Ann. I. H. Poincaré – PR 38, 6 (2002) 1039–1052

S0246-0203(02)01138-X/FLA

MAXIMAL INEQUALITIES VIA BRACKETING WITH ADAPTIVE TRUNCATION

LES INÉGALITÉS MAXIMALES ET LA MÉTHODE FONDÉE SUR L'ENTROPIE À CROCHETS AVEC TRONCATURE ADAPTATIVE

David POLLARD

Statistics Department, Yale University, Box 208290 Yale Station, New Haven, CT 06520, USA

Received 1 June 2001, revised 1 May 2002

ABSTRACT. – The paper provides a recursive interpretation for the technique known as bracketing with adaptive truncation. By way of illustration, a simple bound is derived for the expected value of the supremum of an empirical process, thereby leading to a simpler derivation of a functional central limit theorem due to Ossiander. The recursive method is also abstracted into a framework that consists of only a small number of assumptions about processes and functionals indexed by sets of functions. In particular, the details of the underlying probability model are condensed into a single inequality involving finite sets of functions. A functional central limit theorem of Doukhan, Massart and Rio, for empirical processes defined by absolutely regular sequences, motivates the generalization.

MSC: primary 60E15; secondary 60G07, 60F17

RÉSUMÉ. – Cet article donne une interprétation récursive de la méthode fondée sur l'entropie à crochets ("bracketing") avec troncature adaptative. Pour illustrer cette approche, on montre de manière simple une inégalité maximale pour le processus empirique, laquelle implique un théorème limite central fonctionnel dû à Ossiander. Un autre T.L.C. fonctionnel, concernant le processus empirique associé à une suite absolument régulière de variables aléatoires (dû à Doukhan, Massart et Rio), motive une généralisation de la méthode. Cette généralisation ne requiert qu'un petit nombre d'hypothèses sur les processus et les fonctionnelles indexés par les ensembles de fonctions. En particulier, les détails du modèle probabiliste sont résumés par une seule inégalité pour les ensembles finis de fonctions.

E-mail address: david.pollard@yale.edu (D. Pollard).
URL address: http://www.stat.yale.edu/~pollard.

1. Introduction

In the empirical process literature, many important theorems and inequalities have been derived by a technique known as bracketing. Some of the arguments are long and involved, primarily because they require a delicate balancing act between several sequences of constants. The modern refinements due to the Seattle group (Pyke, Alexander, Bass, and Ossiander – for a discussion of their contributions see Section 6) are the most delicate of all because they combine bracketing with ingenious truncation arguments.

This paper presents a general method for handling bracketing arguments with truncation. By way of illustration, I begin with the important special case of a process constructed from independent random elements $\xi_1, \ldots, \xi_n$ taking values in a space $\mathcal{X}$. For f a real-valued function on $\mathcal{X}$ with each $f(\xi_i)$ integrable, define the centered-sum $\mathcal{S}_n f := \sum_{i \leqslant n}(f(\xi_i) - \mathbb{P}g(\xi_i))$.

Remark. – Throughout the paper I use the de Finetti notation [14, Chapter 1], writing $\mathbb{P}$ for expectations as well as probabilities, and identifying sets with their indicator functions. For example, $\mathbb{P}g\{g > c\}$ might be written as $\mathbb{E}(g\mathbf{1}\{g > c\})$ or as $\int_{g(x)>c} g(x)\mathbb{P}(dx)$ in traditional notation.

Some readers might be more familiar with the standardized form $\nu_n f := \mathcal{S}_n f/\sqrt{n}$, the so-called *empirical process*. Division by $\sqrt{n}$ is natural for the derivation of some limit theorems, particularly so for identically distributed $\{\xi_i\}$, but it would merely complicate the notation for the derivation of uniform approximations to a *centered-sum process* $\{\mathcal{S}_n f\colon\ f \in \mathcal{F}\}$ indexed by a set of functions $\mathcal{F}$ on $\mathcal{X}$.

The approximations in the present paper are derived (via bracketing and truncation arguments) using maps A_δ from $\mathcal{F}$ into finite sets of approximating functions. The main results take the form of bounds for quantities such as $\mathbb{P}\sup_{\mathcal{F}}|\mathcal{S}_n(f - A_\delta f)|$. (In fact, the theorems involve truncated functions, but the modification has only a minor effect on applications.) The behaviour of the process indexed by $\mathcal{F}$ is thereby related to the behaviour of a process $\{\mathcal{S}_n a\colon\ a \in \mathcal{A}\}$ with $\mathcal{A}$ a finite set of functions. Such an approximation underlies functional central limit theorems (fCLTs), functional laws of the iterated logarithm, and the stochastic equicontinuity results that are so useful for asymptotic inference. The rederivation in Section 3 of the fCLT for iid $\{\xi_i\}$, due to Ossiander [12], is typical.

A very simple form of bracketing is often used in textbooks to prove the Glivenko–Cantelli theorem, the most basic example of a uniform law of large numbers. The empirical distribution function F_n for a sample $\xi_1, \ldots, \xi_n$ from a distribution function F on the real line is defined by $F_n(t) := \sum_{i \leqslant n}\{\xi_i \leqslant t\}/n$ for each t in $\mathbb{R}$. That is, $F_n(t)$ denotes the proportion of the observations less than or equal to t. The Glivenko–Cantelli theorem asserts that $\sup_t |F_n(t) - F(t)|$ converges to zero almost surely.

The strong law of large numbers ensures that $F_n(t) - F(t) \to 0$ almost surely, for each fixed t. The bracketing argument then leads to uniform bounds over suitably small intervals, $t_1 \leqslant t \leqslant t_2$, by means of bounds that hold throughout the interval: for such t we have $F_n(t_1) - F(t_2) \leqslant F_n(t) - F(t) \leqslant F_n(t_2) - F(t_1)$. The two bounds converge almost surely to $F(t_1) - F(t_2)$ and $F(t_2) - F(t_1)$. If t_2 and t_1 are close enough together then all

the $F_n(t) - F(t)$ values, for $t_1 \leqslant t \leqslant t_2$, eventually get squeezed close to the origin. If we cover the whole real line by a union of finitely many such intervals, we are able to deduce that $\sup_t |F_n(t) - F(t)|$ is eventually small.

It is more fruitful to think of the increment $F(t_2) - F(t_1)$ as the $\mathcal{L}^1(P)$ distance between the two indicator functions $(-\infty, t_1]$ and $(-\infty, t_2]$, where P is the probability measure corresponding to the distribution function F. The concept of bracketing then has an obvious extension to more general sets of functions $\mathcal{F}$ on a set $\mathcal{X}$. The extension also makes sense for norms on spaces of functions more general than the $\mathcal{L}^1(P)$ norm. In particular, it has proved most useful for various $\mathcal{L}^2$ norms.

