Université de Franche-Comté

Document de synthèse en vue de l'obtention du diplôme

d'HABILITATION À DIRIGER DES RECHERCHES

Contributions à l'étude de la mesure empirique et de la mesure empirique locale. Applications en statistique non paramétrique.

Davit VARRON

Neuf Septembre 2014

Présenté devant le jury composé de

Pr. Philippe BERTHET,	Université Paul Sabatier,	Rapporteur,
Pr. Laurent BORDES,	Université de Pau et des Pays de l'Adour,	Rapporteur,
Pr. Benoît CADRE,	École Normale Supérieure de Rennes,	Rapporteur,
Pr. Jean-Yves DAUXOIS,	Institut National de Sciences Appliquées de Toulouse,	Examinateur
Pr. Célestin KOKONENDJI,	Université de Franche-Comté,	Examinateur
Pr. Aad VAN DER VAART,	University of Leiden,	Rapporteur.

La rédaction d'un manuscrit de synthèse est l'occasion parfaite d'écrire noir sur blanc mes remerciement à tous ceux qui auront joué un rôle très important dans mon début de carrière professionnelle.

Mon premier remerciement revient à Philippe Berthet. Je me garderai toujours en mon souvenir ton cours de DEA concernant les processus empiriques, que j'ai eu la chance de suivre, et qui a provoqué en moi un véritable déclic, voire même un électrochoc. Si ma vocation vers la recherche s'était déclarée dès le début de mes études supérieures, elle ne prit de forme concrète qu'à ce moment précis. J'ai pu faire mes premiers pas en recherche avec toi, pendant que tu encadrais mon mémoire de DEA. Je l'ai fait avec une intensité fortement attribuable à l'enthousiasme contagieux dont tu fais preuve à chaque conversation de maths. Avec le recul des années, je me rends compte que le sujet de mémoire que tu m'avais proposé était d'une profondeur idéale pour maintenir la stimulation de l'étudiant que j'étais, en parsemant le parcours de nouvelles étincelles, au gré des articles, dans lesquels j'ai pu me plonger avec passion. Ce mémoire de DEA fut pour moi un événement majeur de ma carrière de chercheur. Grâce à toi j'ai été mis sur de bons et solides rails.

Mon second remerciement est directement adressé à l'ensemble du Laboratoire de Mathématiques de Besançon. Il m'est superflu de parler de son niveau scientifique, l'AERES s'en est chargée elle même. Mais, quitte à choquer certains, ce n'est pas l'excellence que je remercie ici : c'est l'ambiance. Une ambiance de travail avant tout sereine et conviviale. Avec bien sûr ses couacs occasionnels, mais somme toute bien peu fréquents. Combien d'autres laboratoires pourraient rivaliser avec le LMB sur cet aspect, qui a toute son importance pour des activités d'enseignement et de recherche de qualité ? S'il fallait une note, je mettrais A^+ sans hésiter.

Une mention particulière à vous deux, Jean-Yves Dauxois et Stéphane Chrétien, qui avez par-

ticulièrement contribué à mon accueil au LMB, en 2006. Grâce à vous, qui avez assumé les reponsabilités du master modélisation statistique, à l'époque reposant sur les épaules d'une équipe de statisticiens réduite à 3 membres, j'ai pu prendre mes marques sans aucun stress, pour trouver, à mon rythme, ma place dans l'équipe, et dans le LMB. Je vous en suis extrêmement reconnaissant.

Me voici donc à l'étape de l'habilitation, moment important de la vie d'un chercheur. On m'avait dit : "prends-toi un an à l'avance pour constituer ton jury. Beaucoup de professeurs risquent de refuser, ils ne veulent pas perdre de temps à cela". J'ai donc pris contact dès l'année 2013. Vous avez tous répondu présent, et vous l'avez fait sans hésiter. Philippe, Laurent, Benoît, Jean-Yves, Célestin, c'est un honneur et un plaisir de vous compter, chacun d'entre vous, dans mon jury d'habilitation.

I am very grateful to Pr. van der Vaart, who had the kindness to review my habilitation thesis. Your opinion on my works is of particular importance to me. For this, I give you all my deepest thanks.

Contents

Chapter 1

Introduction

1.1 The central objects

The empirical processes theory, and its applications to statistics, have been the core of my research works. In the two following subsections I will give a definition of two central objects, each of them having connections with distinct subfields of nonparametric statistics.

1.1.1 The general empirical process

Consider a measurable space $(\mathfrak{X}, \mathcal{X})$. Write $\mathfrak{M}$ for the set of all probability measures on $(\mathfrak{X}, \mathcal{X})$. Given $\mathbf{P} \in \mathfrak{M}$, and $g \in L^1(\mathbf{P})$, we shall write

$$\mathbf{P}(g) := \int_{\mathfrak{X}} g d\mathbf{P}.$$

The general empirical process is defined as follows.

— Given a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, consider an independent, identically distributed [i.i.d.] sequence of random variables $(Z_i)_{i\geq 1}$, from $(\Omega, \mathcal{A}, \mathbb{P})$ to $(\mathfrak{X}, \mathcal{X})$. We will denote by $\mathbf{P}_0 \in \mathfrak{M}$ the common law of each Z_i.

— For fixed $n \geq 1$, define the empirical measure

$$\mathbf{P}_n := \frac{1}{n} \sum_{i=1}^{n} \delta_{Z_i},$$

as a (composition) map from Ω to $\mathfrak{M}$.

— Consider a class of functions $\mathcal{G} \subset L^1(\mathbf{P}_0)$. Given $n \geq 1$, define the corresponding empirical process by

$$G_n(\cdot) := g \to \sqrt{n}\big(\mathbf{P}_n(g) - \mathbf{P}_0(g)\big),$$

as a map from Ω to $\mathbb{R}^{\mathcal{G}}$, where the latter is the space of all real functions on $\mathcal{G}$.

Note that $\mathbf{P}_n$ is random, for each $n \geq 1$, while $\mathbf{P}_0$ is not.
In statistics, G_n has a particular relevance, since its stochastic behavior provides a description of the random spatial setting of the sample $(Z_1, \ldots, Z_n)$, which, itself, rules the stochastic behavior of general estimating procedures such as , e.g., Z-estimation, M-estimation, bootstrap procedures, or empirical likelihood procedures (see, [92, Part 3] for a large, yet non exhaustive overview of the possibilities).

1.1.2 The local empirical process

A particular domain of nonparametric statistics concerns smoothing techniques, toward the estimation of Lebesgue densities or conditional expectations. In this framework, it is of interest to understand the stochastic behavior of local empirical processes. Taking $\mathfrak{X} := \mathbb{R}^d$ and a class of functions $\mathcal{G} \subset L^1(\mathbf{P}_0)$, we can define, for $n \geq 1, z \in \mathbb{R}^d,\ h > 0$, and $g \in \mathcal{G}$:

$$\begin{aligned} T_n(g,h,z) :=& \sum_{i=1}^n g_{h,z}(Z_i) - \mathbb{E}\Big(g_{h,z}(Z_i)\Big), \text{ where} \\ g_{h,z}(\mathbf{z}) :=& g\Big(h^{-1}\big(\mathbf{z} - z\big)\Big),\ \mathbf{z} \in \mathbb{R}^d. \end{aligned}$$

This object describes the random setting of the sample *locally* around z, with $h > 0$ (usually called **the bandwidth**) playing the role of the localization strength.
Given a class of functions $\mathcal{G}$, the asymptotic behavior, as $n \to \infty$ together with $h \to 0$, of $T_n(\cdot, h, z)$ (as processes indexed by $\mathcal{G}$) has been intensively investigated in the literature over the past three decades (an illustrative, yet non exhaustive list of references is [8, 9, 11, 17, 18, 20, 21, 22, 23, 24, 37, 38, 41, 64, 65, 66]). It has to be noted that, the relationship between the $T_n(\cdot, h, z)$ and nonparametric estimation should not be narrowed to the Parzen-Rosenblatt and Nadaraya-Watson estimators. For example, estimation by wavelet expansions or by local polynomials are also linked to $T_n(\cdot, h, z)$, and more generally, this is also the case for local M-estimation or Z-estimation, bootstrap procedures, or empirical likelihood procedures.

1.2 Contributions to the study of the general empirical process

In this section, I will sum up my contributions to the general empirical processes theory.

1.2.1 The uniform entropy condition

An important result in empirical processes theory is that a class of functions $\mathcal{G}$ having a square integrable envelope and admitting a *finite uniform entropy integral* is Donsker. That condition, very commonly encountered in practice, has the particularity of involving suprema of packing numbers over all possible $L^2(Q)$-norms induced by probability measures Q. I investigated how these uniform bounds could be used as a basis for proving Donsker and Glivenko-Cantelli theorems for sequences of random measures of the form

$$\widetilde{\mathbf{P}}_n := \sum_{i\geq 1} \beta_{i,n}\delta_{Z_{i,n}}.$$

The initial motivation of this work came from Bayesian nonparametric theory, where discrete priors and posteriors can be represented this way, most particularly the Dirichlet process prior/posterior random measure.

— **Contribution** [105]. I showed that the *finite uniform entropy integral* condition still does adapt very well to the more general form of $\widetilde{\mathbf{P}}_n$, when the $(Z_{i,n})_{i\geq 1}$ are conditionally i.i.d given $(\beta_{i,n})_{i\geq 1}$. Indeed, that uniform entropy numbers condition turns out to be sufficient (beside a minimal integrability condition on the envelope of $\mathcal{G}$) to prove a Glivenko-Cantelli and a Donsker theorem, as soon as the conditional distributions of $Z_{1,n}$ given $(\beta_{i,n})_{i\geq 1}$ converge to a limit in the sense of Sheehy and Wellner [83]. Note that boostraped empirical procedures fall into this framework (see, e.g., [92, Chapter 3.6]). A direct consequence of that result is an alternate proof of Doob's theorem and Bernstein-Von Mises theorem due to James [51], under the topology spanned by $||\cdot||_{\mathcal{G}}$, for posterior distributions of the Dirichlet process prior.

1.2.2 Contributions to empirical likelihood theory

Empirical likelihood is a nonparametric method of building confidence regions, which was developed by Owen [74]. It has many advantages over the usual asymptotic gaussian pivot, and turns out to be a serious competitor to the pivotal bootstrap.

— **Contribution** [15], [107]. With J-Y Dauxois and A. Flesch, we investigated the estimation of functional parameters by empirical likelihood, in a particular setup of lifetime data analysis. The parameter of interest was the joint trajectories of the mean numbers of recurrent events under competing risks, with independent right censoring, and in the presence of a terminal event. These works led to investigating a more general framework, where the functional parameter of interest can be expressed as $T(\mathbf{P}_n)$, with T fulfilling some Gâteau and Hadamard differentiability conditions, relatively to norms of the form $||\cdot||_{\mathcal{G}}$. These works are improvements of already existing results of Bertail [7] and [50], to multisample, multivariate settings, and with an additional care on practical implementation.

1.2.3 Contributions to associated discrete kernel estimation

Assume that the law of Z_1 is *discrete* with support $\mathbb{T}$. The nonparametric estimation of the probability mass function (p.m.f) $f(\cdot)$ by the empirical p.m.f.

$$f_n:\ z \to \mathbf{P}_n(\{z\})$$

can be put into the wider framework of *discrete associated kernel estimators*

$$f_{n,h} : z \to \frac{1}{n}\sum_{i=1}^{n} K_{z,h}(Z_i)$$

where $\mathcal{K} = \{K_{z,h}(\cdot),\ z \in \mathbb{T},\ h \geq 0\}$ is a family of p.m.f on $\mathbb{T}$. Despite its expression, $f_{n,h}$ has nothing to do with the local empirical process.

— **Contribution** [56]. With C. Kokonendji, we investigated the consistency of this class of estimators, for the total variation distance

$$TV(f_{n,h}, f) := \frac{1}{2}\int_{\mathbb{T}} \mid f_{n,h}(x) - f(x) \mid dc(x) + \frac{1}{2}\Big|\int_{\mathbb{T}} f_{n,h}(x)dc(x) - 1\Big|,$$

where c denotes the counting measure on $\mathbb{T}$, under assumptions on $\mathcal{K}$ that are satisfied in practice. We could also derive concentration inequalities for $TV(f_{n,h}, f)$.

1.3 Contributions to the study of local empirical processes

Under an additional condition of uniform boundedness on $\mathcal{G}$, or on integrability of its envelope, both local an general empirical processes are viewed as random elements in $\big(\ell^\infty(\mathcal{G}), ||\cdot||_{\mathcal{G}}\big)$, where $\ell^\infty(\mathcal{G})$ is the space of all bounded real functions on $\mathcal{G}$, and where

$$|| \psi ||_{\mathcal{G}} := \sup_{g\in\mathcal{G}} \mid \psi(g) \mid, \psi \in \ell^\infty(\mathcal{G}).$$

One can very rouhgly divide the existing literature into the following types of results.

1. Asymptotic results
   (a) Donsker theorems and weak approximations.
   (b) Functional limit laws and strong approximations.
   (c) Large deviation principles.
2. Non asymptotic results
   — Concentrations inequalities.
   — Control of first moments by chaining techniques.

Concerning the local empirical process, most of my work takes place in (b) and (c). This is the point of the next subsections.

1.3.1 Refinements of existing functional limit laws

Consider a compact set $H \subset \mathbb{R}^d$, and assume that, for an open set $\mathfrak{O} \supset H$, the following condition is satisfied :

(Hf) Z_1 admits a Lebesgue density on $\mathfrak{D}$ for which there exists a version f that is continuous and bounded away from 0 and ∞ on $\mathfrak{D}$.

Now write, given $n \geq 1$ and $h > 0$:

$$\Theta_n(h) := \left\{ \frac{T_n(\cdot, h, z)}{r_n(h, z)},\ z \in H \right\},$$

Where $r_n(h, z) > 0$. A functional limit law for the local empirical process is any theorem stating the almost sure inner and outer topological limits (in $(\ell^\infty(\mathcal{G}), || \cdot ||_{\mathcal{G}})$) of the sequence $\Theta_n(h_n)$, for appropriate normalizing sequences $r_n(h_n, z)$, where $h_n > 0$ is a deterministic bandwidths sequence.

- When H is finite (say, a singleton $\{z_0\}$), those type of results are called *local* functional limit laws, and $r_n(h, z_0) := \sqrt{2f(z_0)nh^d \log\log(n)}$ is the appropriate type of normalization.
- When H has a nonempty interior, those type of results are called *global* functional limit laws, and $r_n(h, z) := \sqrt{2f(z)nh^d \log(1/h^d)}$ is the appropriate type of normalization.
- When the inner/outer topological limits are related to the large deviations properties of the associated Gaussian processes (Strassen-type sets), those limit laws are called *standard*. This happens when h_n tends to zero, but not too fast. For local limit laws, "not too fast" means
$$nh_n^d/\log\log(n) \to \infty, \tag{1.1}$$
while this term has to be understood as
$$nh_n^d/\log(1/h_n^d) \to \infty, \tag{1.2}$$
for global functional limit laws. Standard functional limit laws entail several (strong) consistency or uniform consistency (over a compact H) results for nonparametric or semiparametric estimators.
- When the inner/outer topological limits are related to the large deviations properties of the associated Poisson processes, those limit laws are called *nonstandard*. This happens when the convergence h_n is slightly too fast, which means replacing the symbol ∞ by a nonzero constant in (1.1) or (1.2). Nonstandard functional limit laws entail several non consistency or non uniform consistency results (in the strong sense) for estimators.

The simplest form of the local empirical process consists in taking $d = 1$, Z_i uniformly distributed in $[0, 1]$, and $\mathcal{G} = \mathcal{G}_0$ as the class of indicators of intervals of the form $[0, t]$, $t \in [0, 1]^d$. The object $T_n(\cdot, h, z)$ is then called a *functional increment of the uniform empirical process*. The pioneering works are due to [65], [22, 21], proving both standard and nonstandard functional limit laws for those functional increments. Since then, the challenge was to extend those results into the following directions

1. To extend $\mathcal{G}_0$ to more general classes of functions. In particular, to confront the usual conditions of *uniform entropy numbers* and *bracketing numbers* with those functional

limit laws.

2. To go past the unidimensional case.
3. To handle other (continuous) distributions for the $(Z_i)_{i\geq 1}$.

My contributions on this field are

— **Contribution 1** [96]. In the local standard functional limit law for the increments of the uniform empirical process, I established that the *clustering rates* of $\Theta_n(h_n)$ to the corresponding Strassen set are of order $\log\log(n)^{-2/3}$. This result is not surprising in regard of a results in global standard limit laws established by Berthet [8, 9]. I also filled a small gap in the Chung-Mogulskii laws for the local empirical process, providing rates of approximation of a certain class of points belonging to the boundary of the Strassen set.

— **Contribution 2** [99, 100]. In the global nonstandard functional laws, I managed to attain the multidimensional setting ($d \geq 2$), under (Hf), for any uniformly bounded class $\mathcal{G}$ with supports included in a common bounded set $[-M, M]^d$, admitting finite bracketing numbers (for the uniform distribution on $[-M, M]^d$). I pointed out a consequence of that result to the asymptotic behavior of estimators of densities by wavelet projections.

— **Contribution 3** [68]. With M. Maumy-Bertrand, we proved the *local* counterpart of Contribution 2, and also extended this local law to the more general object $T_{n,\mathfrak{c}}(\mathbf{g}, \cdot, h, z)$, where

$$T_{n,\mathfrak{c}}(\mathbf{g}, g, h, z) := \sum_{i=1}^{n} g_{h,z}(Z_i)\mathbf{g}(Y_i),$$

for a function $\mathbf{g}$ taking values in $\mathbb{R}^k$, for which we made assumptions of finite conditional exponential moments for $\mathbf{g}(Y_1)$ (note the absence of a centering parameter). Our result was limited to the particulat class $\mathcal{G}_{0,d}$ of indicators of hypercubes of $[0, 1]^d$.

1.3.2 The impact of the in-bandwidth uniformity to the usual functional limit laws

A very central question has animated a significant part of my research works. It can be stated as follows :

What is the impact, on the existing functional limit laws, of requiring an additional in-bandwidth uniformity for the convergences?

More precisely, given two deterministic bandwidth sequences $h_n \leq \mathfrak{h}_n$, can we still determine the almost sure topological limits of

$$\Theta_n := \bigcup_{h_n \leq h \leq \mathfrak{h}_n} \Theta_n(h),$$

and do those limits depend on the widths of the intervals $[h_n, \mathfrak{h}_n]$?
The original motivation comes from nonparametric statistics, where in-bandwidth-uniform consistency results for nonparametric estimators provides a rigorous justification for the use of data driven bandwidth selection procedures. I provided the following answers.

— **Contribution 1** [96]. When two bandwidths $h_n \leq \mathfrak{h}_n$ have different orders of convergence to 0 (i.e $h_n = o(\mathfrak{h}_n) = o(1)$), an asymptotic independence shows up : for fixed z, $T_n(\cdot, h_n, z)$ and $T_n(\cdot, \mathfrak{h}_n, z)$ obey the same large deviation principle and Strassen law as of two independent Wiener processes. I will call this phenomenon *between bandwidths asymptotic independence*. That phenomenon was already partially visible in [19].

— **Contribution 2** [108]. For local standard functional limit laws (see [65, 41]), there is an impact of requiring an in-bandwidth uniformity in $h \in [h_n, \mathfrak{h}_n]$, which is clearly due to the *between bandwidths asymptotic independence*. Indeed, when

$$\log\log(\mathfrak{h}_n/h_n)/\log\log(n) \to \delta \in (0,1], \tag{1.3}$$

the inner and outer topological limits of T_n are not anymore $\emptyset$ and $\mathcal{S}_{\mathcal{G},1}$, but $\mathcal{S}_{\mathcal{G},\delta}$ and $\mathcal{S}_{\mathcal{G},1+\delta}$, where for $a \geq 0$, $\mathcal{S}_{\mathcal{G},a}$ equals $\sqrt{a}$ times the Strassen ball of the associated Gaussian process.

— **Contribution 3** [98]. For global standard functional limit laws ([25],[64]), there is no impact of the uniformity in $h \in [h_n, \mathfrak{h}_n]$. At this level of normalization, nothing is visible.

— **Contribution 4** [93]. With I. van Keilegom, we investigated the global functional limit law for the more general sets

$$\Theta_{n,\mathfrak{c}} = \left\{ \frac{T_{n,\mathfrak{c}}(\cdot,\cdot,h,z) - \mathbb{E}\Big(T_{n,\mathfrak{c}}(\cdot,\cdot,h,z)\Big)}{r_n(h,z)},\ h \in [h_n, \mathfrak{h}_n],\ z \in H \right\}.$$

We did not prove that functional limit law. However, following initial works of Einmahl and Mason [42], we proved a closely related result, which is weaker in its form (because it is non functional), but which covers a larger framework in regression estimation.

1.3.3 Asymptotic results in a spatial setting

A very large part of the asymptotic results in empirical processes theory can be roughly stated as "both the empirical process and the associated Gaussian process share the same property". Wschebor [109] proved that the increments of the wiener process $\epsilon^{-1/2}\big(W(t+\epsilon) - W(t)\big)$ almost surely satisfy the property

$$\forall B \text{ Borel },\ \lim_{\epsilon\to 0} \lambda\bigg(\Big\{t \in [0,1],\ \frac{W(t+\epsilon) - W(t)}{\sqrt{\epsilon}} \in B\Big\}\bigg) = \mu(B),$$

where μ is the standard normal distribution and λ stands for the uniform distribution on $[0,1]$. This almost sure phenomenon can be roughly called an *almost sure spatial convergence in* *distribution* of the increments of the Wiener process.

— **Contribution 1** [103, 101]. I extended that result for the functional increments of the uniform empirical process, proving an almost sure spatial Donsker theorem , toward a Wiener measure. I then extended this result to the general form of $T_n(\cdot, h, z)$, under (Hf), for classes of functions $\mathcal{G}$ admitting a finite uniform entropy numbers integral, and for a large class of probability distributions different from λ (including those admitting a square integrable density, and excluding those having atoms).

— **Contribution 2** [104, 101]. In the same vein, I also established the almost sure spatial local standard functional limit law for Θ_n. I also proved the nonstandard version of that limit law, but restricted to $h_n = \mathfrak{h}_n$, and for the uniform empirical process.

1.3.4 The probabilistic tools that have been developed

For all the above mentioned contribution, the proof relied on extensions and refinement of already existing tools in empirical process theory. Two of them did involve a significant part of my works.

Poissonization techniques

If the sample size is randomized by a Poisson random variable with expectation n, independent of the sequence $(Z_i)_{i\geq 1}$, then the corresponding *poissonized local empirical process*

$$\Pi_n(g, h, z) := \sum_{i=1}^{\eta_n} g_{h,z}(Z_i) - \mathbb{E}\Big(g_{h,z}(Z_i)\Big),$$

inherits properties of random Poisson measures, which often turn out to be the nonasymptotic counterparts of asymptotic properties of the $T_n(\cdot, h, z)$. This properties of independence are very useful for probability calculus. Hence, an effort has been made to approximate the $T_n(\cdot, h, z)$ by the corresponding $\Pi_n(\cdot, h, z)$ during the past decades. For finite H and for very small $h_n = O(\log\log(n)/n)$, this can be achieved by a direct strong approximation [21, 24]. Outside this case, it is possible to obtain weaker but nevertheless useful approximations. One of them, initiated by J. Einmhal [36] and generalized by Deheuvels and Mason [22, 65] provided a bound of the form

$$\mathbb{P}\Big(T_n(\cdot,\cdot,\cdot) \in A\Big) \leq C\mathbb{P}\Big(\Pi_n(\cdot,\cdot,\cdot) \in A\Big),$$

as processes indexed by $\mathcal{G}\times]0, h_0] \times H$, where the constant C does *not* depend on A nor on n. This happens as soon a there is a non null probability that Z_1 falls outside the union of all supports of $g_{h,z},\ g \in \mathcal{G},\ h \leq h_0,\ z \in H$.
This very useful inequality was put in a wider framework by Giné, Mason and Zaitsev [46], for

sums of i.i.d. random variables in a measurable semigroup, proving that

$$\mathbb{P}\Big(\sum_{i=1}^{n} \mathbb{1}_B(X_i)X_i \in A\Big) \leq C\mathbb{P}\Big(\sum_{i=1}^{\eta_n} \mathbb{1}_B(X_i)X_i \in A\Big),$$

as soon as $\mathbb{P}\big(X_1 \notin B\big) > 0$.

— **Contribution 1** [103]. I extended inequality (1.4) to a wider class of *truncating* functions on $(X_i)_{i\geq 1}$, for which the arguments are the *partial sums histories* up to n (or η_n), namely the vector

$$\Big(X_1, X_1 + X_2, \ldots, X_1 + \ldots + X_n\Big).$$

As an example of application, it is now possible to poissonize maximal inequalities for sums of i.i.d random variables.

Though already very useful, inequalities (1.4) and (1.4) cannot make a connection, between $T_n(\cdot,\cdot,\cdot)$ and $\Pi_n(\cdot,\cdot,\cdot)$, of some crucial facts needed for functional limit laws. As a matter of fact, it is not possible to directly translate assertions of the type

$$\epsilon_n = O\Big(\mathbb{P}\big(\Pi_n(\cdot, h_n, z) \in A_n\big)\Big)$$

into an assertion of the same type for $T_n(\cdot, h_n, z)$.

— **Contribution 2** [108]. I proved an approximation result between $T_n(\cdot, h_n, z)$ and $\Pi_n(\cdot, h_n, z)$, by providing deviation probabilities for large exceedances of $|| T_n(\cdot, h_n, z) - \Pi_n(\cdot, h_n, z) ||_{\mathcal{G}}$. Those deviation probabilities opened new possibilities of using Poissonization in the *local standard* functional limit laws. Moreover, these deviation probabilities almost entirely depend on the concentration of η_n around n, and thus hold even if η_n and $(Z_i)_{i\geq 1}$ are not independent, or if η_n is not Poisson.

Uniform large deviation principles

Functional limit laws heavily rely on large deviation principles. Arcones [4, 5] proposed a criterion to prove large deviation principles in $\big(\ell^\infty(\mathcal{G}), || \cdot ||_{\mathcal{G}}\ \big)$. That paradigm decomposes a large deviation principle for a sequence of processes W_n into two parts :

1. First, prove a large deviation principle for processes discretized along any finite grid, $\big(W_n(g_1), \ldots, W_n(g_p)\big)$.
2. Second, control their oscillations on the elements of suitable finite partitions.

Thanks to the paradigm of Arcones [4, 5], Mason [65] showed that it was possible to obtain large deviation principles for $T_n(\cdot, h, z)$ *without* invoking any strong approximation argument.

— **Contribution 1** [96, 98, 104, 108]. I investigated how the arguments of Arcones and

Mason could be extended to obtain *uniform* large deviation principles of the form

$$\overline{\lim_{n\to\infty}} \sup_{\substack{z\in H_n,\\ h\in[h_n,\mathfrak{h}_n]}} \frac{1}{v_n(h)} \log\left(\mathbb{P}^*\Big(\frac{T_n(\cdot,h,z)}{r_n(h)}\in F\Big)\right) \leq -J(F),$$

for any F closed in $\big(\ell^\infty(\mathcal{G}), ||\cdot||_{\mathcal{G}}\big)$, where the presence outer probabilities is due to well known measurability issues for the empirical processes. The uniformity in $h \in [h_n, \mathfrak{h}_n]$ turns out to be crucial for proving in-bandwidth- uniform functional limit laws.

— **Contribution 2** [97, 106]. The criterion of Arcones can be seen as follows :

1. Given a family of finite rank operators $(I_\alpha)_{\alpha\in\Upsilon}$ of a Banach space $(E, ||\cdot||)$, prove a large deviation principle for $\big(I_\alpha(W_n)\big)_{n\geq 1}$ for each $\alpha \in \Upsilon$.
2. Sharply control $||\ I_\alpha(W_n) - W_n\ ||$ by a suitable choice of α.

I showed that this general idea was rigorously applicable in the context of Schauder decomposable spaces, then more generally in spaces having the bounded approximation property.

1.4 Inventory of publications and prepublications

1.4.1 Short notes

The short notes that are preceded by an asterisk have also been the subject of a full lenght article, provided in the next subsection.

1. *Uniformity in h in the functional limit law for the increments of the empirical process indexed by functions. **C. R. Acad. Sci. Paris, Ser. I** 340, p. 453-456 (2005).
2. *A nonstandard uniform functional law of the logarithm for the increments of the multivariate empirical process . **C. R. Acad. Sci. Paris, Ser. I** 343, p. 427-430 (2006).
3. A note on large deviation principles in Schauder decomposable spaces. **C. R. Acad. Sci. Paris, Ser. I** 343, p. 345-348 (2006).
4. Some asymptotic results on density estimators by wavelet projections. **Statistics and Probability Letters**, 78, p. 2517-2521 (2008).

1.4.2 Full length articles

1. Some uniform in bandwidth functional results for the tail uniform empirical and quantile processes. **Annales de l'ISUP**, 50, p. 83-103 (2006)
2. A limited in bandwidth uniformity for the functional limit law of the increments of the empirical process. **Electronic Journal of Statistics**, 2, p. 1043-1064 (2008).

3. A nonstandard uniform functional limit law of the iterated logarithm for the increments of the multivariate empirical distribution function. **Advances and Applications in Statistical Sciences** 1, p. 399-428 (2010).
4. (With I. Van Keilegom) Uniform in bandwidth exact rates for a class of kernel estimators. **Annals of the Institute of Statistical Mathematics** Volume 63, p. 1077-1102 (2011).
5. (With M. Maumy-Bertrand) Non standard functional limit laws for the increments of the compound empirical distribution function. **Electronic Journal of Statistics**, 4, p.1324-1344 (2010).
6. Some new almost sure properties of the increments of the uniform empirical process. **Stochastic Processes and Applications**, 121,No 2, p. 337-356 (2011).
7. Clustering rates and Chung laws of the iterated logarithm for empirical and quantile processes. **Annales de l'ISUP**, 55, No 2-3, pp. 3-26 (2011)
8. A note on weak convergence, large deviations, and the bounded approximation property. To appear in **Theory of probability an its applications**

1.4.3 Submitted articles

1. The almost sure behavior of some spatial repartitions of local empirical processes indexed by functions.
2. A spatial Strassen-type functional limit law for the increments of the local empirical process.
3. Simultaneous confidence bands for some functional plug-in parameters : a computationally feasible approach.
4. The almost sure Kuratowski limit of local empirical processes with variable bandwidths
5. (With J-Y. Dauxois and A. Flesch) Empirical likelihood uniform confidence bands in survival analysis under the assumption of competing risks.
6. Donsker and Glivenko-Cantelli theorems for a class of processes generalizing the empirical process.

1.4.4 Articles in preparation

1. (With C. Kokonendji) Performances of the discrete associated kernel estimators for the total variation distance.

Chapter 2

Uniform large deviations and Poissonization tools for local empirical processes

This chapter is dedicated to the probabilistic tools that I progressively developed during my research works. In this chapter, and in this chapter only, proofs are given in details. The main motivation of providing such a level of details is that taking time to expose them with precision will allow me to give to be a lot more elusive in the next chapter. Another underlying motivation is that those results somehow reflect the very base of my works : almost every result of Chapter 3 has his proof deeply relying on one of the probabilistic tools of the present chapter.

2.1 Uniform large deviation principles

2.1.1 The framework

A large deviation principle (LDP for short) states the rates of convergence to zero of rare events of the form $\mathbb{P}\big(W_n \in A\big)$, where W_n is a sequence of Borel random variables in a topological space. It is an important tool for functional limit laws, because (roughly speaking) it allows to determine the summability or non summability of sequences of such events, and hence apply the Borel-Cantelli Lemma or its converse part. The rates of convergence to zero is determined by a (rate) function J, which describes the *stochastic cost* of regions of E.

Definition 2.1.1 (Good rate function) *Let $(E,\mathfrak{O})$ be a topological space. A map $J:\ E\rightarrow [0,\infty]$ is called a good rate function (or rate function for short) when the sets $\big\{e\in E,\ J(e)\leq a\big\}$, $a\geq 0$ are compact sets of $(E,\mathfrak{O})$.*
For $A\subset E$, we shall write $J(A):=\inf_{e\in A}\ J(e)$.

When the topological space is Polish (metric, complete, separable), many criteria have been established to prove large deviation principles (see, e.g. Lynch and Sethuraman [63] for continuous processes with independent increments under the sup-norm, Pukhalskii [77, Theorem 2] for Skorokhod spaces, the abstract version of the Gärtner-Ellis theorem [27, Theorem 4.5.20, p. 157], Bryc's inverse lemma [27, Chapter 4.4]). Arcones [4, 5] was the first to provide such a paradigm in a nonseparable space : the space $\big(\ell^\infty(T), ||\cdot||_T\big)$, where we remind the reader that, for a set T, $\ell^\infty(T)$ denotes the space of all real bounded functions on T, endowed with the sup norm

$$|| \psi ||_T := \sup_{t\in T} | \psi(t) | . \tag{2.1}$$

He showed that a LDP in $\big(\ell^\infty(T), ||\cdot||_T\big)$ can always be decomposed into two parts :

— For any $p \geq 1$ and $(t_1, \ldots, t_p) \in T^p$, prove the expected LDP for the finite dimensional random vectors $\big(W_n(t_1), \ldots, W_n(t_p)\big)$.

— Prove that the oscillations on a suitable finite partition of T are negligible, in the sense that their probabilities of exceeding arbitrarily small thresholds $\epsilon > 0$ tend to zero at sufficiently fast rates.

Arcones makes an explicit use of the particular structure of $\big(\ell^\infty(T), ||\cdot||_T\big)$, by invoking the Arzela-Ascoli theorem as well as extensions of uniformly continuous functions on a set T_0 to its completion.

Following a brief remark of Mason [65], and after a careful reading of the proofs in [4], I came to the conclusion that all the arguments of Arcones could be extended with no efforts to a larger setup of *uniform* large deviation principles. The problem of non (Borel) measurability of empirical processes is well known. This is why the following definition uses the concept of inner and outer probability measures (see, e.g. [92, Chapter 1]).

Definition 2.1.2 (Uniform large deviation principle) *Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and let $(W_{n,\rho})_{n\geq 1,\, \rho\in\mathcal{H}_n}$ be a sequence of collections of maps from Ω to E. Let J be a rate function on $(E, \mathfrak{D})$. Let $(v_{n,\rho})_{n\geq 1,\, \rho\in\mathcal{H}_n}$ be a sequence of collections of positive real numbers such that*

$$\lim_{n\to\infty} \inf_{\rho\in\mathcal{H}_n} v_{n,\rho} = \infty.$$

We say that $(W_{n,\rho})_{n\geq 1,\, \rho\in\mathcal{H}_n}$ satisfies the uniform large deviation principle (ULDP) for $(v_{n,\rho})_{n\geq 1,\, \rho\in\mathcal{H}_n}$ and J, when :

— For all F closed in $(E, \mathfrak{D})$ we have

$$\overline{\lim_{n\to\infty}} \sup_{\rho\in\mathcal{H}_n} \frac{1}{v_{n,\rho}} \log\Bigg(\mathbb{P}^*\Big(W_{n,\rho} \in F\Big)\Bigg) \leq -J(F).$$

— For all O open in $(E, \mathfrak{D})$ we have

$$\underline{\lim_{n\to\infty}} \inf_{\rho\in\mathcal{H}_n} \frac{1}{v_{n,\rho}} \log\Bigg(\mathbb{P}_*\Big(W_{n,\rho} \in O\Big)\Bigg) \geq -J(O).$$

When the $\mathcal{H}_n$ are singletons, we will rather speak of a large deviation principle (LDP) for a *sequence* $(W_n)_{n\geq 1}$.

The conclusion of my early works was as follows. For $\psi \in \mathbb{R}^T$, $p \geq 1$ and $\vec{t} = (t_1, \dots, t_p)$

$$\Pi_{\vec{t}}(\psi) := \big(\psi(t_1), \dots, \psi(t_p)\big). \tag{2.2}$$

Proposition 2.1.1 (Varron, derived from Arcones, 2004) *Take* $(E, \mathfrak{D})$ *as the Banach space* $(\ell^\infty(T), ||\cdot||_T)$. *Assume that*

1. *(Finite dimensional ULDP) For each* $p \geq 1$ *and* $\vec{t} = (t_1, \dots, t_p) \in T^p$, *the sequence of collections* $\big(\Pi_{\vec{t}}(W_{n,\rho})\big)_{n\geq 1,\, \rho\in\mathcal{H}_n}$ *is Borel measurable, and satisfies the ULDP in* $\mathbb{R}^p$, *for* $(v_{n,\rho})_{n\geq 1,\, \rho\in\mathcal{H}_n}$ *and for a rate function* $J_{\vec{t}}$.
2. *(Asymptotic equicontinuity) For each* $\epsilon > 0$, *there exists a finite partition* $\mathcal{T}$ *of* T *such that, for each* $T' \in \mathcal{T}$, *we have*

$$\overline{\lim_{n\to\infty}} \sup_{\rho\in\mathcal{H}_n} \frac{1}{v_{n,\rho}} \log\left(\mathbb{P}^*\left(\sup_{(t_1,t_2)\in T'^2}\Big|W_{n,\rho}(t_1) - W_{n,\rho}(t_2)\Big| \geq \epsilon\right)\right) \leq -1/\epsilon.$$

Then $(W_{n,\rho})_{n\geq 1,\, \rho\in\mathcal{H}_n}$ *satisfies the uniform large deviation principle (ULDP) for* $(v_{n,\rho})_{n\geq 1,\, \rho\in\mathcal{H}_n}$ *and the rate function*

$$J:\ \psi \to \sup_{p\geq 1,\, \vec{t}\in T^p} J_{\vec{t}}\Big(\Pi_{\vec{t}}(\psi)\Big).$$

Before emphasizing the usefulness of such a criterion for the local empirical process, I will briefly open a parenthesis to a work I made in large deviation theory [97, 106], and which has his foundations in the works of Arcones for $\ell^\infty(T)$.

