

# ANNALES DE L'I. H. P., SECTION B

M. M. RAO

## **Prediction sequences in smooth Banach spaces**

*Annales de l'I. H. P., section B*, tome 8, n° 4 (1972), p. 319-332

[http://www.numdam.org/item?id=AIHPB\\_1972\\_\\_8\\_4\\_319\\_0](http://www.numdam.org/item?id=AIHPB_1972__8_4_319_0)

© Gauthier-Villars, 1972, tous droits réservés.

L'accès aux archives de la revue « Annales de l'I. H. P., section B » (<http://www.elsevier.com/locate/anihpb>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme  
Numérisation de documents anciens mathématiques

<http://www.numdam.org/>

## **Prediction sequences in smooth Banach spaces (\*)**

by

**M. M. RAO**

The Institute for Advanced Study,  
Princeton, New Jersey, 08540.

---

**SUMMARY.** — In this paper, first a general result on the norm convergence of prediction sequences in abstract Banach spaces satisfying some smoothness conditions is presented. Smoothness here refers to the Fréchet differentiability of norm. If the Banach space is a general function space on a measure (or probability) space satisfying conditions implying those of the first theorem, then the pointwise a. e. convergence of the above prediction sequence is also established. Some remarks and illustrations are included to clarify the generality of the results, and its relations with other work on the problem.

---

### **0. INTRODUCTION**

The purpose of this paper is to present two convergence theorems for prediction sequences in general Banach spaces satisfying certain smoothness conditions. The first result involves the strong (or norm) convergence and the second, when the Banach spaces are function spaces on a measure space (i. e. normed Riesz spaces in the sense of W. A. J. Luxemburg and

---

(\*) This work was completed, with a partial support of the NSF grant GP-8777, in the Fall of 1968 when the author was visiting the Indian Statistical Institute, Calcutta; and the final version, with minor revisions, is prepared at The Institute for Advanced Study, Princeton, N. J.

A. C. Zaanen), relates to the pointwise a. e. convergence. Known work on this problem, in  $L^p$ -spaces and in most of the  $L^\Phi$ -spaces of Orlicz, is subsumed here. Moreover, the abstract treatment clarifies the structure, the existence and convergence of (non-linear) prediction and approximation sequence, and the work has independent interest.

Thus the next section contains the notation and preliminaries to be used later in the abstract formulation of the problem. Then the mean and pointwise convergence are presented in sections 2 and 3, respectively. Some illustrative examples and remarks clarifying the generality, and the related computational problems awaiting solution, are given in the last section. Hopefully, this work also brings out the linear and non-linear prediction into focus.

## 1. PRELIMINARIES

If  $(\mathcal{X}, \|\cdot\|)$  is a Banach (or B-) space, the following properties, when present, will be of interest in the present work (notation as in [4] [5]):

- (K): If  $K \subset \mathcal{X}$  is convex and  $U$  is the closed unit ball of  $\mathcal{X}$ , then the diameter of  $(K \cap tU)$  tends to zero as  $t$  tends to the distance from  $K$  to the origin of  $\mathcal{X}$ .
- (D): If  $f \in S^*$ , the unit sphere of the adjoint space  $\mathcal{X}^*$  of  $\mathcal{X}$ , and  $E(f, \delta) = \{x: f(x) \geq 1-\delta\} \cap U$ , then the diameter of  $E(f, \delta)$  tends to zero as  $\delta \searrow 0$ .
- (E): For any convex set  $K \subset \mathcal{X}$ , every sequence  $\{x_n\} \subset K$  such that  $\lim_n \|x_n\| = \inf \{\|x\|: x \in K\}$  has the property that  $\{x_n\}$  is a Cauchy sequence in  $\mathcal{X}$ .
- (F):  $\mathcal{X}$  has the Fréchet differentiable norm, i. e., if  $x \in S$ , the unit sphere of  $\mathcal{X}$ , then  $G(x, y) = \lim_{t \rightarrow 0} [(\|x + ty\| - 1)/t]$  exists uniformly in  $y \in S$ . [This is the « smoothness property » of  $\mathcal{X}$ .]
- (H): If  $\mathcal{X}$  is rotund (= strictly convex),  $\{x_n\} \subset \mathcal{X}$  converges weakly to  $x$ , and  $\|x_n\| \rightarrow \|x\|$  as  $n \rightarrow \infty$ , then  $\|x_n - x\| \rightarrow 0$ .

