Probab. Th. Rel. Fields 78, 523–534 (1988)

Probability Theory and Related Fields

The Strong Law of Large Numbers for *k*-Means and Best Possible Nets of Banach Valued Random Variables

J.A. Cuesta and C. Matran

Department of Statistics, University of Santander and Department of Statistics, University of Valladolid

Summary. Let B be a uniformly convex Banach space, X a B-valued random variable and k a given positive integer number. A random sample of X is substituted by the set of k elements which minimizes a criterion. We found conditions to assure that this set converges a.s., as the sample size increases, to the set of k-elements which minimizes the same criterion for X.

1. Introduction

Sometimes we are confronted with situations in which we must choose a simple random variable (r.v.) in instead of a more complexe one (for instance a continuous r.v.). This happens, e.g., when we try to transmit a continuous signal through a channel which only admits a finite number of states.

The general problem is of the following type: Given the r.v. X, at the first stage, we choose a finite set, H, (with a previously given number of elements) and then to each value x, of X, we must associate an element $\Pi(x)$ in H. Both choices must be made in such a way that the mean discrepancy between x and $\Pi(x)$ becomes as short as possible. This process is sometimes called quantization.

To be more precise: Let B be a Banach space, k a fixed positive integer number and X a B-valued r.v. defined on the probability space (Ω, σ, μ). Let $H=\{h_1, \ldots, h_k\} \subset B$ and $\mathscr{S}=\{S_1, \ldots, S_k\}$ a Borel-measurable partition of B. Now we define the map $\Pi_{H;\mathscr{S}}: B \to B$ by

$$\Pi_{H;\mathscr{S}}(x)=\sum_{i=1}^{k} h_i \cdot I_{S_i}(x)$$

(where I_A denotes the indicator set function of A) and the discrepancy between x and $\Pi_{H;\mathscr{S}}(x)$ by $\Phi(\|x-\Pi_{H;\mathscr{S}}(x)\|)$. Where $\Phi: \mathbb{R}^+ \to \mathbb{R}^+$ is a suitable nondecreasing function. We try to select H_0 and $\mathscr{S}_0$ which minimize the mean value of $\Phi(\|x-\Pi_{H;\mathscr{S}}(x)\|)$, i.e.

$$\int \Phi(\|X-\Pi_{H_0;\mathscr{S}_0}(X)\|)\, d\mu = \inf_{H;\mathscr{S}} \int \Phi(\|X-\Pi_{H;\mathscr{S}}(X)\|)\, d\mu \qquad (1)$$

Note that the nondecreasing character of Φ implies that for each set $H = \{h_1, \ldots, h_k\} \subset B$ there exists a best partition $\mathscr{S}_H = \{S_1, \ldots, S_k\}$ with a minor discrepancy given by

$$S_1 = \{x \in B / \|x - h_1\| \leqq \|x - h_j\|; j = 1, \ldots, k\}$$
$$S_i = \{x \in B / \|x - h_i\| \leqq \|x - h_j\|; j = 1, \ldots, k\} - S_{i-1}; \qquad i = 2, \ldots, k.$$

Consequently it is sufficient to consider the set H and we can ignore $\mathscr{S}_H$ in the notation. A set H_0 verifying (1) will be called a k-mean or Φ^k-quantizer, indistinctly.

A study of the quantization problem appears in [8] which is a special issue devoted to this topic.

Moreover, if X is μ-essentially bounded, we can consider a different method of quantization: choose the set $H_0 = \{h_1, \ldots, h_k\}$ which minimizes the maximum discrepancy, i.e. H_0 should be that

$$\text{ess sup} \|X - \Pi_{H_0}(X)\| = \inf_H (\text{ess sup} \|X - \Pi_H(X)\|).$$

This expression coincides with

$$\text{ess sup}(\inf_{h \in H_0} \|X - h\|) = \inf_H [\text{ess sup}(\inf_{h \in H} \|X - h\|)]. \tag{2}$$

Such a set is called a best k-net (see Garkavi [5]) for all cases of k but if $k = 1$ it is called a Chebysev Center (CH-center in shortened form). Best k-nets can be considered as a particular case of k-means despite of the difference between (2) and (1).

In this paper we prove the Strong Law of Large Numbers (SLLN) (i.e. the a.s. consistency) for a wider class of k-means and for best k-nets. The technique employed is that of Cuesta and Matran [2] which relies on the Skorohod Representation Theorem for weak convergence of measures in Polish spaces.