2.1.2 Large deviations in spaces having the bounded approximation property

When $(E, \mathfrak{D})$ is a Banach space admitting a Schauder basis (or more generally, a Schauder decomposable space), Suquet [86] proved a tightness criterion for sequences of Borel random variables that involves finite dimensional projections on Schauder basis. Following his ideas, and having in mind the similarity between large deviations and weak convergence in $\big(\ell^\infty(T), ||\cdot||_T\big)$, I investigated the possibility of establishing a similar criterion, for large deviations in such spaces. This was achieved by making use of the concept of LD-tightness introduced by Lynch an Sethuraman [63], which is an appropriate tool when $(E, \mathfrak{D})$ is Polish. The next step was then to extend those results to the more general case of spaces $(E, ||\cdot||)$ having the *bounded approximation property* (BAP spaces), for which the definition is as follows.

Definition 2.1.3 (BAP spaces) *Let* $\big(I_\alpha\big)_{\alpha\in\Upsilon}$ *be a net of finite rank linear operators from* E *to* E. *The space* $(E, ||\cdot||)$ *is said to satisfy the bounded approximation property (BAP for* *short) with the net* $\big(I_\alpha\big)_{\alpha\in\Upsilon}$ *when it is a Banach space, and when :*

$$\lim_{\alpha\in\Upsilon}\sup_{e\in K} \| I_\alpha(e)-e \| \quad \textit{for each compact } K,$$
$$\sup_{\alpha\in\Upsilon} ||| I_\alpha ||| < \infty,$$

where $\lim_{\alpha\in\Upsilon}$ *stands for the limit along the net* Υ*, and where :* $||| L ||| := \sup\{| L(x) |, x \in E, \| x \|= 1\}$ *for a linear operator* L *in* E*.*

Note that this framework encompasses $\big(\ell^\infty(T), \|\cdot\|_T\big)$, as well as spaces having a Schauder basis.
Any Banach space $(E, \|\cdot\|)$ can be identified to a closed subspace of $\big(\ell^\infty(T), \|\cdot\|_T\big)$, by taking T as the unit ball of its topological dual. In empirical processes theory, such an identification can lead to wrong tracks, as it is for example the case for the central limit theorem in Banach spaces (see, e.g., [92, p. 92]). The result of my works, however, is that the above-mentioned identification turns out to be an efficient way to extend the results of Arcones to BAP spaces.

Theorem 1 (Varron, 2014 [106]) *Let* $\big(W_n\big)_{n\geq 1}$ *be a sequence of maps from* Ω *to* E*. Consider the following conditions :*

- $(B1)$ *For each* $\alpha\in\Upsilon$*, the sequence* $\big(I_\alpha(W_n)\big)_{n\geq 1}$ *is Borel measurable, and satisfies the LDP in* $(E, \|\cdot\|)$ *for* v_n *and the rate function* J_α*.*
- $(B2)$ *For each* $\epsilon>0$ *we have*

$$\lim_{\alpha\in\Upsilon}\overline{\lim_{n\to\infty}}\ \frac{1}{v_n}\log\mathbb{P}^*\Big(\| W_n - I_\alpha(W_n) \| > \epsilon\Big) \leq -1/\epsilon.$$

- $(b1)$ *The sequence* $(W_n)_{n\geq 1}$ *satisfies the LDP for the rate function* J*.*

Then $(B1)$*,* $(B2)$ *together imply* $(b1)$*, with* J *defined as*

$$J(x) := \sup_{\alpha\in\Upsilon} J_\alpha\big(I_\alpha(x)\big),\ x\in E.$$

Moreover, $(b1)$ *implies both* $(B1)$ *and* $(B2)$ *with*

$$J_\alpha(x) := \inf\big\{J(y),\ I_\alpha(y)=x\big\}.$$

A by product of this result was a functional limit law, in Hölderian topologies, for smoothed versions of the uniform empirical process on the real line, by convolutions with $K(h_n^{-1}\cdot)$, where K is Hölder (see, [106, Theorem 2.4]).

2.1.3 A ULDP for the local empirical process

The extension of the works of Arcones did open several opportunities for my works on functional limit laws for the local empirical process, for which I recall the definition for $n \geq$

$1, z \in \mathbb{R}^d,\ h > 0$, and $g \in \mathcal{G}$:

$$T_n(g,h,z) := \sum_{i=1}^{n} g_{h,z}(Z_i) - \mathbb{E}\Big(g_{h,z}(Z_i)\Big), \text{ where}$$
$$g_{h,z}(\mathbf{z}) := g\Big(h^{-1}\big(\mathbf{z}-z\big)\Big),\ \mathbf{z} \in \mathbb{R}^d.$$

Our assumption on the law of Z_1 is as follows (as indicated in the introduction).

(Hf) Z_1 admits a Lebesgue density on an open set $\mathfrak{O}$ for which there exists a version f that is continuous and bounded away from 0 and ∞ on $\mathfrak{O}$.

We will then consider a compact set $H \subset \mathfrak{O}$. Such a regularity condition as (Hf) allows to have in mind the heuristic (λ standing for the Lebesgue measure)

$$\mathbb{P}\Big(Z_1 \in z + hJ\Big) \sim \lambda(J)h^d f(z), \tag{2.3}$$

for bounded J, as $h \to 0$, uniformly in $z \in H$.
The main advantage of Proposition 2.1.1 is that, taking $\mathcal{H}_n := H \times [h_n, \mathfrak{h}_n]$, it allows to simultaneously control probabilities of rare events of $T_n(\cdot,h,z),\ z \in H,\ h \in [h_n, \mathfrak{h}_n]$, even if the rate of converge to depends on h (for example, it can be of order $h^{J(F)}$ for closed F). Each article I wrote or co-wrote on standard functional limit laws involves the proof of an ad-hoc ULDP, which then is used as the very base of asymptotic probability calculus. At the price of very slightly strengthening the assumptions made in those articles, it is possible to unify all of these ULDP into a single one, which I will present and prove here. Before stating our assumptions on $\mathcal{G}$ we will recall some definitions that are usual in empirical processes theory.

Definition 2.1.4 (Covering numbers) *Let (E,d) be a metric space, and let $E_0 \subset E$ be an arbitrary set. For all $\epsilon > 0$, we denote by $N(\epsilon, E_0, d)$ the minimal (possibly infinite) number of closed d-balls of radius ϵ needed to cover E.*

Definition 2.1.5 (Pointwise separable class) *A class $\mathcal{G}$ of real functions on a set $\mathfrak{X}$ is said to be pointwise separable when there exists a countable subclass $\mathcal{G}_0$ such that, for each $g \in \mathcal{G}$ there exists a sequence $(g_n)_{n\geq 1} \in \mathcal{G}_0^{\mathbb{N}}$ such that $g_n(x) \to g(x)$, as $n \to \infty$, for all $x \in \mathfrak{X}$.*

We will consider a class of functions $\mathcal{G}$ satisfying :

$(Pointw.\ sep.)$	The class $\mathcal{G}$ is pointwise separable on $\mathbb{R}^d$,
$(Bounded)$	There exists $M > 0$ such that $\mid g \mid \leq M$ for each $g \in \mathcal{G}$,
$(Support)$	There exists $M > 0$ such that $g(z) = 0$ for each $g \in \mathcal{G}$ and $z \notin [-M,M]^d$,
$(Unif.\ entropy)$	We have $\int_0^\infty \sup_{Q \text{ probab.}} \sqrt{\log N\big(\epsilon, \mathcal{G}, \|\cdot\|_{Q,2}\big)}\, d\epsilon < \infty$,

where, for any measure Q and $r > 0$, the symbol $|| \cdot ||_{Q,r}$ denotes the corresponding $L^r(Q)$-norm. Note that, when both assumptions $(Bounded)$ and $(Support)$ are made simultaneously, I will implicitly use a parameter $M > 0$ fulfilling both of them.
A combination of assumptions $(Unif.\ entropy)$ and $(Support)$ is largely sufficient to ensure the existence of $L^2(\lambda)$ isonormal Gaussian process $\mathcal{W}_\mathcal{G}$ on $\mathcal{G}$, namely a centered gaussian process having covariance function

$$\mathrm{Cov}\big(\mathcal{W}_\mathcal{G}(g_1), \mathcal{W}_\mathcal{G}(g_2)\big) := \int_{[-M,M]^d} g_1 g_2 d\lambda,$$

for which we can consider the associated rate function

$$J_\mathcal{G} : \psi \to \inf \Big\{ || \mathfrak{g} ||_{\lambda,2}^2,\ g \text{ Borel},\ \psi(\cdot) \equiv g \to \int_{\mathbb{R}^d} \mathfrak{g} g d\lambda \Big\}, \tag{2.4}$$

with the convention $\inf \emptyset = \infty$. It is well known that $J_\mathcal{G}$ rules the large deviation properties of $\mathcal{W}_\mathcal{G}$ (see [28, p. 85]).
In a metric space (E, d) we shall write

$$A^\epsilon := \Big\{ x \in E,\ d(x, A) < \epsilon \Big\},\ A \subset E,\ \epsilon > 0, \text{ with} \tag{2.5}$$
$$d(x, A) := \inf_{y \in A} d(x, y).$$

Our general ULDP can be stated as follows. An indexation in k has been chosen (instead of n), because this ULDP will be systematically used for subsequences. Note that, form now on, unless otherwise specified, each vector space $\mathbb{R}^p$ will be endowed with its Euclidian norm $|| \cdot ||_p$.

Proposition 2.1.2 (Varron, 2008-2014, from [95, 98, 108]) *Let $\tilde{n}_k$ be a strictly increasing sequence of integers, and let $\tilde{h}_k, \tilde{\mathfrak{h}}_k, \epsilon_k$ be three sequences of non negative real numbers such that $\tilde{h}_k \le \tilde{\mathfrak{h}}_k$, $\tilde{\mathfrak{h}}_k \to 0$ and $\epsilon_k \to 0$. Let $(v_k(h))_{k \ge 1, h \in [\tilde{h}_k, \tilde{\mathfrak{h}}_k]}$ be a sequence of collections of nonnegative real numbers such that*

$$\lim_{k \to \infty} \inf_{h \in [\tilde{h}_k, \tilde{\mathfrak{h}}_k]} v_k(h) = \infty, \tag{2.6}$$

$$\mathfrak{r}_k := \inf_{h \in [\tilde{h}_k, \tilde{\mathfrak{h}}_k]} \frac{\tilde{n}_k h^d}{v_k(h)} \to \infty. \tag{2.7}$$

Let $\mathcal{G}$ be a class of functions fulfilling $(Pointw.\ sep.)$, $(Bounded)$, $(Support)$, and $(Unif.\ entropy)$. Then, under (Hf), the following assertions hold :

$$\forall F \textit{ closed in } \big(\ell^\infty(\mathcal{G}), || \cdot ||_\mathcal{G}\big),$$
$$\overline{\lim_{k \to \infty}} \sup_{\substack{z \in H,\ ||z' - z||_d \le \epsilon_k \\ h \in [\tilde{h}_k, \tilde{\mathfrak{h}}_k]}} \frac{1}{v_k(h)} \log\left(\mathbb{P}^*\left(\frac{T_{\tilde{n}_k}(\cdot, h, z')}{\sqrt{2 f(z) \tilde{n}_k h^d v_k(h)}} \in F\right)\right) \le -J_\mathcal{G}(F);$$

$$\forall O\, open\ in\ \big(\ell^\infty(\mathcal{G}), ||\cdot||_{\mathcal{G}}\ \big),$$
$$\varliminf_{k\to\infty} \inf_{\substack{z\in H,\ ||z'-z||_d\le\epsilon_k \\ h\in[\tilde{h}_k,\tilde{\mathfrak{h}}_k]}} \frac{1}{v_k(h)} \log\left(\mathbb{P}_*\left(\frac{T_{\tilde{n}_k}(\cdot,h,z')}{\sqrt{2f(z)\tilde{n}_k h^d v_k(h)}} \in O\right)\right) \ge -J_{\mathcal{G}}(O).$$

Some comments on the assumptions

In the next paragraphs, I will decompose the proof into several steps, with a particular emphasis on the role of each of the assumptions made upon $\mathcal{G}$, the bandwidths, and Z_1. This complete proof can be skipped at first reading, (in that case, the reader can go directly to §2.2) but it is nevertheless recommended to have in mind the following sum up of the different roles of the assumptions :

- The only role of (*Pointw. sep.*) is to make the suprema of empirical processes measurable.
- Assumption $\tilde{\mathfrak{h}}_k \to 0$ in conjunction with (*Support*) and (Hf) makes the covariance structure of the local empirical processes converge to the covariance structure of $\mathcal{W}_{\mathcal{G}}$ uniformly in $z \in H$ (see Lemma 2.1.1 in the sequel).
- Assumption (2.7) has to be understood as follows : by the heuristic (2.3), and by (*Support*), the average number of Z_i that play a role in the expression of $T_{\tilde{n}_k}(\cdot, h, z)$ is roughly $\tilde{n}_k h^d f(z)\lambda([-M,M]^d)$ This average number can be considered as the "true sample size". Hence $v_k(h)$ is required to be large in front of this "sample size", in order to consider deviations sufficiently large to observe Gaussian tails.
- Those Gaussian tails are determined by a concentration inequality, initiated by Talagrand (see Fact 2.1.2). Assumption (*Bounded*) plays its role for the use of that concentration inequality. It also plays a role to use an approximation result due to Zaitsev (see Fact 2.1.1) between finite dimensional marginals of the $T_{\tilde{n}_k}(\cdot, h, z)$ and those of $\mathcal{W}_{\mathcal{G}}$.
- Such a concentration inequality holds around the expectations of suprema of empirical processes. Assumption ($Unif\ entropy$) is essential to control these expectations by the usual chaining argument for Rademacher processes (see also §2.1.5 for a discussion about bracketing conditions).
- Allowing z' to vary in small balls around z is handled by continuity of f. This additional technicality can be ignored at first reading.

Proof : Since $\mathfrak{D}$ is open, $\epsilon_k \to 0$, and by (*Support*), we can assume without loss of generality that

$$\forall k\ge 1,\ h\in[\tilde{h}_k,\tilde{\mathfrak{h}}_k],\ ||\ z'-z\ ||_d\le \epsilon_k,\ \boldsymbol{z}\notin\mathfrak{D},\ g_{h,z'}(\boldsymbol{z})=0. \tag{2.8}$$

Finite dimensional ULDP by Zaitsev's inequality

Following Proposition 2.1.1 we will first prove point 1 of that proposition. To that end, we will use the powerful approximation result of Zaitsev [111] for, e.g., sums of independent and bounded random vectors. For $p\ge 1$, and for two probability measures P and Q on $\mathbb{R}^p$, we shall write :

$$\pi(P,Q,\epsilon) := \sup_{A\ Borel} \max\big\{P(A)-Q(A^\epsilon), Q(A)-P(A^\epsilon)\big\}, \text{for } \epsilon>0.$$

Fact 2.1.1 (Consequence of Zaitsev [110]) *There exists universal constants c_1, c_2 such that, for any $p \geq 1$, $n \geq 1$, $\tau > 0$, and any sequence $(U_i)_{i=1,\dots,n}$ of centered, independent random vectors in $\mathbb{R}^p$ for which each coordinate is almost surely bounded by τ, we have*

$$\forall \epsilon>0,\ \pi(P,Q,\epsilon) \leq c_1 p^2 \exp\Big(-\frac{\epsilon}{c_2 p^2 \tau}\Big), \tag{2.9}$$

where P and Q are respectively the law of $\sum_{i=1}^n U_i$ and of its Gaussian analogue.

Assumption (*Bounded*) plays its first role here, since it entails, for each $k \geq 1$, $z \in H$, $|| z'-z ||_d \leq \epsilon_k$ and $h \in [\tilde{h}_k, \tilde{\mathfrak{h}}_k]$, that the random vectors

$$U_{i,k,h,z,z'} := \left(\frac{g_{1_{h,z'}}(Z_i) - \mathbb{E}\Big(g_{1_{h,z'}}(Z_i)\Big)}{\sqrt{2f(z)\tilde{n}_k h^d v_k(h)}}, \dots, \frac{g_{p_{h,z'}}(Z_i) - \mathbb{E}\Big(g_{p_{h,z'}}(Z_i)\Big)}{\sqrt{2f(z)\tilde{n}_k h^d v_k(h)}}\right)_{i=1\dots,\tilde{n}_k}$$

have each of their coordinates almost surely bounded by

$$\tau_{k,h,z,z'} := \frac{2M}{\sqrt{2f(z)\tilde{n}_k h^d v_k(h)}},$$

from where, for arbitrary $\epsilon > 0$, taking the notations of Fact 2.1.1 :

$$\begin{aligned}\pi\Big(P_{k,h,z,z'}, Q_{k,h,z,z'}, \epsilon\Big) \leq & c_1 p^2 \exp\Big(-\frac{\epsilon\sqrt{2f(z)\tilde{n}_k h^d v_k(h)}}{2Mc_2p^2}\Big),\\ \leq & c_1 p^2 \exp\Big(-\frac{\epsilon\sqrt{2f(z)}\mathfrak{r}_k v_k(h)}{2Mc_2p^2}\Big),\end{aligned}$$

where $\mathfrak{r}_k$ is defined in (2.7), and where $P_{k,h,z,z'}$ and $Q_{k,h,z,z'}$ respectively denote the law of $\sum_{i=1}^{\tilde{n}_k} U_{i,k,h,z,z'}$ and of its Gaussian analogue. Hence, since both (2.6) and (2.7) hold, we have, for each $\epsilon > 0$:

$$\lim_{k\to\infty} \sup_{\substack{z\in H,\ ||z'-z||_d\leq\epsilon_k\\ h\in[\tilde{h}_k,\tilde{\mathfrak{h}}_k]}} \frac{1}{v_k(h)} \log\Big(\pi\big(P_{k,h,z,z'}, Q_{k,h,z,z'}, \epsilon\big)\Big) = -\infty.$$

By standard topological arguments, (2.10) provides sufficiently fast approximations to prove that the sought ULDP for $\big(P_{k,h,z,z'}\big)_{z\in H,\ ||z'-z||_d\leq\epsilon_k,\ h\in[\tilde{h}_k,\tilde{\mathfrak{h}}_k]}$ is the same as that of $\big(Q_{k,h,z,z'}\big)_{z\in H,\ ||z'-z||_d\leq\epsilon_k,\ h\in[\tilde{h}_k}$
Now, each $Q_{k,h,z,z'}$ can be represented as the law of $(2v_k(h))^{-1/2}\Sigma_{h,z,z'}^{1/2}\mathcal{Z}$, where $\mathcal{Z}$ is standard

normal on $\mathbb{R}^p$, and where

$$\Sigma_{h,z,z'}[j,j'] := f(z)^{-1}h^{-d}\mathrm{Cov}\Big(g_{j\,h,z'}(Z_1), g_{j'\,h,z'}(Z_1)\Big), (j,j') \in \{1,\dots,p\}^2.$$

Now write $\Sigma[j,j'] := \int g_j, g_{j'} d\lambda$ for $(j,j') \in \{1,\dots,p\}$. By standard analysis for Gaussian distributions, if we prove that

$$\lim_{k\to\infty} \sup_{\substack{z\in H,\ ||z'-z||_d\le\epsilon_k \\ h\in[\tilde{h}_k,\tilde{\mathfrak{h}}_k]}} ||\ \Sigma_{h,z,z'} - \Sigma\ ||_{p^2} \to 0, \tag{2.10}$$

then $\big(Q_{k,h,z,z'}\big)_{z\in H,\ ||z'-z||_d\le\epsilon_k,\ h\in[\tilde{h}_k,\tilde{\mathfrak{h}}_k]}$ will satisfy the ULDP for the rate function $J_{g_1,\dots,g_p}$ which is defined as the quadratic form on $\mathbb{R}^p$ associated to Σ. Now (2.10) is a consequence of the following crucial remark.

Lemma 2.1.1 *We have*

$$\lim_{k\to\infty} \sup_{\substack{z\in H,\ ||z'-z||_d\le\epsilon_k \\ h\in[\tilde{h}_k,\tilde{\mathfrak{h}}_k]}} \sup_{(g,g')\in\mathcal{G}^2} \left|\frac{1}{f(z)h^d}\mathrm{Cov}\Big(g_{h,z'}(Z_1), g'_{h,z'}(Z_1)\Big) - \int gg'd\lambda\right| = 0,$$

$$\lim_{k\to\infty} \sup_{\substack{z\in H,\ ||z'-z||_d\le\epsilon_k \\ h\in[\tilde{h}_k,\tilde{\mathfrak{h}}_k]}} \sup_{(g,g')\in\mathcal{G}^2} \left|\frac{1}{f(z)h^d}\mathbb{E}\Big(g_{h,z'}(Z_1)g'_{h,z'}(Z_1)\Big) - \int gg'd\lambda\right| = 0.$$

Proof : We use the change of variable $u := z+hv$ in the next calculus for arbitrary $(g,g') \in \mathcal{G}^2$.

$$\begin{aligned}
&\mathrm{Cov}\Big(g_{h,z'}(Z_1), g'_{h,z'}(Z_1)\Big) \\
=& \int_{\mathfrak{O}} g_{h,z'}(u)g'_{h,z'}(u)f(u)du - \int_{\mathfrak{O}} g_{h,z'}(u)f(u)du \times \int_{\mathfrak{O}} g'_{h,z'}(u)f(u)du, \text{ by (2.8)} \\
=& h^d f(z)\int_{[-M,M]^d} g(v)g'(v)\frac{f(z'+hv)}{f(z)}dv \\
&- h^d f(z)^2\int_{[-M,M]^d} g(v)\frac{f(z'+hv)}{f(z)}dv \times h^d\int_{[-M,M]^d} g'(v)\frac{f(z'+hv)}{f(z)}dv,
\end{aligned} \tag{2.11}$$

where assumption (*Support*) plays its role here, since it makes possible to integrate on $[-M,M]^d$. Then assumption (Hf) allows the following crucial argument

$$\lim_{k\to\infty} \sup_{\substack{z\in H,\ ||z'-z||_d\le\epsilon_k \\ h\in[\tilde{h}_k,\tilde{\mathfrak{h}}_k]}} \sup_{v\in[-M,M]^d} \left|\frac{f(z'+hv)}{f(z)} - 1\right| = 0, \tag{2.12}$$

and (2.11) is always negligible in front of the first term because $h \le \tilde{\mathfrak{h}}_k \to 0$. $\square$

Local representation and concentration inequalities

A well understood heuristic about the local uniform empirical process is that, when looking at the spatial repartition of a uniform sample restricted to a small interval $[t, t+h_n]$, with h_n "small", the quantity of observed points is of order nh_n. This is roughly speaking the true observed sample size. For the local empirical process, the same heuristic is appealing, because, by assumption (*Support*), the support of any function $g_{h,z}$ is contained in

$$A(h,z) := z + [-Mh, Mh]^d,\ z \in \mathfrak{O},\ h > 0. \tag{2.13}$$

Hence, the observations falling outside this small ball do not play any role in the value of $T_{\tilde{n}_k}(\cdot, h, z)$. Einmahl and Mason [41] gave a form to this heuristic by noticing the following equality in distributions (fo which we will use the symbol $=_d$) holds for collections of $\mathcal{G}$-indexed processes, for any $k \geq 1,\ h > 0$ and $z' \in H^{\epsilon_k}$,

$$\Big(g_{h,z'}(Z_1), \ldots, g_{h,z'}(Z_{\tilde{n}_k})\Big)_{g\in\mathcal{G}} =_d \Big(\tau_1^{(h,z')} g\big(Y_1^{(h,z')}\big), \ldots, \tau_{\tilde{n}_k}^{(h,z')} g\big(Y_{\tilde{n}_k}^{(h,z')}\big)\Big)_{g\in\mathcal{G}}, \tag{2.14}$$

where

— $\big(Y^{(h,z')}\big)_{i\geq 1}$ is an i.i.d. sequence having the distribution

$$\mathbf{P}_0^{(h,z')}(\cdot) := \mathbb{P}\Big(h^{-1}(Z_1 - z') \in \cdot \Big| Z_1 \in A(h,z')\Big). \tag{2.15}$$

— $\big(\tau_i^{(h,z')}\big)_{i\geq 1}$ is an i.i.d. Bernoulli sequence, independent of $\big(Y_i^{(h,z')}\big)_{i\geq 1}$, and having expectation $a(h,z')$, with

$$a(h,z) := \mathbb{P}\big(Z_1 \in A_{h,z}\big),\ z \in \mathfrak{O},\ h > 0. \tag{2.16}$$

Assumption (*Support*) plays an essential role for that representation, for which a consequence intensively used by Einmahl and Mason [41] is that

$$T_{\tilde{n}_k}(\cdot, h, z') :=_d \left(\sum_{i=1}^{\mathbf{b}_{k,h,z'}} g\big(Y_i^{(h,z')}\big) - \mathbb{E}\Big(g\big(Y_i^{(h,z')}\big)\Big) + \pi_{k,h,z'}\, \mathbb{E}\Big(g\big(Y_i^{(h,z')}\big)\Big)\right)_{g\in\mathcal{G}}, \tag{2.17}$$

where

— $\mathbf{b}_{k,h,z'}$ is a binomial $(\tilde{n}_k, a(h,z'))$ random variable, independent of $(Y_i^{(h,z')})_{i\geq 1}$, with

— $\pi_{k,h,z'}$ is defined as $(\mathbf{b}_{k,h,z'} - \tilde{n}_k a(h,z'))$

Hence, at the price of a randomness of the sample size, of modifying the law of the sample, and up to a remaining term that can be controlled by usual inequalities for binomial distributions, the local empirical process can be represented as a general empirical process indexed by the same class of functions $\mathcal{G}$.

The symmetrizations techniques are very useful in empirical processes theory. Another consequence of (2.14) is the following representation for the symmetrized versions of the local empirical process

$$\left(\sum_{i=1}^{\tilde{n}_k}\sigma_i g_{h,z'}(Z_i)\right)_{g\in\mathcal{G}} =_d \left(\sum_{i=1}^{\mathbf{b}_{k,h,z'}}\sigma_i g\big(Y_i^{(h,z')}\big)\right)_{g\in\mathcal{G}}, \tag{2.18}$$

where, on each side of the representation, $(\sigma_i)_{i\geq 1}$ denotes an i.i.d Rademacher sequence ($\mathbb{P}\big(\sigma_1 = 1\big) = \mathbb{P}\big(\sigma_1 = -1\big) = 1/2$).
Now let us go back to our initial ULDP problem. Straightforward analysis (see, e.g., [5, Lemma 4.1]) shows that (recall (2.2)) :

$$\forall\psi\in\ell^\infty(\mathcal{G}),\quad \sup_{p\geq 1,\ \vec{g}=(g_1,\ldots,g_p)\in\mathcal{G}^p} J_{\vec{g}}\big(\Pi_{\vec{g}}(\psi)\big) = J_{\mathcal{G}}(\psi).$$

Hence, it only remains to prove the (*Asymptotic equicontinuity*) part of Proposition 2.1.1. This has to be done through the uniform control of the oscillations of the $T_{\tilde{n}_k}(\cdot, h, z)$ on the elements of a suitable finite partition of $T = \mathcal{G}$. By (*Unif. entropy*) and (*Support*), $\mathcal{G}$ is totally bounded for the norm $||\cdot||_{\lambda,2}$. Hence it is sufficient to prove that, for any $\epsilon > 0$, we have

$$\lim_{\delta\to 0}\lim_{k\to\infty}\sup_{\substack{z\in H,\ ||z'-z||_d\leq\epsilon_k\\ h\in[\tilde{h}_k,\tilde{\mathfrak{h}}_k]}}\frac{1}{v_k(h)}\log\left(\mathbb{P}\left(\sup_{\substack{(g,g')\in\mathcal{G}^2,\\ ||g-g'||_{\lambda,2}\leq\delta}}\frac{\Big|T_{\tilde{n}_k}(g',h,z') - T_{\tilde{n}_k}(g,h,z')\Big|}{\sqrt{2f(z)\tilde{n}_k h^d v_k(h)}} > \epsilon\right)\right) = -\infty.$$

Note that the preceding inequality involves probabilities (instead of outer probabilities) since (*Pointw. sep.*) is satisfied.
Clearly, (2.19) involves suprema of empirical processes, for which powerful concentration inequalities around their expectations have been established, first by Talagrand [88], and then refined by Bousquet [13] and Klein and Rio [55]. On the other hand, assumptions (*Bounded*) and (*Unif. entropy*), combined with representation (2.18) will provide sharp enough bounds for the expectations of these suprema. The following concentration inequality is due to Talagrand [87], and then improved by Einmahl and Mason to a maximal inequality [42].

Fact 2.1.2 (Talagrand, Einmahl, Mason [87, 42]) *Let $\mathcal{G}$ be a class of measurable real functions on a measurable space $(\mathfrak{X},\mathcal{X})$, fulfilling (Pointw. sep.) and (Bounded) for some $0 < M < \infty$. Let $(Y_i)_{i\geq 1}$ be an i.i.d. sequence of random variables on $\big(\mathfrak{X},\mathcal{X}\big)$ and $(\sigma_i)_{i\geq 1}$ an i.i.d Rademacher sequence independent of the first sequence. Then, for suitable universal constants $A_1, A_2 > 0$ we have, for each $t > 0$ and $n \geq 1$:*

$$\begin{aligned}&\mathbb{P}\left(\max_{m=1,\ldots,n}\sup_{g\in\mathcal{G}}\Big|\sum_{i=1}^m g(Y_i) - \mathbb{E}\Big(g(Y_i)\Big)\Big| \geq A_1\left(\mathbb{E}\Big(\sup_{g\in\mathcal{G}}\big|\sum_{i=1}^n\sigma_i g(Y_i)\ \big|\Big) + t\right)\right)\\ \leq& 2\left(\exp\Big(-\frac{A_2t^2}{n\sigma_{\mathcal{G}}^2}\Big) + \exp\Big(-\frac{A_2 t}{M}\Big)\right),\ \textit{with}\\ \sigma_{\mathcal{G}}^2 :=& \sup_{g\in\mathcal{G}}\mathrm{Var}\Big(g(Y_1)\Big).\end{aligned}$$

That powerful inequality has to be accompanied by a control of first moment. In our particular setup, the local representation (2.18), combined with Dudley's chaining argument for subgaussian processes leads to a simple bound for those expectations. This is the aim of the following lemma.

Lemma 2.1.2 (Varron, from [108]) *There exists a universal constant C_0 such that, for all $k \geq 1$, $z \in \mathfrak{O}$, $h > 0$, and for all class of function $\mathcal{G}$ fulfilling (Pointw. sep.), we have*

$$\mathbb{E}\Big(\sup_{g\in\mathcal{G}}\Big|\sum_{i=1}^{\tilde{n}_k}\sigma_i g_{h,z}(Z_i)\Big|\Big) \leq C_0\int_0^\infty \sup_{Q\ probab.}\sqrt{\log N\big(\epsilon,\mathcal{G},||\cdot||_{Q,2}\big)}d\epsilon \times \sqrt{\tilde{n}_k a(h,z)}. \tag{2.19}$$

Proof : The proof uses arguments that are very similar to those of Dony and Einmahl [31]. Assume without loss of generality that the RHS of (2.19) is finite. We first remark that, by subgaussianity of Rademacher processes (see, e.g., [92, p. 127-128]), we have for all $n \geq 1$, $z \in \mathfrak{O}$, $h > 0$:

$$\begin{aligned}\mathbb{E}\Big(\sup_{g\in\mathcal{G}}\Big|\sum_{i=1}^{n}\sigma_i g\big(Y_i^{(h,z)}\big)\Big|\Big) =&\mathbb{E}_{Y_1^{(h,z)},\ldots,Y_n^{(h,z)}}\mathbb{E}_{\sigma_1,\ldots,\sigma_n}\Big(\sup_{g\in\mathcal{G}}\Big|\sum_{i=1}^{n}\sigma_i g\big(Y_i^{(h,z)}\big)\Big|\Big)\\ \leq&\mathbb{E}_{Y_1^{(h,z)},\ldots,Y_n^{(h,z)}}\Big(C_0\sqrt{n}\int_0^\infty\sqrt{\log N\big(\epsilon,\mathcal{G},||\cdot||_{\mathbf{P}_n^{(h,z)},2}\big)}d\epsilon\Big),\\ \leq&C_0\sqrt{n}\int_0^\infty\sup_{Q\text{ probab.}}\sqrt{\log N\big(\epsilon,\mathcal{G},||\cdot||_{Q,2}\big)}d\epsilon,\end{aligned}$$

where $\mathbf{P}_n^{(h,z)} = n^{-1}\sum_{i=1}^{n}\delta_{Y_i^{(h,z)}}$, and where C_0 denote a universal constant. Next, using representation (2.18) we have, for all $k \geq 1$, $z \in \mathfrak{O}$, $h > 0$:

$$\begin{aligned}&\mathbb{E}\Big(\sup_{g\in\mathcal{G}}\Big|\sum_{i=1}^{\tilde{n}_k}\sigma_i g_{h,z}(Z_i)\Big|\Big)\\ =&\mathbb{E}\Big(\sup_{g\in\mathcal{G}}\Big|\sum_{i=1}^{\mathbf{b}_{k,h,z}}\sigma_i g(Y_i^{h,z})\Big|\Big)\\ \leq&\sqrt{\tilde{n}_k a(h,z)}C_0\int_0^\infty\sup_{Q\text{ probab.}}\sqrt{\log N\big(\epsilon,\mathcal{G},||\cdot||_{Q,2}\big)}d\epsilon.\Box\end{aligned} \tag{2.20}$$

The next Proposition is a combination of the two preceding ones. It is a refinement of existing arguments in the literature [42, 43, 65]. Its contribution consists in giving precise threshold conditions to give a sense to the heuristics "h has to be small and nh^d has to be large to observe Gaussian tails". I also tried properly determine which parameters of the framework (the law of Z_1 and the structural assumptions on $\mathcal{G}$) do determine those thresholds.

