Probab. Theory Relat. Fields 98, 21–45 (1994)

Probability Theory and Related Fields

Local invariance principles and their application to density estimation

Emmanuel Rio

URA D 0743 CNRS, Modélisation stochastique et statistique, Bât. 425 Mathématique, F-91405 Orsay Cedex France

Received February 12, 1992; in revised form July 14, 1993

Summary. Let $x_1, \ldots, x_n$ be independent random variables with uniform distribution over $[0, 1]^d$, and $X^{(n)}$ be the centered and normalized empirical process associated to $x_1, \ldots, x_n$. Given a Vapnik-Chervonenkis class $\mathscr{S}$ of bounded functions from $[0, 1]^d$ into $\mathbb{R}$ of bounded variation, we apply the one-dimensional dyadic scheme of Komlós, Major and Tusnády to get the best possible rate in Dudley's uniform central limit theorem for the empirical process $\{X^{(n)}(h): h \in \mathscr{S}\}$. When $\mathscr{S}$ fulfills some extra condition, we prove there exists some sequence B_n of Brownian bridges indexed by $\mathscr{S}$ such that

$$\sup_{h \in \mathscr{S}} |X^{(n)}(h) - B_n(h)| = O(n^{-1/2} \log n \vee n^{-1/(2d)} \sqrt{K(\mathscr{S}) \log n}) \text{a.s.}$$

where $K(\mathscr{S})$ denotes the maximal variation of the elements of $\mathscr{S}$. This result is then applied to maximal deviations distributions for kernel density estimators under minimal assumptions on the sequence of bandwith parameters. We also derive some results concerning strong approximations for empirical processes indexed by classes of sets with uniformly small perimeter. For example, it follows from Beck's paper that the above result is optimal, up to a possible factor $\sqrt{\log n}$, when $\mathscr{S}$ is the class of Euclidean balls with radius less than r.

Mathematics Subject Classifications (1991): 60 F17, 62 G 05

1 Introduction and results

Throughout the paper, the probability space Ω is assumed to be rich enough in the following sense: there exists an atomless random variable independent of the observations.

Definitions and notations. Let $I = [0, 1]$ and λ be the Lebesgue measure on I^d. Unless we give more specifications, $d \geqq 2$. Let $x_1, x_2, \ldots$ be a sequence of iid random variables with probability law $\mu = f. \lambda$, where f is a continuous and strictly positive function from the closed unit cube I^d into $\mathbb{R}^{+*}$. We call the empirical process the centered and normalized measure $X^{(n)}$ defined by

$$X^{(n)}(g) = n^{-1/2} \sum_{i=1}^{n} \left(g(x_i) - \int_{I^d} g \, d\mu \right) \tag{1.0}$$

for any integrable funtion g.

Let $\mathscr{S}$ be a class of functions from I^d into $[-1, 1]$. Throughout, we assume that $\mathscr{S}$ is a Vapnik–Chervonenkis class of functions. Let us recall the definition of this notion. Let $\mathscr{P}(I^d)$ denote the set of probability laws on I^d. Let P be in $\mathscr{P}(I^d)$ and let ε be in $]0, 1[$. Let $N(\varepsilon, \mathscr{S}, P)$ denote the maximal cardinality of a subset $\mathscr{S}_\varepsilon$ of $\mathscr{S}$ such that, for any distinct elements g, h of $\mathscr{S}_\varepsilon$, $d_P(g, h) = \int |g - h| \, dP > \varepsilon$. Now, let $\mathscr{A}(I^d)$ denote the set of laws on I^d with finite support. We set

$$N(\varepsilon, \mathscr{S}) = \sup_{P \in \mathscr{A}(I^d)} N(\varepsilon, \mathscr{S}, P).$$

$\log N(\varepsilon, \mathscr{S})$ is called the universal ℓ^1-entropy of $\mathscr{S}$ (see Kolchinsky 1981). When

$$N(\varepsilon, \mathscr{S}) \leqq C(\mathscr{S}) \varepsilon^{-d(\mathscr{S})} \tag{1.1}$$

for any $0 < \varepsilon < 1$, for some constants $C(\mathscr{S})$ and $d(\mathscr{S})$, $\mathscr{S}$ is called a Vapnik–Chervonenkis class of functions (VC class). Now, unless we give other specifications, $\mathscr{S}$ is given the metric d_μ. A Brownian bridge indexed by $\mathscr{S}$ is a centered Gaussian process indexed by $\mathscr{S}$ with covariance function $(g, h) \to \mathbb{E}(g(x_1)h(x_1)) - \mathbb{E}(g(x_1))\mathbb{E}(h(x_1))$. Now, let $(\mathscr{S}_n)_{n>0}$ be a sequence of classes of functions. We say that the strong invariance principle holds for $(\mathscr{S}_n)_{n>0}$ with rate (v_n) if there exists some sequence $(B_n)_{n>0}$ of Brownian bridges indexed by $\mathscr{S}_n$ that are almost surely continuous on $(\mathscr{S}_n, d_\mu)$ such that

$$\sup_{g \in \mathscr{S}_n} |X^{(n)}(g) - B_n(g)| = O(v_n) \text{ a.s.}$$

where $(v_n)_{n>0}$ is some sequence converging to 0. When $\mathscr{S}_n = \mathscr{S}$ is a VC class of functions, according to a result of Dudley (1973), there exists a Brownian bridge indexed by $\mathscr{S}$ with almost surely continuous trajectories on $(\mathscr{S}, d_\mu)$. However, in order to get an invariance principle, some measurability condition is needed (see Dudley 1982 for some counterexample). So, from now on, we assume the following measurability condition $(\mathscr{M})$: there exists some Suslin space Y and some mapping T from Y onto $\mathscr{S}$ such that $(x, y) \to T(y)(x)$ is measurable on $\mathbb{R}^d \times Y$. Let us review the main attempts of getting the best possible rate in the strong invariance principle.

When $d = 1$ and $\mathscr{S}$ is the class of intervals, Komlós et al. (1975) proved that v_n may be taken as $n^{-1/2} \log n$. This result is optimal (see Csörgó and Révész's book 1981) and, up to now, it is the only result having important applications.

When $\mathscr{S}$ is a VC class of Borel sets of I^d satisfying a uniform perimetric condition, Massart (1989) proved that v_n may be taken as $n^{-1/(2d)}(\log n)^{3/2}$, via a multivariate extension of the construction of Komlós et al. (1975). The rate of convergence appearing here is nearly optimal when $\mathscr{S}$ is the class of Euclidean balls. Kolchinsky (1991) applied Massart's exponential bounds for the multinomial embedding of Komlós et al. (1975) to the strong approximation of function indexed empirical processes. In a recent paper, he characterizes the accuracy of the empirical process indexed by bounded functions by their accuracy to some Haar expansion and he further improves Massart's exponential bounds of a factor $\sqrt{\log n}$ (see

Kolchinsky 1992). However, these global rates do not provide optimal applications for kernel density estimators (see Konakov and Piterbarg 1984). In order to obtain the limiting behavior of kernel density estimators under minimal assumptions on the bandwith parameters, we shall give an invariance principle with an explicit dependance in the maximal variation of the elements of $\mathscr{S}$ (note that the variation of the characteristic function of a set is equal to the perimeter of this set). Applying this invariance principle to a sequence of classes of functions whose maximal variation decrease to 0 then leads to optimal results, concerning the sequence of bandwith parameters, for the maximal deviation of kernel density estimators.

So, we study VC classes $\mathscr{S}$ of functions with *uniformly bounded variation.* Refining Massart's method as in Rio (1993), we prove invariance principles with an error term depending explicitly on the maximal variation of the elements of $\mathscr{S}$ in Sect. 3. These invariance principles are applied to kernel density estimators in Sect. 4.

Statement of results. Throughout, we assume the elements of $\mathscr{S}$ to be of bounded variation. Recall this means that the partial derivatives of any element of $\mathscr{S}$ are Radon measures. Let $\mathscr{D}_d(I^d)$ denote the space of C^∞ functions with values in $\mathbb{R}^d$ and with compact support included in I^d. $\mathbb{R}^d$ being given the usual Euclidean norm, we set $\| g \|_\infty = \sup_{x \in \mathbb{R}^d} \| g(x) \|$. For any function h of $\mathscr{S}$, we set:

$$K(h, I^d) = \sup_{g \in \mathscr{D}_d(I^d)} \left(\int_{\mathbb{R}^d} h(x) \operatorname{div} g(x) dx / \| g \|_\infty \right).$$

A classical result of distribution theory (see Schwartz 1957) ensures that $K(h, I^d)$ is finite if and only if h is of bounded variation.

Definition. We shall say that the class $\mathscr{S}$ is *uniformly of bounded variation* (*condition* UBV) if $K(\mathscr{S}) = \sup_{h \in \mathscr{S}} K(h, I^d) < +\infty$.

When $\mathscr{S}$ is a class of characteristic functions of Borel sets S, $K(\mathbb{1}_S, I^d)$ is the De Giorgi perimeter of S related to I^d. So, the classes of Borel sets of I^d satisfying condition UBV are exactly the classes of sets with uniformly bounded perimeter. Let us give some examples of such classes. If $\mathscr{S}$ satisfies condition UM, then $\mathscr{S}$ satisfies condition UBV. Hence the class of convex sets and the classes of Borel sets with uniformly Lipschitz boundaries in the sense of Dudley (1974) fulfill condition UBV. However, in order to get optimal invariance principles for classes with uniformly small variation, we need to impose an extra condition: we will assume that $\mathscr{S}$ fufills the condition below, called *condition* LUBV.

Definition. We shall say that the class $\mathscr{S}$ is *locally uniformly of bounded variation* if for any $a \in]0, 1[$, any cube $\mathscr{C} \subset I^d$ with edges of length a parallel to coordinate axes and any h in $\mathscr{S}$, $K(h, \mathscr{C}) \leqq K_{\text{loc}}(\mathscr{S}) a^{d-1}$ for some constant $K_{\text{loc}}(\mathscr{S}) \geqq 1$ depending only on $\mathscr{S}$, where $K(h, \mathscr{C})$ is defined from $\mathscr{C}$ and h exactly as $K(h, I^d)$ is defined from I^d and h.

Let us now state the main result, providing an invariance principle with an explicit error term depending only on $K(\mathscr{S})$ and on the entropy of the class $\mathscr{S}$.

Theorem 1.1 *Let $d \geqq 2$ and $\mathscr{S}$ be a Vapnik-Chervonenkis class of functions from I^d into* $[-1, 1]$ *satisfying* 1.1, *the condition* UBV *for some constant $K(\mathscr{S})$, and the measurability condition* ($\mathscr{M}$). *Then, there exists a Brownian bridge B_n indexed by*

$\mathcal{S}$ *with almost surely continuous trajectories on* $(\mathcal{S}, d_\mu)$ *such that, for any positive* $t \geqq C\log n$,

$$\mathbb{P}\left(\sqrt{n}\sup_{h\in\mathcal{S}}|X^{(n)}(h) - B_n(h)| \geqq C\sqrt{n^{(d-1)/d}K(\mathcal{S})t} + Cc(n)t\right) \leqq e^{-t}$$

where $c(n) = \sqrt{\log n}$ *in the general case,* $c(n) = \sqrt{K(\mathcal{S})}$ *under condition* LUBV *and* C *is a positive constant depending only on* d, $d(\mathcal{S})$ *and* $C(\mathcal{S})$.

We now give an application of Theorem 1.1 to invariance principles for empirical processes indexed by VC classes of Borel sets of I^d. We need to recall some nice properties of these classes. When $\mathcal{S}$ is a VC class of sets,

$$D(\mathcal{S}) = \sup\{D \in \mathbb{N}: \mathrm{Card}(A \cap \mathcal{S}) = 2^D \text{ for some set } A \text{ with } \mathrm{Card}\,A = D\} < \infty,$$

where $\mathrm{Card}\,A$ denotes from now on the cardinality of the finite set A and $A \cap \mathcal{S} = \{A \cap S\colon S \in \mathcal{S}\}$. We call $D(\mathcal{S})$ the entire density of $\mathcal{S}$. This result provides many examples of VC classes. For example, the class of closed half spaces or the class of closed Euclidean balls are VC classes. Moreover, a lemma of Dudley (1978) ensures that the universal entropy $N(\varepsilon, \mathcal{S})$ satisfies: for any $0 < \varepsilon < 1/2$,

$$N(\varepsilon, \mathcal{S}) \leqq C(D)(\varepsilon^{-1}|\log\varepsilon|)^{D(\mathcal{S})}. \tag{1.2}$$

Theorem 1.1 and (1.2) yield the corollary below.

