeta^* > t) \quad \forall t \in \mathbb{R}$.

The function ζ^* is called minimal measurable majorant of ζ, or also called MEASURABLE COVER or ENVELOPE FUNCTION.

Before concluding this section, let us have once more a glance onto the Characterization Theorem 2.3.9:

As already remarked there, the condition (2.3.10) will be fulfilled in most cases due to classical multivariate CLT's. Also, (2.3.11) will be fulfilled by choosing the pseudo-metric d appropriately; e.g. in the case $T = \mathcal{C}, \mathcal{C} \subset \mathcal{X}$ a VCC, let $d := d_\nu, \nu$ being an arbitrary p-measure on $\mathcal{X}$; then the condition (2.3.11) holds as we shall see in Section 4.2.

So, in order to prove (2.3.13), the crucial task is to verify the AEC (2.3.12):

Since Markov's inequality also holds in the case of outer probabilities and outer expectations, respectively, for verifying the AEC it suffices to show

$$\lim_{\delta \to 0} \limsup_{n \to \infty} \mathbb{E}^*\Big(\sup_{t,t' \in T, d(t,t') \leq \delta} |\eta_n(t) - \eta_n(t')| \Big) = 0, \tag{2.3.18}$$

i.e. later on we will have at our disposal the following fact:

2.3.19. REMARK.

The conditions (2.3.10), (2.3.11) and (2.3.18) imply (2.3.13), i.e. $\eta_n \xrightarrow[sep]{\mathcal{L}} \eta_0$, *where* η_0 *has all its sample paths in* $U^b(T, d)$ *and where* $\eta_0 \underset{fidi}{\overset{\mathcal{L}}{=}} \overline{\eta}_0$.

2.3.19 (*with (2.3.18)* instead of the AEC (2.3.12)) leads to essential simplifications in proving FCLT's in Section 7 comparable with the role of (2.1.5) in proving ULLN's.

3 Random Measures Processes (RMP's)

3.1 Empirical processes, partial-sum processes and smoothed empirical processes, respectively, considered as special cases of RMP's

In order to cope in later sections also with processes indexed by classes of functions instead of sets (cf. Section 4.3 for some motivation) the general context will be now as follows:

Let $X = (X, \mathcal{X})$ be an arbitrary measurable space (sample space) and $\mathcal{F}$ be a class of $\mathcal{X}$-measurable functions $f : X \longrightarrow \mathbb{R}$ with $\mathcal{X}$-measurable envelope $F : X \longrightarrow \mathbb{R}_+$ (i.e. $\sup_{f\in\mathcal{F}} |f(x)| \leq F(x) \;\; \forall x \in X$). Let $(w_{nj})_{1\leq j\leq j(n), n\in\mathbb{N}}$ be a triangular array of random p-measures on $\mathcal{X}$ and $(\xi_{nj})_{1\leq j\leq j(n), n\in\mathbb{N}}$ be a triangular array of real-valued rv's.

Random Measure Processes (RMP's) $S_n = (S_n(f))_{f\in\mathcal{F}}$ (of sample size n) (indexed by $\mathcal{F}$) are defined by

$$S_n(f) := \sum_{j\leq j(n)} w_{nj}(f) \cdot \xi_{nj}, \quad f \in \mathcal{F}, \tag{3.1.1}$$

where $w_{nj}(f) := \int_X f dw_{nj}$.

We tacitly assume regularity conditions such as measurability and finiteness of $w_{nj}(F)$ (which implies that the sample paths of S_n are contained in the Banach space

$$l^\infty(\mathcal{F}) := \{x : \mathcal{F} \longrightarrow \mathbb{R} : ||x||_\mathcal{F} := \sup_{f\in\mathcal{F}} |x(f)| < \infty\},$$

endowed with the sup-norm $|| \cdot ||_\mathcal{F}$.

In connection with Uniform Laws of Large Numbers (ULLN) and Functional Central Limit Theorems (FCLT) in Section 6 and 7, respectively, it will be assumed that $j(n) \longrightarrow \infty$ as $n \to \infty$ and that

for all $n \in \mathbb{N}$ the sequence of pairs
$(w_{n1}, \xi_{n1}), ..., (w_{nj(n)}, \xi_{nj(n)})$ is independent.

(Here independence is to be understood in the sense of independence of the rq's $(w_{nj}(f) \cdot \xi_{nj})_{f\in\mathcal{F}}, 1 \leq j \leq j(n)$, for each n, which means that $(w_{nj}(f) \cdot \xi_{nj})_{f\in\mathcal{F}}, 1 \leq j \leq j(n)$, $n \in \mathbb{N}$, are considered as coordinate projections on an appropriately chosen product-p-space $(\Omega, \mathcal{A}, \mathbb{P})$ (cf. Section 5.1 for the definition of independence of rq's and also Section 6.1 below).

However, we do not assume that the above pairs are identically distributed; also dependence within each pair is allowed.

Processes of the form (3.1.1) with $\mathcal{F} = \{1_C : C \in \mathcal{C}\}, \mathcal{C} \subset \mathcal{X}$, and *non-random* p-measures w_{nj}, were first considered in [Al87] and in its present general form in [Zi94] (see also [Va96], Section 2.12.2 for closely related examples).

Now, special cases of RMP's occur when considering

- the classical multivariate partial-sum process of Section 1.3 and 1.4 where $X = I^d, d \geq 1, w_{nj} = \delta_{\underline{j}/n}, \xi_{nj} := n^{-d/2}\xi_{\underline{j}}, \underline{j} \in \{1, ..., n\}^d$, and $\mathcal{F} = \{1_C : C \in \mathcal{C}\}$ with $\mathcal{C} = \{[0, \underline{t}] : \underline{t} \in I^d\}$;

- empirical processes indexed by $\mathcal{F}$ given by $\nu_n(f) := j(n)^{-1} \sum_{j \leq j(n)} f(\eta_{nj})$, $f \in \mathcal{F}$, based on re's η_{nj} in $(X, \mathcal{X})$, where $w_{nj} = \delta_{\eta_{nj}}$ and $\xi_{nj} \equiv j(n)^{-1}$;

- smoothed empirical processes:
  Here X is assumed to be a linear metric space endowed with its Borel σ-field $\mathcal{X}$. Let $\eta_j, j \in \mathbb{N}$, be i i d re's in $(X, \mathcal{X})$ with law ν on $\mathcal{X}$ and ν_n be the empirical measure based on $\eta_1, ..., \eta_n$, i.e. $\nu_n = \sum_{j \leq n} \delta_{\eta_j}$. Now, if ν is "smooth" (one may think of random vectors η_j in $X = \mathbb{R}^d$ whose df has a continuous or even differentiable density w.r.t. Lebesgue measure), it is natural to replace ν_n by a smoothed version $\tilde{\nu_n}$ serving as an estimator for an unknown ν.
  As in [Yu89] we will consider "smoothing through convolution" as explained in Section 6.4 below. As we will see there, this leads to $\tilde{\nu_n} = (\tilde{\nu_n}(f))_{f \in \mathcal{F}}$ which can be represented as RMP's with $w_{nj}(B) = \mu_n(B - \eta_j), B \in \mathcal{X}$, and $\xi_{nj} \equiv n^{-1}$, where $(\mu_n)_{n \in \mathbb{N}}$ is a given sequence of p-measures μ_n on $\mathcal{X}$ with $\mu_n \longrightarrow \delta_0$ weakly (in the sense of weak convergence of Borel measures as in [Bi68]).

3.2 Further examples

3.2.1 Partial-sum processes with random locations

Let $X = (X, \mathcal{X})$ be an arbitrary measurable space, $\mathcal{C} \subset \mathcal{X}$, $(\eta_{nj})_{1 \leq j \leq j(n), n \in \mathbb{N}}$ be a triangular array of re's in $(X, \mathcal{X})$ and $(\xi_{nj})_{1 \leq j \leq j(n), n \in \mathbb{N}}$ be a triangular array of real-valued rv's.

PARTIAL-SUM PROCESSES $S_n = (S_n(C))_{C \in \mathcal{C}}$ WITH RANDOM LOCATIONS are defined by

(3.2.2) $$S_n(C) := \sum_{j \leq j(n)} 1_C(\eta_{nj}) \cdot \xi_{nj}, \qquad C \in \mathcal{C}.$$

These processes were studied in [Ar92], [Gae94] and [Gae94b], being special RMP's with $\mathcal{F} = \{1_C : C \in \mathcal{C}\}$ and $w_{nj} = \delta_{\eta_{nj}}$.

Many examples of natural phenomena like mineral deposits, earthquakes, forest disease, etc. can be modelled by such processes.

3.2.3 The sequential uniform empirical process

(Cf. [Sh86], Chapter 3.5).
Let $\eta_j, j \in \mathbb{N}$, be i i d rv's with $\mathcal{L}\{\eta_j\} = U[0,1]$ (as in 1.1). The SEQUENTIAL UNIFORM EMPIRICAL PROCESS (of sample size n) $K_n = (K_n(s,t))_{(s,t) \in I^2}$ based on $\eta_1, ..., \eta_n$ is defined by

$$K_n(s,t) := n^{-1/2} \sum_{j \leq <ns>} \Big(1_{[0,t]}(\eta_j) - t\Big), \qquad (s,t) \in I^2.$$

Choosing $X := I^2, \mathcal{X} := I^2 \cap \mathcal{B}^2, \mathcal{C} := \{[0,s] \times [0,t] : (s,t) \in I^2\}, \eta_{nj} := (j/n, \eta_j), 1 \leq j \leq j(n) := n,$

and $S_n(C) := n^{-1/2} \sum_{j \le n} \delta_{\eta_{nj}}(C), C \in \mathcal{C}$, we get for $C = [0, s] \times [0, t]$

$$S_n(C) = n^{-1/2} \sum_{j \le n} 1_C((j/n, \eta_j)) = n^{-1/2} \sum_{j \le <ns>} 1_{[0,t]}(\eta_j)$$

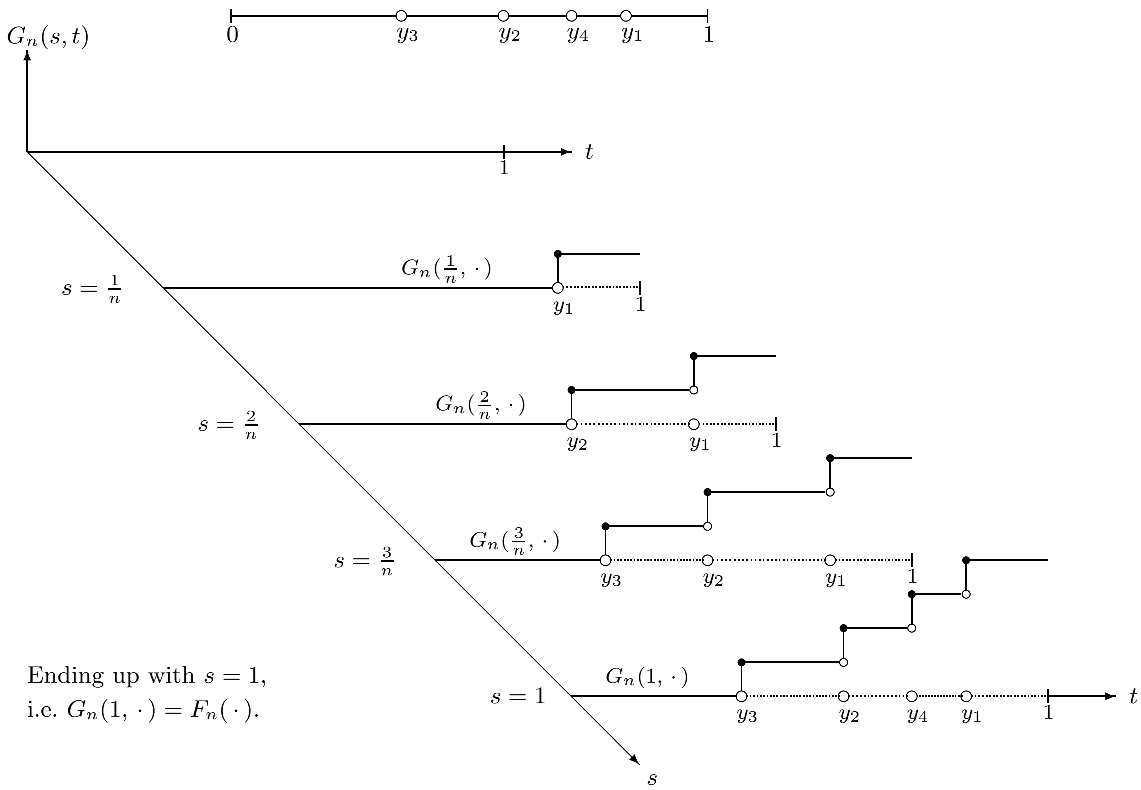
and $\mathbb{E}(S_n(C)) = n^{-1/2} \sum_{j \le n} \mathbb{P}(\eta_{nj} \in C) = n^{-1/2} \sum_{j \le n} \mathbb{P}(\frac{j}{n} \le s, \eta_j \in [0,t]) = n^{-1/2} \sum_{j \le <ns>} \mathbb{P}(\eta_j \in [0,t]) = n^{-1/2} <ns> \cdot t$, whence $S_n(C) - \mathbb{E}(S_n(C)) = K_n(s,t)$, i.e. K_n can be represented as a centered RMP (with $w_{nj} = \delta_{\eta_{nj}}$ and $\xi_{nj} = n^{-1/2}$) indexed by the VCC $\mathcal{C}$.

If one considers instead of K_n the underlying df

$$G_n(s,t) := n^{-1} \sum_{j \le <ns>} 1_{[0,t]}(\eta_j), \qquad (s,t) \in I^2,$$

(in comparison with the edf $F_n(t) = n^{-1} \sum_{j \le n} 1_{[0,t]}(\eta_j),\ t \in I$), then, through the additional parameter s it is possible to visualize the appearance of the data $y_1, ..., y_n$ ($y_j = \eta_j(\omega)$) successively, (therefore the notion "sequential" uniform empirical process) as the following picture may illustrate.

$n = 4:$



As to the function-indexed sequential empirical process (based on an i i d sequence of re's in an arbitrary sample space $(X, \mathcal{X})$ we refer to [Va96], Section 2.12.1 and to [Zi97], Section 7.4; as in the uniform case, also this process can be represented as a centered RMP.

3.2.4 Nonparametric regression

(Cf. [Stu97]). Let η be a re in an arbitrary measurable space $(X, \mathcal{X})$ with law $\mathcal{L}\{\eta\} = \nu$ and let $\mathcal{C} \subset \mathcal{X}$. Let ξ be a rv (defined on the same p-space $(\Omega, \mathcal{A}, \mathbb{P})$ as η) such that $\mathbb{E}(|\xi|) < \infty$. Consider the *regression function*

$$m(y) := \mathbb{E}(\xi|\eta = y), \qquad y \in X,$$

and the corresponding *integrated regression function* indexed by $\mathcal{C}$:

$$\mathbb{I}(C) := \int_C m(y)\,\nu(dy), \qquad C \in \mathcal{C}.$$

Since m is usually ν-a.s. uniquely determined by $\mathbb{I}$, statistical inference may be based on $\mathbb{I}$ instead of m as well.

Now, $\mathbb{I}(C) = \mathbb{E}(1_C(\eta)\mathbb{E}(\xi|\eta)) = \mathbb{E}(\mathbb{E}(1_C(\eta)\xi|\eta)) = \mathbb{E}(1_C(\eta) \cdot \xi)$, whose empirical version (of sample size n) based on i i d pairs (η_j, ξ_j) of re's in $(X \times \mathbb{R}, \mathcal{X} \bigotimes \mathcal{B})$ (where $\mathcal{L}\{\eta_j\} = \nu$ and $\mathcal{L}\{\xi_j\} = \mathcal{L}\{\xi\}$) is given by

$$\mathbb{I}_n(C) := n^{-1} \sum_{j \le n} 1_C(\eta_j)\xi_j, \quad C \in \mathcal{C},$$

where $\mathbb{E}(\mathbb{I}_n(C)) = \mathbb{I}(C)$ for all $C \in \mathcal{C}$. Thus $\mathbb{I}_n$ is a RMP indexed by $\mathcal{C}$ (with $w_{nj} \equiv \delta_{\eta_j}$ and $\xi_{nj} = n^{-1}\xi_j$).

At this place we may also mention another paper by Stute et al. ([Stu98]) where (in our notation) processes R_n of the following form are considered:

$$R_n(C) = n^{-1/2} \sum_{j \le n} 1_C(\eta_j)(\xi_j - m(\eta_j))$$

based on i i d re's (η_j, ξ_j) in $(\mathbb{R}^d \times \mathbb{R}, \mathcal{B}^d \bigotimes \mathcal{B})$ with $C \in \mathcal{C} := \{(-\infty, \underline{y}] : \underline{y} \in \mathbb{R}^d\}$, whence R_n is also a RMP indexed by the VCC $\mathcal{C}$.

3.2.5 Estimation of intensity measures for spatial Poisson processes

This example is taken from [Zi97], Section 7.8; cf. also [Li90].

Let Φ be a Poisson point process on an arbitrary state space $(X, \mathcal{X})$ with finite intensity measure Λ on $\mathcal{X}$, i.e. (based on an underlying p-space $(\Omega, \mathcal{A}, \mathbb{P})$) $\Phi(\omega, \cdot)$ is a measure on $\mathcal{X}$ with values in $\{0, 1, 2, ...\}$ for every fixed $\omega \in \Omega$, $\Phi(\cdot, B)$ is a Poisson rv with parameter $\Lambda(B)$ for every fixed $B \in \mathcal{X}$, and for any disloint $B_1, ..., B_n \in \mathcal{X}, n \in \mathbb{N}$, the rv's $\Phi(\cdot, B_1), ..., \Phi(\cdot, B_n)$ are independent.

In estimating an **unknown** intensity measure Λ on the basis of an i i d sequence $(\Phi_j)_{j \in \mathbb{N}}$ of Poisson point processes Φ_j (with intensity measure Λ) a natural sequence of estimators $\hat{\Lambda}_n, n \in \mathbb{N}$, is $\hat{\Lambda}_n := n^{-1} \sum_{j \le n} \Phi_j$ leading to the corresponding standardized process

$$Z_n(f) := n^{1/2}(\hat{\Lambda}_n(f) - \Lambda(f)), \qquad f \in \mathcal{F},$$

in view of a FCLT for $Z_n = (Z_n(f))_{f\in\mathcal{F}}$, where $\mathcal{F}$ is an appropriate class of measurable functions $f : X \longrightarrow \mathbb{R}$ with $f \in \mathcal{L}_2(X, \mathcal{X}, \Lambda)$; note that $\mathbb{E}(\Phi(f)) = \Lambda(f)$ for all f.
Now, since $Z_n(f) = n^{-1/2}\sum_{j\le n}(\Phi_j(f) - \mathbb{E}(\Phi_j(f))) = S_n(f) - \mathbb{E}(S_n(f))$ with $S_n(f) := \sum_{j\le n} w_{nj}(f)\cdot \xi_{nj}$, where $w_{nj} :\equiv \Phi_j/\Phi_j(X)$ and $\xi_{nj} := n^{-1/2}\Phi_j(X)$, Z_n can be considered as a RMP indexed by $\mathcal{F}$ to which our result in Section 7.1 will apply.

3.2.6 Lévy's Multivariate Brownian motion as a set-indexed process and as limiting process of a sequence of Partial-sum processes with random locations

We will follow here the exposition presented in the paper by Mina Ossiander and Ronald Pyke [Os85]:

Let $B = (B(t))_{t\in\mathbb{R}_+}$ be a Brownian motion (indexed by the parameter space $T = \mathbb{R}_+$), i.e. a mean-zero Gaussian process with independent and stationary increments whose covariance function is given by

$$cov(B(s), B(t)) = s \wedge t, \quad s, t \in \mathbb{R}_+ \tag{1}$$

or equivalently, since $s \wedge t = (|t| + |s| - |t - s|)/2$, by

$$cov(B(s), B(t)) = (|t| + |s| - |t - s|)/2, \quad s, t \in \mathbb{R}_+. \tag{2}$$

Now, in view of (2) Lévy's multivariate Brownian motion ([Lé40], [Lé45]) is defined to be a mean-zero Gaussian process (*random field*) $Z = (Z(\underline{t}))_{\underline{t}\in\mathbb{R}^d}$ with

$$cov(Z(\underline{s}), Z(\underline{t})) = (|\underline{t}| + |\underline{s}| - |\underline{t} - \underline{s}|)/2, \quad \underline{s}, \underline{t} \in \mathbb{R}^d, \tag{3}$$

where $|\cdot|$ is the Euclidian norm in $\mathbb{R}^d, d \ge 1$.

The covariance structure (3) can also be characterized by the isotropic mean square condition

$$\mathbb{E}((Z(\underline{s}) - Z(\underline{t}))^2) = |\underline{t} - \underline{s}|, \qquad \underline{s}, \underline{t} \in \mathbb{R}^d. \tag{4}$$

Notice that $Z(\underline{s})$ and $Z(\underline{t}) - Z(\underline{s})$ are independent if and only if $0, \underline{s}$ and $\underline{t}$ are colinear, so that the independent increments property of one-dimensional Brownian motion has apparently not been fully generalized.

A second generalization of Brownian motion to multi-dimensional time was given by Chentsov ([Che56]); cf. also [Py73]:

Let $W = (W(\underline{t}))_{\underline{t}\in I^d}$ be a mean-zero Gaussian process with covariance structure

$$cov(W(\underline{s}), W(\underline{t})) = \underline{s} \wedge \underline{t}, \quad \underline{s}, \underline{t} \in I^d, \tag{5}$$

where $\underline{s} \wedge \underline{t} := \prod_{i\le d}(s_i \wedge t_i)$ for $\underline{s} = (s_1, ..., s_d)$ and $\underline{t} = (t_1, ..., t_d)$; then, if $s_i \le t_i \;\; \forall 1 \le i \le d$, we have $cov(W(\underline{s}), W(\underline{t}) - W(\underline{s})) = \underline{s} \wedge \underline{t} - \underline{s} \wedge \underline{s} = 0$, i.e. under the natural partial ordering of I^d, the property of independent increments has been retained.

The process W is called *Brownian sheet* (cf. [Py73]). A variant of W is the so-called *tied-down Brownian sheet* $U = (U(\underline{t}))_{\underline{t}\in I^d}$, defined by

$$U(\underline{t}) := W(\underline{t}) - W(\underline{1}) \cdot \prod_{i\leq d} t_i, \quad \underline{t} = (t_1, ..., t_d) \text{ and } \underline{1} = (1, ..., 1)$$

(i.e. for $d = 1 \quad U$ coincides with the Brownian bridge).
Identifying (as in 1.4) each $\underline{t} \in I^d$ with $[0, \underline{t}]^d$, and denoting with ν the Lebesgue measure on I^d, then, with $W([0, \underline{t}]) := W(\underline{t}), \underline{t} \in I^d$, (5) is equivalent to

$$(5') \qquad cov(W([0, \underline{s}]), W([0, \underline{t}])) = \nu([0, \underline{s}] \cap [0, \underline{t}]), \quad \underline{s}, \underline{t} \in I^d.$$

But note however that the restriction of Z onto the parameter space I^d is not identical with W, since for $0, \underline{s}$ and $\underline{t}$ being colinear one has

$$(|\underline{t}| + |\underline{s}| - |\underline{t} - \underline{s}|)/2 = |\underline{s}| \neq \underline{s} \wedge \underline{t},$$

in general. Therefore the following question arises:
Is it possible to represent the Lévy-process Z as a set-indexed process $Z' = (Z'(C))_{C\in\mathcal{C}}$ with an appropriately chosen class $\mathcal{C} = \{C_{\underline{t}} : \underline{t} \in \mathbb{R}^d\}$ such that anagously to $(5')$ the covariance of Z' is given by

$$(6) \qquad cov(Z'(C_{\underline{s}}), Z'(C_{\underline{t}})) = \mu(C_{\underline{s}} \cap C_{\underline{t}})$$

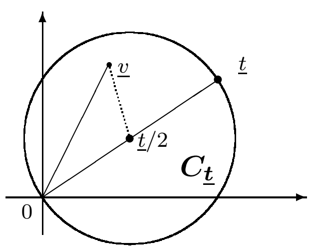
with a suitable p-measure μ?

An answer to this question is given in [Os85]:
Let us restrict ourselves to the unit sphere $S^d := \{\underline{t} \in \mathbb{R}^d : |\underline{t}| \leq 1\}$, i.e. consider Lévy's multivariate Brownian motion $Z = (Z(\underline{t}))_{\underline{t}\in S^d}$ as it is done in [Os85], and, for $\underline{t} \in S^d$ let

$$C_{\underline{t}} := \{\underline{v} \in \mathbb{R}^d : |\underline{v} - \underline{t}/2| \leq |\underline{t}/2|\},$$

so that $C_{\underline{t}}$ is the closed sphere in $\mathbb{R}^d$ having for a diameter the ray from 0 to $\underline{t}$.

$n = 2:$



The family $\{C_{\underline{t}} : \underline{t} \in S^d\}$ plays then (as seen below) in the representaion of the Lévy Brownian motion Z as a set-indexed process the same role as the class of all lower left orthants $[0, \underline{t}] \cap I^d$ do for the W and U processes with parameter set I^d.

Now, let μ be the p-measure on $S^d \cap \mathcal{B}^d$ with density function f_d (w.r.t. Lebesgue measure) given by

$$f_d(\underline{v}) := (c_d \cdot |\underline{v}|)^{-1}, \qquad \underline{v} \in S^d,$$

where $c_d := (2\pi)^{d/2}/\Gamma(d/2)$ is the surface area of S^d. Then we get

(7) $$\text{For any } \underline{s}, \underline{t} \in S^d \quad \mu(C_{\underline{s}} \cap C_{\underline{t}}) = c_d^{-1}\Big(|\underline{t}| + |\underline{s}| - |\underline{t} - \underline{s}|\Big)/2.$$

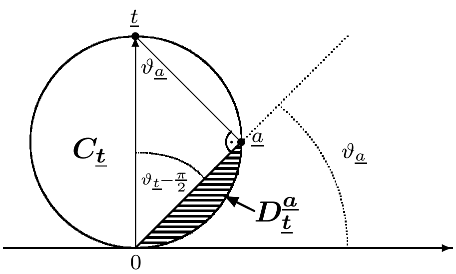
Comparing this with (3) we see: For $\underline{s}, \underline{t} \in S^d$ it follows that $cov(Z(\underline{s}), Z(\underline{t})) = (c_d/2) \cdot \mu(C_{\underline{s}} \cap C_{\underline{t}})$, i.e. with $Z'(C_{\underline{t}}) := (c_d/2)^{-1/2} Z(\underline{t})$ we get

$$cov(Z'(C_{\underline{s}}), Z'(C_{\underline{t}})) = \mu(C_{\underline{s}} \cap C_{\underline{t}}), \qquad \text{i.e. (6).}$$

(Note that scalar multiplication (by $(c_d/2)^{-1/2}$) does not change the process in any essential way.)

Note also that Z' has independent "increments" in the sense that $Z'(C_{\underline{s}})$ and $Z'(C_{\underline{t}})$ are independent if $C_{\underline{s}} \cap C_{\underline{t}} = \varnothing \quad \mu - a.s.$

PROOF OF (7) in the case $d = 2$ (with $c_2 = 2\pi$).
Let $\underline{t} = (t_1, t_2) \in S^2$ and $\underline{a} = (a_1, a_2) \in \partial C_{\underline{t}}$, assuming w.l.o.g. (due to the spherical symmetry of f_d) that $t_1 = 0$ and $a_1 > 0$:



Let $\vartheta_{\underline{a}}$ denote the angle between $\underline{a}$ and the x-axis (w.r.t. $\underline{t}$) and let $D_{\underline{t}}^{\underline{a}}$ be the hatched region in the figure. Representing $\underline{v} = (v_1, v_2)$ by its polar coordinates, we have

$$\underline{v} = (|\underline{v}|\cos\vartheta, |\underline{v}|\sin\vartheta) \in D_{\underline{t}}^{\underline{a}} \iff 0 \le |\underline{v}| \le |\underline{t}|\cos(\vartheta - \frac{\pi}{2}) \text{ and } 0 \le \vartheta \le \vartheta_{\underline{a}}.$$

Thus we get (note that $d\underline{v} = |\underline{v}| d|\underline{v}| d\vartheta$):

$$\begin{aligned}\mu(D_{\underline{t}}^{\underline{a}}) = c_2^{-1} \int_{\mathbb{R}^2} 1_{D_{\underline{t}}^{\underline{a}}}(\underline{v}) \frac{1}{|\underline{v}|}\, d\underline{v} \;&=\; c_2^{-1} \int_0^{\vartheta_{\underline{a}}} \int_0^{|\underline{t}|\cos(\vartheta - \frac{\pi}{2})} \frac{1}{|\underline{v}|} |\underline{v}|\, d|\underline{v}|\, d\vartheta \\ = c_2^{-1} \int_0^{\vartheta_{\underline{a}}} |\underline{t}| \cos(\vartheta - \frac{\pi}{2})\, d\vartheta \;&=\; c_2^{-1} |\underline{t}| \big(\sin(\vartheta_{\underline{a}} - \frac{\pi}{2}) + 1\big) \\ = c_2^{-1}\Big(|\underline{t}| - |\underline{t}|\sin(\frac{\pi}{2} - \vartheta_{\underline{a}})\Big) \;&=\; c_2^{-1}\Big(|\underline{t}| - |\underline{t} - \underline{a}|\Big), \text{ i.e.}\end{aligned}$$

(+) $$\mu(D_{\underline{t}}^{\underline{a}}) = c_2^{-1}\Big(|\underline{t}| - |\underline{t} - \underline{a}|\Big).$$

Considering now $C_{\underline{s}} \cap C_{\underline{t}}$ for $\underline{s} \in S^2$ and $\underline{a} \in \partial C_{\underline{s}} \cap \partial C_{\underline{t}}$, then (where $\underline{a}'$ (see the figure below) takes over the role of $\underline{a}$ before, now with $\underline{s}$ instead of $\underline{t}$)

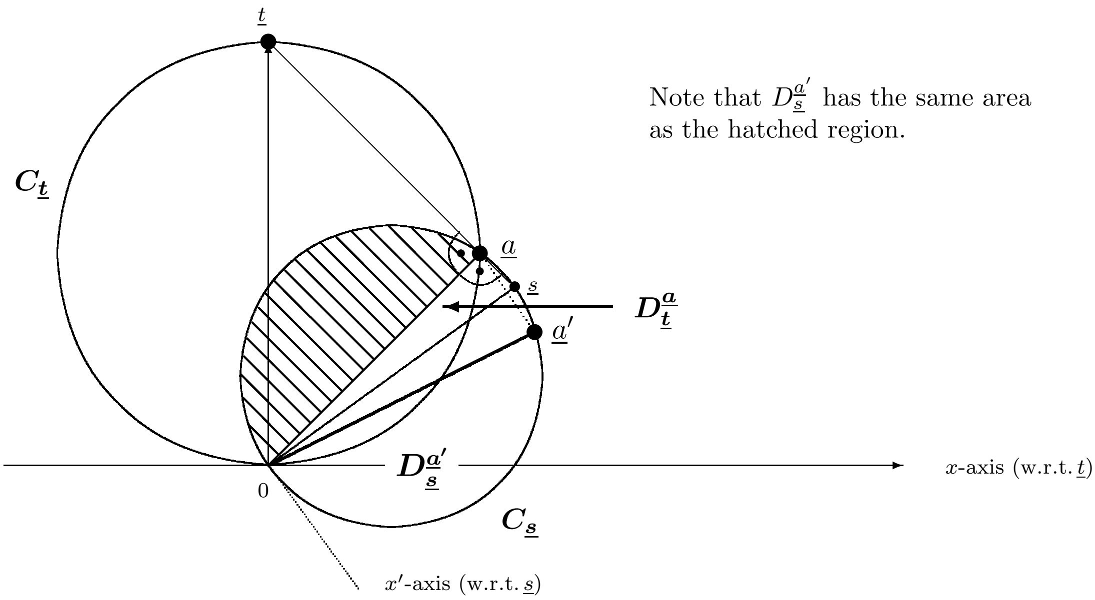
$$\mu(C_{\underline{s}} \cap C_{\underline{t}}) = \mu(D_{\underline{s}}^{\underline{a}'}) + \mu(D_{\underline{t}}^{\underline{a}}),$$

whence by (+) (noticing that $|\underline{s} - \underline{a}'| = |\underline{s} - \underline{a}|$)

$$\mu(C_{\underline{s}} \cap C_{\underline{t}}) = c_2^{-1}\Big(|\underline{s}| - |\underline{s} - \underline{a}'| + |\underline{t}| - |\underline{t} - \underline{a}|\Big) =$$

$$c_2^{-1}\Big[|\underline{t}| + |\underline{s}| - (|\underline{s} - \underline{a}| + |\underline{t} - \underline{a}|)\Big] = c_2^{-1}\big(|\underline{t}| + |\underline{s}| - |\underline{t} - \underline{s}|\big),$$

since $\underline{t} - \underline{a}$ and $\underline{s} - \underline{a}$ are colinear. $\square$



(8) It can be shown that (7) is equivalent to

$$(7') \qquad d_\mu(C_{\underline{s}}, C_{\underline{t}}) = 2c_d^{-1}|\underline{t} - \underline{s}| \qquad \forall \underline{s}, \underline{t} \in S^d$$

(where $d_\mu(A, B) := \mu(A \Delta B)$). Next (as already remarked in (1.5.4)), the class of all closed Euclidian balls form a VCC, whence

$$\mathcal{C} := \{C_{\underline{t}} : \underline{t} \in S^d\}$$

is also a VCC; thus, we can apply the FCLT 2.2.3 with $X = S^d, \mathcal{X} = S^d \cap \mathcal{B}^d$ and $S_n = (S_n(C_{\underline{t}}))_{C_{\underline{t}} \in \mathcal{C}}$, where

$$S_n(C_{\underline{t}}) := n^{-1/2} \sum_{j \leq n} 1_{C_{\underline{t}}}(\eta_j) \cdot \xi_j,$$

the $\xi_j, j \in \mathbb{N}$, being i i d with $\mathbb{E}(\xi_j) = 0$ and $\mathbb{E}(\xi_j^2) = 1$, the $\eta_j, j \in \mathbb{N}$, being i i d with $\mathcal{L}\{\eta_j\} = \mu$ (having Lebesgue density f_d), and where $(\eta_j)_{j \in \mathbb{N}}$ is independent of $(\xi_j)_{j \in \mathbb{N}}$, to obtain the following result (note that both conditions (i) and (ii) in Theorem 2.2.3 are obviously fulfilled here):

$$S_n \xrightarrow[\text{sep}]{\mathcal{L}} Z' \tag{9}$$

i.e. Z' proves to be the limiting process of a sequence of Partial-sum processes with random locations.

In addition, $Z' = (Z'(C_{\underline{t}}))_{C_{\underline{t}} \in \mathcal{C}}$ can be chosen as a process with bounded and uniformly d_μ-continuous sample paths; moreover, $(7')$ shows the existence of a version of the Lévy process $Z = (Z(\underline{t}))_{\underline{t} \in S^d}$ with continuous sample paths, cf. above:

$$Z(\underline{t}) = \left(\frac{c_d}{2}\right)^{1/2} Z'(C_{\underline{t}}) \qquad \text{for } \underline{t} \in S^d.$$

4 Metric entropy and Vapnik-Chervonenkis classes

4.1 Packing and covering numbers; metric entropy

As we have seen in 1.5, Glivenko-Cantelli convergence, i.e. $||\nu_n - \nu||_{\mathcal{C}} \longrightarrow 0 \ \mathbb{P} - a.s.$, fails to hold for all ν if the class $\mathcal{C}$ is "too rich"; cf. our Example 1.5.5.

The same is true in view of the validity of FCLT's:
Consider again 1.5.5 and the corresponding empirical $\mathcal{C}$-process $\beta_n = (\beta_n(C))_{C\in\mathcal{C}}$ with $\beta_n(C) := n^{1/2}\big(\nu_n(C) - \nu(C)\big)$; then: If (as in Theorem 2.2.1) the assertion

$$\beta_n \xrightarrow{\mathcal{L}} \mathbb{G}_\nu \qquad \text{in } (l^\infty(\mathcal{C}), || \cdot ||_{\mathcal{C}})$$

would hold true, the CMT 2.3.16 would imply that

$$\sup_{C\in\mathcal{C}} |\beta_n(C)| \xrightarrow{\mathcal{L}} \sup_{C\in\mathcal{C}} |\mathbb{G}_\nu(C)|,$$

where $\sup_{C\in\mathcal{C}} |\mathbb{G}_\nu(C,\omega)| < \infty$ for all ω, whence $||\nu_n - \nu||_{\mathcal{C}} = n^{-1/2} \sup_{C\in\mathcal{C}} |\beta_n(C)| \xrightarrow{\mathbb{P}} 0$ and also $||\nu_n - \nu||_{\mathcal{C}} \longrightarrow 0 \quad \mathbb{P} - a.s.$ according to 2.1.4, in contradiction to 1.5.5.

Thus, in order to obtain Uniform Laws of Large Numbers (ULLN) and Functional Central Limit Theorems (FCLT) for stochastic processes indexed by general parameter spaces T in Section 6 and 7, respectively, it is clear that T is not allowed to be "too rich".
To be "too rich" will be described through the behaviour of the so-called *metric entropy* of T, assuming that $T = (T,d)$ is a pseudo-metric space.

So, let $T = (T,d)$ be a pseudo-metric parameter space (e.g. $T = \mathcal{C} \subset \mathcal{X}, d = d_\nu, \nu$ p-measure on $\mathcal{X}$, $d_\nu(C_1,C_2) = \nu(C_1 \Delta C_2)$). Following Dudley ([Du84]) a set $\{t_1,...,t_n\} \subset T$ is called a *u-net* (for any given $u > 0$) iff for each $t \in T$ there is some t_i such that $d(t,t_i) \le u$.

This gives raise to define the so-called *covering numbers* of (T,d):

4.1.1. Definition.
For each $u > 0$, let

$$N(u,T,d) := \inf\{n \in \mathbb{N} : \exists u - net \ \{t_1,...,t_n\} \subset T\},$$

i.e. $N(u,T,d)$ is the minimal *number of points in a u-net. $H(\,\cdot\,,T,d) := \log N(\,\cdot\,,T,d)$ is called the* metric entropy *of $T = (T,d)$. (Note that $H(u,T,d)$ is increasing as $u \longrightarrow 0$.)*

(If $d \equiv 0$, that is $d(s,t) = 0 \ \forall s,t \in T$, we put $N(0,T,d) := 1$, whence in this case $N(u,T,d) \equiv 1$ and therefore $H(u,T,d) = 0 \, \forall u \ge 0$. So we may allow u to range in $[0,\infty)$.)

A closely related concept are the so-called *packing numbers* of (T,d). For this, given any $u > 0$, let $D(u,T,d)$ denote the largest m such that for some $t_1,...,t_m \in T \quad d(t_i,t_j) > u$ whenever $i \neq j$. The points $t_1,...,t_m$ may be called *u- distinguishable.*

4.1.2. Remark.
If $\{t_1, ..., t_n\}$ is a maximal u- distinguishable set, then $\{t_1, ..., t_n\}$ is a u-net.

4.1.3. Lemma.
For any $u > 0$

$$D(2u, T, d) \leq N(u, T, d) \leq D(u, T, d).$$

(So, in this sense covering numbers and packing numbers are equivalent concepts.)

As we shall see in Section 6 and 7, respectively, in order to obtain a ULLN for RMP's indexed by $\mathcal{F}$, stochastic boundedness of a sequence of *random* covering numbers of $\mathcal{F}$ (equipped with appropriate *random* pseudo-metrics) will be crucial, wheras, for a FCLT to hold, a uniform integrable L_2-entropy condition will be imposed.

4.2 Vapnik-Chervonenkis classes in arbitrary sample spaces $\boldsymbol{X} = (\boldsymbol{X}, \boldsymbol{\mathcal{X}})$

Let $X = (X, \mathcal{X})$ be an arbitrary measurable space and $\mathcal{C} \subset \mathcal{X}$ be a VCC (see (1.5.9), i.e.

$$\exists s \in \mathbb{N} \text{ s.t. } m^{\mathcal{C}}(s) < 2^s, \text{ where } m^{\mathcal{C}}(n) := \max\{\Delta^{\mathcal{C}}(F) : F \subset X, |F| = n\}$$

for each $n \in \mathbb{N}$, and $\Delta^{\mathcal{C}}(F) := |\{F \cap C : C \in \mathcal{C}\}|$.

Given a VCC $\mathcal{C}$, $v \equiv V(\mathcal{C}) := \min\{s \in \mathbb{N} : m^{\mathcal{C}}(s) < 2^s\}$ is the so-called *Vapnik-Chervonenkis Index* of $\mathcal{C}$.
According to the following lemma VCC's "are of polynomial discrimination".

4.2.1. Lemma *([Vap71]; cf. [Gae83], Lemma 9, p. 27).*
If $\mathcal{C}$ is a VCC, then $m^{\mathcal{C}}(n) \leq n^v$ for all $n \geq 2$.

(Note that for **arbitrary** $\mathcal{C}$ one has $m^{\mathcal{C}}(n) \leq 2^n \ \forall n \in \mathbb{N}$.)

Moreover, as shown by Alexander ([Al84], inequality (1.8)),

$$\mathcal{C} \text{ VCC} \implies (4.2.2): \quad m^{\mathcal{C}}(n) \leq \sum_{j \leq v-1} \binom{n}{j} \leq \left(\frac{ne}{v-1}\right)^{v-1} \quad \forall n \geq v-1.$$

In the following let $(X, \mathcal{X}, \nu)$ be an arbitrary p-space, and, given a class $\mathcal{C} \subset \mathcal{X}$, let $d_\nu(C_1, C_2) := \nu(C_1 \Delta C_2)), C_1, C_2 \in \mathcal{C}$. Then $(T, d) := (\mathcal{C}, d_\nu)$ is a pseudo-metric parameter space. In this situation Dudley ([Du78], Lemma 7.13) proved the following fundamental result:

4.2.3. Lemma.
Let $\mathcal{C} \subset \mathcal{X}$ be a VCC; then

$$N(u, \mathcal{C}, d_\nu) \leq K\, u^{-v} |\log u|^v \qquad \forall 0 < u \leq 1/2,$$

where the constant $0 < K < \infty$ does only depend on $v \equiv V(\mathcal{C})$ **but not on the p-measure ν on $\mathcal{X}$**.

As we shall see in 4.3 below, a corresponding inequality will hold in the more general case of Vapnik-Chervonenkis graph classes $\mathcal{F}$ of $\mathcal{X}$-measurable functions, containing 4.2.3 as a special case (with $\mathcal{F} = \{1_C : C \in \mathcal{C}\}$).

4.2.4. Examples of VCC's *(cf. [Wen81]).*

(a) $X = \mathbb{R}^d, d \geq 1, \mathcal{C} = \{(-\infty, \underline{t}] : \underline{t} \in \mathbb{R}^d\}$ *the class of all "lower-left orthants";* $s = d+1$.

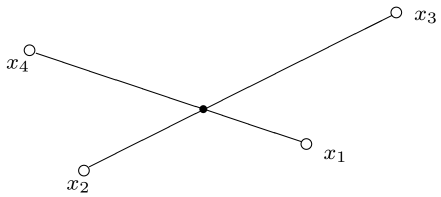
(b) $X = \mathbb{R}^d, d \geq 1, \mathcal{C} = \{\underset{i \leq d}{\times} [s_i, t_i] : -\infty < s_i \leq t_i < \infty\};\ s = 2d+1$.

(c) $X = \mathbb{R}^d, d \geq 1, \mathcal{C} = \{B \subset \mathbb{R}^d : B$ *closed Euclidian ball* $\};\ s = d+2$.

We want to present here an independent nice proof of (c) which I learned from Fleming Topsoe in 1976 (personal communication); the proof is based on the following two auxiliary results (+) and (++):

(+) *RADON'S THEOREM* (cf. [Val64], Thm. 1.26).
Each $F \subset \mathbb{R}^d, d \geq 1$, with $|F| \geq d+2$, can be decomposed into two (disjoint) subsets $F_i, i = 1, 2$, such that $co(F_1) \cap co(F_2) \neq \varnothing$ (where $co(F_i)$ denotes the convex hull of F_i).

For illustration, let $d = 2$:



$F = \{x_1, x_2, x_3, x_4\}$

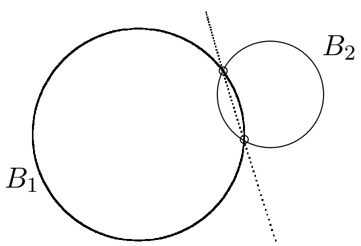
$F_1 = \{x_2, x_3\}$

$F_2 = \{x_1, x_4\}$.

(++) *SEPARATION PROPERTY.*
For any two closed Euclidian balls $B_i \subset \mathbb{R}^d, d \geq 1, i = 1, 2$, one has

$$co(B_1 \backslash B_2) \cap co(B_2 \backslash B_1) = \varnothing\ .$$

For illustration, consider again the case $d = 2$:



Now, in order to prove (c), we must show:

$$\forall F \subset \mathbb{R}^d \text{ with } |F| = s := d+2\ \ \exists F' \subset F \text{ s.t. } F' \neq F \cap B\ \ \forall B \in \mathcal{C}.$$

Suppose to the contrary that there exists an $F \subset \mathbb{R}^d$ with

$$|F| = d+2 \quad \text{s.t.} \quad \forall F' \subset F \;\; \exists B \in \mathcal{C} \;\; \text{with } F' = F \cap B.$$

This implies that for the F_i's in (+) (which decompose F) there exist two closed Euclidian balls $B_i, i = 1, 2$, such that $F_i = F \cap B_i$. Since $F_1 \cap F_2 = \varnothing$ it follows that $F_1 \subset B_1 \backslash B_2$ and $F_2 \subset B_2 \backslash B_1$, whence

$$co(B_1 \backslash B_2) \cap co(B_2 \backslash B_1) \supset co(F_1) \cap co(F_2) \neq \varnothing$$

(by (+)) which contradicts (++). □

Further examples of VCC's are obtained on the basis of the following Lemma (cf. [Du78], Thm. 7.2, and [Po84], Chap. II, Lemma 1.8):

4.2.5. Lemma.
Let $\mathcal{G}$ be an arbitrary m-dimensional vectorspace of real-valued functions g being defined on an arbitrary set X equipped with the σ-field $\mathcal{X} = \mathcal{P}(X)$ (whence each $g \in \mathcal{G}$ is measurable). Then the class $\mathcal{C} := \{\{g \geq 0\} : g \in \mathcal{G}\}$ is a VCC.

PROOF. W.l.o.g. let $|X| \geq m+1$ and let $A = \{x_1, ..., x_s\} \subset X$ be arbitrary with $|A| = s := m+1$; consider the linear map $L : \mathcal{G} \longrightarrow \mathbb{R}^s$, defined by $L(g) := (g(x_1), ..., g(x_s))$. Since $L(\mathcal{G})$ is a linear subspace of $\mathbb{R}^s$ with dimension $\leq m = s-1$, there exists a $v = (v_1, ...v_s) \in \mathbb{R}^s, v \neq 0$, s.t. $v \perp L(\mathcal{G})$, i.e. one has

(+) $$\sum_{i \leq s} v_i g(x_i) = 0 \qquad \forall g \in \mathcal{G}.$$

Now, let $A_+ := \{x_i \in A : v_i \geq 0\}$ and $A_- := \{x_i \in A : v_i < 0\}$, where w.l.o.g. $A_- \neq \varnothing$ (by replacing v through $-v$ otherwise). We are going to show

(++) $$A_+ \neq A \cap \{g \geq 0\} \qquad \forall g \in \mathcal{G},$$

from which the assertion of 4.2.5 follows.

As to (++), suppose to the contrary that there exists a $g \in \mathcal{G}$ s.t. $A_+ = A \cap \{g \geq 0\}$; then

$$\sum_{i \leq s} v_i g(x_i) = \sum_{i : x_i \in A_+} v_i g(x_i) + \sum_{i : x_i \in A_-} v_i g(x_i) > 0$$

in contradiction to (+). □

As an immediate consequence of the definition of a VCC it is clear that any subclass of a VCC is also a VCC.

As a permanence property, we mention here only the following lemma and its corollary:

4.2.6. Lemma *([Du78], Prop. (7.12)).*
Let $\mathcal{C}$ be a VCC and $k \in \mathbb{N}$ be arbitrary, but fixed. Given $C_1, ..., C_k \in \mathcal{C}$, let $\alpha(C_1, ..., C_k)$ be the algebra generated by $C_1, ..., C_k$ and

$$\alpha_k(\mathcal{C}) := \bigcup \{\alpha(C_1, ..., C_k) : C_1, ..., C_k \in \mathcal{C}\};$$

then $\alpha_k(\mathcal{C})$ is also a VCC.

4.2.7. Corollary.
If $\mathcal{C}$ is a VCC, then the classes $\{C\Delta D : C, D \in \mathcal{C}\}$, $\{C\backslash D : C, D \in \mathcal{C}\}$ *and* $\{C \cap D : C, D \in \mathcal{C}\}$, *respectively, form also VCC's.*

As to further examples we refer to Stengle and Yukich [Sten89] and Laskowski [La92]; see also the references in the following more general context.

4.3 Vapnik-Chervonenkis graph classes of $\mathcal{X}$-measurable functions $f : X \longrightarrow \mathbb{R}$

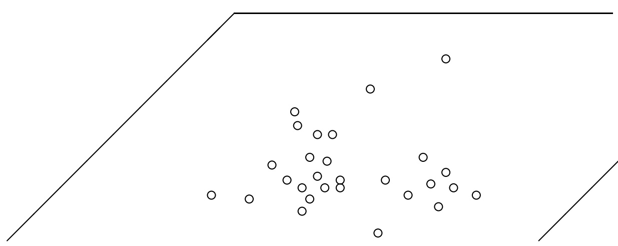
So far we have mainly considered examples of processes indexed by classes $\mathcal{C}$ of sets, where $\mathcal{C} \subset \mathcal{X}, (X, \mathcal{X})$ being a given measurable space.
To motivate the need for extending the index sets from classes of sets to classes of functions, we present the following examples A) and B):

A) POLLARD's k-MEANS CLUSTERING PROCEDURE

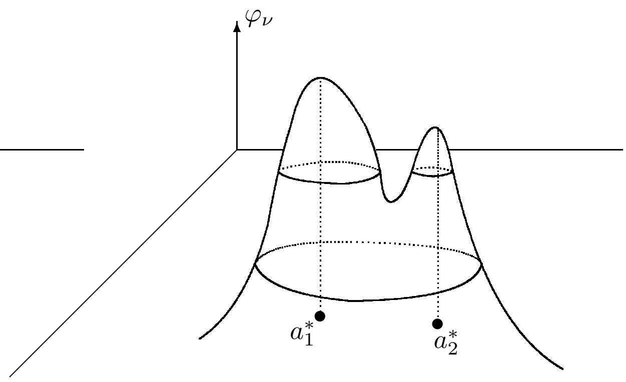
(Cf. [Po84], Example 4, p.9 and Example 29, p. 30; [Po82a] and [Po82b]; [Gae87]; [Ro91] and [Ro95].)

Given data $x_1, ..., x_n \in X = \mathbb{R}^d$ viewed as realizations of i i d re's ξ_j in $(\mathbb{R}^d, \mathcal{B}^d)$ (on a basic p-space $(\Omega, \mathcal{A}, \mathbb{P})$) (i.e. $x_j = \xi_j(\omega)$), let $k \in \mathbb{N}$ be arbitrary but fixed (i.e. k is given in advance).
Suppose that the underlying unknown $\nu := \mathcal{L}\{\xi_j\}$ is "k-modal" (e.g. with Lebesgue-density φ_ν having k modes).

Consider $d = 2$ and $k = 2$ for illustration (i.e. φ_ν bimodal)



$\circ \; = \; \xi_j(\omega), 1 \leq j \leq n.$



a_1^* and a_2^* are the unknown modes of ν (modes of φ_ν).

The question arises how to choose k data-clusters with empirical cluster centers $a_{ni}^* = a_{ni}^*(\omega)$ (based on the data $x_j = \xi_j(\omega), 1 \leq j \leq n$), $1 \leq i \leq k$, such that, as the sample size n tends to infinity, the a_{ni}^* converge $\mathbb{P} - a.s.$ to the unknown modes a_i^* of ν, $1 \leq i \leq k$.

An answer is provided by the *k-means clustering procedure (CP)*:

(CP): Given the data $x_j (= \xi_j(\omega)), 1 \le j \le n, n \in \mathbb{N}$, determine a k-tuple $(a^*_{n1}, ..., a^*_{nk})$ which minimizes the expression

$$n^{-1} \sum_{j \le n} \min_{1 \le i \le k} |x_j - a_i|^2 \tag{4.3.1}$$

over all $(a_1, ..., a_k)$ with $a_i \in \mathbb{R}^d$, and then allocate each x_j to its nearest a^*_{ni}.