We know some references about the SLLN for k-means for $\mathbb{R}^n$-valued random variables (Pollard [10] and Cuesta [1]) and for r.v. valued on compact metric spaces (Sverdrup-Thygeson [13]). Our results, perhaps with some minor technical modifications, cover both cases.

At present we do not know of any references on the SLLN for best k-nets.

2. Notation and Preliminary Results

In this paper, (Ω, σ, μ) is a probability space, $(B, \|-\|)$ is a uniformly convex Banach space and β is the Borel σ-algebra on B. The term B-valued random variable (r.v.) will be reversed for denoting a strongly measurable B-valued function (see Diestel and Uhl [4], p. 41) defined on some probability space. Let P_X be the probablity measure induced by a B-valued r.v., X, on (B, β). It is well known that there exists a unique smallest closed set in B with P_X-probability one. This set is denoted by $S(X)$ and is called the support of P_X, or even, with an abuse of language, the support of X. Recall that $S(X)$ is separable.

Convergences in B will be denoted by $\xrightarrow{\|-\|}$, or simply $\longrightarrow$ for strong convergence and $\xrightarrow{w}$ for weak convergence. The topological dual space of B is represented by B^*.

A property of uniformly convex Banach spaces is the existence and uniqueness of CH-centers for bounded sets in B: Given the bounded set A in B, there exists a unique x_0 in B such that $\sup_{x\in A}\|x-x_0\|=\inf_{y\in B}(\sup_{x\in A}\|x-y\|)$ (see Holmes [7], p. 187). According to the introduction, for a μ-essentially bounded r.v. X, their CH-center coincides with the CH-center of $S(X)$ (which is, of course, bounded in this case). Note that the CH-center of X is defined as the only element Π_∞ in B that verifies

$$\text{ess sup}\,\|X-\Pi_\infty\|=\min_{h\in B}(\text{ess sup}\,\|X-h\|). \tag{3}$$

If we denote by $L_\infty(B)\equiv L_\infty(\Omega,\sigma,\mu,B)$ to the Lebesgue-Bochner space of essentially bounded B-valued r.v. and by $\|-\|_\infty$ to its usual norm, then from (3) we obtain that Π_∞ is the unique L_∞-metric projection of X onto B, i.e.

$$\|X-\Pi_\infty\|_\infty<\|X-h\|_\infty; \quad h\neq\Pi_\infty.$$

From now on $\Phi\colon \mathbb{R}^+\to\mathbb{R}^+$ will be considered continuous and nondecreasing such that (w.l.o.g.) $\Phi(0)=0$ and a k-mean will be a "k-set" $H=\{h_1,\ldots,h_k\}\subset B$ verifying (1) or, what is the same,

$$\int\Phi(\inf_{h_i\in H}\|X-h_i\|)\,d\mu\leqq\int\Phi(\inf_{g_i\in G}\|X-g_i\|)\,d\mu$$

for all k-set $G=\{g_1,\ldots,g_k\}\subset B$.

If we assume that $\int\Phi(\inf_{g_i\in G}\|X-g_i\|)\,d\mu<\infty$ for some k-set, G, in B it is easy to show, with similar techniques to those which will be employed in the proof of Proposition 2, the existence of k-means. For $k=1$, the existence of 1-means (or best Φ-approximants) was proved by Herrndorf [6] for a large class of Banach spaces and functions Φ.

The minimum values in (1) and (3) will be denoted by $V_\Phi^k(X)$ and $V_\infty^k(X)$ respectively.

The Skorohod Representation Theorem for weak convergence of measures on Polish spaces (see Skorohod [12]) will be the key in all the proofs in this work.

Skorohod's Representation Theorem

Denote by (W,α,l) the probability space where $W=(0,1)$, α consists of the Borel subsets of $(0,1)$ and l is the Lebesgue measure. Let $\{P_n\}_n$ be a sequence of probability measures with separable supports defined in the Borel sets in B. If $P_n \xrightarrow{d} P_0$ (this denotes for weak convergence of measures) then there exists a sequence $Y_0, Y_1, Y_2, \ldots$ of B-valued r.v. defined in (W, α, l) such that

a) $P_{Y_n} = P_n$; $n = 0, 1, 2, \ldots$ and
b) $Y_n \to Y_0$ l-a.s.