In what follows, $\#\mathcal{G}$ denotes the cardinality of a set $\mathcal{G}$.

DEFINITION 1. – *Let $\mathcal{U}$ be a vector space of functions equipped with a norm $\|\cdot\|$. Define the bracketing number $N(\delta, \mathcal{F})$ for a subset $\mathcal{F}$ of $\mathcal{U}$ as the smallest N for which there exists a partition of $\mathcal{F}$ into subsets $\mathcal{F}_1, \ldots, \mathcal{F}_N$ and functions $a_1, \ldots, a_N$ and $b_1, \ldots, b_N$ in $\mathcal{U}$ for which $\|b\| \leqslant \delta$ and $|f - a_i| \leqslant b_i$ pointwise when $f \in \mathcal{F}_i$.*

The bracketing defines two maps, A_δ and B_δ, from $\mathcal{F}$ into finite sets of functions: $A_\delta(f) := a_i$ and $B_\delta(f) := b_i$ when $f \in \mathcal{F}_i$. I will refer to $A_\delta(f)$ as the *approximating function*, $R_\delta(f) := f - A_\delta(f)$ as the *remainder*, and $B_\delta(f)$ as the *bracketing function*. The bracketing number $N(\cdot, \mathcal{F})$ is decreasing. It is of use only when finite-valued. Indeed, the most useful bounds require assumptions about the rate of increase of $N(\delta, \mathcal{F})$ as δ tends to zero, as in Ossiander's fCLT.

THEOREM 2 (Ossiander [12]). – *Suppose $\{\xi_i\}$ are independent and identically distributed random elements, each with marginal distribution P. Suppose $\mathcal{F} \subseteq \mathcal{L}^2(P)$ has an envelope F (a measurable function such that $|f(x)| \leqslant F(x)$ for all x and all f in $\mathcal{F}$) for which $PF^2 < \infty$. Let $N_2(\cdot)$ denote the bracketing numbers for $\mathcal{F}$ (under the $\mathcal{L}^2(P)$ norm). If $\int_0^1 \sqrt{\log N_2(x)}\, dx < \infty$ then $\{\nu_n f\colon\ f \in \mathcal{F}\}$ satisfies a fCLT.*

Ossiander derived her theorem from a bound on the tail probabilities for $\sup_{\mathcal{G}} |\nu_n g|$, for various sets of functions $\mathcal{G}$. Close inspection of her proofs, and of proofs for related theorems in the literature, reveals that independence is used only through a bound such as the Bennett inequality for sums of independent random variables [14, Section 11.2]. This inequality implies, for a function $g(\cdot)$ bounded in absolute value by a constant β with $Pg^2 \leqslant \delta^2$, that

$$\mathbb{P}\{|\nu_n g| \geqslant \lambda\delta\} \leqslant 2\exp\left(-\frac{1}{2}\lambda^2\psi\left(n^{-1/2}\beta\lambda/\delta\right)\right), \quad \text{for } \lambda \geqslant 0, \tag{1}$$

where $\psi(x)$ is a specified decreasing, nonnegative function with $\psi(0) = 1$.

The presence of the nuisance factor, $\psi(n^{-1/2}\beta\lambda/\delta)$, complicates the usual chaining argument for tail probabilities. If β and n stay fixed while λ/δ increases, the nuisance factor begins to dominate the bound. It was for this reason that Bass [3] and Ossiander [12] needed to add an extra truncation step to the chaining argument. The truncation keeps $n^{-1/2}\beta\lambda/\delta$ close enough to zero that one can ignore the nuisance factor and act as if $\nu_n g$ has sub-gaussian tails.

As you will see in the Section 2, under Ossiander's assumptions, a similar truncation scheme leads to a maximal inequality in the form of a bound for $\mathbb{P}\sup_{g\in\mathcal{G}}|\mathcal{S}_n g|$ for various $\mathcal{G}$. A proof of the fCLT follows easily (see Section 3).

2. Independent summands

Suppose $\xi_1,\dots,\xi_n$ are independent random variables. Define

$$\|g\|_1 := \sum_{i\leqslant n}\mathbb{P}\big|g(\xi_i)\big| \quad\text{and}\quad \|g\|_2 := \Big(\sum_{i\leqslant n}\mathbb{P}g(\xi_i)^2\Big)^{1/2}. \tag{2}$$

If each ξ_i has distribution P then $\|g\|_1 = nP|g|$ and $\|g\|_2^2 = nPg^2$.

The argument leading to the maximal inequality makes use of independence only through a maximal inequality for finite sets of functions. The method of proof combines an idea of Pisier [13] with the first step in the derivation of the Bennett inequality. It depends on the elementary fact [14, Section 11.2] that the function defined by $\mathcal{E}(x) := 2(\mathrm{e}^x - 1 - x)/x^2$ for $x \neq 0$, and $\mathcal{E}(0) = 1$, is positive and increasing over the whole real line.

LEMMA 3. – *Suppose $\xi_1,\dots,\xi_n$ are independent and $\mathcal{G}$ is a finite set of functions, for each of which $\sup_x|g(x)|\leqslant\beta$ and $\|g\|_2\leqslant\delta$. Then*

$$\mathbb{P}\max_{g\in\mathcal{G}}|\mathcal{S}_n g| \leqslant C_0\delta\sqrt{\log(2\#\mathcal{G})} \quad \textit{if } \beta\leqslant\delta/\sqrt{\log(2\#\mathcal{G})} \quad \textit{where } C_0\approx 1.718.$$

Proof. – Write N for $\#\mathcal{G}$, the cardinality of $\mathcal{G}$. For a fixed function g with $|g|\leqslant\beta$ and $\|g\|_2\leqslant\delta$, temporarily write W_i for $g(\xi_i)$ and μ_i for $\mathbb{P}g(\xi_i)$. For each $t>0$,

$$\mathbb{P}\mathrm{e}^{t\sum_{i\leqslant n}W_i} = \prod_i\Big(1 + t\mathbb{P}W_i + \mathbb{P}\frac{1}{2}t^2W_i^2\mathcal{E}(tW_i)\Big) \leqslant \prod_i\exp\Big(t\mu_i + \frac{1}{2}t^2\mathbb{P}W_i^2\mathcal{E}(t\beta)\Big),$$

which rearranges to the give $\mathbb{P}\exp(t\mathcal{S}_n g)\leqslant\exp(\frac{1}{2}t^2\delta^2\mathcal{E}(t\beta))$. Applying this bound for $\pm g$, for each g in $\mathcal{G}$, we get

$$\begin{aligned}\exp\Big(t\mathbb{P}\max_{\mathcal{G}}|\mathcal{S}_n g|\Big) &\leqslant \mathbb{P}\exp\Big(t\max_{\mathcal{G}}|\mathcal{S}_n g|\Big) \quad \text{by Jensen's inequality}\\ &\leqslant \sum_{g\in\mathcal{G}}\big(\mathbb{P}\exp(t\mathcal{S}_n g) + \mathbb{P}\exp\big(t\mathcal{S}_n(-g)\big)\big)\\ &\leqslant 2N\exp\Big(\frac{1}{2}t^2\delta^2\mathcal{E}(t\beta)\Big).\end{aligned}$$

