Stochastic Processes and their Applications 33 (1989) 163-173
North-Holland

A NOTE ON LIMIT THEOREMS FOR PERTURBED EMPIRICAL PROCESSES

J.E. YUKICH

Department of Mathematics, Lehigh University, Bethlehem, PA 18015, USA

Received 5 May 1988
Revised 16 December 1988

Let X_i, $i\geq 1$, be a sequence of i.i.d. $\mathbb{R}^k$-valued random variables with common distribution P. Let H_n, $n\geq 1$, be a sequence of distribution functions (d.f.) such that $H_n \xrightarrow{w} H_0$, where H_0 is the d.f. of the unit mass at zero. The perturbed empirical d.f. is defined by $\tilde{F}_n(x) := n^{-1}\sum_{i\leq n} H_n(x-X_i)$; $\tilde{P}_n$ denotes the associated perturbed empirical probability measure. Strong laws of large numbers and weak invariance principles are obtained for the perturbed empirical processes $(\tilde{P}_n - P)(f)$, $f\in\mathcal{F}$, where $\mathcal{F}$ denotes a class of functions on $\mathbb{R}^k$. The results extend and generalize those of Winter and Yamato and have applications to non-parametric density estimation.

AMS (1980) Subject Classifications: Primary: 60B12, 60F15; Secondary: 60F17.

laws of large numbers * invariance principles * perturbed empirical distribution functions * metric entropy

1. Introduction

Let X_i, $i\geq 1$, be a sequence of i.i.d. $\mathbb{R}^k$-valued random variables with common law P. P is assumed to be a probability measure on $\mathcal{B}$, the usual Borel σ-algebra. The nth empirical measure for P is

$$P_n := n^{-1}(\delta_{X_1} + \cdots + \delta_{X_n}),$$

where δ_x denotes the unit mass at x. Let $\mathcal{F}$ be a class of real-valued measurable functions on $\mathbb{R}^k$. The empirical process indexed by $\mathcal{F}$ is denoted by

$$(P_n - P)(f) := \int f(\mathrm{d}P_n - \mathrm{d}P), \quad f\in\mathcal{F}.$$

Let H_n, $n\geq 1$, be a sequence of distribution functions; assume that $H_n \xrightarrow{w} H_0$ (i.e., $\int f\,\mathrm{d}H_n \to f(0)$ whenever $f:\mathbb{R}^k\to\mathbb{R}$ is continuous and bounded), where H_0 is the distribution for δ_0. As in Winter (1973) and Yamato (1973), define for all $x\in\mathbb{R}^k$ the perturbed empirical distribution function

$$\tilde{F}_n(x) := n^{-1}\sum_{i\leq n} H_n(x-X_i). \tag{1.1}$$

The estimator $\tilde{F}_n$ differs from the usual empirical distribution function

$$F_n(x) := n^{-1} \sum_{i \leq n} H_0(x - X_i)$$

in that the mass n^{-1} is no longer concentrated at X_i, but is distributed around X_i according to H_n. The assumption $H_n \xrightarrow{w} H_0$ helps insure that the asymptotic behavior of $\tilde{F}_n - F$ will be close to that of $F_n - F$, where F is the distribution function of X.

Winter (1973) and Yamato (1973) have shown that if P is continuous then $\tilde{F}_n$ has the "Glivenko-Cantelli property", that is,

$$\|\tilde{F}_n - F\| \to 0 \quad \text{a.s.}, \tag{1.2}$$

where here and elsewhere $\|f\|$ denotes the essential supremum of the function f.

This note considers limit theorems for the *general* perturbed empirical process

$$(\tilde{P}_n - P)(f) := \int f(\mathrm{d}\tilde{P}_n - \mathrm{d}P), \quad f \in \mathscr{F},$$

where $\tilde{P}_n$ denotes the perturbed probability measure associated with $\tilde{F}_n$ and $\mathscr{F}$ a general class of measurable functions not necessarily of the form $\{1_{(-\infty,x]}: x \in \mathbb{R}\}$ as required in (1.2). Sufficient conditions on $\mathscr{F}$, P and $\tilde{P}_n$ are found such that the following strong law of large numbers (SLLN) holds:

$$|(\tilde{P}_n - P)(f)|_{\mathscr{F}} = |\textstyle\int f(\mathrm{d}\tilde{P}_n - \mathrm{d}P)|_{\mathscr{F}} \to 0 \quad \text{a.s.} \quad \text{as } n \to \infty,$$

where $|\cdot|_{\mathscr{F}}$ denotes supremum norm over $\mathscr{F}$. Here and elsewhere the a.s. convergence is with respect to Pr*, the usual outer measure associated to $\Pr := P^{\infty}$, the infinite product measure on the infinite product of $\mathbb{R}^k$ with itself. (See Dudley, 1984, for details.) The results, which are linked to non-parametric density estimation, extend and generalize those of Winter (1973) and Yamato (1973) and have a variety of applications. In addition to the SLLN we also obtain an invariance principle for the perturbed empirical process.