Let ν_n be the empirical measure based on $\xi_1, ..., \xi_n$ and let for $a_i \in \mathbb{R}^d, 1 \le i \le k$,

$$W(a_1, ..., a_k; \nu_n) := \int_{\mathbb{R}^d} \min_{1 \le i \le k} |x - a_i|^2 \, \nu_n(dx); \tag{4.3.2}$$

then

$$W(a^*_{n1}, ..., a^*_{nk}; \nu_n) = \min_{(a_1, ..., a_k)} W(a_1, ..., a_k; \nu_n). \tag{4.3.3}$$

In the following we confine to the case $d = 1$ and $k = 2$) (i.e. ν being bimodal), and we shall write (a^*_n, b^*_n) and (a^*, b^*) instead of (a^*_{n1}, a^*_{n2}) and (a^*_1, a^*_2), respectively.

Now, consider (instead of a class $\mathcal{C}$ of sets) the following class $\mathcal{F}$ of $\mathcal{B}$-measurable functions $f : \mathbb{R} \longrightarrow \mathbb{R}$:

$$\mathcal{F} := \{ f_{a,b} : (a, b) \in C_M \} \tag{4.3.4}$$

where $f_{a,b}(x) := |x - a|^2 \wedge |x - b|^2, x \in \mathbb{R}$, and $C_M := \big([-M, M] \times \mathbb{R} \big) \cup \big(\mathbb{R} \times [-M, M] \big)$, where $M > 0$ is chosen large enough (see [Po84], p.10).

We assume that

$$\int_{\mathbb{R}} x^2 \, \nu(dx) < \infty,$$

whence $W(a, b; \nu) := \int_{\mathbb{R}} f_{a,b}(x) \, \nu(dx) < \infty \;\; \forall (a, b) \in \mathbb{R}^2$.

Then, by the strong law of large numbers we have $\forall (a, b) \in \mathbb{R}^2$

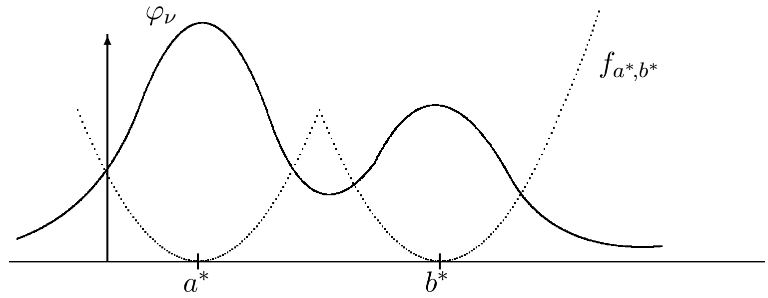
$$W(a, b; \nu_n) = \nu_n(f_{a,b}) = n^{-1} \sum_{j \le n} f_{a,b}(\xi_j) \longrightarrow \nu(f_{a,b}) = W(a, b; \nu) \quad \mathbb{P} - a.s. \tag{4.3.5}$$

Now, a further assumption on ν is needed:

(4.3.6) $\exists (a^*, b^*) \in \mathbb{R}^2$ being uniquely determined up to permutation of its coordinates such that

$$\nu(f_{a^*, b^*}) = \min_{(a,b) \in \mathbb{R}^2} \nu(f_{a,b}).$$

The following picture is to visualize (4.3.6) for ν with density φ_ν:



On the other hand, i.e. on the empirical side, we have by (4.3.3)

(4.3.7) $$\nu_n(f_{a_n^*,b_n^*}) = \min_{(a,b)\in\mathbb{R}^2} \nu_n(f_{a,b}).$$

Thus, (4.3.5) - (4.3.7) gives raise to expect

(4.3.8) $$(a_n^*, b_n^*) \longrightarrow (a^*, b^*) \qquad \mathbb{P}-a.s.$$

i.e. $\mathbb{P}-a.s$ convergence of the empirical cluster centers to the unknown modes of ν.

(4.3.8) can be proved by showing

(4.3.9) $$\sup_{(a,b)\in C_M} |\nu_n(f_{a,b}) - \nu(f_{a,b})| \longrightarrow 0 \quad \mathbb{P}-a.s.$$

As to (4.3.9), $\sup_{(a,b)\in C_M} |\nu_n(f_{a,b}) - \nu(f_{a,b})| = \sup_{f\in\mathcal{F}} |\nu_n(f) - \nu(f)|$ with $\mathcal{F}$ as defined by (4.3.4) being a Vapnik-Chervonenkis graph class (VCGC) (see below for the definition of VCGC's of functions). Thus (4.3.9) proves to be a consequence of a result generalizing Theorem 2.1.6 from VCC's $\mathcal{C}$ to VCGC's $\mathcal{F}$ with $\nu(F) < \infty$, F being an envelope of $\mathcal{F}$; note that in the present case

$$\sup_{(a,b)\in C_M} f_{a,b}(x) \leq F(x) := (x-M)^2 + (x+M)^2 \quad \forall x \in \mathbb{R}$$

and $\nu(F) < \infty$ since $\int_{\mathbb{R}} x^2\, \nu(dx) < \infty$ by assumption. (Cf. [Po84], Example 29, and Section 6.3 below).

To sketch the proof of (4.3.8) on the basis of (4.3.9) one shows at first

(a): For sufficiently large $M > 0$ $(a_n^*, b_n^*) \in C_M$ $\mathbb{P}-a.s.$ $\forall n \geq n_0$.

Then one gets

(b): $\nu_n(f_{a_n^*,b_n^*}) \underset{(4.3.7)}{\leq} \nu_n(f_{a^*,b^*}) \longrightarrow \nu(f_{a^*,b^*})$ $\mathbb{P}-a.s.$ by (4.3.5), where $\mathbb{P}-a.s.$

$\nu(f_{a^*,b^*}) \underset{(4.3.6)}{\leq} \nu(f_{a_n^*,b_n^*}) \underset{\substack{\text{for large } n \\ \text{by (4.3.9) and (a)}}}{\approx} \nu_n(f_{a_n^*,b_n^*})$, which yields

$$\nu(f_{a_n^*,b_n^*}) \longrightarrow \nu(f_{a^*,b^*}) \quad \mathbb{P}-a.s.$$

(c): Finally, since by (4.3.6) (a^*, b^*) is uniquely determined (up to permutation of its coordinates), (b) implies (4.3.8). $\square$

A modification of the k-means clustering procedure for k^*-modal ν with k^* *unknown* to obtain $\mathbb{P}$-a.s. convergence of $k(n)$ to k^* and simultaneously

$$(4.3.8') \qquad (a^*_{n1}, ..., a^*_{nk(n)}) \longrightarrow (a^*_1, ..., a^*_{k^*}) \quad \mathbb{P} - a.s.,$$

where $k(n)$ denotes the number of empirical clusters w.r.t. a modification of the empirical clustering procedure, is contained in [Gae87], [Ro91] and [Ro95].

B) LOCAL EMPIRICAL PROCESSES, STUTE's CONDITIONAL EMPIRICAL PROCESSES AND CONDITIONAL EMPIRICAL DISTRIBUTION FUNCTIONS

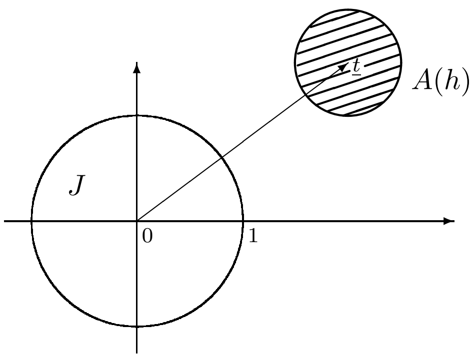
(See [Ei97], [Stu86a] and [Stu86b].)

As pointed out in [Ei97], local empirical processes occur implicitely in the work of Kim and Pollard [Ki90] on cube root asymptotics and of Nolan and Marron [No89] on automatic band width selection; local empirical-type processes arise also in certain interval censoring and deconvolution problems (see Part II of Groeneboom and Wellner [Gro92]).

Let $\xi_j, j \in \mathbb{N}$, be i i d re's in $(\mathbb{R}^d, \mathcal{B}^d), d \geq 1$, defined on a p-space $(\Omega, \mathcal{A}, \mathbb{P})$ with df G.
Let $\underline{t} \in \mathbb{R}^d$ and $J \in \mathcal{B}^d$ be arbitrary, but fixed. Given an invertible bimeasurable transformation $h : \mathbb{R}^d \longrightarrow \mathbb{R}^d$, let

$$(4.3.10) \qquad A(h) := \underline{t} + hJ \qquad (\text{where } hJ := \{h(\underline{x}) : \underline{x} \in J\}).$$

To visualize $A(h)$, let $d = 2, J = E$ (the unit ball in $\mathbb{R}^2$) and $h(\underline{x}) := \frac{1}{2}|\underline{x}|, \underline{x} \in \mathbb{R}^2$:



Next, let $(h_n)_{n \in \mathbb{N}}$ be a sequence of invertible bimeasurable transformations $h : \mathbb{R}^d \longrightarrow \mathbb{R}^d$ and assume for

$$A_n := A(h_n) \quad \text{and} \quad a_n := \mathbb{P}(\xi_j \in A_n),\ n \in \mathbb{N},$$

the following set of conditions

$$(A.i) \quad a_n > 0\ \forall n \in \mathbb{N}, \qquad (A.ii) \quad na_n \longrightarrow \infty \ \text{ and } \ (A.iii) \quad a_n \longrightarrow 0 \ \text{ as } n \longrightarrow \infty.$$

(Note that (A.iii) implies that G is continuous at $\underline{t}$; otherwise (A.iii) may be replaced by (A.iii') $\quad a_n \longrightarrow a$ for some $0 \le a \le 1$.)

For each $n \in \mathbb{N}$, let $\nu_n(\underline{t},\,\cdot\,)$ be defined by

$$\text{(4.3.11)} \qquad \nu_n(\underline{t}, B) := (na_n)^{-1} \sum_{j \le n} I(\xi_j \in \underline{t} + h_n(J \cap B))\,, B \in \mathcal{B}^d$$

(where $I(\,\cdot\,) \equiv 1_{\{\,\cdot\,\}}$).
$\nu_n(\underline{t},\,\cdot\,)$ is called *local empirical measure* at $\underline{t}$.

(Note that $\nu_n(\underline{t},\,\cdot\,)$ need not be a probability measure; one only has that $\mathbb{E}(\nu_n(\underline{t}, \mathbb{R}^d)) = 1$.)

Now, let $\mathcal{F}$ be a class of measurable functions $f : \mathbb{R}^d \longrightarrow \mathbb{R}$ with supports contained in J (i.e. $f(x) = 0 \;\forall x \in \mathbb{R}^d \backslash J \;\forall f \in \mathcal{F}$), and let

$$\text{(4.3.12)} \qquad \nu_n(\underline{t}, f) := \int_J f(x) \nu_n(\underline{t}, dx) \underset{(4.3.11)}{=} (na_n)^{-1} \sum_{j \le n} f(h_n^{-1}(\xi_j - \underline{t})), \quad f \in \mathcal{F},$$

where h_n^{-1} denotes the inverse of h_n.

∘ Note at this place that for any fixed $\underline{t} \in \mathbb{R}^d$ and $J \in \mathcal{B}^d$ the processes $(\nu_n(\underline{t}, f))_{f \in \mathcal{F}}$ can be considered as RMP's (see (3.1.1)) by choosing as random p-measures w_{nj} and as rv's $\xi_{nj}, 1 \le j \le n, n \in \mathbb{N}$,

$$w_{nj} := \delta_{\eta_{nj}} \text{ with } \eta_{nj} := h_n^{-1}(\xi_j - \underline{t}), \qquad \text{and } \xi_{nj} := (na_n)^{-1}.$$

(In fact, for each $f \in \mathcal{F}$, $\sum_{j \le n} w_{nj}(f) \xi_{nj} = (na_n)^{-1} \sum_{j \le n} \int_{\mathbb{R}^d} f(x) \delta_{\eta_{nj}}(dx) = (na_n)^{-1} \sum_{j \le n} f(\eta_{nj}) = (na_n)^{-1} \sum_{j \le n} f(h_n^{-1}(\xi_j - \underline{t})) = \nu_n(\underline{t}, f)$.)

The standardized process $\beta_n^{loc}(\underline{t}) = (\beta_n^{loc}(\underline{t}, f))_{f \in \mathcal{F}}$ with $\beta_n^{loc}(\underline{t}, f) := (na_n)^{1/2} \Big(\nu_n(\underline{t}, f) - \mathbb{E}(\nu_n(\underline{t}, f))\Big)$ is called *local empirical process* at $\underline{t}$ indexed by $\mathcal{F}$.

This setup allows to consider the following interesting examples (see [Ei97]):

4.3.13. Example.
Let $\xi_j, j \in \mathbb{N}$, be i i d rv's on a p-space $(\Omega, \mathcal{A}, \mathbb{P})$ with df G having a continuous density g in a neighborhood of a fixed $t \in \mathbb{R}$.
Set $J := [-\frac{1}{2}, \frac{1}{2}]$, $\mathcal{F} := \{K\}$ with a so-called kernel function K satisfying $K(x) = 0$ if $|x| > \frac{1}{2}$.
Let $h_n(x) := h_n \cdot x, x \in \mathbb{R}$, with $h_n > 0$ and $h_n \longrightarrow 0$ as $n \to \infty$. Then,

$$\nu_n(t, K) \underset{(4.3.12)}{=} (na_n)^{-1} \sum_{j \le n} K\Big(\frac{\xi_j - t}{h_n}\Big) = h_n a_n^{-1} \hat{g}_n(t),$$

where $\hat{g}_n(t) := (nh_n)^{-1} \sum_{j \le n} K\Big(\frac{\xi_j - t}{h_n}\Big)$ is the so-called kernel density estimator of $g(t)$ with window size h_n.

(Note that $h_n^{-1} a_n = h_n^{-1} \mathbb{P}(\xi_j \in t + h_n J) = h_n^{-1} \mathbb{P}(\xi_j \in [t - \frac{1}{2}h_n, t + \frac{1}{2}h_n]) = h_n^{-1} \int_{t - \frac{1}{2}h_n}^{t + \frac{1}{2}h_n} g(x)\, dx \longrightarrow g(t)$ as $n \to \infty$.)

4.3.14. Example.
Let $d = 2$, $\xi_j = (\zeta_j, \eta_j), j \in \mathbb{N}$, iid re's in $(\mathbb{R}^2, \mathcal{B}^2)$ with df G having density $g_{\zeta,\eta}$ and marginal densities g_ζ and g_η, respectively. Choose $J := [-\frac{1}{2}, \frac{1}{2}] \times \mathbb{R}, \underline{t} := (t, 0), t \in \mathbb{R}$, and, for $(z, y) \in \mathbb{R}^2$ let $h_n(z, y) := (h_n \cdot z, y)$ with $h_n > 0$ and $h_n \longrightarrow 0$ as $n \to \infty$. Let K be a kernel function as in 4.3.13 and $\mathcal{F} := \{R\}$ with

$$(+) \qquad R(z, y) := y \cdot K(z) \quad ,(z, y) \in \mathbb{R}^2.$$

Then,

$$\nu_n(\underline{t}, R) \underset{(4.3.12)}{=} (na_n)^{-1} \sum_{j \le n} R(h_n^{-1}(\xi_j - \underline{t})) = (na_n)^{-1} \sum_{j \le n} \eta_j K\Big(\frac{\zeta_j - t}{h_n}\Big),$$

since $h_n^{-1}(\xi_j - \underline{t}) = h_n^{-1}(\zeta_j - t, \eta_j) = (\frac{\zeta_j - t}{h_n}, \eta_j)$ and therefore $R(h_n^{-1}(\xi_j - \underline{t})) = R\Big(\frac{\zeta_j - t}{h_n}, \eta_j\Big) \underset{(+)}{=} \eta_j K\Big(\frac{\zeta_j - t}{h_n}\Big)$. Thus

$$\nu_n(\underline{t}, R) = \hat{m}_n(t) h_n a_n^{-1} \hat{g}_n(t) = \hat{m}_n(t) \nu_n(t, K) \quad (cf.\, 4.3.13),$$

where $\hat{g}_n(t)$ is the kernel density estimator of the marginal density $g_\zeta(t)$ and $\hat{m}_n(t)$ is the kernel regression estimator of $m(t) := \mathbb{E}(\eta | \zeta = t)$ defined by

$$\hat{m}_n(t) := \frac{(nh_n)^{-1} \sum_{j \le n} \eta_j K\Big(\frac{\zeta_j - t}{h_n}\Big)}{\hat{g}_n(t)}.$$

4.3.15. Example.
Keeping up the notation of example 4.3.14, choose now, instead of $\mathcal{F} = \{R\}$, the class $\mathcal{F} = \{f_v : v \in \mathbb{R}\}$ of functions f_v defined by

$$f_v(z, y) := I(y \le v) K(z), \quad (z, y) \in \mathbb{R}^2;$$

then (again with $\underline{t} = (t, 0), t \in \mathbb{R}$)

$$\begin{aligned} \nu_n(\underline{t}, f_v) &:= (na_n)^{-1} \sum_{j \le n} I(\eta_j \le v) K\Big(\frac{\zeta_j - t}{h_n}\Big) = F_n(v|t) h_n a_n^{-1} \hat{g}_n(t) \\ &= F_n(v|t) \nu_n(t, K) \qquad (cf.\, 4.3.13) \end{aligned}$$

with

$$F_n(v|t) := \frac{(nh_n)^{-1} \sum_{j \le n} I(\eta_j \le v) K\Big(\frac{\zeta_j - t}{h_n}\Big)}{\hat{g}_n(t)},$$

which are the conditional empirical distribution functions (of sample size n) first intensively studied by Stute [Stu86a],[Stu86b].

As to an empirical process approach to the uniform consistency of kernel-type function estimators we refer to a very remarkable forthcoming paper by Uwe Einmahl and David Mason [Ei98].

In the following let again $X = (X, \mathcal{X})$ be an arbitrary measurable space and $\mathcal{F}$ be a class of $\mathcal{X}$-measurable functions $f : X \longrightarrow \mathbb{R}$ with $\mathcal{X}$-measurable envelope $F : X \longrightarrow \mathbb{R}_+$ (i.e. $\sup_{f \in \mathcal{F}} |f(x)| \leq F(x) \ \forall x \in X$).
Generalizing the concept of VCC's $\mathcal{C} \subset \mathcal{X}$ (equivalently $\{1_C : C \in \mathcal{C}\}$) to more general classes $\mathcal{F}$ of $\mathcal{X}$-measurable functions $f : X \longrightarrow \mathbb{R}$ leads to

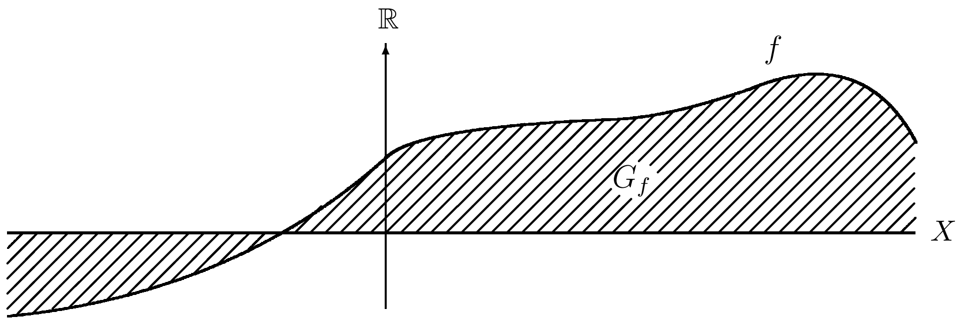
4.3.16. Definition.
$\mathcal{F}$ is called a Vapnik-Chervonenkis graph class (VCGC) *if*

$$\mathcal{R} := \{G_f : f \in \mathcal{F}\}$$

is a VCC in $(X \times \mathbb{R}, \mathcal{X} \bigotimes \mathcal{B})$, where

$$G_f := \{(x, t) \in X \times \mathbb{R} : 0 \leq t \leq f(x) \ \text{ or } \ f(x) \leq t \leq 0\}.$$

$G_f \subset X \times \mathbb{R}$ is the so-called graph region *associated to f.*



NOTE: f $\mathcal{X}$-measurable $\Longrightarrow$ $G_f \in \mathcal{X} \bigotimes \mathcal{B}$ whence the *graph region class* $\mathcal{R}$ is a subclass of $\mathcal{X} \bigotimes \mathcal{B}$.

Given a VCGC $\mathcal{F}$ we denote with $V(\mathcal{R})$ the Vapnik-Chervonenkis Index of the graph region class $\mathcal{R}$ (cf. 4.2) corresponding to $\mathcal{F}$.
Clearly, if $\mathcal{C} \subset \mathcal{X}$ is a VCC, then $\mathcal{F} := \{1_C : C \in \mathcal{C}\}$ is a VCGC with $V(\mathcal{R}) = V(\mathcal{C})$.

Examples of VCGC's as well as permanence properties which allow to construct new VCGC's from known ones are contained in [Po84] (there called "classes of polynomial discrimination") and [Va96], Section 2.6.5.

The present graph regions G_f are called *"between graphs"* in [Va96]; compared with the open *subgraphs* of f, defined by $\{(x, t) : t < f(x)\}$, which led to the concept of *Vapnik-Chervonenkis subgraph classes (VCSGC)* of functions in [Va96], Section 2.6.2, it turns out that $\mathcal{F}$ is a VCGC if and only if it is a VCSGC; see [Va96], Problem 11, p. 152. Thus both concepts are equivalent.

The following fundamental lemma is mentioned in [Al84]. It generalizes lemma 4.2.3 above; but notice that in addition the assumption of ν being a p-measure on $\mathcal{X}$ can be dispensed with. The proof, as carried out by Klaus Ziegler [Zi94], Lemma A4, combines the methods in proving lemma 2.7 in [Al84] and lemma 25 in Section II.5 of [Po84].

4.3.17. Lemma.
Let $\mathcal{F}$ be a VCGC with envelope F and graph region class $\mathcal{R}$. Then there exists a constant $0 < K(v) < \infty$ depending only on $v \equiv V(\mathcal{R})$ such that for all measures ν on $\mathcal{X}$ with $\nu(F) := \int_X F\,d\nu < \infty$

$$N(\varepsilon\nu(F), \mathcal{F}, d_\nu^{(1)}) \leq K(v)\varepsilon^{-(v-1)}|\log\varepsilon|^{v-1} \qquad \forall 0 < \varepsilon \leq \frac{1}{2}.$$

Here, as in 4.2.3, $\log\varepsilon = \log_e\varepsilon \equiv \ln\varepsilon$, and $d_\nu^{(1)}$ is defined by $d_\nu^{(1)}(f,g) := \nu(|f-g|),\ f,g \in \mathcal{F}$.

NOTE: $\lim_{\varepsilon\to 0}|\log\varepsilon|^\alpha\,\varepsilon^\beta = 0\ \forall\alpha,\beta > 0$, so $|\log\varepsilon|^{v-1} \leq \varepsilon^{-(v-1)}$ for small ε, whence

$$N(\varepsilon\nu(F), \mathcal{F}, d_\nu^{(1)}) \leq K(v)\varepsilon^{-2(v-1)} \quad \forall 0 < \varepsilon \leq \frac{1}{2}.$$

Also, in the special case $\mathcal{F} := \{1_C : C \in \mathcal{C}\}, \mathcal{C}$ VCC, $F \equiv 1$, $d_\nu^{(1)} = d_\nu$, ν an arbitrary p-measure on $\mathcal{X}$, 4.3.17 yields a sharpened version of 4.2.3.

PROOF. W.l.o.g. assume $v \geq 2$; let $0 < \varepsilon \leq \frac{1}{2}$ be arbitrary, but fixed, and choose $f_1, ..., f_m \in \mathcal{F}$ (w.l.o.g. $m \geq 2$) s.t.

(1) $\quad d_\nu^{(1)}(f_i, f_j) := \nu(|f_i - f_j|) > \varepsilon\nu(F) \quad$ for $i \neq j$.

Let n be the smallest natural number s.t.

(2) $\quad \frac{1}{2}\exp(2\log m - n\varepsilon/2) < 1.$

Then, by elementary calculations, one gets

(3) $\quad n \leq (1 + 4\log m)/\varepsilon \leq 15Lm/e\varepsilon$

where $La := \max(1, \log a)$.

Now, a stochastic argument comes into play (cf. Dudley's ingenious proof of lemma 4.2.3):

Let μ be the p-measure on $\mathcal{X}$ defined by

$$\mu(A) := \nu(F)^{-1}\int_A F\,d\nu, \quad A \in \mathcal{X},$$

and let $K : X \times \mathcal{B} \longrightarrow [0,1]$ be the stochastic kernel defined by

$$K(x,B) := U[-F(x), F(x)](B), \quad x \in X, B \in \mathcal{B},$$

where $U[a,b]$ denotes the uniform distribution on $[a,b]$. Let $\xi_1, ..., \xi_n$ (with n chosen as above, cf. (2)) be i i d re's in $(X \times \mathbb{R}, \mathcal{X}\bigotimes\mathcal{B})$, defined on a basic p-space $(\Omega, \mathcal{A}, \mathbb{P})$, with $\mathcal{L}\{\xi_j\} = \mu \times K$, where $\mu \times K$ is the p-measure on $\mathcal{X}\bigotimes\mathcal{B}$ defined by

$$\mu \times K(C) := \int_X U[-F(x), F(x)]\big(\{t \in \mathbb{R} : (x,t) \in C\}\big)\,\mu(dx), \qquad C \in \mathcal{X}\bigotimes\mathcal{B}.$$

(To see that K is indeed a stochastic kernel, use the fact that $\mathcal{X}$-measurability of $K(\,\cdot\,, B)$ has only to be checked for sets B of the form $(-\infty, t], t \in \mathbb{R}$. So, fix $t \in \mathbb{R}$ and distinguish between the three cases $t > 0, t = 0$, and $t < 0$. Then, e.g. in the case $t > 0$ we obtain

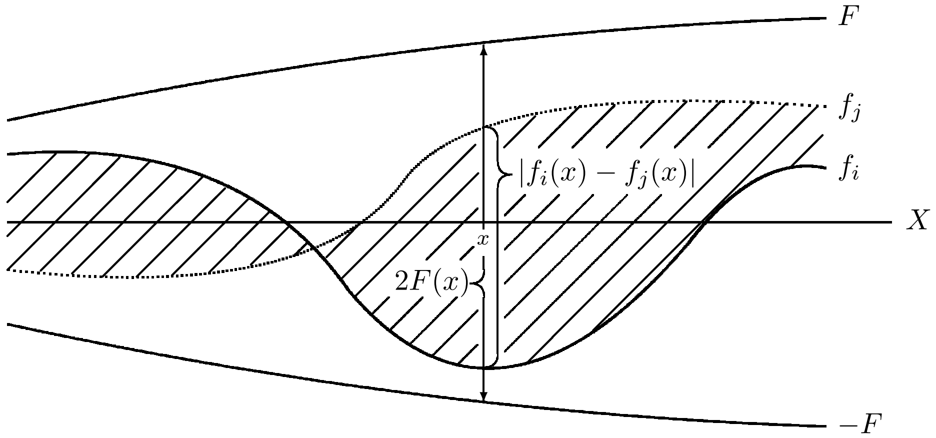
$$K(\,\cdot\,, (-\infty, t]) = 1_{\{F \leq t\}} + ((t+F)/2F)\,1_{\{F \geq t\}}$$

which is $\mathcal{X}$-measurable by the assumed measurability of F.)

Now, let $G_i := G_{f_i}$ be the graph regions of $f_i, 1 \le i \le m$; then (with the convention that in the case $F(x) = 0$ we set $\frac{0}{0} := 0$)

$$U[-F(x), F(x)]\big(\{t \in \mathbb{R} : (x,t) \in G_i \Delta G_j\}\big) = (2F(x))^{-1} |f_i(x) - f_j(x)| \quad \forall x \in X, \tag{4}$$

as the following picture shows (with $G_i \Delta G_j$ being the hatched region).



Furthermore, we have

(5) $\quad \forall \xi_k(\omega), \omega \in \Omega, 1 \le k \le n$, and $\forall 1 \le i, j \le m, i \ne j$

$$S'(\omega) \cap G_i = S'(\omega) \cap G_j \quad \Longleftrightarrow \quad \xi_k(\omega) \notin G_i \Delta G_j \quad \forall 1 \le k \le n,$$

where $S'(\omega) := \{\xi_1(\omega), ..., \xi_n(\omega)\}$.

(Note that $S'(\omega) \cap G_i \ne S'(\omega) \cap G_j \iff \exists 1 \le k_0 \le n$ s.t. $\xi_{k_0}(\omega) \in G_i \Delta G_j$; note also that $|S'(\omega)| \le n$.)

Next, we are going to show that

$$m \le m^{\mathcal{R}}(n) := \max\Big\{ |\{S \cap G_f : f \in \mathcal{F}\}| : S \subset X \times \mathbb{R}, |S| = n \Big\} :$$

For this, consider first any fixed $1 \le i, j \le m, i \ne j$; then

$$\begin{aligned}
\mathbb{P}(\xi_k \notin G_i \Delta G_j \ \forall 1 \le k \le n) &\underset{(\xi_k \text{ iid })}{=} [1 - \mathbb{P}(\xi_1 \in G_i \Delta G_j]^n \ = \ [1 - (\mu \times K)(G_i \Delta G_j]^n \\
&= [1 - \int_X U[-F(x), F(x)]\big(\{t \in \mathbb{R} : (x,t) \in G_i \Delta G_j\}\big)\, \mu(dx)]^n \\
&\underset{(4)}{=} [1 - \int_X (2F(x))^{-1} |f_i(x) - f_j(x)|\, \mu(dx)]^n \\
&= \ [1 - \nu(|f_i - f_j|)(2\nu(F))^{-1}]^n \underset{(1)}{<} (1 - \varepsilon/2)^n \le \exp(-n\varepsilon/2).
\end{aligned}$$

Therefore, according to (5) we get

$$\begin{aligned}\mathbb{P}(\{\omega \in \Omega:\ \exists i \neq j \ \text{ s.t. } S'(\omega) \cap G_i = S'(\omega) \cap G_j\}) \ &< \ \binom{m}{2} \exp(-n\varepsilon/2)\\ \leq \frac{1}{2} m^2 \exp(-n\varepsilon/2) \ = \ \frac{1}{2}\exp(2\log m - n\varepsilon/2) \ &\underset{(2)}{<} \ 1,\end{aligned}$$

whence

$$\mathbb{P}(\{\omega \in \Omega:\ S'(\omega) \cap G_i \neq S'(\omega) \cap G_j \quad \forall 1 \leq i, j \leq m, i \neq j\}) > 0.$$

Therefore there exist $\leq n$ points in $X \times \mathbb{R}$ from which $\mathcal{R}$ picks out m distinct subsets of $S'(\omega_0)$ for some $\omega_0 \in \Omega$. Since $S'(\omega_0) \subset X \times \mathbb{R}$ and $|S'(\omega_0)| \leq n$, it follows (by definition of $m^{\mathcal{R}}(n)$ and the fact that $m^{\mathcal{R}}(n)$ is increasing in n) that $m \leq m^{\mathcal{R}}(n)$, whence by (3)

$$m \leq m^{\mathcal{R}}(15Lm/e\varepsilon).$$

Now, *if* $15Lm/e\varepsilon \geq v-1$, it follows by (4.2.2) (according to which $m^{\mathcal{R}}(n) \leq \left(\frac{ne}{v-1}\right)^{v-1} \ \forall n \geq v-1$) that $m \leq m^{\mathcal{R}}(15Lm/e\varepsilon) \leq (15Lm/(v-1)\varepsilon)^{v-1}$, which implies

(6) $\quad m \leq (30\varepsilon^{-1}\log(15\varepsilon^{-1}))^{v-1}.$

(In fact, since $Lm/(v-1) \leq L(m^{1/(v-1)}) \leq m^{1/2(v-1)}$, we have $m \leq 15^{v-1} m^{1/2} \varepsilon^{-(v-1)}$ and thus $m \leq (15\varepsilon^{-1})^{2(v-1)}$. Hence $m \leq (15L((15\varepsilon^{-1})^{2(v-1)})/(v-1)\varepsilon)^{v-1} = (30\varepsilon^{-1}\log(15\varepsilon^{-1}))^{v-1}$.)

On the other hand, *if* $15Lm/e\varepsilon < v-1$, it follows that $\log m \leq Lm < e\varepsilon(v-1)/15 < \varepsilon(v-1) \leq (v-1)\log(30\varepsilon^{-1}) = \log[(30\varepsilon^{-1})^{v-1}]$, whence $m \leq (30\varepsilon^{-1})^{v-1} \underset{(\varepsilon \leq 1/2)}{\leq} (30\varepsilon^{-1}\log(15\varepsilon^{-1}))^{v-1}$; i.e. (6) holds also true in this case.

But (6) implies (recall $\varepsilon \leq 1/2$)

(7) $\quad m \leq K(v)\varepsilon^{-(v-1)}|\log\varepsilon|^{v-1}$ with

$$K(v) := (30(\log 15 + \log 2)/\log 2)^{v-1}.$$

Finally, taking now $m = m(\varepsilon)$ maximal s.t. (1) is fulfilled (note that $m(\varepsilon) < \infty$ by (6)), we obtain by (7) for the packing numbers $D(\varepsilon\nu(F), \mathcal{F}, d_\nu^{(1)})$ that

$$D(\varepsilon\nu(F), \mathcal{F}, d_\nu^{(1)}) \leq K(v)\varepsilon^{-(v-1)}|\log\varepsilon|^{v-1},$$

which implies the assertion of 4.3.17 according to the second inequality in 4.1.3. □

REMARK.

The above method of proof is very interesting in its own right:

In order to verify that a certain situation holds true, an appropriate stochastic model is constructed within which it is shown that a proper event occurs with positive probability, from which one then infers the **existence** of the situation one was interested in.

As discovered by Alexander [Al87], there is an elegant way to pass from the upper bound in 4.3.17 w.r.t. L_1-entropy (i.e. concerning $d_\nu^{(1)}$) to an analogous result for L_2-entropy, i.e. concerning $d_\nu^{(2)}$ instead of $d_\nu^{(1)}$, where

$$d_\nu^{(2)}(f,g) := \Big(\nu(|f-g|^2)\Big)^{1/2}, \qquad f,g \in \mathcal{F}.$$

This is done by using the elementary inequality

$$(4.3.18) \qquad (a-b)^2 \le 2|a^2\, sign(a) - b^2\, sign(b)| \qquad \forall a,b \in \mathbb{R},$$

where

$$sign(a) := \begin{cases} 1 & , \text{ if } a > 0 \\ 0 & , \text{ if } a = 0 \\ -1 & , \text{ if } a < 0 \end{cases} \quad , \quad a \in \mathbb{R}.$$

4.3.19. Corollary*([Zi94], Cor. A5; see also Lemma 36 in Section II.6 of [Po84] for a different method of proof).*
Let $\mathcal{F}$ be a VCGC with envelope F and graph region class $\mathcal{R}$. Then there exists a constant $0 < K'(v) < \infty$ depending only on $v \equiv V(\mathcal{R})$ such that for all measures ν on $\mathcal{X}$ with $\nu(F^2) < \infty$

$$N(\varepsilon[\nu(F^2)]^{1/2}, \mathcal{F}, d_\nu^{(2)}) \le K'(v)\varepsilon^{-4(v-1)} \qquad \forall\, 0 < \varepsilon \le 1.$$

PROOF. Let $\mathcal{F}' := \{f^2\, sign(f) : f \in \mathcal{F}\}$; then $\forall f,g \in \mathcal{F}$

$$\begin{aligned} d_\nu^{(2)}(f,g)^2 = \nu(|f-g|^2) &= \int_X (f(x)-g(x))^2\, \nu(dx) \\ &\underset{(4.3.18)}{\le} 2\int_X |f^2(x)\, sign(f(x)) - g^2(x)\, sign(g(x))|\, \nu(dx) \\ &= 2d_\nu^{(1)}\Big(f^2\, sign(f), g^2\, sign(g)\Big), \quad \text{i.e.} \end{aligned}$$

$\forall f,g \in \mathcal{F} \qquad d_\nu^{(2)}(f,g)^2 \le 2d_\nu^{(1)}(f',g')$
with $f' := f^2\, sign(f)$ and $g' := g^2\, sign(g)$, whence, by the definition of covering numbers, one gets
(a) $\quad N(\varepsilon[\nu(F^2)]^{1/2}, \mathcal{F}, d_\nu^{(2)}) \le N(\frac{\varepsilon^2}{2}\nu(F^2), \mathcal{F}', d_\nu^{(1)})$.

Now, also $\mathcal{F}'$ is a VCGC with envelope F^2 and $V(\mathcal{R}') = V(\mathcal{R}) \equiv v$, where $\mathcal{R}' := \{G_{f'} : f' \in \mathcal{F}'\}$:
To see this, let $M' := \{(x_i,t_i), 1 \le i \le v\}$ be an arbitrary subset of $X \times \mathbb{R}$ with $|M'| = v$. Set $M := \{(x_i, |t_i|^{1/2}\, sign(t_i)),\, 1 \le i \le v\}$ then $M \subset X \times \mathbb{R}$ and $|M| = v$. Since $\mathcal{F}$ is a VCGC with $V(\mathcal{R}) = v$, there exists an $N \subset M$, $N = \{(x_i, |t_i|^{1/2}\, sign(t_i)),\, i \in J \subset \{1,...,v\}\}$ such that
(b) $\quad N \ne M \cap G_f \qquad \forall f \in \mathcal{F}$.

But (b) implies that $N' \neq M' \cap G_{f'} \quad \forall f' \in \mathcal{F}$, where $N' := \{(x_i, t_i) : i \in J\}$:

For, suppose to the contrary, that $N' = M' \cap G_{f'}$ for some $f' \in \mathcal{F}'$, $f' = f^2\, sign(f)$, i.e. $\{(x_i, t_i) : i \in J\}$ $= \{(x_i, t_i) : 1 \leq i \leq v\} \cap G_{f'}$; then, since

$$(x_i, t_i) \in G_{f'} \quad \Longleftrightarrow \quad (x_i, |t_i|^{1/2}\, sign(t_i)) \in G_f,$$

we get $\{(x_i, |t_i|^{1/2}\, sign(t_i)),\ i \in J\} = \{(x_i, |t_i|^{1/2}\, sign(t_i)),\ 1 \leq i \leq v\} \cap G_f$, which contradicts (b).

Since M' was arbitrary with $|M'| = v$, $\mathcal{F}'$ is a VCGC with $V(\mathcal{R}') \leq v$. In the same way one shows that $V(\mathcal{R}') < v$ would contradict the minimality of v and therefore $V(\mathcal{R}') = V(\mathcal{R})$.

Thus, by the NOTE following lemma 4.3.17 with $\varepsilon' := \varepsilon^2/2$ and $k := 2(v-1)$ we get

$$N(\varepsilon' \nu(F^2), \mathcal{F}', d_\nu^{(1)}) \leq K(v)(\varepsilon')^{-k} \qquad \forall\, 0 < \varepsilon' \leq \frac{1}{2},$$

i.e. $N(\frac{\varepsilon^2}{2}\nu(F^2), \mathcal{F}', d_\nu^{(1)}) \leq 2^k K(v)\varepsilon^{-2k} \quad \forall\, 0 < \varepsilon \leq 1$, whence by (a)

$$N(\varepsilon[\nu(F^2)]^{1/2}, \mathcal{F}, d_\nu^{(2)}) \leq 2^k K(v)\varepsilon^{-4(v-1)} \quad \forall\, 0 < \varepsilon \leq 1. \qquad \square$$

4.3.19 suggests the following definition (cf. [Al87] and [Va96], Condition (2.5.1), p.127):

4.3.20. Definition.

Let $(X, \mathcal{X})$ be a measurable space, $\mathcal{F}$ a class of $\mathcal{X}$-measurable real-valued functions, and let $\mathcal{M}(X, F)$ be the set of all measures ν on $\mathcal{X}$ with $\nu(F^2) < \infty$, where F is an envelope of $\mathcal{F}$. Then $\mathcal{F}$ is said to have uniformly intrgrable L_2-entropy, if

$$\int_0^\infty (\log N(\tau, \mathcal{F}))^{1/2}\, d\tau < \infty,$$

where $N(\tau, \mathcal{F}) := \sup_{\nu \in \mathcal{M}(X,F)} N(\tau[\nu(F^2)]^{1/2}, \mathcal{F}, d_\nu^{(2)})$.

4.3.21. Remark.

If $\mathcal{F}$ has uniformly intrgrable L_2-entropy, then $(\mathcal{F}, d_\nu^{(2)})$ is totally bounded for each $\nu \in \mathcal{M}(X, F)$.

(In fact, if $\int_0^\infty (\log N(\tau, \mathcal{F}))^{1/2}\, d\tau < \infty$, then for each $\nu \in \mathcal{M}(X, F)$ one has

(+) $\qquad N(\tau[\nu(F^2)]^{1/2}, \mathcal{F}, d_\nu^{(2)}) < \infty \quad$ for $\lambda - a.a.\ \tau \in [0, \infty)$ $\qquad$ ($\lambda \equiv$ Lebesgue measure);

but, since $N(\varepsilon, \mathcal{F}, d_\nu^{(2)})$ is increasing as $\varepsilon \to 0$, (+) must hold for all $\tau \in [0, \infty)$)

Finally, it follows from 4.3.19 that each VCGC $\mathcal{F}$ has uniformly intrgrable L_2-entropy.

5 Some fundamental inequalities

5.1 Symmetrization inequality

Before stating some of the inequalities needed later, the following definition concerning the concept of *independence in case of non-measurable maps*, i.e. of random quantities (rq's) is in order.
Guided by [Du83] and [Du84] (cf. also [Ho85]) we define as in [Zi94], Def. 1.2.1:

5.1.1. Definition.
Let $(\Omega, \mathcal{A}, \mathbb{P}) := (\times_{j\in\mathbb{N}} \Omega_j, \bigotimes_{j\in\mathbb{N}} \mathcal{A}_j, \times_{j\in\mathbb{N}} \mathbb{P}_j)$ *be the countable product of p-spaces* $(\Omega_j, \mathcal{A}_j, \mathbb{P}_j)$, *let* V *be an arbitrary set and* $\eta_j : \Omega \longrightarrow V$ *be rq's of the form* $\eta_j(\omega) = h_j(\omega_j)$ *for* $\omega = (\omega_1, \omega_2, ...) \in \Omega, j \in \mathbb{N}$, *with arbitrary rq's* $h_j : \Omega_j \longrightarrow V$.
Then the sequence $(\eta_j)_{j\in\mathbb{N}}$ *is called* independent *(or,* $\eta_j, j \in \mathbb{N}$, *are said to be* independent rq's*).*
The η_j*'s are said to be* independent and identically distributed (i i d)*, if in addition the p-spaces* $(\Omega_j, \mathcal{A}_j, \mathbb{P}_j)$ *as well as the rq's* h_j *defined on them are identical.*

In the case $h_j = id_{\Omega_j}, j \in \mathbb{N}$, *the sequence* $(\eta_j)_{j\in\mathbb{N}}$ *is said to be* canonically formed.

NOTE: If the h_j's (and so the η_j's) are re's in $V = (V, \mathcal{V})$ (with an appropriate σ-field $\mathcal{V}$) then independence in the sense of 5.1.1 is equivalent with the usual concept of independence of re's.
In the context of Definition 5.1.1 the rq.'s $\eta_j, j \in \mathbb{N}$, are also called to be "independently defined"; cf. [Ho02], pointing out that this implies an unpleasant restriction to the underlying probability space $(\Omega, \mathcal{A}, \mathbb{P})$ which causes an unnecessary restriction for the validity of the inequalities. Instead Hoffmann-Jørgensen introduced in [Ho02] a series of concepts of independence which applies to arbitrary probability spaces, and he studied there the exact form of "independence" for non-measurable functions under which the classical inequalities (Lévy's inequality, Ottaviani's inequality, Jensen's inequality, the symmetrization inequalities, the exponential inequality, the subgaussian inequalities, etc.) hold.

In the following, when dealing with stochastic processes $\eta_j = (\eta_j(t))_{t\in T}$ (with common parameter space T), these processes will be considered as rq's with values in $V = \mathbb{R}^T$ or $V = l^\infty(T)$, respectively, and independence of stochastic processes is to be understood in the sense of 5.1.1.
To avoid measurability questions we will (if not stated otherwise) tacitely assume that the parameter spaces T are countable. (If not, one has to work with the "$\mathbb{E}^*$, $\mathbb{P}^*$-calculus"; see e.g. [Zi94] and [Va96].)
Note that, for countable T, the $||\eta_j|| \equiv ||\eta_j||_T := \sup_{t\in T} |\eta_j(t)|$ are re's in $(\bar{\mathbb{R}}_+, \bar{\mathcal{B}}_+)$ ($\bar{\mathbb{R}}_+ := \mathbb{R}_+ \cup \{\infty\}$ endowed with its Borel σ-field $\bar{\mathcal{B}}_+$), whence also $||S_n||_T, n \in \mathbb{N}$, are re's in $(\bar{\mathbb{R}}_+, \bar{\mathcal{B}}_+)$, where $S_n := \sum_{j\leq n} \eta_j$.
To formulate the Symmetrization inequality for independent stochastic processes $\eta_1, ..., \eta_n, n \in \mathbb{N}$, indexed by a common parameter space T, we need the concept of a so-called *Rademacher sequence* $(\varepsilon_j)_{j\leq n}$, which means that the ε_j's are i i d rv's taking only the values $+1$ or -1 with equal probability, i.e. $\mathcal{L}\{\varepsilon_j\} = \frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1$ ($\delta_x \equiv$ Dirac measure at x).

Then, given (as in 5.1.1) $(\Omega', \mathcal{A}', \mathbb{P}') := (\underset{j\le n}{\times}\Omega'_j, \bigotimes_{j\le n}\mathcal{A}'_j, \underset{j\le n}{\times}\mathbb{P}'_j)$ and stochastic processes $\eta_j : \Omega' \longrightarrow \mathbb{R}^T$ with $\eta_j(\omega') = h_j(\omega'_j)$ for $\omega' = (\omega'_1, ..., \omega'_n) \in \Omega'$, $1 \le j \le n$, where $h_j : \Omega'_j \longrightarrow \mathbb{R}^T$ are stochastic processes (indexed by T), let $(\Omega'', \mathcal{A}'', \mathbb{P}'') := (\{-1,1\}^n, \bigotimes_{j\le n}\mathcal{P}(\{-1,1\}), \underset{j\le n}{\times}\mathcal{L}\{\varepsilon_j\})$ and $(\Omega, \mathcal{A}, \mathbb{P}) := (\Omega' \times \Omega'', \mathcal{A}'\bigotimes\mathcal{A}'', \mathbb{P}' \times \mathbb{P}'')$ (i.e. $(\varepsilon_j)_{j\le n}$ is thought to be canonically formed and independent of $(\eta_j)_{j\le n}$, where $\eta_1, ..., \eta_n$ are independent processes). Denoting with $\mathbb{E}_{\omega'}, \mathbb{E}_{\omega''}$ and $\mathbb{E} \equiv \mathbb{E}_{\omega',\omega''}$ expectation of rv's defined on $(\Omega', \mathcal{A}', \mathbb{P}'), (\Omega'', \mathcal{A}'', \mathbb{P}'')$ and $(\Omega, \mathcal{A}, \mathbb{P})$, respectively, then in this setting the following result holds true (cf. [Va96], Lemma 2.3.6 and [Zi94], Lemma 1.3.2):

5.1.2. Symmetrization inequality for independent processes $\boldsymbol{\eta_1, ..., \eta_n, n} \in \mathbb{N}$.
Suppose that $\mathbb{E}_{\omega'}(|\eta_j(t)|) < \infty$ for all $t \in T$ and $1 \le j \le n$; then, for any convex and nondecreasing function $\psi : \mathbb{R}_+ \longrightarrow \mathbb{R}_+$ (put, as usual $\psi(\infty) := \lim_{a\to\infty}\psi(a)$)

$$\mathbb{E}_{\omega'}\Big(\psi\big(\sup_{t\in T}|\sum_{j\le n}(\eta_j(t) - \mathbb{E}_{\omega'}(\eta_j(t)))|\big)\Big) \;\; \le \;\; \mathbb{E}\Big(\psi(2\sup_{t\in T}|\sum_{j\le n}\varepsilon_j\eta_j(t)|)\Big).$$

PROOF. Let $(\delta_1, ..., \delta_n) \in \{-1,1\}^n$ be arbitrary but fixed; consider the decomposition

$$(+)\;\; \sum_{j\le n}(\eta_j(t) - \mathbb{E}_{\omega'}(\eta_j(t))) = \Big(\sum_{\delta_j=1}\delta_j\eta_j(t) + \sum_{\delta_j=-1}\delta_j\mathbb{E}_{\omega'}(\eta_j(t))\Big) - \Big(\sum_{\delta_j=-1}\delta_j\eta_j(t) + \sum_{\delta_j=1}\delta_j\mathbb{E}_{\omega'}(\eta_j(t))\Big).$$

Since for any $M \subset \{1, ..., n\}$ (with $\mathbb{E}_M$ and $\mathbb{E}_{\complement M}$ denoting expectation of rv's indexed by M and $\complement M$, respectively)

$$\begin{aligned}
&\mathbb{E}_{\omega'}\Big(\psi\big(2\sup_{t\in T}|\sum_{j\in M}\delta_j\eta_j(t) + \sum_{j\in\complement M}\delta_j\mathbb{E}_{\omega'}(\eta_j(t))|\big)\Big) \\
&= \mathbb{E}_M\Big(\psi\big(2\sup_{t\in T}|\mathbb{E}_{\complement M}\big(\sum_{j\in M}\delta_j\eta_j(t) + \sum_{j\in\complement M}\delta_j\eta_j(t)\big)|\big)\Big) \\
&\underset{\substack{(\psi \text{ monotone} \\ \text{nondecreasing})}}{\le} \mathbb{E}_M\Big(\psi\big(\mathbb{E}_{\complement M}(2\sup_{t\in T}|\sum_{j\le n}\delta_j\eta_j(t)|)\big)\Big) \\
&\underset{\substack{(\text{Jensen's} \\ \text{inequality})}}{\le} \mathbb{E}_M\mathbb{E}_{\complement M}\Big(\psi\big(2\sup_{t\in T}|\sum_{j\le n}\delta_j\eta_j(t)|\big)\Big) \\
&\underset{(\text{Fubini})}{=} \mathbb{E}_{\omega'}\Big(\psi\big(2\sup_{t\in T}|\sum_{j\le n}\delta_j\eta_j(t)|\big)\Big),
\end{aligned}$$

we obtain with $M = M_1 := \{j \le n : \delta_j = 1\}$ and $M = M_2 := \{j \le n : \delta_j = -1\}$, respectively, by the ineqality

$$\psi(a+b) \le \frac{1}{2}\psi(2a) + \frac{1}{2}\psi(2b) \qquad \forall\, a, b \in \bar{\mathbb{R}}_+$$

(valid since ψ is convex), that

$$\mathbb{E}_{\omega'}\Big(\psi\big(\sup_{t\in T}|\sum_{j\le n}(\eta_j(t)-\mathbb{E}_{\omega'}(\eta_j(t)))|\big)\Big) \underset{(+)}{\le}$$

$$\mathbb{E}_{\omega'}\Big(\psi\Big(\sup_{t\in T}|\sum_{j\in M_1}\delta_j\eta_j(t)+\sum_{j\in \complement M_1}\delta_j\mathbb{E}_{\omega'}(\eta_j(t))| \;+\; \sup_{t\in T}|\sum_{j\in M_2}\delta_j\eta_j(t)+\sum_{j\in \complement M_2}\delta_j\mathbb{E}_{\omega'}(\eta_j(t))|\Big)\Big) \le$$

$$\frac{1}{2}\mathbb{E}_{\omega'}\Big(\psi\big(2\sup_{t\in T}|\sum_{j\in M_1}\delta_j\eta_j(t)+\sum_{j\in \complement M_1}\delta_j\mathbb{E}_{\omega'}(\eta_j(t))|\big)\Big)+\frac{1}{2}\mathbb{E}_{\omega'}\Big(\psi\big(2\sup_{t\in T}|\sum_{j\in M_2}\delta_j\eta_j(t)+\sum_{j\in \complement M_2}\delta_j\mathbb{E}_{\omega'}(\eta_j(t))|\big)\Big)\le$$

$$\mathbb{E}_{\omega'}\Big(\psi(2\sup_{t\in T}|\sum_{j\le n}\delta_j\eta_j(t)|)\Big).$$

Since $(\delta_1, ..., \delta_n)$ was arbitrary, we get

$$\mathbb{E}_{\omega'}\Big(\psi\big(\sup_{t\in T}|\sum_{j\le n}(\eta_j(t)-\mathbb{E}_{\omega'}(\eta_j(t)))|\big)\Big) \le$$

$$\mathbb{E}_{\omega''}\mathbb{E}_{\omega'}\Big(\psi(2\sup_{t\in T}|\sum_{j\le n}\varepsilon_j\eta_j(t)|)\Big) \underset{\text{(Fubini)}}{=}$$

$$\mathbb{E}(\Big(\psi(2\sup_{t\in T}|\sum_{j\le n}\varepsilon_j\eta_j(t))\Big).$$

□

5.2 Maximal inequality for Rademacher averages

The maximal inequality for Rademacher Averages (see 5.2.3 below) is based on ideas exposited by Pisier [Pi83]. The present proofs are mainly due to Klaus Ziegler [Zi94]. The following lemma is a special case of (3.2) in combination with (3.1) in [Po90]:

5.2.1. Lemma.
Given a Rademacher sequence $\varepsilon_1, ..., \varepsilon_N$ and given a finite and non empty subset M of $\mathbb{R}^N$, there exists for each $1 \le p < \infty$ a universal constant $0 < K_p < \infty$ such that

$$\mathbb{E}^{\frac{1}{p}}\Big(\max_{x\in M}|\sum_{j\le N}\varepsilon_j x_j|^p\Big) \;\le\; K_p(1+\log|M|)^{\frac{1}{2}}\cdot\max_{x\in M}\Big(\sum_{j\le N}x_j^2\Big)^{\frac{1}{2}},$$

where $|M|$ denotes the cardinality of M and $x = (x_1, ..., x_N)$.