3. SLLN for k-Means and Best k-Nets

In this section we prove the SLLN for k-means and best k nets (thus for CH-centers). The proofs will be based on the Skorohod Representation Theorem (and more specifically in Proposition 9 below) adding a "Φ-continuity" similar to that L_p-continuity employed in [2] to prove the SLLN for the p-means. In fact, the results in subsection A tend to establish this Φ-continuity while subsection B is devoted to establishing the SLLN's quickly.

A k-mean (resp. a best k-net) of X exists as long as $\int \Phi(\inf_{g \in G} \|X - g\|)\, d\mu < \infty$ for some k-set G in B (resp. as long as $S(X)$ is bounded, see Singer [11]). In particular this guarantees the existence of sample k-means and sample best k-nets.

Recall that Φ: $\mathbb{R}^+ \to \mathbb{R}^+$ is assumed to be nondecreasing, continuous and such that $\Phi(0) = 0$.

A. Auxiliary Lemmata

Some additional notation for this subsection is the following: $\{Z_n\}_{n=0}^{\infty}$, Z_n: $\Omega \to B$, is a sequence of B-valued r.v.'s for which $V_\Phi^k(Z_n) < \infty$; $n = 0, 1, 2, \ldots$. A k-mean (not necessarily unique) of Z_n will be denoted by $H_n = \{h_1^n, \ldots, h_k^n\}$ and we say that H_n converges (resp. converges weakly) to H_0 when there exists a labeling $h_{i_1}^n, \ldots, h_{i_k}^n$ of the points in H_n; $n = 0, 1, 2, 3, \ldots$ such that $h_{i_j}^n \xrightarrow[n \to \infty]{} h_{i_j}^0$ (resp. weakly) for $j = 1, \ldots, k$.

The following lemma shows that the defined convergences verify the, sometimes called, Urysohn condition. We use it frequently. Note that the proof remains valid for any Banach space.

Lemma 1. *Let $\{H_n\}_{n=0}^{\infty}$ a sequence of k-sets in B. Then H_n converges (resp. weakly) to H_0 as $n \to \infty$ iff every subsequence of $\{H_n\}$ admits a new subsequence which converges (resp. weakly) to H_0.*

Proof. Because the strong convergence, in the mentioned sense, is metrizable (it is equivalent to Haussdorff-distance convergence) we will only consider the case of weak convergence.

The necessary condition obviously holds. Hence we assume that $\{H_n\}$ verifies the above condition for showing $H_n \xrightarrow{w} H_0$.

Let $H_n=\{h_1^n, \ldots, h_k^n\}$; $n=0, 1, 2, \ldots$. From the Hahn-Banach Theorem we deduce the existence of $f_1, \ldots, f_k \in B^*$ such that

$$f_i(h_i^0) \neq f_i(h_j^0) \quad \text{for } j \neq i;\ i=1, \ldots, k. \tag{4}$$

We need to choose a good labeling in each H_n, $n=1, 2, \ldots$: Let $n \geqq 1$ and let $n(1)$ be the smaller i such that

$$|f_1(h_i^n)-f_1(h_1^0)| \leqq |f_1(h_j^n)-f_1(h_1^0)| \quad \text{for } j \neq i.$$

We prove that

$$h_{n(1)}^n \xrightarrow{w} h_1^0.$$

Suppose, on the contrary, that $h_{n(1)}^n \overset{w}{\nrightarrow} h_1^0$. Then there exists $f \in B^*$, $\varepsilon > 0$ and a subsequence which we denote as the original one such that

$$|f(h_{n(1)}^n)-f(h_1^0)| > \varepsilon \quad \text{for all } n. \tag{5}$$

From the hypotheses, there exists a "sub-subsequence" $\{H_{n_r}\}$ such that $H_{n_r} \xrightarrow{w} H_0$, hence for some labeling $h_{n'_r(i)}^{n_r} \xrightarrow{w} h_i^0$; $i=1, \ldots, k$. But $n_r(1)$ coincides infinitely often with $n'_r(i)$ for some i, and so, we can consider a new "sub-subsequence", which we denote also with the same notation, such that $h_{n_r(1)}^{n_r} \xrightarrow{w} h_i^0$ for some i.

Moreover we have

$$|f_1(h_{n_r(1)}^{n_r})-f_1(h_1^0)| \leqq |f_1(h_{n'_r(1)}^{n_r})-f_1(h_1^0)| \xrightarrow[n\to\infty]{} 0.$$

Hence $i=1$ by (4) and $h_{n_r(1)}^{n_r} \xrightarrow{w} h_1^0$ contradicting (5).