If  $\mathcal{X}$  is a Hilbert space, it has clearly all these properties. It is known more generally that if  $\mathcal{X}$  is uniformly rotund and smooth, then also all these properties are true for  $\mathcal{X}$ , by [14] (see also [4]). The interrelations between these properties will become clear from the work below.

If  $M \subset \mathcal{X}^*$  is a total (or determining) linear subspace, then the  $M$ -topology of  $\mathcal{X}$ , denoted by  $\sigma(\mathcal{X}, M)$ , is defined by the neighborhood system:

$$N(x: f_1, \dots, f_n, \varepsilon) = \{y \in \mathcal{X}: |f_i(x) - f_i(y)| < \varepsilon, f_i \in M, i = 1, \dots, n\}$$

where  $\varepsilon > 0, x \in \mathcal{X}$ . It is clear that  $\sigma(\mathcal{X}, \mathbf{M})$  is a locally convex topology for  $\mathcal{X}$ . If  $\mathbf{M} = \mathcal{X}^*$ , then the latter is the usual weak topology of  $\mathcal{X}$ . If in the property (H) above, the weak convergence is replaced by the  $\sigma(\mathcal{X}, \mathbf{M})$ -convergence, and the conclusion still holds, then the resulting property will be denoted by  $(H_{\mathbf{M}})$ , in what follows.

The abstract prediction and approximation problems can be stated as: if  $C \subset \mathcal{X}$  is a closed convex set (determined usually by the « observations ») and  $x_0 \in \mathcal{X}$  (a point to be predicted), then find an  $\tilde{x} \in C$ , if it exists, such that  $\|x_0 - \tilde{x}\| = \inf \{ \|x_0 - x\| : x \in C \}$ . When it exists,  $\tilde{x}$  is called a *predictor* (or *best approximation*) of  $x_0$  based on  $C$  relative to the norm  $\|\cdot\|$  of  $\mathcal{X}$ . If  $C_n \subset C_{n+1} \subset \mathcal{X}$ , is a filtering sequence of sets as above, and  $\tilde{x}_n \in C_n$  exists as a predictor of  $x_0$ , under what conditions does  $\tilde{x}_n \rightarrow \tilde{x}_\infty \in C_\infty$ , the closed convex hull of  $\bigcup_n C_n$ , in the norm of  $\mathcal{X}$ ? If  $\mathcal{X}$  is moreover a (normed) Riesz space, when can one assert also the pointwise a. e. convergence of this prediction sequence? These are important questions in the prediction (and approximation) theory. Solutions to these problems, under reasonably general conditions, are presented in the next two sections. Relations between this work and that of the previously known cases is discussed in the last section.

## 2. STRONG CONVERGENCE

The main result of this section is the following theorem which gives the existence, uniqueness and norm convergence of predictors in B-spaces satisfying some of the conditions of the above-section.

**THEOREM 2.1.** — *Let  $(\mathcal{X}, \|\cdot\|)$  be a B-space and  $\mathbf{M} \subset \mathcal{X}^*$  be a norm determining (or total) subspace for  $\mathcal{X}$ . Let  $C_n \subset C_{n+1} \subset \mathcal{X}$  be (norm) closed convex sets and  $C_\infty$  be the closed convex hull of  $\bigcup_n C_n$ . Suppose that : (i) every bounded sequence in  $C_\infty$  is relatively sequentially  $\sigma(\mathcal{X}, \mathbf{M})$ -compact, and (ii)  $\mathcal{X}$  has the property  $(H_{\mathbf{M}})$ . Then, for any  $x_0 \in \mathcal{X}$ , (a) there exists uniquely predictors  $x_n \in C_n$  of  $x_0, 1 \leq n \leq \infty$ , and (b)  $\|x_n - x_\infty\| \rightarrow 0$  as  $n \rightarrow \infty$ . The conditions (i) and (ii) are satisfied (and hence the conclusions (a) and (b) hold) in each of the following « smooth » B-spaces (with  $\mathbf{M} = \mathcal{X}^*$ ) :*

1.  $\mathcal{X}$  is a B-space such that its adjoint  $\mathcal{X}^*$  has (F).

2.  $\mathcal{X}$  is a weakly sequentially complete B-space such that  $\mathcal{X}$  has (H) and  $\mathcal{X}^*$  is separable.

3.  $\mathcal{X}$  is a B-space with property (E).