Corollary 1.1 *Let* $d \geqq 2$. *Let* D *be a positive integer and let* $(\mathcal{S}_n)_{n>0}$ *be a sequence of Vapnik–Chervonenkis classes of Borel sets of* $\mathbb{R}^d$ *with entire densities each bounded by* D, *satisfying the condition* LUBV *for some constant* K_{loc} *and the measurability condition* $(\mathcal{M})$. *Then, the strong invariance principle holds for* $(\mathcal{S}_n)_{n>0}$ *with rate*

$$v_n = n^{-1/(2d)}\sqrt{K(\mathcal{S}_n)\log n} + n^{-1/2}\sqrt{K_{\mathrm{loc}}}\log n \quad \text{a.s.}$$

Remarks. Note that $K(\mathbb{1}_S, I^d)$ is the De Giorgi perimeter of S related to I^d. Hence, the classes of Borel sets of I^d satisfying the condition UBV are exactly the classes of sets with uniformly bounded perimeter.

When $\mathcal{S}_n$ is the class of closed Euclidean balls with radius less than h_n, for some sequence $(h_n)_{n>0}$ converging to 0 and satisfying $nh_n^d > (\log n)^{d/(d-1)}$, the error term is of the order of $(nh_n^d)^{(d-1)/(2d)}\sqrt{\log n}$. By Theorem 1 of Beck's paper (1985), this result is optimal, up to an eventual factor $\sqrt{\log n}$. On the other hand, when $nh_n^d < (\log n)^{d/(d-1)}$, the error term is of the order of $\log n$. From multivariate Erdös-Rényi laws for balls-indexed empirical processes, it follows that this rate cannot be improved when $\lim_{n\to+\infty} nh_n^d/\log n = +\infty$ (see Louani 1992). Moreover, when $(h_n)_{n>0}$ satisfies $\liminf_{n\to+\infty}|\log h_n|/\log n = \delta > 0$ and the usual condition $\lim_{n\to+\infty} nh_n^d/\log n = +\infty$, Corollary 1.1 ensures that

$$\sup_{S\in\mathcal{S}_n}|X^{(n)}(S) - B_n(S)| = o((h_n^d|\log h_n|)^{1/2}) \text{ a.s.}$$

Let π_d denote the volume of the Euclidean unit ball in $\mathbb{R}^d$. According to the above result, we can obtain

$$(2d\pi_d h_n^d|\log h_n|)^{-1/2}\sup_{S\in\mathcal{S}_n}|X^{(n)}(S)| \overset{P}{\to} 1$$

from the corresponding result for the Brownian bridge. Hence, Corollary 1.1 works as soon as the Erdös–Rényi law for the empirical process fails.

In Sect. 4, we give an application of the results of Sect. 3 to the maximal deviation of kernel density estimators. Assume that the strictly positive density f satisfies the additional smoothness condition below:

the density f is β-Hölderian on the closed cube I^d, for some $\beta \in]0, 1]$. (1.3)

Let Ψ be a two-times continuously differentiable kernel function with compact support. Define the Parzen–Rosenblatt estimator of the density f by:

$$f_n(x) = (nh_n^d)^{-1} \sum_{i=1}^{n} \Psi(h_n^{-1}(x - x_i)).$$

Let $\xi_n(x)$ be the normalized deviation field of f_n, i.e.

$$\xi_n(x) = \sqrt{nh_n^d}\,\frac{f_n(x) - \mathbb{E}(f_n(x))}{\sigma_\Psi \sqrt{f(x)}}, \tag{1.4}$$

where $\sigma_\Psi^2 = \int \Psi^2(y)\,dy$. We obtain from Theorem 3.2 of Konakov and Piterbarg (1984) and from an invariance principle for $\xi_n(x)$ which will be stated in Sect. 4 the following result.

Theorem 1.2 *Let $d \geqq 2$ and let $x_1, x_2, \ldots$ be a sequence of* iid *random variables with law μ fulfilling* 1.3. *Let Ψ be a C^2 kernel function from $\mathbb{R}^d$ into $[-1, 1]$ with compact support, satisfying $\int \Psi(y)\,dy = 1$. Let T be any closed Jordan set included in $]0, 1[^d$ with positive Lebesgue measure and let $(h_n)_{n>0}$ be a sequence of bandwith parameters converging to* 0 *and satisfying the condition*

$$0 < \liminf_{n \to +\infty} (\log h_n^{-1}/\log n) \leqq \limsup_{n \to +\infty} (\log h_n^{-1}/\log n) < 1/d. \tag{1.5}$$

Let ξ_n be the deviation field defined by (1.4). *Then there exists some positive γ such that:*

$$\mathbb{P}\left(l_n \sup_{x \in T} |\xi_n(x)| - l_n^2 < t \right)$$

$$= \exp\left\{ -2\exp\left(-t - \frac{t^2}{2l_n^2} \right) \sum_{0 \leqq m \leqq (d-1)/2} e_m l_n^{-2m} \left(1 + \frac{t}{l_n^2} \right)^{d-2m-1} \right\} + O(n^{-\gamma})$$

uniformly in t, where l_n is the maximal root of the equation

$$\lambda(T)(2\pi)^{-(d+1)/2} h_n^{-d} \sqrt{\Lambda}\, l^{d-1} \exp(-l^2/2) = 1.$$

λ is the Lebesgue measure in $\mathbb{R}^d$, $\Lambda = \det(a_{ij})$, $a_{ij} = \sigma_\psi^{-2} \int (\nabla_i \Psi)(x)(\nabla_j \Psi)(x)\,dx$ and $e_m = (-1)^m (d-1)!/(m!\,2^m (d-2m-1)!)$ is mth Hermite polynomial coefficient of order $d-1$.

Remarks. Assume that f belongs to the class $J(I^d, \alpha)$ of α-differentiable functions on I^d, as defined in Dudley (1982). Theorem 1.2 works when $h_n \approx n^{-1/(d+2\alpha)}$ is the *optimal bandwith size* as soon as $\alpha > 0$. Recall Konakov and Piterbarg (1984) had to assume the regularity condition $\alpha > d^2/2$ in their Theorem 1.2.

Note that $l_n \sim \sqrt{2d|\log h_n|}$ as $n \to +\infty$. See Konakov and Piterbarg (1984) for an asymptotic expansion of l_n. Our Theorem 1.2 and their Theorem 3.2 ensure also that

$$\sup_{x \in I^d} l_n^{-1} |\xi_n(x)| \overset{P}{\to} 1 \tag{1.6}$$

if the sequence $(h_n)_{n>0}$ satisfies $\lim_{n\to+\infty} nh_n^d/\log n = +\infty$. This result is optimal. However, in order to obtain the limiting distribution of $l_n \sup_{x\in I^d} |\xi_n(x)| - l_n^2$, we need the stronger condition $\lim_{n\to+\infty} nh_n^d/(\log n)^{2d} = +\infty$ (cf. Theorem 4.1).

The proof of Theorem 1.1 is mainly based on some Bernstein type inequality for the multinomial embedding of Komlós et al. (1975), which is proved in Sect. 2. Next we derive the strong invariance principles for multivariate empirical processes using a multivariate Haar expansion of the functions.

2 Strong approximation

Throughout this section, we will take $d = 1$. Let $(x_1, x_2, \ldots, x_n)$ be an n-sample of the uniform distribution over $[0, 1]$. Let N be a positive integer and $\mathscr{E}_N$ be the class of functions f from $]0, 1]$ into $\mathbb{R}$ such that

$$f = \sum_{i=1}^{2^N} f_i \mathbb{1}_{](i-1)2^{-N}, i2^{-N}]}$$

for some $e(f) = (f_1, f_2, \ldots, f_{2^N})$ with values in $\mathbb{R}^{2^N}$. Now, let $B_n(t)$ be a standard Brownian bridge defined from $(x_1, x_2, \ldots, x_n)$ via Komlós et al.'s method (in fact only the increments of the empirical d.f. and of $B_n(.)$ between time $(i-1)2^{-N}$ and $i2^{-N}$ are defined in a common probability space; a lemma of Skorohod (1976) ensures then that the construction of these two processes may be performed on a *rich enough* probability space Ω). Now, for each f in $\mathscr{E}_N$, we set:

$$Z^{(n)}(f) = \int_0^1 f(u) dB_n(u) .$$

The main aim of this section is to generalize the probabilistic bounds of Komlós et al. (1975) to bounds with an error term depending mainly on the coefficients of the orthogonal expansion of f in the Haar basis on the unit interval.

Notations. Throughout the sequel, the intervals $]l, m]$ are to be interpreted as subsets of $\mathbb{Z}_+$. $\ell^2(\mathbb{Z}_+)$ is given the canonical inner product, denoted by $(.|.)$ $\ell^2(]l, m])$ denotes the subspace of $\ell^2(\mathbb{Z}_+)$ of functions with support included in $]l, m]$. Let $I_{j,k} =]k2^j, (k+1)2^j]$, and let $e_{j,k}$ be the characteristic function of $I_{j,k}$. For any positive integers j and k, we set $\tilde{e}_{j,k} = 2e_{j-1,2k} - e_{j,k}$. Let $\mathscr{B} = \{\tilde{e}_{j,k}: 1 \leq j \leq N, 0 \leq k < 2^{N-j}\}$. Clearly, $\mathscr{B}' = \mathscr{B} \cup \{e_{N,0}\}$ is an orthogonal basis of $\ell^2(]0, 2^N])$ with $(\tilde{e}_{j,k} | \tilde{e}_{j,k}) = 2^j$ and $(e_{N,0} | e_{N,0}) = 2^N$. For any $f \in \mathscr{E}_N$, we define the coefficients $\gamma_{j,k}(f)$ and $\bar{\gamma}_N(f)$ by $\gamma_{j,p}(f) = 2^{-j}(e(f) | \tilde{e}_{j,k})$ and $\bar{\gamma}_N(f) = 2^{-N}(e(f) | \tilde{e}_{N,0})$. Define the Hilbertian pseudonorm $\|f\|_{\mathscr{B}}$ by

$$\|f\|_{\mathscr{B}}^2 = 4 \sum_{j=1}^{N} \sum_{0 \leq k < 2^{N-j}} \gamma_{j,k}^2(f).$$

Let us now state the basic inequality.

Theorem 2.1 *For any vector* $p = (p_1, \ldots, p_N)$ *with positive components such that* $\Sigma_{i=1}^N p_i \leqq 1$, *let* $q_i = (2^i p_i)^{-1}$ *and let*

$$M(p,f) = 4 \sup_{(j,k)} \left\{ \sum_{l<j} q_{j-l} \sum_{m: I_{l,m} \subset I_{j,k}} \gamma_{l,m}^2(f) \right\}.$$

Then, for any f *in* $\mathscr{E}_N$ *and any positive* x,

$$\mathbb{P}\Big(\sqrt{n}\,|X^{(n)}(f) - Z^{(n)}(f)| \geqq (\sqrt{M(p,f)} + C_f)x + \Big(\Big(\sum_{i=1}^N \frac{q_i}{2}\Big)^{1/2} + C_1\Big)\|f\|_{\mathscr{B}}\sqrt{x}\Big) \leqq 2e^{-x}$$

where $C_f = (\sup_{0<x<1} f(x) - \inf_{0<x<1} f(x))/4$ *and* $C_1 = 1 + \sqrt{83/32}$.

Remarks. Theorem 2.1 can be used to obtain numerical constants in the refinement of Komlós et al.'s exponential inequality proved in the paper of Mason and Van Zwet (1987). Let $x_1, x_2, \ldots$ be iid random variables with uniform distribution over [0, 1] and let $F_n(t) = n^{-1}\Sigma_{i=1}^n \mathbb{1}_{(x_i \leqq t)}$ be the associated distribution function. Refining Komlós et al.'s exponential inequality, Mason and Van Zwet (1987) proved that there exists a standard Brownian bridge B_n with almost surely uniformly continuous trajectories such that, for all real d in $[8, n]$, for all positive x,

$$\mathbb{P}\Big(\sup_{0<nt<d} |n(F_n(t) - t) - \sqrt{n}B_n(t)| \geqq ax + b\log d + c\log 2\Big) \leqq \exp(-x),$$

where a, b, and c are positive universal constants. Using the results stated in this section, we prove in a preprint (see Rio 1991) that this inequality holds with $a = 3.26$, $b = 4.86$, and $c = 9.26$. These results further improve the previous numerical results of Bretagnolle and Massart (1989).