Take logarithms then put $t = \sqrt{\log(2N)}/\delta$ to get

$$\mathbb{P}\max_{\mathcal{G}}|\mathcal{S}_n g| \leqslant \delta\sqrt{\log(2N)}\Big(1 + \frac{1}{2}\mathcal{E}\big(\beta\sqrt{\log(2N)}/\delta\big)\Big).$$

The asserted maximal inequality with $C_0 := 1 + \frac{1}{2}\mathcal{E}(1)$ follows. □

The main parts of the proof will involve the calculation of bounds for $\mathbb{P}\sup_{r\in\mathcal{R}}|\mathcal{S}_n r|$ for (possibly infinite) sets $\mathcal{R}$, typically consisting of truncated remainder functions derived from various bracketing approximations to $\mathcal{F}$. To reduce the calculations to finite sets of functions, we will bound each r in absolute value by a truncated bracketing function b.

LEMMA 4. – *Suppose a set of nonnegative functions $\mathcal{B}$ dominates a set of functions $\mathcal{R}$, in the sense that for each $r \in \mathcal{R}$ there is a $b \in \mathcal{B}$ for which $|r| \leqslant b$. Then* $\sup_{r\in\mathcal{R}}|\mathcal{S}_n r| \leqslant \sup_{b\in\mathcal{B}}|\mathcal{S}_n b| + 2\sup_{b\in\mathcal{B}}\|b\|_1$.

Proof. – If $|r| \leqslant b$ then $|\mathcal{S}_n r| \leqslant \sum_{i\leqslant n}(|r(\xi_i)| + \mathbb{P}|r(\xi_i)|) \leqslant \sum_{i\leqslant n}(b(\xi_i) + \mathbb{P}b(\xi_i))$. □

The successive approximations will be combined in such a way that the bounding functions b are not only truncated above but also below, a subtlety that will allow us to bound $\mathcal{L}^1$ norms by $\mathcal{L}^2$ norms.

LEMMA 5. – *For each function b with finite $\mathcal{L}^2$ norm, $\|b\{|b| \geqslant \|b\|_2/t\}\|_1 \leqslant t\|b\|_2$.*

Proof. – $\mathbb{P}\sum_{i\leqslant n}|b(\xi_i)\{|b(\xi_i)| \geqslant \|b\|_2/t\}| \leqslant \mathbb{P}\sum_{i\leqslant n} b(\xi_i)^2/(\|b\|_2/t)$. □

The inequalities from the three lemmas capture everything we have to know about the $\{\xi_i\}$ and the norms in order to derive the main approximation result.

THEOREM 6. – *Let $N(x)$ denote the bracketing number of a set of functions $\mathcal{F}$ under the $\mathcal{L}^2$ norm from* (2). *For a fixed $\delta > 0$, define $\delta_i := \delta/2^i$ and $\beta_i := \delta_i/H(n(\delta_i))$, with $n(y) := N(y)N(y/2)$ and $H(N) := \sqrt{\log(2N)}$. Define*

$$\Delta_i := \mathbb{P}\sup_{f\in\mathcal{F}}\bigl|\mathcal{S}_n\bigl(R_{\delta_i}(f)\{B_{\delta_i}(f) \leqslant \beta_i\}\bigr)\bigr|.$$

Then

$$\Delta_0 \leqslant \Delta_k + 71\int_{\delta_{k+2}}^{\delta_1} H\bigl(N(y)\bigr)\,\mathrm{d}y \quad \textit{for each } k.$$

Remark. – Of course a quantity such as $\sup_{g\in\mathcal{G}}|\mathcal{S}_n g|$ need not be measurable if $\mathcal{G}$ is uncountable. The expectation in the definition of Δ_i should actually be interpreted as an outer expectation. In fact, most of the inequalities needed for the proofs involve upper bounds depending on only finite subsets of $\mathcal{L}^2(P)$, for which the measurability problem disappears.

Proof. – Construct the bracketing approximations for each δ_i, for $i = 0, 1, \ldots, k$. To simplify notation, abbreviate $n(\delta_i)$ to n_i and define $\gamma_i := H(n_i)$. Similarly, abbreviate $A_{\delta_i}(f)$ to A_i, and $B_{\delta_i}(f)$ to B_i, and $R_{\delta_i}(f)$ to R_i, with the argument f understood. Notice that $|R_i| \leqslant B_i$, which implies that $\|R_i\|_2 \leqslant \|B_i\|_2 \leqslant \delta_i$. Write T_i for the truncation region $\{B_{\delta_i}(f) \leqslant \beta_i\}$ and $\mathfrak{m}_{\mathcal{F}}(\cdots)$ for $\mathbb{P}\sup_{f\in\mathcal{F}}|\mathcal{S}_n(\cdots)|$. With this notation we have $\Delta_i = \mathfrak{m}_{\mathcal{F}}(R_i T_i)$. Section 4 will show that the subadditivity property of the functional $\mathfrak{m}_{\mathcal{F}}$ is really what drives the argument.

The key idea behind the Seattle method is captured by a recursive equality,

$$R_i T_i = R_{i+1}T_{i+1} - R_{i+1}T_i^c T_{i+1} + (R_i - R_{i+1})T_i T_{i+1} + R_i T_i T_{i+1}^c,$$

which relates the truncated remainder terms for successive bracketing approximations. Applying $\mathfrak{m}_{\mathcal{F}}$ to the sets of functions on both sides of this equality, we get

$$\Delta_i \leqslant \Delta_{i+1} + \mathfrak{m}_{\mathcal{F}}\big(R_{i+1}T_i^c T_{i+1}\big) + \mathfrak{m}_{\mathcal{F}}\big((R_i - R_{i+1})T_i T_{i+1}\big) + \mathfrak{m}_{\mathcal{F}}\big(R_i T_i T_{i+1}^c\big). \tag{3}$$

Together the three Lemmas will provide bounds for the second, third, and fourth terms on the right-hand side.