The perturbed empirical process is studied by placing it within the standard framework of empirical processes (see Dudley, 1984; Gaenssler, 1983; and Pollard, 1984, for extensive treatments of the subject; see also Giné and Zinn, 1984, 1987, for more recent developments). Limit theorems for perturbed empirical processes essentially amount to limit theorems for empirical processes indexed by a sequence of function classes varying with n, the sample size.

The following definition will be useful (Dudley, 1984); $\mathscr{C}(\mathbb{R}^k)$ denotes the continuous functions on $\mathbb{R}^k$.

Definition 1.1. Given $f, h \in \mathscr{C}(\mathbb{R}^k)$ let $[f, h] := \{g: \mathbb{R}^k \to \mathbb{R}$ such that $f \leq g \leq h$ pointwise$\}$. Given a class of functions $\mathscr{F}$ and a probability measure P on $\mathbb{R}^k$, let $N^{[\ ]}(\varepsilon, \mathscr{F}, P) := \min\{m: \exists f_1, \ldots, f_m$ in $\mathscr{L}^1(\mathbb{R}^k, \mathscr{B}, P)$ such that $\mathscr{F} \subseteq \bigcup_{i=1}^m \{[f_i, f_j]: \int (f_j - f_i)\, \mathrm{d}P \leq \varepsilon\}$. $\log N^{[\ ]}(\varepsilon, \mathscr{F}, P)$ is called *metric entropy with bracketing*; the usual definition, however, does not require continuity of the bracket functions (Dudley, 1984).

2. Preliminary results

Starting with (1.1) it is natural to define the associated perturbed empirical measures by $\tilde{P}_n := \mu_n * P_n$, where μ_n is a sequence of probability measures on $\mathbb{R}^k$. We assume once and for all that $\mu_n \xrightarrow{w} \delta_0$ and that the μ_n are independent of the P_n. It is easier and more convenient to work with the measures $\tilde{P}_n$ instead of the distribution functions $\tilde{F}_n$; working with $\tilde{P}_n$ permits application of the standard theory of empirical processes and also yields results of greater generality.

If $\mathscr{F}$ denotes a class of real-valued functions on $\mathbb{R}^k$ then write $\mathscr{F} \in \mathrm{SLLN}(P, P_n)$ (resp. $\mathscr{F} \in \mathrm{SLLN}(P, \tilde{P}_n)$) iff

$$\left|\int f(\mathrm{d}P_n - \mathrm{d}P)\right|_{\mathscr{F}} \to 0 \quad \text{a.s.} \qquad \left(\text{resp.} \left|\int f(\mathrm{d}\tilde{P}_n - \mathrm{d}P)\right|_{\mathscr{F}} \to 0 \quad \text{a.s.}\right).$$

Throughout, $f_x(y)$ and $f(x+y)$ are used interchangeably and $\tilde{\mathscr{F}}$ denotes the class of translates of all elements of $\mathscr{F}$, i.e. $\tilde{\mathscr{F}} := \{f_x : x \in \mathbb{R}^k, f \in \mathscr{F}\}$.

Notice that for all $f \in \mathscr{F}$,

$$\begin{aligned}\int f(\mathrm{d}\tilde{P}_n - \mathrm{d}P) &= \int f \,\mathrm{d}(\mu_n * P_n) - \int f \,\mathrm{d}P \\ &= \int \int f(x+y)\,\mathrm{d}P_n(y)\,\mathrm{d}\mu_n(x) - \int \int f(y)\,\mathrm{d}P(y)\,\mathrm{d}\mu_n(x).\end{aligned}$$

The triangle inequality yields

$$\left|\int f(\mathrm{d}\tilde{P} - \mathrm{d}P)\right|_{\mathscr{F}} \leq |V(n)|_{\mathscr{F}} + |B(n)|_{\mathscr{F}},$$

where

$$V(n) := \int \int f(x+y)\,\mathrm{d}P_n(y)\,\mathrm{d}\mu_n(x) - \int \int f(x+y)\,\mathrm{d}P(y)\,\mathrm{d}\mu_n(x)$$

and

$$B(n) := \int \int f(x+y)\,\mathrm{d}P(y)\,\mathrm{d}\mu(x) - \int \int f(y)\,\mathrm{d}P(y)\,\mathrm{d}\mu_n(x).$$

$V(n)$ and $B(n)$ are the stochastic and non-stochastic (bias) components, respectively.