Proposition 2.1.3 (Varron, from [108]) *Let $(Z_i)_{i\geq 1}$ be an i.i.d. sequence for which the law of Z_1 fulfills (Hf) for some open set $\mathfrak{O}$. There exist universal constants $C_0, C_1, C_2 > 0$ such that the following assertion is true : Given a compact set $H \subset \mathfrak{O}$, and given $C_3 < \infty$ and $M > 0$, there exists $h_0 > 0$ and $\epsilon_0 > 0$ depending only on M,f and H such that, for any pointwise separable class of functions $\mathcal{G}$ fulfilling*

$$\forall g \in \mathcal{G}, \mid g \mid \leq M\mathbb{1}_{[-M,M]^d}, \tag{2.21}$$

$$\int_0^\infty \sup_{Q \text{ probab.}} \sqrt{\log N(\epsilon, \mathcal{G}, || \cdot ||_{Q,2})} d\epsilon \leq C_3, \tag{2.22}$$

there exists $K_0 > 0$ such that, for each $h > 0$, $u > 0$, $n \geq 1$, $z \in H$ and $z' \in \mathbb{R}^d$ fulfilling $h \leq h_0$, $u > (C_0C_3)^2 2^{d+1} M^d$, $nh^d > K_0 u$ and $|| z' - z ||_d \leq \epsilon_0$, we have :

$$\mathbb{P}\bigg(\max_{m=1,\dots,n} \sup_{g \in \mathcal{G}} \Big| \sum_{i=1}^m g_{h,z'}(Z_i) - \mathbb{E}\Big(g_{h,z'}(Z_i)\Big) \Big| > C_1 \sqrt{f(z) n h^d u} \bigg) \leq 4 \exp\Big(- \frac{C_2 u}{\Delta_{\mathcal{G}}^2} \Big),$$

where

$$\Delta_{\mathcal{G}}^2 := \sup_{g \in \mathcal{G}} || g ||_{\lambda,2}^2 . \tag{2.23}$$

Proof : Fix $M > 0$. By (Hf) we can choose $h_0 > 0$ and $\epsilon_0 > 0$ such that

$$\sup_{z \in H,\, ||v||_d \leq M h_0,\, ||z'-z||_d \leq \epsilon_0} \frac{f(z'+v)}{f(z)} \leq 2. \tag{2.24}$$

Then, by the same arguments as those used in the proof of Lemma 2.1.1 we have, for any $\mathcal{G}$ fulfilling (2.21) :

$$\sup_{0<h\leq h_0} \sup_{z \in H} \sup_{||z'-z||_d \leq \epsilon_0} \sup_{g \in \mathcal{G}} \frac{1}{f(z)h^d} \mathrm{Var}\big(g_{h,z'}(Z_1)\big) \leq 2\Delta_{\mathcal{G}}^2, \tag{2.25}$$

$$\sup_{0<h\leq h_0} \sup_{z \in H} \sup_{||z'-z||_d \leq h_0} \frac{1}{f(z)h^d} a(h,z') \leq 2(2M)^d. \tag{2.26}$$

Now fix $C_3 > 0$. By Lemma 2.1.2 and by (2.26) we see that, for fixed $n \geq 1$, $h \leq \epsilon_0$, $z \in H$ and $|| z' - z ||_d \leq h_0$, we have :

$$C_0 C_3 \sqrt{n a(h,z')} \leq \sqrt{f(z) n h^d u}, \text{ for all } u \geq (C_0C_3)^2 2^{d+1} M^d. \tag{2.27}$$

Moreover, the class

$$\mathcal{G}(h,z') := \Big\{ g_{h,z'},\ g \in \mathcal{G} \Big\},$$

is uniformly bounded by M, and satisfies $\sigma_{\mathcal{G}(h,z')}^2 \leq 2f(z)h^d\Delta_{\mathcal{G}}^2$, by (2.25). Now, for all $h \leq h_0, u \geq (C_0C_3)^2 2^{d+1} M^d$, $|| z' - z ||_d \leq \epsilon_0$, and $n \geq 1$, inserting (2.27) in Fact 2.1.2

(applied to $\mathcal{G}(h,z')$) leads to :

$$\begin{aligned}
&\mathbb{P}\Bigg(\max_{m=1,\dots,n}\sup_{g\in\mathcal{G}}\Big|\sum_{i=1}^{m} g_{h,z'}(Z_i)-\mathbb{E}\Big(g_{h,z'}(Z_i)\Big)\Big|>2A_1\sqrt{f(z)nh^du}\Bigg)\\
\leq& 2\Bigg[\exp\Bigg(-\frac{A_2f(z)nh^du}{n\sigma^2_{\mathcal{G}(h,z')}}\Bigg)+\exp\Bigg(-\frac{A_2\sqrt{f(z)nh^du}}{M}\Bigg)\Bigg]\\
\leq& 4\exp\Bigg(-\frac{A_2u}{2\Delta^2_{\mathcal{G}}}\Bigg),\text{ as soon as }\frac{nh^d}{u}\geq\frac{M^2}{4\Delta^4_{\mathcal{G}}\inf_{z\in H}f(z)}=:K_0. \qquad (2.28)
\end{aligned}$$

This concludes the proof of Proposition 2.1.3.□

Asymptotic equicontinuity

Recall that we aim to prove that, for any $\epsilon>0$, we have

$$\lim_{\delta\to 0}\lim_{k\to\infty}\sup_{\substack{z\in H,\ ||z'-z||_d\leq\epsilon_k\\ h\in[\tilde{h}_k,\tilde{\mathfrak{h}}_k]}}\frac{1}{v_k(h)}\log\Bigg(\mathbb{P}\Bigg(\sup_{\substack{(g,g')\in\mathcal{G}^2,\\ ||g-g'||_{\lambda,2}\leq\delta}}\frac{\Big|T_{\tilde{n}_k}(g',h,z')-T_{\tilde{n}_k}(g,h,z')\Big|}{\sqrt{2f(z)\tilde{n}_kh^dv_k(h)}}>\epsilon\Bigg)\Bigg)=-\infty. \qquad (2.29)$$

We use the notations of Proposition 2.1.3. First, note that, by (*Bounded*) and (*Unif. entropy*), and by standard covering numbers arguments, all the classes

$$\mathcal{G}'_\delta:=\Big\{g-g',\ ||\ g-g'\ ||_{\lambda,2}\leq\delta\Big\},\ \delta>0, \qquad (2.30)$$

do *simultaneously* satisfy assumptions (2.21) and (2.22) with the *same* constants M and C_3. Since $\tilde{\mathfrak{h}}_k\to 0$, and by (2.6), we have, for all large k, and for all $z\in H,\ ||\ z'-z\ ||_d\leq\epsilon_k,\ h\in[\tilde{h}_k,\tilde{\mathfrak{h}}_k]$:

$$\epsilon_k\leq\epsilon_0,\ h\leq h_0,\ v_k(h)\geq(C_0C_3)^2 2^{d+1}M^d,\ \text{and}\ \ ||\ z'-z\ ||_d\leq h_0,$$

which implies, by Proposition 2.1.3

$$\begin{aligned}
&\mathbb{P}\Bigg(\sup_{(g,g')\in\mathcal{G}^2,\ ||g-g'||_{\lambda,2}\leq\delta}\frac{\Big|T_{\tilde{n}_k}(g',h,z')-T_{\tilde{n}_k}(g,h,z')\Big|}{\sqrt{2f(z)\tilde{n}_kh^dv_k(h)}}>\epsilon\Bigg)\\
=&\mathbb{P}\Bigg(\sup_{g\in\mathcal{G}'_\delta}\Big|\sum_{i=1}^{\tilde{n}_k}g_{h,z}(Z_i)\Big|>\epsilon\sqrt{2f(z)\tilde{n}_kh^dv_k(h)}\Bigg)\\
\leq&4\exp\Bigg(-\frac{2A_2\epsilon^2v_k(h)}{2\Delta^2_{\mathcal{G}'_\delta}}\Bigg).
\end{aligned}$$

Since $\Delta^2_{\mathcal{G}'_\delta}=\delta^2$, this proves (2.29) and hence concludes the proof of Proposition 2.1.2.□

2.1.4 Asymptotic independence of two local empirical processes at very different scales

In this subsection, I will give a more precise meaning to the *between bandwidths asymptotic independence* phenomenon which I evoked in the introduction (see, §1.3.2). When investigating the large deviation properties of two independent copies $(\mathcal{W}^1_{\mathcal{G}},\mathcal{W}^2_{\mathcal{G}})$ of $\mathcal{W}_{\mathcal{G}}$ (recall (2.4)), the appropriate rate function on $\ell^\infty(\mathcal{G})\times\ell^\infty(\mathcal{G})$ is the following :

$$J^{\otimes 2}_{\mathcal{G}}(\psi_1,\psi_2):=J_{\mathcal{G}}(\psi_1)+J_{\mathcal{G}}(\psi_2),\ (\psi_1,\psi_2)\in\ell^\infty(\mathcal{G})\times\ell^\infty(\mathcal{G}), \tag{2.31}$$

where the presence of a summation should not be surprising, as large deviations deal with logarithm of probabilities. The following ULDP is an unpublished extension of a result I published in [96] for the local uniform empirical process.

Proposition 2.1.4 (Varron, 2014, extension of [96]) *Under the assumptions and notations of Proposition 2.1.2, if, in addition, $\tilde h_k=o(\tilde{\mathfrak{h}}_k)$ and $v_k(\tilde h_k)\sim v_k(\tilde{\mathfrak{h}}_k)$ as $k\to\infty$, we have :*

$\forall F$ *closed in* $\big(\ell^\infty(\mathcal{G}),||\cdot||_{\mathcal{G}}\big)\times\big(\ell^\infty(\mathcal{G}),||\cdot||_{\mathcal{G}}\big)$,

$$\overline{\lim_{k\to\infty}}\sup_{z\in H,\,||z'-z||_d\le\epsilon_k}\frac{1}{v_k(\tilde h_k)}\log\left(\mathbb{P}^*\left(\left(\frac{T_{\tilde n_k}(\cdot,\tilde h_k,z')}{\sqrt{2f(z)\tilde n_k\tilde h_k^d v_k(\tilde h_k)}},\frac{T_{\tilde n_k}(\cdot,\tilde{\mathfrak{h}}_k,z')}{\sqrt{2f(z)\tilde n_k\tilde{\mathfrak{h}}_k^d v_k(\tilde{\mathfrak{h}}_k)}}\right)\in F\right)\right)\le -J^{\otimes 2}_{\mathcal{G}}(F);$$

$\forall O$ *open in* $\big(\ell^\infty(\mathcal{G}),||\cdot||_{\mathcal{G}}\big)\times\big(\ell^\infty(\mathcal{G}),||\cdot||_{\mathcal{G}}\big)$,

$$\underline{\lim_{k\to\infty}}\inf_{z\in H,\,||z'-z||_d\le\epsilon_k}\frac{1}{v_k(\tilde h_k)}\log\left(\mathbb{P}_*\left(\left(\frac{T_{\tilde n_k}(\cdot,\tilde h_k,z')}{\sqrt{2f(z)\tilde n_k\tilde h_k^d v_k(\tilde h_k)}},\frac{T_{\tilde n_k}(\cdot,\tilde{\mathfrak{h}}_k,z')}{\sqrt{2f(z)\tilde n_k\tilde{\mathfrak{h}}_k^d v_k(\tilde{\mathfrak{h}}_k)}}\right)\in O\right)\right)\ge -J^{\otimes 2}_{\mathcal{G}}(O).$$

Proof : The proof of Proposition 2.1.2 already established the (*Asymptotic equicontinuity*) criterion uniformly in $h\in[\tilde h_k,\tilde{\mathfrak{h}}_k]$, so it remains valid when requiring only simultaneity in $\{\tilde h_k,\tilde{\mathfrak{h}}_k\}$. Hence, we need only to prove the analogue of (*Finite dimensional ULDP*) in $\ell^\infty(\mathcal{G})\times\ell^\infty(\mathcal{G})$. Taking arbitrary $p\ge 1$ and $(g_1,\ldots,g_p)\in\mathcal{G}^p$, the same remark as in the first sentence of the proof holds for the approximation of the law of

$$\mathbf{S}_{k,z,z'}:=\left(\frac{T_{\tilde n_k}(g_1,\tilde h_k,z')}{\sqrt{2f(z)\tilde n_k\tilde h_k^d v_k(\tilde h_k)}},\ldots,\frac{T_{\tilde n_k}(g_p,\tilde h_k,z')}{\sqrt{2f(z)\tilde n_k\tilde h_k^d v_k(\tilde h_k)}},\frac{T_{\tilde n_k}(g_1,\tilde{\mathfrak{h}}_k,z')}{\sqrt{2f(z)\tilde n_k\tilde{\mathfrak{h}}_k^d v_k(\tilde{\mathfrak{h}}_k)}},\ldots,\frac{T_{\tilde n_k}(g_p,\tilde{\mathfrak{h}}_k,z')}{\sqrt{2f(z)\tilde n_k\tilde{\mathfrak{h}}_k^d v_k(\tilde{\mathfrak{h}}_k)}}\right)$$

by their gaussian analogues on $\mathbb{R}^{2p}$. Finally, since $v_k(\tilde h_k)\sim v_k(\tilde{\mathfrak{h}}_k)$, the only assertion to prove is that the covariance matrix of $\sqrt{v_k(\tilde h_k)}\,\mathbf{S}_{k,z,z'}$ converges uniformly in $z\in H$, $||\,z'-z\,||_d\le\epsilon_k$ to a block diagonal matrix having two blocks of size $p\times p$, which is a consequence of the following assertion

$$\forall (g,g') \in \mathcal{G}^2,\ \lim_{k\to\infty} \sup_{z\in H,\ ||z'-z||_d \le \epsilon_k} \frac{1}{f(z)\sqrt{\tilde{h}_k^d \tilde{\mathfrak{h}}_k^d}} \Big| \mathrm{Cov}\Big(g_{\tilde{h}_k,z'}(Z_1), g'_{\tilde{\mathfrak{h}}_k,z'}(Z_1)\Big)\Big| = 0. \tag{2.32}$$

Now (2.32) is proved by noting first that, by Lemma 2.1.1, we have, for fixed $(g,g') \in \mathcal{G}^2$:

$$\begin{aligned}
&\lim_{k\to\infty} \sup_{z\in H,\ ||z'-z||_d \le \epsilon_k} \frac{1}{f(z)\sqrt{\tilde{h}_k^d \tilde{\mathfrak{h}}_k^d}} \Big| \mathbb{E}\Big(g_{\tilde{h}_k,z'}(Z_1)\Big) \mathbb{E}\Big(g'_{\tilde{\mathfrak{h}}_k,z'}(Z_1)\Big)\Big| \\
=& \lim_{k\to\infty} \sqrt{\tilde{h}_k^d \tilde{\mathfrak{h}}_k^d} \sup_{z\in H,\ ||z'-z||_d \le \epsilon_k} f(z) \Big| \frac{1}{f(z)\tilde{h}_k^d} \mathbb{E}\Big(g_{\tilde{h}_k,z'}(Z_1)\Big)\Big| \times \Big| \frac{1}{f(z)\tilde{\mathfrak{h}}_k^d} \mathbb{E}\Big(g'_{\tilde{\mathfrak{h}}_k,z'}(Z_1)\Big)\Big| \\
=& 0.
\end{aligned}$$

On the other hand we have, for each $k \geq 1,\ z \in H,\ ||\ z' - z\ ||_d \leq \epsilon_k$,

$$\begin{aligned}
&\frac{1}{f(z)\sqrt{\tilde{h}_k^d \tilde{\mathfrak{h}}_k^d}} \Big| \mathbb{E}\Big(g_{\tilde{h}_k,z'}(Z_1) g'_{\tilde{\mathfrak{h}}_k,z'}(Z_1)\Big)\Big| \\
=& \frac{\tilde{h}_k^d}{f(z)\sqrt{\tilde{h}_k^d \tilde{\mathfrak{h}}_k^d}} \Big| \int g(u) g'\Big(\frac{\tilde{h}_k}{\tilde{\mathfrak{h}}_k} u\Big) f(z' + \tilde{h}_k u) du \Big| \\
=& \sqrt{\frac{\tilde{h}_k^d}{\tilde{\mathfrak{h}}_k^d}} M^2 (2M)^d \sup_{\substack{z\in H,\ ||z'-z||_d \le \epsilon_k, \\ u\in[-M,M]^d}} \frac{f(z' + \tilde{h}_k u)}{f(z)},
\end{aligned}$$

where we used both (*Bounded*) and (*Support*) in the last inequality. Now, since $\tilde{h}_k = o(\tilde{\mathfrak{h}}_k)$, and by (2.24), that last bound tends to 0, which concludes the proof of (2.32) and also concludes the proof of Proposition 2.1.4. □

2.1.5 Some words on bracketing

Another structural assumption frequently made on a class of functions is that it admits a finite bracketing numbers integral, namely

$$(Bracketing)\quad \int_0^{\infty} \sqrt{\log N_{[]}\big(\epsilon, \mathcal{G}, ||\cdot||_{Q,2}\big)} d\epsilon < \infty,$$

for some particular probability measure Q, (see, e.g., [92, Definition 2.1.6] for more details). For example, a class $\mathcal{G}$ with square integrable envelope is Q-Donsker under this assumption. As shown in the preceding subsection, the local empirical process $T_n(\cdot, h, z)$ can be more or less considered as an empirical process with underlying sample $(Y_i^{(h,z)})_{i\geq 1}$ having distribution

$\mathbf{P}_0^{(h,z)}$, (recall (2.15)). Moreover, we have, under (Hf) :

$$\lim_{h\to 0} \sup_{z\in H} \sup_{(g,g')\in\mathcal{G}^2} \Big| \int gg' d\mathbf{P}_0^{(h,z)} - \int gg' dQ \Big| = 0,$$

where Q is the uniform distribution on $[-M,M]^d$. This convergence of the $\mathbf{P}_0^{(h,z)}$ toward Q for a strong enough seminorm is the key argument to prove that all the results of §2.1.3 are still true when $(Unif.\ entropy)$ is replaced by $(Bracketing)$, with the choice of Q as the uniform distribution on $[-M,M]^d$. Indeed, the reader probably noticed, throughout the reading of §2.1.3, that assumption $(Unif.\ entropy)$ was used at only one single step : when bounding the expectation of suprema of empirical processes over classes of functions fulfilling $(Unif.\ entropy)$ that are contained in small $||\cdot||_{\lambda,2}$-balls (see Lemma 2.1.2). Using, e.g., [91, Lemma 19.34], it is possible to obtain similar bounds for classes of functions fulfilling $(Bracketing)$ that are contained in small $||\cdot||_{\lambda,2}$ brackets.

I hope that those few words can convince the reader that all the results of the present chapter and of Chapter 3 implying the local empirical processes $T_n(\cdot,h,z)$ can be extended, when replacing the uniform entropy number assumptions by their bracketing counterparts, with Q being the uniform distribution on $[-M,M]^d$.

2.1.6 Perspectives

My main perspective of improvement of these ULDP concerns the uniform boundedness of $\mathcal{G}$ (assumption $(Bounded)$). The boundedness of $\mathcal{G}$ plays its crucial role in the use of Talagrand's concentration inequality, so as the second term $\exp(-A_2 t/M)$ become negligible in front of the first (see (2.28)). Following an idea already present in [42], I would like to relax the boundedness assumption to a (high order) integrability of the envelope, perhaps at the price of truncating, by making use of Fuk-Nagaev type inequalities. Einmahl and Li [40] did improve the Fuk-Nagaev inequality which, in its improved form, and applied to empirical processes, can be stated as follows.

Fact 2.1.3 (Fuk-Nagaev type inequality, Einmahl, Li, 2008) *Take the notations of Fact 2.1.2 but, instead of assuming* $(Bounded)$, *suppose that the class* $\mathcal{G}$ *admits an envelope* G *which satisfies* $\mathbb{E}\big(G^p(Z_1)\big) < \infty$ *for some* $p > 2$. *Then given* $\eta \in (0,1]$ *and* $\delta > 0$, *there exists* C *such that, for all* $t > 0$ *and* $n \geq 1$, *we have :*

$$\mathbb{P}\bigg(\max_{m=1,\dots,n} \sup_{g\in\mathcal{G}} \Big| \sum_{i=1}^m g(Z_i) - \mathbb{E}\Big(g(Z_i)\Big) \Big| \geq (1+\eta)\beta_n + t \bigg)$$
$$\leq \exp\bigg(- \frac{t^2}{(2+\delta)n\sigma_{\mathcal{G}}^2} \bigg) + C\mathbb{E}\Big(G^p(Z_1)\Big)\frac{n}{t^p}.$$

2.2 Poissonization

Consider a sequence $(\eta_n)_{n\geq 1}$ of random variables on $(\Omega, \mathcal{A}, \mathbb{P})$. The η_n-randomized local empirical process is defined as follows, with the convention $\sum_{\emptyset} = 0$.

$$\Pi^{\eta_n}(g, h, z) := \sum_{i=1}^{\eta_n} g_{h,z}(Z_i) - \mathbb{E}\Big(g_{h,z}(Z_i)\Big),\ g \text{ Borel and bounded.} \tag{2.33}$$

When each η_n is Poisson with expectation n and independent of $(Z_i)_{i\geq 1}$, the $\Pi^{\eta_n}(\cdot, h, z)$ are called Poissonized versions of the local empirical process (or local Kac processes), and shall be written $\Pi_n(\cdot, h, z)$ for short. The main feature of $\Pi_n(\cdot, h, z)$ is that those processes inherit of a crucial property of Poisson random measures, which, in terms of empirical measures, can be written as follows.

Proposition 2.2.1 *If $(\mathcal{G}_1, \ldots, \mathcal{G}_p)$ are classes of functions respectively vanishing out of disjoint subsets $J_1, \ldots, J_p$, then the family of processes*

$$\Big(g_1 \to \sum_{i=1}^{\eta_n} g_1(Z_i), \ldots, g_p \to \sum_{i=1}^{\eta_n} g_p(Z_i)\Big)$$

is mutually independent, as processes defined respectively on $\mathcal{G}_1, \ldots, \mathcal{G}_p$.

Poissonization can be seen a way to switch from a probability calculus on empirical processes to a related calculus for their Poissonized versions, in order to use the preceding property.

2.2.1 The negligible effect of sample size randomization for large deviation events

Up to far in the literature, Poissonization techniques heavily relied on some interplay between Poisson and binomial distributions (see, e.g., [85, Chapter 8], [92, Chapter 3.5]). My first main contribution to these techniques is that, when considering large deviation events, neither the fact that η_n is Poisson nor that it is independent of the sample are required to obtain suitable good approximations. It is sufficient that the η_n concentrate enough around their expectations n, in regard with the large deviation principle that is involved (hence the "Poissonization" terminology should be relaxed to a general *sample size randomization*). Our next proposition explicitly uses the ULDP framework of Proposition 2.1.2, to make the arguments more clear an coherent. However, it has to be noted that such techniques are not limited to local empirical processes. They can be translated into results of the same type for general empirical processes.

Proposition 2.2.2 (Varron, 2014, from [108]) *Take all the notations and assumptions of*

Proposition 2.1.2. Assume that the sequence $\eta_{\tilde{n}_k}$ fulfills the following condition :

$$\forall \boldsymbol{\delta} > 0,\ \lim_{k\to\infty} \sup_{h\in[\tilde{h}_k,\tilde{\mathfrak{h}}_k]} \frac{1}{v_k(h)} \log\left(\mathbb{P}\Big(\frac{\eta_{\tilde{n}_k}}{\tilde{n}_k} \notin [1-\boldsymbol{\delta}, 1+\boldsymbol{\delta}]\Big)\right) = -\infty, \tag{2.34}$$

then we have, for each $\epsilon > 0$:

$$\lim_{k\to\infty} \sup_{\substack{z\in H,\ ||z'-z||_d\leq \epsilon_k \\ h\in[\tilde{h}_k,\tilde{\mathfrak{h}}_k]}} \frac{1}{v_k(h)} \log\left(\mathbb{P}\Big(\frac{\Big|\Big|\Pi_{\tilde{n}_k}^{\eta_{\tilde{n}_k}}(\cdot,h,z') - T_{\tilde{n}_k}(\cdot,h,z')\Big|\Big|_{\mathcal{G}}}{\sqrt{2f(z)\tilde{n}_k h^d v_k(h)}} > \epsilon\Big)\right) = -\infty.$$

As a consequence, Proposition 2.1.2 is still true with the formal replacement of $T_{\tilde{n}_k}(\cdot,h,z')$ by $\Pi^{\eta_{\tilde{n}_k}}(\cdot,h,z')$.

Before giving a short proof, I would like to point out that, at least in the theory of functional limit laws, assumption (2.34) is always very largely satisfied when needed.

Proof : First fix $B > 0$, and set $\boldsymbol{\delta} := C_2\epsilon^2/(8B\Delta_{\mathcal{G}}^2 C_1^2)$, where C_1 and C_2 are as in Fact 2.1.2. Writing $N_k(\boldsymbol{\delta}) := \big\{n \in \mathbb{N},\ n/\tilde{n}_k \in [1-\boldsymbol{\delta}, 1+\boldsymbol{\delta}]\big\}$, we have, for fixed $k \geq 1,\ z \in H,\ h \in [\tilde{h}_k, \tilde{\mathfrak{h}}_k],\ ||\ z'-z\ ||\leq \epsilon_k$:

$$\begin{aligned}&\mathbb{P}\left(\frac{\Big|\Big|\Pi_{\tilde{n}_k}^{\eta_{\tilde{n}_k}}(\cdot,h,z') - T_{\tilde{n}_k}(\cdot,h,z')\Big|\Big|_{\mathcal{G}}}{\sqrt{2f(z)\tilde{n}_k h^d v_k(h)}} > \epsilon\right)\\
\leq &\mathbb{P}\Big(\eta_{\tilde{n}_k} \notin N_k(\boldsymbol{\delta})\Big) + \mathbb{P}\left(\max_{n\in N_k(\boldsymbol{\delta})} \frac{\sup_{g\in\mathcal{G}}\Big|\sum_{i=n+1}^{\tilde{n}_k} g_{h,z'}(Z_i) - \mathbb{E}\Big(g_{h,z'}(Z_i)\Big)\Big|}{\sqrt{2f(z)\tilde{n}_k h^d v_k(h)}} > \epsilon\right),\end{aligned}$$

with the convention $\sum_{i=p+1}^{q} u_i := -\sum_{i=q+1}^{p} u_i$ for $p > q$. By (2.34) it is sufficient to bound the second term. This is done by applying Proposition 2.1.3, noticing that, for all large enough k, we have :

$$\begin{aligned}&\sup_{h\in[\tilde{h}_k,\tilde{\mathfrak{h}}_k]} h = \tilde{\mathfrak{h}}_k \leq h_0,\ \ \epsilon_k \leq \epsilon_0,\\
&u_k(h) := \frac{\epsilon^2}{4\boldsymbol{\delta} C_1^2} v_k(h) \geq (C_0C_3)^2 2^{d+1} M^d,\ \text{for all } h \in [\tilde{h}_k, \tilde{\mathfrak{h}}_k],\\
&\inf_{h\in[\tilde{h}_k,\tilde{\mathfrak{h}}_k]} \frac{[\boldsymbol{\delta}\tilde{n}_k + 1]h^d}{u_k} = \frac{[\boldsymbol{\delta}\tilde{n}_k + 1]\tilde{h}_k^d}{u_k} \geq K_0,\end{aligned}$$

which yields, $\tilde{n}_k \geq \boldsymbol{\delta}^{-1}$, and uniformly in $h \in [\tilde{h}_k, \tilde{\mathfrak{h}}_k]$, $\| z' - z \| \leq \epsilon_k$:

$$\begin{aligned}
&\mathbb{P}\Bigg(\max_{n \in N_k(\boldsymbol{\delta})} \frac{\sup_{g \in \mathcal{G}} \Big| \sum_{i=n+1}^{\tilde{n}_k} g_{h,z'}(Z_i) - \mathbb{E}\Big(g_{h,z'}(Z_i)\Big)\Big|}{\sqrt{2f(z)\tilde{n}_k h^d v_k(h)}} > \epsilon \Bigg) \\
\leq & 2\mathbb{P}\Bigg(\max_{0 \leq m \leq [\boldsymbol{\delta}\tilde{n}_k+1]} \frac{\sup_{g \in \mathcal{G}} \Big| \sum_{i=1}^{m} g_{h,z'}(Z_i) - \mathbb{E}\Big(g_{h,z'}(Z_i)\Big)\Big|}{\sqrt{2f(z)\tilde{n}_k h^d v_k(h)}} > \epsilon/2 \Bigg) \\
\leq & 2\mathbb{P}\Bigg(\max_{0 \leq m \leq [\boldsymbol{\delta}\tilde{n}_k+1]} \sup_{g \in \mathcal{G}} \Big| \sum_{i=1}^{m} g_{h,z'}(Z_i) - \mathbb{E}\Big(g_{h,z'}(Z_i)\Big)\Big| \geq C_1 \sqrt{f(z)[\boldsymbol{\delta}\tilde{n}_k + 1]h^d \frac{\epsilon^2}{4\boldsymbol{\delta}C_1^2} v_k(h)} \Bigg) \\
\leq & 8\exp\Big(- \frac{C_2 \epsilon^2 v_k(h)}{4\boldsymbol{\delta}C_1^2 \Delta_{\mathcal{G}}^2} \Big) \\
\leq & 8\exp\Big(- 2Bv_k(h)\Big),
\end{aligned}$$

which concludes the proof, as B was arbitrarily chosen. □

2.2.2 A general Poissonization tool for i.i.d. sums in a measurable semigroup

In this subsection I will expose a Poissonization technique which heavily relies on the interplay between binomial and Poisson distributions. Einmahl [36] established a Poissonization technique for the local uniform empirical process. In the setup of general empirical processes, his result can be stated as follows. Recall the general notation $Q(f) = \int f dQ$.

Proposition 2.2.3 (Derived from J. Einmahl, 1986) *Let $J \in \mathcal{X}$ such that $\mathbb{P}\big(Z_1 \in J\big) < 1$. Let $\mathcal{G}$ be a class of functions such that $g(z) = 0$ for each $z \notin J$ and $g \in \mathcal{G}$. There exist a constant C, depending only upon $\mathbb{P}\big(Z_1 \in J\big)$, such that, for any $A \subset \mathbb{R}^{\mathcal{G}}$, which is measurable for each $G_n(\cdot)$, $n \geq 1$, we have :*

$$\forall n \geq 1,\ \mathbb{P}\Big(\sum_{i=1}^{n} \delta_{Z_i}(\cdot) \in A \Big) \leq C\mathbb{P}\Big(\sum_{i=1}^{\eta_n} \delta_{Z_i}(\cdot) \in A \Big),$$

where η_n is Poisson with expectation n, and is independent of the Z_i in the RHS. Moreover, when $\mathbb{P}\big(Z_1 \in J\big) \leq 1/2$, C can be chosen equal to 2.

That result turned out to be crucial when proving global functional limit laws for the local empirical process. Indeed, when $\mathcal{G}$ satisfies (*Support*), and a soon as $\mathbb{P}\big(Z_1 \in H^{Mh_0}\big) < 1$ for some $h_0 > 0$ then all the function $g_{h,z}$, $z \in H, h \in]0, h_0[$, $g \in \mathcal{G}$ have their supports included in $J := H^{Mh_0}$. Hence, Proposition 2.2.3 allows to Poissonize probabilities for events *simultaneously* involving local empirical processes at different points z and at different bandwidths h.

Giné *et. al.* [46] put Einmahl's result in the more abstract framework of sums of i.i.d. random variables $(X_i)_{i\geq 1}$ in a measurable semigroup $(D,+,\mathcal{D})$.

Proposition 2.2.4 (Giné, Mason, Zaitsev, 2001) *Let $(X_i)_{i\geq 1}$ be i.i.d random variables in a measurable semigroup $(D,+,\mathcal{D})$. Let $B\in\mathcal{D}$ such that $\mathbb{P}\big(X_1\in B\big)<1$. Then there exists a constant C such that, for each measurable map H from $(D,\mathcal{D})$ to $[0,\infty[$, we have :*

$$\forall n\geq 1,\ \mathbb{E}\Big(H\Big(\sum_{i=1}^{n}\mathbb{1}_B(X_i)X_i\Big)\Big)\leq C\mathbb{E}\Big(H\Big(\sum_{i=1}^{\eta_n}\mathbb{1}_B(X_i)X_i\Big)\Big),$$

where η_n is Poisson with expectation n, and is independent of the X_i in the RHS. Moreover, when $\mathbb{P}\big(X_1\in B\big)\leq 1/2$, the constant C can be chosen equal to 2.

When working on spatial functional limit laws [103], I was seeking a way to Poissonize maximal inequalities for sums of random variables in a Banach space, namely, to obtain inequalities of the form

$$\forall n\geq 1,\ \mathbb{P}\Big(\max_{m=1,\ldots,n}\Big|\Big|\sum_{i=1}^{m}\mathbb{1}_B(X_i)X_i\Big|\Big|\geq\epsilon\Big)\leq C\mathbb{P}\Big(\max_{m=1,\ldots,\eta_n}\Big|\Big|\sum_{i=1}^{m}\mathbb{1}_B(X_i)X_i\Big|\Big|\geq\epsilon\Big).$$

This research led to a more general Poissonization technique, which may open new possibilities. To enounce it, I will first introduce some notations. We shall write $\widetilde{D}:=\bigcup_{n\geq 1}D^n$.

Definition 2.2.1 (Truncating maps) *Let E_0 be an arbitrary set. A map $\phi:\widetilde{D}\mapsto E_0$ is said to be truncating if, for any $p\geq 2$ and $(d_1,\ldots,d_p)\in D^p$ we have*

$$\phi(d_1,\ldots,d_p,d_p)=\phi(d_1,\ldots,d_p).$$

We shall write, for simplicity of notations,

$$\begin{aligned}\sum_{i=q}^{\rightarrow p}d_i:=&\Big(d_q,d_q+d_{q+1},\ldots,\sum_{i=q}^{p}d_i\Big),\ \text{when } p\geq q,\\ :=&0\ \text{otherwise.}\end{aligned}$$

Hence, the symbol $\sum_{i=1}^{\rightarrow p}$ is understood as the *history record* of partials sums up to p.
My Poissonization result is as follows :

Proposition 2.2.5 (Varron, 2011, from [106]) *Let $(X_i)_{i\geq 1}$ be i.i.d random variables in a measurable semigroup $(D,+,\mathcal{D})$. Let $\widetilde{\mathcal{D}}:=\bigvee_{n\geq 1}\mathcal{D}^{\otimes n}$. Let $(E_0,\mathcal{A}_0)$ be a measurable space and let $\phi:(\widetilde{D},\widetilde{\mathcal{D}})\mapsto[0,\ldots,\infty[$ be measurable. Let $B\in\mathcal{D}$ such that $\mathbb{P}\big(X_1\in B\big)<1$. Then there* *exists a constant C, depending only upon $\mathbb{P}\big(X_1 \in B\big)$, such that, :*

$$\forall n \geq 1,\ \mathbb{E}\bigg(\phi\Big(\sum_{i=1}^{\rightarrow n} \mathbb{1}_B(X_i)X_i\Big)\bigg) \leq C\mathbb{E}\bigg(\phi\Big(\sum_{i=1}^{\rightarrow \eta_n} \mathbb{1}_B(X_i)X_i\Big)\bigg),$$

where η_n is Poisson with expectation n, and is independent of the X_i in the RHS.

This general framework encompasses that of Proposition 2.2.4, by choosing

$$\phi(s_1,\ldots,s_p) = s_p,\ p \geq 1,\ (s_1,\ldots,s_p) \in D^p.$$

It also encompasses maxima of partial sums, by choosing, when D is a normed space :

$$\phi(s_1,\ldots,s_p) = \max_{j=1,\ldots p} \mid\mid s_j \mid\mid,\ p \geq 1,\ (s_1,\ldots,s_p) \in D^p.$$

Proof : Write $p_B := \mathbb{P}(X_1 \in B)$. By an application of Stirling's formula, and since $(1-p_B) > 0$, we have

$$\sup_{n\geq 1} \frac{\frac{\big(n(1-p_B)\big)^{[n(1-p_B)]}}{e^{n(1-p_B)}[n(1-p_B)]!}}{\frac{n^n}{e^n n!}} =: C < \infty, \tag{2.35}$$

and direct calculation for small values of n shows that $C \leq 2$ when $p_B \leq 2$ (see [46, p. 8]). Now denote by $(\tau_i, Y_i)_{i\geq 1}$ an i.i.d. sequence for which Y_i is independent of τ_i, $\mathbb{P}(\tau_i = 1) = 1 - \mathbb{P}(\tau_i = 0) = p_B$ and Y_i has the distribution of X_i conditionally to $X_i \in B$. A simple calculation shows that $\big(\mathbb{1}_B(X_i)X_i\big)_{i\geq 1} =_d \big(\tau_i Y_i\big)_{i\geq 1}$, from where, by conditioning on $(\tau_1,\ldots,\tau_n)$:

$$\mathbb{E}\bigg(\phi\Big(\sum_{i=1}^{\rightarrow n} \mathbb{1}_B(X_i)X_i\Big)\bigg) = \sum_{\mathcal{P}\subset\{1,\ldots,n\}} p_B^{\sharp\mathcal{P}}(1-p_B)^{n-\sharp\mathcal{P}}\ \mathbb{E}\bigg(\phi\Big(\sum_{i=1}^{\rightarrow n} \mathbb{1}_{\mathcal{P}}(i)Y_i\Big)\bigg),$$

with the notation $\sharp\mathcal{P}$ for the number of elements of $\mathcal{P}$.
Now notice that, for fixed $\mathcal{P}$ and for each permutation of indices σ, we have

$$\big(\mathbb{1}_{\mathcal{P}}(1)Y_1,\ldots,\mathbb{1}_{\mathcal{P}}(n)Y_n\big) =_d \big(\mathbb{1}_{\mathcal{P}}(\sigma(1))Y_1,\ldots,\mathbb{1}_{\mathcal{P}}(\sigma(n))Y_n\big).$$

By choosing σ such that $(\mathbb{1}_{\mathcal{P}}(\sigma(1)),\ldots,\mathbb{1}_{\mathcal{P}}(\sigma(n))) = (1,\ldots,1,0,\ldots,0)$ the vector $\sum_{i=1}^{\rightarrow n} \mathbb{1}_{\mathcal{P}}(\sigma(i))Y_i$ has his last $n - \sharp\mathcal{P} + 1$ coordinates equal, from where, since ϕ is truncating :

$$\mathbb{E}\bigg(\phi\Big(\sum_{i=1}^{\rightarrow n} \mathbb{1}_B(X_i)X_i\Big)\bigg) = \sum_{k=0}^{n} \binom{n}{k} p_B^k(1-p_B)^{n-k}\ \mathbb{E}\bigg(\phi\Big(\sum_{i=1}^{\rightarrow k} Y_i\Big)\bigg).$$

The remainder of the calculus continues as in the proof of Lemma 2.1 in [46], namely, introducing independent Poisson random variables η_B and η_{B^C} with respective expectations np_B and $n(1-$

p_B), which are also independent of $(Y_i)_{i\geq 1}$:

$$\begin{aligned}
&\sum_{k=0}^{n}\binom{n}{k}p_B^k(1-p_B)^{n-k}\,\mathbb{E}\Big(\phi\Big(\sum_{i=1}^{\to k}Y_i\Big)\Big)\\
=&\frac{n!e^n}{n^n}\sum_{k=0}^{n}\frac{(np_B)^k}{k!e^k}\frac{\big(n(1-p_B)\big)^{n-k}}{(n-k)!e^{n-k}}\mathbb{E}\Big(\phi\Big(\sum_{i=1}^{\to k}Y_i\Big)\Big)\\
=&\frac{1}{\mathbb{P}\big(\eta_n=n\big)}\sum_{k=0}^{n}\mathbb{E}\Big(\phi\Big(\sum_{i=1}^{\to k}Y_i\Big)\mathbb{1}_{\{\eta_B=k\}}\Big)\mathbb{P}\Big(\eta_{B^C}=n-k\Big)\\
\leq&\frac{\max_{k=0,\ldots,n}\mathbb{P}\big(\eta_{B^C}=n-k\big)}{\mathbb{P}\big(\eta_n=n\big)}\mathbb{E}\Big(\phi\Big(\sum_{i=1}^{\to\eta_B}Y_i\Big)\Big)\\
=&\frac{\mathbb{P}\big(\eta_{B^C}=[n(1-p_B)]\big)}{\mathbb{P}\big(\eta_n=n\big)}\mathbb{E}\Big(\phi\Big(\sum_{i=1}^{\to\eta_B}Y_i\Big)\Big)\\
\leq&C\mathbb{E}\Big(\phi\Big(\sum_{i=1}^{\to\eta_B}Y_i\Big)\Big),\text{ by 2.35.}
\end{aligned}$$

The end of the proof now requires an additional work, which consists in proving that

$$\mathbb{E}\Big(\phi\Big(\sum_{i=1}^{\to\eta}\tau_iY_i\Big)\Big)=\mathbb{E}\Big(\phi\Big(\sum_{i=1}^{\to\eta_B}Y_i\Big)\Big),$$

This is done by using the interplay between Poisson and Binomial distributions :

$$\begin{aligned}
&\mathbb{E}\Big(\phi\Big(\sum_{i=1}^{\to\eta}\tau_iY_i\Big)\Big)\\
=&\sum_{m\geq 0}\mathbb{P}(\eta=m)\sum_{\mathcal{P}\subset\{1,\ldots,m\}}p_B^{\sharp\mathcal{P}}(1-p_B)^{m-\sharp\mathcal{P}}\,\mathbb{E}\Big(\phi\Big(\sum_{i=1}^{\to m}\mathbb{1}_{\mathcal{P}}(i)Y_i\Big)\Big)\\
=&\sum_{m\geq 0}\frac{n^m}{m!}e^{-n}\sum_{k=0}^{m}\binom{m}{k}p_B^k(1-p_B)^{m-k}\,\mathbb{E}\Big(\phi\Big(\sum_{i=1}^{\to k}Y_i\Big)\Big)\\
&\text{(by the same truncating arguments as above)}\\
=&\sum_{k\geq 0}\mathbb{E}\Big(\phi\Big(\sum_{i=1}^{\to k}Y_i\Big)\Big)\sum_{m\geq k}\frac{n^m}{m!}\frac{m!}{k!(n-k)!}p_B^k(1-p_B)^{m-k}e^{-n}\\
=&\sum_{k\geq 0}\mathbb{E}\Big(\phi\Big(\sum_{i=1}^{\to k}Y_i\Big)\Big)e^{-np_B}\frac{(np_B)^k}{k!}\sum_{m'\geq 0}\frac{(n(1-p_B))^{m'}}{m'!}e^{-n(1-p_B)}
\end{aligned}$$

$$= \sum_{k \geq 0} \mathbb{P}(\eta_B = k) \mathbb{E}\left(\phi\Big(\overset{\rightarrow k}{\sum_{i=1}} Y_i \Big) \right),$$

which proves Proposition 2.2.5. □

2.2.3 Perspectives

As a perspective for future works, I would like to investigate the possible applications of Proposition 2.2.5. Since a wide variety of maps do satisfy the *truncating* property, such a Poissonization tool may open new research directions for empirical processes, or more generally, to i.i.d. sums in a semigroup. Some leads of applications are stated in [103].