Proof of Theorem 2.1 Let us recall some results stated in Tusnády's thesis and proved in the paper by Bretagnolle and Massart (1989) concerning the quantile transformation of a random variable with Binomial law. For any d.f. F, let $F^{-1}(t) = \inf\{x: F(x) \geqq t\}$.

Lemma. *Let* Φ *denote the d.f. of a random variable with standard normal law and let* Y *be a random variable with d.f.* Φ. *Denoting by* Φ_n *the d.f. of the binomial law* $B(n, 1/2)$, *set* $B_n = \Phi_n^{-1} \circ \Phi(Y) - n/2$. *Then, the two following inequalities hold*:

(i) $$|B_n| \leqq 1 + (\sqrt{n}/2)|Y|.$$

(ii) $$|B_n - (\sqrt{n}/2)Y| \leqq 1 + (Y^2/8)\,.$$

Let us give an outline of the construction of Komlós et al. (1975). Denote the increments between time $(i-1)2^{-N}$ and $i2^{-N}$ of the empirical process Z_n and of B_n by X_i and Z_i, i.e.

$$X_i = \sum_{k=1}^n \mathbb{1}_{(i-1<2^N x_k \leqq i)} - n2^{-N} \text{ and } Z_i = \sqrt{n}(B_n(i2^{-N}) - B_n((i-1)2^{-N}))\,. \quad (2.0)$$

Let the empirical measures X and Z be defined by:

$$X(.) = \sum_{i=1}^{2^N} X_i \mathbb{1}_{](i-1)2^{-N}, i2^{-N}]} \lambda(.) \text{ and } Z(.) = \sum_{i=1}^{2^N} Z_i \mathbb{1}_{](i-1)2^{-N}, i2^{-N}]} \lambda(.) .$$

Then, with the above notations, for all f in $\mathscr{E}_N$, $X(f) = \sqrt{n} X^{(n)}(f)$ and $Z(f) = \sqrt{n} Z^{(n)}(f)$. For any positive integer j, for any $0 \leqq k < 2^{N-j}$, we set:

$$\tilde{U}_{j,k} = X(\tilde{e}_{j,k}), \text{ and } U_{j,k} = \text{Card}\{i \leqq n: 2^N x_i \in I_{j,k}\} = X(e_{j,k}) + n2^{j-N} .$$

We also set:

$$\tilde{V}_{j,k} = Z(\tilde{e}_{j,k}), V_{j,k} = Z(e_{j,k}), \tilde{\xi}_{j,k} = \tilde{V}_{j,k} \sqrt{n^{-1} 2^{N-j}}.$$

Since $\mathscr{B}$ is a basis of $\ell^2(]0, 2^N])$, it is necessary to define the random variables $\tilde{U}_{j,k}$ only from the corresponding Gaussian increments $\tilde{V}_{j,k}$. So, we will define the random variables by induction on j. Assume that the random variables $(\tilde{U}_{l,m})_{l > j, 0 \leqq m < 2^{N-l}}$ have been already associated to the corresponding random variables $\tilde{V}_{l,m}$. Then, the random variables $(U_{l,m})_{0 \leqq m < 2^{N-l}}$ are determined since the system $(e_{j,k})_{0 \leqq k < 2^{N-j}}$ belongs to the linear span of $\{\tilde{e}_{l,m}: l > j, 0 \leqq k < 2^{N-m}\} \cup \{e_{N,0}\}$.

Now, we claim that the random variables $((U_{j,k} + U_{j,k})/2)_{0 \leqq p < 2^{N-j}}$ are independent and $B(U_{j,k}, 1/2)$-distributed, conditional on the random variables $(U_{j,k})_{0 \leqq k < 2^{N-j}}$. Moreover, since $\mathscr{B}$ is an orthogonal basis, the random variables $(\tilde{\xi}_{j,k})_{0 \leqq k < 2^{N-j}}$ are independent and $N(0,1)$-distributed, conditional on $(U_{j,k})_{0 \leqq k < 2^{N-j}}$. Hence, for all $k \in [0, 2^{N-j}[$, if we define $\tilde{U}_{j,k}$ by:

$$\tilde{U}_{j,k} + U_{j,k} = 2\Phi_{U_{j,k}}^{-1} \circ \Phi(\tilde{\xi}_{j,k}) ,$$

the random variables $(\tilde{U}_{j,k})_{0 \leqq k < 2^{N-j}}$ will have the prescribed conditional distribution (see Bretagnolle and Massart 1989). By Tusnády's lemma, it follows that, for all j and k,

$$|\tilde{U}_{j,k} - (U_{j,k})^{1/2} \tilde{\xi}_{j,k}| \leqq 2 + \tilde{\xi}_{j,k}^2/4 .$$

Now, we use the dyadic decomposition of $Z(f)$ and $X(f)$:

$$X(f) = \sum_{j=1}^{N} \sum_{0 \leqq k < 2^{N-j}} \gamma_{j,k}(f) \tilde{U}_{j,k} \text{ and } Z(f) = \sum_{j=1}^{N} \sum_{0 \leqq k < 2^{N-j}} \gamma_{j,k}(f) \tilde{V}_{j,k}.$$

In order to apply Tusnády's lemma, we define an auxiliary empirical measure Y by:

$$Y(f) = \sum_{j=1}^{N} \sum_{0 \leqq k < 2^{N-j}} \gamma_{j,k}(f) (U_{j,k})^{1/2} \tilde{\xi}_{j,k} .$$

Let $\Delta(f) = X(f) - Z(f)$, and let $\Delta_1(f) = (X - Y)(f)$, $\Delta_2(f) = (Y - Z)(f)$. Clearly, $\Delta(f) = \Delta_1(f) + \Delta_2(f)$. Now, we state a lemma which proves that, in order to control $\Delta(f)$, it will be sufficient to control each of the moment-generating functions of $\Delta_1(f)$ and of $\Delta_2(f)$.

Let A be a random variable with mean zero and with finite moment-generating function in a neighborhood of zero. We set $\gamma_A(t) = \log \mathbb{E}(\exp(tA))$. Then, the Chernoff function h_A of A is the Legendre transform of γ_A ($h_A(\varepsilon) = \sup_{t>0} ((t\varepsilon - \gamma_A(t))$ for all positive ε and $h_A(\varepsilon) = \sup_{t<0}(t\varepsilon - \gamma_A(t))$ otherwise). Clearly, h_A takes its values in $\mathbb{R}^+ \cup \{+\infty\}$ and is a nondecreasing convex function from $\mathbb{R}^+$ into $\mathbb{R}^+ \cup \{+\infty\}$. Define the function h_A^{-1} for any nonnegative x by:

$h_A^{-1}(x) = \inf\{t : h_A(t) \geqq x\}$. From Chernoff's result, it follows that, for any positive x, $\mathbb{P}(A > h_A^{-1}(x)) \leqq \exp(-x)$.

Hence, the result below, which will be proved in Appendix A, yields an exponential bound for the sum of two random variables with finite moment-generating function.

Lemma 2.1 *Let A and B be two centered random variables with finite moment-generating function in a neighborhood of 0. Assume that $\mathbb{P}(A = 0) < 1$ and $\mathbb{P}(B = 0) < 1$. Let h_A, h_B, and h_{A+B} denote respectively the Cramer–Chernoff functions of A, B, and $A + B$. Then, for all positive x, $h_{A+B}^{-1}(x) \leqq h_A^{-1}(x) + h_B^{-1}(x)$.*

For the sake of brevity, we set $\Delta_i(f) = \Delta_i$. Using Lemma 2.1, it will be sufficient to control each of the Laplace transforms of the random variables Δ_i. We control the Laplace transform of Δ_1 via the following lemma, which will be proved in Appendix B.

Lemma 2.2 *For any f in $\mathscr{E}_N$ and for any real t such that $C_f|t| < 1$,*

$$128 C_f^2 \log \mathbb{E}(\exp(t\Delta_1)) \leqq -83 \|f\|_{\mathscr{B}}^2 \log(1 - (C_f t)^2)\,.$$

On the other hand, the control of Δ_2 is ensured by the two lemmas below, which will be proved in Appendix C and Appendix D.

Control of Δ_2. Throughout, we will use the following notations. Let

$$\mathscr{L} = \{(j,k) : 1 \leqq j < N, 0 \leqq k < 2^{N-j}\} \text{ and } \bar{\mathscr{L}} = \mathscr{L} \cup \{(N,0)\}\,. \tag{2.1}$$

Also, let the order relation $\preceq$ be defined on $\bar{\mathscr{L}}$ as follows: if l and m are two elements of $\bar{\mathscr{L}}$, $l \preceq m$ iff $I_l \subset I_m$. We also define $\prec$ by $l \prec m$ iff $l \preceq m$ and $l \neq m$. Let $l = (j,k)$ be any element of $\bar{\mathscr{L}}$: we set $j = |l|$. We set also $|m| - |l| = |m - l|$ for any $l \prec m$. So, with the above notations, we have:

$$\Delta_2 = \sum_{l \in \mathscr{L}} \gamma_l(f)(\sqrt{U_l} - \sqrt{\mathbb{E}(U_l)})\tilde{\xi}_l\,.$$

We control the moment-generating function of Δ_2 by the moment-generating function of a random variable Δ_3 depending only on the Gaussian increments $(\tilde{\xi}_l)_{l \in \bar{\mathscr{L}}}$. This is the purpose of the lemma below.

Lemma 2.3 *For any real t, $\mathbb{E}(\exp(t\Delta_2)) \leqq \mathbb{E}(\exp(t\Delta_3))$ where:*

$$\Delta_3 = \sum_{l \in \mathscr{L}} \gamma_l(f)\tilde{\xi}_l\left(\sqrt{2} + \sum_{m : l \prec m} 2^{-|m-l|/2}|\tilde{\xi}_m|\right).$$

It remains to control the Laplace transform of Δ_3. Let $\Delta_4 = \sqrt{2}\sum_{l \in \mathscr{L}} \gamma_l(f)\tilde{\xi}_l$ and $\Delta_5 = \Delta_3 - \Delta_4$. Using Lemma 2.1 again, it is sufficient to control each of the moment-generating functions of Δ_4 and Δ_5. Since Δ_4 is a Gaussian random variable with zero-mean and variance $\|f\|_{\mathscr{B}}^2/2$, the control of Δ_4 will be ensured via standard arguments of Cramer-Chernoff calculation on the normal random variables. Let us now state an upper bound on the Laplace transform of Δ_5 which depends mainly on the number $M(p,f)$ previously defined in Theorem 2.1, and on the variance term $\|f\|_{\mathscr{B}}^2$.

Lemma 2.4 *Let $p = (p_1, \ldots, p_N)$ be a vector with positive components such that $\sum_{i=1}^{N} p_i \leqq 1$. Let $q_i = (2^i p_i)^{-1}$. For all m in $\mathscr{L}$, we set:*

$$M_m = 4 \sum_{l : l \prec m} q_{|m-l|} \gamma_l^2(f) .$$

Let $M = \sup_{m \in \mathscr{L}} M_m$. Then, for any real t such that $Mt^2 \leqq 1$,

$$\log \mathbb{E}(\exp(t\Delta_5)) \leqq -\frac{\|f\|_{\mathscr{B}}^2}{2M} \sum_{i=1}^{N} q_i \log((1 + \sqrt{1 - Mt^2})/2) .$$

Using Lemma 2.1, Lemma 2.3 and the results already proved in this section, we get:

$$\mathbb{P}(\Delta(f) \geqq h_1^{-1}(x) + h_5^{-1}(x) + \|f\|_{\mathscr{B}}\sqrt{x}) \leqq e^{-x} \tag{2.2}$$

for any positive x, where h_1 and h_5 denote the Cramer-Chernoff functions of the random variables $\Delta_1(f)$ and $\Delta_5(f)$. Using Lemma 2.2 and Lemma 2.4, we prove then that, for all positive t such that $C_f t < 1$,

$$\log \mathbb{E}(\exp(t\Delta_1)) \leqq \frac{83\|f\|_{\mathscr{B}}^2 t^2}{128(1 - C_f t)} \tag{2.3a}$$

and that, for all positive t such that $t\sqrt{M} < 1$,

$$\log \mathbb{E}(\exp(t\Delta_5)) \leqq \left(\sum_{i=1}^{N} q_i\right) \frac{\|f\|_{\mathscr{B}}^2 t^2}{8(1 - t\sqrt{M})} . \tag{2.3b}$$

Proof of (2.3) Taking into account Lemma 2.2, Lemma 2.4, and standard arguments of homogeneity, note that, for all t in $[0, 1[$,

$$-4\log((1 + \sqrt{1 - t^2})/2) \leqq -\log(1 - t^2) \leqq t^2/(1 - t) . \tag{2.4}$$

The proof of (2.4) will be omitted, since it only uses elementary calculations. So, (2.3) holds.