Note that for $|M| = 1$ this is just one of Khintchine's inequalities; see Ledoux and Talagrand [Le91], Lemma 4.1, p.91.

For the proof of 5.2.1 the following proposition is used which is but a reformulation of Lemma 1.6 in [Pi83] for the present purposes.

5.2.2. Proposition.
Let $\xi_1, ..., \xi_n$ be arbitrary nonnegative rv's and Φ be a strictly increasing, nonnegative, convex function

defined on $[0,\infty)$ *such that there are constants* $0 < c_i < \infty$, $1 \leq i \leq n$, *and* $0 < c < \infty$ *with* $\mathbb{E}(\Phi(c_i^{-1}\xi_i)) \leq c$ *for all* $1 \leq i \leq n$. *Then*

$$\mathbb{E}(\max_{1\leq i\leq n} \xi_i) \ \leq \ \Phi^{-1}(cn) \cdot \max_{1\leq i\leq n} c_i$$

(Φ^{-1} being the inverse function of Φ).

PROOF.

$$\Phi\Big(\mathbb{E}(\max_{1\leq i\leq n} \xi_i / max_{1\leq i\leq n} c_i)\Big) \ \leq \ \Phi\Big(\mathbb{E}(\max_{1\leq i\leq n} (\xi_i/c_i))\Big) \ \underset{\substack{\text{(Jensen's}\\ \text{inequality)}}}{\leq}$$
$$\mathbb{E}\Big(\Phi(\max_{1\leq i\leq n} (\xi_i/c_i))\Big) \ = \ \mathbb{E}(\max_{1\leq i\leq n} \Phi(\xi_i/c_i)) \ \leq$$
$$\mathbb{E}\Big(\sum_{i\leq n} \Phi(\xi_i/c_i)\Big) \ = \ \sum_{i\leq n} \mathbb{E}(\Phi(\xi_i/c_i)) \ \leq \ cn.$$

Applying Φ^{-1} to both sides yields the assertion. □

PROOF of 5.2.1. We show at first that the *Rademacher average* on the l.h.s of the stated inequality in 5.2.1 can be dominated by a so-called *Gaussian average*, to which 5.2.2 will be applied.

For this, let $g_1, ..., g_N$ be i i d rv's with $\mathcal{L}\{g_i\} = \mathcal{N}(0,1)$ being independent of $\varepsilon_1, ..., \varepsilon_N$. Note that

$$(+) \qquad \mathcal{L}\{(\varepsilon_1|g_1|, ..., \varepsilon_N|g_N|)\} = \mathcal{L}\{(g_1, ..., g_N)\}.$$

Let $\mu := \mathbb{E}(|g_1|)$; then

$$\mathbb{E}\Big(\max_{x\in M} |\sum_{j\leq N} \varepsilon_j x_j|^p\Big) \ =$$
$$\mu^{-p}\mathbb{E}\Big(\max_{x\in M} |\sum_{j\leq N} \varepsilon_j \mathbb{E}(|g_j|) x_j|^p\Big) \ = \ \mu^{-p}\mathbb{E}\Big(\max_{x\in M} |\sum_{j\leq N} \varepsilon_j \mathbb{E}\Big(|g_j| \ \Big|\varepsilon_1, ..., \varepsilon_N\Big) x_j|^p\Big) \ =$$
$$\mu^{-p}\mathbb{E}\Big(\max_{x\in M} |\mathbb{E}\Big(\sum_{j\leq N} \varepsilon_j |g_j| x_j \ \Big|\varepsilon_1, ..., \varepsilon_N\Big)|^p\Big) \ \underset{\substack{\text{(Jensen's}\\ \text{inequality)}}}{\leq} \ \mu^{-p}\mathbb{E}\Big(\max_{x\in M} \mathbb{E}\Big(|\sum_{j\leq N} \varepsilon_j |g_j| x_j|^p \ \Big|\varepsilon_1, ..., \varepsilon_N\Big)\Big) \ \leq$$
$$\mu^{-p}\mathbb{E}\Big(\mathbb{E}\Big(\max_{x\in M} |\sum_{j\leq N} \varepsilon_j |g_j| x_j|^p \ \Big|\varepsilon_1, ..., \varepsilon_N\Big)\Big) \ =$$
$$\mu^{-p}\mathbb{E}\Big(\max_{x\in M} |\sum_{j\leq N} \varepsilon_j |g_j| x_j|^p\Big) \ \underset{(+)}{=} \ \mu^{-p}\mathbb{E}(\max_{x\in M} |\sum_{j\leq N} g_j x_j|^p).$$

Now, let $\Phi'(u) := \exp(u^{2/p})$ for $u \in [0,\infty)$ and

$$\Phi(u) := \begin{cases} 1 + \frac{\Phi'(u_p)-1}{u_p} & , \text{ for } 0 \leq u \leq u_p \\ \Phi'(u) & , \text{ for } u > u_p \end{cases} ,$$

where $u_p := \left(\frac{p}{2}\right)^{p/2}$; then

($\star$) $\Phi : \mathbb{R}_+ \longrightarrow \mathbb{R}_+$ is a strictly increasing, nonnegative, convex function with $\Phi \le \Phi'$, and

($\star\star$) $\forall v \in \mathbb{R}_+ \quad \Phi^{-1}(v) \le u_p + (\log v)^{p/2}$.

On the other hand, assigning to each $x = (x_1, ..., x_N) \in M$ the rv $\xi_x := \sum_{j\le N} g_j x_j$, we have for each $x \in M$ that $\mathcal{L}\{\xi_x\} = \mathcal{N}(0, c_x^2)$ with $c_x^2 := (\sum_{j\le N} x_j^2)$, and

$$\begin{aligned}
&\mathbb{E}\Big(\Phi(|\xi_x|^p/2^p c_x^p)\Big) \underset{(\Phi\le\Phi')}{\le} \\
&\mathbb{E}\Big(\Phi'(|\xi_x|^p/2^p c_x^p)\Big) \underset{\text{(by def. of } \Phi')}{=} \mathbb{E}\Big(\exp(\xi_x^2/4c_x^2)\Big) \underset{(\mathcal{L}\{\xi_x/c_x\}=\mathcal{N}(0,1))}{=} \\
&(2\pi)^{-1/2}\int_{\mathbb{R}} e^{u^2/4}\, e^{-u^2/2}\, du \;=\; (2\pi)^{-1/2}\int_{\mathbb{R}} e^{-u^2/4}\, du \underset{(u:=\sqrt{2}v)}{=} \\
&\sqrt{2}\,(2\pi)^{-1/2}\int_{\mathbb{R}} e^{-v^2/2}\, dv \;=\; \sqrt{2}.
\end{aligned}$$

Thus 5.2.2 can be applied (cf. ($\star$)) with $|\xi_x|^p, x \in M$, instead of $\xi_i, 1 \le i \le n$, and with $2^p c_x^p, x \in M$, instead of $c_i, 1 \le i \le n$, where $n = |M|$ and $c := \sqrt{2}$.

Hence by 5.2.2 it follows that

$$\begin{aligned}
&\mathbb{E}\Big(\max_{x\in M} |\sum_{j\le N} g_j x_j|^p\Big) \;= \\
&\mathbb{E}\Big(\max_{x\in M} |\xi_x|^p\Big) \;\le\; \Phi^{-1}(\sqrt{2}|M|)\max_{x\in M}(2^p c_x^p) \underset{(\star\star)}{\le} \\
&\Big(u_p + (\log(\sqrt{2}|M|))^{p/2}\Big)\cdot \max_{x\in M}(2^p c_x^p) \underset{\text{(by def. of } c_x)}{=} \\
&2^p u_p \max_{x\in M}(\sum_{j\le N} x_j^2)^{p/2} \;+\; 2^p\Big(\log\sqrt{2} + \log|M|\Big)^{p/2} \max_{x\in M}(\sum_{j\le N} x_j^2)^{p/2} \;\le \\
&\Big(2^p u_p + 2^p(1+\log|M|)^{p/2}\Big)\cdot \max_{x\in M}(\sum_{j\le N} x_j^2)^{p/2} \;\le \\
&\bar{K}_p^p(1+\log|M|)^{p/2}\cdot \max_{x\in M}(\sum_{j\le N} x_j^2)^{p/2} \qquad \text{with } \bar{K}_p^p := 2^p(u_p+1).
\end{aligned}$$

Since $\mathbb{E}^{1/p}\Big(\max_{x\in M} |\sum_{j\le N} \varepsilon_j x_j|^p\Big) \le \mu^{-1}\mathbb{E}^{1/p}\Big(\max_{x\in M} |\sum_{j\le N} g_j x_j|^p\Big)$, as shown above, the assertion of 5.2.1 holds true with $K_p := \mu^{-1}\bar{K}_p$. □

The following maximal inequality will be an essential tool for proving a ULLN for RMP's in Section 6.1 below. As we shall see, it is an easily to be shown consequence of 5.2.1:

5.2.3. Maximal Inequality for Rademacher Averages.
Given any $x_j \in \mathbb{R}^T, 1 \le j \le N, N \in \mathbb{N}$, *let*

$$d_1(s,t) := \sum_{j\le N} |x_j(s) - x_j(t)|, \quad s,t \in T.$$

Then, for each $1 \le p < \infty$, there exists a universal constant $0 < K_p < \infty$ such that for any Rademacher sequence $\varepsilon_1, ..., \varepsilon_N$ and for all $\gamma > 0$

$$\mathbb{E}^{1/p}\Big(\sup_{t\in T}|\sum_{j\le N}\varepsilon_j x_j(t)|^p\Big) \le$$
$$\gamma + K_p(1 + H(\gamma, T, d_1))^{1/2} \cdot \sup_{t\in T}\Big(\sum_{j\le N} x_j^2(t)\Big)^{1/2}$$

(where $H(\,\cdot\,, T, d_1)$ denotes the metric entropy of $T = (T, d_1)$ as defined in 4.1.1).

PROOF. Given any $\gamma > 0$, assume w.l.o.g. that $N(\gamma, T, d_1) < \infty$; then (cf. the definition of $N(\,\cdot\,, T, d_1)$ in 4.1.1) there exists a subset T' of T with $|T'| = N(\gamma, T, d_1)$ such that for each $t \in T$ there exists a $u(t) \in T'$ with $d_1(t, u(t)) \le \gamma$.

Then we get

$$\mathbb{E}^{1/p}\Big(\sup_{t\in T}|\sum_{j\le N}\varepsilon_j x_j(t)|^p\Big) \underset{\text{(Minkowski's Ineq.)}}{\le}$$
$$\mathbb{E}^{1/p}\Big(\sup_{t\in T}|\sum_{j\le N}\varepsilon_j (x_j(t) - x_j(u(t)))|^p\Big) + \mathbb{E}^{1/p}\Big(\sup_{t\in T'}|\sum_{j\le N}\varepsilon_j x_j(t)|^p\Big) \underset{5.2.1}{\le}$$
$$\sup_{t\in T}|\sum_{j\le N}|x_j(t) - x_j(u(t))| + K_p(1 + \log|T'|)^{1/2}\sup_{t\in T}\Big(\sum_{j\le N}x_j^2(t)\Big)^{1/2} \le$$
$$\gamma + K_p(1 + H(\gamma, T, d_1))^{1/2} \cdot \sup_{t\in T}\Big(\sum_{j\le N} x_j^2(t)\Big)^{1/2}.$$

□

5.3 Hoffmann-Jørgensen Inequality

To our knowledge, the Hoffmann-Jørgensen Inequality was originally proved implicitely in [Ho74], Theorem 3.1, for sums of independent and symmetric Banachspace-valued re's (cf. [Le91], Section 6.2).

Here we will consider as before independent (in the sense of 5.1.1) stochastic processes $\eta_j = (\eta_j(t))_{t\in T}$, $j \in \mathbb{N}$, indexed by an arbitrary parameter space T (supposed to be countable for simplicity to avoid measurability considerations). The η_j's will be viewed as rq's with values in $\mathbb{R}^T$ or $l^\infty(T)$, respectively, and $||\eta||$ or $||\sum_{j\le n}\varepsilon_j\eta_j||$ denotes the $\sup_{t\in T}|\eta(t)|$ or $\sup_{t\in T}|\sum_{j\le n}\varepsilon_j\eta_j|$, respectively.

5.3.1. Hoffmann-Jørgensen Inequality
(Cf. [Va96], A.1.5 and [Zi94], Corollary 2.1.3).
Let $\eta_j = (\eta_j(t))_{t\in T}, j \in \mathbb{N}$, be a sequence of independent stochastic processes with common parameter space T and $(\varepsilon_j)_{j\in\mathbb{N}}$ be a canonically formed Rademacher sequence which is independent of $(\eta_j)_{j\in\mathbb{N}}$

(cf. Section 5.1). Let $\Psi : \mathbb{R}_+ \longrightarrow \mathbb{R}_+$ be a nondecreasing function which is absolutely continuous on each interval $[0, a], a > 0$, and which satisfies the so-called Orlicz condition

$$\Psi(2x) \leq C\Psi(x) \qquad \textit{for all } x \in \mathbb{R}_+ \quad \textit{and some } 1 \leq C < \infty$$

(think e.g. of $\Psi(x) = x^p$); then for each $n \in \mathbb{N}$ (as before, we put $\Psi(\infty) := \lim_{a\to\infty} \Psi(a)$)

$$\mathbb{E}\Big(\Psi(||\sum_{j\leq n} \varepsilon_j \eta_j||)\Big) \;\leq\; 2C^2\mathbb{E}\Big(\max_{j\leq n} \Psi(||\eta_j||)\Big) + 2C^2\Psi(s_n)$$

with $s_n := \inf\{s > 0 : \mathbb{P}(||\sum_{j\leq n} \varepsilon_j\eta_j|| > s) \leq (4C^2)^{-1}\}$.

NOTE: This inequality will be an essential tool in proving a uniform law of large numbers (ULLN) for Random Measure Processes in the following Section 6.1. It will be applied there with $\Psi(x) := x^p, 1 \leq p < \infty$. In such a case one can infer L_p-convergence of $\sum_{j\leq n} \varepsilon_j\eta_{nj}$ to zero from its $\mathbb{P}$-stochastic convergence to zero, provided that the η_{nj}'s are asymptotically negligible in the sense that $\lim_{n\to\infty} \mathbb{E}(\max_{j\leq n} ||\eta_{nj}||^p) = 0$, where the latter is e.g. fulfilled, if the $||\eta_{nj}||$'s are bounded by some δ_n with $\lim_{n\to\infty} \delta_n = 0$.

5.4 Further Symmetrization inequalities

(cf. [Va96], Section 2.3 and [Due00], Section 4)

We will consider again stochastic processes $S = (S(t))_{t\in T}$, indexed by T, assuming for simplicity that the parameter space T is countable, whence

$$||S|| \equiv ||S||_T := \sup_{t\in T} |S(t)|$$

will be measurable.
If S is defined on a basic p-space $(\Omega, \mathcal{A}, \mathbb{P})$, $\;S : \Omega \longrightarrow \mathbb{R}^T$, one can enlarge the basic p-space to the product space $(\Omega\times\Omega, \mathcal{A}\bigotimes\mathcal{A}, \mathbb{P}\times\mathbb{P})$ to obtain an *independent copy S' of S* by defining $S'(t)(\omega_1, \omega_2) := S(t)(\omega_2)$, where now also S is considered as defined on the product space, i.e. $S(t)(\omega_1, \omega_2) := S(t)(\omega_1)$.

5.4.1. Lemma.
Let S and S' be independent stochastic processes with common parameter space T both defined on a p-space $(\Omega, \mathcal{A}, \mathbb{P})$ (where S' need not necessarily be an independent copy of S);

i) Assume that there exist constants $\delta, \beta > 0$ such that for all $t \in T$

$$\mathbb{P}(|S'(t)| \leq \delta) \geq \beta,$$

then for arbitrary $\eta > 0$

$$\mathbb{P}(||S|| > \eta) \leq \beta^{-1}\, \mathbb{P}(||S - S'|| > \eta - \delta).$$

ii) Assume $\mathbb{E}(S') \equiv \Big(\mathbb{E}(S'(t))\Big)_{t\in T} = 0 \in \mathbb{R}^T$*. Then*

$$\mathbb{E}(\psi(||S||)) \leq \mathbb{E}\Big(\psi(||S - S'||)\Big)$$

for any convex and monotone increasing function $\psi : \mathbb{R}_+ \longrightarrow \bar{\mathbb{R}}_+ \, (:= [0, \infty])$ *(with* $\psi(\infty) := \lim_{a\to\infty} \psi(a)$*).*

PROOF. *As to i):* W.l.o.g. we may and do assume that $T \subset \mathbb{N}$; let

$$\tau \equiv \tau(S) := \begin{cases} \min\{t \in T : |S(t)| > \eta\}, & \text{if } ||S|| > \eta \\ \min\{T\} & \text{otherwise} \end{cases} .$$

Then τ is a rv such that

$$||S|| > \eta \quad \Longleftrightarrow \quad |S(\tau)| > \eta$$

(where $S(\tau)(\omega) := S(\tau(\omega))(\omega) \quad \forall\, \omega \in \Omega$).
Now,

$$\mathbb{P}(||S - S'|| > \eta - \delta) \;\geq\; \mathbb{P}(|S(\tau) - S'(\tau)| > \eta - \delta) \;\geq\; \mathbb{P}\Big(|S(\tau)| > \eta \text{ and } |S'(\tau)| \leq \delta\Big)$$

(note that $|S(\tau)| > \eta$ together with $|S'(\tau)| \leq \delta$ implies $|S(\tau) - S'(\tau)| > \eta - \delta$).
Now,

$$\begin{aligned} \mathbb{P}\Big(|S(\tau)| > \eta \text{ and } |S'(\tau)| \leq \delta\Big) &= \mathbb{P}\Big(\bigcup_{t\in T} \{\tau = t\} \cap \{|S(\tau)| > \eta\} \cap \{|S'(\tau)| \leq \delta\}\Big) \\ &= \sum_{t\in T} \mathbb{P}\Big(\{\tau = t\} \cap \{|S(\tau)| > \eta\} \cap \{|S'(t)| \leq \delta\}\Big) \\ &= \sum_{t\in T} \mathbb{P}\Big(\{\tau = t\} \cap \{|S(\tau)| > \eta\}\Big) \cdot \mathbb{P}(|S'(t)| \leq \delta\}) \\ &\geq\; \mathbb{P}(|S(\tau)| > \eta) \cdot \beta = \mathbb{P}(||S|| > \eta) \cdot \beta \end{aligned}$$

which proves i).
As to ii): Extending ψ from $\mathbb{R}_+$ to $\mathbb{R}$ by defining $\psi(-r) := \psi(r)$, we can handle ψ as an even convex function $\psi : \mathbb{R} \longrightarrow \bar{\mathbb{R}}_+$, whence

$$\begin{aligned} \mathbb{E}(\psi(||S||)) &= \mathbb{E}(\psi(\sup_{t\in T} |S(t)|)) \underset{\psi:\mathbb{R}\longrightarrow\bar{\mathbb{R}}_+ \text{ mon. increasing}}{=} \mathbb{E}(\sup_{t\in T} \psi(|S(t)|)) \\ &= \mathbb{E}\Big(\sup_{t\in T} \psi(S(t))\Big) \;=\; \mathbb{E}\Big(\sup_{t\in T} \psi(\underbrace{\mathbb{E}(S(t) - S'(t)|S)}_{=\underbrace{\mathbb{E}(S(t)|S)}_{=S(t)} - \underbrace{\mathbb{E}(S'(t)|S)}_{=\mathbb{E}(S'(t))=0}})\Big) \\ &\underset{\substack{\text{Jensen's inequality for conditional} \\ \text{expectations; cf. [Gae77], 5.4.7}}}{\leq} \mathbb{E}\Big(\sup_{t\in T} \mathbb{E}(\psi(S(t) - S'(t))|S)\Big) \end{aligned}$$

$$
\begin{aligned}
&\leq \mathbb{E}\Big(\mathbb{E}\Big(\sup_{t\in T}\psi(S(t)-S'(t))|S\Big)\Big)\\
&= \mathbb{E}\Big(\sup_{t\in T}\psi(S(t)-S'(t))\Big)\\
&\underset{\text{cf. above}}{=} \mathbb{E}\Big(\sup_{t\in T}\psi(|S(t)-S'(t)|)\Big)\\
&\underset{\psi:\mathbb{R}\longrightarrow\bar{\mathbb{R}}_+ \text{ mon. increasing}}{=} \mathbb{E}\Big(\psi(\sup_{t\in T}|S(t)-S'(t)|)\Big)\\
&= \mathbb{E}\Big(\psi(||S-S'||)\Big).
\end{aligned}
$$

□

In view of RMP's with index set $T = \mathcal{F}$ we may consider the even more general model of processes $S_n = (S_n(t))_{t\in T}$ indexed by an arbitrary parameter space T (again supposed to be countable to avoid measurability considerations) given by

$$S_n(t) := \sum_{j\leq j(n)} \eta_{nj}(t),$$

where $\eta_{nj} = (\eta_{nj}(t))_{t\in T}, 1 \leq j \leq j(n), n \in \mathbb{N}$, is a triangular array of rowwise independent stochastic processes (indexed by T).
Now, given in addition a Rademacher sequence $(\varepsilon_j)_{j\in\mathbb{N}}$, assume that for each $n \in \mathbb{N}$

$$\eta_{n1}, ..., \eta_{nj(n)}, \eta'_{n1}, ..., \eta'_{nj(n)}, \varepsilon_1, ..., \varepsilon_{j(n)}$$

are independent and such that

$$\mathcal{L}\{\eta'_{nj}\} = \mathcal{L}\{\eta_{nj}\}, \quad 1 \leq j \leq j(n), n \in \mathbb{N},$$

(i.e. the $\eta'_{nj} = (\eta'_{nj}(t))_{t\in T}$ being independent versions of the processes η_{nj}).
Then, Lemma 5.4.1, applied to $S = S_n$ (or $S = S_n - \mathbb{E}(S_n)$) and $S' = S'_n := \sum\limits_{j\leq j(n)} \eta'_{nj}$ (or $S' = S'_n - \mathbb{E}(S'_n) = S'_n - \mathbb{E}(S_n)$) leads to consider the process

$$S_n - S'_n = \sum_{j\leq j(n)} (\eta_{nj} - \eta'_{nj}).$$

Now, by the assumed independence of $\eta_{n1}, ..., \eta_{nj(n)}, \eta'_{n1}, ..., \eta'_{nj(n)}, n \in \mathbb{N}$, and since $\mathcal{L}\{\eta'_{nj}\} = \mathcal{L}\{\eta_{nj}\}$, $1 \leq j \leq j(n), n \in \mathbb{N}$, we get (where $\xi \overset{\mathcal{L}}{=} \tilde{\xi}$ means that $\mathcal{L}\{\xi\} = \mathcal{L}\{\tilde{\xi}\}$) that

$$\eta_{nj} - \eta'_{nj} \overset{\mathcal{L}}{=} \varepsilon_j(\eta_{nj} - \eta'_{nj})$$

for each $1 \leq j \leq j(n), n \in \mathbb{N}$. Therefore

$$S_n - S'_n \overset{\mathcal{L}}{=} \sum_{j\leq j(n)} \varepsilon_j(\eta_{nj} - \eta'_{nj}) \;=\; S^0_n - \tilde{S}^0_n,$$

where $S_n^0 := \sum\limits_{j\le j(n)} \varepsilon_j \eta_{nj}$ and $\tilde{S}_n^0 := \sum\limits_{j\le j(n)} \varepsilon_j \eta'_{nj}$.

Note that the processes S_n^0 and $\tilde{S}_n^0$ are not independent; they are identically distributed.
We thus obtain the following symmetrization inequalities in continuation of Lemma 5.4.1:

5.4.2. Lemma.
i) For arbitrary $\eta > 0$

$$\mathbb{P}(||S_n - S'_n|| > \eta) \le 2\,\mathbb{P}\Big(||S_n^0|| > \frac{\eta}{2}\Big)$$

ii) For any convex and monotone increasing function $\psi : \mathbb{R} \longrightarrow \bar{\mathbb{R}}_+$ (with $\psi(\infty) := \lim\limits_{a\to\infty} \psi(a)$)

$$\mathbb{E}\Big(\psi(||S_n - S'_n||)\Big) \le \mathbb{E}(\psi(2||S_n^0||)).$$

PROOF. *As to i):* Since

$$||S - S'|| \stackrel{\mathcal{L}}{=} ||S_n^0 - \tilde{S}_n^0|| \le ||S_n^0|| + ||\tilde{S}_n^0||,$$

we get for arbitrary $\eta > 0$ that

$$\mathbb{P}(||S_n - S'_n|| > \eta) \le \mathbb{P}\Big(||S_n^0|| > \frac{\eta}{2}\Big) + \mathbb{P}\Big(||\tilde{S}_n^0|| > \frac{\eta}{2}\Big) \underset{\mathcal{L}\{S_n^0\}=\mathcal{L}\{\tilde{S}_n^0\}}{\le} 2\,\mathbb{P}\Big(||S_n^0|| > \frac{\eta}{2}\Big).$$

As to ii):

$$\begin{aligned}\mathbb{E}\Big(\psi(||S_n - S'_n||)\Big) &= \mathbb{E}\Big(\psi(||S_n^0 - \tilde{S}_n^0||)\Big)\\ &\underset{\psi\text{ mon. increasing}}{\le} \mathbb{E}\Big(\psi\Big(\frac{1}{2}(2||S_n^0|| + 2||\tilde{S}_n^0||)\Big)\Big)\\ &\underset{\psi\text{ convex}}{\le} \frac{1}{2}\,\mathbb{E}(\psi(2||S_n^0||)) + \frac{1}{2}\,\mathbb{E}(\psi(2||\tilde{S}_n^0||))\\ &= \mathbb{E}(\psi(2||S_n^0||)).\end{aligned}$$

□

IMPORTANT NOTE:
According to 5.4.1 ii) and 5.4.2 ii), to obtain a tractable upper bound for $\mathbb{E}(\psi(||S_n||)$ one can seek for an upper bound for

$$\mathbb{E}(\psi(2||S_n^0||)) = \mathbb{E}\Big(\mathbb{E}(\psi(2\,||\sum_{j\le j(n)} \varepsilon_j \eta_{nj}||)|(\eta_{nj}))\Big),$$

which entails (by the assumed independence on (ε_j) and (η_{nj})) to seek for an upper bound for

(5.4.3) $$\mathbb{E}\Big(\psi(2\,||\sum_{j\le j(n)} \varepsilon_j \eta_{nj}||)\Big)$$

with fixed $\eta_{nj} = (\eta_{nj})_{t\in T} \in \mathbb{R}^T$; (cf. with 5.2.3).

If, in addition, the processes $\eta_{n1}, ..., \eta_{nj(n)}$ are identically distributed for each $n \in \mathbb{N}$ (whence also $\eta'_{n1}, ..., \eta'_{nj(n)}$ are identically distributed for each $n \in \mathbb{N}$) one gets the following result:

5.4.4. Lemma.

Assume that for each $n \in \mathbb{N}$ *the processes* $\eta_{n1}, ..., \eta_{nj(n)}$ *are identically distributed and that with* $S_n := \sum_{j \leq j(n)} \eta_{nj}$ $\mathbb{E}(\eta_{n1}) = j(n)^{-1}\mathbb{E}(S_n)$ *exists in* $\mathbb{R}^T$.

Then, for any convex and monotone increasing function $\psi : \mathbb{R}_+ \longrightarrow \bar{\mathbb{R}}_+$ *(with* $\psi(\infty) := \lim_{a \to \infty} \psi(a)$*) and any* $0 < \lambda < 1$

$$\mathbb{E}\Big(\psi(\frac{1}{2}||S_n - \mathbb{E}(S_n)||)\Big) \leq \mathbb{E}(\psi(||S_n^0||))$$
$$\leq \lambda\,\mathbb{E}\Big(\psi(\frac{2}{\lambda}\,||S_n - \mathbb{E}(S_n)||)\Big) + (1-\lambda)\,\mathbb{E}\Big(\psi\Big(\frac{1}{1-\lambda}\,j(n)^{-1}\Big|\sum_{j \leq j(n)} \varepsilon_j\Big| \cdot ||\mathbb{E}(S_n)||\Big)\Big).$$

PROOF. The first inequality follows from 5.4.1 ii) together with 5.4.2 ii): In fact,

$$\mathbb{E}\Big(\psi(||S_n - \mathbb{E}(S_n)||)\Big) \underset{\substack{5.4.1\text{ ii) with } S = S_n - \mathbb{E}(S_n) \\ \text{and } S' = S'_n - \mathbb{E}(S'_n)}}{\leq} \mathbb{E}\Big(\psi(||S_n - \mathbb{E}(S_n) - (S'_n - \underbrace{\mathbb{E}(S'_n)}_{=\mathbb{E}(S_n)})||)\Big)$$
$$= \mathbb{E}\Big(\psi(||S_n - S'_n||)\Big) \underset{5.4.2ii)}{\leq} \mathbb{E}(\psi(2\,||S_n^0||))$$

which yields the first inequality (replacing ψ by $\tilde{\psi}$ with $\tilde{\psi}(x) := \psi(\frac{1}{2}x)$), being also valid if the η_{nj}'s are independent but not necessarily identically distributed. To prove the second inequality, remember that for convex ψ and any $\lambda_i > 0$ with $\sum \lambda_i = 1$

$$(*) \qquad \psi(\sum \lambda_i x_i) \leq \sum \lambda_i \psi(x_i).$$

Now, let $J_1 := \{j \leq j(n) : \varepsilon_j = 1\}$ and $J_2 := \{j \leq j(n) : \varepsilon_j = -1\} = \{1, ..., j(n)\} \backslash J_1$, where $n \in \mathbb{N}$ is arbitrary, but fixed. Then

$$S_n^0 = \sum_{j \leq j(n)} \varepsilon_j(\eta_{nj} - \mathbb{E}(\eta_{n1})) + \sum_{j \leq j(n)} \varepsilon_j \cdot \underbrace{\mathbb{E}(\eta_{n1})}_{=j(n)^{-1}\mathbb{E}(S_n),\ \textit{since the } \eta_{nj}\textit{'s are id}}$$
$$= \sum_{j \in J_1} (\eta_{nj} - \mathbb{E}(\eta_{n1})) - \sum_{j \in J_2} (\eta_{nj} - \mathbb{E}(\eta_{n1})) + E_n \cdot \mathbb{E}(S_n)$$

where $E_n := j(n)^{-1} \sum_{j \leq j(n)} \varepsilon_j$, i.e. $S_n^0 = S_{(J_1)} - S_{(J_2)} + E_n \cdot \mathbb{E}(S_n)$, where

$$S_{(M)} := \sum_{j \in M} (\eta_{nj} - \mathbb{E}(\eta_{n1})) \qquad \text{for } M \subset \{1, ..., j(n)\}.$$

Since $S_n - \mathbb{E}(S_n) = S_{(J_1)} + S_{(J_2)}$ with $S_{(J_1)}$ and $S_{(J_2)}$ being independent and centered, **given $\boldsymbol{J_1}$**, we can apply Lemma 5.4.1 ii) **w.r.t. the conditional distribution** of $(S, S') := (S_{(J_1)}, -S_{(J_2)})$ and

$(S, S') := (S_{(J_2)}, -S_{(J_1)})$ respectively, **given J_1**, to obtain that

$$\begin{aligned}
\mathbb{E}(\psi(||S_n^0||)) &= \mathbb{E}(\psi(||S_{(J_1)} - S_{(J_2)} + E_n \cdot \mathbb{E}(S_n)||)) \\
&\leq \mathbb{E}(\psi(||S_{(J_1)}|| + ||S_{(J_2)}|| + |E_n|\, ||\mathbb{E}(S_n)||)) \\
&= \mathbb{E}(\psi(\frac{\lambda}{2}\frac{2}{\lambda}||S_{(J_1)}|| + \frac{\lambda}{2}\frac{2}{\lambda}||S_{(J_2)}|| + (1-\lambda)\frac{|E_n|}{1-\lambda}\, ||\mathbb{E}(S_n)||)) \\
&\underset{(*)}{\leq} \frac{\lambda}{2}\mathbb{E}(\psi(\frac{2}{\lambda}||S_{(J_1)}||)) + \frac{\lambda}{2}\mathbb{E}(\psi(\frac{2}{\lambda}||S_{(J_2)}||)) + (1-\lambda)\mathbb{E}(\psi(\frac{|E_n|}{1-\lambda}\, ||\mathbb{E}(S_n)||)) \\
&\underset{5.4.1ii)}{\leq} \frac{\lambda}{2}\mathbb{E}(\psi(\frac{2}{\lambda}||S_{(J_1)} + S_{(J_2)}||)) + \frac{\lambda}{2}\mathbb{E}(\psi(\frac{2}{\lambda}||S_{(J_2)} + S_{(J_1)}||)) + (1-\lambda)\mathbb{E}(\psi(\frac{|E_n|}{1-\lambda}\, ||\mathbb{E}(S_n)||)) \\
&= \lambda\, \mathbb{E}(\psi(\frac{2}{\lambda}||S_n - \mathbb{E}(S_n)||)) + (1-\lambda)\mathbb{E}(\psi(\frac{|E_n|}{1-\lambda}\, ||\mathbb{E}(S_n)||)) \qquad \square
\end{aligned}$$

Applying 5.4.4 with $\psi = id_{\mathbb{R}_+}$ yields

5.4.5. Corollary.
Under the assumptions of Lemma 5.4.4 one has

$$\begin{aligned}
\mathbb{E}\Big(\frac{1}{2}||S_n - \mathbb{E}(S_n)||\Big) &\leq \mathbb{E}(||S_n^0||) \\
&\leq 2\mathbb{E}(||S_n - \mathbb{E}(S_n)||) + \mathbb{E}\Big(\big|j(n)^{-1}\sum_{j \leq j(n)} \varepsilon_j\big|\, ||\mathbb{E}(S_n)||\Big) \\
&\leq 2\mathbb{E}(||S_n - \mathbb{E}(S_n)||) + (j(n)^{-1/2}||\mathbb{E}(S_n)||.
\end{aligned}$$

Thus, if $||\mathbb{E}(S_n)|| = o(\sqrt{j(n)})$, *then*

$$\mathbb{E}(||S_n - \mathbb{E}(S_n)||) = o(1) \quad \iff \quad \mathbb{E}(||S_n^0||) = o(1).$$

Shorack and Wellner claimed (cf. [Sh86]) that good inequalities are the key to strong theorems.
So, let us conclude this section presenting some Exponential Inequalities; cf. [Va96], Sections 2.2.1 and 2.2.2, [Po84], Appendix B, and [Due00], Section 6.

5.5 Exponential inequalities

We want to follow mainly [Po84], pp. 191:

Let ξ be a rv defined on a p-space $(\Omega, \mathcal{A}, \mathbb{P})$ with $\mathcal{L}\{\xi\} = \mathcal{N}(0,1)$; then it is known that for all $x > 0$

$$\Big(\frac{1}{x} - \frac{1}{x^3}\Big)\frac{\exp(-\frac{1}{2}x^2)}{\sqrt{2\pi}} \leq \mathbb{P}(\xi \geq x) \leq \frac{1}{x}\frac{\exp(-\frac{1}{2}x^2)}{\sqrt{2\pi}},$$

i.e. the **tail probabilities** $\mathbb{P}(\xi \geq x), x > 0$, are governed by the factor $\exp(-\frac{1}{2}x^2)$.
Now, in view of the classical CLT a similar upper bound for the tail probabilities of a sum of i i d rv's

ξ_i with zero mean and variance 1 should be in order:

Set $S := \xi_1 + ... + \xi_n$; then for $x > 0$ and each $t > 0$

$$\begin{aligned}
\text{(5.5.1)} \qquad \mathbb{P}(S \geq x) &= \mathbb{P}(S - x \geq 0) = \mathbb{P}(tS - tx \geq 0) = \mathbb{P}(\exp(tS - tx) \geq 1) \\
&\underset{\text{(Markov's ineq.)}}{\leq} \mathbb{E}(\exp(tS - tx)) = \exp(-xt)\mathbb{E}(\exp(tS)) \\
&\underset{(\xi_i \text{ indep.})}{=} \exp(-xt) \prod_{i=1}^{n} \mathbb{E}(\exp(t\xi_i))
\end{aligned}$$

The trick will be to find a $t > 0$ that makes the last product small. For $\xi \overset{\mathcal{L}}{=} \mathcal{N}(0,1)$ it is easy to find the best t directly:

For all $t > 0$ one has $\forall\, x > 0$ (as before)

$$\mathbb{P}(\xi \geq x) \leq \mathbb{E}(\exp(t\xi - tx)) = \exp(-xt)\mathbb{E}(\exp(t\xi)) = \exp(-xt)\exp(t^2/2), \quad \text{i.e.}$$

$$\mathbb{P}(\xi \geq 0) \leq \inf_{t>0}[\exp(-xt)\exp(t^2/2)] = \exp(-\frac{1}{2}x^2) \quad \forall\, x > 0.$$

For other than standard normal distributions one has to work harder. One must maneuver (according to (5.5.1) the **moment generating function** $\mathbb{E}(\exp(t\xi_i)), t > 0$, of ξ_i into a tractable form that gives us some clue about which value of t to choose.

5.5.2. Hoeffding's Inequality.
(cf. [Hoe63])
Let $\xi_1, ..., \xi_n$ be independent rv's (defined on $(\Omega, \mathcal{A}, \mathbb{P})$) with zero means and bounded ranges: $a_i \leq \xi_i \leq b_i$ with constants $a_i < 0 < b_i$. Then, for each $x > 0$

$$\mathbb{P}(\sum_{i=1}^{n} \xi_i \geq x) \leq \exp(-2x^2 / \sum_{i=1}^{n} (b_i - a_i)^2).$$

PROOF. According to (5.5.1) we have to bound the moment generating function of ξ_i. Drop the subscript i temporarily. For this, note that with $a \leq \xi \leq b$

$$\exp(t\xi) = \exp\big(\frac{b-\xi}{b-a}ta + \frac{\xi-a}{b-a}tb\big) \underset{(\exp(\cdot)\text{ convex})}{\leq} \frac{b-\xi}{b-a}\exp(ta) + \frac{\xi-a}{b-a}\exp(tb),$$

whence

$$\mathbb{E}(\exp(t\xi)) \underset{(\mathbb{E}(\xi)=0)}{\leq} \frac{b}{b-a}\exp(ta) - \frac{a}{b-a}\exp(tb) \overset{(!)}{=} e^{-u\lambda}\big((1-\lambda) + \lambda e^u\big),$$

where $\lambda := \frac{-a}{b-a} \in (0,1)$ and $u := t(b-a) > 0$.

(*As to (!)*: $u\lambda = \frac{t(b-a)}{b-a} \cdot (-a) = -at$, $1 - \lambda = 1 + \frac{a}{b-a} = \frac{b}{b-a}$ and $\lambda e^u = -\frac{a}{b-a}e^{t(b-a)} = -\frac{a}{b-a}(e^{tb} \cdot e^{-at})$, whence

$$\begin{aligned}
\frac{b}{b-a}e^{ta} - \frac{a}{b-a}e^{tb} &= (1-\lambda)e^{ta} + \lambda e^u e^{ta} \\
&= e^{ta}\big((1-\lambda) + \lambda e^u\big) = e^{-u\lambda}\big((1-\lambda) + \lambda e^u\big).)
\end{aligned}$$

Therefore

$$(+) \qquad \log \mathbb{E}(\exp(t\xi)) \le -\lambda u + \log((1-\lambda) + \lambda e^u) =: L(u)$$

Differentiate twice to obtain

$$L'(u) = -\lambda + \frac{\lambda e^u}{(1-\lambda) + \lambda e^u} = -\lambda + \frac{\lambda}{(1-\lambda)e^{-u} + \lambda}$$

and

$$\begin{aligned} L''(u) &= \frac{-\lambda(1-\lambda)e^{-u} \cdot (-1)}{((1-\lambda)e^{-u} + \lambda)^2} = \frac{\lambda}{(1-\lambda)e^{-u} + \lambda} \cdot \frac{(1-\lambda)e^{-u}}{(1-\lambda)e^{-u} + \lambda} \\ &= \frac{\lambda}{(1-\lambda)e^{-u} + \lambda}\Big(1 - \frac{\lambda}{(1-\lambda)e^{-u} + \lambda}\Big) \le \frac{1}{4}, \end{aligned}$$

since $x(1-x) \le \frac{1}{4} \quad \forall\, x \in \mathbb{R}$.
Expanding by Taylor's theorem we obtain with an appropriate $u^* \in (0, u)$

$$L(u) = \underbrace{L(0)}_{=0} + u \underbrace{L'(0)}_{=0} + \frac{1}{2}u^2 \underbrace{L''(u^*)}_{\le \frac{1}{4}} \le \frac{1}{8}u^2.$$

Applying the last inequality to (+) for each $\xi = \xi_i$ and each $u = t(b_i - a_i)$, $1 \le i \le n$, we get $\mathbb{E}(\exp(t\xi_i)) \le \exp(\frac{1}{8}t^2(b_i - a_i)^2)$, and thus by (5.5.1) we arrive at

$$\mathbb{P}(\sum_{i=1}^n \xi_i \ge x) \le \exp(-xt + \frac{1}{8}t^2 \sum_{i=1}^n (b_i - a_i)^2).$$

Finally, set $t := 4x / \sum\limits_{i=1}^n (b_i - a_i)^2$ to minimize the quadratic form on the rhs of the last inequality which yields the result. $\square$

(In fact, with $C_n := \sum\limits_{i=1}^n (b_i - a_i)^2, \quad -xt + \frac{1}{8}t^2 C_n \underset{(t=4x/C_n)}{=} -\frac{4x^2}{C_n} + \frac{1}{8}\frac{16x^2}{C_n} = -\frac{2x^2}{C_n}$.)

5.5.3. Corollary.
Applying the same arguments to $-\xi_i, 1 \le i \le n$, and combining with the inequality for ξ_i one gets a two-sided bound under the same conditions, namely for each $x > 0$

$$\mathbb{P}(|\sum_{i=1}^n \xi_i| \ge x) \le 2\, \exp(-2x^2 / \sum_{i=1}^n (b_i - a_i)^2).$$

5.5.4. Corollary.
Let $(\varepsilon_i)_{i\in\mathbb{N}}$ be a Rademacher sequence (defined on $(\Omega, \mathcal{A}, \mathbb{P})$) and $(c_i)_{i\in\mathbb{N}}$ be an arbitrary sequence of real numbers. Then for each $n \in \mathbb{N}$ and each $x > 0$

$$\mathbb{P}(|\sum_{i=1}^n \varepsilon_i c_i| \ge x) \le 2\, \exp(-x^2/2 \sum_{i=1}^n c_i^2).$$

PROOF. Follows from 5.5.3 applied to $\xi_i := \varepsilon_i c_i$, $b_i := -a_i := |c_i|$. Note that $\sum\limits_{i=1}^{n} c_i^2 = Var(\sum\limits_{i=1}^{n} \varepsilon_i c_i)$. □

It is this last inequality which is applied in [Po84], pp. 16,26,31,150,164, and [Gee00], Section 3.6.: Uniform laws of large numbers under random entropy conditions, pp. 34 - .

5.5.5. Bennet's Inequality.
(cf. [Ben62])
Let $\xi_1, ..., \xi_n$ be independent rv's (defined on $(\Omega, \mathcal{A}, \mathbb{P})$) with zero means and bounded ranges: $|\xi_i| \leq M < \infty$. Write σ_i^2 for the variance of ξ_i and suppose that $\sum\limits_{i=1}^{n} \sigma_i^2 \leq V < \infty$. Then for each $x > 0$

$$\mathbb{P}(|\sum_{i=1}^{n} \xi_i| \geq x) \leq 2\, \exp(-\frac{1}{2}x^2 V^{-1}\, B(MxV^{-1})).$$

where $B(\lambda) := \frac{(1+\lambda)\log(1+\lambda)-\lambda}{\lambda^2/2}$ *for* $\lambda > 0$.

PROOF. It suffices to establish the corresponding one-sided inequality. The two-sided inequality will follow by combining it with the companion inequality for $-\xi_i, 1 \leq i \leq n$.

In deriving an upper bound for the moment generating function $\mathbb{E}(\exp(t\xi_i))$ we will, as in the proof of Hoeffding's Inequality, drop the subscript i temporarily. So, noticing that $|\xi| \leq M$

$$\begin{aligned}
\mathbb{E}(\exp(t\xi)) &= 1 + \underbrace{\mathbb{E}(\xi)}_{=0} + \sum_{k\geq 2} \frac{t^k}{k!} \underbrace{\mathbb{E}(\xi^2 \xi^{k-2})}_{\text{exists, since } \mathbb{E}(|\xi^k|)\leq M^k < \infty} \\
&\leq 1 + \sum_{k\geq 2} \frac{t^k}{k!} \sigma^2 \cdot M^{k-2} \\
&= 1 + \sigma^2 \cdot g(t) \leq \exp(\sigma^2 \cdot g(t)),
\end{aligned}$$

where $g(t) := (e^{tM} - 1 - tM)/M^2$. Applying this inequality for each $\xi = \xi_i$ and $\sigma^2 = \sigma_i^2, 1 \leq i \leq n$, yields

$$\mathbb{E}(\exp(t\xi_i)) \leq \exp(\sigma_i^2 \cdot g(t)),$$

and thus by (5.5.1) we arrive at

$$\mathbb{P}(\sum_{i=1}^{n} \xi_i \geq x) \leq \exp(-xt + g(t) \cdot \sum_{i=1}^{n} \sigma_i^2) \leq \exp(V \cdot g(t) - xt).$$

Differentiate to find the minimum value

$$t = M^{-1} \log(1 + MxV^{-1})$$

which is positive. It remains to check that with this t and the definition of $g(t)$ the rhs of the last inequality yields the result as stated in 5.5.5. □

5.5.6. Remark.
For the proof of 5.5.5 the condition $|\xi_i| \le M < \infty$ was essential. In [Due00], pp 27 -   it is shown that the one-sided inequality, i.e.

$$\mathbb{P}(\sum_{i=1}^{n} \xi_i \ge x) \le \exp(-\frac{1}{2}x^2 V^{-1} B(MxV^{-1}))$$

holds for each $x > 0$ even under the assumption that $\xi_i \le M < \infty, 1 \le i \le n$.

5.5.7. Remark.
The function $B(\cdot)$ is well behaved:
$B(\lambda)$ is continuous and monotone decreasing in $\lambda > 0$;
$\lambda B(\lambda)$ is continuous and monotone increasing in $\lambda > 0$;
$B(0+) := \lim_{\lambda \searrow 0} B(\lambda) = 1$;
$B(\lambda) = (2 + o(1)) \log(\lambda)/\lambda$ if $\lambda \to \infty$;
For each $\lambda > 0$ $B(\lambda) > \frac{1}{1+\lambda/3}$.
If we replace this lower bound in 5.5.5 we obtain under the same conditions for each $x > 0$

$$\mathbb{P}(|\sum_{i=1}^{n} \xi_i| \ge x) \le 2\,\exp(-\frac{1}{2}x^2/(V + \frac{1}{3}Mx)) \tag{5.5.8}$$

which is known as Bernstein's Inequality *(cf. [Ber24]).*

6 Uniform Laws of Large Numbers (ULLN)

6.1 A ULLN for RMP's

Let again $X = (X, \mathcal{X})$ be an arbitrary measurable space and denote with $M(X)$ the space of all p-measures w on $\mathcal{X}$, equipped with the smallest σ-field $\mathcal{M}$ such that all the maps $w \longmapsto w(B), B \in \mathcal{X}$, are measurable.
Let $\mathcal{F}$ be a class of $\mathcal{X}$-measurable functions $f : X \longrightarrow \mathbb{R}$ with $\mathcal{X}$-measurable envelope $F : X \longrightarrow \mathbb{R}$ (supposed to be countable for simplicity to avoid measurability considerations). Let, as in Section 3, $(w_{nj})_{1\le j\le j(n), n\in\mathbb{N}}$ be a triangular array of *random* p-measures on $\mathcal{X}$, considered as re's in $(M(X), \mathcal{M})$, and let $(\xi_{nj})_{1\le j\le j(n), n\in\mathbb{N}}$ be a triangular array of real-valued rv's (i.e. re's in $(\mathbb{R}, \mathcal{B})$), where $j(n) \longrightarrow \infty$ as $n \to \infty$.

We are going to present a ULLN for RMP's $S_n = (S_n(f))_{f\in\mathcal{F}}$ with

$$S_n(f) := \sum_{j\le j(n)} w_{nj}(f) \cdot \xi_{nj}, \quad f \in \mathcal{F}, \tag{6.1.1}$$

as introduced in Section 3.1, where, as already remarked there, we do assume (cf. 5.1.1) that the processes $(w_{nj}(f) \cdot \xi_{nj})_{f\in\mathcal{F}}$ are defined via coordinate projections on the product p-space

$$(\Omega, \mathcal{A}, \mathbb{P}) := \Big(\underset{\mathbb{N}}{\times} (\underset{j\le j(n)}{\times} (M(X) \times \mathbb{R})) \;,\; \bigotimes_{\mathbb{N}}(\bigotimes_{j\le j(n)} (\mathcal{M}\bigotimes\mathcal{B})) \;,\; \underset{\mathbb{N}}{\times}(\underset{j\le j(n)}{\times} \mathcal{L}\{(w_{nj}, \xi_{nj})\})\Big),$$

whence for all $n \in \mathbb{N}$ the sequence

$$(w_{n1}, \xi_{n1}), ..., (w_{nj(n)}, \xi_{nj(n)})$$

is a sequence of independent but not necessarily identically distributed pairs of re's in $(M(X) \times \mathbb{R}, \mathcal{M}\bigotimes\mathcal{B})$, i.e. the laws $\mathcal{L}\{(w_{nj}, \xi_{nj})\}$ need not be identical; also dependence within each pair is allowed.
(Note that in the notation of definition 5.1.1 we have now that

$$\Omega \ni \omega \longmapsto \eta_{nj}(\omega) = h_{nj}(\omega_{nj}) := h_{nj}((w_{nj}, \xi_{nj})) := (w_{nj}(f) \cdot \xi_{nj})_{f\in\mathcal{F}} \in V := \mathbb{R}^{\mathcal{F}}.)$$

In order to formulate our ULLN we need some more notation:
Given $S_n = (S_n(f))_{f\in\mathcal{F}}$ with $S_n(f)$ as in (6.1.1), let for any $\delta > 0$

$$\mu_{n\delta} := \sum_{j\le j(n)} w_{nj} \cdot |\xi_{nj}| \cdot I(w_{nj}(F)|\xi_{nj}| \le \delta),$$

and let $\bar{d}^{(1)}_{\mu_{n\delta}}$ be the *random* L_1-pseudometric on $\mathcal{F}$ defined by

$$\bar{d}^{(1)}_{\mu_{n\delta}}(f, g) := \sum_{j\le j(n)} |w_{nj}(f) - w_{nj}(g)| \cdot |\xi_{nj}| \cdot I(w_{nj}(F)|\xi_{nj}| \le \delta)$$

for $f, g \in \mathcal{F}$. Finally, for any $\tau > 0$, let $N(\tau, \mathcal{F}, \bar{d}^{(1)}_{\mu_{n\delta}})$ be the *random* covering number of $(\mathcal{F}, \bar{d}^{(1)}_{\mu_{n\delta}})$ (see the definition 4.1.1).