Also each of the cases 1-3. above implies the reflexivity of  $\mathcal{X}$ .

*Remark.* — The hypothesis of the first part of the theorem is strictly weaker than that of 1. to 3. As noted in the last section, there exist certain non-reflexive Orlicz spaces satisfying the hypothesis of the first part. Since even  $C_1$  can span an infinite dimensional subspace, the prediction theory has a more general view point than the classical approximation theory (cf. [2], for the latter).

*Proof.* — For convenience, the proof will be presented in steps.

I. Let  $C \subset \mathcal{X}$  be a closed convex set satisfying (i) and (ii) of the theorem. Then for any  $x_0 \in \mathcal{X}$ , there is a unique  $y_0 \in C$  such that

$$\|x_0 - y_0\| = \inf \{ \|x_0 - y\| : y \in C \}.$$

For, let  $F(x) = \|x - x_0\|$ ,  $x \in C$ . If  $x_0 \in C$ , then take  $y_0 = x_0$ . If  $x_0 \in \mathcal{X} - C$ , then  $\inf \{ F(y) : y \in C \} = d > 0$  since  $C$  is closed. Let  $\{y_n\} \subset C$  be such that  $d = \lim_n F(y_n)$ . Then  $\{y_n\}$  is bounded and, by (i), it is relatively sequentially  $\sigma(\mathcal{X}, \mathbf{M})$ -compact. Since  $C$  is also  $\sigma(\mathcal{X}, \mathbf{M})$ -closed and hence the convex closure of the sequentially compact set  $\{y_n\}$  is complete in this topology, there exists a convergent subsequence  $\{y_{n_i}\}$  and a  $y_0 \in C$  such that  $f(y_{n_i}) \rightarrow f(y_0)$  for all  $f \in \mathbf{M}$ . Hence  $f(x_0 - y_{n_i}) \rightarrow f(x_0 - y_0)$ ,  $f \in \mathbf{M}$ , and so

$$|f(x_0 - y_0)| \leq \liminf_i |f(x_0 - y_{n_i})| \leq \liminf_i \|f\| \|x_0 - y_{n_i}\| = d \|f\|, \quad f \in \mathbf{M}. \quad (1)$$

Since  $\mathbf{M}$  is a determining manifold for  $\mathcal{X}$ , so that the supremum on the left side of (1) for  $\|f\| \leq 1$  gives  $\|x_0 - y_0\|$ , one has

$$0 < d \leq \|x_0 - y_0\| \leq d. \quad (2)$$

Thus  $y_0$  is a minimal element in  $C$ . By (ii),  $\mathcal{X}$  is rotund and hence so is  $C$ . Consequently, the minimum element in  $C$  is also unique.

II. The (unique) minimal elements  $x_n$  in  $C_n$ ,  $1 \leq n \leq \infty$ , (i. e.  $\|x_0 - x_n\|$  is minimum over  $C_n$ ), exist and satisfy:  $\|x_n\| \rightarrow \|x_\infty\|$  as  $n \rightarrow \infty$ , where  $C_n$  are as given in the theorem.

For, since  $C_n \uparrow C_\infty$  is a closed subset, it follows that for each  $n$ ,  $C_n$  satisfies (i) and (ii) since  $C_\infty$  does. So by I, the stated  $x_n \in C_n$  exists. Let  $d_n = F(x_n) = \|x_0 - x_n\|$ . Then  $d_n \geq d_{n+1}$  so that  $d_n \rightarrow d' \geq d = F(x_\infty)$ .

To see that there is equality here, let  $\varepsilon > 0$ . Since  $x_\infty \in C_\infty$ , the closed convex hull of  $\bigcup_n C_n$ , there is an  $n_\varepsilon$  such that  $n \geq n_\varepsilon$  implies the existence of  $x_\varepsilon \in C_n$  with  $\|x_\infty - x_\varepsilon\| < \varepsilon$ . But then

$$d \leq d' \leq d_n = F(x_n) \leq \|x_0 - x_\varepsilon\| \leq F(x_\infty) + \|x_\infty - x_\varepsilon\| < d + \varepsilon. \quad (3)$$

The arbitrariness of  $\varepsilon > 0$  implies the equality:  $d = d'$ .