Chapter 3

Contributions to the study of local empirical processes

My contributions to the study of local empirical processes can be decomposed into three parts :

— The independence between bandwidths phenomenon, ands its consequences to in-bandwidth uniformity in the standard functional limit laws. Those results are stated in §3.2.

— Spatial-type limit results for the local empirical process : endowing a set H with a finite measure μ, I studied the asymptotic repartitions of the $T_n(\cdot, h_n, z)$, when z is distributed by μ. Those results are stated in §3.3.

— A collection of more or less incremental improvements in the existing functional limit laws, without considering in-bandwidth uniformity or spatial repartitions. Those results are stated in the following section.

Each of the presented results is accompanied with some words about the key elements of its proof, with frequent references to Chapter 2.

3.1 Improvements of existing functional limit laws

3.1.1 Some preliminary words on the usual tools used in functional limit laws

In the proof of functional limit laws, several arguments are recurrent. In this subsection, I will briefly explain them, in order to mention them when needed in the sequel.

Blocking arguments

In both local and global functional limit laws, the main idea is to prove that probabilities of events A_n, involving collections of processes $T_n(\cdot, h, z)$, are a summable in n, in order to use the Borel-Cantelli lemma. Systematically, it appears that those probabilities are summable only along subsequences of the form $n_k := [(1+\gamma)^k]$, for $\gamma > 0$. A key argument to interpolate between the n_k is to use maximal inequalities for partial sums of i.i.d. processes. One of them is the Montgomery-Smith inequality, which I chose to write down here, for its concision and simplicity.

Fact 3.1.1 (Montgomery-Smith, 1993) *Let $(E, ||\cdot||)$ be a Banach space, and let $(X_i)_{i\geq 1}$ be an i.i.d. sequence, for which $||\ X_1\ ||$ is Borel measurable. Then for all $n \geq 1$ and for all $t \geq 0$, we have :*

$$\mathbb{P}\Big(\max_{m=1,\dots,n} \Big|\Big|\sum_{i=1}^{m} X_i\Big|\Big| \geq t\Big) \leq 9\mathbb{P}\Big(\Big|\Big|\sum_{i=1}^{n} X_i\Big|\Big| \geq \frac{t}{30}\Big).$$

Interpolating between the n_k is (up to small technicalities due to the variation of h_n with n) is essentially finding bounds, for fixed $\epsilon > 0$, for probabilities

$$\begin{aligned}
&\mathbb{P}\Big(\max_{m=1,\dots,n_k-n_{k-1}} \Big|\sum_{i=1}^{m} g_{h,z}(Z_i) - \mathbb{E}\Big(g_{h,z}(Z_i)\Big)\Big| \geq \epsilon\sqrt{2f(z)n_k h_{n_k}^d \log\log(n_k)}\Big)\\
\leq& 9\mathbb{P}\Big(\Big|\Big|T_{n_k-n_{k-1}}(\cdot,h,z)\Big|\Big|_{\mathcal{G}} \geq \frac{\epsilon}{30}\sqrt{2f(z)n_k h_{n_k}^d \log\log(n_k)}\Big),\\
\leq& 9\mathbb{P}\Big(\Big|\Big|T_{n_k-n_{k-1}}(\cdot,h,z)\Big|\Big|_{\mathcal{G}} \geq \frac{\epsilon}{30\sqrt{\gamma}}\sqrt{2f(z)h_{n_k}^d (n_k-n_{k-1}) \log\log(n_k)}\Big),
\end{aligned}$$

where, in the last inequality, the term $\sqrt{\gamma}$ could be inserted because $n_k - n_{k-1} \leq \gamma n_k$ for all large k. Next, a use, e.g. of Proposition 2.1.3, provides the right deviation bounds, with a choice of γ small enough to make the threshold $\epsilon/\sqrt{\gamma}$ large enough.

Stochastic renewal

Local functional limit laws use the converse part of the Borel-Cantelli lemma. Such a converse part only involves mutually independent events A_n. The $T_n(\cdot,\cdot,\cdot)$ clearly fail to be independent in n. However, along sufficiently spread subsequences n_k, these processes can be proved to be "almost" independent. By sufficiently spread, we can typically choose sequences for which n_{k-1}/n_k tends to 0 (or at least, has a negligible upper limit). The idea is to first prove asymptotic results for the mutually independent sequences of processes

$$\overline{T}_k(g,h,z) := \frac{\sum_{i=n_{k-1}+1}^{n_k} g_{h,z}(Z_i) - \mathbb{E}\Big(g_{h,z}(Z_i)\Big)}{\sqrt{2f(z)n_k h^d \log\log(n_k)}},\ k \geq 1 \tag{3.1}$$

for which the normalization makes senses since $n_k - n_{k-1} \sim n_k$. Then the distance to the original sequence can be rendered negligible by the following calculus

$$\Big|\Big|\overline{T}_k(\cdot,h,z) - \frac{T_{n_k}(\cdot,h,z)}{\sqrt{2f(z)n_k h^d \log\log(n_k)}}\Big|\Big| = \sqrt{\frac{n_{k-1}}{n_k}} \times \frac{T_{n_{k-1}}(\cdot,h,z)}{\sqrt{2f(z)n_{k-1}h^d \log\log(n_k)}},$$

where the second factor is properly normalized to be stochastically controlled, and the first factor $(n_{k-1}/n_k)^{1/2}$ tends to zero.

3.1.2 Perspectives

A careful reading of the proof of Fact 3.1.1 (see [70]) shows that neither the Banach structure, nor the vector space structure does play a role. A perspective of future works wound be to extend Fact 3.1.1 for X_i taking values in an abelian group $(D,+)$, with null element 0_D, and endowed with a norm-like map $||\cdot||$, or more generally, a map $||\cdot||$ taking values in $\mathbb{R}^+$, for which $|| 0_D ||= 0$, and for which, writing $k \odot e := e + \ldots + e$ (k times) and $(-k)\odot e := -(k \odot e)$ for $k \in \mathbb{N}^*$ and $0 \odot e := 0_D$, we have

$$C_1 \mid k \mid^{\alpha_1} || e || \leq || k \odot e || \leq C_2 \mid k \mid^{\alpha_2} || e ||,$$

where neither C_1, C_2 nor α_1, α_2 depends on $k \in \mathbb{Z}$ or $e \in D$.

3.1.3 Clustering rates and Chung-Mogulskii functional limit laws for the local uniform empirical process

When $d = 1$, $\mathcal{G}$ is the particular class

$$\mathcal{G}_0 := \Big\{ \mathbb{1}_{[0,t]},\ t \in [0,1] \Big\},$$

and when the Z_i are uniform on $[0,1]$, the $T_n(\cdot, h, z)$ are identified to the *functional increments of the uniform empirical process*. Those processes were the first to be studied (see, e.g., [85]), because, in addition to being rich enough to have connections with a lot of statistical procedures, they benefit of particularly strong mathematical tools. I will expose some of them in the next paragraph. It first seems convenient to introduce notations that are adapted to the specific framework of the uniform empirical process. We shall write the empirical process

$$\alpha_n(t) := \sqrt{n}(F_n(t) - t),\ \text{where}$$
$$F_n(t) := \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{[0,t]}(U_i),\ t \in [0,1],$$

and the $(U_i)_{n\geq 1}$ are independent, identically distributed (i.i.d) random variables on $(\Omega, \mathcal{A}, \mathbb{P})$, uniformly distributed on $[0,1]$. Define the quantile process by

$$\beta_n(t) = \sqrt{n}\Big(F_n^{-1}(t) - t\Big),\ t \in [0,1],\ \text{where}$$
$$F_n^{-1}(t) := \inf\{u :\ F_n(u) \geq t\}.$$

We shall denote by W a Wiener Process on $[0,1]$, and by B a Brownian bridge on $[0,1]$ (the law of the Brownian bridge is, e.g., defined as the law of $W(\cdot)-W(1)\cdot$ on $[0,1]$). Throughout this subsection we will place ourselves in the metric space $(\ell^\infty([0,1]),||\cdot||_{[0,1]})$.

First tool : Strong approximation by Gaussian processes

Strong approximation of empirical processes by their Gaussian counterparts has been the object of several research works [43, 58, 59, 60, 67, 79, 80, 90]. Among them, the Komlos-Major-Tusnady construction [59, 60, 90] for the uniform empirical process is the only one that provides rates of approximations of order $\log(n)/\sqrt{n}$. These rates of approximations have been proved to hold for the bivariate uniform empirical process [60] (replacing $\log(n)$ by $\log(n)^2$ in the rate of approximation).

Fact 3.1.2 (Komlos-Major-Tusnady [59]) *At the price of enriching the probability space $(\Omega,\mathcal{A},\mathbb{P})$, one can construct a sequence B_n of copies of B, such that, for some constants C_1,C_2,C_3 we have, for all $n\geq 1$ and $z>0$:*

$$\mathbb{P}\Big(\,||\,\alpha_n-B_n\,||_{[0,1]}\geq C_1\frac{\log n+z}{\sqrt{n}}\Big)\leq C_2\exp\big(-C_3z\big).$$

A similar result was proved by the authors [60] for the standard Poisson process on the real line, involving copies W_n of W.
Such a powerful strong approximation result not only allows to prove appropriate large deviation principles, but it also allows to explore second order convergence in these large deviations. For example, while a large deviation result may only give rates of convergence for open ball with fixed radius (namely $O=\psi^\epsilon$ in Proposition 2.1.2), the KMT approximation allows, by Gaussian analysis, to investigate balls with radii tending to 0 as $n\to\infty$ (namely $O=\psi^{\epsilon_n}$, with $\epsilon_n\to 0$).

Second tool : Gaussian analysis for enlarged Strassen sets

The description of the *first tool* leads to the question of small ball probabilities (or more generally, probabilities involving A^{ϵ_n}, for a fixed region A). When $d=1$, Gaussian analysis provides sharp bounds for these probabilities. When $d\geq 2$, generalizing such results is an open an tough problem (see, e.g., [61] for an overview on this field).
Define the Strassen sets

$$\mathcal{S}_1:=\Big\{f\in\ell^\infty([0,1]),\ \exists f'\text{ Borel},\ f(\cdot):=\int_0^\cdot f'(t)dt,\ \int_0^1 f'^2(t)dt\leq 1\Big\},\text{ and}$$

$$\mathcal{S}_2:=\Big\{f(t)\in\mathcal{S}_1,\ f(1)=0\Big\}.$$

Note that $\mathcal{S}_2$ (resp. $\mathcal{S}_1$) is the unit ball of the reproducing kernel Hilbert space of the Brownian bridge (resp. of the Wiener process) on [0,1]. The next inequalities, due to Talagrand, involve small enlargements of those Strassen sets (recall (2.5)).

Fact 3.1.3 (Talagrand [87]) *There exists two constants L_0 and u_0 >0 such that, for any $0 < u < u_0$ and $c > 0$, we have :*

$$\mathbb{P}\Big(B \notin (c\mathcal{S}_2)^u\Big) \leq \exp\Big(\frac{L_0}{u^2} - \frac{cu}{2} - \frac{c^2}{2}\Big).$$

There exist two constants u_1 and L_1 such that, for any $0 < u < u_1$ and $c > 0$, we have

$$\mathbb{P}\Big(W \notin (c\mathcal{S}_1)^u\Big) \leq \exp\Big(\frac{L_1}{u^2} - \frac{cu}{2} - \frac{c^2}{2}\Big).$$

Berthet [8] used inequality (3.2) coupled with the KMT approximation of Fact 3.1.2 to obtain clustering rates in the global standard functional limit laws for the functional increments of the uniform empirical process. In the same spirit, I proved a similar result for the local standard law of Mason [64] and for the functional limit law of Finkelstein [45]. Those are stated in the next two theorems.

Theorem 2 (Varron, 2011 [102]) *There exists a universal constant ϵ_0 >0 such that, for any choice of $\epsilon > \epsilon_0$ we have almost surely, for all large n :*

$$\frac{\alpha_n}{\sqrt{2\log\log(n)}} \in \ \mathcal{S}_2^{\epsilon(\log\log(n))^{-2/3}},$$
$$\frac{\beta_n}{\sqrt{2\log\log(n)}} \in \ \mathcal{S}_2^{\epsilon(\log\log(n))^{-2/3}}.$$

Theorem 3 (Varron, 2011 [102]) *Let h_n be positive real numbers satisfying, as $n \to \infty$,*

$$nh_n \uparrow \infty, \quad \frac{nh_n}{(\log\log(n))^{7/3}} \to \infty, \quad h_n \downarrow 0. \tag{3.2}$$

Then there exists a universal constant $\epsilon_1 > 0$ such that, for any choice of $\epsilon > \epsilon_1$ we have almost surely, for all large n :

$$\frac{\alpha_n(h_n\cdot)}{\sqrt{2h_n\log\log(n)}} \in \mathcal{S}_1^{\epsilon(\log\log(n))^{-2/3}}.$$

If, in addition, $nh_n/(\log\log(n))^{11/3} \to \infty$, then we have, almost surely, for all larg n :

$$\frac{\beta_n(h_n\cdot)}{\sqrt{2h_n\log\log(n)}} \in \mathcal{S}_1^{\epsilon(\log\log(n))^{-2/3}}.$$

Third tool : Small ball probabilities

The two preceding theorems clearly establish outer *clustering rates* in Strassen functional limit laws. Another interesting topic is to establish inner rates of approximations of fixed elements of $\mathcal{S}_1$ (or $\mathcal{S}_2$). Such results are known under the name of functional *Chung-type* *limit laws.* Those inner rates involve small ball probabilities, which have been intensively investigated for the Wiener process W. With the help of the works of de Acosta [2], Deheuvels [17] established a Chung-type limit law for $(\alpha_n(h_n\cdot))_{n\geq 1}$, by showing that, if h_n is a sequence of constants satisfying $nh_n \uparrow \infty$, $h_n \downarrow 0$ and $nh_n/(\log\log(n))^3 \to \infty$, we have, almost surely, for each $f \in \mathcal{S}_1$ satisfying $|| f ||^2_{Hilb} := \int_0^1 f'^2(t)dt < 1$:

$$\varliminf_{n\to\infty} (\log\log(n))\Big|\Big|\frac{\alpha_n(h_n\cdot)}{\sqrt{h_n}b_n} - f\Big|\Big|_{[0,1]} = \frac{\pi}{4\sqrt{1-|| f ||^2_{Hilb}}}.$$

His result, however, left open the problem of finding inner rates when $|| f ||^2_{Hilb}= 1$.
Since the works of Acosta, several researchers managed to obtain small ball probability estimates for elements f pertaining to the border of $\mathcal{S}_1$. Such estimates were obtained for two particular classes of functions, which are defined as follows.

Definition 3.1.1 *Denote by $\mathcal{S}_1^{BV}$ the subset of $\mathcal{S}_1$ of functions f for which $\int_0^1 f'^2(t)dt = 1$, and for which f' admits a version which has a bounded variation on $[0,1]$.*
Denote by $\mathcal{S}_1^{LIV}$ the subset of $\mathcal{S}_1$ of functions f for which $\int_0^1 f'^2(t)dt = 1$, and for which f' admits a version which is locally of infinite variation, namely : there exist $x_1,\ldots,x_k \in [0,1]$ such that, for $\epsilon > 0$ we have ($\mathcal{V}$, denoting the total variation on $[0,1]$)

$$\mathcal{V}\Big(\sum_{i=1}^k f'\mathbb{1}_{\{[0,1]-[x_i-\epsilon,x_i+\epsilon]\}}\Big) < \mathcal{V}\Big(\sum_{i=1}^k f'\mathbb{1}_{\{[x_i-\epsilon,x_i+\epsilon]\}}\Big) = \infty.$$

The works of Csáki [14], Grill [48], Gorn and Lifshits [47], and Berthet and Lifshits [11] on small ball probabilities for Wiener processes can be summed up in a single inequality.

Inequality 3.1.1 (From Csáki, Grill,Gorn, Berthet, Lifshits [14, 48, 47, 11]) *For any $f \in \mathcal{S}_1^{BV} \cup \mathcal{S}_1^{LIV}$, there exists a function $\nabla_f(\cdot)$ from $]0,\infty[$ to itself, as well as a constant $\chi_f > 0$, such that the following assertion holds : for each $\delta > 0$, there exist $\gamma^+ = \gamma^+(\delta, f) > 0$ and $\gamma- = \gamma^-(\delta, f) > 0$ such that for all T sufficiently large :*

$$\mathbb{P}\Big(\nabla_f\Big(\frac{T^2}{2}\Big)\Big|\Big|\frac{W}{T} - f\Big|\Big|_{[0,1]} \leq (1+\delta)\chi_f\Big) \geq \exp\Big(-\frac{T^2}{2} + \gamma^+\frac{\nabla_f^2(T^2/2)}{T^2}\Big),$$

$$\mathbb{P}\Big(\nabla_f\Big(\frac{T^2}{2}\Big)\Big|\Big|\frac{W}{T} - f\Big|\Big|_{[0,1]} \leq (1-\delta)\chi_f\Big) \leq \exp\Big(-\frac{T^2}{2} - \gamma^-\frac{\nabla_f^2(T^2/2)}{T^2}\Big).$$

Note that ∇_f and χ_f do have explicit forms. However, for sake of simplicity, I chose not to give the details of their definitions here.
My contribution to Chung-Mogulskii functional limit laws is as follows. It is the expected counterpart of Berthet's result in the global standard functional limit laws [9].

Theorem 4 (Varron, 2011 [102]) *Let $f \in \mathcal{S}_1^{bv} \cup \mathcal{S}_1^{liv}$ be arbitrary and let h_n be a sequence of real numbers satisfying, as $n \to \infty$,*

$$nh_n \uparrow \infty, \ h_n \downarrow 0,$$

$$\lim_{n\to\infty} \frac{nh_n}{\log\log(n)\nabla_f^2(\log\log(n))} = \infty. \tag{3.3}$$

Then we have, almost surely :

$$\underline{\lim}_{n\to\infty} \nabla_f(\log\log(n))\Big|\Big|\frac{\alpha_n(h_n\cdot)}{\sqrt{2h_n\log\log(n)}} - f\Big|\Big|_{[0,1]} = \chi_f.$$

Some words about the proofs

In each theorem, the growth assumption made upon nh_n is slightly stronger than the usual condition $nh_n/\log\log(n) \to \infty$. The sequence $\log\log(n)$ is replaced by a power of $\log\log(n)$, similarly as in [8, 9]. We imposed that condition so that the KMT strong approximation can be sharp enough to be negligible in front of the small ball/ small enlargements Gaussian analysis. Beside the crucial use of the KMT approximation and the above-mentioned Gaussian inequalities, the proof is routine use of *blocking* and *stochastic renewal* arguments, with contextual minor improvements in regard of the existing literature. Note that, in the original article [102], I needed to impose an additional assumption on h_n, by requiring that this sequence is not too slow, namely

$$h_n \log\log(n) \to 0. \tag{3.4}$$

This assumption was imposed in order to use a Poissonization technique ([102, Lemma 3.2]), which was later made obsolete by the Poissonization written in Proposition 2.2.2. That later Proposition could then get rid of (3.4). To conclude this paragraph, the results involving β_n and $\beta_n(h_n\cdot)$ are obtained by the Bahadur-Kiefer representation [54] and its local version [39, Theorem 5].

3.1.4 Perspectives

Berthet [10] showed that, for the problem of small balls around functions $f \in \mathcal{S}_1^{BV}$, the KMT approximation could be bypassed by making use of estimates of small ball probabilities for Poisson processes with high intensity, due to Shmileva [84]. Using those estimates, and applying them in the context of the global standard functional limit law for the increments of α_n, Berthet [10] could exhibit an uncrossable lower bound for the strong approximation of α_n by Brownian bridges, namely :

Fact 3.1.4 (Berthet, 2010 [10]) *Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, let $(U_i)_{i\geq 1}$ be a sequence of i.i.d. random variable uniformly distributed on $[0,1]$, let $(\alpha_n)_{n\geq 1}$ be the corresponding sequence of uniform empirical processes, and let B_n be a sequence of Brownian bridges on $(\Omega, \mathcal{A}, \mathbb{P})$. Then* *we have, almost surely :*

$$\underline{\lim}_{n\to\infty} \frac{\sqrt{n}}{\log(n)} \mid\mid \alpha_n - B_n \mid\mid_{[0,1]} \geq \frac{1}{6}.$$

That lower bound is explained by the fact that α_n does have an intrinsic dissymmetry at small scales, which makes him approximate more closely the target functions f which satisfy $\int f'(t)^3 dt > 0$ than a Brownian bridge does. That dissymmetry comes from the fact that α_n is the difference between the *stepwise increasing* function F_n and the *continuously decreasing* identity function on $[0, 1]$.
For the local empirical process $\alpha_n(h_n\cdot)$ Mason and van Zwet [66] did prove a strong approximation result which takes in account the fact that $h_n \to 0$, and hence is more precise that using $B_n(h_n\cdot)$ in the KMT approximation (recall Fact 3.1.2).
An interesting perspective is to investigate how the estimates of Shmileva [84] could give information on an uncrossable lower bound for the strong approximations of $\alpha(h_n\cdot)$.

3.1.5 Nonstandard functional limit laws

Let us close this parenthesis on the uniform empirical process, and consider again the general objects $T_n(\cdot, h, z)$ In this subsection, the centering in not necessary, so we will rather focus on the process

$$\widetilde{T}_n(g, h, z) := \sum_{i=1}^{n} g_{h,z}(Z_i)$$

When $H = \{z_0\}$ is a singleton, and when considering a bandwidths sequence h_n fulfilling $nh_n^d \sim c \log\log(n)$ for $c \in]0, \infty[$, the asymptotic behavior of $\widetilde{T}_n(\cdot, h_n, z_0)$ is not Gaussian anymore. Indeed, the inherent large deviation principle and limit set are related to large deviations of Poisson random measures.

Definition 3.1.2 (Rate function of Poisson type) *Let $\mathcal{G}$ be a class of functions fulfilling (Support). Define the following rate function on $\big(\ell^\infty(\mathcal{G}), ||\cdot||_{\mathcal{G}}\big)$.*

$$I_{\mathcal{G}}(\psi) := \inf\Big\{\int_{[-M,M]^d} \mathbf{h}_1\big(\mathfrak{g}\big) d\lambda,\ \forall g \in \mathcal{G},\ \psi(g) = \int_{[-M,M]^d} g\mathfrak{g} d\lambda\Big\},$$

where $\mathbf{h}_1(x) := x\log(x) - x + 1$ for $x \geq 0$ (with an extension by continuity at 0), and $\mathbf{h}_1(x) = \infty$ otherwise.
Define the associated levels

$$\begin{aligned}\Gamma_{\mathcal{G},a} :=& \Big\{\psi \in \ell^\infty(\mathcal{G}),\ I_{\mathcal{G}}(\psi) \leq a\Big\}\\ =& \Big\{g \to \int_{[-M,M]^d} \mathfrak{g} g d\lambda,\ \int_{[-M,M]^d} \mathbf{h}_1\big(\mathfrak{g}\big) d\lambda \leq a\Big\}.\end{aligned}$$

The first (local) nonstandard functional limit law was stated by Deheuvels and Mason [21], for the uniform empirical process. It was then extended by Deheuvels and Mason to a multivariate setting, with $\mathcal{G}$ being a class of indicators of sets which can be included in finite reunions of $||\cdot||_{\lambda,2}$ brackets generated by finite partitions of a set $[-M,M]^d$. (see, [24] for more details). Their two basic ingredients are a strong approximation by a sequence of Poissonized local empirical processes $\widetilde{\Pi}_n(\cdot,h_n,z_0)$, (which is possible since h_n tends rapidly to 0), then the use of the fact that, on an arbitrary partition $(A_1,\ldots,A_p)$ of $[-M,M]^d$ a Poisson random measure defines a mutually independent family. They then use that independence to *tensorize* the large deviation properties of each $\widetilde{\Pi}_n(\mathbb{1}_{A_j},h_n,z_0)$ and hence obtain a large deviation principle for the joint laws $\big(\widetilde{\Pi}_n(\mathbb{1}_{A_j},h_n,z_0)\big)_{j\leq p}$. With M. Maumy-Bertrand [68], we provided an extension of their result to processes of the type $T_{n,\mathfrak{c}}(\mathbf{g},\cdot,h,z)$, where

$$T_{n,\mathfrak{c}}(\mathbf{g},g,h,z) := \sum_{i=1}^{n} g_{h,z}(Z_i)\mathbf{g}(Y_i),$$

for a function $\mathbf{g}$ taking values in $\mathbb{R}^k$, for which we made assumptions of finite conditional exponential moments for $\mathfrak{g}(Y_1)$.
Such a nonstandard functional limit law also occurs when H is a compact with nonempty interior , as showed by Deheuvels and Mason [22] for the uniform empirical process (in which case it is required that $nh_n^d \sim c\log(n)$ instead of $c\log\log(n)$). My contribution (in [100]) is a generalization of the set-indexed local law of Deheuvels and Mason [24] to a global law. In its published form, it is restricted to the class

$$\mathcal{G}_{0,d} := \Big\{\mathbb{1}_{[0,t_1]\times\ldots\times[0,t_d]}(\cdot),\ (t_1,\ldots,t_d)\in[0,1]^d\Big\}.$$

Theorem 5 (Varron, 2010, [100]) *Assume that the law of Z_1 fulfills (Hf) for some open set $\mathfrak{D}$. Let $H\subset\mathfrak{D}$ be a compact set with nonempty interior. Let h_n be a bandwidths sequence such that $nh_n^d\sim c\log(n)$ for some $c\in]0,\infty[$. Then, almost surely, the two following assertions hold*

$$(i)\lim_{n\to\infty}\sup_{z\in H}\inf_{\psi\in\Gamma_{\mathcal{G}_{0,d},cf(z)}}\Big|\Big|\frac{T_n(\cdot,h_n,z)}{nh_n^d}-\psi\Big|\Big|_{\mathcal{G}_{0,d}}=0,$$

$$(ii)\forall z\in H,\ \forall\psi\in\Gamma_{\mathcal{G}_{0,d},cf(z)},\ \lim_{n\to\infty}\inf_{z\in H}\Big|\Big|\frac{T_n(\cdot,h_n,z)}{nh_n^d}-\psi\Big|\Big|_{\mathcal{G}_{0,d}}=0.$$

Corollary 5.1 *The preceding theorem still holds for a class of function $\mathcal{G}$ fulfilling $(Support)$, $(Bounded)$, and which admits finite $||\cdot||_{\lambda,2}$ bracketing numbers.*

Proof of the corollary : Identifying the $\widetilde{T}_n(\cdot,h_n,z)$ to random distribution functions on $[0,1]^d$, and by integration by parts, the preceding theorem can be directly generalized to arbitrary finite classes $\mathcal{G}$ of functions on $[0,1]^d$ that are of bounded variation (in the sense that they are themselves distribution functions of finite signed measures on $[0,1]^d$). This is a consequence of the fact that, for such a function g, the map $\psi \to \int_{[0,1]^d} \psi dg$ is continuous on $\big(\ell^\infty([0,1]^d), ||\cdot||_{[0,1]^d}\big)$. The support $[0,1]^d$ can be extended to any $[-M,M]^d$ with no efforts. Now, since $\mathbf{h}_1(x)$ is finite only for $x \geq 0$ the limit sets $\Gamma_{\mathcal{G},a}$, for any $\mathcal{G}$ fulfilling (*support*) and $a \geq 0$, contain elements of the form $\psi: \;\; g \to \int_{[-M,M]^d} \mathfrak{g} d d\lambda$, where $\mathfrak{g}$ is *nonnegative*. This automatically entails an ordered structure

$$g_1 \prec g \prec g_2 \Rightarrow \psi(g) \in [\psi(g_1), \psi(g_2)],$$

where $\prec$ stands for the partial order on real functions. The same remark holds for the processes $\widetilde{T}_n(\cdot, h_n, z)$. Hence, by a simple bracketing argument as in proof of the *Glivenko-Cantelli theorem under bracketing* (see, [92, p. 122]), Theorem 5 can be directly extended to any class of functions $\mathcal{G}$ satisfying (*Support*) and admitting finite $||\cdot||_{\lambda,2}$ bracketing numbers, as soon as the brackets $[g, g']$ are defined with g and g' of bounded variation. Any bounded measurable function on $[-M,M]^d$ can be $||\cdot||_{\lambda,2}$-approximated from below and from above by functions of bounded variation, so the brackets can be taken without imposing bounded variation.

3.1.6 Perspective

My main perspective is to investigate wether a condition on uniform covering numbers may lead to a variant of Theorem 5. The works of Menneteau [69], giving general conditions for the Poisson large deviations of local empirical measures may be a good starting point.

3.2 Uniform-in-bandwidth functional limit laws

In this section, I will give a description of my contributions to the study of the impact of the in-bandwidth uniformity to standard functional limit laws for the local empirical process. At this stage of the manuscript, it is convenient to remind the notion of Strassen-type sets.

Definition 3.2.1 (Strassen-type set) *Let $\mathcal{G}$ be a class of functions satisfying (Support) and (Unif. entropy). Recall that $J_{\mathcal{G}}$ has been defined in (2.4). Given $a \geq 0$, the associated Strassen-type set is defined by :*

$$\begin{aligned}\mathcal{S}_{\mathcal{G},a} :=& \Big\{\psi \in \ell^\infty(\mathcal{G}),\ J_{\mathcal{G}}(\psi) \leq a\Big\}\\ =& \Big\{g \to \int_{\mathbb{R}^d} \mathfrak{g} g d\lambda,\ \mathfrak{g} \text{ Borel},\ \int_{\mathbb{R}^d} \mathfrak{g}^2 d\lambda \leq a\Big\}.\end{aligned}$$

In the next results, we will make structural assumptions on $\mathcal{G}$ that may be slightly stronger that those used in the ULDP of Propositions 2.1.2 and 2.1.4.
Define

$$\mathcal{G}_\rho := \Big\{g\big(\lambda^{-1}\cdot\big) - g(\cdot),\ \lambda \in [1,\rho],\ g \in \mathcal{G}\Big\}, \tag{3.5}$$

$$\mathcal{G}_{+,\delta} := \Big\{ g\big(\,\cdot + u\big) - g(\cdot),\ \|\, u \,\|_d \leq \delta,\ g \in \mathcal{G} \Big\}. \tag{3.6}$$

The assumptions that will be made are (recalling (2.23)) :

$(Unif.\ entropy\ dilatations)$	There exists $\rho > 1$ such that $\mathcal{G}_\rho$ satisfies $\int_0^\infty \sup_{Q \text{ probab.}} \sqrt{\log N(\epsilon, \mathcal{G}_\rho, \|\cdot\|_{Q,2})} d\epsilon < \infty$;
$(Unif.\ entropy\ translations)$	There exists $\delta > 0$ such that $\mathcal{G}_{+,\delta}$ satisfies $\int_0^\infty \sup_{Q \text{ probab.}} \sqrt{\log N(\epsilon, \mathcal{G}_{+,\delta}, \|\cdot\|_{Q,2})} d\epsilon < \infty$;
$(Contin.\ dilatations)$	$\lim_{\rho \to 1} \Delta^2_{\mathcal{G}_\rho, 2} = 0$;
$(Contin.\ translations)$	$\lim_{\delta \to 0} \Delta^2_{\mathcal{G}_{+,\delta}, 2} = 0$.

Those assumptions allow to use the empirical processes theory in order to make small interpolations between finite grids of points, $z \in H$ or of bandwidths $h \in [h_n, \mathfrak{h}_n]$. As pointed out by Mason [65, Section 3], several classes of functions that are used in practice do satisfy those assumptions. Moreover, as mentioned in §2.1.5, all the results of the present section remain valid when the uniform entropy conditions are replaced by their bracketing counterparts.
In metric spaces, the topological limit of sequence of sets can be defined as follows.

Definition 3.2.2 (Inner and outer topological limits in a metric space) *Let (E, d) be a metric space, and let $(E_n)_{n \geq 1}$ be a sequence of subsets of E. The inner topological limit of $(E_n)_{n \geq 1}$ is defined as*

$$\underset{n\to\infty}{\underline{\lim}}^{top}\, E_n := \Big\{ e \in E,\ \lim_{n \to \infty} d(e, E_n) = 0 \Big\},$$

and the outer topological limit of $(E_n)_{n \geq 1}$ is defined as

$$\underset{n\to\infty}{\overline{\lim}}^{top}\, E_n := \Big\{ e \in E,\ \underline{\lim}_{n \to \infty} d(e, E_n) = 0 \Big\}.$$

When those two sets are equal, we denote them as the topological limit of $(E_n)_{n \geq 1}$.

3.2.1 Asymptotic independence in the local standard functional limit law

Proposition 2.1.4 has a direct consequence in terms of functional limit laws, when considering two local empirical processes sequences $T_n(\cdot, h_n, z_0)$ and $T_n(\cdot, \mathfrak{h}_n, z_0)$, with $h_n = o(\mathfrak{h}_n)$. Our next result is an unpublished generalization of a published result which was restricted to the local uniform empirical process. We will denote the bivariate strassen type sets as follows (recall that $J^{\otimes 2}_{\mathcal{G}}(\psi_1, \psi_2) := J_{\mathcal{G}}(\psi_1) + J_{\mathcal{G}}(\psi_2)$)

$$\mathcal{S}^{\otimes 2}_{\mathcal{G}, a} := \Big\{ (\psi_1, \psi_2) \in \ell^\infty(\mathcal{G}) \times \ell^\infty(\mathcal{G}),\ J^{\otimes 2}_{\mathcal{G}}(\psi_1, \psi_2) \leq a \Big\},\ a \geq 0.$$

Theorem 6 (Extension of Varron, 2006 [96]) *Assume that the law of Z_1 fulfills (Hf) for an open set $\mathfrak{O}$. Let $z_0 \in \mathfrak{O}$. Assume that $\mathcal{G}$ fulfills (Pointw. sep), (Bounded), (Support), (Unif. entropy dilatations) and (Contin. dilatations). Let $h_n \leq \mathfrak{h}_n$ be two subsequences of non negative numbers such that $h_n = o(\mathfrak{h}_n)$ and*

$$\mathfrak{h}_n \downarrow 0,\ nh_n^d \uparrow,\ \frac{nh_n^d}{\log\log(n)} \to \infty.$$

Then the sequence

$$\left(\frac{T_n(\cdot, h_n, z_0)}{\sqrt{2f(z_0)nh_n^d \log\log(n)}}, \frac{T_n(\cdot, \mathfrak{h}_n, z_0)}{\sqrt{2f(z_0)n\mathfrak{h}_n^d \log\log(n)}}\right)_{n\geq 2}$$

almost surely admits $\mathcal{S}_{\mathcal{G},1}^{\otimes 2}$ as cluster set in $\big(\ell^\infty(\mathcal{G}), ||\cdot||_{\mathcal{G}}\big) \times \big(\ell^\infty(\mathcal{G}), ||\cdot||_{\mathcal{G}}\big)$.

Remark : Theorem 6 can be also stated in terms of almost sure topological limits. It states that, almost surely, the outer topological limit of the sequence of singletons

$$\Theta_n^{local\otimes 2}(h_n, \mathfrak{h}_n, z_0) := \left\{\left(\frac{T_n(\cdot, h_n, z_0)}{\sqrt{2f(z_0)nh_n^d \log\log(n)}}, \frac{T_n(\cdot, \mathfrak{h}_n, z_0)}{\sqrt{2f(z_0)n\mathfrak{h}_n^d \log\log(n)}}\right)\right\},$$

is $\mathcal{S}_{\mathcal{G},1}^{\otimes 2}$, while its inner topological limit is $\emptyset$.

Some words about the proof

The essential part of the proof relies on a ULDP which is a consequence of Proposition 2.1.4, taking $H = \{z_0\}$ and $\epsilon_k = 0$. We then use *blocking* and *stochastic renewal* arguments. By looking at the definition of $\mathcal{S}_{\mathcal{G},1}^{\otimes 2}$, Theorem 6 can be roughly interpreted as follows : two local empirical processes at the same point z_0, but at radically different scales, cannot simultaneously hit (neighborhoods of) functions both having a high stochastic cost. If the first hits a (neighborhood of) function ψ with $J_{\mathcal{G}}$ close to 1, then the second has to compensate by hitting (neighborhood of) functions with low stochastic costs.

3.2.2 In bandwidth uniformity in the local standard functional limit law

For stating local standard functional limit laws, we will use the notations :

$$\Theta_n^{local}(\mathbf{h}, \mathfrak{h}, z_0) := \left\{\frac{T_n(\cdot, h, z_0)}{\sqrt{2f(z_0)nh^d \log\log(n)}},\ h \in [\mathbf{h}, \mathfrak{h}]\right\}\ n \geq 2,\ 0 < \mathbf{h} \leq \mathfrak{h}.$$

I will now state my contribution. In order to gain in concision and clarity, I will formulate it in a very slightly weaker form than that written in [108], in order to emphasize the notions of inner and outer topological limits.