Contribution of the third term from (3). As f ranges over the set $\mathcal{F}$, the truncated difference function $(R_i - R_{i+1})T_i T_{i+1} = -(A_i - A_{i+1})T_i T_{i+1}$ ranges over at most n_i distinct functions. Moreover,

$$\big\|(R_i - R_{i+1})T_i T_{i+1}\big\|_2 \leqslant \|R_i\|_2 + \|R_{i+1}\|_2 \leqslant \delta_i + \delta_{i+1}$$

and

$$\begin{aligned} |R_i - R_{i+1}|T_i T_{i+1} &\leqslant B_i T_i + B_{i+1}T_{i+1} \ \leqslant \beta_i + \beta_{i+1} \\ &\leqslant \delta_i/\gamma_i + \delta_{i+1}/\gamma_{i+1} \leqslant (\delta_i + \delta_{i+1})/H(n_i). \end{aligned}$$

Thus the set of functions $\{(R_i - R_{i+1})T_i T_{i+1}\colon\ f \in \mathcal{F}\}$ satisfies the conditions of Lemma 3, which gives

$$\mathfrak{m}_{\mathcal{F}}\big((R_i - R_{i+1})T_i T_{i+1}\big) \leqslant C_0(\delta_i + \delta_{i+1})\gamma_i. \tag{4}$$

Contribution of the second term from (3). The set of functions $\{R_{i+1}T_i^c T_{i+1}\colon\ f \in \mathcal{F}\}$ is potentially infinite, but it is dominated by the set $\{B_{i+1}T_i^c T_{i+1}\colon\ f \in \mathcal{F}\}$, which contains at most n_i nonnegative functions, each bounded above by β_{i+1} and with $\mathcal{L}^2$ norm at most δ_{i+1}. Moreover, by splitting according to which of B_i or B_{i+1} is larger, we get the inequality

$$\begin{aligned} \big\|B_{i+1}T_i^c T_{i+1}\big\|_1 &\leqslant \big\|B_i\{B_i > \beta_i\}\big\|_1 + \big\|B_{i+1}\{B_{i+1} > \beta_i\}\big\|_1 \\ &\leqslant \big\|B_i\{B_i > \|B_i\|_2/\gamma_i\}\big\|_1 + \big\|B_{i+1}\{B_{i+1} > 2\|B_{i+1}\|_2/\gamma_i\}\big\|_1 \\ &\leqslant \delta_i\gamma_i + \frac{1}{2}\delta_{i+1}\gamma_i \quad \text{by Lemma 5.} \end{aligned}$$

From Lemmas 4 and 3 deduce that

$$\mathfrak{m}_{\mathcal{F}}\big(R_{i+1}T_i^c T_{i+1}\big) \leqslant C_0\delta_{i+1}\gamma_i + 2\left(\delta_i\gamma_i + \frac{1}{2}\delta_{i+1}\gamma_i\right). \tag{5}$$

Contribution of the fourth term from (3). The argument is almost the same as for the second term. Each of the dominating functions $B_i T_i T_{i+1}^c$ is bounded above by β_i, has $\mathcal{L}^2$ norm at most δ_i, and

$$\big\|B_i T_i T_{i+1}^c\big\|_1 \leqslant \big\|B_i\{B_i > \beta_{i+1}\}\big\|_1 + \big\|B_{i+1}\{B_{i+1} > \beta_{i+1}\}\big\|_1 \leqslant 5\delta_{i+1}\gamma_{i+1}.$$

Again from Lemmas 4 and 3 deduce that

$$\mathfrak{m}_{\mathcal{F}}\big(R_i T_i T_{i+1}^c\big) \leqslant C_0\delta_i\gamma_i + 10\delta_{i+1}\gamma_{i+1}. \tag{6}$$

Recursive inequality. From inequalities (3), (4), (5), and (6),

$$\Delta_i \leqslant \Delta_{i+1} + (5 + 6C_0)\delta_{i+1}\gamma_{i+1} + 20\delta_{i+2}\gamma_{i+1}.$$

Subadditivity of the square-root function gives

$$\gamma_i = \sqrt{\log\bigl(n(\delta_i)\bigr)} \leqslant \sqrt{\log\bigl(2N(\delta_i)\bigr)} + \sqrt{\log\bigl(2N(\delta_{i+1})\bigr)} \leqslant 2H\bigl(N(\delta_{i+1})\bigr).$$

By repeated substitution we are then left with the inequality

$$\Delta_0 \leqslant \Delta_k + \sum_{i=0}^{k-1}(10 + 12C_0)(\delta_{i+1} - \delta_{i+2})H\bigl(N(\delta_{i+1})\bigr) + 40(\delta_{i+2} - \delta_{i+3})H\bigl(N(\delta_{i+2})\bigr).$$

Monotonicity of the function $y \mapsto H(N(y))$ lets us bound the summands by multiples of integrals of the form $\int\{\delta_{j+1} < y \leqslant \delta_j\}H(N(y))\,dy$, from which the assertion of the theorem follows because $50 + 12C_0 \approx 70.62$. □

COROLLARY 7. – *Under the conditions of the theorem,* $\Delta_0 \leqslant 71\int_0^{\delta/2} H(N(y))\,dy$.

Proof. – Note that $|R_kT_k| \leqslant \beta_k \to 0$ as $k \to \infty$, implying that $\Delta_k \to 0$ for fixed n. □

3. Proof of Ossiander's functional CLT

The theorem asserts convergence in distribution of ν_n to a Gaussian process $\{\nu f\colon f \in \mathcal{F}\}$. To prove her theorem, Ossiander [12] needed to show

(a) finite dimensional convergence: $\{\nu_n g\colon g \in \mathcal{G}\} \rightsquigarrow \{\nu g\colon g \in \mathcal{G}\}$ for each finite subset $\mathcal{F}$;

(b) stochastic equicontinuity: for each $\eta > 0$ and $\varepsilon > 0$, there exists a $\delta > 0$ for which $\mathbb{P}\{\sup_{\|f-g\|<\delta} |\nu_n f - \nu_n g| > \eta\} \leqslant \varepsilon$ for all n large enough. (The supremum runs over all pairs of functions in $\mathcal{F}$ whose $\mathcal{L}^2(P)$ distance is smaller than δ.)

The assumption of identical distributions for the $\{\xi_i\}$ is not crucial for the validity of a fCLT. It ensures that (a) follows directly from the multivariate central limit theorem, and it slightly simplifies the notation. Ossiander's methods also work for more general triangular arrays.