Then we have the following result (cf. [Gae98], Theorem 2.1):

6.1.2. THEOREM (ULLN for RMP's).
Assume that (6.1.3) – (6.1.5) hold, where (for $1 \le p < \infty$)

$$\text{(6.1.3)} \qquad \lim_{n\to\infty} \sum_{j\le j(n)} \mathbb{E}^{\frac{1}{p}}\Big(w_{nj}(F)^p \cdot |\xi_{nj}|^p \cdot I(w_{nj}(F)|\xi_{nj}| > \delta)\Big) = 0 \quad \textit{for all } \ \delta > 0$$

$$\text{(6.1.4)} \qquad \sup_{n\in\mathbb{N}} \sum_{j\le j(n)} \mathbb{E}\Big(w_{nj}(F) \cdot |\xi_{nj}| \cdot I(w_{nj}(F)|\xi_{nj}| \le \delta_1)\Big) < \infty \quad \textit{for some } \ \delta_1 > 0$$

(6.1.5) *For all $\tau > 0$ there exists $\delta \equiv \delta(\tau) > 0$ such that $\big(N(\tau \cdot \mu_{n\delta}(F), \mathcal{F}, \bar{d}^{(1)}_{\mu_{n\delta}}\,)\big)_{n\in\mathbb{N}}$ is stochastically bounded.*

Then

$$\text{(6.1.6)} \qquad \sup_{f\in\mathcal{F}} |S_n(f) - \mathbb{E}(S_n(f))| \ \xrightarrow{L_p}\ 0\ ,$$

where $\xrightarrow{L_p}$ denotes convergence w.r.t. the L_p-metric.

($\big(N(\tau \cdot \mu_{n\delta}(F), \mathcal{F}, \bar{d}^{(1)}_{\mu_{n\delta}}\,)\big)_{n\in\mathbb{N}}$ stochastically bounded means that for all $\rho > 0$ there exists an $M \equiv M(\tau, \rho) < \infty$ such that

$$\limsup_{n\to\infty} \mathbb{P}^*\Big(N(\tau \cdot \mu_{n\delta}(F), \mathcal{F}, \bar{d}^{(1)}_{\mu_{n\delta}}\,) > M\Big) < \rho. \ \ )$$

PROOF. Concerning (6.1.6) we remark that by (6.1.3) we also have (since $||\cdot||_1 \le ||\cdot||_p$)

$$\lim_{n\to\infty} \sum_{j\le j(n)} \mathbb{E}\Big(w_{nj}(F) \cdot |\xi_{nj}| \cdot I(w_{nj}(F)|\xi_{nj}| > \delta)\Big) = 0 \quad \forall \delta > 0,$$

whence by (6.1.4)

$$\sup_{n\in\mathbb{N}} \sum_{j\le j(n)} \mathbb{E}\Big(w_{nj}(F) \cdot |\xi_{nj}|\Big) < \infty,$$

and therefore $\mathbb{E}(|S_n(f)|) < \infty \quad \forall n \in \mathbb{N}$ and $\forall f \in \mathcal{F}$.

Now, by the Symmetrization Inequality 5.1.2 (applied with $\Psi(x) := x^p, x \in \mathbb{R}_+$), it suffices to show that

$$\lim_{n\to\infty} \mathbb{E}\Big(\sup_{f\in\mathcal{F}} |\sum_{j\le j(n)} \varepsilon_j w_{nj}(f)\xi_{nj}|^p\Big) = 0$$

where $(\varepsilon_j)_{j\in\mathbb{N}}$ is a canonically formed Rademacher sequence which is independent of both arrays (w_{nj}) and (ξ_{nj}).

Next, by (6.1.3) there exists a sequence $(\delta_n)_{n\in\mathbb{N}}$ of positive real numbers with $\delta_n \to 0$ and

$$\lim_{n\to\infty} \sum_{j\le j(n)} \mathbb{E}^{\frac{1}{p}}\Big(\mu_{nj}(F)^p \cdot I(\mu_{nj}(F) > \delta_n)\Big) = 0,$$

where we put $\mu_{nj} := w_{nj} \cdot |\xi_{nj}|$ for short. Hence it suffices to show that

$$\lim_{n\to\infty} \mathbb{E}\big(\sup_{f\in\mathcal{F}} |S_{n\delta_n}(f)|^p \big) = 0,$$

where $S_{n\delta_n}(f) := \sum_{j\le j(n)} \varepsilon_j w_{nj}(f)\xi_{nj} \cdot I(\mu_{nj}(F) \le \delta_n)$.

But, since the summands of $S_{n\delta_n}(f)$ are bounded by δ_n (with $\delta_n \longrightarrow 0$ as $n \to \infty$), it follows by application of Hoffmann-Jørgensen's Inequality 5.3.1 (with $\Psi(x) := x^p$, $x \in \mathbb{R}_+$) that it suffices to verify

(a) $$\sup_{f\in\mathcal{F}} |S_{n\delta_n}(f)| \overset{\mathbb{P}}{\longrightarrow} 0.$$

To prove (a), let $\beta > 0$ and $\varepsilon > 0$ be arbitrary but fixed. Let (cf. (6.1.4))

$$C := \sup_{n\in\mathbb{N}} \sum_{j\le j(n)} \mathbb{E}\big(\mu_{nj}(F) \cdot I(\mu_{nj}(F) \le \delta_1)\big).$$

Choose $\tau := \varepsilon\beta/2C$ and take $\delta = \delta(\tau)$ according to (6.1.5).

Now, for $\rho := \varepsilon/2$, let $M = M(\tau, \rho) > 0$ be such that for $A_n := \{N(\tau\mu_{n\delta}(F), \mathcal{F}, \bar{d}^{(1)}_{\mu_{n\delta}})^* > M\}$ we have by (6.1.5) that $\limsup_{n\to\infty} \mathbb{P}(A_n) < \rho$ where the star $(^\star)$ denotes the measurable cover function (cf. (2.3.17)). Then, by Markov's Inequality and Fubini's theorem it follows that

(b) $$\mathbb{P}(\sup_{f\in\mathcal{F}} |S_{n\delta_n}(f)| > \beta) \le \mathbb{P}(A_n) + \beta^{-1}\mathbb{E}(1_{\complement A_n}\mathbb{E}_\varepsilon(\sup_{f\in\mathcal{F}} |S_{n\delta_n}(f)|)),$$

where $\mathbb{E}_\varepsilon$ denotes integration w.r.t. the Rademacher sequence.

Now, for n large enough such that $\delta_n \le \delta$ and $\delta_n \le \delta_1$ (δ_1 as in (6.1.4)) we obtain by the Maximal Inequality for Rademacher Averages with a universal constant $0 < K_1 < \infty$ that

$$\begin{aligned}
&\mathbb{E}_\varepsilon(\sup_{f\in\mathcal{F}} |S_{n\delta_n}(f)|)\\
&\le \tau\mu_{n\delta}(F) +\\
&\quad K_1(1 + N(\tau\mu_{n\delta}(F), \mathcal{F}, \bar{d}^{(1)}_{\mu_{n\delta}}))^{\frac{1}{2}} \cdot \sup_{f\in\mathcal{F}} \Big|\sum_{j\le j(n)} w^2_{nj}(f) \cdot \xi^2_{nj} \cdot I(\mu_{nj}(F) \le \delta_n)\Big|^{\frac{1}{2}}\\
&\le \tau\mu_{n\delta}(F) +\\
&\quad \delta_n^{\frac{1}{2}} K_1(1 + N(\tau\mu_{n\delta}(F), \mathcal{F}, \bar{d}^{(1)}_{\mu_{n\delta}}))^{\frac{1}{2}} \Big(\sum_{j\le j(n)} \mu_{nj}(F) \cdot I(\mu_{nj}(F) \le \delta_1)\Big)^{\frac{1}{2}}.
\end{aligned}$$

(Actually the Maximal Inequality even holds with $\log N(\tau\mu_{n\delta}(F), \mathcal{F}, \bar{d}^{(1)}_{\mu_{n\delta}})$ instead of $N(\tau\mu_{n\delta}(F), \mathcal{F}, \bar{d}^{(1)}_{\mu_{n\delta}})$.)

So by definition of A_n it follows that for large enough n

(c) $$\begin{aligned}&\mathbb{E}(1_{\complement A_n}\mathbb{E}_\varepsilon(\sup_{f\in\mathcal{F}} |S_{n\delta_n}(f)|)) \quad \le\\ &\tau\mathbb{E}(\mu_{n\delta}(F)) \;+\; \delta_n^{\frac{1}{2}} K_1(1 + M)^{\frac{1}{2}}\, \mathbb{E}^{\frac{1}{2}}\big(\textstyle\sum_{j\le j(n)} \mu_{nj}(F) \cdot I(\mu_{nj}(F) \le \delta_1)).\end{aligned}$$

Now, observe that $\mu_{n\delta} \leq \sum_{j\leq j(n)} \mu_{nj} \cdot I(\mu_{nj}(F) \leq \delta_1) + \sum_{j\leq j(n)} \mu_{nj} \cdot I(\mu_{nj}(F) > \delta_1)$ whence by (6.1.3) we have

$$\text{(d)} \qquad \limsup_{n\to\infty} \mathbb{E}(\mu_{n\delta}(F)) \leq C + \limsup_{n\to\infty} \sum_{j\leq j(n)} \mathbb{E}^{\frac{1}{p}}\big(\mu_{nj}(F)^p \cdot I(\mu_{nj}(F) > \delta_1)\big) = C.$$

Hence we obtain by (b) – (d) that

$$\limsup_{n\to\infty} \mathbb{P}(\sup_{f\in\mathcal{F}} |S_{n\delta_n}(f)| > \beta) \leq \rho + \beta^{-1}\tau C = \varepsilon$$

by the choice of ρ and τ. Since ε and β were arbitrary, this implies (a). □

NOTE: Condition (6.1.5) in the theorem can be replaced by (6.1.5)′

$$\text{(6.1.5)}' \qquad \begin{array}{c}\text{For all } \tau > 0 \text{ there exists } \delta \equiv \delta(\tau) > 0 \text{ such that} \\ \big(N(\tau, \mathcal{F}, \bar{d}^{(1)}_{\mu_{n\delta}})\big)_{n\in\mathbb{N}} \text{ is stochastically bounded.}\end{array}$$

Indeed, following the proof of theorem 6.1.2 up to (b) now with $\tau := \varepsilon\beta/2$ and $A_n := \{N(\tau, \mathcal{F}, \bar{d}^{(1)}_{\mu_{n\delta}})^* > M\}$, the Maximal Inequality for Rademacher Averages now gives

$$\mathbb{E}_\varepsilon(\sup_{f\in\mathcal{F}} |S_{n\delta_n}(f)|) \leq$$
$$\tau + K_1(1 + N(\tau, \mathcal{F}, \bar{d}^{(1)}_{\mu_{n\delta}}))^{\frac{1}{2}} \cdot \sup_{f\in\mathcal{F}} \Big|\sum_{j\leq j(n)} w^2_{nj}(f) \cdot \xi^2_{nj} \cdot I(\mu_{nj}(F) \leq \delta_n)\Big|^{\frac{1}{2}}.$$

The result then follows as above.

6.1.7. Remark.
Since for RMP's we did tacitily assume (cf. Section 3.1) measurability and finiteness of $w_{nj}(F)$ for all $1 \leq j \leq j(n)$ and $n \in \mathbb{N}$, the same is true for the random measures $\mu_{n\delta}$, whence $\mu_{n\delta}(F) < \infty$ for all $n \in \mathbb{N}$ and $\delta > 0$. Therefore, it follows from 4.3.17 that in case of VCGC's $\mathcal{F}$, for each $\tau > 0$ there exists a constant $C = C(\tau), 0 < C < \infty$, such that (note that $\bar{d}^{(1)}_{\mu_{n\delta}}(f,g) \leq d^{(1)}_{\mu_{n\delta}} := \mu_{n\delta}(|f-g|)$)

$$\sup_{n\in\mathbb{N}} N(\tau \cdot \mu_{n\delta}(F), \mathcal{F}, \bar{d}^{(1)}_{\mu_{n\delta}}) \leq C,$$

whence the condition (6.1.5) is automatically fulfilled for VCGC's $\mathcal{F}$ with envelope F.

Thus, from Theorem 6.1.2 we get

6.1.8. Corollary.
Let $\mathcal{F}$ be a countable VCGC with envelpoe F. Assume (6.1.3) and (6.1.4), i.e. (for $1 \leq p < \infty$)

$$\lim_{n\to\infty} \sum_{j\leq j(n)} \mathbb{E}^{\frac{1}{p}}\Big(w_{nj}(F)^p \cdot |\xi_{nj}|^p \cdot I(w_{nj}(F)|\xi_{nj}| > \delta)\Big) = 0 \quad \textit{for all } \delta > 0, \textit{ and}$$

$$\sup_{n\in\mathbb{N}} \sum_{j\leq j(n)} \mathbb{E}\Big(w_{nj}(F) \cdot |\xi_{nj}| \cdot I(w_{nj}(F)|\xi_{nj}| \leq \delta_1)\Big) < \infty \quad \textit{for some } \delta_1 > 0.$$

Then

$$\sup_{f\in\mathcal{F}} |S_n(f) - \mathbb{E}(S_n(f))| \xrightarrow{L_p} 0,$$

where $S_n(f) := \sum_{j\le j(n)} w_{nj}(f)\cdot \xi_{nj}, \quad f\in\mathcal{F}$, *and where the processes* $(w_{nj}(f)\cdot\xi_{nj})_{f\in\mathcal{F}}$ *are defined as coordinate projections on the product space* $(\Omega,\mathcal{A},\mathbb{P})$ *as introduced above.*

6.2 ULLN's for partial-sum processes with either fixed or random locations

Let $X = (X,\mathcal{X})$ be an arbitrary measurable space, $\mathcal{C}\subset\mathcal{X}$ a countable VCC, $(\eta_{nj})_{1\le j\le j(n),n\in\mathbb{N}}$ be a triangular array of re's in $(X,\mathcal{X})$ and $(\xi_{nj})_{1\le j\le j(n),n\in\mathbb{N}}$ a triangular array of rv's with $j(n)\longrightarrow\infty$ as $n\to\infty$, such that for each $n\in\mathbb{N}$ the sequence of pairs $(w_{n1},\xi_{n1}),\ldots,(w_{nj(n)},\xi_{nj(n)})$ is independent but not necessarily identically distributed; also the components within each pair need not be independent.

Then, by taking $w_{nj} := \delta_{\eta_{nj}}$ ($\delta_{\eta_{nj}}$ = Dirac measure at η_{nj}) we obtain from 6.1.8 immediately the following result for partial-sum processes with random locations as introduced in Section 3.2.1:

6.2.1. THEOREM *(cf. [Gae94b], Theorem 3.1).*
Assume that the following two conditions are fulfilled:

(6.2.2)
$$\lim_{n\to\infty}\sum_{j\le j(n)}\mathbb{E}\Big(|\xi_{nj}|\cdot I(|\xi_{nj}|>\delta)\Big) = 0 \quad \textit{for all } \delta>0$$

(6.2.3)
$$\sup_{n\in\mathbb{N}}\sum_{j\le j(n)}\mathbb{E}\Big(|\xi_{nj}|\cdot I(|\xi_{nj}|\le\delta_1)\Big) < \infty \quad \textit{for some } \delta_1>0.$$

Then, for the partial-sum processes $S_n = (S_n(C))_{C\in\mathcal{C}}$ *defined by*

$$S_n(C) := \sum_{j\le j(n)} 1_C(\eta_{nj})\cdot\xi_{nj}, \quad C\in\mathcal{C},$$

one has

(6.2.4)
$$\lim_{n\to\infty}\mathbb{E}\Big(\sup_{C\in\mathcal{C}}|S_n(C)-\mathbb{E}(S_n(C))|\Big) = 0\,.$$

6.2.5. Remark.
Note that (6.2.2) and (6.2.3) together imply that $\sup_{n\in\mathbb{N}}\sum_{j\le j(n)}\mathbb{E}(|\xi_{nj}|)<\infty$, *and thus* $\mathbb{E}(S_n(C))$ *exists for all* $n\in\mathbb{N}$ *and* $C\in\mathcal{C}$.

In the *identically distributed (id)* - case, that is, when $\xi_{nj} = j(n)^{-1}\xi_j, 1 \le j \le j(n), n \in \mathbb{N}$, with $j(n) \longrightarrow \infty$ as $n \to \infty$, for some sequence $(\xi_j)_{j\in\mathbb{N}}$ of identically distributed ξ_j, we have for each $\delta > 0$

$$\sum_{j\le j(n)} \mathbb{E}\Big(|\xi_{nj}| \cdot I(|\xi_{nj}| > \delta)\Big) = \mathbb{E}\Big(|\xi_1| \cdot I(|\xi_1| > \delta j(n))\Big)$$

and

$$\sup_{n\in\mathbb{N}} \sum_{j\le j(n)} \mathbb{E}\Big(|\xi_{nj}| \cdot I(w_{nj}(F)|\xi_{nj}| \le \delta)\Big) \le \sup_{n\in\mathbb{N}} \sum_{j\le j(n)} \mathbb{E}(|\xi_{nj}|) = \mathbb{E}(|\xi_1|),$$

whence in the id-case both conditions (6.2.2) and (6.2.3) are fulfilled under the only assumption $\mathbb{E}(|\xi_1|) < \infty$.

From Theorem 6.2.1 together with Remark 6.2.5 we obtain

6.2.6. Corollary.
Let $S_n(C) := j(n)^{-1} \sum_{j\le j(n)} 1_C(\eta_{nj}) \cdot \xi_j,\ \ C \in \mathcal{C},\ \mathcal{C} \subset \mathcal{X}$ being a countable VCC, $(w_{nj})_{1\le j\le j(n), n\in\mathbb{N}}$ be a triangular array of re's in $(X, \mathcal{X})$ with $j(n) \longrightarrow \infty$ as $n \to \infty$, and let $(\xi_j)_{j\in\mathbb{N}}$ be a sequence of identically distributed rv's ξ_j with $\mathbb{E}(|\xi_1|) < \infty$ such that for all $n \in \mathbb{N}$ $\ (w_{n1}, \xi_1), ..., (w_{nj(n)}, \xi_n)$ is a sequence of independent but not necessarily identically distributed pairs of re's in $(X \times \mathbb{R}, \mathcal{X} \bigotimes \mathcal{B})$. Then

$$\lim_{n\to\infty} \mathbb{E}\Big(\sup_{C\in\mathcal{C}} |S_n(C) - \mathbb{E}(S_n(C))|\Big) = 0.$$

Concerning partial-sum processes with fixed locations in $X = I^d \equiv [0,1]^d, d \ge 1$, Theorem 6.2.1 together with Remark 6.2.5 implies the following result (cf. Section 1.3 and 1.4):

6.2.7. Corollary.
Let $S_n(C) := n^{-d} \sum_{\underline{j}\in J_n} 1_C(\underline{j}/n) \cdot \xi_{\underline{j}},\ \ C \in \mathcal{C}$, where $\xi_{\underline{j}}, \underline{j} \in \mathbb{N}^d$, are i i d rv's with $\mathbb{E}(|\xi_{\underline{1}}|) < \infty$, and where $\mathcal{C} \subset I^d \cap \mathcal{B}^d$ is a countable VCC; then

$$\lim_{n\to\infty} \mathbb{E}\Big(\sup_{C\in\mathcal{C}} |S_n(C) - n^{-d}|J_n \cap (nC)| \cdot \mathbb{E}(|\xi_{\underline{1}}|)|\Big) = 0,$$

where $J_n := \{1, ..., n\}^d$ $\quad$ (and $nC := \{nc : c \in \mathcal{C}\}$).

6.2.8. Remark.
Considering, more generally, function-indexed partial-sum processes $S_n = (S_n(f))_{f\in\mathcal{F}}$, defined by

$$S_n(f) := j(n)^{-1} \sum_{j\le j(n)} f(\eta_{nj}) \cdot \xi_j, \ \ f \in \mathcal{F},$$

with $\mathcal{F}$ being a countable and uniformly bounded VCGC (i.e. with envelope $F \equiv M < \infty$), and where $(\eta_{nj})_{1\le j\le j(n), n\in\mathbb{N}}$ is a triangular array of re's in $(X, \mathcal{X})$ with $j(n) \longrightarrow \infty$ as $n \to \infty$, and $\xi_j, j \in \mathbb{N}$, are identically distributed rv's with $\mathbb{E}(|\xi_1|) < \infty$ such that for all $n \in \mathbb{N}$ $\ (\eta_{n1}, \xi_1), ..., (\eta_{nj(n)}, \xi_{j(n)})$ is a sequence of independent but not necessarily identically distributed pairs of re's in $(X \times \mathbb{R}, \mathcal{X} \bigotimes \mathcal{B})$, then in the same way as in the set-indexed case above, Theorem 6.2.1 together with Remark 6.2.5 yields

$$\lim_{n\to\infty} \mathbb{E}\Big(\sup_{f\in\mathcal{F}} |S_n(f) - \mathbb{E}(S_n(f))|\Big) = 0.$$

6.3 ULLN's for empirical processes

Given an arbitrary measurable space $X = (X, \mathcal{X})$, let us consider at first the set-indexed case, i.e. with a countable VCC $\mathcal{C} \subset \mathcal{X}$ as parameter space for the empirical measures $\nu_n = (\nu_n(C))_{C \in \mathcal{C}}$ defined by

$$\nu_n(C) := j(n)^{-1} \sum_{j \leq j(n)} 1_C(\eta_{nj}), \quad C \in \mathcal{C},$$

where $(\eta_{nj})_{1 \leq j \leq j(n), n \in \mathbb{N}}$ is a triangular array of rowwise independent but not necessarily identically distributed re's in $(X, \mathcal{X})$ with $j(n) \longrightarrow \infty$ as $n \to \infty$. Then it follows from Corollary 6.2.6 (with $\xi_j \equiv 1$) that

$$\lim_{n \to \infty} \mathbb{E}\Big(\sup_{C \in \mathcal{C}} |\nu_n(C) - \bar{\nu}_n(C)| \Big) = 0, \tag{6.3.1}$$

where $\bar{\nu}_n(C) := j(n)^{-1} \sum_{j \leq j(n)} \mathbb{P}(\eta_{nj} \in C)$, $C \in \mathcal{C}$.

Especially, if for each $n \in \mathbb{N}$ $\eta_{nj} = \eta_j, 1 \leq j \leq j(n)$, with $\eta_j, j \in \mathbb{N}$, being i i d re's in $(X, \mathcal{X})$ with law ν on $\mathcal{X}$, then for

$$\nu_n(C) := j(n)^{-1} \sum_{j \leq j(n)} 1_C(\eta_j), \quad C \in \mathcal{C},$$

it follows together with (2.1.5) that (cf. Theorem 2.1.6)

$$||\nu_n - \nu||_{\mathcal{C}} := \sup_{C \in \mathcal{C}} |\nu_n(C) - \nu(C)| \longrightarrow 0 \quad \mathbb{P} - a.s. \tag{6.3.2}$$

As to the function-indexed case we get from Corollary 6.1.8 (with $p = 1, w_{nj} := \delta_{\eta_j}$ and $\xi_{nj} := j(n)^{-1}, 1 \leq j \leq j(n), n \in \mathbb{N}$) the following more general result mentioned already in connection with (4.3.9):

6.3.3. THEOREM.
Let $X = (X, \mathcal{X})$ be an arbitrary measurable space, $\eta_j, j \in \mathbb{N}$, be i i d re's in $(X, \mathcal{X})$ with law ν on $\mathcal{X}$ (defined as coordinate projections on the p-space $(\Omega, \mathcal{A}, \mathbb{P}) := (X^{\mathbb{N}}, \mathcal{X}^{\mathbb{N}}, \nu^{\mathbb{N}})$), and let $\mathcal{F}$ be a countable VCGC of $\mathcal{X}$-measurable functions $f : X \longrightarrow \mathbb{R}$ with $\mathcal{X}$-measurable envelope $F : X \longrightarrow \mathbb{R}$ such that $\nu(F) := \int_X F \, d\nu < \infty$. Then, for $\nu_n(f) := j(n)^{-1} \sum_{j \leq j(n)} f(\eta_j)$, $f \in \mathcal{F}$, one has (with $j(n) \longrightarrow \infty$ as $n \to \infty$)

$$\lim_{n \to \infty} \mathbb{E}\Big(\sup_{f \in \mathcal{F}} |\nu_n(f) - \nu(f)| \Big) = 0. \tag{6.3.4}$$

Moreover, by the same reversed martingale argument which led to (2.1.5) one obtains also

$$\sup_{f \in \mathcal{F}} |\nu_n(f) - \nu(f)| \longrightarrow 0 \quad \mathbb{P} - a.s. \tag{6.3.5}$$

PROOF. According to 6.1.8 we have to verify (6.1.3) (with $p = 1$) and (6.1.4), where now $w_{nj}(F) = \delta_{\eta_j}(F) = F(\eta_j)$ and $\xi_{nj} = j(n)^{-1}$, i.e.

(+) $\qquad \lim_{n\to\infty} \sum_{j\le j(n)} \mathbb{E}\Big(F(\eta_j)\cdot j(n)^{-1}\cdot I(F(\eta_j)\, j(n)^{-1} > \delta)\Big) = 0 \quad \forall \delta > 0$, and

(++) $\qquad \sup_{n\in\mathbb{N}} \sum_{j\le j(n)} \mathbb{E}\Big(F(\eta_j)\cdot j(n)^{-1}\cdot I(F(\eta_j)\, j(n)^{-1} \le \delta_1)\Big) < \infty \quad$ for some $\delta_1 > 0$.

As to (+), $\sum_{j\le j(n)} \mathbb{E}\Big(F(\eta_j)\cdot j(n)^{-1}\cdot I(F(\eta_j)\, j(n)^{-1} > \delta)\Big) = \mathbb{E}\Big(F(\eta_1)\cdot I(F(\eta_1) > \delta j(n))\Big) \longrightarrow 0$ as $n\to\infty$, since $\mathbb{E}(F(\eta_1)) = \nu(F) < \infty$ by assumption.

As to (++), $\sup_{n\in\mathbb{N}} \sum_{j\le j(n)} \mathbb{E}\Big(F(\eta_j)\cdot j(n)^{-1}\cdot I(F(\eta_j)\, j(n)^{-1} \le \delta_1)\Big) = \sup_{n\in\mathbb{N}} \mathbb{E}\Big(F(\eta_1)\cdot I(F(\eta_1) \le \delta_1 j(n))\Big) \le \mathbb{E}(F(\eta_1)) < \infty$. $\square$

6.3.6. THEOREM*(A ULLN via Bracketing).*
Let $X = (X,\mathcal{X})$ be an arbitrary measurable space, $\eta_j, j\in\mathbb{N}$, be i i d re's in $(X,\mathcal{X})$ with law ν on $\mathcal{X}$ (defined as coordinate projections on the p-space $(\Omega,\mathcal{A},\mathbb{P}) := (X^{\mathbb{N}},\mathcal{X}^{\mathbb{N}},\nu^{\mathbb{N}})$), and let $\mathcal{F}$ be a countable VCGC of $\mathcal{X}$-measurable functions $f : X \longrightarrow \mathbb{R}$ s.t. for any $\varepsilon > 0$ there exist $m\in\mathbb{N}$ and functions $g_1, h_1, ..., g_m, h_m$ in $\mathcal{L}_1(X,\mathcal{X},\nu)$ such that $\forall\, j\le m \quad g_j \le h_j$ and $\nu(h_j - g_j) \le \varepsilon$, and that $\forall\, f\in\mathcal{F}\ \exists\, j \le m$ with $g_j \le f \le h_j$. Then

$$\lim_{n\to\infty} \mathbb{E}\Big(\sup_{f\in\mathcal{F}} |\nu_n(f) - \nu(f)|\Big) = 0.$$

(As the proof will show, that for not necessarily countable $\mathcal{F}$ $\sup_{f\in\mathcal{F}} |\nu_n(f)-\nu(f)| \le \zeta_n$ with rv's ζ_n satisfying $\mathbb{E}(\zeta_n) \longrightarrow 0$.)

PROOF. Let $f\in\mathcal{F}$ be arbitrary and $g_j \le h_j$ be such that $g_j \le f \le h_j$ and $\nu(h_j - g_j) \le \varepsilon$; then

$$\begin{aligned}(\nu_n - \nu)(f) &\le \nu_n(h_j) - \nu(g_j)\\ &= (\nu_n-\nu)(h_j) + \nu(h_j - g_j)\\ &\le (\nu_n-\nu)(h_j) + \varepsilon,\end{aligned}$$

and

$$\begin{aligned}(\nu_n - \nu)(f) &\ge \nu_n(g_j) - \nu(h_j)\\ &= (\nu_n-\nu)(g_j) - \nu(h_j - g_j)\\ &\ge (\nu_n-\nu)(g_j) - \varepsilon,\end{aligned}$$

whence

$$|(\nu_n-\nu)(f)| \le |(\nu_n-\nu)(h_j)| + \varepsilon + |(\nu_n-\nu)(g_j)|,$$

and therefore

$$\begin{aligned}\mathbb{E}\Big(\sup_{f\in\mathcal{F}} |\nu_n(f) - \nu(f)|\Big) &\le \sum_{j\le m} \mathbb{E}(|(\nu_n-\nu)(g_j)|) + \sum_{j\le m} \mathbb{E}(|(\nu_n-\nu)(h_j)|) + \varepsilon\\ &\longrightarrow \varepsilon \quad \text{as } n\to\infty,\end{aligned}$$

since for any $a \in \mathcal{L}_1(X, \mathcal{X}, \nu)$

$$(*) \qquad \lim_{n\to\infty} \mathbb{E}\Big(|(\nu_n - \nu)(a)|\Big) = 0.$$

As $\varepsilon > 0$ was arbitrary, Theorem 6.3.6 is proved. □

As to (*): Note that for any $a \in \mathcal{L}_1(X, \mathcal{X}, \nu)$

$$(\nu_n - \nu)(a) = \frac{1}{n} \sum_{j \le n} (a(\eta_j) - \mathbb{E}(a(\eta_j)))$$

with $\xi_i := a(\eta_j) - \mathbb{E}(a(\eta_j))$ being i i d and $\mathbb{E}(|\xi_1|) < \infty$. Thus $\{S_n := \frac{1}{n} \sum\limits_{j \le n} \xi_j \; : \; n \in \mathbb{N}\}$ is uniformly integrable, whence by the weak law of large numbers (according to which $S_n \xrightarrow{\mathbb{P}} 0$ together with [Gae77], 1.14.9 assertion (*) follows.

6.4 ULLN's for smoothed empirical processes

Throughout this section X is supposed to be an arbitrary *linear metric space* endowed with its Borel σ-field $\mathcal{X}$.

Let $\eta_j, j \in \mathbb{N}$, be i i d re's in $(X, \mathcal{X})$ with law ν on $\mathcal{X}$ (defined as coordinate projections on the p-space $(\Omega, \mathcal{A}, \mathbb{P}) := (X^{\mathbb{N}}, \mathcal{X}^{\mathbb{N}}, \nu^{\mathbb{N}})$).
Let $\nu_n := n^{-1} \sum_{j \le n} \delta_{\eta_j}$ be the empirical measure based on $\eta_1, ..., \eta_n$, $n \in \mathbb{N}$, viewed as nonparametric estimator (of sample size n) for ν.

If the underlying ν is "smooth" it is natural to use a "smoothed" version $\tilde{\nu}_n$ of ν_n as an estimator for ν, rather than the empirical measure itself.
Following Yukich [Yu89] we consider *smoothing through convolution* as follows:

Given a sequence $(\mu_n)_{n \in \mathbb{N}}$, of p-measures μ_n on $\mathcal{X}$ let

$$\tilde{\nu}_n := \nu_n \star \mu_n$$

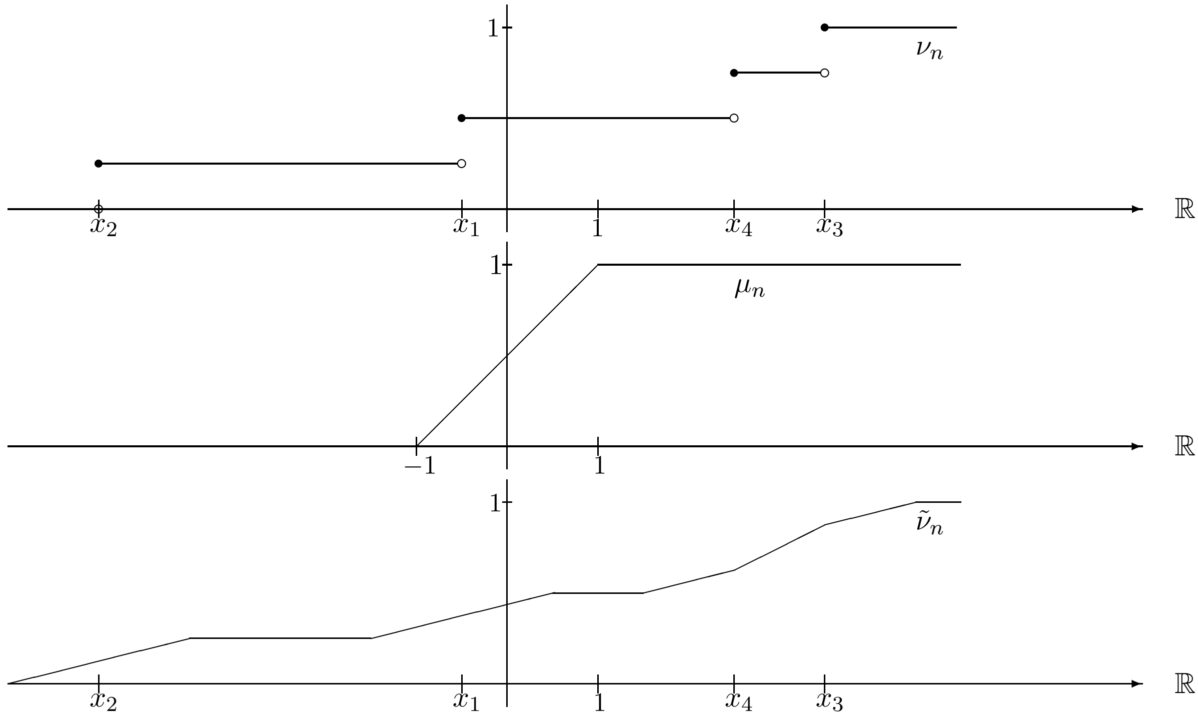
be the so-called *smoothed empirical measure* based on $\eta_1, ..., \eta_n$, i.e.

$$(6.4.1) \qquad \tilde{\nu}_n(B) := \int\limits_X \int\limits_X 1_B(x + y) \nu_n(dx) \mu_n(dy), \quad B \in \mathcal{X}.$$

Note that $\tilde{\nu}_n \equiv \nu_n$ if $\mu_n \equiv \delta_0$ (Dirac measure at 0).

Taking $X = \mathbb{R}$, the following picture shows that by convolution we can turn the discrete empirical measure ν_n into a continuous one. This is not surprising since $\nu_n \star \mu_n$ has a Lebesgue density if μ_n has one.
For illustration we take $n = 4$, μ_n the uniform distribution on $[-1, 1]$ and $x_1, ..., x_n$ a sample from the

rv's $\eta_1, ..., \eta_n$. The picture shows the distribution functions which are also denoted by ν_n, μ_n and $\tilde{\nu}_n$, respectively.



(6.4.1) also includes *kernel smoothing in density estimation.* For this, let us take $X = \mathbb{R}$ for simplicity, and let for each $u \in \mathbb{R}$ $\mu_n((-\infty, u]) := H(\frac{u}{h_n}), h_n > 0$, where

$$H(s) := \int_{-\infty}^{s} K(v)dv, \quad K \geq 0, \quad \int_{\mathbb{R}} K(v)\,dv = 1;$$

(note that in this case $\mu_n \longrightarrow \delta_0$ weakly if $h_n \longrightarrow 0$ as $n \to \infty$).
Then, for each $u \in \mathbb{R}$

$$\begin{aligned}
\tilde{\nu}_n((-\infty, u]) \underset{(6.4.1)}{=} & \int_{\mathbb{R}} \Big[n^{-1} \sum_{j \leq n} 1_{(-\infty,u]}(\eta_j + y) \Big] \mu_n(dy) \\
&= n^{-1} \sum_{j \leq n} \mu_n((-\infty, u - \eta_j]) \quad = \quad n^{-1} \sum_{j \leq n} H(\frac{u - \eta_j}{h_n}) \\
&= n^{-1} \sum_{j \leq n} h_n^{-1} \int_{-\infty}^{u} K(\frac{v - \eta_j}{h_n})\,dv, \qquad \text{i.e.}
\end{aligned}$$

$$\tilde{\nu}_n((-\infty, u]) = \int_{-\infty}^{u} \hat{g}_n(t)\,dt \qquad \forall u \in \mathbb{R} \qquad \text{with}$$

$\hat{g}_n(t) := (nh_n)^{-1}\sum_{j\le n} K(\frac{t-\eta_j}{h_n})$ being the kernel density estimator for the underlying density $g(t)$ of the $\mathcal{L}\{\eta_j\}$'s.

Returning to an arbitrary linear metric space X, let $\mathcal{F}$ be a class of $\mathcal{X}$-measurable functions $f : X \longrightarrow \mathbb{R}$ with $\mathcal{X}$-measurable envelope $F : X \longrightarrow \mathbb{R}$. For each $f \in \mathcal{F}$, put

$$\tilde{\nu}_n(f) := \int_X f\, d\tilde{\nu}_n\ , \tag{6.4.2}$$

tacitly assuming that the integrals of functions $f \in \mathcal{F}$ do exist.
For $\tilde{\nu}_n(f)$ this is the case if $\int_X |f(x+y)|\,\mu_n(dy) < \infty \quad \forall\, x \in X$.

Note that (cf. (6.4.1))

$$\tilde{\nu}_n(f) = \int_X \int_X f(x+y)\nu_n(dx)\mu_n(dy) = n^{-1}\sum_{j\le n}\int_X f(\eta_j+y)\mu_n(dy),\ \text{and} \tag{6.4.3}$$

$$\mathbb{E}(\tilde{\nu}_n(f)) = \nu \star \mu_n(f) \quad \forall f \in \mathcal{F} \text{ and } n \in \mathbb{N}. \tag{6.4.4}$$

(In fact, as to (6.4.4), $\mathbb{E}(\tilde{\nu}_n(f)) = \int_X \mathbb{E}\Big(\int_X f(x+y)\,\nu_n(dx)\Big)\mu_n(dy) = \int_X \mathbb{E}\Big(n^{-1}\sum_{j\le n} f(\eta_j+y)\Big)\,\mu_n(dy) = \int_X\int_X f(x+y)\nu(dx)\mu_n(dy) = \nu\star\mu_n(f)$.)

It will be also tacitly assumed that suprema over $f \in \mathcal{F}$, like $\sup_{f\in\mathcal{F}} |\tilde{\nu}_n(f) - \nu(f)|$, are measurable (being the case by assuming, as in the former sections, that $\mathcal{F}$ is countable, for simplicity).

Now, our aim is to present ULLN's, i.e. sufficient conditions on $\mathcal{F}$ and the smoothing measures $\mu_n, n \in \mathbb{N}$, under which (for $1 \le p < \infty$)

$$\sup_{f\in\mathcal{F}} |\tilde{\nu}_n(f) - \nu(f)| \xrightarrow{L_p} 0. \tag{6.4.5}$$

Concerning $(\tilde{\nu}_n)_{n\in\mathbb{N}}$ as an estimator sequence for an unknown ν, from (6.4.5) one can of course only deduce weak consistency, but, as Pfanzagl [Pf94], p. 188, remarks *strong consistency, i.e. almost sure convergence of an estimator sequence, adds nothing to weak consistency, i.e. convergence in probability, which could be of use on the way to the asymptotic distributions of estimator sequences.* Thus, it is reasonable to seek for sufficient conditions under which (6.4.5) holds true.

Concerning once more the above example of kernel smoothing, (6.4.5) yields $\sup_{u\in\mathbb{R}} |\hat{G}_n(u) - G(u)| \xrightarrow{L_p} 0$, where $\hat{G}_n(u) := \tilde{\nu}_n((-\infty, u]), u \in \mathbb{R}$, and where G is the df of the η_j's.
Fernholz [Fe91] remarks on the estimator $\hat{G}_n$:

Estimators $\hat{G}_n$ derived by integrating density estimators have required less attention. Although estimating a density g by using $\hat{g}_n$ or its distribution function G by using $\hat{G}_n$ are equivalent problems, the error of the corresponding estimator is usually measured in different ways. For density estimation the "L_1 view" (see Devroye and Gyorfi [Dev85]) based on the L_1 error $||\hat{g}_n - g||_1$ has been gaining popularity over the more traditional L_2 approach using $||\hat{g}_n - g||_2$.
In kernel distribution function estimation the discrepancy error between $\hat{G}_n$ and G should be measured in terms of some distance in the space of distribution functions. Metrics such as the supremum norm,

the Prohorov distance, or the Levy distance, provide a useful framework to study the properties of $\hat{G}_n$.

Indeed, from Winter [Win73] and Yamato [Ya73] we have a.s. uniform convergence of $\hat{G}_n$ to G, see also Mack [Ma84] and Prakasa Rao [Pra81].

A more general setting for studying estimators (as already considered in [Win73] and [Ya73]) for a distribution function G is obtained if $\hat{G}_n$ is defined by

$$\hat{G}_n(u) := n^{-1} \sum_{j \leq n} \mu_n((-\infty, u - \eta_j) \quad , \ u \in \mathbb{R},$$

with p-measures μ_n on $\mathcal{B}$ (not necessarily having a density). $\hat{G}_n$ is then called *smoothed* (or *perturbed*) *empirical distribution function* with the above mentioned kernel smoothing as a special case. Note that in general, i.e. for arbitrary linear metric spaces X, smoothing by convolution is its natural extension, since (cf. (6.4.1))

$$\tilde{\nu}_n(B) = n^{-1} \sum_{j \leq n} \int_X 1_B(\eta_j + y) \mu_n(dy) = n^{-1} \sum_{j \leq n} \mu_n(B - \eta_j), \quad B \in \mathcal{X}. \tag{6.4.6}$$

Now, we are going to mention at first the traditional approach towards ULLN's for smoothed empirical measures. We will formulate it for an arbitrary metric space X and for classes $\mathcal{F}$ of $\mathcal{X}$-measurable functions $f : X \longrightarrow \mathbb{R}$ being uniformly bounded, i.e. with $\sup_{f \in \mathcal{F}} \sup_{x \in X} |f(x)| \leq M < \infty$. We do not loose anything if we assume here and in the following that $M = 1$ (which means that the constant function M serves as an envelope of $\mathcal{F}$). Let $\tilde{\mathcal{F}}$ be the class of all *translates* of elements of $\mathcal{F}$, i.e.

$$\tilde{\mathcal{F}} := \{f_x : x \in X, f \in \mathcal{F}\},$$

where $f_x : X \longrightarrow \mathbb{R}$ is defined by $f_x(y) := f(x + y)$, $y \in X$. Now consider the decomposition

$$\tilde{\nu}_n - \nu = \tilde{\nu}_n - \nu \star \mu_n + \nu \star \mu_n - \nu, \tag{6.4.7}$$

where (cf. (6.4.4)) $\mathbb{E}(\tilde{\nu}_n(f)) = \nu \star \mu_n(f) \ \forall f \in \mathcal{F}$, thus $\nu \star \mu_n(f) - \nu(f)$ being the *non-stochastic bias* of $\tilde{\nu}_n(f), f \in \mathcal{F}$.

The decomposition (6.4.7) together with the assumption $\mathcal{F} = \tilde{\mathcal{F}}$ (saying that $\mathcal{F}$ *is closed under translation*) is essential for the following lemma:

6.4.8. Lemma.
Let X be a linear metric space and suppose that $\mathcal{F} = \tilde{\mathcal{F}}$. Assume further that

$$\sup_{f \in \mathcal{F}} |\nu_n(f) - \nu(f)| \xrightarrow{L_p} 0 \quad \text{and} \quad \sup_{f \in \mathcal{F}} |\nu \star \mu_n(f) - \nu(f)| \longrightarrow 0.$$

Then

$$\sup_{f \in \mathcal{F}} |\tilde{\nu}_n(f) - \nu(f)| \xrightarrow{L_p} 0 \ .$$

PROOF. According to (6.4.7) it suffices to show that

$$\sup_{f\in\mathcal{F}} |\tilde{\nu}_n(f) - \nu \star \mu_n(f)| \leq \sup_{f\in\mathcal{F}} |\nu_n(f) - \nu(f)|.$$

For this, let $g \in \mathcal{F}$ be arbitrary; then

$$\begin{aligned}
|\tilde{\nu}_n(g) - \nu \star \mu_n(g)| &= |\nu_n \star \mu_n(g) - \nu \star \mu_n(g)| \\
&= \Big| \int_X \int_X g(x+y)\nu_n(dx)\mu_n(dy) - \int_X \int_X g(x+y)\nu(dx)\mu_n(dy)\Big| \\
&= \Big| \int_X \int_X g_y(x)\nu_n(x)\mu_n(dy) - \int_X \int_X g_y(x)\nu(dx)\mu_n(dy)\Big| \\
&= \Big| \int_X (\nu_n(g_y) - \nu(g_y))\mu_n(dy)\Big| \\
&\leq \int_X \sup_{f\in\mathcal{F}} \big|\nu_n(f) - \nu(f)\big|\mu_n(dy) \\
&= \sup_{f\in\mathcal{F}} \big|\nu_n(f) - \nu(f)\big|.
\end{aligned}$$

□

Concerning the bias-term $\nu \star \mu_n - \nu = \nu \star \mu_n - \nu \star \delta_0$ ($\delta_0 =$ Dirac measure at 0) one shows in the same way that in the case $\mathcal{F} = \tilde{\mathcal{F}}$

(6.4.9) $$\sup_{f\in\mathcal{F}} |\nu \star \mu_n(f) - \nu \star \delta_0(f)| \leq \sup_{f\in\mathcal{F}} |\mu_n(f) - \delta_0(f)|,$$

where, e.g. for separable X and uniformly bounded equicontinuous classes $\mathcal{F}$ $\sup_{f\in\mathcal{F}} |\mu_n(f) - \delta_0(f)| \longrightarrow 0$ if $\mu_n \longrightarrow \delta_0$ weakly (in the sense of weak convergence of Borel p-measures in metric spaces); cf. Theorem 1.12.1 in [Va96].

The conditions of Lemma 6.4.8 are fulfilled e.g. if $X = \mathbb{R}, \mathcal{F} = \{1_{(-\infty,t]}, t \in \mathbb{R}\}$, $\mu_n \longrightarrow \delta_0$ weakly and ν being a continuous p-measure on $\mathcal{B}$ in $\mathbb{R}$. The result in this special case goes back to Winter [Win73] and Yamato [Ya73].

The disadvantage of Lemma 6.4.8 (and (6.4.9)) is that it only holds under the rather restrictive assumption $\mathcal{F} = \tilde{\mathcal{F}}$, a condition which cannot be dispensed with in general; see Example 2.2. in [Gae99].
Note also that assuming the existence of a *real-valued* envelope F of $\mathcal{F}$, the condition $\mathcal{F} = \tilde{\mathcal{F}}$ implies that $\mathcal{F}$ is uniformly bounded (i.e. $\sup_{x\in X} \sup_{f\in\mathcal{F}} |f(x)| < \infty$).

For $X = \mathbb{R}^d, d \leq 1$, Lemma 6.4.8 can be found in [Yu89] with a.s. convergence replacing convergence in the L_p-norm. Also from Yukich [Yu89] we know the following result in the case $X = \mathbb{R}^d, d \geq 1$:

6.4.10. THEOREM *(Yukich).*
Let $X = \mathbb{R}^d, d \geq 1$, and assume $\mu_n \longrightarrow \delta_0$ weakly and that

$\mathcal{F}$ is uniformly bounded

and

$$N^{[]}(\tau,\mathcal{F},\nu) < \infty \quad \textit{for all } \tau > 0 \tag{6.4.11}$$

where $N^{[]}(\tau,\mathcal{F},\nu) := \min\{m \in \mathbb{N} : \exists f_1, ..., f_m : X \longrightarrow \mathbb{R},\ f_i$ *continuous* , ν*-integrable such that for all* $f \in \mathcal{F}$ *there exist* f_i, f_j *with* $f_i \leq f \leq f_j$ *and* $\nu(f_j - f_i) < \tau\}$.

Then

$$\sup_{f\in\mathcal{F}} |\tilde{\nu}_n(f) - \nu(f)| \xrightarrow{L_p} 0 \, .$$

Here one gets rid of the assumption $\mathcal{F} = \tilde{\mathcal{F}}$, but the condition (6.4.11) on the so-called covering number *with bracketing* is rather strong: Taking $\mathcal{F}$ uniformly bounded and $\mu_n \equiv \delta_0$, Theorem 6.4.10 leaves (6.4.11) as a sufficient condition for a ULLN in the case of non-smoothed empirical measures, a sufficient condition which is far away from being necessary (Talagrand [Ta96]), especially in view of the continuity assumption on the f_i's which normally is not involved in the definition of covering numbers with bracketing. As we shall see below, Theorem 6.4.10 will follow from our ULLN 6.4.17 (cf. Lemma 6.4.22).

Next, also not imposing the assumtion $\mathcal{F} = \tilde{\mathcal{F}}$, there is a completely different way to obtain ULLN's for smoothed empirical measures via the *Random Measure Process Aprroach*, being based on our Theorem 6.1.2:

For this, note that $\tilde{\nu}_n(f)$ can be represented as (cf (6.4.3))

$$\tilde{\nu}_n(f) = \sum_{j \leq j(n)} w_{nj}(f) \cdot \xi_{nj}$$

by taking $j(n) := n, w_{nj}(f) := \int_X f(\eta_j + y)\, \mu_n(dy)$, and $\xi_{nj} := n^{-1}$. Thus, in view of the decomposition (6.4.7) together with (6.4.4) Theorem 6.1.2 yields the following ULLN. (Note that the η_j's on which the w_{nj}'s are based are i i d .)

6.4.12. THEOREM.
Let X *be a linear metric space and assume that (6.4.13) – (6.4.16) hold, where (for* $1 \leq p < \infty$*)*

$$\lim_{n\to\infty} \mathbb{E}\Big(w_{n1}(F)^p \cdot I(n^{-1} w_{n1}(F) > \delta)\Big) = 0 \quad \textit{for all } \delta > 0 \tag{6.4.13}$$

$$\sup_{n\in\mathbb{N}} \mathbb{E}\Big(w_{n1}(F) \cdot I(n^{-1} w_{n1}(F) \leq \delta_1)\Big) < \infty \quad \textit{for some } \delta_1 > 0 \tag{6.4.14}$$

(6.4.15) *For all* $\tau > 0$ *there exists* $\delta \equiv \delta(\tau) > 0$ *such that* $\big(N(\tau \cdot \mu_{n\delta}(F), \mathcal{F}, \bar{d}^{(1)}_{\mu_{n\delta}}\,)\big)_{n\in\mathbb{N}}$ *is stochastically bounded.*

$$\sup_{f\in\mathcal{F}} |\nu \star \mu_n(f) - \nu(f)| \longrightarrow 0 \, . \tag{6.4.16}$$

Then

$$\sup_{f\in\mathcal{F}} |\tilde{\nu}_n(f) - \nu(f)| \xrightarrow{L_p} 0 \,.$$

Now again (cf. the Note before 6.1.7) the condition (6.4.15) can be replaced by

$$\text{(6.4.15)}' \qquad \begin{array}{l}\text{For all } \tau > 0 \text{ there exists } \delta \equiv \delta(\tau) > 0 \text{ such that} \\ \big(N(\tau, \mathcal{F}, \bar{d}^{(1)}_{\mu_{n\delta}})\big)_{n\in\mathbb{N}} \quad \text{is stochastically bounded.}\end{array}$$

and, since

$$\bar{d}^{(1)}_{\mu_{n\delta}}(f,g) \le \bar{d}^{(1)}_{\tilde{\nu}_n}(f,g) \quad \forall f,g \in \mathcal{F}$$

even by

$$\text{(6.4.15)}'' \qquad \big(N(\tau, \mathcal{F}, \bar{d}^{(1)}_{\tilde{\nu}_n})\big)_{n\in\mathbb{N}} \;\text{ is stochastically bounded for all }\; \tau > 0$$

where $\bar{d}^{(1)}_{\tilde{\nu}_n}$ is defined by

$$\bar{d}^{(1)}_{\tilde{\nu}_n}(f,g) := \int_X \Big| \int_X (f(x+y) - g(x+y))\mu_n(dy) \Big| \nu_n(dx)$$

for $f, g \in \mathcal{F}$.