III. The sequence  $\{x_n, 1 \leq n \leq \infty\}$  of Step II is  $\sigma(\mathcal{X}, M)$ -convergent.

For, since  $\{x_n\} \subset C_\infty$  is clearly bounded, it is relatively  $\sigma(\mathcal{X}, M)$ -sequentially compact. Hence, as in Step I, there is a  $y_\infty \in C_\infty$  and a subsequence  $\{x_{n_i}\} \subset \{x_n\}$  such that  $f(x_{n_i}) \rightarrow f(y_\infty)$  for all  $f \in M$ , so that (cf. (1)),

$$d \leq \|x_0 - y_\infty\| \leq \liminf_i \|x_0 - x_{n_i}\| = \liminf_i d_{n_i} = d, \quad (4)$$

by Step II. By the rotundity of  $C_\infty$ ,  $y_\infty = x_\infty$ . Repeating the argument for every infinite subsequence, one concludes that each subsequence has the same limit  $x_\infty$ , so that the whole sequence converges to  $x_\infty$  in the  $\sigma(\mathcal{X}, M)$ -topology, since the latter is Hausdorff.

IV. The sequence  $\{x_n, 1 \leq n \leq \infty\}$  above satisfies:  $\|x_n - x_\infty\| \rightarrow 0$ , as  $n \rightarrow \infty$ .

For, the preceding two steps show that  $x_n \rightarrow x_\infty$  in  $\sigma(\mathcal{X}, M)$  and  $\|x_n\| \rightarrow \|x_\infty\|$ , as  $n \rightarrow \infty$ . By (ii) of the hypothesis,  $\mathcal{X}$  has  $(H_M)$  so that, by its definition, these two facts imply that  $\|x_n - x_\infty\| \rightarrow 0$ , as  $n \rightarrow \infty$ . This proves the first (main) part of the theorem. It remains to establish 1.-3.

V. *Proof of 1.* By hypothesis  $\mathcal{X}$  has (F) so that by a theorem of Šmulian [14],  $\mathcal{X}$  has (D) (cf. definitions in Section 1). However, for any B-space, (D) holds if and only if (K) holds and the latter implies reflexivity of the space (cf. [5], p. 113). Moreover (D) implies (H) for any B-space (cf. [4], p. 90). This and the fact that, in a reflexive space, bounded sets are relatively weakly sequentially compact imply (i) and (ii) of the hypothesis with  $M = \mathcal{X}^*$ .

*Proof of 2.* Since  $\mathcal{X}$  has property (H) by hypothesis, conditions (i) and (ii) hold if  $\mathcal{X}$  is shown to be reflexive under the present hypothesis (and  $M = \mathcal{X}^*$ ). The reflexivity is proved as follows.

Let  $A \subset \mathcal{X}$  be a bounded set. If  $\{x_n\}_1^\infty \subset A$  is any sequence, and  $\mathcal{Y}$  is the closed linear span of  $\{x_n\}$ , then its adjoint  $\mathcal{Y}^* \subset \mathcal{X}^*$  is separable since  $\mathcal{X}^*$  is separable by hypothesis. Hence there is a dense denumerable set  $\{y_n^*\} \subset \mathcal{Y}^*$ . Then, by the diagonal process, one may select a convergent subsequence  $\{y^*(x_{n_i})\}_1^\infty$  for all  $y^* \in \{y_n^*\}_1^\infty$ . But by the density of

the latter set in  $\mathcal{Y}^*$ , this implies that  $\{y^*(x_{n_i})\}_1^\infty$  is a Cauchy sequence, for all  $y^* \in \mathcal{Y}^*$ . Since any  $x^* \in \mathcal{X}^*$  restricted to  $\mathcal{Y}^*$  determines a  $y^* \in \mathcal{Y}^*$ , the preceding statement implies that  $\{x_{n_i}\}_1^\infty$  is a weak Cauchy sequence in  $\mathcal{X}$  also. Finally by the weak completeness of  $\mathcal{X}$ , there exists an  $\bar{x} \in \mathcal{X}$  with  $x^*(x_{n_i} - \bar{x}) \rightarrow 0$ , as  $i \rightarrow \infty$ ,  $x^* \in \mathcal{X}^*$ . Since the  $\{x_n\}_1^\infty$  is an arbitrary sequence in  $\mathcal{A}$ , this implies that  $\mathcal{A}$  is relatively weakly sequentially compact and hence, by the Eberlein-Šmulian theorem ([7], p. 430),  $\mathcal{A}$  is relatively weakly compact. Taking  $\mathcal{A}$  as the unit ball of  $\mathcal{X}$ , it follows that  $\mathcal{X}$  is reflexive, by ([7], V. 4.7). Thus 2. is proved.