Square integrability of F ensures that, for each fixed $\varepsilon > 0$,

$$\mathbb{P}\Bigl\{\max_{i\leqslant n} F(\xi_i) > \varepsilon\sqrt{n}\Bigr\} \leqslant nP\{F > \varepsilon\sqrt{n}\} \leqslant P\bigl(F^2\{F > \varepsilon\sqrt{n}\}\bigr) \to 0 \quad \text{as } n \to \infty.$$

The same assertion holds with ε replaced by an ε_n that tends to zero slowly enough. Thus there exists a sequence of constants M_n of order $\mathrm{o}(\sqrt{n})$ for which $\max_{i\leqslant n} F(\xi_i) \leqslant M_n$ with probability tending to one. Define

$$\mathcal{H} = \mathcal{H}_n(\delta) := \bigl\{(f-g)\{F \leqslant M_n\}/\sqrt{n}\colon f, g \in \mathcal{F} \text{ and } P(f-g)^2 < \delta^2\bigr\}.$$

If we show that $\limsup_n \mathbb{P}\sup_{h\in\mathcal{H}} |\mathcal{S}_n h| \to 0$ as $\delta \to 0$ then (b) will follow.

To avoid confusion between norms, write $A_y^*(f)$ and $B_y^*(f)$ for the approximating functions and bracketing functions for $\mathcal{F}$ under the $\mathcal{L}^2(P)$ norm. The corresponding bracketing numbers are given by the function $N_2(\cdot)$. If $h = (f-g)\{F \leqslant M_n\}/\sqrt{n}$ we may take

$$A_y(h) := \big(A_{y/2}^*(f) - A_{y/2}^*(g)\big)\{F \leqslant M_n\}/\sqrt{n},$$
$$B_y(h) := \big(B_{y/2}^*(f) + B_{y/2}^*(g)\big)\{F \leqslant M_n\}/\sqrt{n}.$$

The bracketing number $N(y,\mathcal{H})$ for $\mathcal{H}$ under the $\|\cdot\|_2$ norm from (2) is then smaller than $N_2(y/2)^2$. For y equal to δ we can do much better by redefining $A_\delta(h) \equiv 0$ and $B_\delta(h) \equiv 2F\{F \leqslant M_n\}/\sqrt{n}$, which gives $N(\delta,\mathcal{H}) = 1$. Notice that $B_\delta(h) \leqslant 2M_n/\sqrt{n} \to 0$, which implies that $\{B_\delta(h) \leqslant \beta_0\}$ is equal to the whole space when n is large enough. That is, we can eventually ignore the trunction factor in the definition of Δ_0, and deduce via Corollary 7 that

$$\mathbb{P}\sup_{h\leqslant\mathcal{H}} |\mathcal{S}_n h| = \Delta_0 \leqslant 71 \int_0^{\delta/2} \sqrt{\log(2N_2(y)^2)}\,dy \quad \text{for large enough } n.$$

The integral on the right-hand side converges to zero with δ.

Remark. – We were able to argue directly via Corollary 7 because $\log(2N_2(y)^2)$ increases like $\log(2N_2(y))$. For the analogous results in the next Section we might not have the benefit of a logarithm to counter the squaring of the bracketing number. We could however argue directly from Theorem 6 using the method of Ledoux and Talagrand [10, Theorem 11.6] to avoid the problem caused by working with sets of differences.

4. Generalization

The three lemmas in Section 2 and the method of proof suggest that the theorem really depends only on the relationship between a functional $\mathfrak{m}_{\mathcal{F}}$ and the norms $\|g\|_1$ and $\|g\|_2$. Indeed, the argument extends readily to more general functionals defined for subsets $\mathcal{G}$ of a vector space of functions $\mathcal{U}$. There are also extensions to functionals with properties analogous to tail probabilities and to more complicated truncation schemes, as in Birgé and Massart [5]; but, for simplicity of exposition, I describe only one generalization.

The role of the $\mathcal{L}^2$ norm from Section 2 will be taken over by a general norm $\|\cdot\|$ on $\mathcal{U}$. In fact, we do not need all the properties of a norm: it will suffice that $\|\cdot\|$ is *subadditive*, that is, $\|g_1+g_2\| \leqslant \|g_1\| + \|g_2\|$ for all $g_1, g_2 \in \mathcal{U}$. Similarly, the role of the $\mathcal{L}^1$ norm will be taken over by a second subadditive map ρ from $\mathcal{U}$ into $\mathbb{R}^+$. In place of $\mathfrak{m}_{\mathcal{F}}$, consider a functional $\mathfrak{m}$ that assigns a nonnegative number $\mathfrak{m}(\mathcal{G})$ to each subset $\mathcal{G}$ of $\mathcal{U}$. Assume that the following properties hold.

(i) if $g_1, g_2 \in \mathcal{U}$ and $c \in \mathbb{R}$ then $g_1\{g_2 \leqslant c\} \in \mathcal{U}$ and $g_1\{g_2 > c\} \in \mathcal{U}$
(ii) if $|g_1| \leqslant |g_2|$ pointwise then $\|g_1\| \leqslant \|g_2\|$ and $\rho(g_1) \leqslant \rho(g_2)$
(iii) if subsets $\mathcal{G}, \mathcal{G}', \mathcal{G}''$ of $\mathcal{U}$ are such that each g in $\mathcal{G}$ can be written as a sum $g' + g''$, with $g' \in \mathcal{G}'$ and $g'' \in \mathcal{G}''$, then $\mathfrak{m}(\mathcal{G}) \leqslant \mathfrak{m}(\mathcal{G}') + \mathfrak{m}(\mathcal{G}'')$

(iv) there exist nonnegative, increasing functions $G(N)$ and $H(N)$ for which: if $\mathcal{G}$ is a finite subset of functions from $\mathcal{U}$ for each of which $\|g\| \leqslant \delta$ and $\sup_x |g(x)| \leqslant \beta \leqslant \delta/G(\#\mathcal{G})$ then $\mathfrak{m}(G) \leqslant \delta H(\#\mathcal{G})$

(v) if $\mathcal{H}$ dominates $\mathcal{G}$, in the sense that for each g in $\mathcal{G}$ there is an h in $\mathcal{H}$ for which $|g| \leqslant h$, then $\mathfrak{m}(\mathcal{G}) \leqslant \mathfrak{m}(\mathcal{H}) + \sup_{h\in\mathcal{H}} \rho(h)$

(vi) there is an increasing, nonnegative function D for which $\rho(g\{|g| > \|g\|/t\}) \leqslant \|g\| D(t)$ for each $t > 0$ and $g \in \mathcal{U}$.

Assumption (iii) is the subadditivity property that will allow us to develop a recursive inequality analogous to (3). For example, any functional defined by taking an $\mathcal{L}^p$ norm of $\sup_{g\in\mathcal{G}} |\mathcal{S}_n g|$ is subadditive in the sense of (iii). Assumption (iv) corresponds to Lemma 3, but with the dual role of the function $\sqrt{\log(2N)}$ split between two separate functions, G and H. The extra generality is not needed for the examples discussed in the present paper, but it does serve to clarify the two roles played by $\sqrt{\log(2N)}$ in Theorem 6. Assumption (v) corresponds to Lemma 4, with a slight tidying of constants. Assumption (vi) extends Lemma 5 by allowing a more subtle dependence on t, a generalization motivated by the results of Doukhan, Massart, and Rio [7], as described in the next section. It implies that, for all nonnegative g_1 and g_2 in $\mathcal{U}$,

$$\rho\big(g_1\{g_2 > c\}\big) \leqslant \|g_1\| D\big(\|g_1\|/c\big) + \|g_2\| D\big(\|g_2\|/c\big), \tag{7}$$

an inequality derived via the subadditivity of ρ by splitting according to which of g_1 or g_2 is larger, as in the argument for the second term from (3) in Section 2.