If $\tilde{\mathscr{F}} \in \mathrm{SLLN}(P, P_n)$ then $|V(n)|_{\mathscr{F}} \to 0$ a.s.; if, in addition

$$|B(n)|_{\mathscr{F}} = \left|\int \int f(x+y) - f(y)\,\mathrm{d}P(y)\,\mathrm{d}\mu_n(x)\right|_{\mathscr{F}} \to 0, \tag{2.1}$$

then $\mathscr{F} \in \mathrm{SLLN}(P, \tilde{P}_n)$. In fact, since

$$\left|\int f(\mathrm{d}\tilde{P}_n - \mathrm{d}P)\right|_{\mathscr{F}} = |V(n) + B(n)|_{\mathscr{F}} \geq \big||V(n)|_{\mathscr{F}} - |B(n)|_{\mathscr{F}}\big|,$$

it follows that if $\tilde{\mathscr{F}} \in \text{SLLN}(P, P_n)$ and $\mathscr{F} \in \text{SLLN}(P, \tilde{P}_n)$ then $|B(n)|_{\mathscr{F}} \to 0$, i.e. (2.1) is necessary as well as sufficient for $\mathscr{F} \in \text{SLLN}(P, \tilde{P}_n)$.

Proposition 2.1. *Let $\mathscr{F}$, $\tilde{\mathscr{F}}$ and $\tilde{P}_n$ be as above; suppose $\tilde{\mathscr{F}} \in \text{SLLN}(P, P_n)$. Then*

$$\left|\int f(\mathrm{d}\tilde{P}_n - \mathrm{d}P)\right|_{\mathscr{F}} \to 0 \quad a.s. \qquad \Leftrightarrow \qquad \left|\int\int (f_x - f)\,\mathrm{d}P\,\mathrm{d}\mu_n(x)\right|_{\mathscr{F}} \to 0.$$

This simple proposition is especially useful when $\mathscr{F}$ is closed under translations. For example, we easily deduce the Winter-Yamato version of the Glivenko-Cantelli theorem for perturbed empirical measures.

Corollary 2.2 (Winter, 1973; Yamato, 1973). *Let $\mathscr{F} := \{1_{(-\infty,x]}: x \in \mathbb{R}\}$. Then*

$$\left|\int f(\mathrm{d}\tilde{P}_n - \mathrm{d}P)\right|_{\mathscr{F}} = \|(\tilde{F}_n - F)(x)\| \to 0 \quad a.s. \tag{2.2}$$

for any continuous probability measure P on $\mathbb{R}$.

This results implies that if F has a density and $\hat{f}_n$ is a "typical" non-parametric density estimator and $\hat{F}_n(x) := \int_{-\infty}^{x} \hat{f}_n(t)\,\mathrm{d}t$, then $\hat{F}_n$ converges uniformly to F a.s. Actually, we prove later that (2.2) holds in higher dimensions. For the moment we sketch the proof of (2.2).

Proof. Since $\tilde{\mathscr{F}} \in \text{SLLN}(P, P_n)$ it suffices to show convergence of the bias term (2.1). By continuity of P, $\forall \varepsilon > 0$ $\exists \delta > 0$ such that $P\{[a, b]\} < \frac{1}{2}\varepsilon$ whenever $b - a < \delta$. Find $n_0 := n_0(\delta)$ such that $\forall n \geq n_0$, $\mu_n\{x: |x| > \delta\} < \frac{1}{2}\varepsilon$. Then $\forall n \geq n_0$,

$$\begin{aligned} |B(n)|_{\mathscr{F}} &= \sup_y \int P\{(y, x+y]\}\,\mathrm{d}\mu_n(x) \\ &\leq \int_{-\delta}^{\delta} \sup_y P\{(y, x+y]\}\,\mathrm{d}\mu_n(x) + \tfrac{1}{2}\varepsilon \leq \varepsilon. \qquad \square \end{aligned}$$

We easily deduce two additional corollaries. Let $\text{BL}(1) := \{f: \mathbb{R}^k \to \mathbb{R}$ such that $|f(x) - f(y)| \leq \|x - y\|$ $\forall x, y \in \mathbb{R}^k$ and $\|f\| \leq 1\}$. BL(1) is the set of bounded Lipschitz functions on $\mathbb{R}^k$.

Corollary 2.3. $\text{BL}(1) \in \text{SLLN}(P, \tilde{P}_n)$.

Proof. Since $N^{[\ ]}(\varepsilon, \text{BL}(1), P) < \infty$ $\forall \varepsilon > 0$ it follows that $\text{BL}(1) \in \text{SLLN}(P, P_n)$ (see e.g. Dudley, 1984, Theorem 6.1.5). Since BL(1) is translation invariant it suffices to show $|B(n)|_{\mathscr{F}} \to 0$. We have

$$|B(n)|_{\mathscr{F}} = \left| \int_{\|x\| \le \varepsilon} \left[\int (f_x - f) \, \mathrm{d}P(y) \right] \mathrm{d}\mu_n(x) + \int_{\|x\| > \varepsilon} [\cdots] \, \mathrm{d}\mu_n(x) \right|_{\mathscr{F}}$$

$$\le \int_{\|x\| \le \varepsilon} \|x\| \, \mathrm{d}\mu_n(x) + 2\mu_n\{x \colon \|x\| > \varepsilon\}$$

$$\le \varepsilon + 2\mu_n\{x \colon \|x\| > \varepsilon\} \le 2\varepsilon,$$

for n large enough. □

Corollary 2.4. *Let $f : \mathbb{R}^k \to \mathbb{R}$ be uniformly continuous, $\|f\| < \infty$ and $\mathscr{F} := \{f_x \colon x \in \mathbb{R}\}$. Then $\mathscr{F} \in \mathrm{SLLN}(P, \tilde{P}_n)$.*