Theorem 7 (Varron, 2014, from [108]) *Assume that the law of Z_1 fulfills (Hf). Assume that $\mathcal{G}$ fulfills (Pointw. sep.), (Bounded), (Support), (Unif. entropy dilatations), (Unif. entropy translations), (Contin. dilatations) and (Contin. translations). Let $H = \{z_0\}$ be a singleton. Let $h_n \leq \mathfrak{h}_n$ be two subsequences of non negative numbers such that, for some $\delta \in (0,1]$, we have :*

$$\mathfrak{h}_n \downarrow 0,\ nh_n^d \uparrow,\ \frac{nh_n^d}{\log\log(n)} \to \infty,\ \frac{\log\log(\mathfrak{h}_n/h_n)}{\log\log(n)} \to \delta. \tag{3.7}$$

Then, almost surely, the inner and outer topological limits of $\big(\Theta_n^{local}(h_n, \mathfrak{h}_n, z_0)\big)_{n\geq 2}$ in $\big(\ell^\infty(\mathcal{G}), ||\cdot||_{\mathcal{G}}\big)$ are $\mathcal{S}_{\mathcal{G},\delta}$ and $\mathcal{S}_{\mathcal{G},1+\delta}$.

Sketch of the proof and comments

Let us point out some key arguments of that proof.

Outer bounds

Consider, for arbitrary $\epsilon > 0$, the closed set $F := \ell^\infty(\mathcal{G}) - \mathcal{S}^{2\epsilon}_{\mathcal{G},1+\delta}$, which satisfies $J_{\mathcal{G}}(F) > 1 + \delta + 3\tau$ for some $\tau > 0$ (this comes from the fact that $J_{\mathcal{G}}$ is a good rate function). We will focus on finding upper bounds for the probabilities $\mathbb{P}(A_n)$, where

$$A_n := \bigcup_{h\in[h_n,\mathfrak{h}_n]} \Big\{ \frac{T_n(\cdot, h, z_0)}{\sqrt{2f(z_0)nh^d \log\log(n)}} \in F \Big\}.$$

A natural way to handle this probability of union is to discretize $[h_n, \mathfrak{h}_n]$ into a finite grid,

$$\begin{aligned} h_{n,\ell} :=& \rho_n^\ell h_n,\ \ell = 0,\dots,R_n,\ \text{where, } [\cdot] \text{ denoting the integer part} \\ R_n :=& \Big[\frac{\log(\mathfrak{h}_n/h_n)}{\log(\rho_n)}\Big], \end{aligned} \tag{3.8}$$

and then focus on $\mathbb{P}(A'_n)$, where

$$A'_n := \bigcup_{\ell=0}^{R_n} \Big\{ \frac{T_n(\cdot, h_{n,\ell}, z_0)}{\sqrt{2f(z_0)nh_{n,\ell}^d \log\log(n)}} \in F \Big\}.$$

This strategy is possible if we choose ρ_n such that $\rho_n \to 1$. In that case the oscillations between two consecutive elements $T_n(\cdot, h_{n,\ell}, z_0)$ and $T_n(\cdot, \rho_n h_{n,\ell}, z_0)$ are uniformly (in ℓ) negligible, because they are suprema of empirical processes on classes $\mathcal{G}_\rho$, with ρ close to 1. Those are handled by using Proposition 2.1.3, which is possible because (*Unif. entropy dilatations*) and

(*Contin. dilatations*) have been made. Now, using Bonferroni's inequality, we have

$$\begin{aligned}\mathbb{P}(A_n') \leq & (R_n+1) \max_{\ell=0,\,\dots,R_n} \mathbb{P}\Bigg(\frac{T_n(\cdot, h_{n,\ell}, z_0)}{\sqrt{2f(z_0) n h_{n,\ell}^d \log\log(n)}} \in F \Bigg) \\ \leq & (R_n+1) \exp\Bigg(- \big(J(F)-\tau\big) \log\log(n) \Bigg),\end{aligned}$$

where that last inequality holds for n large enough, by a use of Proposition 2.1.2 with $\tilde{n}_k := n$, $\tilde{h}_k := h_n$, $\tilde{\mathfrak{h}}_k := \mathfrak{h}_n$, $\epsilon_k = 0$, $v_k(h) = \log\log(n)$. A choice of $\rho_n := \exp(\log(n)^{-\tau})$ entails $\log(\rho_n) = \log(n)^{-\tau}$. Now, since $\log\log(\mathfrak{h}_n/h_n)/\log\log(n) \to \delta$ we have $R_n \leq \log(n)^{\delta}\log(n)^{\tau}$, from where :

$$\mathbb{P}(A_n') = O\big(\log(n)^{J(F)-2\tau}\big) = O\big(\log(n)^{-1+\tau}\big).$$

This sequence is summable along any subsequence $n_k = [(1+\gamma)^k]$, for arbitrarily small $\gamma > 0$. Now the interpolation between elements of that subsequence is made by making use of the usual *blocking arguments*. Hence, by the Borel-Cantelli lemma we have

$$\mathbb{P}\Bigg(\overline{\lim_{n\to\infty}} \bigcup_{h\in[h_n,\mathfrak{h}_n]} \Bigg\{ \frac{T_n(\cdot, h, z_0)}{\sqrt{2f(z_0) n h^d \log\log(n)}} \notin \mathcal{S}_{\mathcal{G},1+\delta}^{2\epsilon} \Bigg\} \Bigg) = 0,$$

which proves that, if it exists, the almost sure topological limit of $\Theta_n^{local}(h_n, \mathfrak{h}_n, z_0)$ is included in $\mathcal{S}_{\mathcal{G},1+\delta}$.

Inner bounds

It is already known ([43]) that, when $h_n = \mathfrak{h}_n$, the outer topological limit of $\Theta_n(h_n, \mathfrak{h}_n, z_0)$ (which, in this case, are singletons) is $\mathcal{S}_{\mathcal{G},1}$. So the outer bound $\mathcal{S}_{\mathcal{G},1+\delta}$ is either too rough (because of Bonferroni's inequality), or it has a stochastic explanation. In the next lines, we will provide such an explanation. I already evoked the *between bandwidths asymptotic independence phenomenon*, which was illustrated in Proposition 2.1.4 and Theorem 6. Note that the just mentioned results admit a natural extension to finite collections of local empirical processes. Since $h_n = o(\mathfrak{h}_n)$, it is possible choose $\rho_n \to \infty$ in the discretization of $[h_n, \mathfrak{h}_n]$. This would lead to the exhibition of finite grids $h_{n,\ell}$, $\ell = 0,\ R_n$, with $\max_{\ell \leq R_n} h_{n,\ell+1}/h_{n,\ell} \to 0$. Using the heuristic of independence between bandwidths, this would result in the fact that $T_n^{loc}(h_n, \mathfrak{h}_n, z_0)$ contains R_n "mutually independent" local empirical processes, with $R_n \to \infty$. Hence, even if ψ has a stochastic cost greater that 1 ($J_{\mathcal{G}}(\psi) > 1$) it is still probable enough that at least one of the R_n independent processes does hit a neighborhood of ψ (because "we run many independent trials"). Write ψ^{ϵ} for the open ball with center ψ and radius $\epsilon > 0$ in $\ell^{\infty}(\mathcal{G})$. First, since $J_{\mathcal{G}}$ is a rate function, we have, $J_{\mathcal{G}}(\psi^{\epsilon}) < 1 + \delta - 3\tau$, for some $\tau > 0$. We now explicitly choose

$\rho_n := \exp\big(\log(n)^\tau\big)$ in the discretization of $[h_n, \mathfrak{h}_n]$. Writing

$$B_n := \bigcup_{h\in[h_n,\mathfrak{h}_n]} \Big\{ \frac{T_n(\cdot,h,z_0)}{\sqrt{2f(z_0)nh^d\log\log(n)}} \in \psi^\epsilon \Big\},$$

we can make the following heuristic calculus

$$\begin{aligned}
&\mathbb{P}\bigg(\bigcup_{h\in[h_n,\mathfrak{h}_n]} \Big\{ \frac{T_n(\cdot,h,z_0)}{\sqrt{2f(z_0)nh^d\log\log(n)}} \in \psi^\epsilon \Big\}\bigg) \\
\geq &\mathbb{P}\bigg(\bigcup_{\ell=0}^{R_n} \Big\{ \frac{T_n(\cdot,h_{n,\ell},z_0)}{\sqrt{2f(z_0)nh_{n,\ell}^d\log\log(n)}} \in \psi^\epsilon \Big\}\bigg) \\
\simeq &1 - \prod_{\ell=0}^{R_n}\bigg(1-\mathbb{P}\Big(\frac{T_n(\cdot,h_{n,\ell},z_0)}{\sqrt{2f(z_0)nh_{n,\ell}^d\log\log(n)}} \in \psi^\epsilon \Big)\bigg) \text{ by "independence"} \qquad (3.9)\\
\geq &1-\exp\bigg(R_n \min_{\ell=0,\dots,R_n} \mathbb{P}\Big(\frac{T_n(\cdot,h_{n,\ell},z_0)}{\sqrt{2f(z_0)nh_{n,\ell}^d\log\log(n)}} \in \psi^\epsilon \Big)\bigg), \text{ since } 1-u\leq \exp(-u),\ u\in\mathbb{R} \\
\geq &\frac{1}{2}\wedge\bigg[\frac{1}{2}R_n \min_{\ell=0,\dots,R_n} \mathbb{P}\Big(\frac{T_n(\cdot,h_{n,\ell},z_0)}{\sqrt{2f(z_0)nh_{n,\ell}^d\log\log(n)}} \in \psi^\epsilon \Big)\bigg], \text{ since } 1-\exp(-u)\geq \frac{1}{2}\wedge\frac{u}{2},\ u\geq 0 \\
\geq &\log(n)^{\delta-\tau}\log(n)^{-1-\delta+2\tau} \\
= &\log(n)^{-1+\tau},
\end{aligned}$$

where, again, we used the ULDP of Proposition 2.1.2 and the fact that $\log(n)^{\delta-\tau} \leq R_n$ for all large n. Now by the usual *renewal arguments* used in Strassen laws, we can suppose that the B_n are "sufficiently independent" between sufficiently spread subsequences. This is for example the case for $n_k \sim \exp\big(\log(k)^{\tau/2}\big)$, for which we also have

$$\sum_{k\geq 1} \mathbb{P}(B_{n_k}) = \infty.$$

A use of the converse part of the Borel-Cantelli lemma then leads to

$$\mathbb{P}\Big(\overline{\lim_{k\to\infty}}\, B_{n_k}\Big) = 1,$$

which can conclude the proof of the "outer topological limit" part of Theorem 7.

Explanation of the heuristic of (3.9)

We will explain how the heuristic of (3.9) is justified. As mentioned earlier, Proposition 2.1.4 admits a generalization for more than two processes. Hence, that proposition would be the appropriate tool if R_n was bounded. Since it is not the case, we will approximate these local empirical processes by truly mutually independent processes. First, a use of the Poissonization technique of Proposition 2.2.2 allows to switch from $\big\{T_n(\cdot, h_{n,\ell}, z_0),\ \ell \leq R_n\big\}$ to $\big\{\Pi_n(\cdot, h_{n,\ell}, z_0),\ \ell \leq R_n\big\}$. However, the classes of functions

$$\mathcal{G}_\ell := \Big\{g_{h_{n,\ell},z_0},\ g \in \mathcal{G}\Big\},\ \ell = 0, \ldots, R_n$$

have their supports included respectively in $A(h_{n,\ell}, z_0)$, $\ell = 0, \ldots, R_n$ (recall (2.13)). This family of sets is not mutually disjoint, so we cannot directly use Proposition 2.2.1 to achieve rigorous independence. The last trick is to construct, by successive translations of respective lengths $2Mh_{n,\ell}$, a sequence of points $z_{n,\ell}$ such that, writing $\mid u \mid_d := \max\{\mid u_1 \mid, \ldots, \mid u_d \mid\}$:

$$\begin{aligned}
&z_{n,0} := z_0,\\
&\big(z_{n,\ell} + [-Mh_{n,\ell}, Mh_{n,\ell}]^d\big) \cap \big(z_{n,\ell'} + [-Mh_{n,\ell'}, Mh_{n,\ell'}]^d\big) = \emptyset, \text{ if } \ell \neq \ell', && (3.10)\\
&\forall \ell \leq R_n,\ \mid z_{n,\ell} - z \mid_d \leq 2M \sum_{\ell'=0}^{\ell-1} h_{n,\ell'}, \text{ from where}\\
&\delta_n := \max_{\ell=0,\ldots,R_n} h_{n,\ell}^{-1} \mid z_{n,\ell} - z_0 \mid_d \to 0. && (3.11)
\end{aligned}$$

For example, for $d = 1$, we can recursively set $z_{n,\ell+1} := z_{n,\ell} + 2Mh_{n,\ell}$. Now, by (3.10), the family of processes

$$\left\{\frac{\Pi_n(\cdot, h_{n,\ell}, z_{n,\ell})}{\sqrt{2f(z_0)nh_{n,\ell}^d \log\log(n)}}, \ell = 0, \ldots, R_n\right\}$$

is mutually independent. Moreover, the fact that we take local processes at points different from z_0 does not hurt for using large deviations, because the ULDP of Proposition 2.1.2 allows to take $z' \neq z$, a soon as $\mid\mid z' - z \mid\mid_d \leq \epsilon_n$, for $\epsilon_n \to 0$. Assertion (3.11) is largely sufficient to ensure these proximities between the $z_{n,\ell}$ and z_0 (take $\epsilon_n := \delta_n \mathfrak{h}_n$). What remains to show is a uniform bound for the following probabilities

$$\mathbb{P}\Bigg(\Big|\Big|\Pi_n(\cdot, h_{n,\ell}, z_{n,\ell}) - \Pi_n(\cdot, h_{n,\ell}, z_0)\Big|\Big|_{\mathcal{G}} \geq \epsilon\sqrt{2f(z_0)nh_{n,\ell}^d \log\log(n)}\Bigg),\ \ell = 0, \ldots, R_n,$$

or, equivalently, find bounds for

$$\mathbb{P}\Bigg(\Big|\Big|T_n(\cdot, h_{n,\ell}, z_{n,\ell}) - T_n(\cdot, h_{n,\ell}, z_0)\Big|\Big|_{\mathcal{G}} \geq \epsilon\sqrt{2f(z_0)nh_{n,\ell}^d \log\log(n)}\Bigg),\ \ell = 0, \ldots, R_n,$$

in which case the Poissonization occurs *after* the introduction of the $z_{n,\ell}$. Now since, for any $g \in \mathcal{G}$

$$g_{h_{n,\ell},z_{n,\ell}}(\cdot) = g_{h_{n,\ell},z_0}\big(\cdot + h_{n,\ell}^{-1}(z_{n,\ell} - z_0)\big),$$

we always have (recall (3.6)):

$$\begin{aligned}&\mathbb{P}\bigg(\Big|\Big|T_n(\cdot, h_{n,\ell}, z_{n,\ell}) - T_n(\cdot, h_{n,\ell}, z_0)\Big|\Big|_{\mathcal{G}} \geq \epsilon\sqrt{2f(z_0)nh_{n,\ell}^d \log\log(n)}\bigg)\\ \leq&\mathbb{P}\bigg(\Big|\Big|T_n(\cdot, h_{n,\ell}, z_0)\Big|\Big|_{\mathcal{G}_{+,\delta_n}} \geq \epsilon\sqrt{2f(z_0)nh_{n,\ell}^d \log\log(n)}\bigg),\end{aligned}$$

Now assumptions (*Unif. entropy translations*) and (*Contin. translations*) play their only role here, to control those deviations probabilities, uniformly in $\ell \leq R_n$, through the use of Proposition 2.1.3.□

3.2.3 Asymptotic independence in the global standard functional limit law

Before entering into the subject, I will briefly revisit an important result due to Mason : the global standard functional limit law for the local empirical process.

Removing the Euclidian condition in Mason's global standard functional limit law for the local empirical process

Now consider a compact $H \subset \mathfrak{O}$ with nonempty interior and consider, for a sequence of bandwidths h_n, the collections of processes :

$$\Theta_n^{global}(h_n, H) := \bigg\{\frac{T_n(\cdot, h_n)}{\sqrt{2f(z)nh_n^d \log(1/h_n^d)}},\ z \in H\bigg\}.$$

Mason [65] did prove a functional limit law for $\Theta_n^{global}(h_n, H)$, under the usual Csörgő-Révész-Stute conditions (see (3.12) below). His result was a major step. Indeed, in the setting of *global* functional limit laws, this result was the first to consider the general object $T_n(\cdot, h, z)$, instead of the uniform empirical process, bypassing the use of the KMT approximation. Beside assumptions (*Pointw. sep*), (*Bounded*), (*Support*), (*Contin. translations*) and (*Contin. dilatations*), he had to strengthen the structural assumptions (*Unif. entropy dilatations*) and (*Unif. entropy translations*) to an assumption of polynomial uniform entropy numbers, namely :

$$(Euclidian)\quad \exists C_0 > 0,\ v_0 > 0,\ \forall \epsilon \in (0,1),\ \sup_{Q\ probab} N(\epsilon, \mathcal{F}, ||\cdot||_{Q,2}) \leq C_0\epsilon^{-v_0},$$

where

$$\mathcal{F} := \Big\{g_{h,z}(\cdot),\ z \in \mathbb{R}^d,\ h > 0,\ g \in \mathcal{G}\Big\}.$$

That assumption, though more stringent that, e.g., ($Unif.$ *entropy*), still encompasses Vapnik-Chervonenkis classes of sets, or more generally uniformly bounded V-C subgraph classes of functions (see, e.g. [92, Chapter 2.6]). Mason's result is stronger than an almost sure topological limit, as it was already the case for the uniform empirical process [22]. It is a convergence for the Hausdorff distance.

Definition 3.2.3 (Hausdorff distance) *In a metric space* (E, d)*, the Hausdorff distance between sets is defined as*

$$d_{Haus}(A, B) := \max\Big\{ \sup_{\psi\in A} d(\psi, B); \sup_{\psi\in B} d(\psi, A)\Big\}.$$

Note that a sequence of sets (E_n) has topological limit E when $d_{Haus}(E_n, E) \to 0$, but the converse is not true.

Fact 3.2.1 (Mason, 2004, [65]) *Assume that the law of* Z_1 *fulfills* (Hf) *for a set* $\mathfrak{O}$*. Let* $H \subset \mathfrak{O}$ *be a compact set with nonempty interior. Assume that* $\mathcal{G}$ *fulfills* (*Pointw. sep*)*,* (*Bounded*)*,* (*Support*)*,* (*Euclidian*)*,* (*Contin. translations*) *and* (*Contin. dilatations*)*. Let* h_n *be a sequence of bandwitdhs fulfilling the Csörgő-Révész-Stute conditions :*

$$h_n \downarrow 0,\ nh_n \uparrow,\ \frac{nh_n^d}{\log(1/h_n^d)} \to \infty,\ \frac{\log(1/h_n)}{\log\log(n)} \to \infty.$$

Then, almost surely

$$\lim_{n\to\infty} d_{Haus}\Big(\Theta_n^{global}(h_n, H), \mathcal{S}_{\mathcal{G},1}\Big) = 0.$$

In a discussion following his theorem, Mason wrote that it was "not clear at all wether assumption (*Euclidian*) could be relaxed to a less stringent condition" (see his remarks 1 and 2 in [65]). My contribution to the problem is as follows. I did not send it to publication.

Proposition 3.2.1 (Varron, 2014) *Fact 3.2.1 is still true if* (*Euclidian*) *relaxed to* ($Unif.$ *entropy dilatations*) *and* ($Unif.$ *entropy translations*)*, or their bracketing counterparts.*

Sketch of the proof and comments

In this subsection I will explain the technical difficulty that led Mason to strengthen ($Unif.$ *entrpoy dilatations*) to (*Euclidian*), and how I could (somewhat simply) bypass the difficulty. Second, some aspects of his proof will come in handy to explain some heuristics of my contribution in the next paragraph.

Outer part

To prove the outer part of Fact 3.2.1, we need to control the probabilities

$$\mathbb{P}\Bigg(\exists z \in H,\ \frac{T_n(\cdot, h_n, z)}{\sqrt{2f(z)nh_n^d \log(1/h_n^d)}} \notin \mathcal{S}^{\epsilon}_{\mathcal{G},1}\Bigg).$$

The first idea is, for $n \geq 1$, to discretize H into a finite grid $z_{n,\ell}, \ell = 1, \ldots, M_n$, and hope to control the oscillations between the corresponding $(2f(z_{n,\ell})nh_n^d \log(1/h_n^d))^{-1/2} T_n(\cdot, h_n, z_{n,\ell})$. For some $\delta > 0$, we choose $z_{n,\ell},\ \ell = 1, \ldots, M_n$ such that

$$\forall z \in H,\ \exists \ell \leq M_n,\ \|\, z - z_{n,\ell} \,\|_d \leq \delta h_n, \tag{3.12}$$

hoping to control the oscillations between the grid. Note that we can choose $M_n \sim \lambda(H)(\delta h_n)^{-d}$ in (3.12).
By making use of, e.g., the ULDP of Proposition 2.1.2, with $v_k(h_n) := \log(1/h_n^d)$ we obtain, for some $\eta > 0$

$$\begin{aligned}
&\mathbb{P}\Bigg(\bigcup_{\ell=1}^{M_n} \Big\{\frac{T_n(\cdot, h_n, z_{n,\ell})}{\sqrt{2f(z_{n,\ell})nh_n^d \log(1/h_n^d)}} \notin \mathcal{S}^{\epsilon}_{\mathcal{G},1}\Big\}\Bigg) \\
\leq& M_n \max_{\ell=1,\ldots,M_n} \mathbb{P}\Bigg(\frac{T_n(\cdot, h_n, z)}{\sqrt{2f(z_{n,\ell})nh_{n,\ell}^d \log(1/h_{n,\ell}^d)}} \notin \mathcal{S}^{\epsilon}_{\mathcal{G},1}\Bigg) \\
\leq& \lambda(H)(\delta h_n)^{-d} \exp\Big(-(1+\eta)\log(1/h_n^d)\Big) \\
=& O(h_n^{d\eta}).
\end{aligned}$$

Since $\log(1/h_n)/\log\log(n) \to \infty$, that sequence is summable along any subsequence of the type $n_k \sim (1+\gamma)^k$ (the *blocking arguments* then play their role).
We now turn to controlling the oscillations between the $z_{n,\ell}$. Such a global oscillation is described by the following probabilities, for which it would be sufficient to prove that (if $\delta > 0$ is sufficiently small) those are $O(h_n^{\eta'})$, a $n \to \infty$, for some $\eta' > 0$.

$$\mathbb{P}\Bigg(\max_{\ell=1,\ldots,M_n} \sup_{\|z - z_{n,\ell}\|_d \leq \delta h_n} \frac{\Big\|T_n(\cdot, h_n, z) - T_n(\cdot, h_n, z_{n,\ell})\Big\|_{\mathcal{G}}}{\sqrt{2f(z_{n,\ell})h_n^d \log(1/h_n^d)}} \geq \epsilon\Bigg) \tag{3.13}$$

$$\leq \mathbb{P}\Bigg(\sup_{\substack{g \in \mathcal{G},\ z \in H, \\ \|z' - z\|_d \leq \delta h_n}} \Big|\sum_{i=1}^n g_{h_n,z}(Z_i) - \mathbb{E}\Big(g_{h_n,z}(Z_i)\Big)\Big| \geq \epsilon\sqrt{2\inf_{z \in H} f(z)nh_n^d \log(1/h_n^d)}\Bigg) \tag{3.14}$$

$$\leq \mathbb{P}\Bigg(\Big\|G_n(\cdot)\Big\|_{\mathcal{F}'_{h_n,\delta}} \geq \epsilon\sqrt{2\inf_{z \in H} f(z)h_n^d \log(1/h_n^d)}\Bigg)$$

$$\mathcal{F}'_{h_n,\delta} := \Big\{ g_{h_n,z}(\cdot) - g_{h_n,z}(\cdot + vh_n),\ g \in \mathcal{G},\ z \in H,\ || \ v \ ||_d \leq \delta \Big\}.$$

Since $\mathcal{F}$ satisfies $(Contin.\ dilatations)$, and by a usual change of variable, the class $\mathcal{F}'_{h_n,\delta}$ satisfies

$$\sup_{g \in \mathcal{F}'_{h_n,\delta}} \mathrm{Var}\big(g(Z_1)\big) \leq \mathbf{r}(\delta) h_n^d,$$

where $\lim_{\delta \to 0} \mathbf{r}(\delta) = 0$ and where $\mathbf{r}$ does not depend upon n. In view of the concentration inequality of Talagrand (Fact 2.1.2, with $t \sim \sqrt{nh_n^d \log(1/h_n^d)}$), appropriate exponential bounds will be achieved if it is proved that

$$\mathbb{E}\left(\Big|\Big| G_n(\cdot) \Big|\Big|_{\mathcal{F}'_{h_n,\delta}} \right) \leq \mathbf{r}'(\delta) \sqrt{h_n^d \log(1/h_n^d)}. \tag{3.15}$$

where $\lim_{\delta \to 0} \mathbf{r}'(\delta) = 0$. Mason used assumption $(Euclidian)$ to prove (3.15). Very roughly speaking, since $\sup_{Q\ probab} N(\epsilon, \mathcal{F}, || \cdot ||_{Q,2}) \leq C_0 \epsilon^{-v_0}$ for all $\epsilon > 0$, we have, for some universal constants C and C' :

$$\begin{aligned} &\int_0^{h^d} \sqrt{\log \sup_{Q\ probab} N(\epsilon, \mathcal{F}, || \cdot ||_{Q,2})} d\epsilon \\ &\leq C \int_0^{h^d} \sqrt{\log(1/\epsilon)} \\ &\leq C' \sqrt{h^d \log(1/h^d)}, \end{aligned}$$

for small values of h. All the subtilties to obtain a proper bound for (3.15) are due to Einmahl and Mason [42, Proposition A.1].
To sum up : inequality (3.14) involves suprema of empirical processes over a class of functions where the point z is not fixed, but varies on the whole set H, (contrarily to local functional limit laws). Hence it is impossible to handle the whole class through a local representation such as (2.17), and hence we have to consider empirical processes over classes of function of the form $g_{h,z}$, instead if the class of $\mathcal{G}$ directly. Moreover, without a local representation at hand such as (2.17), the considered empirical processes are built on sample of size n instead of $na(h,z) \sim f(z)nh^d$, which introduce another difficulty for controlling their first moments. That difficulty was overcome by Mason by strengthening $(Unif.\ entropy\, dilatations)$ to $(Euclidian)$.
Now, if we go back to (3.13), we can also bound

$$\mathbb{P}\left(\max_{\ell=1,\ldots,M_n} \sup_{||z - z_{n,\ell}||_d \leq \delta h_n} \frac{\Big|\Big| T_n(\cdot, h_n, z) - T_n(\cdot, h_n, z_{n,\ell}) \Big|\Big|_{\mathcal{G}}}{\sqrt{2f(z_{n,\ell}) h_n^d \log(1/h_n^d)}} \geq \epsilon \right)$$

$$\leq M_n \max_{\ell=1,\dots,M_n} \mathbb{P}\left(\sup_{||z-z_{n,\ell}||_d \leq \delta h_n} \frac{\left|\left|T_n(\cdot,h_n,z_{n,\ell}) - T_n(\cdot,h_n,z_{n,\ell})\right|\right|_{\mathcal{G}}}{\sqrt{2f(z_{n,\ell})h_n^d \log(1/h_n^d)}} \geq \epsilon \right)$$

$$= M_n \max_{\ell=1,\dots,M_n} \mathbb{P}\left(\frac{\left|\left|T_n(\cdot,h_n,z)\right|\right|_{\mathcal{G}_{+,\delta}}}{\sqrt{2f(z_{n,\ell})h_n^d \log(1/h_n^d)}} > \epsilon \right) \tag{3.16}$$

By this simple remark, we see that each of the M_n probabilities can be handled by a local representation around each $z_{n,\ell}$. More precisely, it is possible to use Proposition 2.1.3 uniformly in $z_{n,\ell}$, taking $u := \epsilon \log(1/h_n^d)$. Since Proposition 2.1.3 requires (*Unif. entropy*), it is clear that, under (*Unif. entropy translations*), we have :

$$M_n \max_{\ell=1,\dots,M_n} \mathbb{P}\left(\frac{\left|\left|T_n(\cdot,h_n,z)\right|\right|_{\mathcal{G}_{+,\delta}}}{\sqrt{2f(z_{n,\ell})h_n^d \log(1/h_n^d)}} > \epsilon \right)$$

$$\leq \lambda(H)\delta^{-d} h_n^{d(-1+\epsilon \Delta_{\mathcal{G}_{+,\delta}}^{-2})} \tag{3.17}$$

By (*Contin. translations*) that last bound can be $O(h_n^{\eta'})$, as $n \to \infty$, for some $\eta' > 0$ if δ is chosen small enough in the construction of the $z_{n,\ell}$. This concludes my explanation on how to relax (*Euclidian*).

Inner part

I will now give the idea on how to prove the inner part of Fact 3.2.1. It is sufficient to prove that, for arbitrary $\psi \in \mathcal{S}_{\mathcal{G},1}$ the probabilities

$$\mathbb{P}\left(\bigcap_{z \in H} \left\{ \frac{T_n(\cdot,h_n,z)}{\sqrt{2f(z)nh_n^d \log(1/h_n^d)}} \notin \psi^\epsilon \right\} \right)$$

are summable in n (no subsequence argument is involved here). The idea to prove that summability is that $\Theta_n^{global}(h_n, H)$ contains an increasing number of "almost mutually independent" processes. Contrarily to what happens in Theorem 7, the *independence* does not happens between bandwidths, but *between the points* z. Suppose, that instead of local empirical processes, we were studying their Poissonized versions (this is possible by the usual Poissonization of Einmahl, see, Proposition 2.2.3). Then, by (*Support*), if two points z, z' have distance $|| z - z' ||_d \geq 2Mh_n$, then, by Proposition 2.2.1, the processes $\Pi_n(\cdot, h_n, z)$ and $\Pi_n(\cdot, h_n, z')$ are rigorously independent. This can be generalized to finite collections of points $z_{n,\ell}$ which are at distance at least $|| z - z' ||_d \geq 2Mh_n$ one from each other. Since H has nonempty interior, it is possible construct a number $M_n \sim \lambda(H)h_n^{-d}$ of such points. By independence we then have, writing $J_{\mathcal{G}}(\psi^\epsilon) = 1 - 2\eta$, with $\eta > 0$:

$$\begin{aligned}
&\mathbb{P}\Bigg(\bigcap_{z\in H}\Big\{\frac{\Pi_n(\cdot, h_n, z)}{\sqrt{2f(z)nh_n^d\log(1/h_n^d)}} \notin \psi^\epsilon\Big\}\Bigg)\\
\geq &\mathbb{P}\Bigg(\bigcap_{\ell=1}^{M_n}\Big\{\frac{\Pi_n(\cdot, h_n, z_{n,\ell})}{\sqrt{2f(z_{n,\ell})nh_n^d\log(1/h_n^d)}} \notin \psi^\epsilon\Big\}\Bigg)\\
= &\prod_{\ell=1}^{M_n}\Bigg(1-\mathbb{P}\Big(\frac{\Pi_n(\cdot, h_n, z_{n,\ell})}{\sqrt{2f(z_{n,\ell})nh_n^d\log(1/h_n^d)}} \in \psi^\epsilon\Big)\Bigg)\\
\leq &\exp\Bigg(-M_n\min_\ell\mathbb{P}\Big(\frac{\Pi_n(\cdot, h_n, z_{n,\ell})}{\sqrt{2f(z_{n,\ell})nh_n^d\log(1/h_n^d)}} \in \psi^\epsilon\Big)\Bigg)\\
\leq &\exp\Big(-\lambda(H)h_n^{-d}h_n^{d(1-\eta)}\Big) \text{ by a ULDP}\\
\leq &\exp\Big(-\lambda(H)h_n^{-d\eta}\Big),
\end{aligned}$$

which is summable, since $\log(1/h_n)/\log\log(n) \to \infty$.

The asymptotic independence phenomenon in this framework

The first work of my PhD thesis was to generalize Fact 3.2.1 to couples $(T_n(\cdot, h_n, z), T_n(\cdot, \mathfrak{h}_n, z))$, with $h_n = o(\mathfrak{h}_n)$, generalizing a result of Deheuvels [18] for the uniform empirical process. This result was not sent for publication. Before stating it, I will need the definition of another type of Strassen set, given $\alpha \in]0, 1[$:

$$\mathcal{S}_{\mathcal{G}}^{\otimes 2}(\alpha) := \Big\{(\psi_1, \psi_2) \in \ell^\infty(\mathcal{G}) \times \ell^\infty(\mathcal{G}),\ J_{\mathcal{G}}(\psi_2) \leq 1,\ J_{\mathcal{G}}(\psi_1) + \alpha J_{\mathcal{G}}(\psi_2) \leq 1\Big\}.$$

Also define, for $n \geq 1$ and for two bandwidths $h_n \leq \mathfrak{h}_n$:

$$\Theta_n^{global,\otimes 2}(h_n, \mathfrak{h}_n, H) := \Big\{\Big(\frac{T_n(\cdot, h_n, z)}{\sqrt{2f(z)nh_n^d\log(1/h_n^d)}}, \frac{T_n(\cdot, \mathfrak{h}_n, z)}{\sqrt{2f(z)n\mathfrak{h}_n^d\log(1/\mathfrak{h}_n^d)}}\Big),\ z \in H\Big\}.$$

Theorem 8 (Varron, 2003, [94]) *Assume that the law of Z_1 fulfills (Hf) for a set $\mathfrak{O}$. Let $H \subset \mathfrak{O}$ be a compact set with nonempty interior. Assume that $\mathcal{G}$ fulfills (Pointw. sep.), (Bounded), (Support), (Unif. entropy dilatations), (Unif. entropy translations), (Contin. translations) and (Contin. dilatations). Let $h_n \leq \mathfrak{h}_n$ both fulfilling the Csörgő-Révész-Stute conditions :*

$$\mathfrak{h}_n \downarrow 0,\ nh_n^d \uparrow,\ \frac{nh_n^d}{\log(1/h_n^d)} \to \infty,\ \frac{\log(1/\mathfrak{h}_n)}{\log\log(n)} \to \infty.$$

Assume in addition that

$$\frac{\log(1/\mathfrak{h}_n)}{\log(1/h_n)} \to \alpha \in]0,1[.$$

Then, almost surely, $\left(\Theta_n^{global,\otimes 2}(h_n, \mathfrak{h}_n, H)\right)_{n\geq 2}$ *converges to* $\mathcal{S}_{\mathcal{G}}^{\otimes 2}(\alpha)$ *for the Hausdorff distance.*

Sketch of the proof and comments

The structure of the limit set $\mathcal{S}_{\mathcal{G},1}^{\otimes 2}(\alpha)$ depends on $\alpha := \lim_{n\to\infty} \log(1/\mathfrak{h}_n)/\log(1/h_n)$. That parameter measures how different the rates of convergence to 0 of h_n and $\mathfrak{h}_n$ are (typically, h_n and $\mathfrak{h}_n$ are two different powers of $1/n$). Let us consider the two extreme cases :

— When α gets close to 1, this means that those rates of convergence tend to be similar (but note that h_n is always $o(\mathfrak{h}_n)$, whatever the value of $\alpha < 1$). In that case, the limit set $\mathcal{S}_{\mathcal{G}}^{\otimes 2}(\alpha)$ is close to the limit set exhibited in Theorem 6. Theorem 8 then has the following rough interpretation : if, for some $z \in H$, the process $(2f(z)n\mathfrak{h}_n^d \log(1/\mathfrak{h}_n^d))^{-1/2}T_n(\cdot, \mathfrak{h}_n, z)$ hits a function ψ with stochastic cost $J_{\mathcal{G}}(\psi)$ close to 1, then $(2f(z)nh_n^d \log(1/h_n^d))^{-1/2}T_n(\cdot, h_n, z)$ (at the same point) has to compensate by hitting a function with low cost.

— When α gets close to 0, this means that h_n and $\mathfrak{h}_n$ have very different rates of convergence to zero. In that case, the limit set $\mathcal{S}_{\mathcal{G},1}^{\otimes 2}(\alpha)$ can have elements (ψ_1, ψ_2) with ψ_1 simultaneously having their respective stochastic costs close to 1. Hence, the compensation phenomenon softens : it may happens (in fact, it happens almost surely for all large n) that, for some $z \in H$, both the processes $(2f(z)n\mathfrak{h}_n^d \log(1/\mathfrak{h}_n^d))^{-1/2}T_n(\cdot, \mathfrak{h}_n, z)$ and $(2f(z)nh_n^d \log(1/h_n^d))^{-1/2}T_n(\cdot, h_n, z)$ simultaneously hit (neighborhoods of) functions having respective stochastic costs 1 and $1-\alpha$.