THEOREM 8. – *Let $N(x)$ denote the bracketing number of a set of functions $\mathcal{F} \subseteq \mathcal{U}$ under the norm $\|\cdot\|$. Assume that* (i) *through* (vi) *hold. For a fixed $\delta > 0$, define $\delta_i := \delta/2^i$ and $\beta_i := \delta_i/G(n(\delta_i))$, with $n(y) := N(y)N(y/2)$. Define*

$$\Delta_i := \mathfrak{m}\big\{R_{\delta_i}(f)\big\{B_{\delta_i}(f) \leqslant \beta_i\big\}\colon\ f \in \mathcal{F}\big\}.$$

Then for some universal constant C,

$$\Delta_0 \leqslant \Delta_k + C \int_{\delta_{k+2}}^{\delta_1} H\big(n(y)\big) + D\big(2G\big(n(y)\big)\big)\, dy \quad \textit{for each } k.$$

Outline of proof. – Define A_i, B_i, n_i, R_i, and T_i as in the proof of Theorem 3. From the recursive equality for the truncated remainder $R_i T_i$, argue via (iii) that

$$\begin{aligned}\Delta_i \leqslant{}& \Delta_{i+1} + \mathfrak{m}\big\{-R_{i+1}T_i^c T_{i+1}\colon\ f \in \mathcal{F}\big\}\\ &+ \mathfrak{m}\big\{(R_i - R_{i+1})T_i T_{i+1}\colon\ f \in \mathcal{F}\big\} + \mathfrak{m}\big\{(R_i T_i T_{i+1}^c\colon\ f \in \mathcal{F}\big\}.\end{aligned}$$

For the second term on the right-hand side, invoke (v) for the dominating set of functions $\{B_{i+1}T_i^c T_{i+1}\colon\ f \in \mathcal{F}\}$ then appeal to (7) to derive the bound

$$\delta_{i+1} H\big(n(\delta_i)\big) + \delta_{i+1} D\big(G\big(n(\delta_{i+1})/2\big)\big) + \delta_i D\big(G\big(n(\delta_i)\big)\big).$$

And so on, along the same lines as the proof of Theorem 6. □

5. Absolute regularity

Doukhan, Massart, and Rio [7] – henceforth DMR – established a functional central limit theorem for stationary, absolutely regular sequences $\{\xi_i\}$ of random elements of a Polish space $\mathcal{X}$, each with distribution P. Their method fits into the framework of Theorem 8 with $\mathfrak{m}(\mathcal{F}) = \mathbb{P}\sup_{f\in\mathcal{F}}|\nu_n f|$ and $\rho(g) := 2\sqrt{n}P|g|$. With small modifications, their Lemma 3 gives a maximal inequality as in (iv) and their Lemma 4 gives (vi) for an unusual D. This Section outlines the argument.

The definition of absolute regularity involves a decreasing sequence of mixing coeficients $\{r_q\colon\ q = 0, 1, 2, \ldots\}$. We may assume that $r_q = r(q)$, where $r(\cdot)$ is a continuous, decreasing function on $\mathbb{R}^+$ with $r(0) = 1$ and $r(x) \to 0$ as $x \to \infty$. The function r has a right-continuous, decreasing "inverse" function, defined by $r^{-1}(u) := \inf\{x\colon\ r(x) \leqslant u\}$ for $0 < u < 1$. Similarly, the tail quantile function Q_f for a measurable real function f on $\mathcal{X}$ is defined by

$$Q_f(u) := \inf\bigl\{x\colon\ P\{|f| > x\} \leqslant u\bigr\} \quad \text{for } 0 < u < 1.$$

If U is distributed Uniform$(0, 1)$ then $Q_f(U)$ has the same distribution as $|f|$ under P, a representation that will be needed in Lemma 9. Following [16], DMR defined $\|f\|^2 := \int_0^1 r^{-1}(u)Q_f(u)^2\,du$ for real measurable functions on $\mathcal{X}$. The set $\mathcal{U}$ of all f for which $\|f\| < \infty$ is a vector space for which assumptions (i) and (ii) hold.

As noted by DMR, the precise definition of absolute regularity of the sequence is unimportant. It matters only that there exists a coupling with a process constructed from independent random vectors, as follows. For any positive integer q, break $\{\xi_i\}$ into a sequence of q-vectors $Y_1, Y_2, \ldots$. That is, Y_i has components ξ_j for $j \in \mathcal{N}_i := \{1 + (i-1)q, \ldots, iq\}$. Then there exists a sequence of q-vectors Y_i^* for which: (a) Y_i^* has the same distribution as Y_i, for each i; (b) $\mathbb{P}\{Y_i \neq Y_i^*\} \leqslant r_q$; and (c) $\{Y_{2i}^*\colon\ i = 1, 2, \ldots\}$ are independent and so are $\{Y_{2i-1}^*\colon\ i = 1, 2, \ldots\}$.

If the integer q lies in the range $1 \leqslant q \leqslant n$, properties (a), (b) and (c) let us couple the empirical process ν_n with a sum of two processes $\nu_n^* + \nu_n^{**}$, with ν_n^* constructed from the ξ_j^* variables from the $\mathcal{N}_{2i}$ blocks and ν_n^{**} constructed from the remaining variables, leading to the inequality

$$\mathbb{P}\max_{g\in\mathcal{G}}|\nu_n g| \leqslant \mathbb{P}\max_{g\in\mathcal{G}}|\nu_n^* g| + \mathbb{P}\max_{g\in\mathcal{G}}|\nu_n^{**} g| + 2\beta r_q\sqrt{n} \quad \text{if } \max_{g\in\mathcal{G}}|g| \leqslant \beta.$$

If $\mathcal{G}$ is a set of at most N functions from $\mathcal{U}$, each bounded in absolute value by a constant β and with norm less than δ, we may apply the method of Lemma 3 with W_i equal to a sum $\sum_{j\in\mathcal{N}_i}(g(\xi_j^*) - \mathbb{P}g(\xi_j^*))/\sqrt{n}$, first for even then for odd values of i, in order to bound both $\mathbb{P}\exp(t\nu_n^* g)$ and $\mathbb{P}\exp(t\nu_n^{**} g)$ by expressions of the form $\exp(ct^2\|g\|^2\mathcal{E}(c'q\beta t/\sqrt{n}))$, for constants c and c'. We then deduce that

$$\mathbb{P}\max_{g\in\mathcal{G}}|\nu_n g| \leqslant c_0\delta\ell_N\left(1 + \mathcal{E}\left(\frac{c_1 q\beta\ell_N}{\delta\sqrt{n}}\right) + \frac{q\beta}{\delta}\frac{r(q)\sqrt{n}}{q\ell_N}\right), \tag{8}$$

where c_0 and c_1 are constants and $\ell_N = \ell(N) := \sqrt{1 + \log N}$. (We could take ℓ_N as $\sqrt{\log(2N)}$, but the slightly larger value ensures $\ell_N \geqslant 1$ for all $N \geqslant 1$.) With a slight

increase in the constants, inequality (8) also holds for all q in the continuous range $[1, n]$. With an appropriate choice for q, the inequality will become the desired maximal inequality (iv).