Proof. Since $\mathscr{F} \in \mathrm{SLLN}(P, P_n)$ it suffices to show convergence of the bias term (2.1). Let $0 < \varepsilon < 1$ be arbitrary. By the uniform continuity of f, $\exists \delta > 0$ such that $\forall x, y$ with $\|x - y\| < \delta$ we have $|f(x) - f(y)| < \frac{1}{2}\varepsilon$. Also, $\exists n_0 := n_0(\delta)$ such that for all $n \ge n_0$, $\mu_n\{x \colon \|x\| \le \delta\} \ge 1 - \varepsilon/(4\|f\|)$, where $\|f\|$ is assumed positive, else there is nothing to prove. Therefore

$$|B(n)|_{\mathscr{F}} \le \sup_t \int_{-\delta}^{\delta} \int |f_t(x+y) - f_t(y)| \, \mathrm{d}P(y) \, \mathrm{d}\mu_n(x) + \tfrac{1}{2}\varepsilon.$$

The uniform continuity of f implies that the integrand is at most $\frac{1}{2}\varepsilon$, showing that $|B(n)|_{\mathscr{F}} \le \varepsilon$. □

3. Metric entropy and the SLLN for perturbed empirical processes

In this section we provide sufficient conditions guaranteeing $\mathscr{F} \in \mathrm{SLLN}(P, \tilde{P}_n)$; it is assumed throughout that the elements of $\mathscr{F}$ are bounded by one in the supremum norm.

Recall that Blum (1955), DeHardt (1971) and Dudley (1984) have shown that if $N^{[\ ]}(\varepsilon, \mathscr{F}, P) < \infty \; \forall \varepsilon > 0$ then $\mathscr{F} \in \mathrm{SLLN}(P, P_n)$. (Actually, this was proved without any continuity assumption on the bracket functions.) The following theorem shows that the condition $N^{[\ ]}(\varepsilon, \mathscr{F}, P) < \infty \; \forall \varepsilon > 0$, together with the implicit hypothesis $\mu_n \xrightarrow{\mathrm{w}} \delta_0$, which cannot be removed in general, implies $\mathscr{F} \in \mathrm{SLLN}(P, \tilde{P}_n)$. This result should be considered as the analog of the Blum–DeHardt–Dudley theorem for the perturbed case.

Theorem 3.1. *Let $N^{[\ ]}(\varepsilon, \mathscr{F}, P) < \infty \; \forall \varepsilon > 0$. Then*

$$\left| \int f(\mathrm{d}\tilde{P}_n - \mathrm{d}P) \right|_{\mathscr{F}} \to 0 \quad a.s. \tag{3.1}$$

Proof. Let $\varepsilon > 0$. Clearly, $P_n * \mu_n \xrightarrow{w} P * \delta_0$ a.s. since $\mu_n \xrightarrow{w} \delta_0$ and $P_n \xrightarrow{w} P$ a.s. Thus for each bounded and continuous f on $\mathbb{R}^k$,

$$\left| \int f(\mathrm{d}\tilde{P}_n - \mathrm{d}P) \right| \to 0 \quad \text{a.s.}$$

Without loss of generality we may assume that each bracket function is bounded by one in the sup norm. Thus, if $m := N^{[\ ]}(\varepsilon, \mathscr{F}, P)$ then $\exists n_0 := n_0(\varepsilon)$ such that for all $n \geq n_0$,

$$\max_{i \leq m} \left| \int f_i(\mathrm{d}\tilde{P}_n - \mathrm{d}P) \right| < \varepsilon, \tag{3.2}$$

except on a set with probability less than ε.