Let h^{-1} denote the Legendre transform of the function $\gamma(t) = t^2/(1 - t)$. For any positive x, $h^{-1}(x) = x + 2\sqrt{x}$. Both (2.2), (2.3) and the above inequality then imply Theorem 2.1. □

3 A local invariance principle for empirical processes

Throughout the section, $d \geqq 2$. Using the upper bounds for the embedding of Komlós et al. proved in Sect. 2, we study the rates existing in the strong invariance principle for multivariate empirical processes.

Proof of Theorem 1.1 First, we use the so called multivariate quantile transformation to transform the r.v's x_i into r.v's with uniform distribution over I^d. Second we use the multivariate adaptation of construction method proposed in our previous paper (see Rio 1993) to construct the homogeneous empirical bridge and the corresponding Brownian bridge in a common probability space. Next, using the results of Sect. 2 and the conditions UBV and LUBV, we give uniform exponential bounds in h on the error term $|X^{(n)}(h) - B_n(h)|$, and we derive Theorem 1.1 from these bounds via Massart's oscillation controls (1986) for VC classes.

Transformation of the r.v.'s. Let F denote the multivariate quantile transformation from the closed unit cube onto itself, which turns a r.v. with density f into a uniformly distributed r.v. This transformation will be called Rosenblatt transformation; (see Rosenblatt 1952). Under the assumptions of Theorem 1.1, F is a diffeomorphism from I^d onto I^d.

Assume now that Theorem 1.1 holds for r.v.'s y_i with uniform distribution over I^d and let B_n° denote the corresponding approximating homogeneous Brownian bridges with almost surely continuous trajectories on $(\mathcal{S}, d_\lambda)$. Define $x_i = F^{-1}(y_i)$ and let

$$\mathcal{K} = \{g \circ F^{-1} : g \in \mathcal{S}\} . \tag{3.1}$$

Since F is a diffeomorphism, $\mathcal{K}$ is a VC class of functions satisfying the measurability condition $(\mathcal{M})$, with the same entropy function as $\mathcal{S}$. Moreover, $\mathcal{K}$ satisfies either the condition LUBV for some constant K_{loc} depending only on f or the condition UBV. Hence, Theorem 1.1 in the general case follows from Theorem 1.1 in the special case of the Lebesgue measure and from the fact that the gaussian process B_n defined by $B_n(g) = B_n^\circ(g \circ F^{-1})$ is a Brownian bridge with the prescribed covariance function (to prove this fact, notice that $|\operatorname{Jac} F(x)| = f(x)$). Let us now prove Theorem 1.1 for uniformly distributed r.v.'s.

Construction in a common probability space. Let $\mathcal{R}_d$ denote the class of characteristic functions of closed boxes and let $\bar{\mathcal{S}} = \mathcal{S} \cup \mathcal{R}_d$. Clearly, the so completed class $\bar{\mathcal{S}}$ is a Vapnik-Chervonenkis class of functions. Here, it will be convenient to define the processes $X^{(n)}$ and B_n on this completed class $\bar{\mathcal{S}}$. Let N be the integer such that $2^{Nd} \leqq n < 2^{(N+1)d}$. Divide each coordinate segment into 2^N intervals of the length 2^{-N}. This division generates the partition of $[0, 1]^d$ on cubes volumes of which are equal to 2^{-Nd}. For each $p = (p_1, \ldots, p_d)$ in $\mathbb{Z}_+^d$, let $C_{0,p}$ denote the open cube of volume 2^{-Nd} with lower-left vertice $2^{-N}p = (2^{-N}p_1, \ldots, 2^{-N}p_d)$. We define the increments of B_n and $X^{(n)}$ on the cubes $C_{0,p}$ via the multinomial embedding of Komlós et al. Let σ denote the one to one function from $\mathbb{Z}_+^d$ onto $\mathbb{Z}_+$ already defined in Rio (1993): σ maps the cubes $]0, 2^N]^d$ onto the intervals $]0, 2^{Nd}]$. Hence, by a lemma of Skorohod (1976), there exists a sequence $(x_1, \ldots, x_n)$ of independent random variables uniformly distributed in $[0, 1]^d$ and a standard Brownian bridge with almost surely continuous trajectories on $(\bar{\mathcal{S}}, d_\lambda)$ such that

$$B_n(C_{0,p}) = Z_{\sigma(p)} \text{ and } X^{(n)}(C_{0,p}) = X_{\sigma(p)} , \tag{3.2}$$

where the random variables $(X_i)_{0 < i \leqq 2^{Nd}}$ and $(Z_i)_{0 < i \leqq 2^{Nd}}$ are defined in Sect. 2 by Eqs. (2.0). We now need to recall the definition and the properties of σ.

$\mathbb{Z}^d$ is provided with the usual sum and product. Let

$$J = \{(j_1, j_2, \ldots, j_d) \in \mathbb{N}^d \text{ such that } j_1 \leqq j_2 \leqq \ldots \leqq j_d \leqq j_1 + 1\} .$$

It is obvious that $(j_1, \ldots, j_d) \to (j_1 + j_2 + \cdots + j_d)$ is a one to one mapping from J onto $\mathbb{N}$. For each integer j, we call $(j_1, j_2, \ldots, j_d)$ the unique element of J such that $j = j_1 + j_2 + \cdots + j_d$. Let R_j be the lattice of multiples of $(2^{j_1}, 2^{j_2}, \ldots, 2^{j_d})$: we define the box $C'_{j,p}$ for any p of R_j by (here $\mathbb{1} = (1, 1, \ldots, 1) \in \mathbb{R}^d$):

$$C'_{j,p} = \{x \in \mathbb{Z}^d, p + \mathbb{1} \leqq x \leqq p + (2^{j_1}, \ldots, 2^{j_d})\}.$$

The following lemma holds (see Rio 1993).

Lemma 3.1 *There exists a one to one map σ from $\mathbb{Z}_+^d$ to $\mathbb{Z}_+$ mapping the boxes $C'_{N,0}$ onto the intervals $]0, 2^N]$, and the boxes $(C'_{j,p})_{p\in R_j}$ onto the intervals $I_{j,q} =]q2^j, (q+1)2^j]$.*

We now give exponential bounds on the error term $|X^{(n)}(h) - B_n(h)|$ for the so constructed processes.

Lemma 3.2 *For any positive x and any h in $\mathcal{S}$,*

$$\mathbb{P}(\sqrt{n}|X^{(n)}(h) - B_n(h)| \geqq C_6 n^{(d-1)/(2d)}\sqrt{K(\mathcal{S})x} + C(N)x) \leqq 8\exp(-x)$$

where C_6 is some constant depending only on d, $C(N) = 3 + \sqrt{8dN}$ in the general case, and $C(N) = (3 + 4d^{3/2})K_{\text{loc}}(\mathcal{S})$ under the condition LUBV.

Proof. Let $\Pi_0 h$ be the orthogonal projection of h on the space of functions generated by the characteristic functions of the cubes $C_{0,p}$.

$$\Pi_0 h(x) = \sum_{C_{0,p} \subset [0,1]^d} h_p \mathbb{1}_{C_{0,p}}, \tag{3.3}$$

where $h_p = 2^{Nd}\int_{C_{0,p}} h(x)\,dx$. Let $\sigma^* h$ be the element of $\mathcal{E}_{Nd}$ defined by

$$\sigma^* h(2^{-Nd}\sigma(p)) = h_p. \tag{3.4}$$

Let $D_1(h) = |(X - Z)(\sigma^* h)|$, $D_2(h) = \sqrt{n}|X^{(n)}(h - \Pi_0 h)|$, and $D_3(h) = \sqrt{n}|B_n(h - \Pi_0 h)|$. Clearly

$$\sqrt{n}|X^{(n)}(h) - B_n(h)| \leqq D_1(h) + D_2(h) + D_3(h)\,. \tag{3.5}$$

First, we control $D_1(h)$. By Theorem 2.1, the exponential bounds on $D_1(h)$ depend mainly on $\|\sigma^* h\|_{\mathcal{B}}$. Now, we claim that for any function h of bounded variation on the unit cube $[0,1]^d$,

$$\|\sigma^* h\|_{\mathcal{B}}^2 \leqq 16.2^{N(d-1)} K(h, I^d)\,. \tag{3.6}$$

Proof of (3.6) By definition of σ^*,

$$\gamma_{j,k}(\sigma^* h) = 2^{Nd-j}\left(\sum_{\sigma(p)\in I_{j-1,2k}} h_p - \sum_{\sigma(p)\in I_{j-1,2k+1}} h_p\right).$$

For any $p \in R_j$, let $\bar{C}_{j,p} = \bigcup_{p' \in C'_{j,p}} \bar{C}_{0,p}$. Assume that $j = rd - l$ for some positive r and some l in $]0, d]$. By Lemma 3.1, there exists some p in R_j such that

$$\gamma_{j,k}(\sigma^* h) = 2^{Nd-j} \int_{C_{j-1,p}} (h(x) - h(x + (0, \ldots, 0, 2^{j_l - N}, 0, \ldots, 0)))\,dx. \tag{3.7}$$

Since h takes its values in $[-1, 1]$, the coefficients $|\gamma_{j,k}(\sigma^* h)|$ are each bounded by 1. Hence, taking into account the above equality, we get

$$\sum_{0 \leqq k < 2^{Nd-j}} \gamma_{j,k}^2(\sigma^* h) \leqq \theta(h, j)\,, \tag{3.8}$$

where

$$\theta(h,j) = 2^{Nd-j} \int_{I^{l-1}\times[0,1-2^{j_l-N}]\times I^{d-l}} |h(x) - h(x+(0,\ldots,0,2^{j_l-N},0,\ldots,0))|\,dx\,.$$

Concluding the proof needs regularization arguments. Let ϕ be a positive function be longing to the set C_0^∞ of C^∞ real-valued functions with a compact support contained in $]-1,1[^d$ and satisfying $\int\phi(x)dx = 1$. Let ϕ_ε be defined from ϕ by $\phi_\varepsilon = \varepsilon^{-d}\phi(\varepsilon^{-1}x)$ and let $h_\varepsilon = h*\phi_\varepsilon$. Let $I_\varepsilon = [\varepsilon, 1-\varepsilon]$. It is straightforward to prove that, for any $\varepsilon > 0$,

$$K(h, I^d) \geqq \int_{I_\varepsilon^d} \|\nabla h_\varepsilon(x)\|\,dx\,. \tag{3.9}$$

Since h takes its values in $[-1,1]$, it is obvious that

$$\theta(h,j) = \lim_{\varepsilon\to 0} 2^{Nd-j} \int_{I_\varepsilon^{l-1}\times[\varepsilon,1-\varepsilon-2^{j_l-N}]\times I_\varepsilon^{d-l}} |h_\varepsilon(x) - h_\varepsilon(x+(0,\ldots,0,2^{j_l-N},0,\ldots,0))|\,dx.$$

Therefore, we have:

$$\theta(h,j) \leqq 2^{Nd-j+j_l-N} \limsup_{\varepsilon\to 0} \int_{I_\varepsilon^d} |\nabla_l h_\varepsilon(x)|\,dx\,.$$

Then, using (3.9) and summing on j the above inequality, we get (3.6). □

In order to apply Theorem 2.1, it remains to give an upper bound on $M(p, \sigma^* h)$. Under the assumptions of Theorem 1.1, we set:

$$p_i = \frac{1}{2}\left(\frac{1}{Nd} + \frac{1}{i(i+1)}\right).$$

With the above choice of $(p_1, \ldots, p_{Nd})$, $\sum_{i\leqq Nd} q_i < 16$. Recall $\sigma^* h$ takes its values in $[-1,1]$ and $\mathscr{B}$ is an orthogonal system of $\ell^2(\mathbb{Z}_+)$. Hence

$$\sum_{l<j} \sum_{\{m: I_{l,m}\subset I_{j,k}\}} 2^{l-j}\gamma_{l,m}^2(\sigma^* h) \leqq 1.$$

Since $1/p_i < 2Nd$, it follows that

$$M(p,f) \leqq 8Nd \text{ and } \sum_{i=1}^{Nd} \frac{q_i}{2} < 8\,. \tag{3.10}$$

On the other hand, under the condition LUBV, using the same arguments as in the proof of (3.6), we obtain:

$$2^{l-j} \sum_{\{m: I_{l,m}\subset I_{j,k}\}} \gamma_{l,m}^2(\sigma^* h) \leqq 4.2^{(l-j)/d} K_{\mathrm{loc}}(\mathscr{S})$$

for any $l < j$ and any h in $\mathscr{S}$. Now, let $p_i = \frac{1}{i(i+1)}$. Summing on l the above inequality, we have:

$$M(p,f) \leqq 16d^3 K_{\text{loc}}(\mathscr{S}) \text{ and } \sum_{i=1}^{Nd} \frac{q_i}{2} < 4 . \tag{3.11}$$

From (3.6), (3.10), (3.11) and Theorem 2.1, it follows that, for any positive x and any h in $\mathscr{S}$,

$$\mathbb{P}(D_1(h) \geqq 24.2^{N(d-1)/2}\sqrt{K(\mathscr{S})x} + C(d,N)x) \leqq 2\exp(-x) \tag{3.12}$$

where $C(d,N) = 1 + \sqrt{8dN}$ in the general case, and $C(d,N) = (1 + 4d^{3/2})K_{\text{loc}}(\mathscr{S})$ under the condition LUBV.