Now we define $n(2)$. Let $n(2)$ be the smaller i such that

$$|f_2(h_i^n)-f_2(h_2^0)| \leqq |f_2(h_j^n)-f_2(h_2^0)| \quad \text{for } j \neq i.$$

The same reasoning as above allows us to conclude that $h_{n(2)}^n \xrightarrow{w} h_2^0$. Then (4) implies that there exists a positive integer number N such that for $n \geqq N$: $n(1) \neq n(2)$ and the labeling is correct up to this point.

We define $n(3), \ldots, n(k)$ in a similar way and obtain the lemma.

Proposition 2. *Assume that*

a) $Z_n \to Z_0$ *μ-a.e.*

b) H_0 *is unique*

c) $\int \Phi(\inf_{h\in H_0} \|Z_n-h\|)\, d\mu \to \int \Phi(\inf_{h\in H_0} \|Z_0-h\|)\, d\mu$.

Then

$$H_n \xrightarrow{w} H_0.$$

Proof. First note that:

$$\Phi(\infty) \overset{\text{def}}{=} \lim_{x\to\infty} \Phi(x) \geqq \Phi(y) \quad \text{for all } y \text{ in } \mathbb{R}. \tag{6}$$

Moreover, by definition of $V_\Phi^k(Z_n)$ and c)

$$\varliminf_{n\to\infty} V_\Phi^k(Z_n) \leqq V_\Phi^k(Z_0). \tag{7}$$

Now choose a subsequence of $\{H_n\}$ which we denote as the initial one. We will prove that $I=\{i\in\{1,\dots,k\}/\varliminf_n \|h_i^n\|<\infty\}\neq\emptyset$.

On the contrary, suppose $I=\emptyset$. Since Φ is positive, continuous and increasing, applying Fatou's Lemma and (7), we have:

$$V_\Phi^k(Z_0) \geqq \varliminf_n \int \Phi(\inf_{i=1,\dots,k} \|Z_n-h_i^n\|)\, d\mu$$

$$\geqq \int \varliminf_n \Phi(\inf_{i=1,\dots,k} (\|h_i^n\|-\|Z_n\|))\, d\mu$$

$$= \int \Phi(\inf_{i=1,\dots,k} \{\varliminf_n (\|h_i^n\|-\|Z_n\|)\}) d\mu = \int \Phi(\infty)\, d\mu. \tag{8}$$

And then (6) implies that H_0 is not unique. (In fact it implies that every k-set is a k-mean of Z_0.) To continue, since $I\neq\emptyset$, we can reach a new subsequence (which we also denote as $\{H_n\}$) such that $J=\{i\in\{1,\dots,k\}/\{h_i^n\}$ converges weakly to some $h_i\}$ is not empty and for $i\notin J$: $\|h_i^n\|\to\infty$.

We will prove $J=I=\{1,\dots,k\}$ and $\{h_1,\dots,h_k\}=H_0$.

In fact, from a), for μ-a.e. $\omega\in\Omega$:

$$Z_n(\omega)-h_i^n \xrightarrow{w} Z_0(\omega)-h_i; \qquad i\in J.$$

Then a known property of weak convergence entails

$$\varliminf \|Z_n(\omega)-h_i^n\| \geqq \|Z_0(\omega)-h_i\| \quad \mu\text{-a.e.} \tag{9}$$

and a development similar to that in (8) shows that

$$V_\Phi^k(Z_0) \geqq \int \Phi(\inf_{i=1,\dots,k} (\underline{\lim}_n \|Z_n - h_i^n\|))\, d\mu$$

$$\geqq \int \Phi(\inf_{i\in J} \|Z_0 - h_i\|)\, d\mu. \tag{10}$$

From which the uniqueness of H_0 proves that $J = I = \{1, \dots, k\}$ and $\{h_1, \dots, h_k\} = \{h_1^0, \dots, h_k^0\} \equiv H_0$.

Finally the application of Lemma 1 establishes the proposition.

The following corollaries tend to show three situations in which we can also assure the strong convergence.

Corollary 3. *In the hypotheses of Proposition 2; if $B \equiv \mathbb{R}^n$, then H_n converges to H_0.*

Proof. It follows, trivially, from the fact that in $\mathbb{R}^n$ both types of convergence coincide.