*Proof of 3.* If  $\mathcal{X}$  has (E), then, by ([8], Th. 3),  $\mathcal{X}$  is reflexive and has (H). Since (F) and (H) are dual in reflexive spaces (cf. [4], p. 90), this means  $\mathcal{X}^*$  has (F) and the hypothesis of 1. is satisfied, and so the conclusion follows. Note that each case implies reflexivity of  $\mathcal{X}$ .

This completes the proof of the theorem.

*Alternate proof of 2.* The following interesting argument has been pointed out to me by Professor C. Foiaş, after listening to a talk on this material in Bucarest in May 1969. Since  $\mathcal{X}^*$  is separable, the  $\mathcal{X}^*$  (or weak)-topology of the closed unit sphere  $S$  of  $\mathcal{X}$  is a metric topology by ([7], V. 5.2), and only then. Also the weak sequential completeness of  $\mathcal{X}$  implies that  $S$  is then a complete metric space. By ([7], V. 5.1), since  $\mathcal{X}^*$  is separable, the  $\mathcal{X}^*$  (or weak-star)-topology of  $S^{**}$ , the closed unit sphere of  $\mathcal{X}^{**}$ , is a metric topology. But  $S^{**}$  is also compact in the same topology and hence, being a compact metric space, it is also complete. But the natural embedding of  $S$  in  $S^{**}$  is  $\mathcal{X}^*$ -dense by ([7], V. 4.6). Since  $S$  and  $S^{**}$  are both complete metric spaces, it follows that  $S = S^{**}$  under the natural embedding. But this means that  $\mathcal{X}(= \mathcal{X}^{**})$  is reflexive. Since  $\mathcal{X}$  has (H), the result follows.

It is of interest to remark that the reflexivity of  $\mathcal{X}$  in 2. is false if  $\mathcal{X}$  is not weakly complete. For instance, take  $\mathcal{X} = c_0$ , the space of real sequences converging to zero, so that  $\mathcal{X}^* = l_1$ , the space of summable sequences, which is separable but not reflexive. Thus the following conclusion can be stated for reference.

**PROPOSITION 2.2.** — *A Banach space with a separable dual is reflexive if and only if it is weakly sequentially complete.*

The preceding special cases yield some further implications extending the tables of [4] and [8]:

**PROPOSITION 2.3.** — *For a Banach space  $\mathcal{X}$ , the following statements are equivalent: (i)  $\mathcal{X}$  has (D), (ii)  $\mathcal{X}$  has (K), (iii)  $\mathcal{X}$  has (E) (iv)  $\mathcal{X}$  is reflexive*

and has (H), (v) the adjoint space  $\mathcal{X}^*$  has (F). If, moreover,  $\mathcal{X}$  is separable, then each of the above is equivalent to: (vi)  $\mathcal{X}^*$  is separable and  $\mathcal{X}$  is weakly sequentially complete and has (H).

The proof of this proposition uses a few more properties of metric geometry (cf. [4] [5]), and is similar to the proof of 1., of Theorem 2.1, and will be omitted. This result shows that the assumptions (i) and (ii) of the theorem hold also in the cases (i) (ii) and (iv) of the above proposition.

The next section will be devoted to the pointwise convergence of the sequence  $\{x_n, 1 \leq n \leq \infty\}$  if  $\mathcal{X}$  is a (normed) Riesz space on a measure space.