Now, we make the control of $D_2(h)$ and of $D_3(h)$. By definition of $X^{(n)}$,

$$D_2(h) = \left| \sum_{i=1}^{n} h(x_i) - \Pi_0 h(x_i) \right| .$$

So, $D_2(h)$ is the absolute value of a sum of iid zero-mean random variables each bounded by 1. To apply Bernstein's inequality, we need to control the variance of each random variable $h(x_i) - \Pi_0 h(x_i)$. Since $h - \Pi_0 h$ takes its values in $[-1, 1]$,

$$\mathbb{E}((h(x_i) - \Pi_0 h(x_i))^2) \leqq \overline{\lim_{\varepsilon \to 0}} V_\varepsilon$$

where

$$V_\varepsilon = \sum_{p \leqq 2^N(1,\ldots,1)} \int_{C_{0,p} \cap I_\varepsilon^d} |h_\varepsilon(x) - h_{\varepsilon,p}|\, dx .$$

From Remark 1.3 of Miranda's paper (1964) it follows that

$$V_\varepsilon \leqq r(\varepsilon) + 2^{-N} \sum_{l=1}^{d} \int_{I_\varepsilon^d} |\nabla_l h_\varepsilon(x)|\, dx$$

for some $r(\varepsilon)$ converging to 0 as $\varepsilon \to 0$. Therefore, by (3.9), we get:

$$\operatorname{Var}(h(x_i) - \Pi_0 h(x_i)) \leqq d2^{-N} K(h, I^d) . \tag{3.13}$$

Hence, by (3.13) and by Bernstein's inequality, for any positive x, for each h in $\mathscr{S}$,

$$\mathbb{P}(D_2(h) \geqq n^{(d-1)/(2d)}\sqrt{2d2^d K(\mathscr{S})x} + 2x) \leqq 2\exp(-x) . \tag{3.14}$$

Some calculation and (3.13) also show that, for any positive x,

$$\mathbb{P}(D_3(h) \geqq n^{(d-1)/(2d)}\sqrt{2d2^d K(\mathscr{S})x}) \leqq 2\exp(-x) . \tag{3.15}$$

Both inequalities (3.12), (3.14) and (3.15) then yield Lemma 3.2. □

Uniform control. Using Lemma 3.2 and Massart's oscillation control, we will prove Theorem 1.1. Let $\varepsilon = 1/n$. Since $\mathscr{S}$ is a Vapnik–Chervonenkis class of functions, there exists an ε-net $\mathscr{S}_\varepsilon$ with cardinality no more than $C(\mathscr{S})n^{d(\mathscr{S})}$. Let $\mathscr{U}_\varepsilon$ be the class of functions defined by: $\mathscr{U}_\varepsilon = \{h - g\colon (h, g) \in \mathscr{S} \times \mathscr{S}, d_\lambda(h, g) \leqq \varepsilon\}$. Let

$$D_1 = \sup_{h \in \mathscr{S}_\varepsilon} \sqrt{n}|X^{(n)}(h) - B_n(h)|, D_2 = \sup_{g \in \mathscr{U}_\varepsilon} \sqrt{n}|X^{(n)}(g)|, D_3 = \sup_{g \in \mathscr{U}_\varepsilon} \sqrt{n}|B_n(g)|.$$

Clearly

$$\sup_{h \in \mathscr{S}} \sqrt{n}|X^{(n)}(h) - B_n(h)| \leqq D_1 + D_2 + D_3.$$

We now control D_2 and D_3. $\mathscr{U}_\varepsilon$ is a Vapnik–Chervonenkis class of functions fulfilling condition $(\mathscr{M})$. Moreover, $N(\delta, \mathscr{U}_\varepsilon) \leqq (N(\delta/2, \mathscr{S}))^2$, for any positive δ. Hence, by Proposition 3.7 of Massart's paper (1986), with $U = 1, \sigma^2 = 1/n$, we get:

$$\mathbb{P}(D_2 \geqq \sqrt{2(1 + \log_2 n)x} + 2x) = O(n^{40d(\mathscr{S})})\exp(-x) \tag{3.16}$$

where $\log_2 n = \log\log n$ (since the functions of $\mathscr{S}$ take their values in $[-1, 1]$, the dimension $d^{(2)}(\mathscr{S})$ of universal (2)-entropy used by Massart satisfies $d^{(2)}(\mathscr{S}) \leqq 2d(\mathscr{S})$). From Theorem 4.1 of Massart's paper (1986), it follows that inequality (3.16) still holds for the random variable D_3 related to the Brownian bridge B_n. Hence, combining the above inequalities and Lemma 3.2 with the upper bound on Card $\mathscr{S}_\varepsilon$, we get:

$$\mathbb{P}\left(\sup_{h \in \mathscr{S}} \sqrt{n}|X^{(n)}(h) - B_n(h)| \geqq C_6 n^{(d-1)/(2d)}\sqrt{K(\mathscr{S})x} + 4\sqrt{x\log_2 n} + (4 + C(N))x\right) = O(n^{40d(\mathscr{S})})\exp(-x). \tag{3.17}$$

A straightforward application of (3.17) then yields Theorem 1.1. □

4 Deviation fields of kernel density estimators

In this section, we give an application of the results of Sect. 3 to the investigation of maximal deviation of kernel density estimators. We prove an invariance principle for the deviation field associated with f_n. So, let Ψ be a Lipschitzian kernel function from $\mathbb{R}^d$ into $[-1, 1]$ with compact support, satisfying $\int \Psi(y)dy = 1$. For any positive real h, let Ψ_h be the mapping defined by $\Psi_h(y) = \Psi(h^{-1}y)$. Now, for any reals a and b fulfilling $0 < a < b \leqq 1$, let $\mathscr{K}_{a,b}$ be the class of functions

$$\mathscr{K}_{a,b} = \{\Psi_h(. - x)\colon x \in \mathbb{R}^d, h \in [a, b]\}.$$

Theorem 1.2 follows from the invariance principle below and from Theorem 3.2 of Konakov and Piterbarg (1984).

Theorem 4.1 *Let* $d \geqq 2$ *and let* $x_1, x_2, \ldots$ *be a sequence of independent random variables with common law* μ *fulfilling* 1.3 *and* Ψ *be a Lipschitzian kernel function from* $\mathbb{R}^d$ *into* $[-1, 1]$ *with compact support and satisfying* $\int \Psi(y)dy = 1$. *Let* $(h_n)_{n>0}$ *be a sequence of bandwith parameters converging to* 0, *satisfying* $\lim_{n\to+\infty} nh_n^d/\log n = +\infty$. *Let* ξ_n *be the deviation field defined by* (1.4). *Then, there* *exists a sequence* $(W_n)_{n>0}$ *of standard Wiener processes indexed by* $\mathscr{K}_n = \mathscr{K}_{h_n/2, 2h_n}$ *with almost surely continuous trajectories on* $(\mathscr{K}_n, d_\lambda)$ *such that*:

$$\sup_{x \in \mathbb{R}^d} |\sigma_\Psi \xi_n(x) - h_n^{-d/2} W_n(\psi_{h_n}(x - .))|$$

$$= O(h_n^\beta \sqrt{\log n} + (nh_n^d)^{-1/(2d)} \sqrt{\log n} + (nh_n^d)^{-1/2} \log n) \text{ a.s.}$$

Proof. The proof uses Theorem 1.1 and the properties of the classes $\mathscr{K}_{h/4, h}$: as Pollard (1982) does, we prove that $\mathscr{K}_{h/4, h}$ is a VC class.

Lemma 4.1 *Let* Ψ *be a kernel function fulfilling the above assumptions. Then, for any* $h \in [0, 1]$, $\mathscr{K}_{h/4, h}$ *is a Vapnik–Chervonenkis class of functions satisfying the condition* LUBV, *for some constant* K_{loc} *depending only on the kernel function* Ψ, *and the mild measurability condition* $(\mathscr{M})$. *Moreover, there exists some constant C depending only on* Ψ *such that, for any positive* ε *and any* $h \in [0, 1]$,

$$N(\varepsilon, \mathscr{K}_{h/4, h}) \leqq C\varepsilon^{-2-d} .$$

Proof of Lemma 4.1 Clearly, $\mathscr{K}_{0,1} = \bigcup_{a>0} \mathscr{K}_{a,1}$ satisfies the mild measurability condition $(\mathscr{M})$. Now, using the Lipschitz condition on Ψ, we prove that $\mathscr{K}_{0,1}$ satisfies the condition LUBV. Let $\mathscr{C}$ be any cube of $\mathbb{R}^d$ with edges of length ε. Since Ψ is a Lipschitzian function, Ψ has, almost everywhere, partial derivatives uniformly bounded on $\mathbb{R}^d$. Hence,

$$K(\Psi_h(. - x), \mathscr{C}) \leqq \int_{\mathscr{C}} \|\nabla \Psi_h(y - x)\| \, dy \leqq C_\Psi h^{-1} \inf(\varepsilon^d, h^d) . \tag{4.1}$$

It follows that $\mathscr{K}_{0,1}$ satisfies condition LUBV for some constant K_{loc} depending only on Ψ. It remains to control the universal entropy.