Corollary 4. *In the hypotheses of Proposition 2, if Φ is strictly increasing, then H_n converges to H_0.*

Proof. In Proposition 2 we have proved that $H_n \xrightarrow{w} H_0$. Hence we can write inequalities (9) and (10) for the whole original sequence and for any subsequence:

$$\underline{\lim} \|Z_n(\omega) - h_i^n\| \geqq \|Z_0(\omega) - h_i^0\| \quad \mu\text{-a.e.} \tag{9*}$$

$$V_\Phi^k(Z_0) \geqq \int \Phi(\inf_{i=1,\dots,k} (\underline{\lim}_n \|Z_n - h_i^n\|))\, d\mu$$

$$\geqq \int \Phi(\inf_{i=1,\dots,k} \|Z_0 - h_i^0\|)\, d\mu. \tag{10*}$$

But, by definition of $V_\Phi^k(Z_0)$, (10*) is, really, an equality, so the strictly increasing character of Φ implies that (9*) is also an equality.

From this, for $i = 1, \dots, k$ (and relabeling if necessary), if we choose a subsequence of $\{h_i^n\}$, there exists a sub-subsequence (that we denote with the same notation) such that:

$$Z_n(\omega) - h_i^n \xrightarrow{w} Z_0(\omega) - h_i^0$$

$$\|Z_n(\omega) - h_i^n\| \longrightarrow \|Z_0(\omega) - h_i^0\|$$

and a well known property of the uniformly convex Banach spaces allows us to conclude that this subsequence verifies:

$$Z_n(\omega) - h_i^n \xrightarrow{\|-\|} Z_0(\omega) - h_i^0$$

and then,

$$h_i^n \xrightarrow{\|-\|} h_i^0.$$

Once more, Lemma 1 implies the result.

The following proposition is contained in Lemma 8 in [3].

Proposition 5. *Let Z_0 be a B-valued essentially bounded r.v. defined in (Ω, σ, μ). Let $C\in\sigma$ such that $\mu(C)>0$, S be the support of Z_0 restricted to C and h_0 be the CH-center of S.*

Let $\xi>0$ and denote $V=\operatorname{ess\,sup}\{\|Z_0(\omega)-h_0\|;\ \omega\in C\}$, then there exists $t_0>0$ such that for all $t\leqq t_0$ there exists $\delta>0$ ($\delta\equiv\delta(t)$) such that, if $h\in B$ and $\|h-h_0\|>\xi$, then

$$\mu\{\omega\in C/\|Z_0(\omega)-h\|\geqq V+t\}>\delta.$$

Corollary 6. *In the hypotheses of Proposition 2; if Φ is a convex function, then H_n converges to H_0.*

Proof. Notice that Φ increasing and convex implies the existence of $m\geqq 0$ such that Φ is constant in $[0, m]$ and strictly increasing in (m, ∞).

As the case $m=0$ was considered in the Corollary 4, here we suppose $m>0$.

As occurred in Corollary 4 inequalities (9*) and (10*) are true for the whole sequence and every subsequence.

Let $i\in\{1, \dots, k\}$. We denote $\Sigma_i=\{\omega/\Phi(\|Z_0(\omega)-h_i^0\|)<\Phi(\|Z_0(\omega)-h_j^0\|);\ j\neq i\}$. The uniqueness of H_0 implies that $\mu(\Sigma_i)>0$.

We have two possibilities:

a) $\mu\{\omega\in\Sigma_i/\|Z_0(\omega)-h_i^0\|>m\}>0$.

Φ strictly increasing in (m, ∞) implies that there exists ω_0 in Σ_i such that (9*) is an equality for ω_0. Then, the result follows from the same reasoning as in Corollary 4.

b) $\Sigma_i\subset\{\omega/\|Z_0(\omega)-h_i^0\|\leqq m\}$ μ-a.e.

From the uniqueness of H_0 we obtain that:

$$\begin{aligned}&-\operatorname{ess\,sup}\{\|Z_0(\omega)-h_i^0\|;\ \omega\in\Sigma_i\}=m\\ &-h_i^0 \text{ is the CH-center of } S(Z_0)\cap Z_0(\Sigma_i).\end{aligned}\tag{11}$$

Let us suppose that $\|h_i^n-h_i^0\|\nrightarrow 0$.

Then (w.l.o.g.) we can write that there exists $\xi>0$ such that for every $n\in\mathbb{N}$: $\|h_i^n-h_i^0\|>\xi$.

Lemma 5 implies that there exists $\delta>0$ and $t_1>0$ such that if

$$\alpha_n=\{\omega\in\Sigma_i/\|Z_0(\omega)-h_i^n\|>m+t_1\}$$

then $\mu(\alpha_n)>\delta$ for every n in $\mathbb{N}$.