### 3. POINTWISE CONVERGENCE

If  $\mathcal{X}$  is a Banach function space on a measure space  $(\Omega, \Sigma, P)$ , then it makes sense to ask about the pointwise a. e. convergence of the prediction sequence in addition to norm convergence. The normed Riesz spaces considered in [10] appear to be the most general spaces, which include the Orlicz (hence Lebesgue) spaces, suitable for this purpose. If  $\mathcal{J}$  is the class of measurable scalar functions on  $(\Omega, \Sigma, P)$ , then  $\rho: \mathcal{J} \mapsto \bar{\mathbb{R}}^+$  is said to be a function norm if (i)  $\rho(af) = |a| \rho(f) = |a| \rho(|f|)$ , (ii)  $\rho(f_1 + f_2) \leq \rho(f_1) + \rho(f_2)$  and (iii)  $\rho(f) = 0$  if and only if  $f = 0$ , a. e., where, in (i),  $a$  is any scalar. Let  $L^\rho(\Sigma)$  denote the subspace of  $\mathcal{J}$  such that  $f \in L^\rho(\Sigma)$  whenever  $\rho(f) < \infty$ . When functions a. e. equal are identified and « function » is used to mean a member of its equivalence class then, with the usual abuse of notation,  $L^\rho(\Sigma)$  becomes a normed linear space and is complete when  $\rho$  satisfies the triangle inequality for infinite sums in  $L^\rho$ , called the *Riesz-Fischer property*. A stronger condition sufficient for the present purposes is the *Fatou property*:  $0 \leq f_n \uparrow f$  a. e., implies  $\rho(f_n) \uparrow \rho(f)$ . Then  $L^\rho(\Sigma)$  is a B-space. See ([15], Ch. 15), for a good treatment of  $L^\rho(\Sigma)$  spaces where they are called normed Köthe spaces. [In a forthcoming monograph, by W. A. J. Luxemburg and A. C. Zaanen, containing an extensive account of these spaces, they are renamed, as in [10], Riesz spaces.] Thus the Riesz space  $L^\rho$  is complete.

Let  $\rho': \mathcal{J} \mapsto \bar{\mathbb{R}}^+$ , called the *associate norm* of  $\rho$ , be defined by

$$\rho'(f) = \sup \left\{ \int_{\Omega} |fg| dP: \rho(g) \leq 1 \right\}, \tag{5}$$

and let  $\rho'' = (\rho')'$ , the second associate. It is clear that  $\rho', \rho''$  are also function norms having the Fatou property. It is an important result (of Halperin-Luxemburg) that, when  $P$  is  $\sigma$ -finite and  $\rho$  has the Fatou pro-

perty,  $\rho = \rho''$  ([15], p. 470). This result holds true also when  $P$  is a more general measure (e. g. localizable). On the other hand,  $\rho'$  (and hence  $\rho''$ ) can be trivial ( $\equiv 0$ ) when  $P$  is arbitrary and the above relation may be false. Since no special properties of  $P$  are needed here,  $\rho = \rho''$  will be assumed below. This is automatic for the Orlicz (hence Lebesgue) spaces. [The reader may take  $P$  as a probability measure. But the treatment is not really simplified even with this restriction.]

**PROPOSITION 3.1.** — *Let  $L^\rho(\Sigma)$  be a Riesz space on a measure space  $(\Omega, \Sigma, P)$ , such that  $\rho = \rho''$ . Let  $f, f_1, f_2$  in  $L^\rho(\Sigma)$  be such that  $f \wedge f_j = 0$ ,  $j = 1, 2$ , i. e.,  $f$  and  $f_j$  have disjoint supports. If  $\rho(f_1) \geq \rho(f_2)$ , then  $\rho(f + f_1) \geq \rho(f + f_2)$ .*

*Proof.* — Since  $f \wedge f_i = 0$  implies  $\rho(f + f_i) = \rho(|f| + |f_i|)$  by the function norm property of  $\rho$ , the following computation holds.

$$\begin{aligned} \rho''(|f| + |f_1|) &= \sup \left\{ \int_{\Omega} (|f| + |f_1|) |g| dP : \rho'(g) \leq 1 \right\}, \\ &= \sup \left\{ \int_{\Omega} |f| |g_1| dP + \int_{\Omega} |f_1| |g_2| dP : g = g_1 + g_2, \right. \\ &\quad \left. \rho'(g) \leq 1, f \wedge g_2 = 0 = f_1 \wedge g_1 \right\}, \\ &\geq \sup \left\{ \int_{\Omega} |f| |g_1| dP + \int_{\Omega} |f_2| |g_2| dP : g = g_1 + g_2, \right. \\ &\quad \left. \rho'(g) \leq 1, f \wedge g_2 = 0 = f_2 \wedge g_1 \right\}, \\ &= \sup \left\{ \int_{\Omega} (|f| + |f_2|) |g| dP : \rho'(g) \leq 1 \right\}, \\ &= \rho''(|f| + |f_2|) = \rho(f + f_2), \text{ since } \rho = \rho''. \end{aligned} \tag{6}$$

This completes the proof.

A simple modification of the above proof, together with the closed graph theorem, shows that the following result is valid.