Next, for each $f \in \mathscr{F}$ find a bracket $[f_i, f_j]$ such that $f_i \leq f \leq f_j$. Then by (3.2) and the definition of a bracket we have on a set with probability greater than $1 - \varepsilon$ (Dudley, 1984),

$$\begin{aligned} |(\tilde{P}_n - P)(f)| &\leq |(\tilde{P}_n - P)(f - f_i)| + |(\tilde{P}_n - P)(f_i)| \\ &\leq |(\tilde{P}_n + P)(f_j - f_i)| + \varepsilon \leq |(\tilde{P}_n - P)(f_j - f_i)| + 3\varepsilon \leq 5\varepsilon. \end{aligned}$$

This holds for all $f \in \mathscr{F}$ and thus (3.1) holds as desired. □

Applications

Theorem 3.1 applies to a variety of classes of functions. For example, if $\mathscr{F} := \{x \to \mathrm{e}^{itx}: |t| \leq 1\}$, then for any P, $N^{[\ ]}(\varepsilon, \mathscr{F}, P) < \infty\ \forall \varepsilon > 0$ (smooth the brackets of Yukich, 1985), implying a SLLN for the perturbed empirical characteristic function:

$$\limsup_{n \to \infty} \sup_{|t| \leq 1} \left| \int \mathrm{e}^{itx}(\mathrm{d}\tilde{P}_n - \mathrm{d}P) \right| = 0 \quad \text{a.s.}$$

Next, let $\mathscr{C} := \{(-\infty, x]: -\infty, x \in \mathbb{R}^k\}$, $\mathscr{F} := \{1_C: C \in \mathscr{C}\}$ and P a probability measure on $\mathbb{R}^k$ satisfying the following "continuity" condition (δA denotes the boundary of the set A):

$$\forall C \in \mathscr{C}, \quad P(\delta C) = 0. \tag{3.3}$$

Note that (3.3) is satisfied iff for all hyperplanes H perpendicular to the axes $P\{H\} = 0$. Under (3.3) it can be shown that $N^{[\ ]}(\varepsilon, \mathscr{F}, P) < \infty\ \forall \varepsilon > 0$ (use the results of Blum, 1955; DeHardt, 1971; together with standard smoothing operations to generate a finite number of *continuous* brackets). This implies a SLLN for the multidimensional perturbed empirical distribution function. This readily extends the Winter–Yamato results to higher dimensions; their methods are tied to the properties of distribution functions on $\mathbb{R}$ and do not easily admit an extension. We have thus obtained a generalization of Corollary 2.2.

Corollary 3.2. *Let $\tilde{F}_n$ denote the perturbed empirical distribution function $\tilde{F}_n(x) := \tilde{P}_n\{(-\infty, x]\}$, $x \in \mathbb{R}^k$, where P satisfies condition* (3.3). *Then*

$$\|\tilde{F}_n - F\| \to 0 \quad a.s.$$

4. Weak invariance principles for perturbed empirical processes

Extensions of the above methods yield a weak invariance principle (abbreviated WI) for $(\tilde{P}_n - P)(f)$, $f \in \mathscr{F}$. Say that $\mathscr{F} \in \mathrm{WI}(P, P_n)$ iff $\mathscr{F}$ is G_PBUC and there exist processes $Y_j(f)$, $f \in \mathscr{F}$, where Y_j are independent copies of a suitable Gaussian process G_P, such that $\forall \varepsilon > 0$,

$$\lim_{n\to\infty} \Pr{}^*\left\{\max_{m\leq n} \left| n^{-1/2} \sum_{j\leq m} (f(X_j) - P(f) - Y_j(f))\right|_{\mathscr{F}} > \varepsilon\right\} = 0,$$

or, equivalently

$$\lim_{n\to\infty} \Pr{}^*\left\{\max_{m\leq n} \left| n^{-1/2}[m(P_m(f) - P(f)) - \sum_{j\leq m} Y_j(f)]\right|_{\mathscr{F}} > \varepsilon\right\} = 0.$$

This condition is equivalent to $\mathscr{F}$ having the P-Donsker property (see Dudley, 1984; Dudley and Philipp, 1983; Gaenssler, 1983; Giné and Zinn, 1984, 1987; Pollard, 1984; Talagrand, 1988, for notation, details and sufficient conditions implying weak invariance principles). Note that the Y_j a.s. have values in the Banach space B of bounded functions on $\mathscr{F}$ equipped with the supremum norm $|\cdot|_{\mathscr{F}}$. Say that $\mathscr{F} \in \mathrm{WI}(P, \tilde{P}_n)$ iff $\forall \varepsilon > 0$,

$$\lim_{n\to\infty} \Pr{}^*\left\{\max_{m\leq n} \left| \int n^{-1/2} \sum_{j\leq m} (f_x(X_j) - P(f) - Y_j(f))\,\mathrm{d}\mu_n(x)\right|_{\mathscr{F}} > \varepsilon\right\} = 0,$$

or, equivalently,

$$\lim_{n\to\infty} \Pr{}^*\left\{\max_{m\leq n} \left| n^{-1/2}[m(P_m * \mu_n(f) - P(f)) - \sum_{j\leq m} Y_j(f)]\right|_{\mathscr{F}} > \varepsilon\right\} = 0.$$

One could also say that $\mathscr{F} \in \mathrm{WI}(P, \tilde{P}_n)$ if μ_n in the above condition is replaced by μ_m, but this approach leads to technically more complicated arguments in what follows and will not be pursued here. The following is most useful when $\mathscr{F}$ is translation invariant; formally speaking, it may be considered as a weak invariance principle analog of Proposition 2.1.