From the definition of Σ_i we deduce that there exists $t_2>0$ such that:

$$\mu\{\omega\in\Sigma_i/\|Z_0(\omega)-h_j^0\|>m+t_2;\ j\neq i\}>\mu(\Sigma_i)-\delta/2.$$

We denote by Σ_i^* to this set.

W.l.o.g. we can consider $t_1=t_2=t$ and $\delta/2<\mu(\Sigma_i)$.

From (9*) and the definition of Σ_i^* we obtain that for μ-a.e. ω in Σ_i^* there exists $N(\omega)$ such that for every $n\geqq N(\omega)$:

$$\|Z_n(\omega)-h_j^n\|>m+t,\qquad j\neq i.$$

Now let $n \in \mathbb{N}$ and define $A_n = \{\omega \in \Sigma_i^* / N(\omega) \leqq n\}$. As $A_n \uparrow \Sigma_i^*$ μ-a.e., there exists $N_1 \in \mathbb{N}$ such that $\mu(A_{N_1}) > \mu(\Sigma_i) - \delta/2$ and, from the definition of A_{N_1} ·, we have

$$\text{for all } \omega \in A_{N_1};\ \forall n \geqq N_1:\ \|Z_n(\omega) - h_j^n\| > m + t; \quad \text{if } j \neq i.$$

On the other hand applying Egoroff's Theorem we obtain a set C in σ with $\mu(C) > 1 - \delta/2$ and $\|Z_n - Z_0\| \to 0$ uniformly in C. I.e. there exists $N_2 \in \mathbb{N}$ such that if $n \geqq \mathbb{N}_2$, $\|Z_n(\omega) - Z_0(\omega)\| < t/2$ for every ω in C.

Let $N = \sup(N_1, N_2)$; $n \geqq N$ and $\omega \in \alpha_n \cap A_{N_1} \cap C$. From the definition of these sets

$$\inf_{i=1,\dots,k} \|Z_n(\omega) - h_j^n\| > m + t/2. \tag{12}$$

Finally the same reasoning as in (8) leads us to the following inequalities:

$$\begin{aligned} V_\Phi^k(Z_0) &\geqq \varliminf_n V_\Phi^k(Z_n) \\ &\geqq \varliminf_n \int_{\Sigma_i^c} \inf_j \Phi(\|Z_n - h_j^n\|)\, d\mu + \varliminf_n \int_{\Sigma_i} \inf_j \Phi(\|Z_n - h_j^n\|)\, d\mu \\ &\geqq \int_{\Sigma_i^c} \inf_j \Phi(\|Z_o - h_j^0\|)\, d\mu + \varliminf_n \int_{\alpha_n \cap A_{N_1} \cap C} \inf_j \Phi(\|Z_n - h_j^n\|)\, d\mu. \end{aligned}$$

Remember that Φ is equal to 0 in $[0, m]$ and strictly increasing in (m, ∞). Then from (11) and (12) we can continue in the following way:

$$\begin{aligned} V_\Phi^k(Z_0) &\geqq \int_{\Sigma_i^c} \inf_j \Phi(\|Z_0 - h_j^0\|)\, d\mu + \varliminf_n \Phi(m + t/2) \cdot \mu(\alpha_n \cap A_{N_1} \cap C) \\ &> \int_{\Sigma_i^c} \inf_j \Phi(\|Z_0 - h_j^0\|)\, d\mu + \int_{\Sigma_i} \Phi(\|Z_0 - h_i^0\|)\, d\mu = V_\Phi^k(Z_0) \end{aligned}$$

which is impossible. Then $\|h_i^n - h_i^0\| \xrightarrow[n \to \infty]{} 0$.

We end this subsection with a proposition devoted to best k-nets. It might be interesting to emphasize that this proposition is obtained trivially from Corollary 6 because the best k-nets are particular cases of k-means.

Lemma 7. *Let Z_0 be essentially bounded and such that there exists a unique best k-net for Z_0: $H_0 = \{h_1^0, \dots, h_k^0\}$. Let*

$$B_i = \{x \in S(Z_0) / \|x - h_i^0\| < \|x - h_j^0\|;\ j \neq i\}; \quad i = 1, \dots, k.$$

Then h_i^0 is the CH-center of B_i and

$$V_\infty^k(Z_0) = \operatorname{ess\,sup}\{\|x - h_i^0\|;\ x \in B_i\}; \quad i = 1, \dots, k.$$

Proof. Both statements are obtained easily from the definition of V_∞^k and the uniqueness of H_0 by a standard reasoning.