DMR established a functional central limit theorem for subsets $\mathcal{F}$ of $\mathcal{U}$ for which $\int_0^1 \sqrt{\log N(x, \mathcal{F})}\, dx < \infty$, for the covering numbers under their new norm, and with envelope F for which $\|F\| < \infty$. They assumed that $\sum_q r_q < \infty$, which implies $\int_0^1 r^{-1}(u)\, du < \infty$, thereby ensuring that the function $R(x) := \int_0^{r(x)} r^{-1}(u)\, du$ is continuous and decreases to zero as x tends to infinity. With these functions, we can define a suitable D for assumption (vi).

LEMMA 9. – *For each f in $\mathcal{U}$ and each $x > 0$ define $\|f\|_x^2 := \int_0^{r(x)} r^{-1}(u) Q_f(u)^2\, du$. Then $P|f|\{|f| > \|f\|_x/\sqrt{R(x)}\} \leqslant \|f\|_x \sqrt{r(x)/x}$.*

Proof. – First note that $\|f\|_x^2 \geqslant R(x) Q_f(r(x))^2$, because Q_f is a decreasing function. Thus the quantity on the left-hand side of the asserted inequality is less than

$$\int_0^1 Q_f(u)\{Q_f(u) > Q_f(r(x))\}\, du \leqslant \int_0^1 Q_f(u)\{u < r(x)\}\, du$$
$$\leqslant \int_0^1 \sqrt{r^{-1}(u)/x}\, Q_f(u)\{u < r(x)\}\, du,$$

the second inequality following from the fact that $r^{-1}(u) > x$ when $u < r(x)$. The Cauchy–Schwarz inequality completes the proof. □

If we replace $\|f\|_x$ in the lemma by the larger $\|f\|$, we get a weaker inequality that suggests we should define D indirectly by putting

$$D(t) := 2\sqrt{nr(x)/x} \quad \text{when } t = \sqrt{R(x)}. \tag{9}$$

The definition makes sense for all t in the range $0 \leqslant t \leqslant \sqrt{R(0)}$. It will turn out that we only need to consider such values of t. Indeed, the largest t needed for the proofs is $2G(n_k)$. We keep this value within the required range by defining $G(N) := \frac{1}{2}\sqrt{R(q_N)}$ for a value q_N that will be determined by the requirements of the maximal inequality (iv). These choices give $D(2G(N)) = 2\sqrt{nr(q_N)/q_N}$ and $1/G(N) \leqslant 2/\sqrt{q_N r(q_N)}$, because $R(x) \geqslant xr(x)$ for all $x \geqslant 0$.

How should we choose $q = q_N$ to balance the requirements of assumptions (iv) and (vi)? At best we can make the right-hand side of (8) smaller than a multiple of $\delta\ell_N$ by keeping β/δ smaller than a multiple of $\min(\sqrt{n}/(q\ell_N), \ell_N/(\sqrt{n}r_q))$. One term in the minimum decreases as q gets larger, the other increases. We get the largest range for β/δ by balancing the terms: choose q equal to the value q_N for which $r(q_N)/q_N = \ell_N^2/n$, an equality that defines a unique value in the range $[1, n]$ when $\ell_N^2 \leqslant nr(1)$. (The upper bound on q_N comes from the fact that $\ell_N^2/n \geqslant 1/n \geqslant r(n)/n$.) Provided β/δ is smaller than $1/G(N) := 2/\sqrt{R(q_N)} \leqslant 2/\sqrt{q_N r(q_N)}$, we then bound the right-hand side of (8) for $q = q_N$ by $c_0\delta\ell_N(1 + \mathcal{E}(2c_1) + 2)$. That is, assumption (iv) holds with $H(N)$ a constant multiple of ℓ_N and $G(N) = \frac{1}{2}\sqrt{R(q_N)}$, provided we consider only values of N

for which $\ell_N^2 \leqslant nr(1)$. We also have $D(2G(N)) = 2\ell_N$. An appeal to Theorem 8 then gives the bound

$$\Delta_0 := \mathfrak{m}\big(R_0(f)\{B_0(f) \leqslant \beta_0\}\big) \leqslant \Delta_k + C' J(\delta) \quad \text{where } J(\delta) := \int_0^\delta \ell\big(n(x)\big)\,dx.$$

The assumed finiteness of $\int_0^1 \sqrt{\log N(x,\mathcal{F})}\,dx$ ensures that $J(\delta)$ converges to zero as δ tends to zero. We have only to choose k so that Δ_k is suitably small and $\ell(n_k)^2 \leqslant nr(1)$. The largest k for which $\sqrt{n}\delta_k \leqslant J(\delta)$ will suffice if δ is small enough. With that choice we have $\delta_k \ell(n_k) \leqslant J(\delta_k) = \mathrm{o}(1) = \mathrm{o}(\sqrt{n}\delta_k)$, and, by (iv) and (vi) applied to $\{(f - A_k(f))\{B_k(f) \leqslant \beta_k\}\colon f \in \mathcal{F}\}$,

$$\Delta_k \leqslant \mathfrak{m}_{\mathcal{F}}\big(B_k\{B_k \leqslant \beta_k\}\big) + 2\sqrt{n}\max_{\mathcal{F}} P\big(B_k\{B_k \leqslant \beta_k\}\big) \leqslant \delta_k H(N_k) + 2\sqrt{n}\delta_k,$$

which is smaller than some constant multiple of $J(\delta)$.