Proposition 4.1. *Let $\mathscr{F}$, $\tilde{\mathscr{F}}$, P and $\tilde{P}_n$ be as above; $\tilde{P}_n := \mu_n * P_n$. Suppose $\tilde{\mathscr{F}} \in \mathrm{WI}(P, P_n)$ and*

$$\Delta_n := \max_{m\leq n} \left| \int n^{-1/2} \sum_{j\leq m} (Y_j(f) - Y_j(f_x))\,\mathrm{d}\mu_n(x)\right|_{\mathscr{F}} \xrightarrow{\Pr^*} 0.$$

Then $\mathscr{F} \in \mathrm{WI}(P, \tilde{P}_n)$ iff $n^{1/2}|B(n)|_{\mathscr{F}} \to 0$, that is, iff

$$n^{1/2}\left| \int \int (f - f_x)\,\mathrm{d}P\,\mathrm{d}\mu_n(x)\right|_{\mathscr{F}} \to 0 \quad \text{as } n\to\infty.$$

Proof (sufficiency). Denote by D_n the quantity

$$D_n := \max_{m\le n}\left\| \int n^{-1/2}\sum_{j\le m} (f_x(X_j)-P(f)-Y_j(f))\,\mathrm{d}\mu_n(x)\right\|_{\mathscr{F}}.$$

We need to show $D_n \xrightarrow{\Pr^*} 0$. The triangle inequality yields

$$D_n \le \max_{m\le n}\left\| \int n^{-1/2}\sum_{j\le m} (f_x(X_j)-P(f_x)-Y_j(f_x))\,\mathrm{d}\mu_n(x)\right\|_{\mathscr{F}} + \Delta_n$$
$$+\max_{m\le n}\left\| \int n^{-1/2}\sum_{j\le m} P(f_x-f)\,\mathrm{d}\mu_n(x)\right\|_{\mathscr{F}}.$$

The second and third terms in this inequality converge to zero by hypothesis. Applying the Dudley–Philipp invariance principle (Dudley and Philipp, 1983, Theorem 1.3) to the P-Donsker class $\tilde{\mathscr{F}}$, it follows that the first term converges to zero, concluding sufficiency. Necessity follows approximately as in the proof of Proposition 2.1. □

The following theorem provides a weak invariance principle for the perturbed empirical process. Letting $e_P(f,g) := (\int (f-g)^2\,\mathrm{d}P)^{1/2}$, assume from now on that $\mathscr{F}$ satisfies the following regularity condition for P:

$$\forall\varepsilon>0\ \exists 0<\eta<\varepsilon \text{ such that } \forall f\in\mathscr{F}\ \forall\|x\|<\eta,\quad e_P(f,f_x)<\varepsilon. \tag{4.1}$$

Theorem 4.2. *Let $\mathscr{F}$ be P-Donsker; assume that $\mathscr{F}$ is translation invariant and satisfies (4.1) for P. If $n^{1/2}|B(n)|_{\mathscr{F}}\to 0$, that is, if*

$$n^{1/2}\left\|\int\int (f-f_x)\,\mathrm{d}P\,\mathrm{d}\mu_n(x)\right\|_{\mathscr{F}} \to 0 \quad \text{as } n\to\infty,$$

then $\mathscr{F}\in \mathrm{WI}(P,\tilde{P}_n)$.

It is now a simple matter to obtain a weak invariance principle for the perturbed empirical distribution function on $\mathbb{R}^k$, $k\ge 1$ (cf. Puri and Ralescu, 1986, for a pointwise version of a similar result). This result, like Corollary 2.2, is of interest in density estimation and implies a weak invariance principle for the normalized distribution function of a "typical" non-parametric density estimator. We are unaware of similar multidimensional results.

Corollary 4.3. *Let $\mathscr{F} := \{1_{(-\infty,x]}: x\in\mathbb{R}^k\}$ and P a probability measure on $\mathbb{R}^k$ satisfying (4.1). If*

$$n^{1/2}\sup_{t\in\mathbb{R}^k}\int\int 1_{(t,t+x]}(y)\,\mathrm{d}P(y)\,\mathrm{d}\mu_n(x)\to 0 \quad \text{as } n\to\infty,$$

then $\mathscr{F}\in\mathrm{WI}(P,\tilde{P}_n)$. It follows that for all $\varepsilon>0$,

$$\lim_{n\to\infty}\Pr^*\left\{\sup_x\left| n[\tilde{F}_n(x)-F(x)] - n^{-1/2}\sum_{j\le n} Y_j(1_{(-\infty,x]})\right| > \varepsilon\right\} = 0.$$

Proof of Theorem 4.2

By Proposition 4.1 it suffices to show that $\Delta_n \xrightarrow{\Pr^*} 0$; this will follow from well-known properties of Gaussian processes, Lévy's maximal inequality and the regularity condition (4.1).