Next, take a closer look at the case when $\mathcal{F}$ is uniformly bounded. Then $\{n^{-1}w_{nj}(F) > \delta\} = \varnothing$ and $\{n^{-1}w_{nj}(F) \le \delta\} = \Omega$ for each $\delta > 0$ and large enough n. Thus (6.4.13) and (6.4.14) are fulfilled in this case. Furthermore, for every $\delta > 0$ we have $\bar{d}^{(1)}_{\mu_{n\delta}} = \bar{d}^{(1)}_{\tilde{\nu}_n}$ for large enough n. So Theorem 6.4.12 yields

6.4.17. THEOREM.
Let X be a linear metric space and suppose that $\mathcal{F}$ is uniformly bounded. Assume that (6.4.16) and (6.4.18) hold, where

$$\text{(6.4.18)} \qquad \textit{For all } \tau > 0 \quad \big(N(\tau, \mathcal{F}, \bar{d}^{(1)}_{\tilde{\nu}_n})\big)_{n\in\mathbb{N}} \;\textit{ is stochastically bounded.}$$

Then (for each $1 \le p < \infty$)

$$\sup_{f\in\mathcal{F}} |\tilde{\nu}_n(f) - \nu(f)| \xrightarrow{L_p} 0 \,.$$

Since, for uniformly bounded $\mathcal{F}$ (with $F \equiv 1$ w.l.o.g.),

$$N(\tau, \mathcal{F}, \bar{d}^{(1)}_{\tilde{\nu}_n}) \;=\; N(\tau \cdot \mu_{n\delta}(F), \mathcal{F}, \bar{d}^{(1)}_{\mu_{n\delta}})$$

for large enough n, we get from Theorem 6.4.17 together with 6.1.7 the following result in case of uniformly bounded VCGC's $\mathcal{F}$:

6.4.19. THEOREM.
Let X be a linear metric space and let $\mathcal{F}$ be a uniformly bounded VCGC. Assume uniform convergence to zero of the bias-term, i.e. $\sup_{f\in\mathcal{F}} |\nu \star \mu_n(f) - \nu(f)| \longrightarrow 0$.
Then (for each $1 \le p < \infty$)

$$\sup_{f\in\mathcal{F}} |\tilde{\nu}_n(f) - \nu(f)| \xrightarrow{L_p} 0 \, .$$

In the context of smoothed empirical measures or processes, respectively, one usually assumes $\mu_n \longrightarrow \delta_0$ weakly. Note however that in our theorems we did not assume weak convergence of $(\mu_n)_{n\in\mathbb{N}}$ in advance. This does does not follow from (6.4.16) nor does $\mu_n \longrightarrow \delta_0$ weakly imply (6.4.16) as can be seen by the following example:

6.4.20. Example.
Let $X := \mathbb{R}, \mathcal{F} := \{1_{(-\infty,t]} : t \in \mathbb{Q}\}$, $\nu = \delta_0$ and $\mu_n := \delta_{\frac{1}{n}}, n \in \mathbb{N}$. Then $\mu_n \longrightarrow \delta_0$ weakly, but (6.4.16) does not hold; in fact, $\sup_{f\in\mathcal{F}} |\nu \star \mu_n(f) - \nu(f)| \equiv 1$.

On the other hand, since (6.4.13) – (6.4.15) are fulfilled, this example also shows that (6.4.16) cannot be dispensed with, in general, for our theorems 6.4.12, 6.4.17 and 6.4.19 to hold true, since in the present case $\mathbb{E}(\sup_{f\in\mathcal{F}} |\tilde{\nu}_n(f) - \nu(f)|) \equiv 1$.

However, if $\mathcal{F}$ is "smooth" we can deduce (6.4.16) from $\mu_n \longrightarrow \delta_0$ weakly (**without assuming $\boldsymbol{\mathcal{F} = \tilde{\mathcal{F}}}$**; cf. (6.4.9) and the remarks made there).
Assuming X to be separable, we obtain the following result:

6.4.21. THEOREM.
Let X be a separable linear metric space and let $\mathcal{F}$ be a uniformly bounded equicontinuous VCGC. Suppose that $\mu_n \longrightarrow \delta_0$ weakly. Then (for each $1 \le p < \infty$)

$$\sup_{f\in\mathcal{F}} |\tilde{\nu}_n(f) - \nu(f)| \xrightarrow{L_p} 0 \, .$$

PROOF. According to Theorem 6.4.19 it suffices to verify (6.4.16), i.e. $\sup_{f\in\mathcal{F}} |\nu\star\mu_n(f) - \nu(f)| \longrightarrow 0$:
For each bounded and continuous $f : X \longrightarrow \mathbb{R}$ we have by dominated convergence that

$$\begin{aligned} \nu \star \mu_n(f) &= \int_X \int_X f(x+y)\mu_n(dy)\nu(dx) \\ &= \int_X \int_X f_x(y)\mu_n(dy)\nu(dx) \longrightarrow \int_X f_x(0)\nu(dx), \end{aligned}$$

since $\mu_n \longrightarrow \delta_0$ weakly and $f_x : X \longrightarrow \mathbb{R}$ is also bounded and continuous for all $x \in X$, where

$$\int_X f_x(0)\nu(dx) = \nu(f),$$

whence $\nu \star \mu_n \longrightarrow \nu$ weakly.
Applying now Theorem 1.12.1 in [Va96] yields (6.4.16). □

6.4.22. Lemma.

Let X be a separable linear metric space and let $\mathcal{F}$ be uniformly bounded satisfying the condition (6.4.11) in Yukich's theorem, i.e. $N^{[]}(\tau, \mathcal{F}, \nu) < \infty$ for all $\tau > 0$. Suppose that $\mu_n \longrightarrow \delta_0$ weakly. Then (6.4.18) and (6.4.16) hold true, whence Yukich's theorem comes up as a special case of Theorem 6.4.17.

PROOF. Given any $\tau > 0$ let $f_1, ..., f_m$ be continuous, ν-integrable and bounded (note that $\mathcal{F}$ is assumed to be uniformly bounded) such that for all $f \in \mathcal{F}$ there exist f_i, f_j with $f_i \le f \le f_j$ and $\nu(f_j - f_i) < \tau$ (note that $N^{[]}(\tau, \mathcal{F}, \nu) < \infty$).
Now, for all f_i, f_j with

$$[f_i, f_j] := \{f \in \mathcal{F} : f_i \le f \le f_j\} \ne \varnothing$$

choose $g_{ij} \in [f_i, f_j]$. Then, given $f \in \mathcal{F}$ and f_i, f_j with $f \in [f_i, f_j]$ and $\nu(f_j - f_i) < \tau$, we have

$$\tilde{\nu}_n(|f - g_{ij}|) \le \tilde{\nu}_n(f_j - f_i) = \nu_n \star \mu_n(f_j - f_i) \longrightarrow \nu \star \delta_0(f_j - f_i) \text{ a.s.},$$

since $\mu_n \longrightarrow \delta_0$ weakly and (cf. e.g. [Gae79], Section 1.5) $\nu_n \longrightarrow \nu$ weakly a.s.; note that $f_j - f_i$ is bounded and continuous.
Since $\nu \star \delta_0(f_j - f_i) = \nu(f_j - f_i) < \tau$, it follows that

$$\limsup_{n\to\infty} N(\tau, \mathcal{F}, d^{(1)}_{\tilde{\nu}_n}) \le m^2 \quad \text{a.s.},$$

whence $\Big(N(\tau, \mathcal{F}, d^{(1)}_{\tilde{\nu}_n})\Big)_{n\in\mathbb{N}}$ is stochastically bounded and therefore also $\Big(N(\tau, \mathcal{F}, \bar{d}^{(1)}_{\tilde{\nu}_n})\Big)_{n\in\mathbb{N}}$, since

$$\bar{d}^{(1)}_{\tilde{\nu}_n}(f, g) \le d^{(1)}_{\tilde{\nu}_n}(f, g) := \tilde{\nu}_n(|f - g|) \quad \text{ for } f, g \in \mathcal{F}.$$

So we conclude that (6.4.18) holds.

Next, from $f \in [f_i, f_j]$ and $\nu(f_j - f_i) < \tau$ we can also conclude that

$$\begin{aligned}|\nu \star \mu_n(f) - \nu(f)| \le \max\{&|\nu \star \mu_n(f_j) - \nu(f_j)| + |\nu(f_j) - \nu(f)|,\\ &|\nu \star \mu_n(f_i) - \nu(f_i)| + |\nu(f_i) - \nu(f)|\},\end{aligned}$$

thus

$$\sup_{f\in\mathcal{F}} |\nu \star \mu_n(f) - \nu(f)| \le \max\{|\nu \star \mu_n(f_j) - \nu(f_j)| : 1 \le j \le m\} + \tau.$$

But $\nu \star \mu_n(f_j) - \nu(f_j) \longrightarrow 0$ for all $j = 1, ..., m$, since $\nu \star \mu_n \longrightarrow \nu$ weakly and the f_j's are bounded and continuous. So we get

$$\limsup_{n\to\infty} \sup_{f\in\mathcal{F}} |\nu \star \mu_n(f) - \nu(f)| \le \tau.$$

Since $\tau > 0$ was arbitrary, this gives (6.4.16). □

Finally, in the non-smoothed case (i.e. with $\mu_n \equiv \delta_0$) one has the following deep result on empirical measures ν_n which we deduce from Talagrand [Ta96]; here X is not required to be a linear metric space.

6.4.23. THEOREM *(Talagrand).*
Let $(X, \mathcal{X}, \nu)$ be a complete p-space and $\mathcal{F}$ be a uniformly bounded class of $\mathcal{X}$-measurable functions $f : X \longrightarrow \mathbb{R}$. Then the condition

(6.4.24) $$\big(N(\tau, \mathcal{F}, d^{(1)}_{\nu_n}\,)\big)_{n\in\mathbb{N}} \quad \textit{is stochastically bounded for all } \tau > 0$$

(with $d^{(1)}_{\nu_n}(f,g) := \nu_n(|f-g|)$, $f, g \in \mathcal{F}$) is necessary and sufficient for

$$\sup_{f\in\mathcal{F}} |\nu_n(f) - \nu(f)| \xrightarrow{L_1} 0.$$

In view of this result it is tempting to see what comes up in the smoothed case. The following result is contained in [Gae00]:

6.4.25. THEOREM.
Let X be a linear metric space endowed with its Borel σ-field $\mathcal{X}$ such that $(X, \mathcal{X}, \nu)$ is complete, and let $\mathcal{F}$ be a uniformly bounded class of $\mathcal{X}$-measurable functions $f : X \longrightarrow \mathbb{R}$ which is closed under translations, i.e. $\tilde{\mathcal{F}} = \mathcal{F}$. Suppose that $\sup_{f\in\mathcal{F}} |\mu_n(f) - f(0)| \longrightarrow 0$. Then the following statements are equivalent:

a) $\big(N(\tau, \mathcal{F}, d^{(1)}_{\nu_n}\,)\big)_{n\in\mathbb{N}}$ *is stochastically bounded for all $\tau > 0$*
b) $\big(N(\tau, \mathcal{F}, \bar{d}^{(1)}_{\tilde{\nu}_n})\big)_{n\in\mathbb{N}}$ *is stochastically bounded for all $\tau > 0$*
c) $\sup_{f\in\mathcal{F}} |\tilde{\nu}_n(f) - \nu(f)| \xrightarrow{L_1} 0$
d) $\sup_{f\in\mathcal{F}} |\nu_n(f) - \nu(f)| \xrightarrow{L_1} 0.$

6.5 A more general ULLN

As in Section 5.4, let us consider the general model of stochastic processes

$$S_n = (S_n(t))_{t\in T}, \qquad n \in \mathbb{N},$$

indexed by an arbitrary parameter space T (again supposed to be countable to avoid measurability considerations) given by

$$S_n(t) := \sum_{j\le j(n)} \eta_{nj}(t), \quad t \in T,$$

where $\eta_{nj} = (\eta_{nj}(t))_{t\in T}, 1 \le j \le j(n), n \in \mathbb{N}$, is a triangular array of rowwise independent stochastic processes (indexed by T) and with $j(n) \to \infty$ as $n \to \infty$.
We are going to present conditions under which (cf. (2.1.5))

(6.5.1) $$\lim_{n\to\infty} \mathbb{E}(||S_n - \mathbb{E}(S_n)||) = 0,$$

where again $||S_n - \mathbb{E}(S_n)|| = ||S_n - \mathbb{E}(S_n)||_T := \sup_{t\in T} |S_n(t) - \mathbb{E}(S_n(t))|$.

According to the first inequality in 5.4.4, being valid also for not necessarily identically distributed but still independent η_{nj} (cf. its proof based on the two Symmetrization Lemmata 5.4.1 ii) and 5.4.2 ii)) (6.5.1) is shown by verifying

$$\lim_{n\to\infty} \mathbb{E}(||S_n^0||) = 0, \tag{6.5.2}$$

where $S_n^0 := \sum\limits_{j\le j(n)} \varepsilon_j \eta_{nj}$ with a Rademacher sequence $(\varepsilon_j)_{j\in\mathbb{N}}$ being independent of (η_{nj}).

To verify (6.5.2) the rv $||S_n^0|| \equiv ||S_n^0||_T := \sup\limits_{t\in T} |S_n^0(t)|$ will be approximated by $||S_n^0||_{\hat{T}_n}$, where $\hat{T}_n$ will be a (usually finite) subset of T. Then, with the random pseudo-metric

$$\hat{\rho}_n(s,t) \equiv \hat{\rho}_n(s,t,\eta_{n1},...,\eta_{nj(n)}) := \sum_{j\le j(n)} |\eta_{nj}(s) - \eta_{nj}(t)|, \quad s,t \in T, \tag{6.5.3}$$

we get

$$\begin{aligned} ||S_n^0|| &\le \sup_{t\in T} \inf_{s\in\hat{T}_n} \big(|S_n^0(s)| + |S_n^0(t) - S_n^0(s)|\big) \\ &\le ||S_n^0||_{\hat{T}_n} + \sup_{t\in T} \inf_{s\in\hat{T}_n} |S_n^0(t) - S_n^0(s)| \\ &\le ||S_n^0||_{\hat{T}_n} + \sup_{t\in T} \hat{\rho}_n(t,\hat{T}_n), \end{aligned} \tag{6.5.4}$$

where $\hat{\rho}_n(t,\hat{T}_n) := \inf\limits_{s\in\hat{T}_n} \hat{\rho}_n(t,s)$.

Remember below the Definition 4.1.1 of covering numbers $N(u,T,\hat{\rho}_n)$ being the minimal number of closed $\hat{\rho}_n$ - balls with radius u which cover T.

6.5.5. THEOREM.

Assume the following conditions (6.5.6) and (6.5.7), where $(\delta_n)_{n\in\mathbb{N}}$ is some sequence of positive real numbers with $\delta_n \to 0$:

$$\mathbb{E}\Big(\sum_{j\le j(n)} ||\eta_{nj}||\Big) = O(1) \quad \textit{and} \quad \mathbb{E}\Big(\sum_{j\le j(n)} ||\eta_{nj}|| \cdot I(||\eta_{nj}|| > \delta_n)\Big) = o(1) \tag{6.5.6}$$

$$\log N(u,T,\hat{\rho}_n)^* = o_{\mathbb{P}}(\delta_n^{-1}) \quad \forall\, u > 0. \tag{6.5.7}$$

Then (6.5.1), i.e. $\lim\limits_{n\to\infty} \mathbb{E}(||S_n - \mathbb{E}(S_n)||) = 0$ *holds true.*

The following two special cases are included:

Assume (with $K_n := n\delta_n\,, \delta_n \to 0$)

$$\mathbb{E}\Big(\sum_{j\le j(n)} ||\eta_{nj}||\Big) = O(1) \quad \text{and} \quad \lim_{n\wedge K_n\to\infty} \mathbb{E}\Big(\sum_{j\le j(n)} ||\eta_{nj}|| \cdot I(||\eta_{nj}|| > \frac{K_n}{n})\Big) = 0 \tag{6.5.6'}$$

and

(6.5.7') $$\log N(u,T,\hat\rho_n)^* = o_{\mathbb{P}}(\sqrt{n}) \quad \forall\, u>0.$$

Then (6.5.1) holds true.

Assume on the other hand

(6.5.6") $$\mathbb{E}\Big(\sum_{j\le j(n)} ||\eta_{nj}||\Big) = O(1) \quad \text{and} \quad \mathbb{E}\Big(\sum_{j\le j(n)} ||\eta_{nj}||\cdot I(||\eta_{nj}||>u)\Big) = o(1) \quad \forall\, u>0$$

and that $\log N(u,T,\hat\rho_n)$ is stochastically bounded, i.e.

(6.5.7") $$logN(u,T,\hat\rho_n)^* = O_{\mathbb{P}}(1) \quad \forall\, u>0.$$

Then (6.5.1) holds true.

PROOF of Theorem 6.5.5. Let $\tilde\eta_{nj} := \eta_{nj}\cdot I(||\eta_{nj}||\le\delta_n)$ and $\tilde S_n := \sum\limits_{j\le j(n)} \tilde\eta_{nj}$. Then

$$\begin{aligned}
\mathbb{E}(\underbrace{||S_n-\mathbb{E}(S_n)-(\tilde S_n-\mathbb{E}(\tilde S_n)||}_{\le ||S_n-\tilde S_n||+\underbrace{||\mathbb{E}(\tilde S_n)-\mathbb{E}(S_n)||}_{\le\mathbb{E}(||S_n-\tilde S_n||)}}) &\le 2\,\mathbb{E}(||S_n-\tilde S_n||) \\
&= 2\,\mathbb{E}\Big(||\sum_{j\le j(n)} \eta_{nj}\cdot I(||\eta_{nj}||>\delta_n)||\Big) \\
&\le 2\sum_{j\le j(n)} \mathbb{E}(||\eta_{nj}||\cdot I(||\eta_{nj}||>\delta_n)) \underset{(6.5.6)}{=} o(1).
\end{aligned}$$

In addition, $\forall\, n\in\mathbb{N}$ and $\forall\, s,t\in T$

$$\hat{\tilde{\rho_n}}(s,t) := \hat\rho_n(s,t,\tilde\eta_{n1},...,\tilde\eta_{nj(n)}) \le \hat\rho_n(s,t,\eta_{n1},...,\eta_{nj(n)}) = \hat\rho_n(s,t),$$

whence (cf. Def. 4.1.1)

$$N(u,T,\hat{\tilde{\rho_n}}) \le N(u,T,\hat\rho_n) \qquad \forall\, u>0.$$

Thus, in view of (6.5.7), we may and do assume w.l.o.g. that $\forall\, n\in\mathbb{N}$ and $1\le j\le j(n)$

(*) $$||\eta_{nj}|| \le \delta_n.$$

Now, as already remarked above, (6.5.1) will be shown by verifying (6.5.2). For this,

$$\begin{aligned}
\mathbb{E}(||S_n^0||^2) &\le \mathbb{E}\Big(\sum_{j,j'=1}^{j(n)} ||\eta_{nj}|| \cdot ||\eta_{nj'}||\Big) \\
&= \sum_{j\le j(n)} \mathbb{E}(||\eta_{nj}||^2) + \sum_{j,j'=1}^{j(n)} I(j\neq j')\underbrace{\mathbb{E}(||\eta_{nj}|| \cdot ||\eta_{nj'}||)}_{=\mathbb{E}(||\eta_{nj}||)\cdot\mathbb{E}(||\eta_{nj'})||} \\
&\underset{(*)}{\le} \delta_n \cdot \mathbb{E}\Big(\sum_{j\le j(n)} ||\eta_{nj}||\Big) + \Big(\mathbb{E}(\sum_{j\le j(n)} ||\eta_{nj}||)\Big)^2 \\
&\underset{(6.5.6)}{=} O(1).
\end{aligned}$$

Thus, for arbitrary $\varepsilon > 0$,

$$\begin{aligned}
\mathbb{E}(||S_n^0||) &\le \varepsilon + \mathbb{E}(||S_n^0|| \cdot I(||S_n^0|| > \varepsilon)) \\
&\underset{\text{Cauchy-Schwarz}}{\le} \varepsilon + \sqrt{\mathbb{E}(||S_n^0||^2) \cdot \mathbb{P}(||S_n^0|| > \varepsilon)} \\
&= \varepsilon + O(\sqrt{\mathbb{P}(||S_n^0|| > \varepsilon)}.
\end{aligned}$$

Therefore, to verify (6.5.2) it suffices to show that

(**) $$\lim_{n\to\infty} \mathbb{P}(||S_n^0|| > \varepsilon) = 0 \qquad \forall\, \varepsilon > 0.$$

For this, given an arbitrary $\varepsilon > 0$, let $\hat{T}_n \equiv \hat{T}_n(\eta_{n1}, ..., \eta_{nj(n)})$ be a random subset of T such that $|\hat{T}_n| = N(\frac{\varepsilon}{2}, T, \hat{\rho}_n)$ and $\sup\limits_{t\in T} \hat{\rho}_n(t, \hat{T}_n) \le \frac{\varepsilon}{2}$.

Then, by (6.5.4)

$$||S_n^0|| \le \frac{\varepsilon}{2} + ||S_n^0||_{\hat{T}_n}.$$

Now,

$$\mathbb{P}(||S_n^0|| \ge \varepsilon) = \mathbb{E}(I(||S_n^0|| \ge \varepsilon)) = \mathbb{E}(\underbrace{\mathbb{E}(I(||S_n^0|| \ge \varepsilon)|\eta_{n1}, ..., \eta_{nj(n)})}_{=:\zeta_n\le 1\mathbb{P}-\text{a.s.}}),$$

and so, by Lebesgue's dominated convergence theorem, (**) will be proved by showing

(***) $$\zeta_n \xrightarrow{\mathbb{P}} 0.$$

For this, using Corollary 5.5.4, we get $\forall\,\varepsilon>0$

$$\begin{aligned}
\zeta_n &= \mathbb{P}\Big(\{||S_n^0||\geq\varepsilon\}|\eta_{n1},...,\eta_{nj(n)}\Big)\\
&\leq \mathbb{P}\Big(\{||S_n^0||_{\hat{T}_n}\geq\frac{\varepsilon}{2}\}\,|\,\eta_{n1},...,\eta_{nj(n)}\Big)\\
&\leq \sum_{t\in\hat{T}_n}\mathbb{P}(\{|S_n^0(t)|\geq\frac{\varepsilon}{2}\}\,|\,\eta_{n1},...,\eta_{nj(n)})\\
&\underset{(|\hat{T}_n|=N(\frac{\varepsilon}{2},T,\hat{\rho}_n))}{\leq} N(\frac{\varepsilon}{2},T,\hat{\rho}_n)\cdot\sup_{t\in T}\mathbb{P}(\{|S_n^0(t)|\geq\frac{\varepsilon}{2}\}\,|\,\eta_{n1},...,\eta_{nj(n)})\\
&\overset{(!)}{\leq} 2N(\frac{\varepsilon}{2},T,\hat{\rho}_n)\sup_{t\in T}\exp\Big(-\frac{\varepsilon^2}{8\sum\limits_{j\leq j(n)}\eta_{nj}^2(t)}\Big)\\
&\underset{(*)}{\leq} 2N(\frac{\varepsilon}{2},T,\hat{\rho}_n)\exp\Big(-\frac{\varepsilon^2}{8\delta_n\sum\limits_{j\leq j(n)}||\eta_{nj}||}\Big)\\
&= 2\,\exp\Big[\underbrace{-\frac{1}{\delta_n}}_{\to-\infty}\Big(\varepsilon^2\,\underbrace{(\,8\sum_{j\leq j(n)}||\eta_{nj}||\,)^{-1}}_{=O_{\mathbb{P}}(1)\text{ by (6.5.6)}}-\underbrace{\delta_n\log N(\frac{\varepsilon}{2},T,\hat{\rho}_n)}_{=o_{\mathbb{P}}(1)\text{ by (6.5.7)}}\Big)\Big]\\
&\xrightarrow{\mathbb{P}} 0.
\end{aligned}$$

□

As to (!): $(\varepsilon_j),(\eta_{nj})$ are independent, so [Gae77], 5.3.22 can be applied:

$$\begin{aligned}
\mathbb{P}\Big(\{|S_n^0(t)|\geq\frac{\varepsilon}{2}\}\,|\,\eta_{n1},...,\eta_{nj(n)}\Big) &= \mathbb{P}\Big(\{|\sum_{j\leq j(n)}\varepsilon_j\eta_{nj}(t)|\geq\frac{\varepsilon}{2}\}\,|\,\eta_{n1},...,\eta_{nj(n)}\Big)\\
&= \mathbb{P}_\varepsilon\Big(\{|\sum_{j\leq j(n)}\varepsilon_j\eta_{nj}(t,\omega)|\geq\frac{\varepsilon}{2}\}\Big)\quad\text{for }\mathbb{P}-\text{ a.a. }\omega\\
&\underset{(5.5.4\text{ with }x=\frac{\varepsilon}{2})}{\leq} 2\,\exp\Big(\frac{\varepsilon^2}{8\sum\limits_{j\leq j(n)}\eta_{nj}^2(t,\omega)}\Big)\quad\text{for }\mathbb{P}-\text{ a.a. }\omega.
\end{aligned}$$

Strengthening the condition (6.5.7) will lead to a refinement of Theorem 6.5.5 with an application to density estimation as carried out in [Due00] Sections 8.4 and 8.5.
For this one needs the following lemma, where the basic idea of its proof will be Le Cam's "square root trick" via Giné and Zinn (cf. [Cam83], [Gi84]).

6.5.8. Lemma.
Assume $||\eta_{nj}||\leq\delta_n\ \ \forall\,1\leq j\leq j(n)$ *with* $n\in\mathbb{N}$ *being arbitrary but fixed. Let*

$$V_n:=\sup_{t\in T}\mathbb{E}\Big(\sum_{j\leq j(n)}\eta_{nj}^2\Big).$$

Then $\mathbb{P}\Big(\sup\limits_{t\in T}\sum\limits_{j\leq j(n)}\eta_{nj}^2\geq 8\tau\Big)\leq\dfrac{2}{1-V_n/\tau}\,\mathbb{E}\Big(\exp[\log N(\frac{\tau}{32\delta_n},T,\hat{\rho}_n)-\frac{\tau}{8\delta_n^2}]\wedge 1\Big)\qquad\forall\,\tau>V_n.$

PROOF: Let $(\varepsilon_j)_{j\in\mathbb{N}}$ be a Rademacher sequence, independent of (η_{nj}), and $\forall\, n \in \mathbb{N}$ $J_1 := \{j \le j(n) : \varepsilon_j = 1\}$ and $J_2 := \{j \le j(n) : \varepsilon_j = -1\} = \{1, ..., j(n)\}\backslash J_1$,

$$S_K := \sum_{j\in K} \eta_{nj}^2 \qquad \text{for } K \subset \{1, ..., j(n)\},$$

and $S := S_{\{1,...,j(n)\}} = \sum_{j\le j(n)} \eta_{nj}^2$. Since $\mathcal{L}\{S_{J_1}\} = \mathcal{L}\{S_{J_2}\}$ we have

$$\begin{aligned}\mathbb{P}\Big(\sup_{t\in T} \sum_{j\le j(n)} \eta_{nj}^2(t) \ge 8\tau\Big) &= \mathbb{P}(||S|| \ge 8\tau) \\ &\le 2\,\mathbb{P}(||S_{J_1}|| \ge 4\tau) = \; 2\,\mathbb{P}(||\sqrt{S_{J_1}}|| \ge 2\sqrt{\tau}).\end{aligned}$$

Now, for any $t \in T$,

$$(*) \qquad \begin{aligned}\mathbb{P}(\{\sqrt{S_{J_1}(t)} \ge \sqrt{\tau}\}\,|\,\varepsilon_1, ..., \varepsilon_{j(n)}) &\underset{\text{(Markov's inequality)}}{\le} \frac{\mathbb{E}(S_{J_1}(t)\,|\,\varepsilon_1, ..., \varepsilon_{j(n)})}{\tau} \\ &\le \frac{\mathbb{E}(S(t))}{\tau} \le \frac{V_n}{\tau},\end{aligned}$$

since

$$\begin{aligned}\mathbb{E}(S_{J_1}(t)\,|\,\varepsilon_1, ..., \varepsilon_{j(n)}) &= \mathbb{E}\Big(\sum_{j\in J_1} \eta_{nj}^2(t)\,|\,\varepsilon_1, ..., \varepsilon_{j(n)}\Big) \\ &= \sum_{j\in J_1} \mathbb{E}(\eta_{nj}^2(t)\,|\,\varepsilon_1, ..., \varepsilon_{j(n)}) \\ &= \sum_{j\in J_1} \mathbb{E}(\eta_{nj}^2(t)) \le \sum_{j\le j(n)} \mathbb{E}(\eta_{nj}^2(t)) \\ &= \underbrace{\mathbb{E}\Big(\sum_{j\le j(n)} \eta_{nj}^2(t)\Big)}_{=\mathbb{E}(S(t))} \underset{\text{(by def. of } V_n)}{\le} V_n.\end{aligned}$$

Next, since $\sqrt{S_{J_1}}$ and $\sqrt{S_{J_2}}$ are independent, given $\varepsilon_1, ..., \varepsilon_{j(n)}$, we can apply Lemma 5.4.1 i) w.r.t. the conditional distribution on the lhs. of (*) to obtain

$$(**) \qquad \mathbb{P}(||\sqrt{S_{J_1}}|| \ge 2\sqrt{\tau}) \le \frac{1}{1 - V_n/\tau}\,\mathbb{P}(||\sqrt{S_{J_1}} - \sqrt{S_{J_2}}|| \ge \sqrt{\tau}).$$

Now, $\forall\, s,t\in T$

$$\begin{aligned}
\Big|\Big(\sqrt{S_{J_1}(s)}-\sqrt{S_{J_2}(s)}\Big)-\Big(\sqrt{S_{J_1}(t)}-\sqrt{S_{J_2}(t)}\Big)\Big| &\le \Big|\sqrt{S_{J_1}(s)}-\sqrt{S_{J_1}(t)}\Big|+\Big|\sqrt{S_{J_2}(s)}-\sqrt{S_{J_2}(t)}\Big| \\
&\le \Big(\sum_{j\in J_1}(\eta_{nj}(s)-\eta_{nj}(t))^2\Big)^{1/2}+\Big(\sum_{j\in J_2}(\eta_{nj}(s)-\eta_{nj}(t))^2\Big)^{1/2} \\
&\le \Big(4\sum_{j\le j(n)}(\eta_{nj}(s)-\eta_{nj}(t))^2\Big)^{1/2} \\
&\underset{(||\eta_{nj}||\le\delta_n)}{\le} \Big(8\,\delta_n \underbrace{\sum_{j\le j(n)}|\eta_{nj}(s)-\eta_{nj}(t)|}_{=\hat\rho_n(t,s)\text{ (by def. of }\hat\rho_n\text{ in (6.5.3))}}\Big)^{1/2}.
\end{aligned}$$

Therefore, $\forall\, s,t\in T$

$$\Big|\Big(\sqrt{S_{J_1}(s)}-\sqrt{S_{J_2}(s)}\Big)-\Big(\sqrt{S_{J_1}(t)}-\sqrt{S_{J_2}(t)}\Big)\Big|\le\sqrt{8\delta_n\hat\rho_n(t,s)}.$$

So, choosing $\hat T_n=\hat T_n(\eta_{n1},...,\eta_{nj(n)})\subset T$ such that $|\hat T_n|=N(\frac{\tau}{32\delta_n},T,\hat\rho_n)$ and $\sup\limits_{t\in T}\hat\rho_n(t,\hat T_n)\le\frac{\tau}{32\delta_n}$, then, in the same way as in (6.5.4) one gets

$$||\sqrt{S_{J_1}}-\sqrt{S_{J_2}}||\le||\sqrt{S_{J_1}}-\sqrt{S_{J_2}}||_{\hat T_n}+\sup_{t\in T}\underbrace{\sqrt{8\delta_n\hat\rho_n(t,\hat T_n)}}_{\le\sqrt{\frac{\tau}{4}}=\frac{\sqrt{\tau}}{2}}.$$

Therefore, we can proceed as in the proof of Theorem 6.5.5 using Corollary 5.5.4 to obtain

$$\begin{aligned}
\mathbb{P}(\{||\sqrt{S_{J_1}}-\sqrt{S_{J_2}}||\ge\sqrt{\tau}\}\,|\,\eta_{n1},...,\eta_{nj(n)}) &\le \mathbb{P}(\{||\sqrt{S_{J_1}}-\sqrt{S_{J_2}}||_{\hat T_n}\ge\frac{\sqrt{\tau}}{2}\}\,|\,\eta_{n1},...,\eta_{nj(n)}) \\
&\le \sum_{t\in\hat T_n}\mathbb{P}(\{|\sqrt{S_{J_1}(t)}-\sqrt{S_{J_2}(t)}|\ge\frac{\sqrt{\tau}}{2}\}\,|\,\eta_{n1},...,\eta_{nj(n)}) \\
&\le \sum_{t\in\hat T_n}\mathbb{P}(\{|S_{J_1}(t)-S_{J_2}(t)|\ge\frac{\sqrt{\tau}}{2}\Big(\sqrt{S_{J_1}(t)}+\sqrt{S_{J_2}(t)}\}\Big)\,|\,\eta_{n1},...,\eta_{nj(n)}) \\
&\le \sum_{t\in\hat T_n}\mathbb{P}(\{|S_{J_1}(t)-S_{J_2}(t)|\ge\frac{\sqrt{\tau}}{2}\sqrt{S(t)}\,|\,\eta_{n1},...,\eta_{nj(n)}) \\
&= \sum_{t\in\hat T_n}\mathbb{P}(\{|\sum_{j\le j(n)}\varepsilon_j\eta_{nj}^2(t)|\ge\frac{\sqrt{\tau}}{2}\sqrt{S(t)}\,|\,\eta_{n1},...,\eta_{nj(n)}) \\
&\underset{\text{Corollary 5.5.4}}{\le} 2\sum_{t\in\hat T_n}\exp\Big(-\frac{S(t)\tau}{8\sum\limits_{j\le j(n)}\eta_{nj}^4(t)}\Big) \\
&\underset{(||\eta_{nj}||\le\delta_n)}{\le} 2\,|\hat T_n|\,\exp\Big(-\frac{\tau}{8\delta_n^2}\Big)
\end{aligned}$$

from which the result follows after integration w.r.t. (η_{nj}). $\square$

6.5.9. Theorem.
Under the assumptions of Lemma 6.5.8 one has for $S_n := \sum_{j\le j(n)} \eta_{nj}$ *that*

$$\mathbb{P}(||S_n - \mathbb{E}(S_n)|| \ge 3\eta) \le 16\,\mathbb{E}\Big(\exp[\log N(\frac{\tau}{32\delta_n}, T, \hat\rho_n) - \frac{\tau}{8\delta_n^2}] \wedge 1\Big)$$
$$+ 8\,\mathbb{E}\Big(\exp[\log N(\frac{\eta}{2}, T, \hat\rho_n) - \frac{\eta^2}{64\tau}] \wedge 1\Big)$$

for all $\eta \ge \sqrt{2V_n}$ *and all* $\tau > 2V_n$.

PROOF: Since $\forall\, t \in T$

$$Var(S_n(t)) = \sum_{j\le j(n)} Var(\eta_{nj}(t)) \le \sum_{j\le j(n)} \mathbb{E}(\eta_{nj}^2(t)) \le V_n := \sup_{t\in T} \mathbb{E}\Big(\sum_{j\le j(n)} \eta_{nj}^2(t)\Big),$$

we have $\forall\, t \in T$ and $S_n'(t) := \sum_{j\le j(n)} \eta_{nj}'(t)$, $\;(\eta_{nj}' = (\eta_{nj}'(t))_{t\in T}$ being independent versions of the processes η_{nj} as in Lemma 5.4.2)

$$\mathbb{P}(|S_n'(t) - \mathbb{E}(S_n'(t))| > \eta) \underset{\text{(Tschebyscheff-Ineq.)}}{\le} \frac{Var(S_n'(t))}{\eta^2} = \frac{Var(S_n(t))}{\eta^2} \le \frac{V_n}{\eta^2},$$

whence

$$\mathbb{P}(|S_n'(t) - \mathbb{E}(S_n'(t))| \le \eta) \ge 1 - \frac{V_n}{\eta^2}.$$

Thus an application of Lemma 5.4.1 i) (with $S_n - \mathbb{E}(S_n)$ and $S_n' - \mathbb{E}(S_n') = S_n' - \mathbb{E}(S_n)$ instead of S and S', respectively), together with Lemma 5.4.2 i) yields

$$\begin{aligned}\mathbb{P}(||S_n - \mathbb{E}(S_n)|| \ge 3\eta) &\le \frac{1}{1 - V_n/\eta^2}\,\mathbb{P}(||S_n - S_n'|| \ge 2\eta)\\ &\le \frac{2}{1 - V_n/\eta^2}\,\mathbb{P}(||S_n^0|| \ge \eta).\end{aligned}$$

Now, as in the proof of Theorem 6.5.5 one gets

$$(*)\qquad \mathbb{P}(\{||S_n^0|| \ge \eta\}\,|\,\eta_{n1}, ..., \eta_{nj(n)}) \le 2N(\frac{\eta}{2}, T, \hat\rho_n) \exp\Big(- \frac{\eta^2}{8X_n}\Big),$$

where $X_n := \sup_{t\in T} \sum_{j\le j(n)} \eta_{nj}^2(t)$. Therefore

$$\begin{aligned}\mathbb{P}(||S_n^0|| \ge \eta) &= \mathbb{E}(1_{\{||S_n^0||\ge\eta\}})\\ &= \mathbb{E}(1_{\{||S_n^0||\ge\eta\}} \cdot 1_{\{X_n \ge 8\tau\}}) + \mathbb{E}(1_{\{||S_n^0||\ge\eta\}} \cdot 1_{\{X_n < 8\tau\}})\\ &\le \mathbb{P}(X_n \ge 8\tau) + \mathbb{E}(1_{\{||S_n^0||\ge\eta\}} \cdot 1_{\{X_n < 8\tau\}}),\end{aligned}$$

where

$$\mathbb{P}(X_n \geq 8\tau) \leq \frac{2}{1 - V_n/\tau}\, \mathbb{E}\Big(\exp[\log N(\frac{\tau}{32\delta_n}, T, \hat{\rho}_n) - \frac{\tau}{8\delta_n^2}] \wedge 1\Big)$$

according to Lemma 6.5.8 and

$$\begin{aligned}\mathbb{E}(1_{\{||S_n^0|| \geq \eta\}} \cdot 1_{\{X_n < 8\tau\}}) &= \mathbb{E}\Big(\underbrace{\mathbb{E}(1_{\{||S_n^0|| \geq \eta\}} \cdot 1_{\{X_n < 8\tau\}} \,|\, \eta_{n1}, ..., \eta_{nj(n)})}_{=1_{\{X_n<8\tau\}} \cdot \mathbb{E}(1_{\{||S_n^0||\geq\eta\}}|\eta_{n1},...,\eta_{nj(n)}) \wedge 1} \Big) \\ &\leq 2\, \mathbb{E}\Big(\exp[\log N(\frac{\eta}{2}, T, \hat{\rho}_n) - \frac{\eta^2}{64\tau}] \wedge 1\Big).\end{aligned}$$

Note that

$$\begin{aligned}\mathbb{E}(1_{\{||S_n^0||\geq\eta\}}|\eta_{n1}, ..., \eta_{nj(n)}) \wedge 1 &= \mathbb{P}(\{||S_n^0|| \geq \eta\}|\eta_{n1}, ..., \eta_{nj(n)}) \wedge 1 \\ &\underset{(*)}{\leq} [2\, N(\frac{\eta}{2}, T, \hat{\rho}_n) \exp(-\frac{\eta^2}{8X_n})] \wedge 1 = 2\, \exp[\log N(\frac{\eta}{2}, T, \hat{\rho}_n) \exp(-\frac{\eta^2}{8X_n})] \wedge 1 \\ &\underset{\text{on } \{X_n<8\tau\}}{\leq} 2\, \exp[\log N(\frac{\eta}{2}, T, \hat{\rho}_n) - \frac{\eta^2}{64\tau}] \wedge 1.\end{aligned}$$

It follows that

$$\begin{aligned}\frac{2}{1 - V_n/\eta^2}\, \mathbb{P}(||S_n^0|| \geq \eta) &\leq \frac{2}{1 - V_n/\eta^2}\, \frac{2}{1 - V_n/\tau}\, \mathbb{E}\Big(\exp[\log N(\frac{\tau}{32\delta_n}, T, \hat{\rho}_n) - \frac{\tau}{8\delta_n^2}] \wedge 1\Big) \\ &\quad + \frac{2}{1 - V_n/\eta^2}\, 2\, \mathbb{E}\Big(\exp[\log N(\frac{\eta}{2}, T, \hat{\rho}_n) - \frac{\eta^2}{64\tau}] \wedge 1\Big),\end{aligned}$$

where for all η, τ with $\eta^2 \wedge \tau \geq 2V_n$ as assumed, $\frac{2}{1-V_n/\eta^2}\, \frac{2}{1-V_n/\tau} \leq 16$, and $\frac{4}{1-V_n/\eta^2} \leq 8$, which proves Theorem 6.5.9. $\square$

6.5.10. Corollary.
Let the assumptions of Lemma 6.5.8 be fulfilled with $\delta_n := \frac{1}{n}, n \in \mathbb{N}$, and assume that for certain constants $0 < A < \infty, 0 < B < \infty$, and $\forall\, n \in \mathbb{N}$

(6.5.11) $$N(\varepsilon, T, \hat{\rho}_n) \leq A \cdot \varepsilon^{-B} \qquad \mathbb{P} - a.s.\ \forall\, 0 < \varepsilon \leq 1.$$

Then, for sufficiently large $K > 0$

$$\sum_{n\geq 1} \mathbb{P}\Big(\{||S_n - \mathbb{E}(S_n)|| \geq K(\sqrt{\log(n)\, V_n} \vee \frac{\log n}{n})\}\Big) < \infty.$$

PROOF: The assertion will follow from Theorem 6.5.9 by taking

$$\begin{aligned}\tau = \tau_n &:= \gamma_\tau \cdot (V_n \vee \frac{\log n}{n^2}), \\ \eta = \eta_n &:= \sqrt{\gamma_\eta \log(n) \cdot \tau_n} = \sqrt{\gamma_\tau\, \gamma_\eta} \sqrt{\log(n)\, V_n \vee \frac{\log n}{n}}\end{aligned}$$

with sufficiently large constants $\gamma_\tau\,\gamma_\eta > 0$; then the assumption $\eta_n^2 \wedge \tau_n \geq 2V_n$ in Theorem 6.5.9 will be fulfilled for all $n \geq 2$ (by choosing $\gamma_\tau > 2$ and $\gamma_\eta > 2/\log 2$), whence an application of Theorem 6.5.9 yields the result since according to (6.5.11) $\mathbb{P}$-a.s. $\forall\, n \in \mathbb{N}$ and sufficiently large γ_τ, γ_η

$$\log N(\frac{n\tau_n}{32}, T, \hat{\rho}_n) - \frac{n^2\tau_n}{8} \leq const. - B\log(\frac{\log n}{n}) - \frac{\gamma_\tau \log n}{8} \leq const. + (B - \frac{\gamma_\tau}{8})\log n,$$

and

$$\log N(\frac{\eta_n}{2}, T, \hat{\rho}_n) - \frac{\eta_n^2}{64\tau_n} \leq const. + (B - \frac{\gamma_\eta}{64})\log n,$$

where for sufficiently large γ_τ and γ_η $\quad B - \frac{\gamma_\tau}{8} < -1$ and $B - \frac{\gamma_\eta}{64} < -1$ whence for sufficiently large n and appropriate $a > 1$ $\quad\mathbb{P}$-a.s.

$$\exp[\log N(\frac{n\tau_n}{32}, T, \hat{\rho}_n) - \frac{n^2\tau_n}{8}] \wedge 1 \leq n^{-a}$$

as well as

$$\exp[\log N(\frac{\eta_n}{2}, T, \hat{\rho}_n) - \frac{\eta_n^2}{64\tau_n}] \wedge 1 \leq n^{-a}$$

proving Corollary 6.5.10. $\square$

6.6 Application in density estimation

Let us reconsider the Example 0.17, i.e. the kernel density estimator $\hat{g}_n$ for g based on i i d re's η_j in $(\mathbb{R}^d, \mathcal{B}^d), d \geq 1$, with unknown Lebesgue density g, defined by

$$\hat{g}_n(t) := h_n^{-d}\frac{1}{n}\sum_{j\leq n} K(\frac{t-\eta_j}{h_n}) \;=\; h_n^{-d}\int\limits_{\mathbb{R}^d} K(\frac{t-y}{h_n})\,\nu_n(dy)\;, \qquad t \in \mathbb{R}^d.$$

where ν_n is the empirical measure based on $\eta_1, ..., \eta_n$. Remember the decomposition mentioned in Example 0.17, i.e.

$$\hat{g}_n(t) - g(t) = \underbrace{\hat{g}_n(t) - \mathbb{E}(\hat{g}_n(t))}_{\substack{\text{random part}\\ \text{“fluctuation at } t\text{”}}} + \underbrace{\mathbb{E}(\hat{g}_n(t)) - g(t)}_{\text{Bias (at } t)}$$

and consider now as index set the set

$$\mathcal{F} := \{K\Big(\frac{t-\cdot}{h_n}\Big) \;:\; t \in \mathbb{R}^d\}.$$

which turns out to be a VCGC (cf. (0.19)), whence by the fundamental Lemma 4.3.17 the condition (6.5.11) in Corollary 6.5.10 is fulfilled with $T = \mathcal{F}$:
In fact, in the present situation we have $\forall\, n \in \mathbb{N}$ with $f_t(\cdot) := K\Big(\frac{t-\cdot}{h_n}\Big) \in \mathcal{F},\;\; t \in \mathbb{R}^d$, and for $\nu = \nu_n$

(for which $\nu(F) = 1$ for the envelope $F \equiv 1$ which can be taken as envelope for $\mathcal{F}$ provided that (cf. our assumptions (6.6.3) below) $||K|| = \sup\limits_{z \in \mathbb{R}^d} |K(z)| \leq 1$):

$$\begin{aligned} d_\nu^{(1)}(f_t, f_s) &= \nu_n(|f_t - f_s|) \\ &= \frac{1}{n} \sum_{j \leq n} \Big| K(\frac{t - \eta_j}{h_n}) - K(\frac{s - \eta_j}{h_n}) \Big| \\ &= \hat{\rho}_n(f_t, f_s) \underset{(\text{cf. } (6.5.3))}{:=} \sum_{j \leq n} \Big| \eta_{nj}(f_t) - \eta_{nj}(f_s) \Big| \end{aligned}$$

with $\eta_{nj}(f_t) := \frac{1}{n} K(\frac{t - \eta_j}{h_n}) \quad \forall\, s, t \in \mathbb{R}^d$, whence the condition (6.5.11) in Corollary 6.5.10 is fulfilled. Reconsider now $\forall\, t \in \mathbb{R}^d$

$$\hat{g}_n(t) = h_n^{-d} S_n(f_t) := h_n^{-d} \sum_{j \leq n} \eta_{nj}(f_t).$$

Since (cf. (6.6.3) below) $||K|| \leq 1$ will be assumed, we have

$$||\eta_{nj}|| \equiv ||\eta_{nj}||_{\mathcal{F}} = \sup_{f_t \in \mathcal{F}} |\eta_{nj}(f_t)| = \frac{1}{n} \sup_{t \in \mathbb{R}^d} |K(\frac{t - \eta_j}{h_n})| \leq \frac{1}{n},$$

whence the assumptions of Lemma 6.5.8 are fulfilled with $\delta_n = \frac{1}{n}$; furthermore, in the present situation,

$$\begin{aligned} \mathbb{E}\Big(\sum_{j \leq n} \eta_{nj}^2(f_t) \Big) &= \frac{1}{n} \mathbb{E}\Big(K^2(\frac{t - \eta_j}{h_n}) \Big) \\ &= \frac{1}{n} \int\limits_{\mathbb{R}^d} K^2(\frac{t - \eta_j}{h_n}) g(y)\, dy \\ &\underset{(||K|| \leq 1)}{\leq} \frac{1}{n} h_n^d\, ||g|| \int\limits_{\mathbb{R}^d} |K(z)|\, dz = O(\frac{h_n^d}{n}), \end{aligned}$$

whence in view of Lemma 6.5.8

$$V_n = \sup_{f_t \in \mathcal{F}} \mathbb{E}\Big(\sum_{j \leq n} \eta_{nj}^2(f_t) \Big) \leq O(\frac{h_n^d}{n}).$$

Before, we have used finiteness of $||g|| := \sup\limits_{t \in \mathbb{R}^d} |g(t)|$; this is guaranteed by assuming e.g. Lipschitz-continuity of g, i.e.

(6.6.1) $\qquad |g(x) - g(y)| \leq L|x - y| \quad$ for some constant $0 < L < \infty$ and $\forall\, x, y \in \mathbb{R}^d$.

Thus, Corollary 6.5.10 together with the Borel-Cantelli-Lemma ([Gae77] 1.16.7 (i)) yields

(6.6.2) $\qquad ||\hat{g}_n - \mathbb{E}(\hat{g}_n)|| = h_n^{-d} ||S_n - \mathbb{E}(S_n)|| = h_n^{-d} O\Big(\sqrt{h_n^d \frac{\log n}{n}} \vee \frac{\log n}{n} \Big) \qquad \mathbb{P} - \text{a.s.}$

(In fact, according to Corollary 6.5.10, for sufficiently large $K > 0$

$$\sum_{n\in\mathbb{N}} \mathbb{P}\Big(\underbrace{\{||S_n - \mathbb{E}(S_n)|| \geq K(\sqrt{V_n \log(n)} \vee \frac{\log n}{n})\}}_{=:A_n}\Big) < \infty$$

whence by Borel-Cantelli $\mathbb{P}(\limsup A_n) = 0$, i.e. $\mathbb{P}(\bigcup_m \bigcap_{n\geq m} \complement A_n) = 1$, where

$$\complement A_n = \{||S_n - \mathbb{E}(S_n)|| < K(\sqrt{V_n \log(n)} \vee \frac{\log n}{n})\} \quad \text{and} \quad V_n = O(\frac{h_n^d}{n}),$$

from which (6.6.2) follows.)
Concerning the Bias, i.e. $||\mathbb{E}(\hat{g}_n) - g||$ (with $\mathbb{E}(\hat{g}_n)(t) := \mathbb{E}(\hat{g}_n(t))$), one gets under the assumptions (6.6.1) and

(6.6.3) $$||K|| \leq 1, \quad \int_{\mathbb{R}^d} K(z)\, dz = 1 \quad \text{and} \quad \int_{\mathbb{R}^d} |z|\, |K(z)|\, dz < \infty$$

that

(6.6.4) $$||\mathbb{E}(\hat{g}_n) - g|| \leq L\, h_n \int_{\mathbb{R}^d} |z|\, |K(z)|\, dz = O(h_n).$$

(In fact, $\forall\, t \in \mathbb{R}^d$

$$\begin{aligned} \mathbb{E}(\hat{g}_n(t)) - g(t) &= h_n^{-d} \int_{\mathbb{R}^d} K(\frac{t-y}{h_n})(g(y) - g(t))\, dy \\ &= \int_{\mathbb{R}^d} K(z)(g(t - h_n z) - g(t))\, dz \\ &\leq L\, h_n \int_{\mathbb{R}^d} |z|\, |K(z)|\, dz. \quad) \end{aligned}$$

As one can see from (6.6.2) and (6.6.4), the Bias is decreasing with $h_n \searrow 0$, but at the same time the fluctuation, i.e. $||\hat{g}_n - \mathbb{E}(\hat{g}_n)||$, is increasing. So, to obtain a balance between both effects one may consider the equation

$$h_n = h_n^{-d} \sqrt{h_n^d \frac{\log n}{n}}$$

which yields

$$h_n = \Big(\frac{\log n}{n}\Big)^{1/(d+2)}.$$

Thus (!)

$$||\hat{g}_n - g|| = O\Big(\Big(\frac{\log n}{n}\Big)^{1/(d+2)}\Big) \qquad \mathbb{P}-\text{a.s.},$$

if $h_n/\Big(\frac{\log n}{n}\Big)^{1/(d+2)} \longrightarrow c$ for some $0 < c < \infty$.

7 Functional Central Limit Theorems (FCLT)

7.1 A FCLT for RMP's

Our starting point in this section is the same as in 6.1 with the aim to present a Functional Central Limit Theorem (FCLT) for Random Measure Processes (RMP's) $S_n = (S_n(f))_{f\in\mathcal{F}}$, where

$$S_n(f) := \sum_{j\le j(n)} w_{nj}(f)\cdot\xi_{nj}, \quad f\in\mathcal{F}, \text{ with } j(n)\longrightarrow\infty \text{ as } n\to\infty,$$

assuming again that the processes $(w_{nj}(f)\cdot\xi_{nj})_{f\in\mathcal{F}}$ are given via coordinate projections on the product p-space $(\Omega,\mathcal{A},\mathbb{P})$ as defined in 6.1.

We tacitly assume regularity conditions such as measurability and finiteness of $w_{nj}(F)$ and now even the same of $w_{nj}(F^2)$ (with $F: X\longrightarrow\mathbb{R}$ being an $\mathcal{X}$-measurable envelope of $\mathcal{F}$)
As already remarked in 3.1, this implies that the sample paths os S_n are contained in the Banachspace

$$l^\infty(\mathcal{F}) := \{x:\mathcal{F}\longrightarrow\mathbb{R} \ : \ ||x||_{\mathcal{F}} := \sup_{f\in\mathcal{F}}|x(f)| < \infty\}$$

endowed with the sup-norm $||\cdot||_{\mathcal{F}}$, and it also implies in view of the condition (7.1.4) imposed in our FCLT 7.1.3 below that also $\sup_{f\in\mathcal{F}}\mathbb{E}(|S_n(f)|)<\infty$ for sufficiently large n.