Proposition 8. *Let Z_0 be essentially bounded and such that it admits a unique best k-net: H_0. Suppose that:*

a) *for every n: $S(Z_n) \subset S(Z_0)$.*

b) *$Z_n \to Z_0$ μ-a.e.*

Let H_n a best k-net of Z_n; $n=1,2,\ldots$. Then:

$$H_n \xrightarrow{\|-\|} H_0 .$$

Proof. Let Φ be a convex function such that $\Phi(x)=0$ if $x \leqq V_\infty^k(Z_0)$ and Φ is strictly increasing in the interval $(V_\infty^k(Z_0); \infty)$.

Lemma 7 and the uniqueness of H_0 imply that H_0 is the only k-mean (for Φ) of Z_0.

a) implies that H_n is a k-net (for Φ) for Z_n; $n=1, 2, \ldots$.

Then the result follows from Corollary 6.

Note that the hypotheses of Proposition 8 do not imply the convergence in the L_∞-norm of Z_n to Z_0 but they are sufficient to obtain the continuity of the best L_∞-metric-approximation.

B. SLLN Theorems

As we have already mentioned, the propositions in the preceeding subsection allow us to obtain quickly (with the help, of course, of the Skorohod Theorem) the SLLN both for k-means and best k-nets.

As in subsection A we begin by presenting the peculiar notation for this subsection.

From now on we consider a fixed B-valued r.v., X_0, with a probability distribution P_0 and a sequence $\{X_n\}_n$ of independent B-valued r.v. with the same distribution as X_0 (i.e. $P_{X_n}=P_{X_0}=P_0$). Also, w.l.o.g. we suppose that $X_0; X_1; X_2; \ldots$ are defined in (Ω, σ, μ) and we will use P_n^ω to denote the empirical probability distribution (i.e. P_n^ω gives mass $1/n$ to each $X_i(\omega)$, $i=1, \ldots, n$).

It is well known that $\mu\{\omega/P_n^\omega \xrightarrow{d} P_0\}=1$ (see Parthasarathy [9], p. 53). Therefore, we have the following consequence of the Skorohod Representation Theorem:

Proposition 9. *For μ-a.e. $\omega \in \Omega$ there exists a sequence $Y_0^\omega, Y_1^\omega, Y_2^\omega, \ldots$ of B-valued r.v. defined in (W, α, l) such that*

a) $P_{Y_0^\omega}=P_0$; $P_{Y_n^\omega}=P_n$; $n=1, 2, \ldots$.

b) $Y_n^\omega \to Y_0^\omega$ *l-a.s.*

The set of μ-probability one which this proposition gives, will be denoted in the sequel by Ω^*.

Note that the k-means and the best k-nets of a random variable only depend on its probability distribution. Then the k-means (resp. the best k-nets) of Y_0^ω

coincide with those of X_0. Moreover, if $\omega \in \Omega^*$, we can identify the empirical k-means (or the best k-nets) with the corresponding ones for the random variables which we have obtained in Proposition 9 and we have the following theorem:

Theorem 10. *Let $\{H_n^\omega\}$ be a sequence of empirical k-means and suppose that X_0 admits a unique k-mean: H_0. Then*

$$H_n^\omega \xrightarrow{w} H_0 \qquad \text{for } \mu\text{-a.e. } \omega \in \Omega.$$

Proof. We obtain this theorem from Proposition 2 applied to the sequence $\{Y_n^\omega\}_n$ and to the r.v. Y_0^ω.

In fact, as we have noted previously, for every $\omega \in \Omega^*$; Y_0^ω admits a unique k-mean which is H_0. Then for any ω in Ω^*, Y_0^ω verifies the hypothesis b) in Proposition 2.

The SLLN for real-valued r.v. permit us to conclude that

$$\int \Phi(\inf_{h \in H_0} \|Y_n^\omega - h\|)\, d\mu \to \int \Phi(\inf_{h \in H_0} \|Y_0^\omega - h\|)\, d\mu$$

for μ-a.e. ω in Ω^*.

From Proposition 9, a) is also true.

Then Proposition 2 implies the result.

Corollary 11 (Strong Law of Large Numbers for k-means). *In the hypotheses of the preceeding theorem, if $B = \mathbb{R}^n$ or Φ is strictly increasing or Φ is convex, then*

$$H_n \xrightarrow{\|\,-\,\|} H_0 \qquad \text{for } \mu\text{-a.e. } \omega \in \Omega.$$

Proof. This corollary is obtained from Corollaries 4–6 with a similar reasoning to that in Theorem 10.