First, notice that $\max_{m\le n} |n^{-1/2}\sum_{j\le m} Y_j(f)|_{\mathscr{F}}$ is stochastically bounded.

Lemma 4.4. *For all $\beta > 0$ $\exists C < \infty$ such that for all $n \geq 1$,*

$$\Pr^*\left\{\max_{m\le n} n^{-1/2}\left|\sum_{j\le m} Y_j(f)\right|_{\mathscr{F}} \le C\right\} \ge 1-\beta.$$

Proof. This follows from the generalized Lévy maximal inequality for symmetric Banach space valued random variables (Dudley and Philipp, 1983, Lemmas 2.5 and 2.9) Markov's inequality, the Fernique–Slepian lemma (Giné and Zinn, 1984, Theorem 2.17(a)) and $\mathbb{E}|Y_1(f)|_{\mathscr{F}} < \infty$. □

Next, let $\varepsilon > 0$ and $\eta := \eta(\varepsilon)$ be defined by (4.1). For small ε it will be shown that

$$\max_{m\le n} \sup_{\|x\|<\eta} \left| n^{-1/2} \sum_{j\le m} (Y_j(f) - Y_j(f_x))\right|_{\mathscr{F}}$$

is small with high probability. The precise statement actually shows a little more.

Lemma 4.5. *There exists $D := D(\varepsilon)\downarrow 0$ as $\varepsilon \downarrow 0$ such that for all $n \geq 1$,*

$$\lim_{\varepsilon\downarrow 0} \Pr^*\left\{\max_{m\le n} \sup_{f,g:e_P(f,g)<\varepsilon} \left|n^{-1/2}\sum_{j\le m}(Y_j(f)-Y_j(g))\right| > D(\varepsilon)\right\} = 0.$$

Proof. Given $\varepsilon > 0$, let $H(\varepsilon, \mathscr{F}, e_P)$ denote the logarithm of the covering number of $\mathscr{F}$ with respect to the pseudo-metric e_P (Dudley, 1984). Find $D := D(\varepsilon)\downarrow 0$ such that the quantity

$$c(\varepsilon) := \left\{4\mathbb{E}^* \sup_{e_P(f,g)<\varepsilon} |Y_1(f) - Y_1(g)| + 13\varepsilon H^{1/2}(\tfrac{1}{2}\varepsilon, \mathscr{F}, e_P)\right\} \Big/ D(\varepsilon)$$

converges to zero as $\varepsilon \downarrow 0$. (To see that this is possible, use $\mathbb{E}|Y_1(f)|_{\mathscr{F}} < \infty$, $\lim_{\varepsilon\downarrow 0} \sup_{f,g:e_P(f,g)<\varepsilon} |Y_1(f) - Y_1(g)| = 0$ a.s., dominated convergence on the first term and Sudakov's minorization on the second term.)

Note that Markov's inequality and Fernique's comparison theorem (Fernique, 1985, Corollaire 1.6) applied to the Gaussian processes Y_1 and $m^{-1/2}\sum_{j\le m} Y_j$ imply that

$$\max_{m\le n} \Pr^*\left\{\sup_{f,g:e_P(f,g)<\varepsilon} \left|n^{-1/2}\sum_{j\le m}(Y_j(f)-Y_j(g))\right| > D\right\} \le c(\varepsilon).$$

The generalized Lévy maximal inequality (Dudley and Philipp, 1983, Lemmas 2.5 and 2.7) applied to the random variables $X_j(f,g) := Y_j(f) - Y_j(g)$, Markov's inequality and a second application of Fernique's comparison theorem therefore imply that

$$\begin{aligned}
&\Pr^*\left\{\max_{m\le n} \sup_{f,g:e_P(f,g)<\varepsilon} \left| n^{-1/2} \sum_{j\le m} (Y_j(f) - Y_j(g))\right| > D\right\} \\
&\le (1-c(\varepsilon))^{-1} \Pr^*\left\{\sup_{f,g:e_P(f,g)<\varepsilon} \left| n^{-1/2} \sum_{j\le n} (Y_j(f) - Y_j(g))\right| > \tfrac{1}{4}D\right\} \\
&\le (1-c(\varepsilon))^{-1} 4D^{-1}\mathbb{E}^*\left\{\sup_{f,g:e_P(f,g)<\varepsilon} \left| n^{-1/2} \sum_{j\le n} (Y_j(f) - Y_j(g))\right|\right\} \\
&\le (1-c(\varepsilon))^{-1} 4c(\varepsilon).
\end{aligned}$$

Since $c(\varepsilon)$ converges to zero as $\varepsilon \downarrow 0$ the proof of Lemma 4.5 is complete. □

To complete the proof of Theorem 4.2 combine Lemmas 4.4 and 4.5 as follows. Given $\beta > 0$ it suffices to show $\Pr^*\{\Delta_n > 2\beta\} < 2\beta$ for n sufficiently large. Choose $\varepsilon > 0$ small enough so that $D(\varepsilon) \le \beta$ and $(1-c(\varepsilon))^{-1}4c(\varepsilon) < \beta$; let $\eta := \eta(\varepsilon)$ be defined by condition (4.1).