Thus, for sufficiently large n, the processes $S_n-\mathbb{E}(S_n)$ can be viewed as rq's in $S := (l^\infty(\mathcal{F}), ||\cdot||_{\mathcal{F}})$, and to obtain a FCLT for $S_n-\mathbb{E}(S_n)$ amounts to present further sufficient conditions on $\mathcal{F}$ and on both triangular arrays (w_{nj}) and (ξ_{nj}) under which

$$S_n-\mathbb{E}(S_n)\xrightarrow{\mathcal{L}}\mathbb{G} \quad \text{in } S = l^\infty(\mathcal{F})$$

in the sense of (2.3.2) with a limiting re $\mathbb{G} = (G(f))_{f\in\mathcal{F}}$ in $(S,\mathcal{B}(S))$ being a mean-zero Gaussian process.

If, in addition, $\mathbb{G}$ is separable, we write as in 2.3 $\underset{\text{sep}}{\xrightarrow{\mathcal{L}}}$ instead of $\xrightarrow{\mathcal{L}}$. We will focus here on

(7.1.1) $$S_n-\mathbb{E}(S_n)\underset{\text{sep}}{\xrightarrow{\mathcal{L}}}\mathbb{G} \quad \text{in } l^\infty(\mathcal{F})$$

with $\mathbb{G}$ having all its sample paths in the subspace $U^b(\mathcal{F},d)$ of S, where

$$U^b(\mathcal{F},d) := \{x\in l^\infty(\mathcal{F}) : x \text{ uniformly } d\text{-continuous}\},$$

in order to apply our Characterization Theorem of $\mathcal{L}$-Convergence 2.3.9 with d being a pseudo-metric on $\mathcal{F}$ such that $(\mathcal{F},d)$ is totally bounded.

Remember that $U^b(\mathcal{F},d)$ is a separable subspace of S if and only if $(\mathcal{F},d)$ is totally bounded ([Gae90], Corollary 2).

Before focussing on (7.1.1), some general comments are in order in comparing our $\underset{\text{sep}}{\xrightarrow{\mathcal{L}}}$ -convergence with related concepts found in the literature (see e.g. [Va96]):

For this, let $\eta_n : \Omega \longrightarrow l^\infty(\mathcal{F}), n \geq 1$, be arbitrary rq's with $\eta_n \xrightarrow[\text{sep}]{\mathcal{L}} \mathbb{G}$, where $\mathbb{G}$ has all its sample paths in some separable subspace S_0 of $l^\infty(\mathcal{F})$. Then the law $\mathcal{L}\{\mathbb{G}\}$ of $\mathbb{G}$ is a *Radon measure* on $\mathcal{B}(l^\infty(\mathcal{F}))$, i.e. for each $B \in \mathcal{B}(l^\infty(\mathcal{F}))$

$$\mathcal{L}\{\mathbb{G}\}(B) = \sup\{\mathcal{L}\{\mathbb{G}\}(K) \ : \ K \subset B, K \text{ compact}\}.$$

To see this, note that according to [Bi68], p. 9,

$$\mathcal{L}\{\mathbb{G}\} \text{ Radon measure} \quad \Longleftrightarrow \quad \mathcal{L}\{\mathbb{G}\} \ \textit{tight}$$

(i.e. $\forall \varepsilon > 0 \, \exists K = K_\varepsilon \subset l^\infty(\mathcal{F}), K$ compact, s.t. $\mathcal{L}\{\mathbb{G}\}(K) \geq 1 - \varepsilon$). So it remains to show that $\mathfrak{S} := \mathcal{L}\{\mathbb{G}\}$ is tight:
For this, let D be a countable and dense subset of S_0; let $j \in \mathbb{N}$ and $\varepsilon > 0$ be arbitrary. Then the open balls $B(x, \frac{1}{j}) \subset l^\infty(\mathcal{F}), x \in D$, (with center x and radius $\frac{1}{j}$) form a cover of S_0 and therefore (due to the σ-continuity of $\mathfrak{S}$) there exist $x_1, ..., x_{n_j} \in D$ such that

$$\mathfrak{S}(\bigcup_{i=1}^{n_j} B(x_i, \frac{1}{j})) \geq 1 - \varepsilon/2^j.$$

Put $G_j := \bigcup_{i=1}^{n_j} B(x_i, \frac{1}{j})$; then $\bigcap_{j\in\mathbb{N}} G_j$ is totally bounded and $\mathfrak{S}(\bigcap_{j\in\mathbb{N}} G_j) = 1 - \mathfrak{S}(\bigcup_{j\in\mathbb{N}} \complement G_j) \geq 1 - \sum_{j\in\mathbb{N}} \mathfrak{S}(\complement G_j) \geq 1 - \varepsilon$. Since also $K := (\bigcap_{j\in\mathbb{N}} G_j)^c$ is totally bounded and complete (as a closed subset of the complete space $l^\infty(\mathcal{F})$, K is compact with $\mathfrak{S}(K) \geq 1 - \varepsilon$, which proves tightness since $\varepsilon > 0$ was chosen arbitrary.

On the other hand, if $\mathcal{L}\{\mathbb{G}\}$ is a Radon measure, whence tight, and if $\eta_n \xrightarrow{\mathcal{L}} \mathbb{G}$ in the sense of (2.3.2) (with $S = l^\infty(\mathcal{F})$), it follows that there exists a stochastic process $\bar{\mathbb{G}} = (\bar{G}(f))_{f\in\mathcal{F}}$ defined on an appropriate p-space $(\bar{\Omega}, \bar{\mathcal{A}}, \bar{\mathbb{P}})$ with sample paths in a separable subspace S_0 of $l^\infty(\mathcal{F})$ such that $\eta_n \xrightarrow[\text{sep}]{\mathcal{L}} \bar{\mathbb{G}}$, where $\bar{\mathbb{G}} \underset{\text{fidi}}{\overset{\mathcal{L}}{=}} \mathbb{G}$:
In fact, $\mathcal{L}\{\mathbb{G}\}$ tight $\Longrightarrow$ $S_0 := supp\mathcal{L}\{\mathbb{G}\}$ σ-compact and therefore separable; then, taking $(\bar{\Omega}, \bar{\mathcal{A}}, \bar{\mathbb{P}}) := (S_0, \mathcal{B}(S_0), \mathcal{L}\{\mathbb{G}\})$ and $\bar{G}(f)(x) := \pi_f(x) := x(f)$ for $x \in S_0$ the assertion follows (see [Gae77], Lemma 7.2.31).

Finally, let us mention also (without proof) the following result (see [Va96], Section 1.12, and [Gi97], Corollary 1.5):

Let $\eta_n : \Omega \longrightarrow l^\infty(\mathcal{F}), n \geq 1$, be arbitrary rq's, $\eta_0 : \Omega \longrightarrow l^\infty(\mathcal{F})$ be $\mathcal{A}, \mathcal{B}(l^\infty(\mathcal{F}))$-measurable with $\mathcal{L}(\eta_0\}$ being tight; then

(7.1.2) $$\eta_n \xrightarrow{\mathcal{L}} \eta_0 \ \text{ in the sense of (2.3.2) (with } S = l^\infty(\mathcal{F})) \iff d_{BL}(\eta_n, \eta_0) \longrightarrow 0,$$

where $d_{BL}(\eta_n, \eta_0) := \sup\{|\mathbb{E}^*(H(\eta_n)) - \mathbb{E}(H(\eta_0))| : H \in BL_1(l^\infty(\mathcal{F}))\}$ with

$$BL_1(l^\infty(\mathcal{F})) := \Big\{H : l^\infty(\mathcal{F}) \longrightarrow \mathbb{R} \ : \ \sup_{x\in l^\infty(\mathcal{F})} |H(x)| \leq 1, \ \sup_{x,y\in l^\infty(\mathcal{F}), x\neq y} \frac{|H(x) - H(y)|}{||x - y||_\mathcal{F}} \leq 1\Big\}.$$

Now, the Functional Central Limit Theorem (FCLT) for Random Measure Processes (RMP's) reads as follows (cf. [Zi97], Theorem 6.1 together with Remark 6.2):

7.1.3. THEOREM *(FCLT for RMP's).*
Let $X = (X, \mathcal{X})$ be an arbitrary measurable space and $\mathcal{F}$ be a class of $\mathcal{X}$-measurable functions $f : X \longrightarrow \mathbb{R}$ with $\mathcal{X}$-measurable envelope $F : X \longrightarrow \mathbb{R}$ (supposed to be countable to avoid measurability considerations). Assume that $\mathcal{F}$ has uniformly integrable L_2-entropy (cf. 4.3.20) and that there is some pseudometric d on $\mathcal{F}$ such that $(\mathcal{F}, d)$ is totally bounded. Assume further that the following conditions (7.1.4) – (7.1.6) are fulfilled:

For each $\rho > 0$ *there exists* $\delta_n = \delta_n(\rho) > 0, n \in \mathbb{N}$, *with* $\delta_n \longrightarrow 0$ *such that*

$$\limsup_{n\to\infty} \sum_{j\le j(n)} \mathbb{E}\Big(w_{nj}(F)|\xi_{nj}| \cdot I(w_{nj}(F)|\xi_{nj}| > \delta_n)\Big) \le \rho \tag{7.1.4}$$

$$\lim_{\alpha\to 0} \limsup_{n\to\infty} \sup_{d(f,g)\le\alpha} \sum_{j\le j(n)} \mathbb{E}\Big((w_{nj}(f) - w_{nj}(g))^2 \xi_{nj}^2 \cdot I(w_{nj}(F)|\xi_{nj}| \le \delta_n)\Big) = 0 \tag{7.1.5}$$

$$\sup_{n\in\mathbb{N}} \sum_{j\le j(n)} \mathbb{E}\Big(w_{nj}(F^2)\xi_{nj}^2 \cdot I(w_{nj}(F)|\xi_{nj}| \le \delta_n)\Big) < \infty\,. \tag{7.1.6}$$

Assume in addition, that there exists a mean-zero Gaussian process $\bar{\mathbb{G}} = (\bar{G}(f))_{f\in\mathcal{F}}$ such that $S_n - \mathbb{E}(S_n) \xrightarrow[\text{fidi}]{\mathbb{P}} \bar{\mathbb{G}}$.

Then there exists a mean zero Gaussian process $\mathbb{G} = (G(f))_{f\in\mathcal{F}}$ with sample paths in $U^b(\mathcal{F}, d)$ (being a separable subspace of $(l^\infty(\mathcal{F}), ||\cdot||_{\mathcal{F}})$ such that

$$S_n - \mathbb{E}(S_n) \xrightarrow[sep]{\mathcal{L}} \mathbb{G} \quad \text{in } l^\infty(\mathcal{F}) \quad \text{and} \quad \bar{\mathbb{G}} \overset{\mathcal{L}}{\underset{\text{fidi}}{=}} \mathbb{G}. \tag{7.1.7}$$

PROOF. Concerning (7.1.7) we remark (as already mentioned above) that by (7.1.4) for sufficiently large n

$$\sup_{f\in\mathcal{F}} \mathbb{E}(|S_n(f)|) \le \sum_{j\le j(n)} \mathbb{E}(w_{nj}(F)|\xi_{nj}|) \le$$
$$\sum_{j\le j(n)} \mathbb{E}\Big(w_{nj}(F)|\xi_{nj}| \cdot I(w_{nj}(F)|\xi_{nj}| > \delta_n)\Big) + j(n)\delta_n < \infty.$$

Now, since $(\mathcal{F}, d)$ is assumed to be totally bounded and since $S_n - \mathbb{E}(S_n) \xrightarrow[\text{fidi}]{\mathbb{P}} \bar{\mathbb{G}}$ by assumption, it follows in view of our CT$\mathcal{L}$-C 2.3.9 together with Remark 2.3.19 that it remains to show

(a) $\lim_{\alpha\to 0} \limsup_{n\to\infty} \mathbb{E}\Big(\sup_{d(f,g)\le\alpha} \big|S_n(f) - S_n(g) - (\mathbb{E}(S_n(f)) - \mathbb{E}(S_n(g)))\big|\Big) = 0$.

For this, according to the Symmetrization Inequality 5.1.2 it suffices to show

(b) $\lim_{\alpha\to 0} \limsup_{n\to\infty} \mathbb{E}\Big(\sup_{d(f,g)\le\alpha} |\sum_{j\le j(n)} \varepsilon_j \xi_{nj} \cdot (w_{nj}(f) - w_{nj}(g))|\Big) = 0,$

where $(\varepsilon_j)_{j\in\mathbb{N}}$ is a canonically formed Rademacher sequence being independent of both arrays (w_{nj}) and (ξ_{nj}).

Let $\rho > 0$ be arbitrary and $\delta_n = \delta_n(\rho) > 0, n \in \mathbb{N}$, with $\delta_n \longrightarrow 0$, fulfilling (7.1.4) – (7.1.6). Then

$$\mathbb{E}\Big(\sup_{d(f,g)\le\alpha} \big| \sum_{j\le j(n)} \varepsilon_j \, \xi_{nj} \cdot (w_{nj}(f) - w_{nj}(g)) \big| \Big)$$

$$\le \mathbb{E}\Big(\sup_{d(f,g)\le\alpha} \big| \sum_{j\le j(n)} \varepsilon_j \, \xi_{nj} \cdot (w_{nj}(f) - w_{nj}(g)) \cdot I(w_{nj}(F)|\xi_{nj}| \le \delta_n) \big| \Big)$$

$$+ 2 \sum_{j\le j(n)} \mathbb{E}\Big(w_{nj}(F)|\xi_{nj}| \cdot I(w_{nj}(F)|\xi_{nj}| > \delta_n) \Big),$$

and so, because of (7.1.4), it remains to show

(c) $\lim_{\alpha\to 0} \limsup_{n\to\infty} \mathbb{E}\Big(\sup_{d(f,g)\le\alpha} \big| \sum_{j\le j(n)} \varepsilon_j \, \xi_{nj} (w_{nj}(f) - w_{nj}(g)) \cdot I(w_{nj}(F)|\xi_{nj}| \le \delta_n) \big| \Big) = 0.$

For this, let, for $f, g \in \mathcal{F}$,

$$\sigma^2_{n\delta_n}(f,g) := \sum_{j\le j(n)} \mathbb{E}\Big((w_{nj}(f) - w_{nj}(g))^2 \xi^2_{nj} \cdot I(w_{nj}(F)|\xi_{nj}| \le \delta_n) \Big).$$

With this definition of $\sigma_{n\delta_n}$ condition (7.1.5) reads as follows:

(d) $\lim_{\beta\to 0} \limsup_{n\to\infty} \sup_{d(f,g)\le\beta} \sigma^2_{n\delta_n}(f,g) = 0.$

But (d) allows us to switch in (c) from the pseudo-metric d to $\sigma_{n\delta_n}$, i.e. in doing so we have to show

(e) $\lim_{\alpha\to 0} \limsup_{n\to\infty} \mathbb{E}\Big(\sup_{\sigma_{n\delta_n}(f,g)\le\alpha} \big| \sum_{j\le j(n)} \varepsilon_j \, \xi_{nj} (w_{nj}(f) - w_{nj}(g)) \cdot I(w_{nj}(F)|\xi_{nj}| \le \delta_n) \big| \Big) = 0.$

Now

$$\rho^2_{n\delta_n}(f,g) := \sum_{j\le j(n)} (w_{nj}(f) - w_{nj}(g))^2 \xi^2_{nj} \cdot I(w_{nj}(F)|\xi_{nj}| \le \delta_n)$$

$$\le \sum_{j\le j(n)} w_{nj}((f-g)^2) \, \xi^2_{nj} \cdot I(w_{nj}(F)|\xi_{nj}| \le \delta_n) \;=\; \tilde{\mu}_{n\delta_n}((f-g)^2)$$

for all $f, g \in \mathcal{F}$ with

$$\tilde{\mu}_{n\delta_n}(f) := \sum_{j\le j(n)} w_{nj}(f) \, \xi^2_{nj} \cdot I(w_{nj}(F)|\xi_{nj}| \le \delta_n),$$

where

$$\sigma^2_{n\delta_n}(f,g) = \mathbb{E}(\rho^2_{n\delta_n}(f,g)) \quad \text{for all } f, g \in \mathcal{F}.$$

By this, we arrived at a situation which allows us to apply Ziegler's Maximal Inequality ([Zi97], Theorem 3.1, applied here with $\Phi_{nj}(f) := w_{nj}(f) \, \xi^2_{nj} \cdot I(w_{nj}(F)|\xi_{nj}| \le \delta_n)$) according to which there exist universal constants $0 < K_i < \infty, i = 1, 2$, such that for all $\alpha > 0$

(f) $\mathbb{E}\big(\sup_{\sigma_{n\delta_n}(f,g)\le\alpha} \big| \sum_{j\le j(n)} \varepsilon_j \, \xi_{nj} \cdot (w_{nj}(f) - w_{nj}(g)) \cdot I(w_{nj}(F)|\xi_{nj}| \le \delta_n) \big| \big)$

$$\le K_1 \cdot A(n,\alpha) \cdot B(n) \;+\; K_2 \cdot C(n,\alpha)$$

with

$$A(n,\alpha) := \alpha^{-1}\mathbb{E}^{*\frac{1}{2}}\Big(\max_{j\le j(n)}\sup_{f,g\in\mathcal{F}}|\xi_{nj}(w_{nj}(f)-w_{nj}(g))|\cdot I(w_{nj}(F)|\xi_{nj}|\le\delta_n)\cdot[\tilde\mu_{n\delta_n}(F^2)]^{1/2}\cdot l_n(1)\Big)$$

$$B(n) := \mathbb{E}^{*\frac{1}{2}}\Big(\tilde\mu_{n\delta_n}(F^2)\ (l_n(1))^2\Big)$$

and $C(n,\alpha) := \mathbb{E}^*\Big(\max\{1,[\tilde\mu_{n\delta_n}(F^2)]^{1/2}\}\cdot l_n(\alpha)\Big)$,

where $l_n(\alpha)$ is the random integral, defined by

$$l_n(\alpha) := \int_0^\alpha\Big(\log N\big(\tau\,(\tilde\mu_{n\delta_n}(F^2))^{1/2},\mathcal{F},d^{(2)}_{\tilde\mu_{n\delta_n}}\big)\Big)^{1/2}d\tau.$$

Here, we have according to (7.1.6) that

(g) $\sup_{n\in\mathbb{N}}\mathbb{E}(\tilde\mu_{n\delta_n}(F^2)) = \sum_{j\le j(n)}\mathbb{E}\Big(w_{nj}(F^2)\xi^2_{nj}\cdot I(w_{nj}(F)\,|\xi_{nj}|\le\delta_n)\Big)<\infty,$

whence $\tilde\mu_{n\delta_n}\in\mathcal{M}(X,F)$ for all $n\in\mathbb{N}$ a.s., and thus a.s.

(h) $l_n(1)\le\int_0^\infty\Big(\log\big[\sup_{\nu\in\mathcal{M}(X,F)}N\big(\tau\,(\nu(F^2))^{1/2},\mathcal{F},d^{(2)}_\nu\big)\big]\Big)^{1/2}d\tau<\infty,$

since $\mathcal{F}$ has uniformly integrable L_2-entropy.
The latter also implies that a.s

(i) $l_n(\alpha)\le\int_0^\alpha\Big(\log\big[\sup_{\nu\in\mathcal{M}(X,F)}N\big(\tau\,(\nu(F^2))^{1/2},\mathcal{F},d^{(2)}_\nu\big)\big]\Big)^{1/2}d\tau\longrightarrow 0$ as $\alpha\longrightarrow 0$.

Now, concerning $A(n,\alpha)$, note that

$$\max_{j\le j(n)}\sup_{f,g\in\mathcal{F}}|\xi_{nj}(w_{nj}(f)-w_{nj}(g))|\cdot I(w_{nj}(F)|\xi_{nj}|\le\delta_n)\le 2\cdot\delta_n,$$

which implies by (g) and (h) that for all $\alpha>0$

$$\limsup_{n\to\infty}A(n,\alpha)\cdot B(n)=0.$$

Finally, by (g) and (i) we get

$$\lim_{\alpha\to 0}\limsup_{n\to\infty}C(n,\alpha)=0,$$

which completes the proof of (e) according to (f). □

It is easily seen that the conditions (7.1.4) – (7.1.6) become much simpler in case of a uniformly bounded index set $\mathcal{F}$ (with envelope $F\equiv 1$ w.l.o.g.) In this case we obtain immediately from Theorem 7.1.3:

7.1.8. Corollary *(cf. [Zi97], Corollary 6.3).*
Assume that $\mathcal{F}$ is uniformly bounded and has uniformly integrable L_2-entropy, and that there is some pseudo-metric d on $\mathcal{F}$ such that $(\mathcal{F}, d)$ is totally bounded. Assume further that the following conditions (7.1.4)′ – (7.1.6)′ are fulfilled:

$$(7.1.4)' \qquad \lim_{n\to\infty} \sum_{j\le j(n)} \mathbb{E}\Big(|\xi_{nj}| \cdot I(|\xi_{nj}| > \delta)\Big) = 0 \quad \textit{for all} \quad \delta > 0$$

(Noticing that (7.1.4)′ implies the existence of a sequence $(\delta_n)_{n\in\mathbb{N}}$ of positive real numbers δ_n such that $\delta_n \longrightarrow 0$ and $\lim_{n\to\infty} \sum_{j\le j(n)} \mathbb{E}\Big(|\xi_{nj}| \cdot I(|\xi_{nj}| > \delta_n)\Big) = 0$.)

$$(7.1.5)' \qquad \lim_{\alpha\to 0} \limsup_{n\to\infty} \sup_{d(f,g)\le\alpha} \sum_{j\le j(n)} \mathbb{E}\Big((w_{nj}(f) - w_{nj}(g))^2 \xi_{nj}^2 \cdot I(|\xi_{nj}| \le \delta_1)\Big) = 0 \textit{ for some } \delta_1 > 0$$

$$(7.1.6)' \qquad \sup_{n\in\mathbb{N}} \sum_{j\le j(n)} \mathbb{E}\Big(\xi_{nj}^2 \cdot I(|\xi_{nj}| \le \delta_2)\Big) < \infty \textit{ for some } \delta_2 > 0$$

There exists a mean-zero Gaussian process $\bar{\mathbb{G}} = (\bar{G}(f))_{f\in\mathcal{F}}$ such that

$$S_n - \mathbb{E}(S_n) \xrightarrow[\textit{fidi}]{\mathbb{P}} \bar{\mathbb{G}}.$$

Then there exists a mean-zero Gaussian process $\mathbb{G} = (G(f))_{f\in\mathcal{F}}$ with sample paths in $U^b(\mathcal{F}, d)$ (being a separable subspace of $(l^\infty(\mathcal{F}), || \cdot ||_{\mathcal{F}})$ such that

$$(7.1.7) \qquad S_n - \mathbb{E}(S_n) \xrightarrow[\textit{sep}]{\mathcal{L}} \mathbb{G} \quad \textit{in } l^\infty(\mathcal{F}) \quad \textit{and} \quad \bar{\mathbb{G}} \underset{\textit{fidi}}{\overset{\mathcal{L}}{=}} \mathbb{G}.$$

where again

$$S_n(f) := \sum_{j\le j(n)} w_{nj}(f) \cdot \xi_{nj}, \quad f \in \mathcal{F}.$$

Let us consider next the special case where $w_{nj} = \delta_{\eta_{nj}}$, $(\eta_{nj})_{1\le j\le j(n), n\in\mathbb{N}}$ being a triangular array of re's in $(X, \mathcal{X})$ in order to present tractable conditions under which (7.1.8) holds true.

7.1.6. Corollary *(cf. [Zi97], Corollary 6.4).*
Let $\mathcal{F}$ be as in Corollary 7.1.8. Let $w_{nj} = \delta_{\eta_{nj}}$ where $(\eta_{nj})_{1\le j\le j(n), n\in\mathbb{N}}$ is a triangular array of re's in $(X, \mathcal{X})$ with laws $\nu_{nj} := \mathcal{L}\{\eta_{nj}\}$ on $\mathcal{X}$, and suppose now, in addition to the basic independence assumption for the pairs $(\eta_{n1}, \xi_{n1}), ..., (\eta_{nj(n)}, \xi_{nj(n)})$, that for each $n \in \mathbb{N}$ and $1 \le j \le j(n)$ also η_{nj} and ξ_{nj} are independent. Assume further that there is some p-measure ν on $\mathcal{X}$ and constants

$0 < c_i < \infty$ *such that the following four conditions are fulfilled:*

(7.1.4)′
$$\lim_{n\to\infty} \sum_{j\le j(n)} \mathbb{E}\Big(|\xi_{nj}| \cdot I(|\xi_{nj}| > \delta)\Big) = 0 \quad \textit{for all} \quad \delta > 0$$

(7.1.7)
$$\lim_{\alpha\to 0} \limsup_{n\to\infty} \sup_{d_\nu^{(2)}(f,g)\le\alpha} \sum_{j\le j(n)} \nu_{nj}((f-g)^2) \cdot \mathbb{E}(\xi_{nj}^2 \cdot I(|\xi_{nj}| \le \delta_1)) = 0 \textit{ for some } \delta_1 > 0$$

(7.1.8)
$$\lim_{n\to\infty} \sum_{j\le j(n)} \nu_{nj}(f\cdot g) \cdot \mathbb{E}(\xi_{nj}^2 \cdot I(|\xi_{nj}| \le \delta_2)) = c_1\nu(f\cdot g) \quad \textit{for all } f,g \in \mathcal{F} \textit{ and some } \delta_2 > 0$$

(7.1.9)
$$\lim_{n\to\infty} \sum_{j\le j(n)} \nu_{nj}(f) \cdot \nu_{nj}(g) \Big(\mathbb{E}(\xi_{nj} \cdot I(|\xi_{nj}| \le \delta_3))\Big)^2 = c_2\nu(f)\cdot\nu(g) \textit{ for all } f,g \in \mathcal{F} \textit{ and some } \delta_3 > 0.$$

Then, with $S_n(f) := \sum_{j\le j(n)} f(\eta_{nj}) \cdot \xi_{nj}$, $f \in \mathcal{F}$,

$$S_n - \mathbb{E}(S_n) \xrightarrow[sep]{\mathcal{L}} \mathbb{G}_\nu \quad \textit{in } l^\infty(\mathcal{F}),$$

where $\mathbb{G}_\nu = (G_\nu(f))_{f\in\mathcal{F}}$ *is a mean zero Gaussian process with sample paths in* $U^b(\mathcal{F}, d_\nu^{(2)})$ *and* $cov(G_\nu(f), G_\nu(g)) = c_1\nu(f\cdot g) - c_2\nu(f)\cdot\nu(g)$ *for* $f,g \in \mathcal{F}$.

PROOF. Note first that (7.1.5)′ coincides with (7.1.7) in the present case since $\mathbb{E}\Big((w_{nj}(f) - w_{nj}(g))^2\xi_{nj}^2 \cdot I(|\xi_{nj}| \le \delta_1)\Big) = \mathbb{E}((w_{nj}(f) - w_{nj}(g))^2) \cdot \mathbb{E}(\xi_{nj}^2 \cdot I(|\xi_{nj}| \le \delta_1)) \underset{(w_{nj}=\delta_{\eta_{nj}})}{=} \nu_{nj}((f-g)^2) \cdot \mathbb{E}(\xi_{nj}^2 \cdot I(|\xi_{nj}| \le \delta_1))$ for each n and $1 \le j \le j(n)$. Secondly, assuming w.l.o.g. $F \equiv 1 \in \mathcal{F}$, (7.1.8) (with $f = g = F \equiv 1$) implies (7.1.6)′. Therefore, the assertion follows from Corollary 7.1.8, since under the present conditions one can verify (7.1.8) in the same way as it was done (in the set-indexed case) within the proof of Theorem 2.2 in [Gae94], part (a). □

7.2 FCLT's for partial-sum processes with either fixed or random locations

Let $X = (X, \mathcal{X})$ be an arbitrary measurable space, $\mathcal{C} \subset \mathcal{X}$ a countable VCC being w.l.o.g. closed under the formation of symmetric differences (cf. 4.2.7). Note that $\mathcal{F} := \{1_C : C \in \mathcal{C}\}$ has uniformly integrable L_2-entropy according to 4.3.21.

Let $w_{nj} = \delta_{\eta_{nj}}$, $(\eta_{nj})_{1\le j\le j(n), n\in\mathbb{N}}$ being a triangular array of re's in $(X, \mathcal{X})$ (with $j(n) \longrightarrow \infty$ as $n \to \infty$) and $(\xi_{nj})_{1\le j\le j(n), n\in\mathbb{N}}$ a triangular array of rv's such that for each $n \in \mathbb{N}$ the sequence of pairs $(\eta_{n1}, \xi_{n1}), ..., (\eta_{nj(n)}, \xi_{nj(n)})$ is independent but not necessarily identically distributed (and where the components within each pair need not be independent). Let $d := d_\nu$ for some p-measure ν on $\mathcal{X}$ (where $d_\nu(C,D) := \nu(C\Delta D) = d_\nu^{(2)}(1_C, 1_D)$ for $C, D \in \mathcal{C}$); note that $(\mathcal{C}, d_\nu)$ is totally bounded (cf. 4.2.3). Then, specializing Corollary 7.1.8 to the present case, we obtain the following result for partial-sum processes $S_n = (S_n(C))_{C\in\mathcal{C}}$ with random locations as introduced in Section 3.2.1, i.e. with

$$S_n(C) := \sum_{j\le j(n)} 1_C(\eta_{nj})\cdot\xi_{nj} \quad , \; C\in\mathcal{C}:$$

7.2.1. THEOREM *(cf. [Gae94], Theorem 2.11).*
Assume that the following conditions are fulfilled:

$$(7.1.4)' \qquad \lim_{n\to\infty}\sum_{j\le j(n)} \mathbb{E}\Big(|\xi_{nj}|\cdot I(|\xi_{nj}|>\delta)\Big) = 0 \text{ for all } \delta>0$$

$$(7.2.2) \qquad \lim_{\alpha\to 0}\limsup_{n\to\infty}\sup_{C\in\mathcal{C}:\nu(C)\le\alpha}\sum_{j\le j(n)} \mathbb{E}\Big(1_C(\eta_{nj})\xi_{nj}^2\cdot I(|\xi_{nj}|\le\delta_1)\Big) = 0 \text{ for some } \delta_1>0$$

$$(7.1.6)' \qquad \sup_{n\in\mathbb{N}}\sum_{j\le j(n)} \mathbb{E}\Big(\xi_{nj}^2\cdot I(|\xi_{nj}|\le\delta_2)\Big) < \infty \text{ for some } \delta_2>0$$

(7.1.8)′ *There exists a mean-zero Gaussian process $\bar{\mathbb{G}} = (\bar G(C))_{C\in\mathcal{C}}$ such that*

$$S_n - \mathbb{E}(S_n) \xrightarrow[fidi]{\mathbb{P}} \bar{\mathbb{G}}.$$

Then there exists a mean-zero Gaussian process $\mathbb{G} = (G(C))_{C\in\mathcal{C}}$ with sample paths in $U^b(\mathcal{C}, d_\nu)$ (being a separable subspace of $(l^\infty(\mathcal{C}), ||\cdot||_{\mathcal{C}})$) such that

$$S_n - \mathbb{E}(S_n) \xrightarrow[sep]{\mathcal{L}} \mathbb{G} \quad \text{in } l^\infty(\mathcal{C}) \qquad \text{and } \bar{\mathbb{G}} \underset{fidi}{\overset{\mathcal{L}}{=}} \mathbb{G}.$$

Specializing Corollary 7.1.6 in the same way as just done with Corollary 7.1.8 to the set-indexed case yields the following result for $S_n = (S_n(C))_{C\in\mathcal{C}}$ with

$$S_n(C) := \sum_{j\le j(n)} 1_C(\eta_{nj})\cdot\xi_{nj} \quad , \; C\in\mathcal{C},$$

under the additional assumption that for each $n\in\mathbb{N}$ and $1\le j\le j(n)$ also η_{nj} and ξ_{nj} are independent.

7.2.3. THEOREM *(cf. [Gae94], Theorem 2.2).*
Suppose that there is some p-measure ν on $\mathcal{X}$ and constants $0<c_i<\infty$ such that the following four

conditions are fulfilled:

$$(7.1.4)' \qquad \lim_{n\to\infty} \sum_{j\le j(n)} \mathbb{E}\Big(|\xi_{nj}|\cdot I(|\xi_{nj}|>\delta)\Big) = 0 \quad \textit{for all} \quad \delta>0$$

$$(7.2.4) \qquad \lim_{\alpha\to 0}\limsup_{n\to\infty} \sup_{C\in\mathcal{C}:\nu(C)\le\alpha} \sum_{j\le j(n)} \nu_{nj}(C)\cdot\mathbb{E}(\xi_{nj}^2\cdot I(|\xi_{nj}|\le\delta_1)) = 0 \textit{ for some } \delta_1>0$$

(7.2.5)

$$\lim_{n\to\infty} \sum_{j\le j(n)} \nu_{nj}(C\cap D)\cdot\mathbb{E}(\xi_{nj}^2\cdot I(|\xi_{nj}|\le\delta_2)) = c_1\nu(C\cap D) \;\; \textit{for all } C,D\in\mathcal{C} \textit{ and some } \delta_2>0$$

(7.2.6)

$$\lim_{n\to\infty} \sum_{j\le j(n)} \nu_{nj}(C)\cdot\nu_{nj}(D)\Big(\mathbb{E}(\xi_{nj}\cdot I(|\xi_{nj}|\le\delta_3))\Big)^2 = c_2\nu(C)\cdot\nu(D) \textit{ for all } C,D\in\mathcal{C} \textit{ and some } \delta_3>0.$$

Then

$$S_n - \mathbb{E}(S_n) \xrightarrow[sep]{\mathcal{L}} \mathbb{G}_\nu \quad \textit{in } l^\infty(\mathcal{F}),$$

where $\mathbb{G} = (G(C))_{C\in\mathcal{C}}$ *is a mean-zero Gaussian process with sample paths in* $U^b(\mathcal{C}, d_\nu)$ *and* $cov(G_\nu(C), G_\nu(D)) = c_1\nu(C\cap D) - c_2\,\nu(C)\cdot\nu(D)$ *for* $C,D\in\mathcal{C}$.

From Theorem 7.2.3 we get the following result which was already mentioned in Section 2 (see Theorem 2.2.3) and used at the end of Section 2.3.6.

7.2.7. Corollary *(cf. [Gae94], Theorem 2.15).*
Let $\xi_{nj} = j(n)^{-1}\xi_j$ *for each* $1\le j\le j(n)$ *and* $n\in\mathbb{N}$ *(with* $j(n)\longrightarrow\infty$ *as* $n\to\infty$*), the* ξ_j*'s being i i d rv's with* $\mathbb{E}(\xi_1)=0$ *and* $\mathbb{E}(\xi_1^2)=1$. *Let* $(\eta_{nj})_{1\le j\le j(n),n\in\mathbb{N}}$ *being a triangular array of rowwise independent but not necessarily identically distributed re's in* $(X,\mathcal{X})$ *which is independent of the sequence* $(\xi_j)_{j\in\mathbb{N}}$. *Suppose that there is some p-measure* ν *on* $\mathcal{X}$ *such that the following two conditions are fulfilled (with* $\nu_{nj} := \mathcal{L}\{\eta_{nj}\}$*):*

(i) $\quad \lim_{n\to\infty} j(n)^{-1}\sum_{j\le j(n)}\nu_{nj}(C\cap D) \;=\; \nu(C\cap D) \quad$ *for all* $\;C,D\in\mathcal{C}$

(ii) $\quad \lim_{\alpha\to 0}\limsup_{n\to\infty}\sup_{C\in\mathcal{C}:\nu(C)\le\alpha} j(n)^{-1}\sum_{j\le j(n)}\nu_{nj}(C) \;=\; 0.$

Then

$$\Big(j(n)^{-1/2}\sum_{j\le j(n)} 1_C(\eta_{nj})\cdot\xi_j\Big)_{C\in\mathcal{C}} \xrightarrow[sep]{\mathcal{L}} \mathbb{G}_\nu,$$

where $\mathbb{G}_\nu = (G_\nu(C))_{C\in\mathcal{C}}$ *is a mean-zero Gaussian process with sample paths in* $U^b(\mathcal{C}, d_\nu)$ *and* $cov(G_\nu(C), G_\nu(D)) = \nu(C\cap D)$ *for* $C,D\in\mathcal{C}$.

PROOF. According to Theorem 7.2.3 we have to verify the conditions $(7.1.4)'$, (7.2.4), (7.2.5) with $c_1=1$ and (7.2.6) with $c_2=0$ to get the assertion of 7.2.7.

As to (7.1.4)′: For each $\delta > 0$ we have $\sum_{j\le j(n)} \mathbb{E}(|\xi_{nj}| \cdot I(|\xi_{nj}| > \delta)) \le \delta^{-1} \sum_{j\le j(n)} \mathbb{E}(\xi_{nj}^2 \cdot I(|\xi_{nj}| > \delta)) = \delta^{-1}\mathbb{E}(\xi_1^2 \cdot I(|\xi_1| > \delta j(n)^{1/2})) \longrightarrow 0$, since $\mathbb{E}(\xi_1^2) < \infty$.

As to (7.2.5): $\sum_{j\le j(n)} \nu_{nj}(C\cap D)\cdot\mathbb{E}(\xi_{nj}^2\cdot I(|\xi_{nj}| \le \delta_2)) = [j(n)^{-1}\sum_{j\le j(n)} \nu_{nj}(C\cap D)]\cdot\mathbb{E}(\xi_1^2\cdot I(|\xi_1| \le \delta_2 j(n)^{1/2}))$, where by (i) $\lim_{n\to\infty} j(n)^{-1}\sum_{j\le j(n)} \nu_{nj}(C\cap D) = \nu(C\cap D)$ and $\lim_{n\to\infty}\mathbb{E}(\xi_1^2\cdot I(|\xi_1| \le \delta_2 j(n)^{1/2})) = \mathbb{E}(\xi_1^2) = 1$. This proves (7.2.5) with $c_1 = 1$.

As to (7.2.6): Since $\limsup_{n\to\infty}\sum_{j\le j(n)} \nu_{nj}(C)\cdot\nu_{nj}(D)\cdot\Big(\mathbb{E}(\xi_{nj}\cdot I(|\xi_{nj}| \le \delta_3))\Big)^2 \le \limsup_{n\to\infty}\sum_{j\le j(n)} \Big(\mathbb{E}(\xi_{nj}\cdot I(|\xi_{nj}| \le \delta_3))\Big)^2 = \lim_{n\to\infty}\Big(\mathbb{E}(\xi_1\cdot I(|\xi_1| \le \delta_3 j(n)^{1/2}))\Big)^2 = \Big(\mathbb{E}(\xi_1)\Big)^2 = 0$, we get (7.2.6) with $c_2 = 0$.

As to (7.2.4): Since $\sum_{j\le j(n)} \nu_{nj}(C)\cdot\mathbb{E}(\xi_{nj}^2\cdot I(|\xi_{nj}| \le \delta_1)) = [j(n)^{-1}\sum_{j\le j(n)} \nu_{nj}(C)]\cdot\mathbb{E}(\xi_1^2\cdot I(|\xi_1| \le \delta_1 j(n)^{1/2}))$, it follows that

$$\sup_{C\in\mathcal{C}:\nu(C)\le\alpha}\sum_{j\le j(n)} \nu_{nj}(C)\cdot\mathbb{E}(\xi_{nj}^2\cdot I(|\xi_{nj}| \le \delta_1)) = \Big[\sup_{C\in\mathcal{C}:\nu(C)\le\alpha} j(n)^{-1}\sum_{j\le j(n)} \nu_{nj}(C)\Big]\cdot\mathbb{E}(\xi_1^2\cdot I(|\xi_1| \le \delta_1 j(n)^{1/2}))$$

whence by (ii) and the fact that $\lim_{n\to\infty}\mathbb{E}(\xi_1^2\cdot I(|\xi_1| \le \delta_1 j(n)^{1/2})) = \mathbb{E}(\xi_1^2) = 1$ condition (7.2.4) is also fulfilled. □

NOTE: Corollary 7.2.7 can also be proved more directly by application of Corollary 7.1.8.

Considering as in 6.2.8 function-indexed partial-sum processes $S_n = (S_n(f))_{f\in\mathcal{F}}$, defined by

$$S_n(f) := j(n)^{-1}\sum_{j\le j(n)} f(\eta_{nj})\cdot\xi_j,\quad f\in\mathcal{F},$$

$\mathcal{F}$ being countable, uniformly bounded, having uniformly integrable L_2-entropy (whence $(\mathcal{F}, d)$ is totally bounded w.r.t. $d = d_\nu^{(2)}$ for each p-measure ν on $\mathcal{X}$ according to 4.3.21), Corollary 7.1.6 yields the following result (cf. [Zi97], 7.2):

7.2.8. THEOREM.
Let $(\eta_{nj})_{1\le j\le j(n),n\in\mathbb{N}}$ being a triangular array of rowwise independent (but not necessarily identically distributed) re's in $(X,\mathcal{X})$, $(\xi_j)_{j\in\mathbb{N}}$ be a sequence of i i d rv's ξ_j with $\mathbb{E}(\xi_1) = 0$ and $\mathbb{E}(\xi_1^2) = 1$, such that the whole array (η_{nj}) is independent of the sequence (ξ_j). Suppose that there is some p-measure ν on $\mathcal{X}$ such that the following conditions are fulfilled (again with $\nu_{nj} := \mathcal{L}\{\eta_{nj}\}$):

$$\text{(7.2.9)}\qquad \lim_{n\to\infty} j(n)^{-1}\sum_{j\le j(n)} \nu_{nj}(f\cdot g) = \nu(f\cdot g)\quad \textit{for all}\quad f,g\in\mathcal{F}$$

$$\text{(7.2.10)}\qquad \lim_{\alpha\to 0}\limsup_{n\to\infty}\sup_{d_\nu^{(2)}(f,g)\le\alpha} j(n)^{-1}\sum_{j\le j(n)} \nu_{nj}((f-g)^2) = 0.$$

Then

$$\Big(j(n)^{-1/2}\sum_{j\le j(n)} f(\eta_{nj})\cdot\xi_j\Big)_{f\in\mathcal{F}} \xrightarrow[sep]{\mathcal{L}} \mathbb{G}_\nu\quad \textit{in}\ \ l^\infty(\mathcal{F}),$$

where $\mathbb{G}_\nu = (G_\nu(f))_{f\in\mathcal{F}}$ *is a mean-zero Gaussian process with sample paths in* $U^b(\mathcal{F}, d_\nu^{(2)})$ *and* $cov(G_\nu(f), G_\nu(g)) = \nu(f\cdot g)$ *for* $f, g \in \mathcal{F}$.

PROOF. The assertion follows from Corollary 7.1.6 in an analogous way as in the proof of 7.2.7. $\square$

Concerning on the other hand partial-sum processes $S_n = (S_n(f))_{f\in\mathcal{F}}$ with

$$S_n(f) := \sum_{j\le j(n)} f(\eta_j)\cdot\xi_{nj}, \quad f\in\mathcal{F},$$

where $\eta_j, j\in\mathbb{N}$, are i i d re's in $(X,\mathcal{X})$ with $\mathcal{L}\{\eta_j\}=\nu$, and where $(\xi_{nj})_{1\le j\le j(n),n\in\mathbb{N}}$ (with $j(n)\longrightarrow\infty$ as $n\to\infty$) is a triangular array of rowwise independent (but not necessarily identically distributed) rv's such that the whole array (ξ_{nj}) is independent of the sequence (η_j), the following result is mentioned in [Zi97], 4.4:

7.2.11. THEOREM.
Let $\mathbb{E}(\xi_{nj}) = 0$ *for all* $1\le j\le j(n)$ *and* $n\in\mathbb{N}$. *Assume that* $\mathcal{F}$ *has uniformly integrable* L_2*-entropy and that* $\nu(F^2)<\infty$, *where* F *denotes the envelope of* $\mathcal{F}$. *Suppose that the following conditions are fulfilled:*

(i) $\lim_{n\to\infty}\sum_{j\le j(n)}\mathbb{E}\Big(F^2(\eta_j)\,\xi_{nj}^2\cdot I(F(\eta_j)|\xi_{nj}|>\delta)\Big)=0$ *for all* $\delta>0$
(Lindeberg-type condition)

(ii) $\lim_{n\to\infty}\sum_{j\le j(n)}\mathbb{E}(\xi_{nj}^2)=1$.

Then

$$S_n \xrightarrow[sep]{\mathcal{L}} \mathbb{G}_\nu \quad in\ \ l^\infty(\mathcal{F}),$$

where $\mathbb{G}_\nu = (G_\nu(f))_{f\in\mathcal{F}}$ *is a mean-zero Gaussian process with sample paths in* $U^b(\mathcal{F}, d_\nu^{(2)})$ *and* $cov(G_\nu(f), G_\nu(g)) = \nu(f\cdot g)$ *for* $f, g \in \mathcal{F}$.

PROOF. We are going to apply Theorem 7.1.3 with $w_{nj}=\delta_{\eta_j}$ and $d=d_\nu^{(2)}$. For thist, one has to verify that $S_n \xrightarrow[\text{fidi}]{\mathbb{P}} \mathbb{G}_\nu$, but this follows from the classical multivariate CLT for triangular arrays. So it remains to verify (7.1.4) – (7.1.6):

As to (7.1.4): $\sum_{j\le j(n)}\mathbb{E}\Big(w_{nj}(F)|\xi_{nj}|\cdot I(w_{nj}(F)|\xi_{nj}|>\delta_n)\Big)=\sum_{j\le j(n)}\mathbb{E}\Big(F(\eta_j)\,|\xi_{nj}|\cdot I(F(\eta_j)|\xi_{nj}|>\delta_n)\Big)\le\delta_n^{-1}\sum_{j\le j(n)}\mathbb{E}\Big(F^2(\eta_j)\,\xi_{nj}^2\cdot I(F(\eta_j)|\xi_{nj}|>\delta_n)\Big)\longrightarrow 0$ with an appropriate chosen sequence $(\delta_n)_{n\in\mathbb{N}}\subset\mathbb{R}_+$ with $\delta_n\longrightarrow 0$, since by (i)

$$\lim_{n\to\infty}\delta^{-1}\sum_{j\le j(n)}\mathbb{E}\Big(F^2(\eta_j)\,\xi_{nj}^2\cdot I(F(\eta_j)|\xi_{nj}|>\delta)\Big)=0 \quad \text{for all } \delta>0.$$

As to (7.1.6): $\sup_{n\in\mathbb{N}} \sum_{j\le j(n)} \mathbb{E}\Big(w_{nj}(F^2)\xi_{nj}^2 \cdot I(w_{nj}(F)|\xi_{nj}| \le \delta_n)\Big) \le \sup_{n\in\mathbb{N}} \sum_{j\le j(n)} \mathbb{E}(F^2(\eta_j)) \cdot \mathbb{E}(\xi_{nj}^2) = \nu(F^2) \cdot \sup_{n\in\mathbb{N}} \sum_{j\le j(n)} \mathbb{E}(\xi_{nj}^2) < \infty$ by (ii), since $\nu(F^2) < \infty$.

As to (7.1.5) (with $d = d_\nu^{(2)}$): $\sup_{d_\nu^{(2)}(f,g)\le\alpha} \sum_{j\le j(n)} \mathbb{E}\Big((w_{nj}(f) - w_{nj}(g))^2\xi_{nj}^2 \cdot I(w_{nj}(F)|\xi_{nj}| \le \delta_n)\Big) \le \sup_{d_\nu^{(2)}(f,g)\le\alpha} \sum_{j\le j(n)} \mathbb{E}\Big((f(\eta_j) - g(\eta_j))^2\Big) \cdot \mathbb{E}(\xi_{nj}^2) = \sup_{d_\nu^{(2)}(f,g)\le\alpha} \nu((f-g)^2) \cdot \sum_{j\le j(n)} \mathbb{E}(\xi_{nj}^2) \le \alpha^2 \sum_{j\le j(n)} \mathbb{E}(\xi_{nj}^2)$ which implies (7.1.5). □

Finally, concerning function-indexed partial-sum processes with fixed locations and index set $\mathcal{F}$ being countable, uniformly bounded and having uniformly integrable L_2-entropy, we obtain from Theorem 7.2.8:

7.2.12. THEOREM *(cf. [Zi97], 7.3).*
Let $(X, \mathcal{X}) = (I^d, I^d \cap \mathcal{B}^d), d \ge 1,$ ($I^d \equiv [0,1]^d$) and consider (cf. (1.4.2))

$$S_n(f) := n^{-d/2} \sum_{\underline{j}\in J_n} f(\underline{j}/n)\xi_{\underline{j}} \quad , f \in \mathcal{F},$$

($J_n := \{1, ..., n\}^d$), where the $\xi_{\underline{j}}, \underline{j} \in \mathbb{N}^d$, are i i d with $\mathbb{E}(\xi_{\underline{j}}) = 0$ and $\mathbb{E}(\xi_{\underline{j}}^2) = 1$. Let ν be the restriction of the d-dimensional Lebesgue measure λ^d on $I^d \cap \mathcal{B}^d$ and suppose that the following two conditions are fulfilled:

$$\lim_{n\to\infty} n^{-d} \sum_{\underline{j}\in J_n} \delta_{\underline{j}/n}(f \cdot g) = \lambda^d(f \cdot g) \quad \textit{for all} \quad f, g \in \mathcal{F} \tag{7.2.13}$$

$$\lim_{\alpha\to 0} \limsup_{n\to\infty} \sup_{\lambda^d((f-g)^2)\le\alpha} n^{-d} \sum_{\underline{j}\in J_n} \delta_{\underline{j}/n}((f-g)^2) = 0. \tag{7.2.14}$$

Then

$$S_n \xrightarrow[sep]{\mathcal{L}} \mathbb{G}_\nu \quad \textit{in} \quad l^\infty(\mathcal{F}),$$

where $\mathbb{G}_\nu = (G_\nu(f))_{f\in\mathcal{F}}$ is a mean-zero Gaussian process with sample paths in $U^b(\mathcal{F}, d_{\lambda^d}^{(2)})$ and $cov(G_\nu(f), G_\nu(g)) = \lambda^d(f \cdot g)$ for $f, g \in \mathcal{F}$.

In the set-indexed case, i.e. for (cf. (1.4.2))

$$S_n(C) = n^{-d/2} \sum_{\underline{j}\in J_n} 1_C(\underline{j}/n)\xi_{\underline{j}} \quad , C \in \mathcal{C},$$

with $\mathcal{C} \subset I^d \cap \mathcal{B}^d$, attempts to find natural conditions under which (7.2.13) and (7.2.14) hold have been made in [Al87]; cf. also [Gae94], Remark 2.16 and the results contained in [Va96], Section 2.12.2: Partial-Sum Processes on Lattices.

7.3 FCLT's for empirical processes

Let $X = (X, \mathcal{X})$ be again an arbitrary measurable space (sample space) and $(\eta_{nj})_{1\le j\le j(n), n\in\mathbb{N}}$ (with $j(n) \longrightarrow \infty$ as $n \to \infty$) be a triangular array of re's in $(X, \mathcal{X})$ assumed to be rowwise independent (but not necessarily identically distributed) with law $\mathcal{L}\{\eta_{nj}\} = \nu_{nj}$). Let $\mathcal{F}$ be a class of $\mathcal{X}$-measurable functions $f : X \longrightarrow \mathbb{R}$ with $\mathcal{X}$-measurable envelope $F : X \longrightarrow \mathbb{R}$; $\mathcal{F}$ being countable for simplicity. Assume that $\mathcal{F}$ has uniformly integrable L_2-entropy and that there is some pseudo-metric d on $\mathcal{F}$ such that $(\mathcal{F}, d)$ is totally bounded (e.g. $d = d_\nu^{(2)}$ for some p-measure ν on $\mathcal{X}$ with $\nu(F^2) < \infty$; cf. 4.3.21).

We are going to apply our FCLT for RMP's 7.1.3 with $w_{nj} = \delta_{\eta_{nj}}$ and $\xi_{nj} = j(n)^{-1/2}$ to obtain the following

FCLT for empirical processes in the non-i i d -case,

i.e. for $S_n = (S_n(f))_{f\in\mathcal{F}}$ with

$$S_n(f) := j(n)^{-1/2} \sum_{j\le j(n)} (f(\eta_{nj}) - \nu_{nj}(f)), \quad f \in \mathcal{F}.$$

7.3.1. THEOREM *(cf. [Zi97], 4.2).*
Assume $S_n \xrightarrow[fidi]{\mathbb{P}} \bar{\mathbb{G}}$, *where* $\bar{\mathbb{G}} = (\bar{G}(f))_{f\in\mathcal{F}}$ *is a mean-zero Gaussian process. Let*

$$\bar{\nu}_n := j(n)^{-1} \sum_{j\le j(n)} \nu_{nj}\ , \qquad a_n(\alpha) := \sup_{d(f,g)\le\alpha} \Big(\bar{\nu}_n((f-g)^2)\Big)^{1/2}, \alpha > 0,$$

and suppose that the following conditions are fulfilled:

(7.3.2) $\qquad \sup_{n\in\mathbb{N}} \bar{\nu}_n(F^2) < \infty$ *(whence S_n has its sample paths in $l^\infty(\mathcal{F})$)*

(7.3.3) $\qquad \lim_{\alpha\to 0} \limsup_{n\to\infty} a_n(\alpha) = 0$

(7.3.4) $\qquad \lim_{n\to\infty} j(n)^{-1} \sum_{j\le j(n)} \mathbb{E}\Big(F^2(\eta_{nj}) \cdot I(F(\eta_{nj}) > \delta j(n)^{1/2})\Big) = 0$ *for all* $\delta > 0$.