**PROPOSITION 3.2.** — *Let  $L^\rho(\Sigma)$  be a Riesz space such that  $L^\rho(\Sigma) = L^{\rho''}(\Sigma)$ . Then, for  $f, f_1, f_2$  as in Proposition 3.1,  $\rho(f_1 + f) \geq \gamma^2 \rho(f_2 + f)$  for a unique constant  $0 < \gamma \leq 1$  depending only on  $\rho$ .  $\gamma = 1$  implies the Fatou property of  $\rho$ .*

The details will be omitted as this proposition will not be used below. The main result of this section can now be stated:

**THEOREM 3.3.** — *Let  $L^\rho(\Sigma)$  be a Riesz space on  $(\Omega, \Sigma, P)$  and  $\rho = \rho''$ . Let  $\{f_0, f_n, n \geq 1\} \subset L^\rho(\Sigma)$  and  $\mathcal{B}_n = \sigma(f_1, \dots, f_n)$ , the  $\sigma$ -field generated by  $f_1, \dots, f_n$ . Let  $\mathcal{M}_n = L^\rho(\mathcal{B}_n) \subset L^\rho(\Sigma)$ . Suppose  $\rho$  is such that the following two conditions hold for  $L^\rho(\Sigma)$  when  $M$  is the closed subspace of  $L^\rho(\Sigma)$  spanned by all the step functions and  $\mathcal{B}_\infty = \sigma\left(\bigcup_n \mathcal{B}_n\right)$  where  $\rho'$  is the associate norm of  $\rho$ : (i)  $L^\rho(\Sigma)$  has  $(H_M)$ , and (ii)  $N(g\chi_{A_n}) \searrow 0$  for every  $g \in M$  as  $A_n \searrow \varnothing$ , where  $A_n \in \Sigma$  and  $N(g) = \sup \left\{ \int_\Omega |fg| dP : f \in L^\rho(\mathcal{B}_\infty), \rho(f) \leq 1 \right\}$ ,  $\chi_A$  being the indicator function of  $A$ . Then the following conclusions hold: (a) there exists a unique predictor  $h_n \in \mathcal{M}_n$  of  $f_0$ , for each  $n$ , (b)  $\rho(h_n - h_\infty) \rightarrow 0$  as  $n \rightarrow \infty$ , where  $h_\infty \in \mathcal{M}_\infty = L^\rho(\mathcal{B}_\infty)$  is the predictor of  $f_0$ ; and (c)  $h_n \rightarrow h_\infty$ , a. e.  $[P]$ , provided  $P$  has the finite subset property, i. e.,  $P(A) > 0$  implies the existence of  $B \subset A$ ,  $B \in \Sigma$  and  $0 < P(B) < \infty$ ; (d) condition (ii) is implied by the following one: (ii') each element  $g$  of  $M$  has an absolutely continuous norm, i. e.  $\rho'(g\chi_{A_n}) \searrow 0$  as  $A_n \searrow \varnothing$ ,  $A_n \in \Sigma$ ,  $n \rightarrow \infty$ .*

*Proof.* — Since  $M \subset L^\rho(\Sigma) \subset (L^\rho(\Sigma))^*$  where the second inclusion is an isometric imbedding, then it is known (and follows from a standard computation) that  $M$  is norm determining for  $L^\rho(\Sigma)$  (and so is total for  $L^\rho(\Sigma)$ ). If now  $C_n = \mathcal{M}_n$ :  $1 \leq n \leq \infty$ , then this is a sequence of closed convex and increasing subspaces of  $L^\rho(\Sigma)$ . It follows from the totality of  $M$  on  $L^\rho(\Sigma)$  and ([10], Thm. 3.2) that  $L^\rho(\Sigma)$  is  $\sigma(L^\rho, M)$ -sequentially complete and hence so is  $\mathcal{M}_n$  for each  $1 \leq n \leq \infty$ . [This can be false if  $M$  is replaced by  $L^\rho(\Sigma)$  itself.] From (ii) it follows that, in view of ([10], Th. 5.1), any bounded sequence of  $\mathcal{M}_\infty$  is  $\sigma(L^\rho, M)$ -sequentially compact. Since by (i)  $L^\rho(\Sigma)$  has  $(H_M)$ , the hypothesis of Theorem 2.1 is satisfied. Consequently, conclusions (a) and (b) follow from that result. Note also that if (ii') holds, then, since the Hölder inequality implies  $N(g) \leq \rho'(g)$  for each  $g \in M$ , the absolute continuity of  $\rho'$  on  $M$  trivially implies (ii). Thus it remains to prove (c) under the additional condition that  $P$  has the finite subset property. This will be established in the next three steps by an extension and modification of the proof of [12] (cf. also [1]). The computations of the latter will also be clarified.