Let p be the greatest integer such that $ph \leqq 1$. Divide each coordinate segment into p intervals of length $1/p$. This division generates the partition of I^d on cubes $(\mathscr{C}_k)_{k \leqq (p, \ldots, p)}$, of volume equal to $p^{-d} \sim h^d$. Let P denote any probability on I^d and let ε be any positive real less than 1. Let $\mathscr{A} = \{k \leqq (p, \ldots, p)$ such that $P(\mathscr{C}_k) > \varepsilon\}$. Clearly, $\text{Card}\,\mathscr{A} \leqq \varepsilon^{-1}$. Now, define a pseudometric d_ψ on $[h/4, h] \times \mathbb{R}^d$ by:

$$d_\Psi((h_0, x_0), (h_1, x_1)) = \| \Psi_{h_0}(. - x_0) - \Psi_{h_1}(. - x_1) \|_\infty . \tag{4.2}$$

Since Ψ is a function with compact support, the ε-capacity of the set $\mathscr{X}$ of reals x such that the support of $\Psi_{h'}(. - x)$ intersects the cube $\mathscr{C}_k$ for some h' with $h/4 \leqq h' \leqq h$ and for some $k \in \mathscr{A}$, with respect to the usual metric on $\mathbb{R}^d$ is bounded by $C_2 \varepsilon^{-1} h^d$. When x does not belong to $\mathscr{X}$, $\int |\Psi_{h'}(. - x)| dP \leqq C_3 \varepsilon$. Moreover, the Lipschitz norm of the elements of $\mathscr{K}_{h/4,h}$ is uniformly bounded by M/h for some positive M. Using the above bound on the ε-capacity of $\mathscr{X}$, we obtain:

$$N(C_3 \varepsilon, [h/4, h] \times \mathscr{X}, d_\psi) \leqq C_4 \varepsilon^{-2-d}. \tag{4.3}$$

Since P is a probability law, by (4.2) and (4.3), $N(C_3 \varepsilon, \mathscr{K}_{h/4, h}, P) \leqq C_4 \varepsilon^{-2-d}$, therefore completing the proof of Lemma 4.1. □

Proof of Theorem 4.1 Since μ has a continuous and strictly positive density on the closed unit cube I^d, it follows from Corollary 1.1 that

$$\sup_{g \in \mathscr{S}_n} |X^{(n)}(g) - B_n(g)| = O(n^{-1/(2d)} \sqrt{h_n^{d-1} \log n} + \log n) \text{ a.s.} \tag{4.4}$$

for some sequence $(B_n)_{n>0}$ of Brownian bridges with a.s. continuous trajectories on $(\mathscr{K}_n, d_\mu)$. Since Ω is rich enough, there exists a sequence $(W_n)_{n>0}$ of Gaussian processes on Ω with covariance function $(g, h) \to \int g(x)h(x)f(x)dx$ such that

$$B_n(g) = W_n(g) - W_n(I^d)\int g(x)f(x)\,dx \tag{4.5}$$

for any positive n. Now, for all g in $\mathscr{S}_n$, $\int g(x)f(x)dx = O(h_n^d)$ as $n \to +\infty$. Hence, by (4.4) and (4.5),

$$\sup_{g \in \mathscr{K}_n} |X^{(n)}(g) - W_n(g)| = O(h_n^d\sqrt{\log n} + n^{-1/(2d)}\sqrt{h_n^{d-1}\log n} + \log n)\text{ a.s.} \tag{4.6}$$

For any class $\mathscr{S}$ of functions, let $\sqrt{f}\mathscr{S}$ be the class defined by $\sqrt{f}\mathscr{S} = \{\sqrt{f}g : g \in \mathscr{S}\}$. By the above equality, there exists a sequence $(G_n)_{n>0}$ of homogeneous Wiener fields on I^d with a.s. continuous trajectories on $(\sqrt{f}\mathscr{K}_n \cup \mathscr{K}_n, d_\lambda)$ such that $G_n(\sqrt{f}g) = W_n(g)$ for any $g \in \mathscr{K}_n$. So, recalling that $f^{-1/2}$ is uniformly bounded on I^d, it is easy to see that Theorem 4.1 follows from (4.6) and lemma below:

Lemma 4.2 *Under the assumptions of Theorem* 4.1,

$$\sup_{\substack{x \in R^d \\ h_n/2 \leq h \leq 2h_n}} |G_n(\sqrt{f(\cdot)}\,\Psi_h(. - x)) - \sqrt{f(x)}\,G_n(\Psi_h(. - x))| = O(h_n^{\beta + d/2}\sqrt{\log n})\text{ a.s.}$$

Proof. Let $(\eta(g))_{g \in \mathscr{K}_n}$ be the Gaussian process defined from G_n by

$$\eta(\Psi_h(. - x)) = G_n(\sqrt{f(.)}\,\Psi_h(. - x)) - \sqrt{f(x)}\,G_n(\Psi_h(. - x))\,. \tag{4.7}$$

Let ρ be the usual metric on $\mathscr{K}_n$ related to η $(\rho(g_0, g_1) = |E((\eta(g_0) - \eta(g_1))^2)|^{1/2})$. Since $\mathscr{K}_n$ is a VC class, for any positive ε,

$$N(\varepsilon, \mathscr{K}_n, \rho) \leqq C_6\varepsilon^{-D} \tag{4.8}$$

for some positive D. So, in order to control the maximal deviation of η it will be sufficient to control the maximal standard error $\sigma(\eta) = \sup_{g \in \mathscr{K}_n}\rho(g, g)$. Now, using the Hölder condition on f, it is straightforward to see that

$$\sup_{y \in \mathbb{R}^d} |\sqrt{f(y)}\,\Psi_h(y - x) - \sqrt{f(x)}\,\Psi_h(y - x)| \leqq C_7 h^\beta$$

for some constant C_7. Hence, integrating on x, we get: $\sigma^2(\eta) \leqq C_8 h_n^{2\beta + d}$ for some positive constant C_8. Both the above inequality and (4.8) together with Lemma 2.4 of Konakov and Piterbarg (1984) imply Lemma 4.2, therefore completing the proof of Theorem 4.2. □

Appendix: Proof of the lemmas of Sect. 2

A. Proof of Lemma 2.1

Let $\mathscr{A}$ denote the class of random variables with zero-mean and finite moment-generating function over $]t_0, +\infty[$ for some negative t_0, such that $\gamma_A(t) = \log E(\exp(tA))$ satisfies: there exists an $\varepsilon > 0$ such that $\gamma_A''(t) \geqq \varepsilon$ for all $t > 0$.

First, we prove that Lemma 2.1 holds for any (A, B) in $\mathscr{A} \times \mathscr{A}$. Let $\gamma_\beta(t) = \beta\gamma_A(t/\beta) + (1-\beta)\gamma_B(t/(1-\beta))$. By Hölder's inequality, for each β and each positive, t, $\gamma_{A+B}(t) \leqq \gamma_B(t)$.

Now, let $h_\beta(\varepsilon) = \sup_{t>0}(t\varepsilon - \gamma_\beta(t))$ denote the Legendre transform of γ_β. From the above inequality, it follows that $h_{A+B}^{-1}(u) \leqq h_\beta^{-1}(u)$ for any positive u. Hence, it is sufficient to prove that, for any positive u, there exists some β in $]0, 1[$ such that

$$h_\beta^{-1}(u) \leqq h_A^{-1}(u) + h_B^{-1}(u) . \tag{A.1}$$

Now, let $x = \gamma'_A(t/\beta)$ and let $y = \gamma'_B(t/(1-\beta))$. The usual Cramer–Chernoff calculation yields:

$$h_\beta(x+y) = \beta h_A(x) + (1-\beta)h_B(y) . \tag{A.2}$$

Clearly, it is sufficient to prove that, for any positive u, there exists (t, β) such that $x = \gamma'_A(t/\beta)$ and $y = \gamma'_B(t/(1-\beta))$ fulfill the equations

$$h_A(x) = u \text{ and } h_B(y) = u. \tag{A.3}$$

Because (A.3) and (A.2) imply $h_\beta^{-1}(u) \leqq h_A^{-1}(u) + h_B^{-1}(u)$. Now, (A.3) holds iff there exists $t > 0$ and $\beta \in]0, 1[$ such that:

$$u = h_A \circ \gamma'_A(t/\beta) = h_B \circ \gamma'_B(t/(1-\beta)) .$$

Since A belongs to $\mathscr{A}$, γ_A is an analytic convex function on $\mathbb{R}^+$. Recalling that $\gamma''_A(t) \geqq \varepsilon$ for some positive ε, it is easy to see that $g_A = h_A \circ \gamma'_A$ is a continuous increasing function from $\mathbb{R}^+$ onto $\mathbb{R}^+$ with $g'_A(x) \geqq x\varepsilon$. Hence, for any positive t, the function $\beta \to g(t, \beta) = g_B(t/(1-\beta)) - g_A(t/\beta)$ is a one to one continuous increasing function from $]0, 1[$ onto $\mathbb{R}$. Moreover, $g(t, .)$ fulfills the assumptions of the implicit function theorem. Hence, there exists a unique continuous function $t \to \beta(t)$ such that $g(t, \beta(t)) = 0$.

Let $u(t) = h_A \circ \gamma'(t/\beta(t))$. Clearly $t \to u(t)$ is a continuous function such that $u(t)$ satisfies the Eqs. (A.3). Clearly,

$$\min(g_A(2t), g_B(2t)) \leqq u(t) \leqq \max(g_A(2t), g_B(2t)) .$$

Hence, $\lim_0 u(t) = 0$, $\lim_{+\infty} u(t) = +\infty$ and (A.3) follows. Hence, Lemma 2.1 holds for any (A, B) in $\mathscr{A} \times \mathscr{A}$. We prove now Lemma 2.1 in the general case.

The general case. Let I_A, I_B and I_{A+B} denote the respective domains of the moment generating functions of A, B, and $A + B$. Let Y be a real random variable with standard normal law, independent of (A, B). For any positive ε, let

$$A_\varepsilon = A\mathbb{1}_{\varepsilon A<1} - \mathbb{E}(A\mathbb{1}_{\varepsilon A<1}) + \varepsilon Y$$

and let B_ε denote the corresponding random variable associated with B. Clearly, A_ε and B_ε belong to $\mathscr{A}$. Now, by the Beppo-Levi lemma, for any nonnegative t,

$$\lim_{\varepsilon\to 0^+} \mathbb{E}(\exp(tA\mathbb{1}_{\varepsilon A<1})) = \mathbb{E}(\exp(tA))$$

and

$$\lim_{\varepsilon\to 0^+} \mathbb{E}(\exp(tA\mathbb{1}_{\varepsilon A<1} + tB\mathbb{1}_{\varepsilon B<1})) = \mathbb{E}(\exp(tA + tB)) .$$

It follows that the functions γ_{A_ε} and $\gamma_{A_\varepsilon + B_\varepsilon}$ converge pointwise to the respective functions γ_A and γ_{A+B} on $\mathbb{R}^+$, with the convention $\gamma_A(t) = +\infty$ if

$\mathbb{E}(\exp(tA)) = +\infty$. Then, using some arguments of convexity, it is straightforward to prove that, for any compact subset $\mathscr{K}$ of $\mathring{I}_A \cap \mathbb{R}^+$, γ_{A_ε} converges uniformly to γ_A over $\mathscr{K}$ as $\varepsilon \to 0^+$. Let D_A and D_{A+B} be the respective domains of h_A and h_{A+B}. From the above results, it follows that, for any positive $x \notin \partial D_A$, $\lim_{\varepsilon \to 0^+} h_{A_\varepsilon}(x) = h_A(x)$, with the convention $h_A(x) = +\infty$ if $x \notin D_A$, and using exactly the same arguments, for any positive $x \notin \partial D_{A+B}$, $\lim_{\varepsilon \to 0^+} h_{A_\varepsilon + B_\varepsilon}(x) = h_{A+B}(x)$. Since the above functions are convex, it follows that, for any positive u, $h_{A_\varepsilon}^{-1}(u)$ converges to $h_A^{-1}(u)$, $h_{B_\varepsilon}^{-1}(u)$ converges to $h_B^{-1}(u)$, and $h_{A_\varepsilon + B_\varepsilon}^{-1}(u)$ converges to $h_{A+B}^{-1}(u)$. Hence, recalling that $(A_\varepsilon, B_\varepsilon)$ satisfies the prescribed inequality, we obtain Lemma 2.1. □

B. *Proof of Lemma 2.2*

We may w.l.o.g. assume that f takes its values in $[0, 1]$. Define the nonincreasing family of fields $(\mathscr{F}_j)_{0 < j \leqq N}$ by $\mathscr{F}_N = \{\emptyset, \Omega\}$. For any $j \leqq N$, let

$$\mathscr{F}_{j-1} = \mathscr{F}_j \vee \sigma\{\tilde{\xi}_{j,k} : k \in [0, 2^{N-j}[\} .$$

Clearly, the random variables $(U_{j,k})_{0 \leqq k < 2^{N-j}}$ are $\mathscr{F}_j$-measurable. So, the random variables $(\tilde{U}_{j,k} - (U_{j,k})^{1/2}\tilde{\xi}_{j,k})_{0 \leqq k < 2^{N-j}}$ are independent and symmetric, conditional on $\mathscr{F}_j$. For convenience, we set $\gamma_{j,k} = \gamma_{j,k}(f)$. Define:

$$D_{j,1} = \sum_{0 \leqq k < 2^{N-j}} \gamma_{j,k}(\tilde{U}_{j,k} - (U_{j,k})^{1/2}\tilde{\xi}_{j,k}) .$$

By definition, $\Delta_1 = \sum_{j=1}^N D_{j,1}$. Now, from the above remark and from (ii) of Tusnady's lemma, it follows that, for any positive integer j,

$$\mathbb{E}(\exp(tD_{j,1}) | \mathscr{F}_j) \leqq \mathbb{E}\left(\prod_{0 \leqq k < 2^{N-j}} \cosh(t\gamma_{j,k}(2 + \tilde{\xi}_{j,k}^2/4)) | \mathscr{F}_j \right) .$$