Then

$$S_n \xrightarrow[sep]{\mathcal{L}} \mathbb{G} \quad \text{in } l^\infty(\mathcal{F}),$$

where $\mathbb{G} = (G(f))_{f\in\mathcal{F}}$ *is a mean-zero Gaussian process with sample paths in* $U^b(\mathcal{F}, d)$ *and* $\mathbb{G} \overset{\mathcal{L}}{\underset{fidi}{=}} \bar{\mathbb{G}}$.

PROOF. The proof runs along the same pattern as the proof of Theorem 7.2.11. According to Theorem 7.1.3 we have to verify (7.1.4) – (7.1.6) (with $w_{nj} = \delta_{\eta_{nj}}$ and $\xi_{nj} = j(n)^{-1/2}$):

As to (7.1.4): $\sum_{j\le j(n)} \mathbb{E}\Big(w_{nj}(F)|\xi_{nj}|\cdot I(w_{nj}(F)|\xi_{nj}| > \delta_n)\Big) = \sum_{j\le j(n)} \mathbb{E}\Big(F(\eta_{nj})\, j(n)^{-1/2} \cdot I(F(\eta_{nj}) > \delta_n j(n)^{1/2})\Big) \le \delta_n^{-1} j(n)^{-1} \sum_{j\le j(n)} \mathbb{E}\Big(F^2(\eta_{nj}) \cdot I(F(\eta_{nj}) > \delta_n j(n)^{1/2})\Big) \longrightarrow 0$ with an appropriate chosen sequence $(\delta_n)_{n\in\mathbb{N}} \subset \mathbb{R}_+$ with $\delta_n \longrightarrow 0$, since by (7.3.4)

$$\delta^{-1} j(n)^{-1} \sum_{j\le j(n)} \mathbb{E}\Big(F^2(\eta_{nj}) \cdot I(F(\eta_{nj}) > \delta j(n)^{1/2})\Big) \longrightarrow 0$$

as $n \longrightarrow \infty$ for all $\delta > 0$.

As to (7.1.6): $\sup_{n\in\mathbb{N}} \sum_{j\le j(n)} \mathbb{E}\Big(w_{nj}(F^2)\xi_{nj}^2 \cdot I(w_{nj}(F)|\xi_{nj}| \le \delta_n)\Big) = \sup_{n\in\mathbb{N}} \sum_{j\le j(n)} \mathbb{E}(F^2(\eta_{nj})j(n)^{-1} \cdot I(F(\eta_{nj}) \le \delta_n j(n)^{1/2})\Big) \le \sup_{n\in\mathbb{N}} j(n)^{-1} \sum_{j\le j(n)} \mathbb{E}(F^2(\eta_{nj})) = \sup_{n\in\mathbb{N}} j(n)^{-1} \sum_{j\le j(n)} \nu_{nj}(F^2) = \sup_{n\in\mathbb{N}} \bar{\nu}_n(F^2) < \infty$ by (7.3.2).

As to (7.1.5): $\sup_{d(f,g)\le\alpha} \sum_{j\le j(n)} \mathbb{E}\Big((w_{nj}(f)-w_{nj}(g))^2\xi_{nj}^2 \cdot I(w_{nj}(F)|\xi_{nj}| \le \delta_n)\Big) = \sup_{d(f,g)\le\alpha} \sum_{j\le j(n)} \mathbb{E}\Big((f(\eta_{nj})-g(\eta_{nj}))^2 j(n)^{-1} \cdot I(F(\eta_{nj}) \le \delta_n j(n)^{1/2})\Big) \le \sup_{d(f,g)\le\alpha} j(n)^{-1} \sum_{j\le j(n)} \mathbb{E}\Big((f(\eta_{nj})-g(\eta_{nj}))^2\Big) = \sup_{d(f,g)\le\alpha} j(n)^{-1} \sum_{j\le j(n)} \nu_{nj}((f-g)^2) = \sup_{d(f,g)\le\alpha} \bar{\nu}_n((f-g)^2) = a_n^2(\alpha)$,
from which (7.1.5) follows according to (7.3.3). □

Replacing the triangular array (η_{nj}) by a sequence $(\eta_j)_{j\in\mathbb{N}}$ of i i d re's in $(X, \mathcal{X})$ with law $\mathcal{L}\{\eta_j\} = \nu$, we obtain from Theorem 7.3.1 the following FCLT for *empirical $\mathcal{F}$-processes* in the i i d -case, i.e. for $\beta_n = (\beta_n(f))_{f\in\mathcal{F}}$ with $\beta_n(f) := n^{-1/2} \sum_{j\le n}(f(\eta_j) - \nu(f)) = n^{1/2}(\nu_n(f) - \nu(f))$, where $\nu_n(f) := n^{-1} \sum_{j\le n} f(\eta_j)$ (cf. 2.2.1 in the set-indexed case):

7.3.5. THEOREM.
Suppose that $\mathcal{F}$ has uniformly integrable L_2-entropy and that $\nu(F^2) < \infty$ ($\mathcal{F}$ being countable for simplicity). Then

$$\beta_n \xrightarrow[sep]{\mathcal{L}} \mathbb{G}_\nu \quad \text{in } l^\infty(\mathcal{F}),$$

where $\mathbb{G}_\nu = (G_\nu(f))_{f\in\mathcal{F}}$ is a mean-zero Gaussian process with sample paths in $U^b(\mathcal{F}, d_\nu^{(2)})$ and $cov(G_\nu(f), G_\nu(g)) = \nu(f \cdot g) - \nu(f) \cdot \nu(g)$ for $f, g \in \mathcal{F}$.

PROOF. $\beta_n \xrightarrow[\text{fidi}]{\mathbb{P}} \mathbb{G}_\nu$ follows by the classical multivariate CLT. The conditions (7.3.2) and (7.3.3) are obviously fulfilled. As to (7.3.4) we have in the present case

$$\lim_{n\to\infty} n^{-1} \sum_{j\le n} \mathbb{E}\Big(F^2(\eta_j) \cdot I(F(\eta_j) > \delta n^{1/2})\Big) = \lim_{n\to\infty} \mathbb{E}\Big(F^2(\eta_1) \cdot I(F(\eta_1) > \delta n^{1/2})\Big) = 0,$$

since $\mathbb{E}(F^2(\eta_1)) = \nu(F^2) < \infty$. Thus Theorem 7.3.1 yields the assertion. □

7.3.6. REMARK.
Concerning VCGC's $\mathcal{F}$ (having uniformly integrable L_2-entropy according to 4.3.21) with $\nu(F^2) < \infty$, the assertion of Thoerem 7.3.5 holds true, especially for $\mathcal{F} = \{1_C : C \in \mathcal{C}\}, \mathcal{C} \subset \mathcal{X}$ being a countable VCC; see Theorem 2.2.1.

7.4 FCLT's for smoothed empirical processes

Throughout this section X is supposed to be an arbitrary *linear metric space* endowed with its Borel σ-field $\mathcal{X}$. The basic situation is the same as in Section 6.4, i.e. given i i d re's $\eta_j, j \in \mathbb{N}$, in $(X, \mathcal{X})$ with law $\mathcal{L}\{\eta_j\} = \nu$ on $\mathcal{X}$ we consider the smoothed empirical measures

$$\tilde{\nu}_n := \nu_n \star \mu_n \ , \quad n \in \mathbb{N},$$

indexed by classes $\mathcal{F}$ of $\mathcal{X}$-measurable functions $f : X \longrightarrow \mathbb{R}$ with $\mathcal{X}$-measurable envelope $F : X \longrightarrow \mathbb{R}$ assuming that $\nu(F^2) := \int_X F^2 \, d\nu < \infty$. Remember from 6.4 that

$$\text{(7.4.1)} \qquad \tilde{\nu}_n(f) = \int_X \int_X f(x+y) \nu_n(dx) \mu_n(dy) = n^{-1} \sum_{j \leq n} \int_X f(\eta_j + y) \mu_n(dy)$$

$$\text{and} \quad \mathbb{E}(\tilde{\nu}_n(f)) = \nu \star \mu_n(f) \quad \forall f \in \mathcal{F} \text{ (cf. (6.4.3) and (6.4.4)).}$$

Also our decomposition from 6.4 will be again important, i.e.

$$\text{(7.4.2)} \qquad \tilde{\nu}_n - \nu = \tilde{\nu}_n - \nu \star \mu_n + \nu \star \mu_n - \nu,$$

noticing that $\tilde{\nu}_n - \nu \star \mu_n = \tilde{\nu}_n - \mathbb{E}(\tilde{\nu}_n)$ is a mean-zero RMP and where $\nu \star \mu_n - \nu$ is the *non-stochastic* bias term.

As in 6.4 let $\tilde{\mathcal{F}}$ be the class of all translates f_x of elements f of $\mathcal{F}$ (with $f_x(y) := f(x+y), y \in X$). **Without imposing the condition $\mathcal{F} = \tilde{\mathcal{F}}$** we are going to apply our FCLT for RMP's 7.1.3 with (cf. (7.4.1))

$$w_{nj}(f) := \int_X f(\eta_j + y) \, \mu_n(dy), \quad \textit{and} \quad \xi_{nj} := n^{-1/2} \ , 1 \leq j \leq j(n) := n, n \in \mathbb{N},$$

to obtain sufficient conditions under which

$$\text{(7.4.3)} \qquad \Big(\sqrt{n}(\tilde{\nu}_n(f) - \nu(f))\Big)_{f \in \mathcal{F}} \xrightarrow[\text{sep}]{\mathcal{L}} \mathbb{G} \quad \text{in } l^\infty(\mathcal{F}),$$

where $\mathbb{G} = (G(f))_{f \in \mathcal{F}}$ is a mean-zero Gaussian process with sample paths in $U^b(\mathcal{F}, d_\nu^{(2)})$.
Remember that $U^b(\mathcal{F}, d_\nu^{(2)})$ is a separable subspace of $(l^\infty(\mathcal{F}), || \cdot ||_{\mathcal{F}})$ if and only if $(\mathcal{F}, d_\nu^{(2)})$ is totally bounded ([Gae90], Corollary 2) where the latter is true for classes $\mathcal{F}$ having uniformly integrable L_2-entropy (cf. 4.3.21).

Now, in view of (7.4.2), Theorem 7.1.3 yields immediately the following FCLT. (Note that the η_j's on which the w_{nj}'s are based are i i d .)

7.4.4. THEOREM *(cf. [Ro97], Theorem 3.2.2).*
Let X be a linear metric space and let $\mathcal{F}$ have uniformly integrable L_2-entropy. Assume that the following conditions (7.4.5) – (7.4.8) are fulfilled:

For each $\rho > 0$ there exists $\delta_n \equiv \delta_n(\rho), n \in \mathbb{N},\ \delta_n \to 0$ such that

$$\limsup_{n\to\infty} \sqrt{n}\cdot \mathbb{E}\Big(w_{n1}(F)\, I(w_{n1}(F) > \delta_n\sqrt{n})\Big) \le \rho \tag{7.4.5}$$

$$\lim_{\alpha\to 0}\limsup_{n\to\infty} \sup_{d_\nu^{(2)}(f,g)\le\alpha} \mathbb{E}\Big((w_{n1}(f) - w_{n1}(g))^2)\cdot I(w_{n1}(F) \le \delta_n\sqrt{n})\Big) = 0 \tag{7.4.6}$$

$$\sup_{n\in\mathbb{N}} \mathbb{E}\Big(w_{n1}(F^2)\cdot I(w_{nj}(F) \le \delta_n\sqrt{n})\Big) < \infty \tag{7.4.7}$$

$$\sup_{f\in\mathcal{F}} \sqrt{n}|\nu \star \mu_n(f) - \nu(f)| \longrightarrow 0\ . \tag{7.4.8}$$

Assume, in addition, that there exists a mean-zero Gaussian process $\bar{\mathbb{G}} = (\bar{\mathbb{G}}(f))_{f\in\mathcal{F}}$ such that $\sqrt{n}(\tilde{\nu}_n - \nu \star \mu_n) \xrightarrow[fidi]{\mathbb{P}} \bar{\mathbb{G}}$.

Then there exists a mean-zero Gaussian process $\mathbb{G} = (\mathbb{G}(f))_{f\in\mathcal{F}}$ with sample paths in $U^b(\mathcal{F}, d_\nu^{(2)})$ such that

$$\sqrt{n}(\tilde{\nu}_n - \nu) \xrightarrow[sep]{\mathcal{L}} \mathbb{G} \qquad \text{in } l^\infty(\mathcal{F}) \text{ and } \mathbb{G} \overset{\mathcal{L}}{\underset{fidi}{=}} \bar{\mathbb{G}}.$$

Before going further, let us have a view on a FCLT for smoothed empirical processes **under the condition $\boldsymbol{\mathcal{F} = \tilde{\mathcal{F}}}$**:
From van der Vaart [Va94] one gets the following result. For this, $\mathcal{F}$ is called a *ν-Donsker class* if (cf. Theorem 7.3.5) $\Big(\sqrt{n}(\nu_n(f) - \nu(f))\Big)_{f\in\mathcal{F}} \xrightarrow[\text{sep}]{\mathcal{L}} \mathbb{G}_\nu$ in $l^\infty(\mathcal{F})$.

7.4.9. THEOREM.
Let $X = \mathbb{R}^d, d \ge 1$, and assume $\mathcal{F} = \tilde{\mathcal{F}}$. Let $\mathcal{F}$ be ν-Donsker and μ_n p-measures on $\mathcal{B}^d$ with $\mu_n \longrightarrow \delta_0$ weakly. Suppose that the following two conditions are fulfilled:

$$\sup_{f\in\mathcal{F}} \int_X \Big[\int_X (f(x+y) - f(x))\mu_n(dy)\Big]^2 \nu(dx) \longrightarrow 0 \tag{7.4.10}$$

$$\sup_{f\in\mathcal{F}} \sqrt{n}\big|\nu \star \mu_n(f) - \nu(f)\big| \longrightarrow 0 \tag{7.4.11}$$

Then

$$\sqrt{n}(\tilde{\nu}_n - \nu) \xrightarrow[sep]{\mathcal{L}} \mathbb{G} \qquad \text{in } l^\infty(\mathcal{F}),$$

where $\mathbb{G} = (\mathbb{G}(f))_{f\in\mathcal{F}}$ is a mean-zero Gaussian process with sample paths in $U^b(\mathcal{F}, d_\nu^{(2)})$.

NOTE: In [Va94] the μ_n's are even allowed to be *random* p-measures on $\mathcal{B}^d$. On the other hand it should be noted that $\mathcal{F}$ being a ν-Donsker class does in general not imply that $\tilde{\mathcal{F}}$ is also a ν-Donsker class (cf. Example 2.2.8 in [Ro97]) as it is imposed in 7.4.9 via the $\mathcal{F} = \tilde{\mathcal{F}}$ - assumption. See also [Ro99] for a comparison with a result of Yukich [Yu92] obtained in the case $X = \mathbb{R}^d, d \ge 1, \mathcal{F} = \tilde{\mathcal{F}}, \mathcal{F}$ a ν-Donsker class. The method of proof in [Va94] and [Yu92], respectively, is completely different from our approach via RMP's (see [Ro99] for a discussion). Their key method consists of showing asymptotic (stochastic) equivalence of the empirical process $\sqrt{n}(\nu_n - \nu)$ and the unbiased smoothed empirical processes $\sqrt{n}(\tilde{\nu}_n - \nu \star \mu_n)$ in order to apply the Cramér-Slutzky-type result (cf. Theorem 2.3.15).

Now, in view of Theorem 7.4.9 we will present in the following more tractable conditions compared with those in Theorem 7.4.4, but again **without imposing the condition $\boldsymbol{\mathcal{F} = \tilde{\mathcal{F}}}$**:

7.4.12. THEOREM *(cf. [Ro97], Theorem 3.2.3).*
Let X be a linear metric space and $\mathcal{F}$ have uniformly integrable L_2-entropy. Assume that the following conditions (7.4.13) and (7.4.14) are fulfilled:

$$(7.4.13) \qquad \sup_{f \in \mathcal{F}} \int_X \Big[\int_X (f(x+y) - f(x)) \mu_n(dy) \Big]^2 \nu(dx) \longrightarrow 0$$

$$(7.4.14) \qquad \sup_{f \in \mathcal{F} \cup \{F^3\}} \sqrt{n} \big| \nu \star \mu_n(f) - \nu(f) \big| \longrightarrow 0.$$

Then

$$\sqrt{n}(\tilde{\nu}_n - \nu) \xrightarrow[sep]{\mathcal{L}} \mathbb{G}_\nu \qquad \text{in } l^\infty(\mathcal{F}),$$

where $\mathbb{G}_\nu = (\mathbb{G}_\nu(f))_{f \in \mathcal{F}}$ is a mean-zero Gaussian process with sample paths in $U^b(\mathcal{F}, d_\nu^{(2)})$ and $cov(G_\nu(f), G_\nu(g)) = \nu(f \cdot g) - \nu(f) \cdot \nu(g)$ for $f, g \in \mathcal{F}$.

PROOF. To prove this result, one shows (7.4.5) – (7.4.7), fidi-convergence and then one applies Theorem 7.4.4. This is carried out in [Ro99].

NOTE: Condition (7.4.13) is just (7.4.10), whereas (7.4.14) is apparently a bit stronger than (7.4.11) implying convergence also for F^3. In (7.4.14) it is tacitly understood that $\nu \star \mu_n(F^3)$ and $\nu(F^3)$ exist. (7.4.14) can be replaced by (7.4.11) if, in addition, $\nu \star \mu_n(F^{2+\varepsilon}) \longrightarrow \nu(F^{2+\varepsilon})$ for some $\varepsilon > 0$. On the other hand, as already mentioned in Section 6.4, the condition $\mathcal{F} = \tilde{\mathcal{F}}$ implies that $\mathcal{F}$ is uniformly bounded. But for uniformly bounded $\mathcal{F}$ (7.4.14) reduces to (7.4.11), and so we get finally the following result:

7.4.15. THEOREM *(cf. [Ro97], Theorem 3.2.4).*
Let X be a linear metric space and let $\mathcal{F}$ be uniformly bounded having uniformly integrable L_2-entropy. Assume that the conditions (7.4.8) and (7.4.13) are satisfied. Then

$$\sqrt{n}(\tilde{\nu}_n - \nu) \xrightarrow[sep]{\mathcal{L}} \mathbb{G}_\nu \qquad \text{in } l^\infty(\mathcal{F}),$$

where $\mathbb{G}_\nu = (\mathbb{G}_\nu(f))_{f \in \mathcal{F}}$ is a mean-zero Gaussian process with sample paths in $U^b(\mathcal{F}, d_\nu^{(2)})$ and $cov(G_\nu(f), G_\nu(g)) = \nu(f \cdot g) - \nu(f) \cdot \nu(g)$ for $f, g \in \mathcal{F}$.

7.5 A uniform FCLT for the unbiased smoothed empirical process

As in Section 7.4 X is supposed to be an arbitrary *linear metric space* endowed with its Borel σ-field $\mathcal{X}$ and $\eta_j, j \in \mathbb{N}$, are i i d re's in $(X, \mathcal{X})$ with law $\mathcal{L}\{\eta_j\} = \nu$ on $\mathcal{X}$.
Let us consider first the non-smoothed empirical process $\mathbb{G}_n^\nu := \sqrt{n}(\nu_n - \nu)$ indexed by a (countable, for

simplicity) class $\mathcal{F}$ of $\mathcal{X}$-measurable functions $f : X \longrightarrow \mathbb{R}$ with $\mathcal{X}$-measurable envelope $F : X \longrightarrow \mathbb{R}$ such that $\nu(F^2) < \infty$.

As we know from (7.1.2) and the general comments preceeding (7.1.2) we have

$$\mathbb{G}_n^\nu \xrightarrow[\text{sep}]{\mathcal{L}} \mathbb{G}_\nu \quad \text{in} \quad l^\infty(\mathcal{F}) \quad \Longleftrightarrow \quad d_{BL}(\mathbb{G}_n^\nu, \mathbb{G}_\nu) \longrightarrow 0,$$

where $\mathbb{G}_\nu = (G_\nu(g))_{f \in \mathcal{F}}$ is a mean-zero Gaussian process with sample paths in $U^b(\mathcal{F}, d_\nu^{(2)})$ and

$$cov(G_\nu(f), G_\nu(g)) = \nu(f \cdot g) - \nu(f) \cdot \nu(g) \quad \text{for} \;\; f, g \in \mathcal{F},$$

calling (as in Section 7.4) $\mathcal{F}$ to be a ν-Donsker class in this case.

Now, concerning the question whether the class $\mathcal{F}$ is also a *uniform Donsker class*, i.e whether $\sup_{\nu \in \mathcal{M}^1(X)} d_{BL}(\mathbb{G}_n^\nu, \mathbb{G}_\nu) \longrightarrow 0$ (and $(\mathcal{F}, d_\nu^{(2)})$ is totally bounded uniformly in ν), where $\mathcal{M}^1(X)$ denotes the class of all p-measures on $\mathcal{X}$, the following result is known (see [Gi97], Theorem 5.3 and [Gi91]):

7.5.1. THEOREM.
Let X be an arbitrary measurable space and $\mathcal{F}$ be uniformly bounded having uniformly integrable L_2-entropy. Then (with $\mathcal{L}\{\mathbb{G}_\nu\}$ being tight)

$$\sup_{\nu \in \mathcal{M}^1(X)} d_{BL}(\mathbb{G}_n^\nu, \mathbb{G}_\nu) \longrightarrow 0. \tag{7.5.2}$$

Uniform Donsker classes were e.g. studied by Sheehy and Wellner [She92] (who also studied in detail (7.5.2) with the supremum taken over subclasses of $\mathcal{M}^1(X)$) and by Giné and Zinn [Gi91]. They showed that (putting measurablilty questions aside) a so-called *uniformly pregaussian class* $\mathcal{F}$ (saying $\mathcal{F}$ is UPG) is a uniform Donsker class.

$\mathcal{F}$ is UPG means that the following two conditions are fulfilled:

$$\sup_{\nu \in \mathcal{M}^1(X)} \mathbb{E}(||\mathbb{Z}_\nu||_{\mathcal{F}}) < \infty \tag{7.5.3}$$

$$\lim_{\delta \to 0} \sup_{\nu \in \mathcal{M}^1(X)} \mathbb{E}\Big(\sup\{|Z_\nu(f) - Z_\nu(g)| : f, g \in \mathcal{F}, \; d_\nu^{(2)}(f,g) \le \delta\}\Big) = 0. \tag{7.5.4}$$

In both conditions $\mathbb{Z}_\nu$ can be replaced by $\mathbb{G}_\nu$ (see [Gi97], Theorem 5.3).

Here $\mathbb{Z}_\nu = (Z_\nu(f))_{f \in \mathcal{F}}$ stands for a mean-zero Gaussian process with tight law $\mathcal{L}\{\mathbb{Z}_\nu\}$ on $\mathcal{B}(l^\infty(\mathcal{F}))$ whose covariance structure is given by

$$cov(Z_\nu(f), Z_\nu(g)) = \nu(f \cdot g) \quad \text{for} \;\; f, g \in \mathcal{F}.$$

From [Va96], Example 1.5.10 it follows that also $\mathbb{Z}_\nu$ (as $\mathbb{G}_\nu$) can be chosen to have its sample paths in $U^b(\mathcal{F}, d_\nu^{(2)})$; note that $d_\nu^{(2)}$ coincides with the so-called *intrinsic pseudo-metric* $\rho_{\mathbb{Z}_\nu}$ (on $\mathcal{F}$) for $\mathbb{Z}_\nu$, i.e.

$$\rho_{\mathbb{Z}_\nu}(f,g) := \mathbb{E}^{1/2}(|Z_\nu(f) - Z_\nu(g)|^2) = d_\nu^{(2)}(f,g) \quad \text{for } f, g \in \mathcal{F}.$$

Now we are going to establish an analogous result as (7.5.2) for smoothed empirical processes under conditions similar to (7.5.3) and (7.5.4) replacing $\mathbb{G}_n^\nu$ by the unbiased smoothed empirical process

$$\tilde{\mathbb{G}}_n^\nu := (\tilde{G}_n^\nu(f))_{f\in\mathcal{F}},$$

where $\tilde{G}_n^\nu(f) := \sqrt{n}(\tilde{\nu}_n(f) - \nu \star \mu_n(f)),\ f \in \mathcal{F}$.

7.5.5. THEOREM *(cf. [Ro97], Theorem 4.4).*
Let X be a linear metric space and $\mathcal{F}$ be uniformly bounded. Suppose that for every $\nu \in \mathcal{M}^1(X)$ there is a mean-zero Gaussian process $\mathbb{G}_\nu = (G_\nu(h))_{h\in\mathcal{F}\cup\mathcal{G}}$ with tight law $\mathcal{L}\{\mathbb{G}_\nu\}$ on $\mathcal{B}(l^\infty(\mathcal{F}\cup\mathcal{G}))$ where

$$\mathcal{G} := \{g_{f,\mu_n} : f \in \mathcal{F}, n \in \mathbb{N}\} \quad \textit{with } g_{f,\mu_n}(x) := \int_X f(x+y)\mu_n(dy), x \in X.$$

Assume that the following conditions (7.5.6) – (7.5.8) are fulfilled:

$$\sup_{\nu\in\mathcal{M}^1(X)} \mathbb{E}(||\mathbb{G}_\nu||_\mathcal{F}) < \infty \tag{7.5.6}$$

$$\lim_{\delta\to 0} \sup_{\nu\in\mathcal{M}^1(X)} \mathbb{E}\Big(\sup\big\{|G_\nu(h_1) - G_\nu(h_1)| : h_1, h_2 \in \mathcal{F}\cup\mathcal{G}, d_\nu^{(2)}(h_1,h_2) \le \delta\big\}\Big) = 0 \tag{7.5.7}$$

$$\sup_{\nu\in\mathcal{M}^1(X)} \sup_{f\in\mathcal{F}} \int_X \big[\int_X (f(x+y) - f(x))\mu_n(dy)\big]^2 \nu(dx) \longrightarrow 0. \tag{7.5.8}$$

Then

$$\sup_{\nu\in\mathcal{M}^1(X)} d_{BL}(\tilde{\mathbb{G}}_n^\nu, \mathbb{G}_\nu) \longrightarrow 0. \tag{7.5.9}$$

In (7.5.9) not only $\tilde{\mathbb{G}}_n^\nu$ but also $\mathbb{G}_\nu$ (restricted to the index set $\mathcal{F}$) is considered as a process with sample paths in $l^\infty(\mathcal{F})$ whose law $\mathcal{L}\{\mathbb{G}_\nu\}$ is tight on $\mathcal{B}(l^\infty(\mathcal{F}))$; so d_{BL} in (7.5.9) stands for the bounded Lipschitz distance based on $l^\infty(\mathcal{F})$ (and not on $l^\infty(\mathcal{F}\cup\mathcal{G})$).

PROOF. We follow the lines of proof of (7.5.2) as given in [Gi97], respectively the lines of proof of Theorem 2.3 in [Gi91] under the conditions (7.5.3) and (7.5.4) using Gaussian comparison methods.
First, we show that for each $\tau > 0$

$$\sup_{\nu\in\mathcal{M}^1(X)} N(\tau, \mathcal{F}, d_\nu^{(2)}) < \infty. \tag{7.5.10}$$

As to (7.5.10), according to Sudakov's Inequality (cf. [Va96], A.2.5) there exists a constant $0 < K < \infty$ such that for every $\nu \in \mathcal{M}^1(X)$

$$\Big(\log N(\tau, \mathcal{F}, \rho_{\mathbb{Z}^\nu})\Big)^{1/2} \le K \cdot \mathbb{E}(||\mathbb{Z}_\nu||_\mathcal{F})$$

(with $\rho_{\mathbb{Z}_\nu}(f,g) := \mathbb{E}^{1/2}(|Z_\nu(f) - Z_\nu(g)|^2) = d_\nu^{(2)}(f,g) \quad \forall\, f,g \in \mathcal{F}$)
whence

$$\Big(\log N(\tau, \mathcal{F}, d_\nu^{(2)})\Big)^{1/2} \le K \cdot \mathbb{E}(||\mathbb{Z}_\nu||_{\mathcal{F}}).$$

Now, let g be a standardnormal distributed rv which is independent of $\mathbb{G}_\nu$; then

$$\mathcal{L}\{\mathbb{G}_\nu + g \cdot \nu\} = \mathcal{L}\{\mathbb{Z}_\nu\}$$

(as can be seen by computing covariances), so

$$\mathbb{E}(||\mathbb{Z}_\nu||_{\mathcal{F}}) \le \mathbb{E}(||\mathbb{G}_\nu||_{\mathcal{F}}) + \mathbb{E}(|g|) \cdot \sup_{f \in \mathcal{F}} |\nu(f)|,$$

whence (7.5.6) together with $\mathcal{F}$ being uniformly bounded yields (7.5.10).

Next, let $k := \sup_{\nu \in \mathcal{M}^1(X)} N(\tau, \mathcal{F}, d_\nu^{(2)})$, and for each $\nu \in \mathcal{M}^1(X)$ let $f_1, ..., f_k \in \mathcal{F}$ denote the centers of the $d_\nu^{(2)}$-balls with radius τ that cover $\mathcal{F}$. (Note that, of course, $f_1, ..., f_k$ depend on ν.) Then for each $f \in \mathcal{F}$ let $\pi_\tau(f) \in \{f_1, ..., f_k\}$ be such that

$$d_\nu^{(2)}(\pi_\tau(f), f) \le \tau \quad (\text{ where w.l.o.g. } \ \pi_\tau(f_i) = f_i \ \ \forall\, i = 1, ..., k).$$

This allows us to define the processes $\tilde{\mathbb{G}}_n^\nu(\pi_\tau) = (\tilde{G}_n^\nu(\pi_\tau)(f))_{f \in \mathcal{F}}$ and $\mathbb{G}_\nu(\pi_\tau) = (G_\nu(\pi_\tau)(f))_{f \in \mathcal{F}}$ with sample paths in $l^\infty(\mathcal{F})$ by

$$\begin{aligned} \tilde{G}_n^\nu(\pi_\tau)(f) &:= \tilde{G}_n^\nu(\pi_\tau(f)) \\ \text{and} \quad G_\nu(\pi_\tau)(f) &:= G_\nu(\pi_\tau(f)) \quad , f \in \mathcal{F}. \end{aligned}$$

Then, for each $H \in BL_1(l^\infty(\mathcal{F}))$ we have the decomposition

$$\begin{aligned} \big|\mathbb{E}^*(H(\tilde{\mathbb{G}}_n^\nu)) - \mathbb{E}(H(\mathbb{G}_\nu))\big| &\le \big|\mathbb{E}^*(H(\tilde{\mathbb{G}}_n^\nu)) - \mathbb{E}(H(\tilde{\mathbb{G}}_n^\nu(\pi_\tau)))\big| \\ &\quad + \big|\mathbb{E}(H(\tilde{\mathbb{G}}_n^\nu(\pi_\tau))) - \mathbb{E}(H(\mathbb{G}_\nu(\pi_\tau)))\big| \\ &\quad + \big|\mathbb{E}(H(\mathbb{G}_\nu(\pi_\tau))) - \mathbb{E}(H(\mathbb{G}_\nu))\big| \\ &=: I_{1n} + I_{2n} + I_{3n}. \end{aligned}$$

We will show

$$\limsup_{\tau \to 0} \limsup_{n \to \infty} \sup_{\nu \in \mathcal{M}^1(X)} \sup_{H \in BL_1(l^\infty(\mathcal{F}))} I_{in} = 0 \text{ for } i = 1,2,3. \tag{7.5.11}$$

As to I_{3n}: Since $H \in BL_1(l^\infty(\mathcal{F}))$ we have

$$\begin{aligned} &\big|\mathbb{E}(H(\mathbb{G}_\nu(\pi_\tau))) - \mathbb{E}(H(\mathbb{G}_\nu))\big| \\ &\le \mathbb{E}(||\,\mathbb{G}_\nu(\pi_\tau) - \mathbb{G}_\nu||_{\mathcal{F}}) \ \le \ \mathbb{E}(\sup\{|\mathbb{G}_\nu(f) - \mathbb{G}_\nu(g)| : f,g \in \mathcal{F}, d_\nu^{(2)}(f,g) \le \tau\}), \end{aligned}$$

so (7.5.7) yields (7.5.11) for $i = 3$.

As to I_{2n}: Let $\nu \in \mathcal{M}^1(X)$ and $H \in BL_1(l^\infty(\mathcal{F}))$ be arbitrary. Then there exists a bounded Lipschitz function $L : \mathbb{R}^k \longrightarrow \mathbb{R}$ s.t.

$$H(x(\pi_\tau)) = L((x(f_1), ..., x(f_k))) \quad \forall\, x \in l^\infty(\mathcal{F})$$

with $x(\pi_\tau)(f) := x(\pi_\tau(f)), f \in \mathcal{F}$. So we obtain

$$\begin{aligned} &\big|\mathbb{E}(H(\tilde{\mathbb{G}}_n^\nu(\pi_\tau))) - \mathbb{E}(H(\mathbb{G}_\nu(\pi_\tau)))\big| \\ &\le d_{BL}\big((\tilde{\mathbb{G}}_n^\nu(f_1), ..., \tilde{\mathbb{G}}_n^\nu(f_k))^t, (\mathbb{G}_\nu(f_1), ..., \mathbb{G}_\nu(f_k))^t\big), \end{aligned}$$

where the superscript t denotes the transposed vector and here d_{BL} is the bounded Lipschitz metric on the space of all p-measures (laws) on $\mathcal{B}^k$ in $\mathbb{R}^k$.
Now $(\tilde{\mathbb{G}}_n^\nu(f_1), ..., \tilde{\mathbb{G}}_n^\nu(f_k))^t = \sqrt{n}\, n^{-1} \sum_{j \le n} \zeta_{nj}$ with

$$\zeta_{nj} = \begin{pmatrix} \int_X f_1(\eta_j + y)\mu_n(dy) - \nu \star \mu_n(f_1) \\ \vdots \\ \int_X f_k(\eta_j + y)\mu_n(dy) - \nu \star \mu_n(f_k) \end{pmatrix} \quad , j = 1, ..., n,\ n \in \mathbb{N}.$$

Let $\mathbb{V}_{n1} = (\mathbb{V}_{n1}(i,l))_{1 \le i,j \le k}$ and $\Sigma_k = (\Sigma_k(i,l))_{1 \le i,j \le k}$ denote the covariance matrix of ζ_{nj} and $(\mathbb{G}_\nu(f_1), ..., \mathbb{G}_\nu(f_k))^t$, respectively, where

$$\Sigma_k(i,l) = \nu(f_i \cdot f_l) - \nu(f_i) \cdot \nu(f_l) \quad , 1 \le i, l \le k.$$

According to the triangle inequality

$$\begin{aligned} &d_{BL}\big((\tilde{\mathbb{G}}_n^\nu(f_1), ..., \tilde{\mathbb{G}}_n^\nu(f_k))^t, (\mathbb{G}_\nu(f_1), ..., \mathbb{G}_\nu(f_k))^t\big) \\ &\le d_{BL}\big((\tilde{\mathbb{G}}_n^\nu(f_1), ..., \tilde{\mathbb{G}}_n^\nu(f_k))^t, \mathcal{N}_k(0, \mathbb{V}_{n1})\big) \ + \ d_{BL}\big(\mathcal{N}_k(0, \mathbb{V}_{n1}), \mathcal{N}_k(0, \Sigma_k)\big). \end{aligned}$$

Note that $\mathcal{L}\{(\mathbb{G}_\nu(f_1), ..., \mathbb{G}_\nu(f_k))^t\} = \mathcal{N}_k(0, \Sigma_k)$.
Now the components $\int_X f_i(\eta_j + y)\mu_n(dy), i = 1, ..., k$, of ζ_{nj} are rv's which are bounded by 1 (since $\mathcal{F}$ is assumed to be uniformly bounded with envelope $F \equiv 1$ w.l.o.g.) for any $f_i \in \mathcal{F}$, so this bound does not depend on ν. An application of Lemma 2.1 in [Gi91] now gives

$$\lim_{n \to \infty} \sup_{\nu \in \mathcal{M}^1(X)} d_{BL}\big((\tilde{\mathbb{G}}_n^\nu(f_1), ..., \tilde{\mathbb{G}}_n^\nu(f_k))^t, \mathcal{N}_k(0, \mathbb{V}_{n1})\big) = 0.$$

From Lemma 2.2 in [Gi91] we have

$$d_{BL}\big(\mathcal{N}_k(0, \mathbb{V}_{n1}), \mathcal{N}_k(0, \Sigma_k)\big) \le C \cdot \sup_{1 \le i,l \le k} |\mathbb{V}_{n1}(i,l) - \Sigma_k(i,l)|$$

with a constant C depending only on k.
Keeping in mind that the $f_1, ..., f_k$ (and therefore also $\mathbb{V}_{n1}$ and Σ_k) depend on ν we are going to show that

(7.5.12) $$\sup_{\nu \in \mathcal{M}^1(X)} \sup_{1 \le i,l \le k} |\mathbb{V}_{n1}(i,l) - \Sigma_k(i,l)| \longrightarrow 0,$$

whence $\limsup_{n\to\infty} \sup_{\nu\in\mathcal{M}^1(X)} d_{BL}\big(\mathcal{N}_k(0,\mathbb{V}_{n1}), \mathcal{N}_k(0,\Sigma_k)\big) = 0$, which completes the proof of (7.5.11) for $i = 2$.

As to (7.5.12):

$$\mathbb{V}_{n1}(i,l) = \int_X \big[\int_X f_i(x+y)\mu_n(dy) - \nu\star\mu_n(f_i)\big]\cdot\big[\int_X f_l(x+y)\mu_n(dy) - \nu\star\mu_n(f_l)\big]\nu(dx)$$

which is equal to (inserting $f_i(x)$ and $f_l(x)$, respectively)

$$\begin{aligned}
&\int_X \big[\int_X f_i(x+y)\mu_n(dy) - f_i(x)\big]\cdot\big[\int_X f_l(x+y)\mu_n(dy) - f_l(x)\big]\nu(dx) \\
&+ \int_X \big[f_i(x) - \nu\star\mu_n(f_i)\big]\cdot\big[\int_X f_l(x+y)\mu_n(dy) - f_l(x)\big]\nu(dx) \\
&+ \int_X \big[f_l(x) - \nu\star\mu_n(f_l)\big]\cdot\big[\int_X f_i(x+y)\mu_n(dy) - f_i(x)\big]\nu(dx) \\
&+ \int_X \big[f_l(x) - \nu\star\mu_n(f_l)\big]\cdot\big[f_i(x) - \nu\star\mu_n(f_i)\big]\nu(dx) \\
&=: I_{n1}(f_i,f_l) + I_{n2}(f_i,f_l) + I_{n3}(f_i,f_l) + I_{n4}(f_i,f_l).
\end{aligned}$$

The Cauchy-Schwarz inequality together with (7.5.8) yields

$$\sup_{\nu\in\mathcal{M}^1(X)} \sup_{1\le i,l\le k} I_{n1}(f_i,f_l) \longrightarrow 0.$$

Next,

$$\begin{aligned}
|I_{n4}(f_i,f_l) - \Sigma_k(i,l)| &= |I_{n4}(f_i,f_l) - (\nu(f_i\cdot f_l) - \nu(f_i)\cdot\nu(f_l))| \\
&= |\nu\star\mu_n(f_i) - \nu(f_i)|\cdot|\nu\star\mu_n(f_l) - \nu(f_l)|
\end{aligned}$$

$$\le \sup_{f\in\mathcal{F}} \Big|\int_X\int_X (f(x+y) - f(x))\mu_n(dy)\nu(dx)\Big|^2 \le \sup_{f\in\mathcal{F}} \int_X \big[\int_X (f(x+y) - f(x))\mu_n(dy)\big]^2 \nu(dx),$$

whence by (7.5.8)

$$\sup_{\nu\in\mathcal{M}^1(X)} \sup_{1\le i,l\le k} |I_{n4}(f_i,f_l) - \Sigma_k(i,l)| \longrightarrow 0.$$

From this, the Cauchy-Schwarz Inequality and (7.5.8) aggain, we also have

$$\sup_{\nu\in\mathcal{M}^1(X)} \sup_{1\le i,l\le k} I_{nj}(f_i,f_l) \longrightarrow 0 \quad \text{for } j = 2,3.$$

This proves (7.5.12).

As to I_{1n} $\Big(:= \big|\mathbb{E}^*(H(\tilde{\mathbb{G}}_n^\nu)) - \mathbb{E}(H(\tilde{\mathbb{G}}_n^\nu(\pi_\tau)))\big|\Big)$:
Since $H \in BL_1(l^\infty(\mathcal{F}))$ it suffices to show that

$$\lim_{\tau\to 0} \limsup_{n\to\infty} \sup_{\nu\in\mathcal{M}^1(X)} \mathbb{P}\big(\sup\{|\tilde{\mathbb{G}}_n^\nu(f_1) - \tilde{\mathbb{G}}_n^\nu(f_2)| : f_1, f_2 \in \mathcal{F}, d_\nu^{(2)}(f_1,f_2) \le \tau\} > \delta\big) = 0$$

for all $\delta > 0$.
Since for each $f \in \mathcal{F}$

$$\tilde{\mathbb{G}}_n^\nu(f) = \mathbb{G}_n^\nu(g_{f,\mu_n}) \quad (\text{with } g_{f,\mu_n}(x) := \int_X f(x+y)\mu_n(dy) \quad , x \in X)$$

we have

$$\begin{aligned} &\sup\{|\tilde{\mathbb{G}}_n^\nu(f_1) - \tilde{\mathbb{G}}_n^\nu(f_2)| : f_1, f_2 \in \mathcal{F}, d_\nu^{(2)}(f_1, f_2) \le \tau\} \\ &= \sup\{|\mathbb{G}_n^\nu(g_{f_1,\mu_n}) - \mathbb{G}_n^\nu(g_{f_2,\mu_n})| : f_1, f_2 \in \mathcal{F}, d_\nu^{(2)}(f_1, f_2) \le \tau\} \\ &\le \sup\{|\mathbb{G}_n^\nu(f_1) - \mathbb{G}_n^\nu(f_2)| : f_1, f_2 \in \mathcal{F}, d_\nu^{(2)}(f_1, f_2) \le \tau\} \;+\; 2 \cdot \sup_{f \in \mathcal{F}} |\mathbb{G}_n^\nu(f) - \mathbb{G}_n^\nu(g_{f,\mu_n})|. \end{aligned} \tag{7.5.13}$$

We show first

(a) $\limsup_{n\to\infty} \sup_{\nu \in \mathcal{M}^1(X)} \mathbb{P}\big(\sup_{f\in\mathcal{F}} |\mathbb{G}_n^\nu(f) - \mathbb{G}_n^\nu(g_{f,\mu_n})| > \delta\big) = 0 \quad \forall \delta > 0.$

Let $\delta > 0$ be arbitrary but fixed. Then by Markov's inequality and the Symmetrization Inequality 5.1.2 we obtain

$$\begin{aligned} &\mathbb{P}\big(\sup_{f\in\mathcal{F}} |\mathbb{G}_n^\nu(f) - \mathbb{G}_n^\nu(g_{f,\mu_n})| > \delta\big) \\ &\le \delta^{-1}\sqrt{n}\;\; \mathbb{E}(\sup_{f\in\mathcal{F}} |\nu_n(f - g_{f,\mu_n}) - \nu(f - g_{f,\mu_n})|) \\ &\le 2\,\delta^{-1}\sqrt{n}\;\; \mathbb{E}(\sup_{f\in\mathcal{F}} |n^{-1} \sum_{j\le n} \varepsilon_j (f - g_{f,\mu_n})(\eta_j)|), \end{aligned}$$

where $\varepsilon_1, \varepsilon_2, ...$ is a canonically formed Rademacher sequence which is independent of $(\eta_j)_{j\in\mathbb{N}}$.

Now, by Lemma 2.9.1 in [Va96] we can replace the ε_j's by a sequence of i i d rv's g_j with $\mathcal{L}\{g_j\} = \mathcal{N}(0,1)$, to obtain the following upper bound (by taking expectations w.r.t. the g_j's (denoted by $\mathbb{E}_g$) and the η_j's (denoted by $\mathbb{E}_\nu$) seperately:

$$\begin{aligned} &\mathbb{P}\big(\sup_{f\in\mathcal{F}} |\mathbb{G}_n^\nu(f) - \mathbb{G}_n^\nu(g_{f,\mu_n})| > \delta\big) \\ &\le C \cdot \mathbb{E}_\nu\; \mathbb{E}_g\big(\sup_{f\in\mathcal{F}} |n^{-1/2} \sum_{j\le n} g_j \cdot (f - g_{f,\mu_n})(\eta_j)|\big) \end{aligned}$$

where the constant C depends on δ but not on n.
Now , for fixed realizations $\eta_1(\omega), ..., \eta_n(\omega)$ consider the process $\mathbb{Z}_{\nu_n}^\omega = (Z_{\nu_n}^\omega(f))_{f\in\mathcal{F}_n}$ with

$$\begin{aligned} Z_{\nu_n}^\omega(f) &:= n^{-1/2} \sum_{j\le n} g_j \cdot f(\eta_j(\omega)) \quad , f \in \mathcal{F}_n, \\ \text{and} \quad \mathcal{F}_n &:= \{f, g_{f,\mu_n} : f \in \mathcal{F}\}. \end{aligned}$$

Then $\mathbb{Z}_{\nu_n}^\omega$ is a mean-zero Gaussian process with

$$cov(Z_{\nu_n}^\omega(f_1), Z_{\nu_n}^\omega(f_2)) = \nu_n(f_1 \cdot f_2, \omega) \quad \text{for } f_1, f_2 \in \mathcal{F}_n,$$

where $\nu_n(f_1 \cdot f_2, \omega) := n^{-1} \sum_{j \le n} \delta_{\eta_j(\omega)}(f_1 \cdot f_2)$.

Considering instead the process (indexed also by $\mathcal{F}_n$)

$$\mathbb{G}^\omega_{\nu_n} + g \cdot \nu_n(\,\cdot\,, \omega) = (G^\omega_{\nu_n}(f) + g \cdot \nu_n(f, \omega))_{f \in \mathcal{F}_n},$$

where $\mathbb{G}^\omega_{\nu_n} = (G^\omega_{\nu_n}(f))_{f \in \mathcal{F}_n}$ is a mean-zero Gaussian process with

$$cov(G^\omega_{\nu_n}(f_1), G^\omega_{\nu_n}(f_2)) = \nu_n(f_1 \cdot f_2, \omega) - \nu_n(f_1, \omega) \cdot \nu_n(f_2, \omega)$$

for $f_1, f_2 \in \mathcal{F}_n$, and where g with $\mathcal{L}\{g\} = \mathcal{N}(0,1)$ is independent of $\mathbb{G}^\omega_{\nu_n}$, we have (as can be seen by computing covariances) that $\forall \omega$

$$(\star) \qquad \mathbb{Z}^\omega_{\nu_n} \underset{\text{fidi}}{\overset{\mathcal{L}}{=}} \mathbb{G}^\omega_{\nu_n} + g \cdot \nu_n(\,\cdot\,, \omega).$$

Thus $\forall \omega$

$$\begin{aligned}
&\mathbb{E}_g\big(\sup_{f \in \mathcal{F}} |n^{-1/2} \sum_{j \le n} g_j \cdot (f - g_{f,\mu_n})(\eta_j(\omega))|\big) \\
&\underset{(\star)}{\le} \mathbb{E}\big(\sup_{f \in \mathcal{F}} |G^\omega_{\nu_n}(f) - G^\omega_{\nu_n}(g_{f,\mu_n})|\big) + \mathbb{E}(|g|) \cdot \sup_{f \in \mathcal{F}} |\nu_n(f - g_{f,\mu_n}, \omega)| \\
&\le \sup_{\nu \in \mathcal{M}^1(X)} \mathbb{E}(\sup_{f \in \mathcal{F}} |G_\nu(f) - G_\nu(g_{f,\mu_n})|) + \mathbb{E}(|g|) \cdot \sup_{\nu \in \mathcal{M}^1(X)} \sup_{f \in \mathcal{F}} |\nu(f - g_{f,\mu_n})|.
\end{aligned}$$

Now let $\varepsilon > 0$ be arbitrary and (using (7.5.7)) choose $\delta > 0$ s.t.

$$\sup_{\nu \in \mathcal{M}^1(X)} \mathbb{E}\big(\sup\big\{|G_\nu(h_1) - G_\nu(h_2)| : h_1, h_2 \in \mathcal{F} \cup \mathcal{G}, d^{(2)}_\nu(h_1, h_2) \le \delta\big\}\big) \le \varepsilon.$$

Then by (7.5.8) for large enough n we have for all $f \in \mathcal{F}$

$$\begin{aligned}
&\sup_{\nu \in \mathcal{M}^1(X)} (d^{(2)}_\nu(f, g_{f,\mu_n}))^2 \\
&= \sup_{\nu \in \mathcal{M}^1(X)} \nu((f - g_{f,\mu_n})^2) \;=\; \sup_{\nu \in \mathcal{M}^1(X)} \int_X \big[\int_X (f(x+y) - f(x)) \mu_n(dy)\big]^2 \nu(dx) \le \delta^2,
\end{aligned}$$

and so

$$\begin{aligned}
&\limsup_{n \to \infty} \sup_{\nu \in \mathcal{M}^1(X)} \mathbb{E}_\nu\, \mathbb{E}_g\big(\sup_{f \in \mathcal{F}} |n^{-1/2} \sum_{j \le n} g_j \cdot (f - g_{f,\mu_n})(\eta_j)|\big) \\
&\le \limsup_{n \to \infty} \sup_{\nu \in \mathcal{M}^1(X)} \mathbb{E}(\sup_{f \in \mathcal{F}} |G_\nu(f) - G_\nu(g_{f,\mu_n})|) \;+\; \mathbb{E}(|g|) \limsup_{n \to \infty} \sup_{\nu \in \mathcal{M}^1(X)} \sup_{f \in \mathcal{F}} |\nu(f - g_{f,\mu_n})| \\
&= \limsup_{n \to \infty} \sup_{\nu \in \mathcal{M}^1(X)} \mathbb{E}(\sup\{|G_\nu(f) - G_\nu(g_{f,\mu_n})| : f \in \mathcal{F}, d^{(2)}_\nu(f, g_{f,\mu_n}) \le \delta\}) + 0 \\
&\le \sup_{\nu \in \mathcal{M}^1(X)} \mathbb{E}\big(\sup\big\{|G_\nu(h_1) - G_\nu(h_2)| : h_1, h_2 \in \mathcal{F} \cup \mathcal{G}, d^{(2)}_\nu(h_1, h_2) \le \delta\big\}\big) \;\le\; \varepsilon,
\end{aligned}$$

where we have used that

$$\begin{aligned}
&\limsup_{n\to\infty} \sup_{\nu\in\mathcal{M}^1(X)} \sup_{f\in\mathcal{F}} |\nu(f - g_{f,\mu_n})| \\
\le &\limsup_{n\to\infty} \sup_{\nu\in\mathcal{M}^1(X)} \sup_{f\in\mathcal{F}} \Big[\int_X \big[\int_X (f(x+y) - f(x))\mu_n(dy)\big]^2 \nu(dx)\Big]^{1/2} = 0
\end{aligned}$$

according to (7.5.8). Thus (a) is proved.
To conclude the proof of (7.5.11) for $i = 1$, we still have to show (see (7.5.13)) that

$$\lim_{\tau\to 0} \limsup_{n\to\infty} \sup_{\nu\in\mathcal{M}^1(X)} \mathbb{P}\big(\sup\{|G_n^\nu(f_1) - G_n^\nu(f_2)| : f_1, f_2 \in \mathcal{F}, d_\nu^{(2)}(f_1, f_2) \le \tau\} > \delta\big) = 0$$

for all $\delta > 0$.
This is proved by similar techniques. Since this expression, however, does not involve any smoothing operations we refer to [Gi97] for a proof.
So we have shown (7.5.11) for $i = 1$, too, and the theorem is proved. □

Finally, from [Va96], Theorem 2.8.3 we have that (7.5.6) and (7.5.7) are fulfilled if $\mathcal{F}\cup\mathcal{G}$ has uniformly integrable L_2-entropy, which in turn is implied if $\mathcal{F}\cup\mathcal{G}$ is a VCGC. So, Theorem 7.5.5 yields

7.5.14. THEOREM.
Let X be a linear metric space and let $\mathcal{F}$ be uniformly bounded. Suppose that $\mathcal{F}\cup\mathcal{G}$ has uniformly integrable L_2-entropy and that (7.5.8) is fulfilled. Then

$$\sup_{\nu\in\mathcal{M}^1(X)} d_{BL}(\tilde{\mathbb{G}}_n^\nu, \mathbb{G}_\nu) \longrightarrow 0.$$

7.5.15. REMARK.
As we will see in the next section, the results of Section 7.5 are important in the area of bootstrapping empirical processes (see Section 8.4 and the literature cited there).