Theorem 12 (Strong Law of Large Numbers for best k-nets). *Let us suppose that X_0 is essentially bounded and $\{H_n\}$ be a sequence of empirical k-nets. If X_0 admits a unique best k-net, H_0, then:*

$$H_n \xrightarrow{\|\,-\,\|} H_0 \qquad \text{for } \mu\text{-a.e. } \omega \in \Omega.$$

Proof. The support of P_n^ω is contained, for all $n \in \mathbb{N}$, *in* $S(X_0)$ for μ-a.e. $\omega \in \Omega$. Then this theorem can be proved analogously to the preceding ones.

We end this paper with an example in which H_0 is not unique and the sequence of empirical k-nets (which moreover are unique μ-a.e.) does not converge for μ-a.e. $\omega \in \Omega$. This example shows that to obtain SLLN for k-nets without uniqueness in the theoretical best k-net requires some other additional hypotheses. Furthermore, as the best k-nets are particular k-means, this example shows that this is also true for k-means.

Example. Let $k=2$ and X_0 be a real-valued r.v. such that

$$P_{X_0}(A)=P_0(A)=\int_{A\cap([0,2]\cup[3,4])} 1/6\,dx+\tfrac{1}{2}\cdot I_A(2).$$

Let $\{X_n\}$ be a sequence of independent, identically distributed r.v. whose distribution is that of X_0 and $\{F_n^\omega\}$ be the sequence of empirical distributions functions. Let

$$M_n=\sup(X_1,\dots,X_n) \quad \text{and} \quad m_n=\inf(X_1,\dots,X_n); \qquad n=1,2,\dots.$$

Let $\omega\in\Omega$. It is easy to prove that F_n^ω admits a unique best k-net unless if there $i\leqq n$ such that $X_i(\omega)=\frac{1}{2}(M_n(\omega)+m_n(\omega))$; but this does not happens for μ-a.e. ω in Ω, from certain index onwards. Let $H_n^\omega=\{h_1^n(\omega),h_2^n(\omega)\}$ a best 2-net for F_n^ω, $n=1,2,\dots$ and suppose that $h_1^n\leqq h_2^n$. It is not too difficult to obtain the following: for μ-a.e. ω in Ω, $\{h_1^n\}$ converges to 1 and $\{h_2^n\}$ has two cluster points: 3 and 3.5.

References

1. Cuesta, J.A.: Medidas de centralización multidimensionales (ley fuerte de los grandes números). Trab. Estad. e Invest. Oper. **35**, 3–16 (1984)
2. Cuesta, J.A., Matran, C.: Strong laws of large numbers in abstract spaces via Skorohod's Representation Theorem. Sankhyà, Ser. A, **84**, 98–103 (1986)
3. Cuesta, J.A., Matran, C.: Uniform consistency of r-means. Stat. Probab. Lett. to appear
4. Diestel, J., Uhl, JJ.: Vector measures. Providence: American Math Society (1977)
5. Garkavi, A.L.: The best possible net and the best possible cross-selection of a set in a normed space. Izv. Akad. Nauk. SSSR Ser. Mat. **26**, 87–106 (1962). (English translation in: Am. Math. Soc. Transl. (2), **39**, 111–132 (1964))
6. Herrndorf, N.: Approximation of vector-valued random variables by constants. J. Approximation Theory **37**, 175–181 (1983)
7. Holmes, R.B.: A course on optimization and best approximation. Berlin Heidelberg New York: Springer 1972
8. IEEE Trans. Inform. Theory **28** (1982)
9. Parthasarathy, K.R.: Probability measures on metric spaces. New York: Academic Press 1967
10. Pollard, D.: Strong consistency of k-means clustering. Ann. Statist. **9**, 135–140 (1981)
11. Singer, I.: Best approximation in normed linear spaces by elements of linear subspaces. New York Heidelberg Berlin: Springer 1970
12. Skorohod, A.V.: Limit theorems for stochastic processes. Theor. Probab. Appl. **1**, 261–290 (1956)
13. Sverdrup-Thygeson, H.: Strong law of large numbers for measures of central tendency and dispersion in compact metric spaces. Ann. Statist. **9**, 141–145 (1981)

Received May 13, 1985; received in revised form December 29, 1987