By the assumed convergence $\mu_n \xrightarrow{w} \delta_0$, there is an $n_0 := n_0(\beta)$ such that for all $n \ge n_0$,

$$\mu_n\{x: \|x\| > \eta\} \le \beta/(2C), \tag{4.2}$$

C as in Lemma 4.4. For all $n \ge n_0$, Lemma 4.4 and inequality (4.2) imply that

$$\begin{aligned}
&\max_{m\le n} \left\| \int_{\|x\|>\eta} n^{-1/2} \sum_{j\le m} (Y_j(f) - Y_j(f_x))\, \mathrm{d}\mu_n(x) \right\|_{\mathcal{F}} \\
&\le \int_{\|x\|>\eta} \max_{m\le n} n^{-1/2} \left\| \sum_{j\le m} (Y_j(f) - Y_j(f)) \right\|_{\mathcal{F}} \mathrm{d}\mu_n(x) \\
&\le \int_{\|x\|>\eta} 2C\, \mathrm{d}\mu_n(x) \le \beta,
\end{aligned} \tag{4.3}$$

except perhaps on a set with probability smaller than β.

Additionally, since $D(\varepsilon) \le \beta$, Lemma 4.5 implies that for all $n \ge 1$,

$$\begin{aligned}
&\Pr^*\left\{\max_{m\le n} \left\| \int_{\|x\|\le\eta} n^{-1/2} \sum_{j\le m} (Y_j(f) - Y_j(f_x))\, \mathrm{d}\mu_n(x) \right\|_{\mathcal{F}} > \beta\right\} \\
&\le \Pr^*\left\{\max_{m\le n} \sup_{\|x\|\le\eta} \left\| n^{-1/2} \sum_{j\le m} (Y_j(f) - Y_j(f_x)) \right\|_{\mathcal{F}} > \beta\right\} \\
&\le 4c(\varepsilon)/(1-c(\varepsilon)).
\end{aligned} \tag{4.4}$$

Combining (4.3) and (4.4) gives for $n \ge n_0$, $\Pr^*\{\Delta_n > 2\beta\} \le \beta + 4c(\varepsilon)/(1-c(\varepsilon)) < 2\beta$, as desired. □

Acknowledgement

The author gratefully thanks Professors P. Gaennsler and W.M. Huang for helpful comments and remarks. The author also thanks the referee for his careful and detailed comments and for pointing out some critical gaps in the original version.

References

J.R. Blum, On the convergence of empiric distribution functions, Ann. Math. Statist. 26 (1955) 527-529.

J. DeHardt, Generalizations of the Glivenko-Cantelli theorem, Ann. Math. Statist. 42 (1971) 2050-2055.

R.M. Dudley, A course on empirical processes, in: Lecture Notes in Math., Vol. 1097 (Springer, Berlin, 1984) pp. 1-142.

R.M. Dudley and W. Philipp, Invariance principles for sums of Banach space valued random elements and empirical processes, Z. Wahrsch. Verw. Gebiete 62 (1983) 509-552.

X. Fernique, Sur la convergence étroite des mesures Gaussiennes, Z. Wahrsch. Verw. Gebiete 68 (1985) 331-336.

P. Gaenssler, Empirical processes, in: Inst. Math. Statist. Lecture Notes, Vol. 3 (1983).

E. Giné and J. Zinn, Some limit theorems for empirical processes, Ann. Probab. 12 (1984) 929-989.

E. Giné and J. Zinn, Lectures on the central limit theorem for empirical processes, in: Lecture Notes in Math., Vol. 1221 (Springer, Berlin, 1987) pp. 50-113.

D. Pollard, Convergence of Stochastic Processes (Springer, Berlin, 1984).

M.L. Puri and S.S. Ralescu, Central limit theorem for perturbed empirical distribution functions evaluated at a random point, J. Multivariate Anal. 19 (1986) 273-279.

M. Talagrand, Donsker classes of sets, Probab. Theory Rel. Fields 78 (1988) 169-191.

B.B. Winter, Strong uniform consistency of integrals of density estimators, Canadian J. (1973) 247-253.

B.B. Winter, Convergence rate of perturbed empirical distribution functions, J. Appl. Probab. 16 (1979) 163-173.

H. Yamato, Uniform convergence of an estimator of a distribution function, Bull. Math. Statist. 15 (1973) 69-78.

J.E. Yukich, Weak convergence of the empirical characteristic function, Proc. Amer. Math. Soc. 95 (1985) 470-473.