As in Section 3, we can eliminate the effect of the indicator $\{B_0 \leqslant \beta_0\}$ from Δ_0 by means of an initial truncation based on the finiteness of $\|F\|$. For each fixed C, the sequence $M_n = \|F\|_{x_n}/\sqrt{R(x_n)}$, where x_n is defined by the equalities $r(x_n)/x_n = C/n$, has the property

$$PF\{F > M_n\} \leqslant \|F\|_{x_n}\sqrt{r(x_n)/x_n} = \mathrm{o}\big(n^{-1/2}\big) \quad \text{by Lemma 9.}$$

If we let C tend to infinity slowly enough with n, we get sequences $\{x_n\}$ and $\{M_n\}$ for which $nr(x_n)/x_n \to \infty$ and

$$\mathbb{P}\sup_{f\in\mathcal{F}}\big|\nu_n\big(f\{F > M_n\}\big)\big| \leqslant 2\sqrt{n}PF\{F > M_n\} \to 0.$$

Eventually M_n will be smaller than the truncation level $\beta_0 := 2\delta/\sqrt{R(q_{n(\delta)})}$, no matter how small we choose δ. Indeed, $q_{n(\delta)}$ is defined by the equality $r(q_{n(\delta)})/q_{n(\delta)} = \ell(n(\delta))/n = \mathrm{o}(r(x_n)/x_n)$. Eventually we must have $q_{n(\delta)} > x_n$ and hence $R(q_{n(\delta)}) \leqslant R(x_n)$. When we also have $\|F\|_{x_n} < 2\delta$ then it follows that $M_n < \beta_0$.

The rest of the argument leading to the functional central limit theorem follows the method outlined in Section 3.

6. Some history

Bracketing arguments have long been used to prove fCLTs: for example, the original paper of Donsker [6, near his Eq. 2.11] applied a version of the method.

Dudley [8] used the concept of metric entropy with bracketing for general classes of sets in order to prove a fCLT for empirical processes indexed by classes of sets. He later [9] extended the result to classes of functions with an envelope having a finite pth moment, for some $p > 2$. His method involved an initial truncation at a level much smaller than $\sqrt{n}$ and it required an assumption on the bracketing numbers stronger than Ossiander's condition.

Pyke [15] used a similar truncation to prove a CLT for processes indexed by sets. This result was refined first by Bass and Pyke [4], and then by Alexander and Pyke [1]. The second paper added the refinement of multiple levels of truncation (the stratification argument on page 589), to partition a partial-sum process into a sum of bounded processes, thereby obtaining the fCLT under the natural second moment and bracketing conditions. They cited the preprint form of Bass [3], who also applied stratification to prove a functional LIL for set-indexed processes. Ossiander [12, pp. 899, 903] stated that her chaining argument was adapted from the Bass paper. In a private communication, Ron Pyke explained to me that the history is more complicated than suggested by the publication dates:

> Ken Alexander saw the paper of Pyke [15], and realized how to improve the truncation technique used there. He applied the improvement in a 1984 paper. With Pyke he wrote another paper [1] – see the remarks at the end of the paper. Bass [3] applied the truncation to set-indexed partial-sum processes (the paper was not written up before December 1984). Bass and Pyke [4] (in a paper written around 1983, Pyke believes) recognized the truncation problem; but they didn't use the best form of truncation. Mina Ossiander worked on her dissertation during the spring and summer of 1984, producing her thesis – later published as Ossiander [12] – and a technical report in November–December of that year. Starting from the preprint form of [1], she developed a more general form of the truncation argument. There were many discussions between Ossiander and Bass. The final publication dates are not indicative of the true order in which work was carried out, because of delays in refereeing.

In view of this information, I think it is fair to spread the credit for the truncation method between all the members of the Seattle group.

My involvement with the method began in early 1985, with a study of [3] and Ossiander's thesis. By mid 1987, I realized that the argument could be thought of as a recursive procedure, an idea that I circulated in unpublished preprints. The generalization to dependent variables by Doukhan, Massart, and Rio [7] later suggested to me the possibility of the abstract version of the method, as presented in Section 4. The method has also been extended by Andersen et al. [2], replacing the concept of a bracketing number by the concept of a majorizing measure.

Acknowledgements

I thank the referee for suggestions that led to an improvement of the paper through sacrifice of generality for clarity.

REFERENCES

[1] K.S. Alexander, R. Pyke, A uniform central limit theorem for set-indexed partial-sum processes with finite variance, Ann. Probab. 14 (1986) 582–597.

[2] N.T. Andersen, E. Giné, M. Ossiander, J. Zinn, The central limit theorem and the law of the iterated logarithm for empirical processes under local conditions, Z. Wahrscheinlichkeitstheorie Verw. Geb. 77 (1988) 271–306.

[3] R.F. Bass, Law of the iterated logarithm for set-indexed partial-sum processes with finite variance, Z. Wahrscheinlichkeitstheorie Verw. Geb. 70 (1985) 591–608.
[4] R.F. Bass, R. Pyke, Functional law of the iterated logarithm and uniform central limit theorem for partial-sum processes indexed by sets, Ann. Probab. 12 (1984) 13–34.
[5] L. Birgé, P. Massart, Rates of convergence for minimum contrast estimators, Probab. Theory Related Fields 97 (1993) 113–150.
[6] M.D. Donsker, Justification and extension of Doob's heuristic approach to the Kolmogorov–Smirnov theorems, Ann. Math. Statist. 23 (1952) 277–281.
[7] P. Doukhan, P. Massart, E. Rio, Invariance principle for absolutely regular processes, Ann. Institut H. Poincaré 31 (1995) 393–427.
[8] R.M. Dudley, Central limit theorems for empirical measures, Ann. Probab. 6 (1978) 899–929.
[9] R.M. Dudley, Donsker classes of functions, in: M. Csörgő, D.A. Dawson, J.N.K. Rao, A.K.Md.E. Saleh (Eds.), Statistics and Related Topics, North-Holland, Amsterdam, 1981, pp. 341–352.
[10] M. Ledoux, M. Talagrand, Probability in Banach Spaces: Isoperimetry and Processes, Springer, New York, 1991.
[11] P. Massart, Rates of convergence in the central limit theorem for empirical processes, Ann. Institut H. Poincaré 22 (1986) 381–423.
[12] M. Ossiander, A central limit theorem under metric entropy with L_2 bracketing, Ann. Probab. 15 (1987) 897–919.
[13] G. Pisier, Some applications of the metric entropy condition to harmonic analysis, in: Lecture Notes in Mathematics, Vol. 995, Springer, New York, 1983, pp. 123–154.
[14] D. Pollard, A User's Guide to Measure Theoretic Probability, Cambridge University Press, Cambridge, 2001.
[15] R. Pyke, A uniform central limit theorem for partial-sum processes indexed by sets, in: J.F.C. Kingman, G.E.H. Reuter (Eds.), Probability, Statistics and Analysis, Cambridge University Press, Cambridge, 1983, pp. 219–240.
[16] E. Rio, Covariance inequalities for strongly mixing processes, Ann. Institut H. Poincaré 29 (1993) 587–597.