I will now give a heuristic argument to understand why this strange phenomenon occurs :

1. As showed in the heuristics of the proof of Fact 3.2.1, H contains, for each n, a number $M_n \sim \lambda(H)\mathfrak{h}_n^{-d}$ of points $z_{n,\ell}$ (distant one from each other of at least $\mathfrak{h}_n$), which define "mutually independent processes" $(2f(z_{n,\ell})n\mathfrak{h}_n^d \log(1/\mathfrak{h}_n^d))^{-1/2}T_n(\cdot, \mathfrak{h}_n, z_{n,\ell})$. Given ψ_2 fulfilling $J_{\mathcal{G}}(\psi_2) \leq 1$, the probability that at least one of these processes hits ψ_2 is high enough because "we run M_n independent trials".
2. But, for a given small value $\delta > 0$, since $\mathfrak{h}_n = o(\mathfrak{h}_n)$ it is possible, for each $z_{n,\ell}$, to construct a number $m_n \sim (\delta\mathfrak{h}_n/h_n)^d \sim \delta^d h_n^{-d(1-\alpha)}$ of points $z_{n,\ell,q}$, $q \leq m_n$, which are distant from each other of at least h_n (hence defining "independent" processes $(2f(z_{n,\ell,q})nh_n^d \log(1/h_n^d))^{-1/2}T_n(\cdot, h_n, z_{n,\ell,q})$ but which are still at distance at most $\delta\mathfrak{h}_n$ of their corresponding point $z_{n,\ell}$. That proximity with respect to the scale $\mathfrak{h}_n$ allows to approximate all the $(2f(\underline{z_{n,\ell,q}})n\mathfrak{h}_n^d \log(1/\mathfrak{h}_n^d))^{-1/2}T_n(\cdot, \mathfrak{h}_n, \underline{z_{n,\ell,q}})$, $q \leq m_n$ by $(2f(\underline{z_{n,\ell}})n\mathfrak{h}_n^d \log(1/\mathfrak{h}_n^d))^{-1/2}T_n(\cdot, \mathfrak{h}_n, \underline{z_{n,\ell}})$.
3. Roughly speaking, each $z_{n,\ell}$ is accompanied by $m_n \sim \delta^d h_n^{-d(1-\alpha)}$ independent trials for the realization of $(2f(z_{n,\ell})nh_n^d \log(1/h_n^d))^{-1/2}T_n(\cdot, h_n, z_{n,\ell})$. Hence, even if, for fixed $\ell \leq M_n$ the process $(2f(z_{n,\ell})n\mathfrak{h}_n^d \log(1/\mathfrak{h}_n^d))^{-1/2}T_n(\cdot, \mathfrak{h}_n, z_{n,\ell})$ hits a target ψ_2 with high stochastic,

there are still $h_n^{-d(1-\alpha+J(\psi_1))}$ independent possibilities that $(2f(z_{n,\ell})nh_n^d\log(1/h_n^d))^{-1/2}T_n(\cdot,h_n,z_{n,\ell})$ hits a target ψ_1, which allows ψ_1 to have a non negligible stochastic cost, less than $1-\alpha$.

3.2.4 In bandwidth uniformity for the global standard limit law

My second main result concerning in-bandwidth uniformity is related to global standard functional limit laws for the local empirical process.

Theorem 9 (Varron, 2008 [98]) *Let $h_n \leq \mathfrak{h}_n$ be two bandwidths sequences fulfilling (3.12). Then, under the assumptions of Fact 3.2.1, we have, almost surely*

$$\lim_{n\to\infty}\sup_{h\in[h_n,\mathfrak{h}_n]} d_{Haus}\Big(\Theta_n^{global}(h,H),\mathcal{S}_{\mathcal{G},1}\Big)=0.$$

Later Deheuvels and Ouadah [26] did remove condition $\log(1/\mathfrak{h}_n)/\log\log(n)\to\infty$, for the uniform empirical process, at the price of relaxing the almost sure convergence to a convergence in probability. **Sketch of the proof and comments** :
Let us consider the outer part of the proof. As I tried to emphasize in a brief description of the proof of Fact 3.2.1, one of the key aspects of the problem is that, given a sequence h_n fulfilling (3.12), the probability of

$$E_n(\epsilon,h_n):=\Big\{\Theta_n^{global}(h_n,H)-\mathcal{S}_{\mathcal{G},1}^{\epsilon}\neq\emptyset\Big\}$$

is (for fixed $\epsilon>0$) of order h_n^{η}, for some $\eta>0$. Now, proving (3.18) would necessarily require a discretization of the intervals $[h_n,\mathfrak{h}_n]$. As it was already shown in the proof of Theorem 7, assumptions (*Unif. entropy dilatations*) and (*Contin. dilatations*) allow to control the oscillations between bandwidths of the form $[h,\rho h]$, as soon as $\rho>1$ is small enough. We will therefore define, for fixed $\rho>1$, the grid $h_{n,\ell}:=\rho^{\ell}h_n,\ \ell\leq R_n$, with $R_n\leq\log(\mathfrak{h}_n/h_n)/\log(\rho)$. Now, using the ULDP of Proposition 2.1.2, with $\tilde{n}_k:=n$ and $v_k(h):=\log(1/h^d)$ it is possible to uniformly control the probabilities of $E_n(\epsilon,h_{n,\ell}),\ \ell\leq R_n$, and hence use Bonferroni's inequality

$$\begin{aligned}
&\mathbb{P}\Big(\bigcup_{\ell=0}^{R_n}\Theta_n^{global}(h_{n,\ell},H)-\mathcal{S}_{\mathcal{G},1}^{\epsilon}\neq\emptyset\Big)\\
\leq&\sum_{\ell=0}^{R_n}\mathbb{P}\big(E_n(\epsilon,h_{n,\ell})\big)\\
\leq&\sum_{\ell=0}^{R_n}h_{n,\ell}^{\eta}\\
=&h_n^{\eta}\sum_{\ell=0}^{R_n}\rho^{\eta\ell}
\end{aligned}$$

$$\begin{aligned}&\le h_n^\eta \frac{\rho^{\eta(R_n+1)}}{\rho^\eta - 1}\\&=\frac{\rho^\eta}{\rho^\eta - 1} h_n^\eta \exp\big(\delta R_n \log(\rho)\big)\\&=\frac{\rho^\eta}{\rho^\eta - 1}\mathfrak{h}_n^\eta.\end{aligned}$$

Which is of the same order as if we had taken $h_n = \mathfrak{h}_n$. The same arguments hold for the inner part of the proof.
Hence, contrarily to what happens in the local standard functional limit law, summing all these probabilities does not hurt at all, since those probabilities form a geometric sequence. In other words the probability of $E_n(\epsilon, \mathfrak{h}_n)$ is preponderant in front of the others, which roughly speaking means that

$$\sup_{h\in[h_n,\mathfrak{h}_n]} d_{Haus}\Big(\Theta_n^{global}(h), \mathcal{S}_{\mathcal{G},1}\Big) \simeq d_{Haus}\Big(\Theta_n^{global}(\mathfrak{h}_n), \mathcal{S}_{\mathcal{G},1}\Big).$$

3.2.5 Perspectives

My main perspective is to investigate how these functional limit laws can be extended to classes that fail to be uniformly bounded. This perspective is completely embedded in the perspective of improvements of the ULDP of the preceding chapter (see §2.1.6).

3.2.6 A nonfunctional uniform-in-bandwidth limit law for an object related to nonparametric regression

Several techniques developed for the local empirical processes can be used out of the context of functional limit laws. This is typically the case for concentration inequalities coupled with control of first moments, as it was pointed out by Einmahl and Mason [42]. In this article, they could prove the exact almost sure rates of convergence for objects related to nonparametric density/ regression estimation, which is described as follows.
Given an i.i.d sample $(Y_i, Z_i)_{i=1,\dots,n}$ taking values in $\mathbb{R}^{d'} \times \mathbb{R}^d$, with the same distribution as a vector (Y, Z), we assume that the law of (Y, Z) satisfies (Hf) for some product of open sets $\mathfrak{O} = \mathfrak{O}_1 \times \mathfrak{O}_2$, with $\mathfrak{O}_1 \subset \mathbb{R}^{d'}$ and $\mathfrak{O}_2 \subset \mathbb{R}^d$. Now write

$$V_n(g,h,z) := f_Z(z)^{-1/2}\sum_{i=1}^n \Big[\big(c_g(z)\bullet g(Y_i) + d_g(z)\big)K_{h,z}(Z_i) - \mathbb{E}\Big(\big(c_g(z)\bullet g(Y_i) + d_g(z)\big)K_{h,z}(Z_i)\Big)\Big], \tag{3.18}$$

where
- K denotes a kernel
- $h > 0$ is a bandwidth
- g is a Borel function from $\mathbb{R}^{d'}$ to $\mathbb{R}^k$

— f_Z is (a version of) the density of Z
— c_g maps $\mathbb{R}^d$ to $\mathbb{R}^k$ and d_g maps $\mathbb{R}^d$ to $\mathbb{R}$
— The symbol $\bullet$ stands for the canonical inner product on $\mathbb{R}^k$.

As pointed out by Einmahl and Mason [42], those objects are of statistical interest, since they encompass the main stochastic term of several estimators, ranging from the Parzen-Rosenblatt density estimator to the kernel estimator of the conditional distribution functions.
Given a class $\mathcal{G}$, of functions from $\mathbb{R}^{d'}$ to $\mathbb{R}^k$, we will write, for $\ell = 1, \ldots, k$ the class $\mathcal{G}_\ell := \pi_\ell(\mathcal{G})$, where $\pi_\ell(x_1, \ldots, x_\ell, \ldots, x_k) := x_\ell$ for $(x_1, \ldots, x_k) \in \mathbb{R}^k$. The following structural assumptions are commonly fulfilled in practice.

$(VCsubgraph)$	Each class $\mathcal{G}_\ell$ is a pointwise separable VC subgraph,
$(Envelope)$	Each class $\mathcal{G}_\ell$ has a finite valued measurable envelope function G_ℓ satisfying, for some $p \in (2, \infty]$: $\max_{\ell=1,\ldots,k} \sup_{y \in \mathfrak{O}_1} \|\| G_\ell(\cdot) \|\|_{\mathcal{L}_{Y\mid Z=z},p} < \infty$,

where $|| G_\ell(\cdot) ||_{\mathcal{L}_{Y|Z=z},p}$ is the L^p-norm of G_ℓ under the distribution of $Y \mid Z = z$ (which is uniquely defined if we use the continuous version of $f_{Y,Z}$). For a definition of a VC subgraph class, we refer, e.g., to [92, p.141].
The assumptions made upon the collections $\big(c_g(\cdot)\big)_{g\in\mathcal{G}}$ and $\big(d_g(\cdot)\big)_{g\in\mathcal{G}}$ are as follows.

$(H\mathcal{C})$ The classes of functions $\mathcal{D}_1 := \{c_g,\ g \in \mathcal{G}\}$ and $\mathcal{D}_2 := \{d_g,\ g \in \mathcal{G}\}$ are uniformly bounded and uniformly equicontinuous on $\mathfrak{O}_2$.

The assumption on the kernel K, involves the class.

$$\mathcal{K} := \Big\{K_{h,z},\ h > 0,\ z \in \mathbb{R}^d\Big\}.$$

$(HK1)$ K has bounded variation and the class $\mathcal{K}$ is VC subgraph,
$(HK2)$ $K(s) = 0$ when $s \notin [-1/2, 1/2]^d$,
$(HK3)$ $\int_{\mathbb{R}^d} K(s)ds = 1$.

Einmahl and Mason have studied the almost sure asymptotic behavior of $\sup_{z\in H} \mathbf{\Delta}_n(\mathcal{G}, h_n, z)$, with

$$\mathbf{\Delta}_n(\mathcal{G}, h, z) := \sup_{g\in\mathcal{G}} \frac{\mid V_n(g, h, z) \mid}{\sqrt{2f(z)nh^d \log(1/h^d)}}$$

(recall (3.18)), along a bandwidth sequence $(h_n)_{n\geq 1}$ that satisfies the following conditions

$$(HV)\quad h_n \downarrow 0,\ n{h_n}^d \uparrow \infty,\ \log(1/h_n)/\log\log(n) \to \infty,\ h_n^d\big(n/\log(1/h_n)\big)^{1-2/p} \to \infty,$$

where p is as in condition $(Envelope)$.Let us define

$$\Delta^2(g, z) :=\mathbb{E}\Big(\big(< c_g(z), g(Y) > +d_g(z)\big)^2 \mid Z = z\Big),\ z \in \mathbb{R}^d,\ g \in \mathcal{G}, \tag{3.19}$$

$$\Delta^2(\mathcal{G}, z) := \sup_{g\in\mathcal{G}\in H} \Delta^2(g, z),\ z \in H, \tag{3.20}$$
$$\Delta^2(\mathcal{G}) := \sup_{z\in H} \Delta^2(\mathcal{G}, z).$$

Under the above mentioned assumptions, Einmahl and Mason have proved the following theorem, λ denoting the Lebesgue measure.

Theorem 10 (Einmahl, Mason, 2000) *Assume that the law of (Y, Z) fulfills (Hf) for some product of open set $\mathfrak{O}_1 \times \mathfrak{O}_2$ with $\mathfrak{O}_1 \subset \mathbb{R}^{d'}$ and $\mathfrak{O}_2 \subset \mathbb{R}^d$. Assume that $(VC$ subgraph$)$, $(Envelope)$, $(H\mathcal{C})$ and $(HK1) - (HK3)$ are satisfied. Let h_n be a sequence fulfilling (HV). Then, given a compact $H \subset \mathfrak{O}_2$ with nonempty interior we have, almost surely :*

$$\lim_{n\to\infty} \sup_{z\in H} \mathbf{\Delta}_n(\mathcal{G}, h_n, z) = \Delta(\mathcal{G}) \mid\mid K \mid\mid_{\lambda,2} .$$

Their proof uses the same discretization techniques as in Theorem 7. The fact that the classes $\mathcal{G}_\ell$ are not necessarily bounded, but satisfy (*Envelope*) is handled by a truncating argument involving the Fuk-Nagaev inequality (see Fact 2.1.3).
With I. van Keilegom, we worked on proving the in-bandwidth uniform version of Theorem 10, by using discretizations of $[h_n, \mathfrak{h}_n]$ as in the proof of Theorem 9. We obtained the following result.

Theorem 11 (Varron, van Keilegom, 2011, [93]) *Assume that the law of (Y, Z) fulfills (Hf) for some product of open set $\mathfrak{O}_1 \times \mathfrak{O}_2$ with $\mathfrak{O}_1 \subset \mathbb{R}^{d'}$ and $\mathfrak{O}_2 \subset \mathbb{R}^d$. Assume that $(VC$ subgraph$)$, $(Envelope)$, $(H\mathcal{C})$ and $(HK1) - (HK3)$ are satisfied. Let $(h_n)_{n\geq 1}$ and $(\mathfrak{h}_n)_{n\geq 1}$ be two sequences of constants fulfilling (HV). Then, given a compact $H \subset \mathfrak{O}_2$ with nonempty interior we have, almost surely*

$$\lim_{n\to\infty} \sup_{h\in[h_n,\mathfrak{h}_n]} \sup_{z\in H} \Big| \frac{\mathbf{\Delta}_n(\mathcal{G}, h, z)}{\Delta(\mathcal{G}, z)} - \mid\mid K \mid\mid_{\lambda,2} \Big| = 0.$$

In fact, our result in [93] has a weaker form that assertion (3.21), but is can be extended with no major technical difficulty.
Since their initial work in [42], Einmahl and Mason (sometimes in collaboration with Dony) did prove in-bandwidth-uniform results for objects such as $V_n(g, h, z)$ (see [30, 31, 32, 34, 33]). Our contribution with regard of these results is that we could make *explicit* the limits and normalizations, at the price of slightly strengthening conditions upon h_n an $\mathfrak{h}_n$.

3.2.7 Perspectives

A first perspective of improvement will be to relax the $(VC$ *subgraph*$)$ assumption to a condition of the type $(Unif.\ entropy)$. At the light of the improvement of Fact 3.2.1, I am pretty certain that such an extension is reachable.
A second perspective of improvement would be to relax the assumption that the whole vector

(Y, Z) admits a density. That condition is somewhat unsatisfactory, because the smoothing is only made on the Z_i.
A third tough challenge is to establish *Bickel-Rosenblatt* limit laws for such estimators, namely proving that, for some sequence $a_n, b_n \to \infty$

$$a_n \left[\sup_{h\in[h_n,\mathfrak{h}_n]} \sup_{z\in H} \left| \mathbf{\Delta}_n(\mathcal{G}, h, z) - \Delta(\mathcal{G}, z) \parallel K \parallel_{\lambda,2} \right| \right] - b_n \to_{\mathcal{L}} \mathfrak{U},$$

where $\mathfrak{U}$ is a probability distribution on $\mathbb{R}$. Obtaining such results would have a more significant practical impact that the functional limit laws, since a convergence in distribution may allow calibrations. Such a limit law was proved by Bickel and Rosenblatt [12] for the Parzen-Rosenblatt density estimator, for univariate distributions (i.e. $\mathcal{G} = \{1\}$, $d = 1$ and $h_n = \mathfrak{h}_n$). In that case $a_n = \log(1/h_n^d)$ and $b_n = \log\log(1/h_n)$. In their proof they make use of two crucial tools

- The KMT strong approximation.
- The analysis of the distributions of suprema of Gaussian processes over increasing intervals, following ideas of Pickands [75].

It is essentially the KMT approximation that prevents a generalization to $d \geq 3$ of their methods. I am currently working on using the modern tools in empirical processes theory (concentration inequalities and approximations results of Zaitsev) to extend such a limit law to $d \geq 3$. I however, came across to a stumbling block, which can be described as follows.

- Poissonization works, so we can directly study the $\Pi_n(1, h_n, z)$.
- Such a level of precision imposes the analysis of

$$\max_{\ell=0,\ldots,M_n} \Pi_n(1, h_n, z_{n,\ell}), \tag{3.21}$$

  where, in the definition of the $z_{n,\ell}$, we choose $\delta_n \sim \delta \log(1/h_n^d)^{-1}$ instead of $\delta > 0$ (as in 3.12).
- The joint law of those M_n random variables cannot be approximated by the joint law of Gaussian random variables : the dimensionality explodes with n, which makes Zaitsev's bounds unusable. However, it is possible to make a sharp enough approximation separately on subblocks of cardinalities $\log(1/h_n^d)$. This defines $\delta^{-d} h_n^{-d} \log(1/h_n^d)$ sub blocks. The distances between two consecutive sub blocks is δh_n, with δ small. Hence, there is a non negligible dependency between the blocks.
- In their proof, Bickel and Rosenblatt could directly work on a joint Gaussian law for all those blocks (thanks to the KMT approximation). Even if those Gaussian blocks are not mutually independent, the authors could work on the covariance structure of these random variables, and conclude to a convergence of these maxima by making use of results of Pickands [75].
- Without a strong approximation, I do no know how to translate their Gaussian analysis (a control of the covariance structure) to a Poisson analysis. The underlying reason is

that, for $\int fgd\lambda = 0$ implies $\mathcal{W}_{\mathcal{G}}(f) \perp\!\!\!\perp \mathcal{W}_{\mathcal{G}}(g)$, while, for a Poisson process, such an independence is guaranteed only when f and g have disjoint supports.

3.3 Spatial-type limit laws

This section is dedicated to another type of functional limit laws (as well as a Donsker theorem) which are different from the usual functional limit laws for local empirical processes, in the sense that they are spatial.

3.3.1 A spatial Donsker theorem

First results for the uniform empirical process

Wschebor [109] discovered the following property of the Wiener process W on $[0, 1]$: almost surely, for each $0 \le a < b < 1$, and for each Borel set $B \subset \mathbb{R}$,

$$\frac{1}{(b-a)}\lambda\Big(\Big\{z \in [a,b],\ \epsilon^{-1/2}\big(W(z+\epsilon) - W(z)\big) \in B\Big\}\Big) \underset{\epsilon\to 0}{\to} \mathbb{P}\Big(\mathcal{N}(0,1) \in B\Big).$$

Recall that λ denotes the Lebesgue measure. That result was later extended to a much wider class of processes by Azaïs and Wschebor [6]. It can be interpreted as an almost sure spatial convergence in distribution. Here, the word "spatial" refers to the fact that it involves the almost sure limits of (Lebsesgue) measures of random sets of points of $[0, 1]$.
A first question of interest is to establish the same limit result for the *functional* increments of the Wiener process, namely

$$\epsilon^{1/2}\big(W(z+\epsilon\cdot) - W(z)\big) \in \ell^\infty([0,1]), \tag{3.22}$$

hence replacing a limit normal distribution by a Wiener measure.
A second natural question is to determine how this property of the Wiener process is shared by the empirical process α_n (see §3.1.3).
My first result in that direction [103] is stated as follows, writing

$$\Delta_n(\cdot, h_n, z) := h_n^{-1/2}\big(\alpha_n(z + h_n\cdot) - \alpha_n(z)\big),$$

as processes on $[0, 1]$, and writing λ^* for the outer Lebesgue measure on $[0, 1]$.

Theorem 12 (From Varron, 2011 [103]) *Assume that :*

$$h_n \downarrow 0,\ \ nh_n \uparrow \infty,\ \ \underline{\lim}_{n\to\infty} \log(1/h_n)/\log\log(n) > 1. \tag{3.23}$$

Then almost surely, for each interval $I = [a, b]$ with $0 \le a < b < 1$, the following assertion is *true :*

$$\text{For each closed set } F \text{ of } \Big(\ell^\infty([0,1]), ||\cdot||_{[0,1]}\Big), \text{ we have}$$
$$\overline{\lim_{n\to\infty}} \frac{\lambda^*\big(\{z\in I,\ \Delta_n(\cdot,h_n,z)\in F\}\big)}{\lambda(I)} \leq \mathbb{P}(W\in F).$$

Strong approximation under $||\cdot||_{[0,1]}$ (such as the KMT approximation) is a too heavy mathematical tool to handle this problem, since we only need to control Lebesgue measures. Hence, it is not surprising that the preceding theorem has a direct extension to the multivariate uniform empirical process on $[0,1]^d$, as it is stated in its complete form in [103].

Sketch of the proof and comments

In his original result, Wschebor used the property of independence of increments of the Wiener process. His proof is decomposed as follows : By a separability argument, we only need to prove that, for any fixed bounded Lipschitz function ϕ, from $\mathbb{R}$ to $\mathbb{R}$, we have

$$\lim_{\epsilon\to 0} Z(\phi,\epsilon) =_{a.s.} \mathbb{E}\Big(\phi(\mathfrak{Z})\Big)$$

with

$$Z(\phi,\epsilon) := \frac{1}{\lambda(I)}\int_I \phi\Big(\sqrt{\epsilon}\big(W(z+\epsilon)-W(z)\big)\Big)dz,$$

and with $\mathfrak{Z}$ denoting a standard normal random variable.

— (*Expectation*) The expectation of the random variables of interest is simply computed as

$$\begin{aligned}\mathbb{E}\Big(Z(\phi,\epsilon)\Big) =& \frac{1}{\lambda(I)}\int_I \mathbb{E}\bigg(\phi\Big(\sqrt{\epsilon}\big(W(z+\epsilon)-W(z)\big)\Big)\bigg)dz\\ =&\mathbb{E}\Big(\phi(\mathfrak{Z})\Big).\end{aligned}$$

Hence, the almost sure limit to establish is simply the expectation of the $Z(\phi,\epsilon)$, which does not depend upon ϵ.

— (*Variance*) Now the independence of increments plays its role in the computation of the related variance (equality (3.24) below) :

$$\begin{aligned}&\mathrm{Var}\Big(Z(\phi,\epsilon)\Big)\\ =&\frac{1}{\lambda(I)^2}\int_{I^2}\mathrm{Cov}\bigg(\phi\Big(\sqrt{\epsilon}\big(W(z+\epsilon)-W(z)\big)\Big),\phi\Big(\sqrt{\epsilon}\big(W(z'+\epsilon)-W(z')\big)\Big)\bigg)dzdz'\\ =&\frac{1}{\lambda(I)^2}\int_{|z-z'|\leq\epsilon}\mathrm{Cov}\bigg(\phi\Big(\sqrt{\epsilon}\big(W(z+\epsilon)-W(z)\big)\Big),\phi\Big(\sqrt{\epsilon}\big(W(z'+\epsilon)-W(z')\big)\Big)\bigg)dzdz'\end{aligned} \tag{3.24}$$

$=O(\epsilon)$, as $\epsilon \to 0$.

— ($Interpolation$) The preceding variance bound proves the required almost sure convergence along each sequence $\epsilon_n = n^{-(1+\delta)}$, with $\delta > 0$. Then, the almost sure Hölder property of W ensures the interpolation between the ϵ_n.

Those three key arguments can be extended with no efforts to obtain the same limit result for the functional increments in (3.22).
The idea of the proof of Theorem 12 is to use similar arguments for the $\Delta_n(\cdot, h_n, z)$.

— The counterpart of Step ($Expectation$) is simple : for each $z \in [0, b[$, and for $h_n \leq [1-b]$, the law of $\Delta_n(\cdot, h_n, z)$ is exactly the law of $\Delta_n(\cdot, h_n, 0)$, hence, writing, for any bounded Lipschitz function ϕ on $\big(\ell^\infty([0,1]), ||\cdot||_{[0,1]}\ \big)$:

$$Z_n(\phi, h_n) := \frac{1}{\lambda(I)} \int_I \phi\big(\Delta_n(\cdot, h_n, z)\big) dz, \tag{3.25}$$

we have

$$\begin{aligned}\mathbb{E}\Big(Z_n(\phi, h_n)\Big) =& \frac{1}{\lambda(I)} \int_I \mathbb{E}\Big(\phi\big(\Delta_n(\cdot, h_n, 0)\big)\Big) dz \\ =& \mathbb{E}\Big(\phi\big(\Delta_n(\cdot, h_n, 0)\big)\Big).\end{aligned}$$

But, since $nh_n \to \infty$, Donsker's theorem for the local uniform empirical process (see [64]) states that

$$\lim_{n\to\infty} \mathbb{E}\Big(\phi\big(\Delta_n(\cdot, h_n, 0)\big)\Big) = \mathbb{E}\Big(\phi\big(W(\cdot)\big)\Big),$$

which provides an adapted counterpart of ($Expectation$). Note that the preceding arguments *are not rigorous*, since $\phi\big(\Delta_n(\cdot, h_n, 0)\big)$ may fail to be Borel measurable. The proof can nevertheless be handled by considering only a subclass of measurable functions ϕ, involving evaluations at a countable number of elements of $\mathcal{G}$. However, for concision, I will skip all the technicalities used to tackle such a lack of measurability, and state all the subsequent arguments "as if $\phi\big(\Delta_n(\cdot, h_n, 0)\big)$ was measurable".

— The counterpart of ($Variance$) is based on Poissonization, in order to take advantage of the independent increments of a Poisson process on the real line. By a simple union argument, it can be supposed that $\mathbb{P}\big(U_1 \in I\big) < 1/2$ without losing generality. In that case, the random variables $Z_n(\phi, h_n)$ are entirely determined by the values U_i falling in I^{h_n}. In view of the Poissonization technique in Proposition 2.2.4, it is hence not surprising that one can Poissonize the variance of $Z_n(\phi, h_n)$ (beside some small technicalities that are voluntarily skipped - for more details, see [103, Proposition 3.2]). This leads to the following calculus, writing

$$\Delta\Pi_n(\cdot, h, z) := \frac{\sum_{i=1}^{\eta_n} \big(\mathbb{1}_{[z,z+h\cdot]}(U_i) - h\cdot\big)}{\sqrt{nh}},$$

where η_n is a Poisson random variable with expectation n and independent of $(U_i)_{i\geq 1}$.

$$\begin{aligned}&\mathrm{Var}\Big(\int_I \phi\big(\Delta\Pi_n(\cdot, h_n, z)\big)dz\Big)\\ =&\int_{I^2} \mathrm{Cov}\Big(\phi\big(\Delta\Pi_n(\cdot, h_n, z)\big), \phi\big(\Delta\Pi_n(\cdot, h_n, z')\big)\Big)dzdz'\\ =&O(h_n), \text{ by independence of the increments}\end{aligned}$$

— The counterpart of (*Interpolation*) is the use of the usual *blocking arguments*, which hold between consecutive terms of geometrical subsequences $n_k \sim (1+\gamma)^k$. Combining those arguments with the preceding variance calculus concludes the proof, but under the technical condition that h_{n_k} is summable in k for any geometric subsequences n_k. This explains why I had to impose condition $\underline{\lim}_{n\to\infty} \log(1/h_n)/\log\log(n) > 1$ in (3.23), which excludes sequences h_n tending too slowly to 0. I do not know if this condition can be bypassed or if some new phenomenon arises when, e.g., $\underline{\lim}_{n\to\infty} \log(1/h_n)/\log\log(n) < 1$.

Extension to more general local empirical processes

The preceding results was written here for the concision and clarity of the main arguments of its proof. I could, however, generalize it to the wider class of objects $T_n(\cdot, h_n, z)$, under assumptions that are very similar to those used in the results of §3.2. I could also use the same trick as in the proof of Theorem 9, to enrich the generalization with an additional uniformity in $h \in [h_n, \mathfrak{h}_n]$.

Theorem 13 (From Varron, 2014, [101]) *Assume that Z_1 satisfies (Hf) for some $\mathfrak{O}$. Let H be a set with nonempty interior fulfilling $\overline{H} \subset \mathfrak{O}$ ($\overline{H}$ denoting the closure of H). Let μ be a continuous probability measure on H having a version of Lebesgue density which is uniformly bounded. Let h_n and $\mathfrak{h}_n$ be two bandwidths sequences fulfilling*

$$\mathfrak{h}_n \downarrow 0,\ nh_n^d \uparrow \infty,\ \textit{as well as}$$

$$\underline{\lim}_{n\to\infty} \log(1/\mathfrak{h}_n^d)/\log\log(n) > 1.$$

Assume that $\mathcal{G}$ fulfills (Unif. entropy dilatations), (Contin dilatations), (Pointw. sep.) (Support), and admits an envelope that is square integrable under the Lebesgue measure. Then the following assertion holds with probability one : for each closed set F of $\big(\ell^\infty(\mathcal{G}), ||\cdot||_{\mathcal{G}}\big)$, we *have*

$$\overline{\lim_{n\to\infty}} \sup_{h\in[h_n,\mathfrak{h}_n]} \mu^*\left(\left\{z \in H,\ \frac{T_n(\cdot,h,z)}{\sqrt{f(z)nh^d}} \in F\right\}\right) \le \mathbb{P}\big(\mathcal{W}_{\mathcal{G}} \in F\big),$$

where $\mathcal{W}_{\mathcal{G}}$ *is the isonormal Gaussian process spanned by* $\big(\mathcal{G}, ||\cdot||_{\lambda,2}\big)$.

Note that the preceding theorem holds under conditions that are slightly weaker that those written here, but somewhat cumbersome to expose. For more details see [101].

Sketch of the proof an comments

Write, for a bounded Lipschitz function ϕ :

$$Z_n(\phi,h) := \int_H \phi\Big(\frac{T_n(\cdot,h_n,z)}{\sqrt{f(z)nh_n^d}}\Big) d\mu(z),$$

ignoring measurability issues for simplicity of the presentation.

— The usual *blocking arguments* still hold for the general objects $T_n(\cdot,h,z),\ h \in [h_n,\mathfrak{h}_n],\ z \in H$.

— The Poissonization technique of Giné et. al (Proposition 2.2.4) adapts with very few efforts to the objects $T_n(\cdot,h,z)$ (for more details, see [101, Proposition 2]).
By assumption (*Support*), we have $\Pi_n(\cdot,h,z) \perp\!\!\!\perp \Pi_n(\cdot,h,z')$ as soon as $||\ z-z'\ ||_d \ge 2Mh$. Hence, the ($Variance$) argument can be easily generalized. The assumption that μ admits a bounded density ensures that the measure of the diagonal $\mu^{\otimes 2}\big(\big\{(z,z') \in H^2,\ ||\ z-z'\ ||_d \le 2Mh\big\}\big)$ is $O(h^d)$ as $h \to 0$. Since $\mathfrak{h}_n \to 0$ we obtain

$$\mathrm{Var}\big(Z_n(\phi,h)\big) \le Ch^d, \text{ uniformly in } h \in [h_n,\mathfrak{h}_n],$$

for a constant C, and for n large enough. In addition, since the involved variances are of order h^d, it is possible to use the discretization procedure $h_{n,\ell} = \rho^\ell h_n,\ \ell \le R_n$ and sum up the corresponding variances to obtain a sum

$$\sum_{\ell=0}^{R_n} \mathrm{Var}\big(Z_n(\phi,h_{n,\ell})\big) \le \frac{C}{\rho^d - 1}\mathfrak{h}_n^d. \tag{3.26}$$

Hence requiring a uniformity in $h_{n,\ell}$ does not hurt. The interpolation arguments between consecutive $h_{n,\ell}$ are standard, and involve (*Unif. entropy dilatations*) and (*Contin. dilatations*).

— It remains to prove an adapted counterpart of the (*Expectation*) arguments. Consider the problem of proving

$$\begin{aligned}
&\mathbb{E}\Big(\phi(\mathcal{W}_{\mathcal{G}})\Big)\\
&= \lim_{n\to\infty} \mathbb{E}\Big(\int_I \phi\Big(\frac{T_n(\cdot,h_n,z)}{\sqrt{f(z)nh_n^d}}\Big)d\mu(z)\Big)\\
&= \lim_{n\to\infty} \int_I \mathbb{E}\Big(\phi\Big(\frac{T_n(\cdot,h_n,z)}{\sqrt{f(z)nh_n^d}}\Big)\Big)d\mu(z). \qquad (3.27)
\end{aligned}$$

The law of $T_n(\cdot,h_n,z)$ is not constant in $z \in H$, as it was the case for the uniform empirical process. However, by the dominated convergence theorem (under the measure μ) it is sufficient to prove that

$$\lim_{n\to\infty} \mathbb{E}\Big(\phi\Big(\frac{T_n(\cdot,h_n,z)}{\sqrt{f(z)nh_n^d}}\Big)\Big) = \mathbb{E}\Big(\phi\Big(\mathcal{W}_{\mathcal{G}}\Big)\Big),$$

for μ-almost each z. That convergence is simply Donsker's theorem for local empirical processes, which is a consequence of Einmahl and Mason [41].

Because our result involves a uniformity in $h \in [h_n, \mathfrak{h}_n]$, such a uniformity has to be added in (3.27), and hence raises additional technicalities that will not be detailed in this manuscript. For more details, see [101].

3.3.2 Spatial functional limit laws

Since the preceding results are spatial-type Donsker theorems, it seemed natural to me to investigate spatial-type standard/nonstandard functional limit laws. After first results concerning the uniform empirical process [103], I could obtain the following spatial standard functional limit law, which is the spatial counterpart of Theorem 7.

Theorem 14 (Varron, 2014 [104]) *Assume that the law of Z_1 satisfies (Hf) for some open set $\mathfrak{O}$. Let $H \subset \mathfrak{O}$ be a compact set and let μ be a probability measure on H. Let $h_n \leq \mathfrak{h}_n$ be two non random sequences fulfilling :*

$$(HV) \quad nh_n^d \uparrow,\ \mathfrak{h}_n \downarrow 0,\ nh_n^d/\log\log(n) \to \infty,\ \lim_{n\to\infty} \frac{\log\log(\mathfrak{h}_n/h_n)}{\log\log(n)} = \delta \in]0,1].$$

Then, under assumption (Bounded), (Support), (Pointw. sep.), (Unif. entropy dilatations), (Contin. dilatiations), (Unif. entropy translations), (Contin. translations), the following assertions hold with probability one :

$$\mathbb{P}\Bigg(\mu\Big(\Big\{z \in H,\ \overset{top}{\lim_{n\to\infty}} \Theta_n^{local}(h_n,\mathfrak{h}_n,z) = \mathcal{S}_{\mathcal{G}}^{1+\delta}\Big\}\Big) = 1\Bigg) = 1; \qquad (3.28)$$

$$\mathbb{P}\Bigg(\mu\bigg(\Big\{z \in H,\ \underset{n\to\infty}{\overset{top}{\lim}}\ \Theta_n^{local}(h_n, \mathfrak{h}_n, z) = \mathcal{S}_{\mathcal{G}}^{\delta}\Big\}\bigg) = 1\Bigg) = 1.$$

Sketch of the proof and comments

The arguments of §3.3.1 show that, for particular bounded functions ϕ, we have almost surely $Z_n(\phi, \mathfrak{h}_n) \to \mathbb{E}\Big(\phi\big(\mathcal{W}_{\mathcal{G}}\big)\Big)$. A functional limit law relies of large deviations, which provide rates of convergence to zero of probabilities of rare events. Hence, my first idea was to replace the function ϕ by sequences of functions ϕ_n of indicators of rare events (i.e. $\phi_n\big(T_n(\cdot, \mathfrak{h}_n, z)\big) \to 0$ almost surely) and see how the ($Expectation$) and ($Variance$) arguments could be translated in this context. This work took form in a first result for the uniform empirical process (see [103]). However, that methodology imposed the condition

$$\underline{\lim}_{n\to\infty} \frac{\log(1/\mathfrak{h}_n^d)}{\log\log(n)} > 2,$$

which was required to obtain better rates of approximations between $Z_n(\phi_n, \mathfrak{h}_n)$ and their expectations, in order to have

$$Z_n(\phi_n, \mathfrak{h}_n) = O_{a.s.}\Big(\mathbb{E}\big(Z_n(\phi_n, \mathfrak{h}_n)\big)\Big), \tag{3.29}$$

the latter tending to zero as $n \to \infty$. With approximation results such as (3.29), I could then obtain almost sure estimates for $Z_n(\phi_n, \mathfrak{h}_n)$ which are "almost sure μ-large deviations". I then used these large deviations under μ, coupled with a blocking argument to obtain such a spatial functional limit law. In a sense, I did my best to mimic the arguments of the usual Strassen laws, adapting them to a spatial setting.
Afterwards, when I tried to generalize my result to the more general objects $T_n(\cdot, \mathfrak{h}_n, z)$, I realized that I did not actually need to mimic those arguments. Indeed, there is a different way to prove it, by noticing that some crucial random sequences are almost surely monotonic, and hence only their convergences in probability need to be proved. Let us illustrate this through one particular step of the proof. To prove the inner part of (3.28), we need to prove that

$$\mu\bigg(\bigcup_{n_0\geq 1}\bigcap_{n\geq n_0}\bigcap_{h\in[h_n,\mathfrak{h}_n]}\Big\{z \in H,\ \frac{T_n(\cdot, h, z)}{\sqrt{2f(z)nh^d\log\log(n)}} \notin \psi^{\epsilon}\Big\}\bigg) = 0,\ a.s.$$

Since, for each n_0, the sequence

$$\mu\bigg(\bigcap_{n=n_0}^{n_0+p}\bigcap_{h\in[h_n,\mathfrak{h}_n]}\Big\{z \in H,\ \frac{T_n(\cdot, h, z)}{\sqrt{2f(z)nh^d\log\log(n)}} \notin \psi^{\epsilon}\Big\}\bigg).$$

is almost surely decreasing in p, it is sufficient to prove the convergence to 0 (as $p \to \infty$) of

$$\mathbb{E}\Bigg(\mu\Big(\bigcap_{n=n_0}^{n_0+p}\bigcap_{h\in[h_n,\mathfrak{h}_n]}\Big\{z\in H,\ \frac{T_n(\cdot,h,z)}{\sqrt{2f(z)nh^d\log\log(n)}}\notin\psi^\epsilon\Big\}\Big)\Bigg)$$
$$\leq \sup_{z\in H}\mathbb{P}\Big(\bigcap_{n=n_0}^{n_0+p}\bigcap_{\ell=0}^{R_n}\Big\{\frac{T_n(\cdot,h_{n,\ell},z)}{\sqrt{2f(z)nh_{n,\ell}^d\log\log(n)}}\notin\psi^\epsilon\Big\}\Big),$$

where we introduced a sufficiently spread grid $h_{n,\ell} := \rho_n^\ell h_n,\ \ell \leq R_n$, with $\rho_n \to \infty$. By the usual *stochastic renewal argument*, the first intersection symbol can be converted to a product of probabilities of independent events (along sufficiently spread subsequences n_k), this leads to the following heuristic :

$$\sup_{z\in H}\mathbb{P}\Big(\bigcap_{n=n_0}^{n_0+p}\bigcap_{\ell=0}^{R_n}\Big\{\frac{T_n(\cdot,h_{n,\ell},z)}{\sqrt{2f(z)nh_{n,\ell}^d\log\log(n)}}\notin\psi^\epsilon\Big\}\Big)$$
$$\leq \exp\Bigg(-\inf_{z\in H}\sum_{n=n_0}^{n_0+p}\mathbb{P}\Big(\bigcup_{\ell=0}^{R_n}\Big\{\frac{T_n(\cdot,h_{n,\ell},z)}{\sqrt{2f(z)nh_{n,\ell}^d\log\log(n)}}\in\psi^\epsilon\Big\}\Big)\Bigg). \tag{3.30}$$

Now the idea is to use the same arguments as those used to prove the heuristic (3.9), with the formal replacement of z_0 by z. An additional uniformity in $z \in H$ is required though, but that uniformity is perfectly handled by the ULDP and the Poissonization techniques of Propositions 2.1.2 and 2.2.2. The obtained lower bounds are large enough to ensure the required convergence to 0, along the sequence n_k used in the *stochastic renewal argument*. We can see that, using those kind of arguments, it is not needed to have precise almost sure estimates for random sequences of the form $Z_n\big(\phi_n, h_{n,\ell}\big)$. For example, looking at (3.30), it is sufficient to obtain lower large deviations bounds for $T_n(\cdot, h_{n,\ell}, z)$, which hold uniformly in $z \in H$ and $h_{n,\ell},\ \ell \leq R_n$. Since precise estimates are not required, so are the assumption on $\underline{\lim}_{n\to\infty}\log(1/\mathfrak{h}_n)/\log\log(n)$, and so is the assumption that μ has a bounded density, because the ($Variance$) step can be bypassed.□.