Hence, recalling that the random variables $\tilde{\xi}_{j,k}$ are independent and $N(0, 1)$-distributed, we get, by induction on j:

$$\mathbb{E}(\exp(t\Delta_1)) \leqq \prod_{j=1}^N \prod_{0 \leqq k < 2^{N-j}} \mathbb{E}(\cosh(t\gamma_{j,k}(2 + \tilde{\xi}_{j,k}^2/4))) . \tag{B.1}$$

Moreover, the random variables $\tilde{\xi}_{j,k}$ have distribution function Φ. Let us define the convex function ϱ by $\varrho(x) = +\infty$ if $|x| \geqq 1$ and

$$\varrho(x) = \frac{1}{2}\left(\frac{e^{4x}}{\sqrt{1-x}} + \frac{e^{-4x}}{\sqrt{1+x}} \right)$$

otherwise. A few calculation proves that, for all random variable ξ with distribution function Φ, $\mathbb{E}(\cosh((2 + \xi^2/4)x)) = \varrho(x/2)$. It follows that

$$\log \mathbb{E}(\exp(t\Delta_1)) \leqq \sum_{j=1}^N \sum_{0 \leqq k < 2^{N-j}} \log \varrho(\gamma_{j,k} t/2) . \tag{B.2}$$

Now, we claim that, for any $x \in]-1, 1[$,

$$\log \varrho(x) \leqq -\frac{83}{8} \log(1 - x^2) . \tag{B.3}$$

Proof of (B.3) By definition of ϱ,

$$\sqrt{1-x^2}\,\varrho(x) = \cosh(4x)((1+\tanh(4x))\sqrt{1+x} + (1-\tanh(4x))\sqrt{1-x})/2 .$$

Define $l(x) = (\sqrt{1-x} + \sqrt{1+x})/2$. First we claim that, for any $-1 < x < 1$,

$$l(x) \leqq 1 - x^2/8 \text{ and } \tanh(4x)(\sqrt{1+x} - \sqrt{1-x}) \leqq 4x^2 . \tag{B.4}$$

The proof of the left-hand inequality will be omitted. We prove only the right-hand inequality. Clearly, it is sufficient to prove that, for any $0 < x < 1$, $\tanh(4x) \leqq 4xl(x)$, where $l(.)$ is the already defined mapping. Since $y \to \tanh y$ is a continuous decreasing mapping, (B.4) follows from $4x \leqq \operatorname{argtanh}(4xl(x))$, with the convention $\operatorname{argtanh} y = +\infty$ if $y \geqq 1$. Now, it is obvious that, for any $0 < y < 1$, $\operatorname{argtanh} y \geqq y(1 + y^2/3)$. Moreover, $(l(x))^2 \geqq 1/2$. Therefore, (B.4) follows from

$$1 \leqq l(x)(1 + 8x^2/3) . \tag{B.5}$$

Let $\alpha = 1 + (8x^2/3)$. Since $l(x) \geqq 1/\sqrt{2}$, (B.5) holds if $\alpha \geqq \sqrt{2}$. When $\alpha < \sqrt{2}$, (B.5) holds if $2 \leqq \alpha^2(1 + \sqrt{1-x^2})$, e.g. if $4(1-\alpha^2) \leqq -\alpha^4 x^2$. Since $x^2 = 3(\alpha - 1)/8$, the above inequality is equivalent to

$$4(1+\alpha) \geqq 3\alpha^4/8 .$$

When $1 < \alpha < \sqrt{2}$, the term on right hand in the is less than 3/2, which concludes the proof of (B.4).

Then, noting that $\cosh y \leqq \exp(y^2/2)$, we infer that $\sqrt{1-x^2}\,\varrho(x) \leqq \exp(79x^2/8)$. Now, for any real x, $x^2 \leqq -\log(1-x^2)$, and (B.3) follows.

Now, recall that the mapping f takes its values on $[0, 1]$. Hence, for all j, for all p, $|\gamma_{j,p}| \leqq 1/2$. Furthermore, $x \to -\log(1-x^2)$ is a convex function of x^2. We infer that, for all real t such that $|t| < 1$,

$$\log \mathbb{E}(\exp(4t\Delta_1)) \leqq -\frac{83}{8}\|f\|_{\mathscr{B}}^2 \log(1-t^2) ,$$

and Lemma 2.2 follows clearly from this inequality □

C. Proof of Lemma 2.3

Using (i) of Tusnády's lemma, we will prove that:

$$|\sqrt{U_l} - \sqrt{\mathbb{E}(U_l)}| \leqq \sqrt{2} + \sum_{m : l \prec m} 2^{-|m-l|/2} |\tilde{\xi}_m| . \tag{C.1}$$

Proof of (C.1) For all m in $\mathscr{L}$, let $m+1$ denote the element of $\bar{\mathscr{L}}$ such that $m \prec m+1$ and $|m+1| = |m| + 1$. Let $s(l)$ be the element of $\mathscr{L}$ defined by:

$$s(l) = \inf\{m \in \mathscr{L} \text{ such that } l \preceq m \text{ and } U_{m+1} \neq 0\} .$$

Clearly,

$$\sqrt{U_l} - \sqrt{\mathbb{E}(U_l)} = \sum_{s(l) \preceq m \prec (N, 0)} 2^{-(1+|m-l|)/2}(\sqrt{2U_m} - \sqrt{U_{m+1}}) . \tag{C.2}$$

Now, recall that $U_{m+1} - 2U_m$ is, up to the sign, equal to $\tilde{U}_{m+1}$. Moreover, there is no loss of generality in proving (C.1) to assume that $\tilde{U}_{m+1} = U_{m+1} - 2U_m$, which we shall do throughout the proof of (C.1). By (i) of Tusnády's lemma, $|\tilde{U}_{m+1}| \leqq |\tilde{\xi}_{m+1}|\sqrt{U_{m+1}} + 2$. Thus, we have:

$$|\sqrt{2U_m} - \sqrt{U_{m+1}}| \leqq |\tilde{\xi}_{m+1}| + (\sqrt{2U_m} + \sqrt{U_{m+1}})^{-1} \inf(2, |\tilde{U}_{m+1}|) , \quad \text{(C.3)}$$

from which it follows that, for any l in $\mathscr{L}$,

$$|\sqrt{U_l} - \sqrt{\mathbb{E}(U_l)}| \leqq \sup(H_l, K_l) + \sum_{m:l \prec m} 2^{-|m-l|/2} |\tilde{\xi}_m| , \quad \text{(C.4)}$$

where the random variables H_l and K_l are defined by:

$$H_l = \sum_{m:s(l) \preceq m \prec (N,0)} 2^{-(1+|m-l|)/2} \mathbb{1}_{(\tilde{U}_{m+1} > 0)} \inf(2, \tilde{U}_{m+1})(\sqrt{2U_m} + \sqrt{U_{m+1}})^{-1}$$

and

$$K_l = \sum_{m:s(l) \preceq m \prec (N,0)} 2^{-(1+|m-l|)/2} \mathbb{1}_{(\tilde{U}_{m+1} < 0)} \inf(2, -\tilde{U}_{m+1})(\sqrt{2U_m} + \sqrt{U_{m+1}})^{-1} .$$

First, we determine an upper bound for K_l. For convenience, we set $u = U_{m+1}$ and $x = -\tilde{U}_{m+1}$. Then, with the above notations,

$$\inf(2, -\tilde{U}_{m+1})(\sqrt{2U_m} + \sqrt{U_{m+1}})^{-1} = \inf(2, x)/(\sqrt{u} + \sqrt{u+x}) .$$

Now, recall that $(U_m)_{m \geqq l}$ is a nondecreasing sequence of integers. Hence u and x are positive integers with $0 < x \leqq u$, and we easily get: $\inf(2, x)/(\sqrt{u} + \sqrt{u+x}) \leqq 2 - \sqrt{2}$. Hence, we have:

$$K_l \leqq (2 - \sqrt{2}) \sum_{i>0} 2^{-i/2} \leqq \sqrt{2} . \quad \text{(C.5)}$$

It remains to prove that $H_l \leqq \sqrt{2}$. Let $\mathscr{H}$ denote the set of nondecreasing sequences of natural numbers. For all $a = (a_i)_{i \geqq 0}$ in $\mathscr{H}$, let

$$H(a) = \sum_{i \geqq 0} \mathbb{1}_{a_{i+1} > 2a_i} 2^{-(1+i)/2} (\sqrt{a_{i+1}} + \sqrt{2a_i})^{-1} \inf(2, a_{i+1} - 2a_i)$$

and set $H_{\max} = \sup_{a \in \mathscr{H}} H(a)$. Clearly, $H_l \leqq H_{\max}$. Clearly, the function $H(.)$ takes its supremum on the subset $\mathscr{H}_0$ of $\mathscr{H}$ of sequences $a = (a_i)_{i \geqq 0}$ such that, for all natural i, either $a_{i+1} = 2a_i + 1$ or $a_{i+1} = 2a_i + 2$. Then

$$H(a) = \sum_{i \geqq 0} \sqrt{2^{-1-i} a_{i+1}} - \sqrt{2^{-i} a_i} .$$

An elementary calculation proves that, for all a in $\mathscr{H}_0$, $H(a) \leqq \sqrt{2 + a_0} - \sqrt{a} \leqq \sqrt{2}$, therefore completing the proof of (C.1). Now, let us finish off the proof of Lemma 2.3. For all natural $j < N$, define $\Delta_{2,j}$ and $\Delta_{3,j}$ by:

$$\Delta_{2,j} = \sum_{l:|l| \leqq j} \gamma_l(f)(\sqrt{U_l} - \sqrt{\mathbb{E}(U_l)})\tilde{\xi}_l$$

and

$$\Delta_{3,j} = \sum_{l:|l| \leqq j} \gamma_l(f)\tilde{\xi}_l \left(\sqrt{2} + \sum_{m:l \prec m} 2^{-|m-l|/2} |\xi_m| \right) .$$

Let $R_{2,j}(t) = \mathbb{E}(\exp(t\Delta_{2,j})|\mathscr{F}_j)$ and $R_{3,j}(t) = \mathbb{E}(\exp(t\Delta_{3,j})|\mathscr{F}_j)$. It remains to prove that, for all $j < N$, for all real t,

$$R_{2,j}(t) \leqq R_{3,j}(t) . \tag{C.6}$$

Let $\mathscr{G}_j$ be the σ-field generated by the family of random variables $\{|\tilde{\xi}_l|: l \in \bar{\mathscr{L}}, |l| > j\}$. For all l in $\bar{\mathscr{L}}$, let ε_l be the sign of $\tilde{\xi}_l$. Now, we prove, by induction on j, that, for all $j < N$, for all real t, $R_{3,j}(t)$ is a $\mathscr{G}_j$-measurable function and $R_{2,j}(t) \leqq R_{3,j}(t)$. Clearly, the above assumption holds true if $j = 1$. Now, assume that the assumption holds true for any $i < j$. Let $D_{2,j} = \Delta_{2,j} - \Delta_{2,j-1}$ and $D_{3,j} = \Delta_{3,j} - \Delta_{3,j-1}$. Clearly, $D_{2,j}$ is an $\mathscr{F}_{j-1}$-measurable random variable. Hence, we have:

$$\mathbb{E}(\exp(t\Delta_{2,j})|\mathscr{F}_{j-1}) \leqq R_{3,j-1}(t)\exp(tD_{2,j}) .$$

By definition of $D_{2,j}$,

$$D_{2,j} = \sum_{l:|l|=j} \gamma_l(f)\varepsilon_l|\tilde{\xi}_l|(\sqrt{U_l} - \sqrt{\mathbb{E}(U_l)}) .$$

Then, noting that the random variables $\{|\tilde{\xi}_l|(\sqrt{U_l} - \sqrt{\mathbb{E}(U_l)}): l \in \mathscr{L}, |l| = j\}$ are $(\mathscr{F}_j \vee \mathscr{G}_{j-1})$-measurable and that the random variables $(\varepsilon_l)_{l \in \mathscr{L}:|l|=j}$ are independent and symmetric, given $(\mathscr{F}_j \vee \mathscr{G}_{j-1})$, and recalling that $R_{3,j-1}(t)$ is $\mathscr{G}_{j-1}$-measurable, we get:

$$\mathbb{E}(\exp(t\Delta_{2,j})|\mathscr{F}_j \vee \mathscr{G}_{j-1}) \leqq R_{3,j-1}(t) \prod_{l:|l|=j} \cosh(\gamma_l(f)\tilde{\xi}_l(\sqrt{U_l} - \sqrt{\mathbb{E}(U_l)})) .$$