8 Bootstrapping

8.1 Introduction

Bootstrapping is a resampling technique where one replaces the original data given as observations $y_1, ..., y_n$ of i i d re's $\eta_1, ..., \eta_n$ in a sample space $X = (X, \mathcal{X})$ with law $\mathcal{L}\{\eta_j\} \equiv \nu$ on the σ-Algebra $\mathcal{X}$ in X by the so-called bootstrap sample $\eta_1^*, ...\eta_n^*$ where the η_j^*'s are also assumed to be i i d re's in X with law $\mathcal{L}\{\eta_j^*\} = Q_n, 1 \leq j \leq n, n \in \mathbb{N}$, and where Q_n may depend on the original data. In fact, the η_j^*'s, $1 \leq j \leq n, n \in \mathbb{N}$, depend on n; if necessary, we indicate this by writing $\eta_{n1}^*, ..., \eta_{nn}^*$.
A typical example of Q_n is the empirical measure ν_n based on the observations of $\eta_1, ..., \eta_n$, i.e. in case of $\mathcal{L}\{\eta_j^*\} = \nu_n$ we get instead of the empirical process $\mathbb{G}_n^\nu := \sqrt{n}(\nu_n - \nu)$ the so-called *bootstrapped empirical process* (of sample size n)

$$\mathbb{G}_n^{\nu_n} = \sqrt{n}\Big(n^{-1} \sum_{j \leq n} \delta_{\eta_j^*} - \nu_n\Big)$$

based on $\eta_1^*, ..., \eta_n^*$, $n \in \mathbb{N}$, to be thought as defined on a basic p-space $(\Omega^*, \mathcal{A}^*, \mathbb{P}^*)$, i.e. depending on the realizations $y_j = \eta_j(\omega)$ of the original re's η_j (defined on a basic p-space $(\Omega, \mathcal{A}, \mathbb{P})$), the processes $\mathbb{G}_n^{\nu_n}, n \in \mathbb{N}$, are given as

$$(*) \qquad \mathbb{G}_n^{\nu_n}(\omega^*) = \sqrt{n}\Big(n^{-1} \sum_{j \leq n} \delta_{\eta_j^*(\omega^*)} - \nu_n(\omega)\Big)$$

where $\nu_n(\omega) = n^{-1} \sum\limits_{j \leq n} \delta_{y_j},\ y_j = \eta_j(\omega), 1 \leq j \leq n$, are to be considered as constants in (*).
This principle proves to be useful when considering the problem of approximating the law of certain statistics $T(\eta_1, ..., \eta_n; \nu)$ with unknown ν by the law of $T(\eta_1^*, ..., \eta_n^*; \nu_n)$, being efficient at least for "smooth" functionals T in view of the fact (cf. Section 6.3) that ν_n approximates ν as $n \longrightarrow \infty$; in addition, in this connection the law of $T(\eta_1^*, ..., \eta_n^*; \nu_n)$ may either be computed directly or can be approximated by Monte Carlo simulation methods.

In Efron's fundamental paper [Ef79] this procedure was given the name *bootstrap.* It works in parametric and in non-parametric settings as we will see in the following sections.

8.2 On the construction of confidence intervals for an unknown real-valued parameter by bootstrapping

We will follow here mainly the presentation in [Fa87]. Given i i d rv's $\eta_j, j \in \mathbb{N}$, with distribution function (df) F, defined on a basic p-space $(\Omega, \mathcal{A}, \mathbb{P})$, assume that an unknown parameter ϑ can be represented as a functional of F, i.e. $\vartheta = T(F)$ with a properly choosen T defined on the space of all df's.
For emample, $\vartheta = \mathbb{E}(\eta_1)$ is representable as the *mean-value functional*

$$(8.2.1) \qquad \vartheta = T(F) := \int_{\mathbb{R}} x \, dF(x).$$

Then, to obtain an estimate ϑ_n for ϑ, based on $\eta_1, ..., \eta_n$, using the *plug in method*, one simply replaces the unknown F in (8.2.1) by the empirical df (edf) F_n, based on $\eta_1, ..., \eta_n$, i.e.

$$\vartheta_n := T(F_n). \tag{8.2.2}$$

In case of the mean-value functional (8.2.1), which will be considered exclusively in this section,

$$T(F_n) = n^{-1} \sum_{j \leq n} \eta_j.$$

To obtain confidence intervals for ϑ based on ϑ_n one would need the knowledge of the df G_n, where

$$G_n(t) := \mathbb{P}\Big(n^{1/2}\left(T(F_n) - T(F)\right) \leq t\Big), \qquad t \in \mathbb{R}. \tag{8.2.3}$$

In fact, knowing G_n one could proceed as follows:

Given $\alpha \in (0,1)$, choose (a minimal) $d_\alpha = d_\alpha(n)$ and (a maximal) $b_\alpha = b_\alpha(n)$ such that

$$G_n(d_\alpha) \geq 1 - \frac{\alpha}{2} \quad \text{and} \quad G_n(b_\alpha - 0) \leq \frac{\alpha}{2}$$

to obtain with

$$[T(F_n) - \frac{d_\alpha}{\sqrt{n}}\,,\; T(F_n) - \frac{b_\alpha}{\sqrt{n}}]$$

confidence intervals of level $1 - \alpha$. In fact

(8.2.4)

$$\begin{aligned}
\mathbb{P}\Big(T(F) \in [T(F_n) - \frac{d_\alpha}{\sqrt{n}}\,,\; T(F_n) - \frac{b_\alpha}{\sqrt{n}}]\Big) &= \mathbb{P}\Big(T(F) \leq T(F_n) - \frac{b_\alpha}{\sqrt{n}}\Big) - \mathbb{P}\Big(T(F) < T(F_n) - \frac{d_\alpha}{\sqrt{n}}\Big) \\
&= \mathbb{P}\Big(n^{1/2}(T(F_n) - T(F)) \geq b_\alpha\Big) - \mathbb{P}\Big(n^{1/2}(T(F_n) - T(F)) > d_\alpha\Big) \\
&= 1 - G_n(b_\alpha - 0) - (1 - G_n(d_\alpha)) \\
&= G_n(d_\alpha) - G_n(b_\alpha - 0) \geq 1 - \frac{\alpha}{2} - \frac{\alpha}{2} = 1 - \alpha
\end{aligned}$$

But this does not work without knowing G_n, i.e. F. Here bootstrapping comes into play:

Choosing so-called independent *bootstrap-rv's* $\eta_j^*, j \in \mathbb{N}$, defined on another p-space $(\Omega^*, \mathcal{A}^*, \mathbb{P}^*)$, where for each $n \in \mathbb{N}$ the $\eta_j^*, 1 \leq j \leq n$, are i d with df F_n, one replaces in (8.2.3) F_n by the edf F_n^* based on $\eta_1^*, ..., \eta_n^*$, and F by $F_n(\cdot, \omega)$, based on $\eta_1(\omega), ..., \eta_n(\omega)$, to obtain as a *nonparametric estimate* for G_n the function

$$G_n^*(t) := \mathbb{P}^*\Big(n^{1/2}\left(T(F_n^*) - T(F_n(\cdot, \omega))\right) \leq t\Big), \qquad t \in \mathbb{R},$$

i.e. for each fixed $\omega \in \Omega$ G_n^* is defined within the *bootstrap model* by

$$G_n^*(t, \omega) := \mathbb{P}^*\Big(\{\omega^* \in \Omega^* : \; n^{1/2}\left(T(F_n^*(\cdot, \omega^*)) - T(F_n(\cdot, \omega))\right) \leq t\}\Big), \qquad t \in \mathbb{R}, \tag{8.2.5}$$

with $F_n^*(s, \omega^*) := n^{-1} \sum_{j \leq n} 1_{(-\infty, s]}(\eta_j^*(\omega^*)), \quad s \in \mathbb{R}.$

Note that in (8.2.5) the term $T(F_n(\cdot, \omega))$ is to be considered as a constant within the bootstrap model

for each fixed $\omega \in \Omega$.
Now, in view of (8.2.5), for each $n \in \mathbb{N}$ and given $\eta_1(\omega), ..., \eta_n(\omega)$ for fixed $\omega \in \Omega$, one can determine (a minimal) $d^*_\alpha = d^*_\alpha(n, \omega)$ and (a maximal) $b^*_\alpha = b^*_\alpha(n, \omega)$ such that

$$G^*_n(d^*_\alpha(n, \omega), \omega) \geq 1 - \frac{\alpha}{2} \quad \text{and} \quad G^*_n(b^*_\alpha(n, \omega) - 0, \omega) \leq \frac{\alpha}{2}.$$

to obtain

$$\text{(8.2.6)} \quad \mathbb{P}^*\Big(T(F_n(\cdot, \omega) \in [T(F^*_n) - \frac{d^*_\alpha}{\sqrt{n}}\,,\ T(F^*_n) - \frac{b^*_\alpha}{\sqrt{n}}]\Big) = G^*_n(d^*_\alpha(n, \omega), \omega) - G^*_n(b^*_\alpha(n, \omega) - 0, \omega) \geq 1 - \alpha.$$

At this point it is essential to note that in the present situation (see below for its verification):

$$\text{(8.2.7)} \qquad d^*_\alpha(n, \cdot) - d_\alpha(n) \longrightarrow 0 \quad \mathbb{P} - a.s. \quad \text{and} \quad b^*_\alpha(n, \cdot) - b_\alpha(n) \longrightarrow 0 \quad \mathbb{P} - a.s.$$

Therefore the intervals

$$\text{(8.2.8)} \qquad [T(F_n) - \frac{d^*_\alpha}{\sqrt{n}}\,,\ T(F_n) - \frac{b^*_\alpha}{\sqrt{n}}], \quad n \in \mathbb{N},$$

constitute a sequence of $\mathbb{P} - a.s.$ consistent estimators for the intervals

$$[T(F_n) - \frac{d_\alpha}{\sqrt{n}}\,,\ T(F_n) - \frac{b_\alpha}{\sqrt{n}}], \quad n \in \mathbb{N},$$

yielding thus in view of (8.2.4) a sequence of $\mathbb{P} - a.s.$ confidence intervals of asymptotic level $1 - \alpha$ for the unknown parameter $\vartheta = T(F)$.

To verify (8.2.7) in the case of $0 < \sigma^2 := V(\eta_1) < \infty$ one uses the following result in [Si81]:

8.2.9. THEOREM.
Assume that $0 < \sigma^2 := V(\eta_1) < \infty$*; then*

$$\text{(8.2.10)} \qquad \sup_{t \in \mathbb{R}} |G^*_n(t, \omega) - G_n(t)| \longrightarrow 0 \quad \textit{for } \mathbb{P} - a.s.\ \omega \in \Omega.$$

On the other hand, according to the central limit theorem, one has that

$$G_n(t) \longrightarrow G_0(t) := \Phi(\frac{t}{\sigma}) \qquad \forall\, t \in \mathbb{R}$$

(Φ being the standard normal df).
Therefore, by (8.2.10), for $\mathbb{P} - a.a.\ \omega \in \Omega$ we get

$$\text{(8.2.11)} \qquad G^*_n(t, \omega) \longrightarrow G_0(t) \qquad \forall\, t \in \mathbb{R}.$$

Since G_0 is continuous and strictly monotone increasing, it follows from (8.2.11) (cf. [Wit70], Satz 2.11, S. 53) that, with $d_\alpha(n)$ and $b_\alpha(n)$ denoting the $(1 - \frac{\alpha}{2})$ - quantile and $\frac{\alpha}{2}$ - quantile of G_n (cf. (8.2.4)) and with $d^*_\alpha(n, \omega)$ and $b^*_\alpha(n, \omega)$ denoting the $(1 - \frac{\alpha}{2})$ - quantile and $\frac{\alpha}{2}$ - quantile of $G^*_n(\cdot, \omega)$ (cf. (8.2.6)), the assertion (8.2.7) is verified.

IMPORTANT NOTE: For the proof of (8.2.7) continuity and strict monotone increasing of the limiting df G_0 was essential.

8.3 Bootstrapping empirical processes

Let us reconsider at first the uniform empirical process α_n of Section 1.1 based on i i d rv's η_j defined on $(\Omega, \mathcal{A}, \mathbb{P})$ with $\mathcal{L}\{\eta_j\} = U[0,1]$, where we have seen (cf. (1.1.6)) that α_n converges in law to the Brownian Bridge B^0 as $n \longrightarrow \infty$. According to Theorem 2.3.9 this implies (cf. (2.3.12))

$$\lim_{\delta \to 0} \limsup_{n \to \infty} \mathbb{P}(w_{\alpha_n}(\delta) \geq \varepsilon) = 0 \qquad \forall\, \varepsilon > 0. \tag{8.3.1}$$

Now, it is well known that for an arbitrary rv ξ with df F the law $\mathcal{L}\{\xi\}$ of ξ can be represented as $\mathcal{L}\{F^{-1}(\eta)\}$ where $\mathcal{L}\{\eta\} = U[0,1]$ and

$$F^{-1}(s) := \inf\{t \in \mathbb{R} \ : \ F(t) \geq s\}, \qquad 0 \leq s \leq 1.$$

Thus, for a sequence of i i d rv's $\xi_j, j \in \mathbb{N}$, with df F one gets

$$\mathcal{L}\Big\{\Big(n^{1/2}\left(T(F_n) - T(F)\right)\Big)_{t \in \mathbb{R}}\Big\} = \mathcal{L}\Big\{\big(\alpha_n(F(t))\big)_{t \in \mathbb{R}}\Big\}, \tag{8.3.2}$$

where F_n is the edf based on $\xi_1, ..., \xi_n$.

Considering the bootstrap-procedure, given the data in form of $\eta_1(\omega), ..., \eta_n(\omega)$ (with the η_j's as before), let $\xi_j^*(= \xi_{nj}^*)$, $1 \leq j \leq n$, be i i d bootstrap rv's with df

$$G_n(\cdot, \omega) := n^{-1} \sum_{j \leq n} 1_{[0,\cdot]}(\eta_j(\omega)),$$

let F_n^* be the edf based on $\xi_1^*, ..., \xi_n^*$ and let

$$\alpha_n^*(\omega) = \Big(\alpha_n^*(s, \omega)\Big)_{s \in [0,1]}$$

be the *bootstrapped uniform empirical process.* i.e.

$$\alpha_n^*(s, \omega) = n^{1/2}\Big(F_n^*(s) - G_n(s, \omega)\Big), \qquad 0 \leq s \leq 1;$$

then

$$\alpha_n^*(\omega) \xrightarrow{\mathcal{L}} B^0 \qquad \text{for } \mathbb{P} - a.s.\ \omega. \tag{8.3.3}$$

To prove (8.3.3) we may and do assume (cf. (8.3.2)) that

$$\alpha_n^*(\omega) = \Big(\alpha_n(G_n(s, \omega))\Big)_{s \in [0,1]}.$$

So, we must show that for $\mathbb{P} - a.s.\ \omega$

$$\Big(\alpha_n(G_n(s, \omega))\Big)_{s \in [0,1]} \xrightarrow{\mathcal{L}} B^0. \tag{8.3.4}$$

For this, it suffices to show (in view of the fact $\alpha_n \xrightarrow{\mathcal{L}} B^0$) that for $\mathbb{P}-a.s.\ \omega$

$$(+) \qquad \mathbb{P}\Big(\sup_{s\in[0,1]} |\alpha_n(G_n(s,\omega)) - \alpha_n(s)| \geq \varepsilon\Big) \longrightarrow 0 \text{ as } n \to \infty \quad \forall\, \varepsilon > 0$$

(cf. [Gae77] 8.6.2).
Now, given an arbitrary $\varepsilon > 0$ and $||G_n(\cdot,\omega) - id_{[0,1]}|| := \sup_{s\in[0,1]} |G_n(s,\omega) - s|$, we have

$$\mathbb{P}\Big(\sup_{s\in[0,1]} |\alpha_n(G_n(s,\omega)) - \alpha_n(s)| \geq \varepsilon\Big) \leq \mathbb{P}\Big(w_{\alpha_n}\big(||G_n(\cdot,\omega) - id_{[0,1]}||\big) \geq \varepsilon\Big),$$

where again w_{α_n} denotes the oscillation-modulus of α_n.
But, by the classical Glivenko-Cantelli theorem, for $\mathbb{P}-a.s.\ \omega$

$$||G_n(\cdot,\omega) - id_{[0,1]}|| \longrightarrow 0 \qquad \text{as } n \to \infty,$$

whence, according to (8.3.1), for $\mathbb{P}-a.s.\ \omega$

$$\mathbb{P}\Big(w_{\alpha_n}\big(||G_n(\cdot,\omega) - id_{[0,1]}||\big) \geq \varepsilon\Big) \longrightarrow 0 \qquad \text{as } n \to \infty,$$

which proves (+).
For a different proof see [Bi81].
In what follows, let now $X = (X, \mathcal{X})$ be an arbitrary measurable space (sample space) and $\eta_j, j \in \mathbb{N}$, be i i d re's in X with law $\mathcal{L}\{\eta_j\} \equiv \nu$ on X, defined as coordinate projections on $(\Omega, \mathcal{A}, \mathbb{P}) := (X^{\mathbb{N}}, \mathcal{X}^{\mathbb{N}}, \nu^{\mathbb{N}})$. Let ν_n be the empirical measure based on $\eta_1, ..., \eta_n$, and, given the data in form of $y_j = \eta_j(\omega), 1 \leq j \leq n, n \in \mathbb{N}$, let $\eta_j^*(= \eta_{nj}^*), 1 \leq j \leq n, n \in \mathbb{N}$, be i i d *bootstrap re's* in X, defined on another p-space $(\Omega^*, \mathcal{A}^*, \mathbb{P}^*)$, where for each $n \in \mathbb{N}$ the $\eta_j^*, 1 \leq j \leq n$, are i i d with law

$$\mathcal{L}\{\eta_j^*\}(B) = \nu_n(B,\omega) := n^{-1} \sum_{j\leq n} 1_B(\eta_j(\omega)), \qquad B \in \mathcal{X}.$$

Then, the *bootstrapped empirical process* $\mathbb{G}_n^{\nu_n}$ is (for fixed ω) defined by

$$(8.3.5) \qquad \mathbb{G}_n^{\nu_n}(\omega) := n^{1/2}\Big(\nu_n^* - \nu_n(\cdot,\omega)\Big),$$

where ν_n^* denotes the empirical measure based on $\eta_1^*, ..., \eta_n^*$. Now, given e.g. a VCGC $\mathcal{F}$ with $\nu(F^2) < \infty$, we know from Theorem 7.3.5 that for the empirical process $\mathbb{G}_n^\nu = \big(G_n^\nu(f)\big)_{f\in\mathcal{F}}$ indexed by $\mathcal{F}$, where $G_n^\nu(f) := n^{1/2}\Big(\nu_n(f) - \nu(f)\Big), f \in \mathcal{F}$,

$$(8.3.6) \qquad \mathbb{G}_n^\nu \xrightarrow[\text{sep}]{\mathcal{L}} \mathbb{G}_\nu \qquad \text{in } l^\infty(\mathcal{F}),$$

where $\mathbb{G}_\nu = \big(G_\nu(f)\big)_{f\in\mathcal{F}}$ is a mean-zero Gaussian process with sample paths in $U^b(\mathcal{F}, d_\nu^{(2)})$ and $cov(G_\nu(f), G_\nu(g)) = \nu(f \cdot g) - \nu(f) \cdot \nu(g),\ f, g \in \mathcal{F}$. On the other side, given the data $y_j = \eta_j(\omega)$

(for any fixed ω) one may ask in view of (8.3.3) whether in the present general situation an analogous result holds true for the bootstrapped empirical process $\mathbb{G}_n^{\nu_n}(\omega) = \big(G_n^{\nu_n}(f,\omega)\big)_{f\in\mathcal{F}}$, where (cf. (8.3.5)) $G_n^{\nu_n}(f,\omega) := n^{1/2}\Big(\nu_n^*(f) - \nu_n(f,\omega)\Big)$ and $\nu_n(f,\omega) := n^{-1}\sum\limits_{j\le n} f(\eta_j(\omega)),\;\; f\in\mathcal{F}$. The answer is contained in the following theorem proved by Giné and Zinn in 1990 (cf. also [Gi96], Section 4):

8.3.7. THEOREM *([Gi90]).*
Under the usual measurability assumptions the following two statements are equivalent:

(a) $\quad \mathbb{G}_n^{\nu} \xrightarrow[sep]{\mathcal{L}} \mathbb{G}_\nu \qquad$ *in* $l^\infty(\mathcal{F})$ *and* $\nu(F^2) < \infty$

(b) $\quad \mathbb{G}_n^{\nu_n}(\omega) \xrightarrow[sep]{\mathcal{L}} \mathbb{G}_\nu \qquad$ *in* $l^\infty(\mathcal{F})$ *for* $\mathbb{P} - a.a.\ \omega \in \Omega.$

(As before, F denotes the envelope of $\mathcal{F}$.)

When specializing to $\mathcal{F} = \{1_C \,:\, C \in \mathcal{C}\}$ with $\mathcal{C} \subset \mathcal{X}$ being a VCC, the validity of (b) in Theorem 8.3.7 was shown in [Gae86] and used in [Gae90b] to construct confidence bands for probability distributions on VCC $\mathcal{C}$ of sets in arbitrary sample spaces $X = (X, \mathcal{X})$, yielding as result that a confidence band of asymptotic level $1-\alpha$ for $\nu = (\nu(C))_{C\in\mathcal{C}}$ is given for $\mathbb{P} - a.a.\ \underline{y} := (y_j = \eta_j(\omega))_{j\in\mathbb{N}}$ by

$$\underline{y} \longmapsto \Big\{\nu_n(C,\underline{y}) \pm n^{-1/2} c_\alpha^*(n,\underline{y}), C \in \mathcal{C}\Big\},$$

where $c_\alpha^*(n,\underline{y}) := \inf\{t \in \mathbb{R}_+ \,:\, H_n^*(t,\underline{y}) \ge 1-\alpha\}$ with

$$H_n^*(t,\underline{y}) := \mathbb{P}^*\Big(\sup_{C\in\mathcal{C}} |G_n^{\nu_n}(1_C,\underline{y})| \le t\Big), \qquad t \in \mathbb{R}_+,$$

provided that (cf. the Important Note concerning the parametric case at the end of Section 8.2) the following two conditions (C_1) and (C_2) are fulfilled for

$$H_0(t) := \mathbb{P}\Big(\sup_{C\in\mathcal{C}} |G_\nu(1_C)| \le t\Big), \qquad t \in \mathbb{R}_+ :$$

$(C_1)\quad$ H_0 is continuous

$(C_2)\quad$ H_0 is strictly monotone increasing.

In his Diploma-Thesis Molnár [Mo02] proved rigorously what was expected, namely that for countable VCC's $\mathcal{C} \subset \mathcal{X}$ both conditions (C_1) and (C_2) are fulfilled. We owe thanks to Dick Dudley for his guidance in finding the proof presented by Péter Molnár.

8.4 Smoothed empirical processes and the bootstrap

In this last section we want to present a short sketch on bootstrapping smoothed empirical processes. For details we refer to our paper [Gae03].

The context is the same as in Section 7.4: So, throughout X is now supposed to be an arbitrary *linear metric space* endowed with its Borel σ-field $\mathcal{X}$. Let $\eta_j, j \in \mathbb{N}$, be i i d re's in $X = (X, \mathcal{X})$ with law $\mathcal{L}\{\eta_j\} \equiv \nu$ on X, defined as coordinate projections on the p-space

$$(\Omega, \mathcal{A}, \mathbb{P}) := (X^{\mathbb{N}}, \mathcal{X}^{\mathbb{N}}, \nu^{\mathbb{N}}),$$

and let $\nu_n := n^{-1} \sum\limits_{j \le n} \delta_{\eta_j}$ be the empirical measure based on $\eta_1, ..., \eta_n, n \in \mathbb{N}$. Consider the *smoothed empirical measure* $\tilde{\nu}_n := \nu_n \star \mu_n$ (with given non-random $\mu_n \in \mathcal{M}^1(X)$), $n \in \mathbb{N}$, and the *bootstrapped empirical process* (already mentioned in Section 8.1)

$$\mathbb{G}_n^{\nu_n} := \sqrt{n}\Big(n^{-1} \sum_{j \le n} \delta_{\eta_j^*} - \nu_n\Big)$$

based on $\eta_1^*, ..., \eta_n^*, n \in \mathbb{N}$, defined on a basic p-space $(\Omega^*, \mathcal{A}^*, \mathbb{P}^*)$, where the η_j^*'s are assumed to be independent re's in X with $\mathcal{L}\{\eta_j^*\} = \nu_n, 1 \le j \le n, n \in \mathbb{N}$. To be precise, given e.g. a class $\mathcal{F}$ of $\mathcal{X}$-measurable functions $f : X \longrightarrow \mathbb{R}$ (with $\mathcal{X}$-measurable envelope $F : X \longrightarrow \mathbb{R}$ s.t. $\nu(F^2) < \infty$) the processs $\mathbb{G}_n^{\nu_n} = \big(G_n^{\nu_n}(f)\big)_{f \in \mathcal{F}}$ indexed by $\mathcal{F}$ is defined for any fixed $y_j = \eta_j(\omega), 1 \le j \le n, \omega \in \Omega$, by

$$\mathbb{G}_n^{\nu_n}(\omega^*, f) := \sqrt{n}\Big(n^{-1} \sum_{j \le n} f(\eta_j^*(\omega^*)) - \nu_n(f, \omega)\Big) \tag{8.4.1}$$

for each $\omega^* \in \Omega^*$ and $f \in \mathcal{F}$, where $\nu_n(f, \omega) := n^{-1} \sum\limits_{j \le n} f(y_j)$, $n \in \mathbb{N}$, are considered as constants in (8.4.1).

Now we are in the position of deriving the following result for the *smoothed bootstrapped empirical process* (of sample size n) given by

$$\sqrt{n}\Big(n^{-1} \sum_{j \le n} (\delta_{\eta_j^*} \star \mu_n) - \nu_n \star \mu_n\Big). \tag{8.4.2}$$

For this, it is important to note by comparing this process with the unbiased smoothed empirical process $\tilde{\mathbb{G}}_n^{\nu}$ of Section 7.5 that the process (8.4.2) arises through bootstrapping $\tilde{\mathbb{G}}_n^{\nu}$:
In fact, given

$$\tilde{\mathbb{G}}_n^{\nu} := \sqrt{n}(\tilde{\nu}_n - \nu \star \mu_n) = \sqrt{n}\Big((n^{-1} \sum_{j \le n} \delta_{\eta_j}) \star \mu_n - \nu \star \mu_n\Big)$$

$$= \sqrt{n}\Big(\sum_{j \le n} (\delta_{\eta_j} \star \mu_n) - \nu \star \mu_n\Big)$$

based on i i d re's η_j in X with $\mathcal{L}\{\eta_j\} = \nu$ and then replacing ν by ν_n and η_j by η_j^* with $\mathcal{L}\{\eta_j^*\} = \nu_n$ we get the process (8.4.2) which will be denoted by $\tilde{\mathbb{G}}_n^{\nu_n}$ (and also called *bootstrapped smoothed empirical process* (of sample size n).

To this process one can apply Theorem 7.5.14 to obtain

8.4.3. THEOREM *(cf. [Gae03]), Theorem 2.4).*
Let the conditions of Theorem 7.5.14 be satisfied. Then

$$d_{BL}(\tilde{\mathbb{G}}_n^{\nu_n}, \mathbb{G}_\nu)^* \xrightarrow{\mathbb{P}} 0.$$

List of Symbols

$i\,i\,d$	independent identically distributed	1
re	random element	1
$\mathcal{L}\{\xi\}$	law of ξ	1
p-space	probability space	1
$\mathcal{X}^{\mathbb{N}} \equiv \bigotimes_{\mathbb{N}} \mathcal{X}$	product σ-field	1
$\nu^{\mathbb{N}} \equiv \times_{\mathbb{N}} \nu$	product measure	1
ν_n	empirical measure	1
δ_x	Dirac measure	1
edf	empirical distribution function	2
rv	random variable	2
α_n	uniform empirical process	3
$U[0,1]$	uniform distribution on $[0,1]$	3
$\xrightarrow{\mathbb{P}}$	convergence in probability	4
FCLT	functional central limit theorem	5
$\mathbb{P}^*$	outer probability	5
w.r.t.	with respect to	6
CLT	central limit theorem	6
$\square$	end of proof	7
$\log N^{[\,]}(\cdot, \mathcal{F}, d_\nu^{(2)})$	metric entropy with bracketing	9
$\mathcal{P}(X)$	class of all subsets of X (power set of X)	11
$I = [0,1]$	unit intervall	14
1_A	indicator function	14
$\xrightarrow{\mathcal{L}}$	convergence in law	14
a.s.	almost sure	14
$\mathbb{Q}$	rational numbers	15
$D = D(I)$	function space	15
$C = C(I)$	function space	15
ρ	sup-metric	15
s	Skorokhod's metric	15
$\mathcal{B}(\rho)$	Borel σ-field (in (D,ρ))	15
$\mathcal{B}(s)$	Borel σ-field (in (D,s))	15
$\mathcal{B}_b(\rho)$	σ-field (in (D,ρ)) generated by the open ρ-balls	15
π_t	projection	15
$\mathcal{B}(C,\rho)$	Borel σ-field (in (C,ρ))	15
$\mathcal{B}_b(C,\rho)$	σ-field (in (C,ρ)) generated by the open ρ-balls	15
$\sigma(\{\pi_t : t \in I\})$	σ-field generated by the projections	15
$B = (B(t))_{t\in[0,1]}$	Brownian Motion	16
$B^\circ = (B^\circ(t))_{t\in[0,1]}$	Brownian Bridge	16
$t_1 \wedge t_2$	minimum of t_1 and t_2	16
$C^b(D)$	class of all bounded, s-continuous, real functions on D	16
$C_b^b(D)$	class of all bounded, ρ-continuous, $\mathcal{B}_b(\rho)$-measurable, real functions on D	16

References

[Al84] Alexander, K.S. (1984). Probability inequalities for empirical processes and a law of the iterated logarithm for empirical processes. Ann Probab. **12**, 1041 - 1067.

[Al85] Alexander, K.S. (1985). Recent Results on the Central Limit Theorem for Empirical Processes, with Applications. Proceedings of the 45$^{\text{th}}$ Session, Amsterdam ISI meeting, Vol. LI, Book 4, 25.3 1-14.

[Al86] Alexander, K.S. and Pyke, R. (1986). A uniform central limit theorem for set-indexed partial-sum processes with finite variance. Ann. Probab. **14**, 582 - 597.

[Al87] Alexander, K.S. (1987). Central limit theorems for stochastic processes under random entropy conditions. Probab. Theory Relat. Fields **75**, 351 - 378.

[An87] Andersen, N.T. and Dobrič, V. (1987). The central limit theorem for stochastic processes. Ann. Probab. **15**, 164 - 177.

[Ar92] Arcones, M., Gaenssler, P. and Ziegler, K. (1992). Partial-Sum processes with random locations and indexed by Vapnik-Chervonenkis classes of sets in arbitrary sample spaces. In: Probability in Banach Spaces **8**, 379 - 389, R.M. Dudley, Marjorie G. Hahn and J. Kuelbs (Eds.), Birkhäuser.

[Ba85] Bass, R.F. and Pyke, R. (1985). The space $D(\mathcal{A})$ and weak convergence for set-indexed processes. Ann. Probab. **13**, 860 - 884.

[Ben62] Bennet, G. (1962). Probability Inequalities for the sum of independent random variables. J. Amer. Statist. Assoc. **57**, 33 - 45.

[Ber24] Bernstein, S. (1924). Sur une modification de l'inégalité de Tschebichef. Annals Science Institute Sov. Ukraine Sect. Math. I (Russian, French Summary).

[Bi71] Bickel, P.J. and Wichura, M.J. (1971). Convergence criteria for multiparameter stochastic processes and some applications. Ann. Math. Statist. **42**, 1656 - 1670.

[Bi81] Bickel, P.J. and Freedman, D.A. (1981). Some asymptotic theory for the bootstrap. Ann. Statist. **9**, 1196 - 1217.

[Bi68] Billingsley P. (1968). Convergence of Probability Measures. Wiley.

[Cam83] Le Cam, L. (1983). A remark on empirical measures. In: Festschrift for E.L. Lehmann, 305 - 327, P. Bickel, K. Doksum and J. Hodges (Eds.), Belmont, CA: Wadsworth.

[Ca33] Cantelli, F.P. (1933). Sulla determinazione empirica leggi di probabilitá. Giorn. Ist. Ital Attuari **4**, 421 - 424.

[Che56] Chentsov, N.N. (1956). Lévy Brownian motion for several parameters and white noise. Theor. Prob. Appl. **2**, 265 - 266.

[De71] Dehardt, J. (1971). Generalizations of the Glivenko-Cantelli theorem. Ann. Math. Statist. **42**, 2050 - 2055.

[Dev85] Devroye, L. and Gyorfi, L. (1985). Nonparametric density estimation. The L_1 view. Wiley.

[Do49] Doob, J.L. (1949). Heuristic approach to the Kolmogorov-Smirnov theorems. Ann. Math. Statist. **20**, 393 - 403.

[Don51] Donsker, M.D. (1951). An invariance principle for certain probability limit theorems. Mem. Amer. Math. Soc. **6**, 1 - 12.

[Don52] Donsker, M.D. (1952). Justification and extension of Doob's heuristic approach to the Kolmogorov-Smirnov theorems. Ann. Math. Statist. **23**, 277 - 281.

[Du66] Dudley, R.M. (1966). Weak convergence of probabilities on nonseparable metric spaces and empirical measures on Euclidian spaces. Ill. J. Math. **10**, 109 - 126.

[Du78] Dudley, R.M. (1978). Central limit theorems for empirical measures. Ann. Probab. **6**, 899 - 929; Correction: Ann. Probab. **7**, 909 - 911.

[Du83] Dudley, R.M. and Philipp, W. (1983). Invariance principles for sums of Banach space valued random elements and empirical processes. Z. Wahrsch. verw. Gebiete **62**, 509 - 552.

[Du84] Dudley, R.M. (1984). A course on empirical processes. Springer Lecture Notes in Math. **1097**, 1 - 142.

[Du92] Dudley, R.M. (1992). Uniform Central Limit Theorems. Manuscript, MIT.

[Du99] Dudley, R.M. (1999). Uniform Central Limit Theorems. Cambridge Studies in advanced mathematics **63**. Cambridge University Press, Cambridge.

[Du99a] Dudley, R.M. (1999). Notes on Empirical Processes. MaPhySto - Lecture Notes no. 4.

[Due00] Duembgen, L. (2000). Empirische Prozesse. Skripten zur Mathematischen Statistik Nr. 35, Gesellschaft zur Förderung der Mathematischen Statistik, Münster.

[Dur80] Durst, M. and Dudley, R.M. (1980). Empirical Processes, Vapnik-Chervonenkis classes and Poisson Processes. Prob. and Math. Statist. (Wroclaw) **1**, Fasc. 2, 109 - 115.

[Ef79] Efron B. (1979). Bootstrap methods: Another look at the Jacknife. Ann. Statist. **7**, 1-26.

[Ef82] Efron B. (1982). The Jacknife, the Bootstrap, and other Resampling Plans. CBMS-NSF Regional Conference Series in Applied Mathematics, **38**, Society for Industrial and Applied Mathematics, Philadelphia.

[Ei97] Einmahl, U. and Mason, D. (1997). Gaussian approximation of local empirical processes indexed by functions. Probab. Theory Relat. Fields **107**, 283 - 311.

[Ei98] Einmahl, U. and Mason, D. (1998). An Empirical Process Approach to the Uniform Consistency of Kernel-Type Function Estimators. Preprint.

[El79] Elker, J., Pollard, D. and Stute, W. (1979). Glivenko-Cantelli theorems for classes of convex sets. Adv. Appl. Probab. **11**, 820 - 833.

[Fa87] Falk, M. (1987). Bootstrap: Eine aktuelle statistische Idee. Mitteilungen der Mathematischen Gesellschaft in Hamburg, Band XI, Heft **4**, 475 - 486.

[Fe91] Fernholz, L.T. (1991). Almost sure convergence of smoothed empirical distribution functions. Scand. J. Statist. **18**, 255 - 262.

[Gae77] Gaenssler, P. and Stute, W. (1977). Wahrscheinlichkeitstheorie. Springer Hochschultext.

[Gae79] Gaenssler, P. and Stute, W. (1979). Empirical Processes: A survey of results for independent and identically distributed random variables. Ann. Probab. **7**, 193 - 243.

[Gae83] Gaenssler, P. (1983). Empirical Processes. IMS Lecture Notes - Monograph Series, Vol 3, Hayward.

[Gae86] Gaenssler, P. (1986). Bootstrapping empirical measures indexed by Vapnik-Chervonenkis classes of sets. In: Prob. Theory and Math. Stat. Vol **1**, pp. 467 - 481, Prohorov et al. (eds.), VNU Science Press.

[Gae87] Gaenssler, P. (1987). On a modification of the k-means clustering procedure. Proc. Fifth Int. Symp. on Data Analysis and Informatics, E. Diday (Ed.), 365 - 371. North Holland.

[Gae87b] Gaenssler, P. and Stute, W. (1987). Seminar on Empirical Processes. DMV-Seminar, Bd. **9**, pp. 1-110.

[Gae90] Gaenssler, P. and Schneemeier, W. (1990). On Functional Limit theorems for a Class of Stochastic Processes Indexed by Pseudo-Metric Parameter Spaces. In: Probability in Banach Spaces, **9**. U. Haagerup, J. Hoffmann-Jørgensen and N.J. Nielsen (Eds.). Progress in Probability **20**, pp. 82 - 100, Birkhäuser.

[Gae90b] Gaenssler, P. (1990). Confidence bands for probability distributions on Vapnik-Chervonenkis classes of sets in arbitrary sample spaces using the bootstrap. In: Bootstrapping and Related Techniques, Proceedings, Trier, FRG, 1990. Springer-Lecture Notes in Economics and Mathematical Systems, Vol. **376**, 1992, 57 - 61.

[Gae92] Gaenssler, P. (1992). On weak convergence of certain processes indexed by pseudo-metric parameter spaces with applications to empirical processes. Transactions of the 11 th. Prague conference, pp. 49 - 78, Academia.

[Gae94] Gaenssler, P. (1994). On recent developments in the theory of set-indexed processes. In: Asymptotic Statistics. P. Mandl and M. Hušková (Eds.), pp. 87 - 109. Physica-Verlag.

[Gae94b] Gaenssler, P. and Ziegler, K. (1994). A uniform law of large numbers for set-indexed processes with applications to empirical and partial-sum processes. In: Probability in Banach Spaces, **9**, 385 - 400. J. Hoffmann-Jørgensen, J. Kuelbs and M.B. Marcus (Eds.), Birkhäuser.

[Gae98] Gaenssler, P., Rost, D. and Ziegler, K. (1998). On Random Measure Processes with Application to Smoothed Empirical Processes. In: High Dimensional Probability I, 93-102. Progress in Probability, Vol. **43**, E. Eberlein, M. Hahn and M.Talagrand (Eds.), Birkhäuser.

[Gae99] Gaenssler, P. and Rost, D. (1999). On uniform laws of large numbers for smoothed empirical measures. Preprint, Univ. of Munich.

[Gae00] Gaenssler, P. and Rost, D. (2000). On uniform laws of large numbers for smoothed empirical measures. In: High Dimensional Probability II, 79 - 87. Progress in Probability, Vol **47**. E. Giné, D.M. Mason and J.A. Wellner (Eds.), Birkhäuser, Boston.

[Gae03] Gaenssler, P. and Rost, D. (2003). Smoothed Empirical Processes and the Bootstrap. In: High Dimensional Probability III, 313-320. Progress in Probability, Vol. **55**. J. Hoffmann-Jørgensen, M.B. Marcus and J.A. Wellner (Eds.), Birkhäuser.

[Gee00] van de Geer, S. (2000). Empirical Processes in M-Estimation. Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, Cambridge.

[Gi84] Giné, E. and Zinn, J. (1984).Some limit theorems for empirical processes. Ann. Probab. **12**, 929 - 989.

[Gi86] Giné, E. and Zinn, J. (1986). Lectures on the Central Limit Theorem for Empirical processes. Springer Lecture Notes in Math. **1221**, 50 - 113.

[Gi90] Giné, E. and Zinn, J. (1990). Bootstrapping general empirical measures. Ann. Probab. **18**, 851 - 869.

[Gi91] Giné, E. and Zinn, J. (1991). Gaussian characterization of uniform Donsker classes of functions. Ann. Probab. **19**, 758 - 782.

[Gi96] Giné, E. (1996). Empirical processes and applications: An overview. Bernoulli **2** (1), 1 - 28.

[Gi97] Giné, E. (1997). Lectures on some aspects of the bootstrap. École d'Été de Calcul de Probabilités de Saint-Flour XXVI – 1996. In: Giné, Grimmett and Saloff-Coste: Lectures on Probability Theory and Statistics. Springer Lecture Notes in Math. **1665**.

[Gl33] Glivenko, V. (1933). Sulla determinazione empirica della legge di probabilitá. Giorn. Ist. Ital Attuari **4**, 92 - 99.

[Gro92] Groeneboom, P. and Wellner, J.A. (1992). Information bounds and nonparametric maximum likelihood estimation. Birkhäuser, DMV Seminar Band 19.

[Ho74] Hoffmann-Jørgensen, J. (1974). Sums of independent Banach space valued random variables. Studia Math. **52**, 159 - 189.

[Ho84] Hoffmann-Jørgensen, J. (1984). Stochastic Processes on Polish Spaces. Unpublished manuscript.

[Ho85] Hoffmann-Jørgensen, J. (1985). Necessary and sufficient conditions for the uniform law of large numbers. Springer Lecture Notes in Math. **1153**, 258 - 272.

[Ho91] Hoffmann-Jørgensen, J. (1991). Stochastic Processes on Polish Spaces. Various Publication Series Vol. 39, Aarhus Universitet.

[Ho02] Hoffmann-Jørgensen, J. (2002). Perfect Independence and Stochastic Inequalities. MaPhySto Research Report no. 12, April 2002, p. 1-71. Centre for Mathematical Physics and Stochastics, University of Aarhus, Denmark. To appear in High Dimensional Probability III, Proceesdings of the Third International Symposium at Sandbjerg Estate, June 24-28, 2002, Denmark; published by Birkhäuser.

[Hoe63] Hoeffding, W. (1963). Probability inequalities for sums of random variables. J. Amer. Statist. Assoc. **58**, 13 - 30.

[Hu67] Huber, P.J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. Proc. Fifth Berkeley Symp. Math. Statis. Prob. **1**, 221 - 233. Univ. of California Press.

[Ke61] Kelley, J.L. (1961). General Topology. van Nostrand.

[Ki90] Kim, J. and Pollard, D. (1990). Cube root asymptotics. Ann. Statist. **18**, 191 - 219.

[La92] Laskowski, M.C. (1992). Vapnik-Chervonenkis classes of definable sets. J. London Math. Soc. II, Ser. **45**, No 2, 377 - 384.

[Le91] Ledoux, M. and Talagrand, M. (1991). Probability in Banach spaces. Springer.

[Lé40] Lévy, P. (1940). Le mouvement Brownien plan. Amer. J. Math. **62**, 487 - 550.

[Lé45] Lévy, P. (1945). Sur le mouvement brownien dépendant de plusieurs parameters. C.R. Acad. Sci. Paris **220**, 420 - 422.

[Li90] Liese, F. (1990). Estimation of Intensity Measures of Poisson Point Processes. In: Transactions 11 th. Prague Conf., Vol. A, pp. 121 - 139. Kluwer.

[Ma84] Mack, Y.P.C. (1984). Remarks on some smoothed empirical distribution functions and processes. Bull. of Informatics and Cybernetics **21**, 29 - 35.

[Mo02] Molnár, P. (2002). Zur Konstruktion von Konfidenzbändern für Wahrscheinlichkeiten in beliebigen Stichprobenräumen mittels Bootstrap. Diploma Thesis, Ludwig-Maximilians Universität München.

[Ne71] Neuhaus, G. (1971). On weak convergence of stochastic processes with multidimensional time parameter. Ann. Math. Statist. **42**, 1285 - 1295.

[No89] Nolan, D. and Marron, J.S. (1989). Uniform consistency of automatic and location-adaptive delta-sequence estimators. Probab. Theory Relat. Fields **80**, 619 - 632.

[Os84] Ossiander, M. (1984). Weak convergence and the law of the iterated logarithm of partial-sum processes indexed by points in a metric space. University of Washington Ph.D. Dissertation.

[Os85] Ossiander, M. and Pyke, R. (1985). Lévy's Brownian motion as a set-indexed process and a related central limit theorem. Stoch. Proc. and Their Appl. **21**, 133 - 145.

[Os87] Ossiander, M. A central limit theorem for under metric entropy with L_2 bracketing. Ann. Probab. **15**, 897 - 919.

[Pf94] Pfanzagl, J. (1994). Parametric Statistical Theory. de Gruyter Textbook.

[Pi83] Pisier, G. (1983). Some applications of the metric entropy condition to harmonic analysis. Banach spaces, Harmonic Analysis, and Probability Theory. Springer Lecture Notes in Math. **995**, 123 - 154.

[Po81] Pollard, D. (1981). Limit theorems for empirical processes. Z. Wahrsch. verw. Gebiete **57**, 181 - 195.

[Po82a] Pollard, D. (1982). Quantization and the method of k-means. IEEE Trans. Information Theory, Vol. IT-**28**, No 2, 199 - 205.

[Po82b] Pollard, D. (1982). A central limit theorem for k-means clustering. Ann. Probab. **10**, 919 - 926.

[Po84] Pollard, D. (1984). Convergence of Stochastic Processes. Springer.

[Po90] Pollard, D. (1990). Empirical Processes: Theory and Applications NSF-CBMS Regional Conference Series in Probability and Statistics, Vol 2, Hayward, CA: Institute of Mathematical Statistics.

[Pra81] Prakasa Rao, B.L.S. (1981). Convergence of smoothed empirical processes for independent random variables. Aligarh J. of Statist. **1**, 13 - 18.

[Pro56] Prohorov, Yu V. (1956). Convergence of random processes and limit theorems in probability theory. Theor. Probab. Appl. **1**, 157 - 214.

[Py72] Pyke, R. (1972). Empirical Processes. Jeffrey-Williams Lectures 1968 - 1972, Canadian Mathematical Congress, Montreal (1972), 13 - 43.

[Py73] Pyke, R. (1973). Partial sums of matrix arrays and Brownian sheets. In: D.G. Kendall and E.F. Harding, Eds., Stochastic Analysis, pp. 331 - 448. Wiley.

[Py84] Pyke, R. (1984). Asymptotic results for empirical and partial-sum processes: A review. The Canadian Journal of Statistics **12**, No. 4, 241 - 264.

[Py88] Pyke, R. and Wilbour, D.C. (1988). New approaches for goodness-of-fit tests for multidimensional data. In: Statistical Theory and data Analysis II, pp. 139 - 154. Elsevier Science Publishers (North Holland).

[Ra62] Ranga Rao, R.R. (1962). Relations between weak and uniform convergence of measures with applications. Ann. Math. Statist. **33**, 659 - 680.

[Re70] Rényi, A. (1970). Probability Theory. North Holland.

[Ro91] Rost, D. (1991). Das k-mean Clusterverfahren mit Gewichtsfunktion. Dissertation. Univ. of Munich.

[Ro95] Rost, D. (1995). A Simulation Study of the Weighted k-means Cluster Procedure. J. Statist. Computation and Simulation **53**, 51 - 63.

[Ro97] Rost, D. (1997). Approximation results for smoothed empirical processes: A new approach. Habilitationsschrift, Ludwig-Maximilians-Universität München.

[Ro99] Rost, D. (1999). Limit theorems for Smoothed Empirical Processes. Preprint.

[Ro99a] Rost, D. (1999). A consistency theorem for smoothed bootstrapped empirical processes, Preprint.

[Ro00] Rost, D. (2000). Limit theorems for smoothed empirical processes. Progress in Probability, Vol **47**. In: High Dimensional Probability II, 107 - 113. E. Giné, D.M. Mason and J.A. Wellner (Eds.) Birkhäuser, Boston.

[Sha95] Shao, J. and Tu, D. (1995). The Jacknife and the Bootstrap. Springer Series in Statistics.

[She92] Sheehy, A. and Wellner, J.A. (1992). Uniform Donsker classes of functions. Ann. of Probab. **20**, 1983 - 2030.

[Sh86] Shorack, G.R. and Wellner, J.A. (1986). Empirical Processes with Applications to Statistics. Wiley.

[Si81] Sing, K. (1981). On the asymptotic accuracy of Efron's bootstrap. Ann. Statist. **9**, 1187 - 1195.

[St90] Strobl, F. (1990). Zur Verteilungskonvergenz von Zufallsgrössen. Schriftliche Hausarbeit. Universität München.

[St94] Strobl, F. (1994). Zur Theorie empirischer Prozesse unter besonderer Berücksichtigung des Bootstrap-Verfahrens und Fréchet-differenzierbarer statistischer Funktionale. Dissertation, Univ. of Munich.

[St95] Strobl, F. (1995). On the reversed sub-martingale property of empirical discrepancies in arbitrary sample spaces. J. Theoret. Probab. **8**, 825 - 831.

[Ste78] Steele, M. (1978). Empirical discrepancies and subadditive processes. Ann. Probab. **6**, 118 - 127.

[Sten89] Stengle, G. and Yukich, J.E. (1989). Some New Vapnik-Chervonenkis classes. Ann. Statist. **17**, 1441 - 1446.

[Str71] Straf, M.L. (1971). Weak convergence of stochastic processes with several parameters. Proc. Sixth Berkeley Symp. Math. Statist. Prob. **2**, 187 - 221. Univ. of California Press.

[Stu86a] Stute, W. (1986). Conditional empirical processes. Ann. Statist. **14**, 638 - 647.

[Stu86b] Stute, W. (1986). On almost sure behavior of conditional empirical distribution functions. Ann. Probab. **14**, 891 - 901.

[Stu88] Stute, W. (1988). Empirische Prozesse in der Datenanalyse. Jber. d. Dt. Math.-Verein. **90**, 129-144.

[Stu97] Stute, W. (1997). Nonparametric model checks for regression. Ann. Statist. **25**, 613 - 641.

[Stu98] Stute, W., Thies, S. and Li-Xiang Zhu (1998). Model checks for regression: An innovation process approach. Ann. Statist. **26**, 1916 - 1934.

[Ta96] Talagrand, M (1996). The Glivenko-Cantelli Problem, Ten years later. J. of Theoret. Probab. **9**, 371 - 384.

[Va94] van der Vaart, A.W. (1994). Weak convergence of smoothed empirical processes. Scand. J. Statist. **21**, 501-504.

[Va96] van der Vaart, A.W. and Wellner, J.A. (1996). Weak convergence and empirical processes. Springer Series in Statistics.

[Val64] Valentine, F. (1964). Convex Sets. McGraw Hill.

[Vap71] Vapnik, V.N. and Chervonenkis, A.Y. (1971). On uniform convergence of the frequencies of events to their probabilities. Theor. Probab. **16**, 264 - 280.

[We92] Wellner, J.A. (1992). Empirical Processes in Action. A Review. Internat. Statist. Review **60**, 3, 247 - 269.

[Wen81] Wenocur, R.S. and Dudley, R.M. (1981). Some special Vapnik-Chervonenkis classes. Discrete Math. **33**, 313 - 318.

[Wi68] Wichura, M.J. (1968). On the weak convergence of non-Borel probabilities on a metric space. Ph.D. dissertation, Columbia University.

[Win73] Winter, B.B. (1973). Strong uniform consistency of integrals of density estimators. Canadian J. of Statist. **1**, 247-253.

[Wit70] Witting, H. and Nölle, G. (1970). Angewandte Mathematische Statistik. B.G. Teubner Stuttgart.

[Wo60] Wolfowitz, J. (1960). Convergence of the empiric distribution function on half spaces. In: Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling. Stanford Univ. Press.

[Ya73] Yamato, H. (1973). Uniform convergence of an estimator of a distribution function. Bull. Math. Statist. **15**, 69-78.

[Yu89] Yukich, J.E. (1989). A note on limit theorems for perturbed empirical processes. Stoch. Proc. and Their Applications. **33**, 163-173.

[Yu92] Yukich, J.E. (1992). Weak convergence of smoothed empirical processes. Scand. J. Statist. **19**, 271-279.

[Zi94] Ziegler, K. (1994). On Functional Central Limit Theorems and Uniform Laws of Large Numbers for Sums of Independent Processes. Dissertation, Univ. of Munich.

[Zi97] Ziegler, K. (1997). Functional Central Limit Theorems for Triangular Arrays of Function-indexed Processes under Uniformly Integrable Entropy Conditions. J. of Multivariate Analysis **62**, 233 - 272.