3.3.3 Perspectives

As already mentioned for problems involving the local empirical process, an important perspective would be to relax the (*Bounded*) assumptions to high-order integrability of the envelope.
Another perspective is as follows : if we compare the assumptions in Theorems 13 and 14, we can see that the second does not require μ to be continuous, nor any condition upon $\log(1/\mathfrak{h}_n)/\log\log(n)$. It is somewhat surprising because functional limit laws usually require

stronger assumptions than Donsker theorems. I will continue my investigations in the direction of removing those two conditions. Note, however, that Theorem 13 automatically excludes discrete measures with finitely many atoms, because le limit measure takes all possible values in $[0, 1]$.
A crucial perspective is to investigate the possible connections between those spatial limit laws and some statistical procedures.

Chapter 4

Contributions to the study of the general empirical process, and its applications to statistics

In this chapter, the central object will be the general empirical process, namely :

$$G_n(\cdot) := g \to \sqrt{n}\big(\mathbf{P}_n(g) - \mathbf{P}_0(g)\big),$$

as defined in the introduction. My works on that object can be decomposed into three parts :

— In Bayesian nonparametrics, several priors can be seen as random probability measures that are almost surely discrete. An almost surely discrete random probability measure can be seen as $\sum_{n\geq 1} \beta_i \delta_{Y_i}$, where both the probability weights $(\beta_i)_{i\geq 1}$ and the locations Y_i are random. Since the empirical measure $\mathbf{P}_n$, as well as its various bootstrapped versions (see, e.g., [92, Chapter 3.6]) belong to that class of objects, I investigated how far some usual techniques in empirical processes theory could be used in the study of that wider class of random measures. I found that, beside some non surprising envelope conditions, condition (*Unif. entropy*), with the presence of the envelope in the integrals, is robust enough to obtain Donsker and Glivenko-Cantelli theorems, when the Y_i are conditionally i.i.d. given the weights. Those results are stated in §4.1.

— The empirical likelihood technique, initiated by Thomas and Grunkemeier [89] is an alternative to Gaussian pivot or bootstrap procedures for the construction of confidence regions. Roughly speaking, it consists in building confidence regions around an estimator by slightly unbalancing the empirical weights $1/n$ in the definition of the estimator. With J-Y Dauxois and A. Flesch, we worked on using the empirical processes theory to prove that a general class of empirical likelihood procedures is consistent. Those results are stated in §4.2.

— The discrete associated kernel estimator is a general class of nonparametric estimators of the probability mass function of a discrete random variable. Despite the "kernel" terminology, such estimators are not related to the local empirical measure, due to the discrete nature of the data. With C. Kokonendji, we investigated the asymptotic properties of these estimators under the total variation distance. Those results are stated in §4.3.

4.1 A Donsker and a Glivenko-Cantelli theorem for a class of random measures related to the empirical measure

4.1.1 The framework

For $\mathbf{p} = (p_i)_{i\geq 1} \in \mathbb{R}^{\mathbb{N}}$ and $r > 0$, we shall write the (possibly infinite) value

$$\| \mathbf{p} \|_{c,r} := \Big(\sum_{i\geq 1} | p_i |^r \Big)^{1/r}.$$

Assume that $(\Omega, \mathcal{A}, \mathbb{P})$ is a complete probability space (in the sense that every $\mathbb{P}$-negligible set belongs to $\mathcal{A}$), and let

$$\mathbb{S} := \Big\{ \mathbf{p} = (p_i)_{i\geq 1} \in [0, \infty[^{\mathbb{N}},\ \| \mathbf{p} \|_{c,1} < \infty \Big\},$$

be the cone of positive summable sequences, which will be endowed with the product Borel σ-algebra (denoted by Bor). Consider a sequence of $\mathbb{S}$-indexed collections of probability measures $\big(\mathbf{P}_{n,\mathbf{p}}\big)_{n\geq 1, \mathbf{p}\in\mathbb{S}}$ on a measurable space $\big(\mathfrak{X}, \mathcal{X}\big)$. Note that those $\mathbf{P}_{n,\mathbf{p}}$ are *not random*. For fixed n, consider a random variable $\boldsymbol{\beta}_n = (\beta_{i,n})_{i\geq 1}$ from $(\Omega, \mathcal{A}, \mathbb{P})$ to $(\mathbb{S}, Bor)$ and a sequence $\mathbf{Y}_n = \big(Y_{i,n}\big)_{i\geq 1}$ of random variables from $(\Omega, \mathcal{A}, \mathbb{P})$ to $\big(\mathfrak{X}, \mathcal{X}\big)$, for which the conditional law given $\boldsymbol{\beta}_n = \mathbf{p}$ is $\mathbf{P}_{n,\mathbf{p}}^{\otimes\mathbb{N}}$ (we also make the assumption that those conditional laws properly define a Markov kernel). We then define the random element

$$Pr_n := \sum_{i\geq 1} \beta_{i,n} \delta_{Y_{i,n}},$$

as a map from Ω to $\mathfrak{M}$ (recall that the latter is the set of all probability measures on $\big(\mathfrak{X}, \mathcal{X}\big)$). Obtaining asymptotic results (as $n \to \infty$) for Pr_n is of interest for at least two reasons. The first of them is that Pr_n generalizes the usual empirical measure, and also related objects from bootstrap theory, such that the empirical bootstrap, or more generally, the exchangeable bootstrap empirical measure (see [92, Section 3.6.2]). Our main motivation to consider such a generalization comes from the second reason : these kind of random measures play a role in the representation of posterior distributions in Bayesian nonparametrics. The simplest can be described as follows : endow $\mathfrak{M}$ with the Borel σ-algebra $\mathcal{M}$ spanned by the weak topology. Given a probability measure $\overline{\alpha}$ on $\big(\mathfrak{X}, \mathcal{X}\big)$ and a concentration parameter $M > 0$, the Dirichlet process distribution $DP(M, \overline{\alpha})$ on $\big(\mathfrak{M}, \mathcal{M}\big)$ admits the following representation, due to Sethuraman [82] :

$$DP(M, \overline{\alpha}) =_d \sum_{i\geq 1} \beta_i \delta_{Y_i}.$$

Here, the Y_i are i.i.d with common distribution $\overline{\alpha}$ and $\beta_i := U_i \prod_{j=1}^{i-1}(1-U_j)$, with $(U_i)_{i\geq 1}$ being an i.i.d. sample, independent of $(Y_i)_{i\geq 1}$, and having the $Beta(1, M)$ distribution. It is also well known [44, Section 3, Theorem 1] that, if a random probability measure Pr has distribution $DP(M, \overline{\alpha})$ and if $(X_1, \ldots, X_n)$ has law $\mathbf{P}^{\otimes n}$ given $Pr = \mathbf{P}$ (for Pr almost all $\mathbf{P}$), then a version of the posterior distribution of Pr given $(X_1, \ldots, X_n)$ is the map

$$(x_1, \ldots, x_n) \to DP\Big(M + n, \theta_n \overline{\alpha} + (1-\theta_n)\overline{\alpha}_{(x_1,\ldots,x_n)}\Big),$$

where $\theta_n := M/(M+n)$ and $\overline{\alpha}_{(x_1,\ldots,x_n)} := n^{-1}\sum_{i\leq n} \delta_{x_i}$. A more general class of random measures admitting such a representation is the class of *normalized homogenous completely random measures* with no fixed atoms (see, e.g., [49, p. 84-85]).

4.1.2 A Glivenko-Cantelli theorem

We will consider a class of functions $\mathcal{F}$ which satisfies (*Pointw. sep.*). We will write F for its minimal measurable envelope. For $r > 0$, we define the space $\mathcal{E}_{\mathcal{F},r}$ as follows : a map $\Psi : \Omega \to \mathbb{R}^{\mathcal{F}}$ belongs to $\mathcal{E}_{\mathcal{F},r}$ if and only if $\Psi(f)$ defines a Borel random variable on $(\Omega, \mathcal{A}, \mathbb{P})$ for each $f \in \mathcal{F}$, and if $||| \Psi |||_{\mathcal{F},r}^r := \mathbb{E}\big(|| \Psi ||_{\mathcal{F}}^r \big) < \infty$. Under the assumption that $\mathbb{E}\big(F(Y_{1,n})\big) < \infty$ for all n, it is possible to define the process

$$G_{\mathbf{Y}_n, \boldsymbol{\beta}_n}(\cdot) :\ f \to \sum_{i\geq 1} \beta_{i,n}\Big(f(Y_{i,n}) - \mathbb{E}\big(f(Y_{i,n}) \mid \boldsymbol{\beta}_n\big)\Big)$$

as the limit, as $k \to \infty$, of the truncated sequence

$$G^k_{\mathbf{Y}_n, \boldsymbol{\beta}_n}(\cdot) :\ f \to \sum_{i=1}^{k} \beta_{i,n}\Big(f(Y_{i,n}) - \mathbb{E}\big(f(Y_{i,n}) \mid \boldsymbol{\beta}_n\big)\Big)$$

in the Banach space $\big(\mathcal{E}_{\mathcal{F},1}, ||| \cdot |||_{\mathcal{F}}\big)$. Note that this limit also holds in $\mathcal{E}_{\mathcal{F},r}$ for each $r \geq 1$ such that $\mathbb{E}\big(F(Y_{1,n})^r\big) < \infty$. My first contribution is a Glivenko-Cantelli theorem.

Theorem 15 (Varron, 2014, [105]) *Assume that* $|| \boldsymbol{\beta}_n ||_{c,1} = 1$ *almost surely for all* n, *and* $|| \boldsymbol{\beta}_n ||_{c,2} \to 0$ *in probability. Suppose that*

$$\lim_{M\to\infty} \overline{\lim_{n\to\infty}}\ \mathbb{E}\Big(F(Y_{1,n})\mathbb{1}_{\{F(Y_{1,n})\geq M\}}\Big) = 0. \tag{4.1}$$

Also assume that, for each $\epsilon > 0$ and $M > 0$, we have, as $n \to \infty$:

$$\log\Big(N\big(\epsilon, \mathcal{F}_M, ||\cdot||_{\overline{P}(\boldsymbol{\beta}_n,\mathbf{Y}_n),1}\,\big)\Big) = o_{\mathbb{P}}\Big(\,||\,\boldsymbol{\beta}_n\,||_{c,2}^{-2}\,\Big),$$

where, $\mathcal{F}_M := \big\{f\mathbb{1}_{\{F\leq M\}},\ f\in\mathcal{F}\big\}$ *and*

$$\overline{P}(\mathbf{p},\mathbf{y}) := \sum_{i\geq 1} p_i\delta_{y_i},\ \textit{for}\ \mathbf{p}\in\mathbb{S},\ \mathbf{y}\in\mathfrak{X}^{\mathbb{N}}.$$

Then

$$\mathbb{E}\Big(\,||\,G_{\mathbf{Y}_n,\boldsymbol{\beta}_n}\,||_{\mathcal{F}}\,\Big)\to 0. \tag{4.2}$$

Proof : The proof of 15 is an adaptation of the usual symmetrization and conditioning arguments for Glivenko-Cantelli theorems (see, e.g., [92, p. 123]). No consequent technical difficulty does show up along these steps.
Note that, choosing $\beta_{i,n} := n^{-1}$ when $i \leq n$ and 0 otherwise leads to the usual Glivenko-Cantelli theorem under random entropy conditions (see, e.g., [92, p.123]), except for the almost sure counterpart of (4.2). Indeed, that almost sure convergence deeply relies on a reverse submartingale structure, which is not guaranteed under the general conditions of Theorem 15.

4.1.3 A Donsker theorem

My second result is a Donsker Theorem. Since, for fixed n, the $Y_{i,n}$ are only conditionally independent, such a result will not involve the Gaussian analogues of the $G_{\mathbf{Y}_n,\boldsymbol{\beta}_n}$, but *mixtures* $W_{\mathbf{Y}_n,\boldsymbol{\beta}_n}$ of the $\mathbf{P}_{n,\mathbf{p}}$-Brownian bridges by $\boldsymbol{\beta}_n$. Such mixtures of Gaussian processes can fail to be measurable with respect to the Borel σ-algebra of $\big(\ell^\infty(\mathcal{F}), ||\cdot||_{\mathcal{F}}\,\big)$. This lack of measurability will be tackled by introducing outer expectations (see, e.g. [92, Chapter 1.2]). In the next theorem, the $\mathbb{E}^*$ symbol, when applied to $G_{\mathbf{Y}_n,\boldsymbol{\beta}_n}$, denotes the outer expectation taking $(\Omega,\mathcal{A},\mathbb{P})$ as underlying space. When applied to $W_{\mathbf{Y}_n,\boldsymbol{\beta}_n}$, that symbol should be understood as taking the canonical probability space $\mathbb{R}^{\mathcal{F}}$, endowed with the product Borel σ-algebra, and endowed with the probability measure exhibited by Kolmogorov's extension theorem. For more details, see [105].

Theorem 16 (Varron, 2014, [105]) *Assume that, for each $n \geq 1$, $||\,\boldsymbol{\beta}_n\,||_{c,2}^2 = 1$ with probability one, and that $||\,\boldsymbol{\beta}_n\,||_{c,4}\to 0$ in probability. Also assume that $\mathbb{E}\big(F^2(Y_{1,n})\big) < \infty$ for all n, and that :*

$$\lim_{M\to\infty}\overline{\lim_{n\to\infty}}\,\mathbb{E}\Big(F^2(Y_{1,n})\mathbb{1}_{\{F(Y_{1,n})\geq M\}}\Big) = 0,$$
$$\int_0^\infty\sqrt{\log\Big(\sup_{Q\ probab.} N\big(\epsilon\,||\,F\,||_{Q,2}, \mathcal{F}, ||\cdot||_{Q,2}\,\big)\Big)}d\epsilon < \infty.$$

Then, for all $n \geq 1$, $W_{\mathbf{Y}_n,\boldsymbol{\beta}_n}$ is almost surely bounded, and $||\,W_{\mathbf{Y}_n,\boldsymbol{\beta}_n}\,||_{\mathcal{F}}$ is Borel measurable.

Also assume that, for a semimetric ρ_0 that make $\mathcal{F}$ totally bounded we have :

$$\lim_{\delta\to 0}\ \overline{\lim_{n\to\infty}}^{\mathbb{P}^*} \sup_{(f_1,f_2)\in\mathcal{F}^2,\rho_0(f_1,f_2)\le\delta} \| f_1 - f_2 \|_{\mathbf{P}_{n,\boldsymbol{\beta}_n},2} = 0, \tag{4.3}$$

where the symbol $\overline{\lim}_{n\to\infty}^{\mathbb{P}^*}$ stands for the lim sup in outer probability. Then

$$d_{BL}\Big(G_{\mathbf{Y}_n,\boldsymbol{\beta}_n}, W_{\mathbf{Y}_n,\boldsymbol{\beta}_n}\Big) := \sup_{\phi\in BL1}\Big|\mathbb{E}^*\Big(\phi\big(G_{\mathbf{Y}_n,\boldsymbol{\beta}_n}\big)\Big) - \mathbb{E}^*\Big(\phi\big(W_{\mathbf{Y}_n,\boldsymbol{\beta}_n}\big)\Big)\Big| \to 0,$$

where $BL1$ is the set of all 1-Lipschitz functions on $\big(\ell^\infty(\mathcal{F}), \|\cdot\|_{\mathcal{F}}\big)$ that are bounded by 1.

Remark : When $(Y_{i,n})_{i\ge 1} \perp\!\!\!\perp \boldsymbol{\beta}_n$, condition (4.3) is implied by condition (ii) in Sheehy and Wellner [83, Theorem 3.1], namely, for some probability measure $\mathbf{P}_0$:

$$\sup_{(f,g)\in\mathcal{F}^2} \max\Big\{ \mid \mathbf{P}_0^{(n)}\big((f-g)^2\big) - \mathbf{P}_0\big((f-g)^2\big) \mid, \mid \mathbf{P}_0^{(n)} - \mathbf{P}_0(f) \mid \\ \mid \mathbf{P}_0^{(n)}(f^2) - \mathbf{P}_0(f^2) \mid \Big\} \to 0, \tag{4.4}$$

where $\mathbf{P}_0^{(n)}$ stands for the law of $Y_{1,n}$. In this particular case, Theorem 16 turns out to be a partial generalization of Sheehy and Wellner [83, Theorem 3.1], where the authors proved, among other results, a Donsker theorem for sequences of $\mathcal{F}$-indexed empirical processes, for which the law of the sample varies with n. The main advantage of our result is that it extends to random (possibly infinite) convex combinations of the $\delta_{Y_{i,n}}$.
As a consequence of Theorem 16, I could obtain an alternate proof of the Bernstein-Von Mises theorem under $\|\cdot\|_{\mathcal{F}}$-topologies, for the Dirichlet process prior. Such a result was already obtained by James [51], under the slightly more stringent assumption that $\mathcal{F}$ is VC subgraph.

4.1.4 Sketch of the proof and comments

The measurability of $\| W_{\mathbf{Y}_n,\boldsymbol{\beta}_n}(\cdot) \|_{\mathcal{F}}$ uses the fact that $\mathcal{F}$ is pointwise separable in combination with the dominated convergence theorem, which implies (under assumptions of integrability of the envelope) that there exists a countable $\mathcal{F}_0 \subset \mathcal{F}$ which, *for almost each* $\omega\in\Omega'$, is dense in $\big(\mathcal{F}, \|\cdot\|_{\mathbf{P}_{n,\boldsymbol{\beta}_n(\omega)},2}\big)$. The gaussian structure conditionally to $\boldsymbol{\beta}_n$ does the rest.
The proof of Theorem 16 is then decomposed into two parts :

Step 1: convergence of the marginals

Write d_{LP} for the Levy-Prokhorov distance between (Borel) probability measure on $\mathbb{R}^p$, generated by $\|\cdot\|_p$, namely, for two probability measures P and Q :

$$d_{LP}(P,Q) := \inf\big\{\lambda > 0,\ \pi(P,Q,\lambda) \le \lambda\big\},\ \text{where we remind that}$$

$$\pi(P,Q,\lambda) := \sup_{A\ Borel} \max\left\{P(A) - Q(A^\lambda), Q(A) - P(A^\lambda)\right\}, \text{for } \lambda > 0,$$

The first step consists in controlling the distance of the marginals of $G_{\mathbf{Y}_n,\boldsymbol{\beta}_n}$ and those of $W_{\mathbf{Y}_n,\boldsymbol{\beta}_n}$. Recall that $\Pi_{\mathbf{f}}(\psi) = \big(\psi(f_1), \dots, \psi(f_p)\big)$, as defined in (2.2).

Proposition 4.1.1 *For each $p \geq 1$ and $\mathbf{f} \in \mathbb{R}^p$, we have*

$$d_{LP}\big(\Pi_{\mathbf{f}}(G_{\mathbf{Y}_n,\boldsymbol{\beta}_n}), \Pi_{\mathbf{f}}(W_{\mathbf{Y}_n,\boldsymbol{\beta}_n})\big) \to 0.$$

The main argument of the proof of Proposition 4.1.1 is a generalization, to infinite convolutions, of the approximation results of Zaitsev [111] (a weaker form of Zaitsev's result can be found in the present manuscript, in Fact 2.1.1). Such a generalization is simple since Zaitsev's bounds do not depend on the sample size.

Step 2 : asymptotic equicontinuity

The second step consists in showing that $(W_{\mathbf{Y}_n,\boldsymbol{\beta}_n})_{n\geq 1}$ and $(G_{\mathbf{Y}_n,\boldsymbol{\beta}_n})_{n\geq 1}$ are asymptotically equicontinuous. First, from assumption (4.3), the the equicontinuity arguments boil down to showing that

$$\forall \epsilon > 0,\ \lim_{\delta\to 0}\ \overline{\lim_{n\to\infty}}^{\mathbb{P}} \sup_{\substack{(f_1,f_2)\in\mathcal{F}^2,\\ ||f_1-f_2||_{\mathbf{P}_{n,\boldsymbol{\beta}_n},2}<\delta}} \Big|G_{\mathbf{Y}_n,\boldsymbol{\beta}_n}(f_1) - G_{\mathbf{Y}_n,\boldsymbol{\beta}_n}(f_2)\Big| = 0, \tag{4.5}$$

$$\forall \epsilon > 0,\ \lim_{\delta\to 0}\ \overline{\lim_{n\to\infty}}^{\mathbb{P}} \sup_{\substack{(f_1,f_2)\in\mathcal{F}^2,\\ ||f_1-f_2||_{\mathbf{P}_{n,\boldsymbol{\beta}_n},2}<\delta}} \Big|W_{\mathbf{Y}_n,\boldsymbol{\beta}_n}(f_1) - W_{\mathbf{Y}_n,\boldsymbol{\beta}_n}(f_2)\Big| = 0. \tag{4.6}$$

The fact that those two suprema are measurable can be showed by density of $\mathcal{F}_0$ simultaneously for the pointwise convergence and for the norms $||\cdot||_{\mathbf{P}_{n,\boldsymbol{\beta}_n}(\omega),2}$, for almost all ω.
Assertion (4.6) is a consequence of the fact that $W_{\mathbf{Y}_n,\boldsymbol{\beta}_n}$ is, conditionally to $\boldsymbol{\beta}_n$, Gaussian with intrinsic semi metric $||\cdot||_{\mathbf{P}_{n,\boldsymbol{\beta}_n},2}$. Assertion (4.5) is a consequence, through symmetrization, of the fact that, conditionally to $\boldsymbol{\beta}_n$ and $\mathbf{Y}_n$, the process

$$\sum_{i\geq 1} \epsilon_i \beta_{i,n} \delta_{Y_{i,n}}(\cdot)$$

is subgaussian with intrinsic semimetric $\overline{P}(\boldsymbol{\beta}_n, \mathbf{Y}_n)$ (recall (4.2)). Then condition ($Unif.\ entropy$) plays its role.

4.1.5 Perspectives

A commonly used class of nonparametric discrete priors is the class of *normalized completely random measures*. A normalized *homogenous* completely random measure admits the representation (4.1), with the $Y_{i,n}$ being conditionally i.i.d given the weights. In [52], James, Lijoi and Prünster did prove a very general result, by giving the general formula for (a version of) the posterior distribution (given a sample $(X_1, \ldots, X_n)$) of a normalized completely random measure.
Such a formula does not directly imply objects such as (4.1) but mixtures of such random measures with respect to latent variables $U(X_1, \ldots, X_n)$.
A very appealing and apparently tough challenge would be to investigate the posterior consistency and rate of convergence, under $|| \cdot ||_{\mathcal{F}}$ of normalized *homogenous* completely random measures. James [51] already showed that it is not always the case by proving general consistency/unconsistency results when considering the Poisson-Dirichlet process as a prior. When the underlying law of the sample is continuous, the Dirichlet process prior is the only member of that family that benefits of $|| \cdot ||_{\mathcal{F}}$-consistency (except for trivial classes).

4.2 Empirical processes theory for empirical likelihood

4.2.1 The empirical likelihood principle

Empirical likelihood was first introduced by Thomas and Grunkemeier [89] in a setup of survival analysis. Owen ([73, 71, 72]) then generalized their concept and initiated a fruitful theory in estimation (see [74] for an overview). The underlying idea is simple to explain : consider an i.i.d. sample $(Y_1, \ldots, Y_n)$ taking values in $\mathbb{R}^p$, and admitting a nondegenerate variance, and consider the (most) simple problem of building confidence regions for their expectation θ_0. Instead of building ellipsoids centered at the point estimator

$$\hat{\theta}_n := \frac{1}{n} \sum_{i=1}^{n} Y_i,$$

we will construct a region as follows

$$\mathcal{R}_{n,u} := \Big\{ \sum_{i=1}^{n} p_i Y_i,\ (p_i)_{i=1,\ldots,n}\ probab.\ weights,\ \prod_{i=1}^{n} np_i \geq u \Big\},$$

where $u \in (0, 1)$ is a specified threshold. Owen showed that, under those simple assumptions, there is a way to calibrate u_α given a specified asymptotic confidence level $1-\alpha$. More precisely, he showed the following convergence in distribution.

Fact 4.2.1 (Owen, 1988) *Assume that Y_1 admits a variance that is nondegenerate. Then*

$$EL_n := \max\Big\{\prod_{i=1}^{n} np_i,\ (p_i)_{i=1,\ldots,n}\ \textit{probab. weights},\ \sum_{i=1}^{n} p_i Y_i = \theta_0\Big\} \to_{\mathcal{L}} \exp\big(-\chi_p^2\big). \qquad (4.7)$$

Note that

$$\theta_0 \in \mathcal{R}_{n,u} \Leftrightarrow EL_n \geq u.$$

One of the advantages of this empirical likelihood method is that one doesn't need to estimate the variance in order to build the confidence region. Moreover, the shape of those regions strongly depends on the geometry of the data whereas the classical central limit theorem gives ellipse-shaped confidence regions. Finally, the confidence band always lies inside of the convex hull defined by the data.
Since Owen's grounding result, several researchers did contribute to the extension of this method toward the estimation of parameters that are not expectations of finite dimensional data.

4.2.2 Estimations of linear, trajectory-valued functionals

With J.Y Dauxois and A. Flesch [15], we investigated the case where the parameter of interest is a multivariate trajectory from a set $\mathfrak{T}$ to $\mathbb{R}^p$. We will write that parameter $\theta_0(\cdot)$. The setup is then the space $\big(\ell_\infty(\mathfrak{T})\big)^p$, endowed with the supremum norm :

$$\|\ \mathbf{g}\ \|_{\mathfrak{T},p} := \sup_{t\in\mathfrak{T}} \|\ \mathbf{g}(t)\ \|_p,\ \mathbf{g} \in \big(\ell_\infty(\mathfrak{T})\big)^p.$$

The norm $\|\cdot\|_{\mathfrak{T},p}$ is natural when one aims toward establishing confidence regions that are uniform in $t \in \mathfrak{T}$. Assume that we have a natural estimator $\hat{\theta}_n(\cdot)$ which can be written

$$\hat{\theta}_n(\cdot) = \frac{1}{n}\sum_{i=1}^{n} L_n(Y_i,\cdot), \qquad (4.8)$$

where the $t \to L_n(Y_i,t)$ are observed. The simplest example is when $\theta_0(\cdot) = \mathbb{E}\big(L_0(Y_1,\cdot)\big)$ for a specified nonrandom map $L_0(\cdot,\cdot)$, which is not known by the statistician, but which is estimated by $L_n(\cdot,\cdot)$.
When $\mathfrak{T} = \{t_0\}$ is a singleton, Hjort *et al.* [50] gave very general conditions upon L_n to obtain a limit distribution for

$$\widetilde{EL}_n := \max\Big\{\prod_{i=1}^{n} np_i,\ (p_i)_{i=1,\ldots,n}\ \textit{prob. weights},\ \sum_{i=1}^{n} p_i L_n(Y_i,t_0) = \theta_0(t_0)\Big\}.$$

Following their ideas, we could derive a convergence for processes (in $\big(\ell_\infty(\mathfrak{T})\big)^p$) of the type

$$\widetilde{EL}_n(\cdot):\ t \to \max\Big\{\prod_{i=1}^n np_i,\ (p_i)_{i=1,\dots,n}\ prob.\ weights,\ \sum_{i=1}^n p_i L_n(Y_i,t)=\theta_0(t),\Big\}$$

toward a stochastic process which is the square Euclidian norm of an $\mathbb{R}^p$-valued gaussian process on $\mathfrak{T}$. Such a limit law permits to construct asymptotic confidence bands for $\theta_0(\cdot)$.

Original motivation from lifetime data analysis

The initial motivation of our work came from a particular inference problem in survival analysis : the inference of the mean number of different competing recurrent events. That framework was former introduced by Dauxois and Sencey [16], and can be described as follows : We observe i.i.d. replicates of the following random phenomenon :

— Consider a finite collection of counting processes on a time interval, $\big(N_j^\star(\cdot))_{j=1,\dots,k}\big)$. Those coutning processes represent k competing, recurrent events (for example, different type of nosocomial infections that can be contracted by hospitalized patients). Let $N_j^\star(t)$ denote the number of total events of type j up to time t.
— Let D be a terminal event and C an independent censorship. We make no assumption on the dependence structure between the counting processes and D. We suppose that $N_j^\star$ is almost surely bounded by B and has no jump after time D.
— We observe the processes $N_j^\star$ until $X = D \wedge C$ and define $\delta = I(D \le C)$, where $I(\cdot)$ denotes the indicator function. δ informs us of the reason why the observation was stopped.

Let us denote by S the survival function of D, that is $S(t) = \mathbb{P}(D > t)$. Write also :

$$Y(t) = I(X \ge t),$$
$$N_j(t) = N_j^\star(t \wedge C).$$

The observed data are i.i.d. replicates $(N_{i,j}(\cdot), X_i, \delta_i)_{j\le k}$ of $(N_j(\cdot), X, \delta)_{j\le k}$, where the processes are observed on $[0,\tau]$, and τ is a fixed constant chosen so that $\mathbb{P}(C > \tau)\mathbb{P}(X > \tau) > 0$. In order to build confidence regions for the mean functions $t \mapsto \mu_j(t) = \mathbb{E}(N_j^\star(t))$, we use the estimators $\widehat{\mu}_j(t)$ from Dauxois and Sencey [16] :

$$\widehat{\mu}_j(t) = \frac{1}{n}\sum_{i=1}^n \int_0^t \frac{\widehat{S}(u^-)}{\bar{Y}(u)} dN_{i,j}(u),$$

where $\widehat{S}(\cdot)$ is the Kaplan-Meier estimator of $S(\cdot)$, the survival function of D (see, for example, [3] for a definition of $\widehat{S}(\cdot)$). Such an estimator can typically be described as (4.8).

4.2.3 Estimation of nonlinear functionals

Among the possible generalizations of Owen's results, one particular approach is to consider parameters of interests that can be written $\theta_0 := T(\mathbf{P}_0)$, where $\mathbf{P}_0$ is the common law of the observed sample $(Y_1, \ldots, Y_n)$, and where T is a specified map. In that case, the idea is to build the following type of confidence region.

$$\mathcal{R}_{n,u} := \left\{ T\Big(\sum_{i=1}^{n} p_i \delta_{Y_i} \Big),\ (p_i)_{i=1,\ldots,n} \text{ prob. wghts },\ \prod_{i=1}^{n} np_i \geq u \right\}, \tag{4.9}$$

and determine their asymptotic covering probabilities, under the smallest possible assumptions upon the functional T.
The first results establishing an ad-hoc convergence in distribution such as (4.7) were already present in Owen [73], but they were limited to real-valued data, together with T being Frechet differentiable in $\big(D(\mathbb{R}), ||\cdot||_{\mathbb{R}}\big)$, where $D(\mathbb{R})$ stand for the space of bounded CADLAG functions on $\mathbb{R}$. Frechet differentiability is a rather strong assumption, which is rarely satisfied in practice.
The next major step in that direction was taken by Bertail [7], where he showed that, if a class of functions $\mathcal{F}$ satisfies $(Unif.\ entropy)$ and admits a square integrable envelope then a limit result such as (4.7) holds for functional T that are *Hadamard differentiable* with respect to $||\cdot||_{\mathcal{F}}$, tangentially to $\mathcal{C}(\mathcal{F}, ||\cdot||_{\mathbf{P}_0,2})$, the latter being the space of $||\cdot||_{\mathbf{P}_0,2}$ continuous functions on $\mathcal{F}$.
Following an initial work in A. Flesch's PhD thesis, I worked on extending Bertail's theorem under three directions :

- I considered the more general case where θ_0 is a trajectory (as in the preceding subsection).
- In practice, computing $\mathcal{R}_{n,u}$ is computationally too intensive (an exponential n complexity). Following the original idea of A. Flesch, I investigated the consistency of other confidence regions $\mathcal{R}_{n,u}^{LP}$, which involve a local linearization procedure of T around $\mathbf{P}_n$, and which makes the computation times reasonable.
- I considered the case where the estimator is built on several (say k) mutually independent samples. More precisely, $\theta_0 = T(\mathbf{P}_{0,1}, \ldots, \mathbf{P}_{0,k})$ is estimated by $T(\mathbf{P}_{n_1,1}, \ldots, \mathbf{P}_{n_k,k})$, using the corresponding empirical measures. Note that some steps were already taken in the multisample direction, for estimating differences of means or proportions ([53, Theorem 1], [78, statement following (2.6)], and also [62, Theorem 1] in the multivariate case), but the corresponding proofs were never entirely written.