Hence, using (C.1) we obtain:

$$\begin{aligned}\mathbb{E}(\exp(t\Delta_{2,j})|\mathscr{F}_j \vee \mathscr{G}_{j-1}) &\leqq R_{3,j-1}(t)\mathbb{E}(\exp(tD_{3,j})|\mathscr{G}_{j-1}) \\ &\leqq \mathbb{E}(\exp(t\Delta_{3,j})|\mathscr{F}_j \vee \mathscr{G}_{j-1}).\end{aligned}$$

Then, taking the expectation conditionally on $\mathscr{F}_j$ in the above inequality, we get (C.6) and Lemma 2.3. □

D. *Proof of Lemma 2.4*

Clearly, in order to prove Lemma 2.4, we may w.l.o.g. assume that $t = 1$. Now, recall the definition of Δ_5:

$$\Delta_5 = \Delta_3 - \Delta_4 = \sum_{l \in \mathscr{L}} \gamma_l(f)\tilde{\xi}_l \sum_{m:l \prec m} 2^{-|m-l|/2}|\tilde{\xi}_m| .$$

By Cauchy–Schwarz inequality,

$$\sum_{m:l \prec m} 2^{-|m-l|/2}|\tilde{\xi}_m| \leqq \eta_l, \text{ where } \eta_l = \left(\sum_{m:l \prec m} q_{|m-l|}\tilde{\xi}_m^2\right)^{1/2} .$$

We infer that, for any real t, $\mathbb{E}(\exp(\Delta_5)|\mathscr{G}_0) \leqq \mathbb{E}(\exp(\Delta_{6,0})|\mathscr{G}_0)$, where the random variables $\Delta_{6,j}$ are defined by $\Delta_{6,j} = \sum_{l:|l|>j}\gamma_l(f)\tilde{\xi}_l\eta_l$ for any nonnegative integer j. Hence, we have:

$$\mathbb{E}(\exp(\Delta_5)) \leqq \mathbb{E}(\exp(\Delta_{6,0})) .$$

It remains to control the moment-generating function of $\Delta_{6,0}$. Clearly,

$$\mathbb{E}(\exp(\Delta_{6,0})|\mathscr{F}_1) = \mathbb{E}\left(\exp\left(\frac{1}{2}\sum_{l:|l|=1} \gamma_l^2(f)\eta_l^2\right)\exp(t\Delta_{6,1})\right).$$

Notations. Let $\{Y_{l,m}:(l,m)\in\bar{\mathscr{L}}\times\bar{\mathscr{L}}, l\leqq m\}$ be a family of independent $N(0,1)$-distributed random variables, independent of the family $\{\tilde{\xi}_l: l\in\bar{\mathscr{L}}\}$ and define the random variables $\{\zeta_{l,m}:l\prec m\}$ by $\zeta_{l,m}=\gamma_l(f)q_{|m-l|}^{1/2}Y_{l,m}$. For any m in $\bar{\mathscr{L}}$ with $|m|\geqq 2$, we set: $\zeta_m=\sum_{l\prec m}\zeta_{l,m}$. When $|m|=1$, we set $\zeta_m=0$. We also set $\log\mathbb{E}(\exp K)=L(K)$ for any real-valued random variable K.

With the above notations, we get:

$$L(\Delta_{6,0}) = L\left(\Delta_{6,1} + \sum_{|m|>1}\sum_{\substack{|l|=1\\ l\prec m}} \zeta_{l,m}\tilde{\xi}_m\right).$$

Define the function φ by $\varphi(M)=(2/(1+\sqrt{1-M}))^{1/2}$. Lemma 2.4 follows from the inequality below: for any j in $[1,N]$,

$$L(\Delta_{6,0}) \leqq L\left(\varphi(M)\left(\sum_{|l|\leqq j}\zeta_l Y_{l,l} + \sum_{|m|>j}\sum_{\substack{|l|\leqq j\\ l\prec m}}\zeta_{l,m}\tilde{\xi}_m\right) + \Delta_{6,j}\right). \tag{D.1}$$

Proof of (D.1) Clearly, the random variables $\{\zeta_{l,m}:|l|=1, |m|>1\}$ are independent of $\mathscr{F}_1$. Since $\varphi(M)\geqq 1$, we infer that (D.1) holds if $j=1$. Now, by definition of ζ_m, for any $j\geqq 1$,

$$\sum_{|m|>j}\sum_{\substack{|l|\leqq j\\ l\prec m}}\zeta_{l,m}\tilde{\xi}_m = \sum_{|m|>j+1}\sum_{\substack{|l|\leqq j\\ l\prec m}}\zeta_{l,m}\tilde{\xi}_m + \sum_{|l|=j+1}\zeta_l\tilde{\xi}_l.$$

Hence,

$$\varphi(M)\sum_{|m|>j}\sum_{\substack{|l|\leqq j\\ l\prec m}}\zeta_{l,m}\tilde{\xi}_m + \Delta_{6,j} = \sum_{|l|=j+1}(\varphi(M)\zeta_l + \gamma_l(f)\eta_l)\tilde{\xi}_l$$

$$+ \varphi(M)\sum_{|m|>j+1}\sum_{\substack{|l|\leqq j\\ l\prec m}}\zeta_{l,m}\tilde{\xi}_m + \Delta_{6,j+1}.$$

Since the r.v's $\{\tilde{\xi}_l, \zeta_l:|l|=j+1\}$ are independent with respective Gaussian distributions $N(0,1)$ and $N(0,M_l/4)$ conditionally to $\mathscr{F}_{j+1}\vee\sigma\{\zeta_{l,m}:l\prec m, |l|\leqq j, |m|>j+1\}$, integrating the variables $\{\tilde{\xi}_l,\zeta_l:|l|=j+1\}$, and using the above equality, we get:

$$\begin{aligned} &L\left(\varphi(M)\left(\sum_{|l|\leqq j}\zeta_l Y_{l,l} + \sum_{|m|>j}\sum_{\substack{|l|\leqq j\\ l\prec m}}\zeta_{l,m}\tilde{\xi}_m\right) + \Delta_{6,j}\right) \\ &= L\left(\varphi(M)\left(\sum_{|l|\leqq j+1}\zeta_l Y_{l,l} + \sum_{|m|>j+1}\sum_{\substack{|l|\leqq j\\ l\prec m}}\zeta_{l,m}\tilde{\xi}_m\right)\right. \\ &\quad \left. + \sum_{|m|>j+1}\sum_{\substack{|l|=j+1\\ l\prec m}}\frac{\zeta_{l,m}\tilde{\xi}_m}{\sqrt{1-M_l\varphi^2(M)/4}} + \Delta_{6,j+1}\right). \end{aligned}$$

Then, noting that $\varphi(M) = (1 - M\varphi^2(M)/4)^{-1/2}$ and recalling that, for any l in $\bar{\mathscr{L}}$, $M_l \leqq M$, we easily infer from the above inequality that:

$$L\left(\varphi(M)\left(\sum_{|l| \leqq j} \zeta_l Y_{l,l} + \sum_{|m| > j} \sum_{\substack{|l| \leqq j \\ l \prec m}} \zeta_{l,m} \tilde{\xi}_m\right) + \Delta_{6,j}\right)$$
$$\leqq L\left(\varphi(M)\left(\sum_{|l| \leqq j+1} \zeta_l Y_{l,l} + \sum_{|m| > j+1} \sum_{\substack{|l| \leqq j+1 \\ l \prec m}} \zeta_{l,m} \tilde{\xi}_m\right) + \Delta_{6,j+1}\right).$$

Hence (D.1) holds for any $j \geqq 1$. Now, applying (D.1) with $j = N$, we obtain:

$$L(\Delta_5(g)) \leqq L\left(\varphi(M) \sum_{l \in \bar{\mathscr{L}}} \zeta_l Y_{l,l}\right)$$

for any g in $\mathscr{E}_N$. Now, recall that the random variables $\{\zeta_l, Y_{l,l} : l \in \bar{\mathscr{L}}\}$ are independent with respective Gaussian laws $N(0, M_l/4)$ and $N(0, 1)$. Hence, using the above inequality and noting that $(1 - M_l\varphi^2(M)/4)^{-1/2} \leqq (\varphi(M))^{M_l/M}$, we get:

$$L(\Delta_5(g)) \leqq -\frac{1}{2M} \sum_{l \in \bar{\mathscr{L}}} M_l \log((1 + \sqrt{1 - M})/2)$$

for all M in $]0, 1]$, and Lemma 2.4 easily follows. □

Remerciements. Je voudrais remercier J. Bretagnolle pour ses conseils techniques sur la preuve du théorème 2.1, ainsi que P. Massart et P. Doukhan pour m'avoir suggéré de travailler sur cette question.

References

Beck, J.: Lower bounds on the approximation of the multivariate empirical process. Z. Wahrscheinlichkeitstheor. Verw. Geb. **70**, 289–306 (1985)

Bretagnolle, J., Massart, P.: Hungarian constructions from the nonasymptotic viewpoint. Ann. Probab. **17**, 239–256 (1989)

Csörgó, M., Revész, P.: Strong approximations in probabilities and statistics. New York: Academic Press 1981

Dudley, R.M.: Sample functions of the Gaussian process. Ann. Probab. **1**, 66–103 (1973)

Dudley, R.M.: Metric entropy of some classes of sets with differentiable boundaries. J. Approximation Theory **10**, 227–236 (1974)

Dudley, R.M.: Central limit theorems for empirical measures. Ann. Probab. **6**, 899–929 (1978)

Dudley, R.M.: A course on empirical processes Henneguin, L. (ed.). In: Ecole d'été de probabilités de Saint-Flour XII-1982. (Lect. Notes Math., vol. 1097, pp. 1–142) Berlin Heidelberg New York: Springer 1984

Kolchinsky, V.I.: On the central limit theorem for empirical measures (in Russian). Teor. Veroyatn. Mat. Stat. Kiev **24**, 63–75 (1981)

Kolchinsky, V.I.: Accuracy of the approximation of an empirical process by a Brownian bridge (in Russian). Sib. Mat. J. **32–4**, 48–60 (1991)

Kolchinsky, V.I.: Komlós-Major-Tusnády approximations for the general empirical process and Haar expansions of classes of functions. (Preprint 1992)

Komlós, J., Major, P., Tusnády, G.: An approximation of partial sums of independent rv's and the sample df. I. Z. Wahrscheinlichkeitstheor. Verw. Geb. **32**, 111–131 (1975)

Konakov, V.D., Piterbarg, V.I.: On the convergence rate of maximal deviations distributions for kernel regression estimates. J. Multivariate Anal. **15**, 279–294 (1984)

Louani, D.: Limit laws of Erdös-Rényi-type indexed by sets. Stat. Decis. **10**, 291–305 (1992)

Mason, D.M., Van Zwet, W.R.: A refinement of the KMT inequality for the uniform empirical process. Ann. Probab. **15**, 871–884 (1987)
Massart, P.: Rates of convergence in the central limit theorem for empirical processes. Ann. Inst. Henri Poincaré, Probab. Stat. **22**, 381–423 (1986)
Massart, P.: Strong approximation for multivariate empirical and related processes, via K.M.T. constructions. Ann. Probab. **17**, 266–291 (1989)
Miranda, M.: Distribuzioni aventi derivate misure insemi di perimetro localmente finito. Ann. Sc. Norm. Super. Pisa **18**, 27–56 (1964)
Pollard, D.: Rates of strong uniform convergence. (Preprint 1982)
Rio, E.: Strong approximation for set-indexed partial sum Processes, via K.M.T. constructions. I. Ann. Probab. **21**, 759–790 (1993)
Rio, E.: Local invariance principles and its application to density estimation. Prépubl. Math. Univ. Paris-Sud 91-71 (1991)
Rosenblatt, M.: Remarks on a multivariate transformation. Ann. Math. Statist. **23**, 470–472 (1952)
Schwartz, L.: Théorie des distributions. Paris: Hermann 1957
Skorohod, A.V.: On a representation of random variables. Theory Probab. Appl. **21**, 628–632 (1976)