4.2.4 The underlying probabilistic result

The two preceding paragraphs are voluntarily evasive, because introducing all the ad-hoc notations and assumptions would take a lot of space for an already quite long manuscript. It is however possible to point out the key probabilistic result which is behind those two contributions, without much notations. We will write $\boldsymbol{n} := (n_1, \ldots, n_k) \in \mathbb{N}^{*k}$, and we will write the simplex-type sets :

$$\mathbb{S}_{\boldsymbol{n},u} := \Big\{ \mathbf{p} = (p_{i,j})_{j\leq k, i\leq n_j},\ \forall j \leq k,\ \sum_{i=1}^{n_j} p_{i,j} = 1,$$
$$\forall i \leq n_j, p_{i,j} \in [0,1],\ \prod_{j=1}^{k}\prod_{i=1}^{n_j} n_j p_{i,j} \geq u \Big\}, u \in [0,1].$$

Theorem 17 (From Dauxois, Flesch and Varron, 2014 [15, 107]) *On a probability space* $(\Omega, \mathcal{A}, \mathbb{P})$*, let* $\big(W_{i,j,\boldsymbol{n}}(\cdot)\big)_{\boldsymbol{n}\in\mathbb{N}^{*k},\, j\leq k,\, i\leq n_j}$ *be a triangular array taking values in* $\big(\ell_\infty(\mathfrak{T})\big)^p$*. For* $t \in \mathfrak{T}$ *and* $\boldsymbol{n} \in \mathbb{N}^{*k}$*, write*

$$\mathbf{EL}_{\boldsymbol{n}}(t) := \max\Big\{ \prod_{j=1}^{k}\prod_{i=1}^{n_j} n_j p_{i,j}\ \mathbf{p} \in \mathbb{S}_{\boldsymbol{n},0},\ \sum_{j=1}^{k}\sum_{i=1}^{n_j} p_{i,j} W_{i,j,\boldsymbol{n}}(t) = 0 \Big\}.$$

Let $V_j(\cdot)$*,* $j \leq k$ *be deterministic functions on* $\mathfrak{T}$*, whose values are symmetric matrices, and for which*

$$0 < \inf_{t\in\mathfrak{T},\ \lambda\in S(\mathbb{R}^p)} \max_{j\leq k} \lambda \bullet V_j(t)\lambda \leq \sup_{t\in\mathfrak{T},\ \lambda\in S(\mathbb{R}^p)} \max_{j\leq k} \lambda \bullet V_j(t)\lambda < \infty.$$

Let $(U_1(\cdot), \ldots, U_k(\cdot))$ *be* $\mathbb{R}^p$*-valued stochastic processes. Assume that, for some positive array* $(a_{\boldsymbol{n}})_{\boldsymbol{n}\in\mathbb{N}^{*k}}$ *being bounded away from 0, we have, when* $\min\{n_1, \ldots, n_k\} \to \infty$ *:*

$$\text{(A0*)}\ \inf_{t\in\mathfrak{T},\ \lambda\in S(\mathbb{R}^k)} \max_{j\leq k,\ i\leq n_j} \mathbb{1}_{\{\lambda\bullet W_{i,j,\boldsymbol{n}}(t)>0\}} \to_{\mathbb{P}} 1,$$

$$\text{(A1*)}\ \Big(\sum_{i=1}^{n_j} W_{i,j,\boldsymbol{n}}(\cdot)\Big)_{j\leq k} \to_{\mathcal{L}} \big(U_j(\cdot)\big)_{j\leq k},$$

$$\text{(A2*)}\ \max_{j\leq k} \Big|\Big| a_{\boldsymbol{n}} \sum_{i=1}^{n_j} W_{i,j,\boldsymbol{n}}(\cdot)^{\otimes 2} - V_j(\cdot) \Big|\Big|_{\mathfrak{T},p^2} \to_{\mathbb{P}} 0,$$

$$\text{(A3*)}\ \max_{j\leq k,\ i\leq n_j} \Big|\Big| W_{i,j,\boldsymbol{n}}(\cdot) \Big|\Big|_{\mathfrak{T},p} = o_{\mathbb{P}}(a_{\boldsymbol{n}}^{-1}).$$

Then, when $\min\{n_1, \ldots, n_k\} \to \infty$ *together with*

$$\Big(\sum_{j=1}^{k} n_j\Big)^{-1} \boldsymbol{n} \to (\alpha_1, \ldots, \alpha_k) \in (0,1]^k,$$

we have

$$-2a_{\boldsymbol{n}}^{-1} \log \mathbf{EL}_{\boldsymbol{n}}(\cdot) \to_{\mathcal{L}} U(\cdot) \bullet V^{-1}(\cdot) U(\cdot),$$

where $U(\cdot) := \sum_{j=1}^{k} \alpha_j^{-1} U_j(\cdot)$ *and* $V(\cdot) = \sum_{j=1}^{k} \alpha_j^{-2} V_j(\cdot)$*.*

4.2.5 Sketch of the proof and comments

In this paragraph, I will try to emphasize the aspects of the proof that make it non trivial in regard of the existing literature.
Let us go back to the basic ingredients of Owen's results. Consider the setup of Fact 4.2.1, and the expression EL_n as defined in (4.7). To simplify the notations, we will consider that $\mu_0 = \mathbb{E}\big(Y_1\big) = 0$. As soon as $\mu_0 = 0$ lies in the convex-hull of the sample (which is asymptotically guaranteed since Y_1 has a nondegenerate covariance matrix). It is possible to express EL_n through a Lagrange multiplier $\lambda^* = \lambda^*(Y_1, \ldots, Y_n)$. More precisely, EL_n can be expressed as $\prod_{i=1}^{n} np_i^*$, where :

$$p_i^* := \frac{1}{n}\frac{1}{1+\lambda^*\bullet Y_i},$$

and λ^* satisfies

$$\sum_{i=1}^{n}\frac{1}{1+\lambda^*\bullet Y_i}Y_i = 0.$$

Then λ^* is controlled by making use of the fact that the Y_i satisfy the central limit theorem.
Now, if we try to use the same representation for $t \to \mathbf{EL}_{\boldsymbol{n}}(t)$, we obtain $\mathbf{EL}_{\boldsymbol{n}}(t) = \prod_{j=1}^{k}\prod_{i=1}^{n_j} n_j p_{i,j}^*(t)$, with

$$\begin{aligned}
&p_{i,j,\boldsymbol{n}}^*(t)\gamma_j^*(t) + Np_{i,j,\boldsymbol{n}}^*(t)\lambda_{\boldsymbol{n}}^*(t)\bullet W_{i,j,\boldsymbol{n}}(t) = 1,\ j\le k,\ i\le n_j, \qquad (4.10)\\
&\sum_{j=1}^{k}\sum_{i=1}^{n_j} p_{i,j,\boldsymbol{n}}^*(t)W_{i,j,\boldsymbol{n}}(t) = 0,\\
&\big(p_{i,j,\boldsymbol{n}}^*(t)\big)_{j\le k,\ i\le n_j} \in \mathbb{S}_{\boldsymbol{n},0},\ \text{ and } (\gamma_j^*)_{j=1,\ldots,k}\in\mathbb{R}^k.
\end{aligned}$$

Summing (4.10) in i for fixed j entails

$$\gamma_j^*(t) + N\lambda_{\boldsymbol{n}}^*(t)\bullet\widetilde{W}_{j,\boldsymbol{n}}(t) = n_j,\ \text{ with } \widetilde{W}_{j,\boldsymbol{n}}(t) := \sum_{i=1}^{n_j} p_{i,j,\boldsymbol{n}}^*(t)W_{i,j,\boldsymbol{n}}(t).$$

This leads to the expression

$$\begin{aligned}
p_{i,j,\boldsymbol{n}}^*(t) :=&\frac{1}{n_j + N\lambda_{\boldsymbol{n}}^*(t)\bullet(W_{i,j,\boldsymbol{n}}(t)-\widetilde{W}_{j,\boldsymbol{n}}(t))}\\
=&\frac{1}{n_j}\frac{1}{1+\alpha_{j,\boldsymbol{n}}^{-1}\lambda_{\boldsymbol{n}}^*(t)\bullet(W_{i,j,\boldsymbol{n}}(t)-\widetilde{W}_{j,\boldsymbol{n}}(t))}, \qquad (4.11)
\end{aligned}$$

with $\alpha_{j,\boldsymbol{n}} := n_j/N$.

- The first technical difficulty brought by a multisample setting is the annoying presence of the extra term $\widetilde{W}_{j,\boldsymbol{n}}(t)$, itself depending on $(p^*_{i,j,\boldsymbol{n}})_{i\leq n_j}$, in equality (4.11).
- The second technical difficulty is to adapt the arguments of Owen to trajectories. Basically :
  - Condition $(A0^*)$ allows the exhibition of the lagrange multipliers λ^* (convex hull condition).
  - The central limit theorem for the Y_i is replaced by condition $(A1^*)$, which is now a convergence of stochastic processes.
  - The fact that the variance of Y_1 is nondegenerate is replaced by $(A2^*)$.
  - The fact that Y_1 is square integrable is replaced by $(A3^*)$.

4.2.6 Perspectives

Instead of considering explicit plug-in estimators $T(\mathbf{P}_{n_1,1},\ldots,\mathbf{P}_{n_k,k})$, the next natural step is to investigate M estimators of the general form

$$\hat{\theta}_n(\cdot):\ t\to \mathrm{argmin}_{\theta\in\Theta}L\big(\mathbf{P}_{n_1,1},\ldots,\mathbf{P}_{n_k,k},\theta,t\big),$$

where L is typically a loss function, and to determine under which regularity conditions upon L it is possible to use the general empirical likelihood principle by unbalancing the empirical weights in the $\mathbf{P}_{n_j,j}$.
Another interesting perspective would be to investigate the asymptotic properties of the "most unbalanced empirical measure", namely

$$\overline{P}_n := \sum_{i=1}^{n} p^*_{i,n}\delta_{Y_i},$$

where the $p^*_{i,n}$ are the maximizing weights of $\prod_{i=1}^{n} np_i$ under a constraint of the form

$$\sum_{i=1}^{n} p_i f_j(Y_i)=\mathbb{P}(f_j),\ j=1,\ldots,p$$

For example, given a class of functions $\mathcal{F}$, and given $(f_1,\ldots,f_k)\in\mathcal{F}^k$, is it possible to obtain a limit distribution for the processes

$$\overline{G}_n := v_n\Big(\overline{P}_n(\cdot)-\mathbf{P}_n(\cdot)\Big),$$

in $\ell^\infty(\mathcal{F})$, for some normalizing sequence v_n?

4.3 Performances of the discrete associated kernel estimators for the total variation distance

4.3.1 The setup

Let $\big(\mathbb{T}, \mathcal{B}, c\big)$ be a measured space, where $\mathbb{T}$ is countable, $\mathcal{B}$ is the σ-algebra of all subsets of $\mathbb{T}$, and c is the counting measure on $\mathbb{T}$. Given an i.i.d. sample $(X_1, \ldots, X_n)$ taking values in $\mathbb{T}$, we are interested in estimating the *probability mass function* [p.m.f.] $f : x \to \mathbb{P}\big(X = x\big)$ where X stands for a generic random variable having the common distribution of the X_i. The natural estimator of f is the empirical mass function, namely :

$$f_n : x \to \mathbf{P}_n(\{x\}). \tag{4.12}$$

Recently, Kokonendji *et al.* [57] and Kokonendji and Abdous [1] introduced the *discrete associated kernels* density estimator, which can be described as follows : let $\mathcal{K} := \big\{K_{x,h},\ x \in \mathbb{T},\ h \geq 0\big\}$ be a collection of p.m.f on $\mathbb{T}$. For a bandwidth $h \geq 0$, we define :

$$g_{n,h}(x) := \frac{f_{n,h}(x)}{\int_{\mathbb{T}} f_{n,h}(u)dc(u)}, \text{ where}$$
$$f_{n,h}(x) := \frac{1}{n}\sum_{i=1}^{n} K_{x,h}(X_i),\ x \in \mathbb{T}.$$

Several simulation studies [57] showed promising results, comparing some of those estimators with the traditional "bar plot" f_n.
In [1], a first study on the asymptotic properties of such methods showed that their pointwise consistency holds as soon as each $K_{x,h}$ converges, when $h \to 0$, towards the Dirac distribution δ_x in the following sense :

$$\forall k = 1, 2,\ \forall x \in \mathbb{T},\ \lim_{h\to 0} \int u^k K_{x,h}(u)dc(u) = x^k, \tag{4.13}$$

and

$$\forall x \in \mathbb{T},\ \limsup_{h\to 0} \mathbb{E}\Big(\big(K_{x,h}(X)\big)^3\Big) < \infty.$$

Beside the pointwise strong consistency, the authors also established the pointwise asymptotic normality of $g_{n,h}(x)$ (see [1, Theorem 2.4 and 2.5]). With C. Kokonendji, we established asymptotic results for the total variation distance between $g_{n,h}$ and f, namely, the random variable

$$TV\big(g_{n,h}, f\big) := \sup_{A\subset\mathbb{T}} \Big| \int_A g_{n,h}dc - \int_A f dc\Big| = \frac{1}{2} \parallel g_{n,h} - f \parallel_{c,1}. \tag{4.14}$$

4.3.2 The concentration of the normalizing constant

In a first time, we focused on the problem of the normalizing constant

$$C_{n,h} := \frac{1}{n}\sum_{i=1}^{n} Y_{i,h}, \text{ where}$$
$$Y_{i,h} := \int_{\mathbb{T}} K_{x,h}(X_i)dc(x).$$

Note that, without any further assumptions, it is not even guaranteed that $C_{n,h}$ is a.s. finite. From a statistical point of view, several questions immediately arise :

— For given $h \geq 0$, does $Y_{1,h}$ admit a first/second moment?
— Do we have $\mathbb{E}(Y_{1,h}) \to 1$ as $h \to 0$?
— Do we have, for any sequence $h_n \to 0$, a convergence of C_{n,h_n} to 1, in probability or almost surely? Can we get concentration inequalities?

To answer these questions, we proposed the following sets of assumptions for $\ell > 0$:

(HK_0) For each $x \in \mathbb{T}$, we have $K_{x,h}(x) \to 1$ as $h \to 0$.
$(HK_1(\ell))$ There exist a collection of p.m.f $\{g_y(\cdot),\ y \in \mathbb{T}\}$ and a positive function $\mathfrak{h}$ such that $\mathbb{E}\Big(\mathfrak{h}^{\ell}(X)\Big) < \infty$ and, for each $h \in [0,1], (x,y) \in \mathbb{T}^2$, we have $K_{x,h}(y) \leq \mathfrak{h}(y) g_y(x)$;
(HK_2) The function $y \mapsto \sup_{h\in[0,1]} \int_{x\in\mathbb{T}} K_{x,h}(y)dc(x)$ is bounded on $\mathbb{T}$ by a constant C.

When $\mathbb{T} = \mathbb{Z}^d$, several associated kernels can be represented as

$$K_{x,h}(y) := \mathbb{K}_h(x-y),\ h \geq 0,\ (x,y) \in \mathbb{Z}^d \times \mathbb{Z}^d, \tag{4.15}$$

where $\{\mathbb{K}_h,\ h \geq 0\}$ is a family of p.m.f on $\mathbb{Z}^d$. Our first result is stated as follows.

Theorem 18 (Kokonendji and Varron, 2014, [56]) *Let $\mathcal{K} = \{K_{x,h},\ h \geq 0\ x \in \mathbb{T}\}$ be a family of associated kernels admitting representation (4.15). Then $Y_{1,h} \equiv C_{n,h} \equiv 1$ for all $h \geq 0$ and $n \geq 1$, and (HK_2) is satisfied with $C = 1$.*
Under (HK_0) and $(HK_1(\ell))$, for $\ell \in \mathbb{N}^*$, we have $\mathbb{E}(Y_{1,h}^{\ell}) < \infty$ for all $h \in [0,1]$, and hence $\mathbb{E}(C_{n,h}^{\ell}) < \infty$ for all $n \geq 1$ and $h \in [0,1]$. Moreover, we have

$$\lim_{h\to 0} \mathbb{E}(Y_{1,h}^{\ell}) = 1.$$

As a consequence, we have, under (HK_0) and $(HK_1(2))$:

$$\lim_{h\to 0} \mathrm{Var}(Y_{1,h}) = 0, \text{ and hence}$$
$$\lim_{h\to 0} \sup_{n\geq 1} \mathbb{E}\Big(\big(C_{n,h}-1\big)^2\Big) = 0.$$

Under (HK_2), we have, for all $t>0$ and $h \in [0,1]$:

$$\mathbb{P}\Big(C_{n,h} - \mathbb{E}(Y_{1,h}) \geq t\Big) \leq \exp\Big[-\frac{n\Delta^2(h)}{C^2}\mathbf{h}\Big(\frac{C}{\Delta^2(h)}\Big)\Big]$$
$$\mathbb{P}\Big(C_{n,h} - \mathbb{E}(Y_{1,h}) \leq -t\Big) \leq \exp\Big[-\frac{n\Delta^2(h)}{C^2}\mathbf{h}\Big(\frac{C}{\Delta^2(h)}\Big)\Big],$$

where C is mentioned in (HK_2), $\Delta^2(h) := \mathrm{Var}(Y_{1,h})$, and $\mathbf{h}(x) = (x+1)\ln(x+1) - x,\ x>0$.

Remark : I has to be noticed that, while condition (HK_0) is less stringent than (4.13), this is not the case for conditions $(HK_1(\ell))$ and (HK_2). Those conditions may be considered as strong at first glance. However, each associated kernel introduced up to far in the literature does satisfy $(HK_1(2))$, sometimes at the price of making light moment assumptions on X.

4.3.3 Results for the total variation distance

Our second result is focused on $\|\, g_{n,h} - f \,\|_{c,1}$, note, that, in light of the preceding theorem, and by the inequality

$$\|\, g_{n,h} - f \,\|_{c,1} \leq C_{n,h}^{-1} \,\|\, f_{n,h} - f \,\|_{c,1} + |\, C_{n,h}^{-1} - 1 \,|,$$

the main object to concentrate our attention on is $\|\, f_{n,h} - f \,\|_{c,1}$. We first proved the convergence to 0 of its expectation.

Theorem 19 (Kokonendji and Varron, 2014, [56]) *Under $(HK_1(1))$ we have :*

$$\lim_{h\to 0} \int_{\mathbb{T}} \Big|\mathbb{E}\big(K_{x,h}(X)\big) - f(x) \mid dc(x) = 0,$$
$$\lim_{n\to\infty} \sup_{h\in[0,1]} \int_{\mathbb{T}} \mathbb{E}\Big(\Big|f_{n,h}(x) - \mathbb{E}\big(K_{x,h}(X)\big) \mid \Big) dc(x) = 0.$$

Moreover, if $(HK_1(2))$ satisfied and if

$$\int_{\mathbb{T}} \mathfrak{h}(y)\sqrt{f(y)}dc(y) < \infty, \ \textit{then}$$

$$\sup_{h\in[0,1]} \int_{\mathbb{T}} \mathbb{E}\Big(\mid f_{n,h}(x) - \mathbb{E}\big(K_{x,h}(X)\big) \mid \Big) dc(x) = O(n^{-1/2}).$$

We pursue this subsection with a concentration inequality for $\|\, f_{n,h} - f \,\|_{c,1}$. The first one is based on the almost sure boundedness of $Y_{1,h}$, while the second is based on the finiteness of a ϕ_α norm, which is closely related to the tail probabilities of $Y_{1,h}$. Recall that the ϕ_α norm of a random variable Z is defined as

$$|| Z ||_{\phi_\alpha} := \inf \Big\{ B > 0, \ \mathbb{E}\Big(\exp\Big(\frac{| Z |^\alpha}{B^\alpha} \Big) \Big) \leq 2 \Big\}, \ \alpha > 0,$$

with the implicit convention $\inf \emptyset = \infty$.

Theorem 20 (Kokonendji and Varron, 2014, [56]) *Under (HK_2) we have, for each $t \geq 0$ and $h \in [0,1]$.*

$$\mathbb{P}\Big(|| f_{n,h} - f ||_{c,1} \geq t + \mathbb{E}\Big(|| f_{n,h} - f ||_{c,1} \Big) \Big) \leq \exp\Big(- \frac{nt^2}{2C^2} \Big),$$
$$\mathbb{P}\Big(|| f_{n,h} - f ||_{c,1} \leq -t + \mathbb{E}\Big(|| f_{n,h} - f ||_{c,1} \Big) \Big) \leq \exp\Big(- \frac{nt^2}{2C^2} \Big).$$

Now assume that, for some value of $\alpha > 0$, we have $|| \widetilde{Y}_{1,h} ||_{\phi_\alpha} < \infty$ for all $h \in [0,1]$, where

$$\widetilde{Y}_{i,h} := \int_{x \in \mathbb{T}} \Big| K_{x,h}(X_i) - \mathbb{E}\big(K_{x,h}(X_i) \big) \Big| dc(x), \ i \geq 1, \ h \in [0,1].$$

Then, for any $\eta \in [0,1]$ and $\delta > 0$, there exists $K = K(\alpha, \eta, \delta)$ such that, for all $n \geq 1$, $t \geq 0$ and $h \in [0,1]$, we have

$$\mathbb{P}\Big(\int_{\mathbb{T}} \Big| f_{n,h}(x) - \mathbb{E}\big(f_{n,h}(x)\big) \Big| dc(x) \geq t + (1+\eta)\mathbb{E}\Big(\int_{\mathbb{T}} \Big| f_{n,h}(x) - \mathbb{E}\big(f_{n,h}(x)\big) \Big| dc(x) \Big) \Big)$$
$$\leq \exp\Big(- \frac{nt^2}{8(1+\delta)\sigma^2(h)} \Big) + 3 \exp\Big(- \frac{n^\alpha t^\alpha}{K \, || \widetilde{Y}_{1,h} ||_{\phi_\alpha} \log(n)^{1/\alpha}} \Big), \ \textit{and}$$
$$\mathbb{P}\Big(\int_{\mathbb{T}} \Big| f_{n,h}(x) - \mathbb{E}\big(f_{n,h}(x)\big) \Big| dc(x) \leq -t + (1-\eta)\mathbb{E}\Big(\int_{\mathbb{T}} \Big| f_{n,h}(x) - \mathbb{E}\big(f_{n,h}(x)\big) \Big| dc(x) \Big) \Big)$$
$$\leq \exp\Big(- \frac{nt^2}{8(1+\delta)\sigma^2(h)} \Big) + 3 \exp\Big(- \frac{n^\alpha t^\alpha}{K \, || \widetilde{Y}_{1,h} ||_{\phi_\alpha} \log(n)^{1/\alpha}} \Big), \ \textit{where}$$
$$\sigma^2(h) := \sup_{A \subset \mathbb{T}} \mathrm{Var}\Big(\int_A K_{x,h}(X) dc(x) \Big).$$

As a consequence, for any sequence $h_n \to 0$, we have $|| f_{n,h_n} - f ||_{c,1} \to 0$ almost completely.

Bibliography

[1] B. Abdous and C. Kokonendji. Consistency and asymptotic normality for discrete associated-kernel estimator. *Afr. Diasp. Journ. Math*, 8(2):63–70, 2009.

[2] Acosta (De), A. Small deviations in the functional central limit theorem with applications to the functional laws of he iterated logarithm. *Ann. Probab.*, 11(2):78–101, 1983.

[3] P. K. Andersen, Ø. Borgan, R. D. Gill, and N. Keiding. *Statistical models based on counting processes*. Springer Series in Statistics. Springer-Verlag, New York, 1993.

[4] M.A. Arcones. The large deviation principle of stochastic processes, Part 1 . *Theory Probab. Appl.*, 47(4):567–583, 2003.

[5] M.A. Arcones. The large deviation principle of stochastic processes, Part 2. *Theory Probab. Appl.*, 48(1):19–44, 2004.

[6] J.M. Azais and M. Wschebor. Almost sure oscillation of certain random processes. *Bernoulli*, 2(3):257–270, 1996.

[7] P. Bertail. Empirical likelihood in some semiparametric models. *Bernoulli*, 12(2):299–331, 2006.

[8] P. Berthet. On the rate of clustering to the Strassen set for increments of the empirical process. *J. Theoret. Probab.*, 10(3):557–579, 1997.

[9] P. Berthet. Inner rates of coverage of Strassen type sets by increments of the uniform empirical and quantile processes. *Stochastic Process. Appl.*, 115(3):493–537, 2005.

[10] P. Berthet. Functional chung laws for small increments of the empirical process and a lower bound in the strong invariance principle. *Period. Math. Hungar.*, 61(1):67–102, 2010.

[11] P. Berthet and M. Lifshits. Some exact rates in the functional law of the iterated logarithm. *Ann. Inst. H. Poincaré Probab. Statist.*, 38:811–824, 2002.

[12] P.J. Bickel and M Rosenblatt. On some global measures of the deviations of density function estimates. *Ann. Statist.*, 1:1071–1095, 1973.

[13] O. Bousquet. A Bennett concentration inequality and its application to suprema of empirical processes. *C.R. Acad. Sci. Paris, Ser. I*, 334:495–500, 2002.

[14] E. Csàki. A relation between Chung's and Strassen's laws of the iterated logarithm. *Z. Wahrsch. Verw. Gebiete*, 54(3):287–301, 1980.

[15] J.Y. Dauxois, A. Flesch, and D. Varron. Non-parametric Tests for Recurrent Events under Competing Risks. *Preprint.*, 2014.

[16] J.Y. Dauxois and S. Sencey. Non-parametric Tests for Recurrent Events under Competing Risks. *Scand. J. Statist.*, 36(4):649 – 670, 2009.

[17] P. Deheuvels. Chung type functional laws of the iterated logarithm for tail empirical processes. *Ann. Inst. H. Poincaré Probab. Statist.*, 36:583–616, 2000.

[18] P. Deheuvels. Strong approximation of quantile process by iterated Kiefer processes. *Ann. Probab.*, 28(2):909–945, 2000.

[19] P. Deheuvels, U. Einmahl, and D. Mason. Asymptotic independence of the local empirical process indexed by functions. In *High dimensional probability, II*, pages 183–205, 2000.

[20] P. Deheuvels and M. Lifshits. Probabilities of hiting of shifted small balls by centered Poisson processes. *J. Math. Sci. (N. Y.)*, 118(6):5541–5554, 2003.

[21] P. Deheuvels and D.M. Mason. Nonstandard functional laws of the iterated logarithm for tail empirical and quantile processes. *Ann. Probab.*, 18(4):1693–1722, 1990.

[22] P. Deheuvels and D.M. Mason. Functional laws of the iterated logarithm for the increments of empirical and quantile processes. *Ann. Probab.*, 20(3):1248–1287, 1992.

[23] P. Deheuvels and D.M. Mason. Functional laws of the iterated logarithm for local empirical processes indexed by sets. *Ann. Probab.*, 22(3):1619–1661, 1994.

[24] P. Deheuvels and D.M. Mason. Nonstandard local empirical processes indexed by sets. *J. Statist. Plann. Inference*, 23(1):355–387, 1995.

[25] P. Deheuvels and D.M. Mason. General asymptotic confidence bands based on kernel-type function estimators. *Stat. Inference Stoch. Process.*, 7(3):225–277, 2004.

[26] P. Deheuvels and S. Ouadah. Uniform-in-Bandwidth Functional Limit Laws. *J. Theoret. Probab.*, pages 1–25, 2011.

[27] A. Dembo and O. Zeitouni. *Large deviations techniques and applications.* Jones and Bartlett Publishers, Boston, MA, 1993.

[28] J.D. Deuschel and D.W. Stroock. *Large deviations.* Wiley, New York, 1968.

[29] L. Devroye and G. Lugosi. *Combinatorial methods in density estimation.* Springer, New York, 2001.

[30] J. Dony and U. Einmahl. Weighted uniform consistency of kernel density estimators with general bandwidth sequences. *Electron. Journ. Probab.*, 11:844–859, 2006.

[31] J. Dony and U. Einmahl. Uniform in bandwidth consistency of kernel regression estimators at a fixed point. In *High Dimensional Probability V*, pages 308–325, 2009.

[32] J. Dony, U. Einmahl, and D. Mason. Uniform in bandwidth consistency of local polynomial regression function estimators. *Austr. J. Stat.*, 35:105–120, 2006.

[33] J. Dony and D. Mason. Uniform in bandwidth consistency of kernel estimators of the tail index. *Extremes*, 13(3):353–371, 2010.

[34] J. Dony and D. M. Mason. Uniform in bandwidth consistency of conditional U-statistics. *Bernoulli*, 14(4):1108–1133, 2008.

[35] R. M. Dudley. A course on empirical processes. In *Ecole d'Eté de Probabilités de Saint-Flour XII-1982*, pages 1–142. 1984.

[36] J.H.J. Einmahl. *Multivariate empirical processes.* PhD thesis, Ph.D dissertation, Katolieke Universiteit Nigmegen, 1986.

[37] J.H.J. Einmahl. The as behavior of the weighted empirical process and the lil for the weighted tail empirical process. *Ann. Probab.*, 20(2):681 – 695, 1992.

[38] J.H.J. Einmahl. Poisson and gaussian approximation of weighted local empirical processes. *Stochastic Process. and their Appl.*, 70(1):31 – 58, 1997.

[39] J.H.J. Einmahl and D.M. Mason. Strong limit theorems for weighted quantile processes. *Ann. Probab.*, 16(4):1626, 1988.

[40] U. Einmahl and D. Li. Characterization of lil behavior in banach space. *Trans. of the Amer. Math. Soc.*, 360(12):6677–6693, 2008.

[41] U. Einmahl and D.M. Mason. Gaussian approximation of local empirical processes indexed by functions. *Probab. Theory Related Fields*, 107(3):283–311, 1997.

[42] U. Einmahl and D.M. Mason. An empirical process approach to the uniform consistency of kernel-type function estimators. *J. Theoret. Probab.*, 13:1–37, 2000.

[43] U. Einmahl and D.M. Mason. Uniform in bandwidth consistency of variable bandwidth kernel estimators. *Ann. Statist.*, 33(3):1380–1403, 2005.

[44] T. S. Ferguson. A bayesian analysis of some nonparametric problems. *Ann. statist.*, pages 209–230, 1973.

[45] H. Finkelstein. The law of iterated logarithm for empirical distributions. *Ann. Math. Statist.*, 42:607–625, 1971.

[46] E Giné, D.M. Mason, and A. Zaitsev. The L_1 -norm density estimator process. *Ann. Probab.*, 31(2):719–768, 2003.

[47] N. Gorn and M. Lifshits. Chung's law and the Csàki function. *J. Theoret. Probab.*, 12:399–420, 1999.

[48] K. Grill. A lim inf result in Strassen's law of the iterated logarithm. *Probab. Theory Related Fields*, 89:149–157, 1991.

[49] N.L. Hjort, C. Holmes, P. Müller, and S.G. Walker. *Bayesian nonparametrics*, volume 28. Cambridge University Press, 2010.

[50] N.L. Hjort, I.W. Mc Keague, and I. Van Keilegom. Extending the scope of empirical likelihood. *Ann. Statist.*, 37(3):1079–1111, 2009.

[51] L. F. James. Large sample asymptotics for the two-parameter poisson-dirichlet process. In *Pushing the limits of contemporary statistics: contributions in honor of Jayanta K. Ghosh*, pages 187–199. Institute of Mathematical Statistics, 2008.

[52] L. F. James, A. Lijoi, and I. Prünster. Posterior analysis for normalized random measures with independent increments. *Scand. J. Statist.*, 36(1):76–97, 2009.

[53] Bing-Yi Jing. Two-sample empirical likelihood method. *Stat. Proba. Letters*, 24(4):315 – 319, 1995.

[54] J. Kiefer. On Bahadur's representation of sample quantiles. *Ann. Math. Statist.*, 38(5):1323–1342, 1967.

[55] T. Klein. Une inégalité de concentration à gauche pour les processus empiriques. *C.R. Acad. Sci. Paris, Ser. I*, 2002.

[56] C. Kokonendji and D. Varron. Performance of the discrete associated kernel methods for the total variation distance. *Preprint*, 2014.

[57] C.C. Kokonendji, T. Senga Kiessé, and S. Zocchi. Discrete triangular distributions and non-parametric estimation for probability mass function. *J. Nonparametr. Stat.*, 19:241–254, 2007.

[58] V. Koltchinskii. "komlós-major-tusnády approximation for the general empirical process and haar expansions of classes of functions". *Journ. Theoret. Probab.*, 7(1):73–118, 1994.

[59] J. Komlós, P. Major, and G. Tusnády. An approximation of partial sums of independent r.v.'s and the sample d.f.I. *Z. Wahrsch. Verv. Gebiete*, 32:111–131, 1975.

[60] J. Komlós, P. Major, and G Tusnády. An approximation of partial sums of independent r.v.'s and the sample d.f.II. *Z. Wahrsch. Verv. Gebiete*, 34:33–58, 1977.

[61] M. Lifshits. *Lectures on Gaussian processes.* Springer, 2012.

[62] Y. Liu, C. Zou, and R. Zhang. Empirical likelihood for the two-sample mean problem. *Stat. Proba. Letters*, 78(5):548–556, 2008.

[63] J. Lynch and J. Sethuraman. Large deviations for processes with independent increments. *Ann. Probab.*, 15(2):610–627, 1987.

[64] D.M. Mason. A strong invariance principle for the tail empirical process. *Ann. Inst. H. Poincaré Probab. Statist.*, 24(4):491–506, 1988.

[65] D.M. Mason. A uniform functional law of the iterated logarithm for the local empirical process. *Ann. Probab.*, 32(2):1391–1418, 2004.

[66] D.M. Mason and W.R. van Zwet. A refinement of the KMT inequality for the uniform empirical process. *Ann. Probab.*, 15:871–884, 1987.

[67] P. Massart. Strong approximation for multivariate empirical and related processes, via kmt constructions. *Ann. Probab*, 17(1):266–291, 1989.

[68] M. Maumy and D. Varron. Non standard functional limit laws for the increments of the compound empirical distribution function. *Electron. J. Stat.*, 4:1324–1344, 2010.

[69] L. Menneteau. Large deviations for local empirical measures. In *High dimensional probability, II*, 1999.

[70] S.J. Montgomery-Smith. Comparison of sums of identically distributed random vectors. *Probab. Math. Statist.*, 14:281–285, 1993.

[71] A. B. Owen. Empirical likelihood ratio confidence regions. *Ann. Statist.*, 18(1):90–120, 1990.

[72] A. B. Owen. Empirical likelihood for linear models. *Ann. Statist.*, 19(4):1725–1747, 1991.

[73] A.B. Owen. Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, 75(2):237–249, 1988.

[74] A.B. Owen. *Empirical Likelihood.* Chapman and Hall/CRC, 2001.

[75] J.S. III Pickands III. Upcrossing probabilities for gaussian processes. *Trans.Amer.Math.Soc.*, (145):51–73, 1960.

[76] J. Pitman. Combinatorial stochastic processes. Technical report, Springer.

[77] A. Pukhalskii. On the theory of large deviations. *Theory Probab. Appl.*, 38(3):490–497, 1993.

[78] Jing Qin. Semi-empirical likelihood ratio confidence intervals for the difference of two sample means. *Ann. Inst. Stat. Math.*, 46(1):117–126, 1994.

[79] P. Révész. On strong approximation of the multidimensional empirical process. *Ann. Probab.*, 4(5):729–743, 1976.

[80] E. Rio. Vitesses de convergence dans le principe d'invariance faible pour la fonction de répartition empirique multivariée. *C.R. Acad. Sci. Paris, Ser. I*, 322(2):169–172, 1996.

[81] A. Rouault. Lois de zipf et sources markoviennes. In *Ann. Inst. H. Poincaré Probab. Statist.*, volume 14, pages 169–188, 1978.

[82] J Sethuraman. A constructive definition of dirichlet measures. *Statist. Sinica*, 1994.

[83] Anne Sheehy and Jon A Wellner. Uniform donsker classes of functions. *Ann. Probab.*, pages 1983–2030, 1992.

[84] E. Yu. Shmileva. Small ball probabilities for a centered poisson process of high intensity. (russian). *Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI)*, 298:280–303, 2003.

[85] G.R. Shorack and J.A. Wellner. *Empirical Processes and applications to statistics.* Springer, 1986.

[86] C. Suquet. Tightness in Schauder decomposable Banach spaces. *Amer. Math. Soc. Transl.*, 193(2):201–224, 1999.

[87] M. Talagrand. New Gaussian estimates for enlarged balls. *Geom. Funct. Anal.*, 3(5):503–525, 1993.

[88] M. Talagrand. Sharper bounds for Gaussian and empirical processes. *Ann. Probab.*, 22:28–76, 1994.

[89] D. R. Thomas and G. L. Grunkemeier. Confidence interval estimation of survival probabilities for censored data. *Journ. Amer. Statist. Assoc.*, 70(352):865–871, 1975.

[90] G. Tusnady. A remark on the approximation of the sample df in the multidimensional case. *Period. Math. Hungar.*, 8:53–55, 1977.

[91] A.W. Van der Vaart. *Asymptotic Statistics.* Cambridge University Press, 1998.

[92] A.W. Van der Vaart and J.A. Wellner. *Weak convergence and empirical processes.* Springer, 1996.

[93] I. Van Keilegom and D. Varron. Uniform in bandwidth exact rates for a class of kernel estimators. *Ann. Inst. Statist. Math.*, 63(6):1077–1102, 2011.

[94] D. Varron. *Lois fonctionnelles uniformes du logarithme itéré pour les accroissements du processus empirique généralisé. Lois limites de type Chung-Mogulskii pour le processus empirique uniforme local.* PhD thesis, Univ. Paris VI, 2004.

[95] D. Varron. Asymptotic independence in the uniform functional limit law for increments of the empirical process with significatively different bandwidth rates. *Preprint*, 2005.

[96] D. Varron. Some uniform in bandwidth functional results for the tail uniform empirical and quantile processes. *Ann. I.S.U.P.*, 50(1-2):83–103, 2006.

[97] D. Varron. Une remarque concernant les principes de grandes déviations dans les espaces schauder décomposables. *C.R. Acad. Sci. Paris, Ser. I*, 343(5):345–348, 2006.

[98] D. Varron. A limited in bandwidth uniformity for the functional limit law for the increments of the empirical process. *Electron. J. Statist.*, 2:1043–1064, 2008.

[99] D. Varron. Some asymptotic results on density estimators by wavelet projections. *Stat. Proba. Letters*, 78(15):2517–2521, 2008.

[100] D. Varron. A nonstandard uniform functional limit law for the increments of the multivariate empirical distribution function. *Adv. Appl. Stat. Sci.*, 1(2):399–428, 2010.

[101] D. Varron. The almost sure behavior of certain spatial distributions of local empirical processes indexed by functions. *Arxiv preprint*, 1106.0192, 2011.

[102] D. Varron. Clustering rates and chung type functional laws of the iterated logarithm for empirical and quantile processes. *Ann. I.S.U.P.*, 55(2-3):3–26, 2011.

[103] D. Varron. Some new almost sure results on the functional increments of the uniform empirical process. *Stoch. Proc. Appl.*, 121(2):337 – 356, 2011.

[104] D. Varron. A spatial strassen-type functional limit law for the increments of the local empirical process. *Preprint*, 2011.

[105] D. Varron. Donsker and glivenko-cantelli theorems for a class of processes generalizing the empirical process. *Preprint*, 2014.

[106] D. Varron. A note on weak convergence, large deviations, and the bounded approximation property. *Theory Probab. Appl.*, in press, 2014.

[107] D. Varron. Simultaneous confidence bands for some functional plug-in parameters : a computationally feasible approach. *Preprint*, 2014.

[108] D. Varron. The almost sure Kuratowski limit of some local emprical pocesses with variable bandwidths. *Preprint*, 2014.

[109] M. Wschebor. Sur les accroissements du processus de Wiener. *C.R. Acad. Sci. Paris, Ser. I*, 315(12):1293–1296, 1992.

[110] A.Yu. Zaitsev. Estimates of the Levy-Prokhorov distance in the multidimensional central limit theorem for random variables with finite exponential moment. *Theory Probab. Appl.*, 31(2):203–220, 1987.

[111] A.Yu. Zaitsev. On the Gaussian approximation of convolutions under multidimensional analogues of S.N. Bernstein's inequality. *Probab. Theory Related Fields*, 74(4):535–566, 1987.