I. Let  $A_i \in \mathcal{B}_i$ ,  $i = 1, \dots, n$ ,  $\bigcup_{i=1}^n A_i = \Omega$ ,  $A_i$  disjoint. Then for the predictors  $h_1, \dots, h_n$  of  $f_0$  one has

$$\rho(f_0 - h_1) \geq \rho(f_0 - \sum_{i=1}^n h_i \chi_{A_i}) \geq \rho(f_0 - h_n). \tag{7}$$

*Proof of I.* Since  $h_n$  and  $\sum_{i=1}^n h_i \chi_{A_i}$  are in  $L^p(\mathcal{B}_n)$ , and  $h_n$  being the closest element of  $f_0$  in this space, the last inequality of (7) is always true. To prove the first one, note that

$$f_0 - h_1 = \sum_{i=1}^n (f_0 - h_1) \chi_{A_i}$$

and that  $(f_0 - h_1) \chi_{A_j} \wedge (f_0 - h_1) \chi_{A_i} = 0$  for  $i \neq j$ . Moreover, from the definition of norm (using  $\rho = \rho''$ ) it follows that  $\rho(f_0 \chi_{A_i} - h_i \chi_{A_i}) \leq \rho(f_0 \chi_{A_i} - h_1 \chi_{A_i})$ , since on  $L^p(\mathcal{B}_i(A_i))$ ,  $h_i \chi_{A_i}$  is closest to  $f_0 \chi_{A_i}$  (and not  $h_1 \chi_{A_i}$ ), where  $\mathcal{B}_i(A_i)$  is the trace of  $\mathcal{B}_i$  on  $A_i$ . These two statements imply, via Proposition 3.1, that the first inequality of (7) is true if  $n=2$  and the general case follows by iteration or induction.

II. Let  $A_i \in \mathcal{B}_{n+i}$ ,  $i = 1, \dots, k$  with  $\bigcup_{i=1}^k A_i = \Omega$ ,  $A_i$  disjoint. If

$$g_{n,k} = \sum_{i=1}^k h_{n+i} \chi_{A_i},$$

then for any  $\varepsilon > 0$ , there is an  $n_\varepsilon$  such that  $n \geq n_\varepsilon$  implies

$$\rho(h_n - g_{n,k}) < \varepsilon. \tag{8}$$

*Proof of II.* Since  $g_{n,k}, h_{n+k} \in L^p(\mathcal{B}_{n+k})$  and  $\mathcal{B}_n \subset \mathcal{B}_{n+k}$ ,  $k \geq 1$ , (7) implies:

$$d_n = \rho(f_0 - h_n) \geq \rho(f_0 - g_{n,k}) \geq \rho(f_0 - h_{n+k}) \geq \rho(f_0 - h_\infty) = d_0. \tag{9}$$

By (b), and as in Step II of Theorem 2.1,  $h_n \rightarrow h_\infty$  in norm and  $d_n \rightarrow d_0$ , as  $n \rightarrow \infty$ . Hence  $\rho(f_0 - g_{n,k}) \rightarrow d_0$ , for each  $k$ , as  $n \rightarrow \infty$ . With this information, it will be shown that  $(g_{n,k} - h_\infty) \rightarrow 0$  in the  $\sigma(L^p, M)$ -topology, as  $n \rightarrow \infty$  so that, by an argument used in the proof of Theorem 2.1, its norm convergence can be deduced.

Let  $k$  be fixed. Since  $\{g_{n,k}\} \subset \mathcal{M}_\infty$  and is bounded there, it is relatively sequentially  $\sigma(L^p, M)$ -compact. So there exists  $g_{\infty,k} \in \mathcal{M}_\infty$  such that a subsequence  $g_{n_i,k} \rightarrow g_{\infty,k}$  in this topology, as  $i \rightarrow \infty$ . Since  $M$  is total, this implies (with the preceding result) that

$$d_0 \leq \rho(f_0 - g_{\infty,k}) \leq \varliminf_i \rho(f_0 - g_{n_i,k}) = d_0. \tag{10}$$

Hence  $g_{\infty,k}$  is also a predictor of  $f_0$  in  $\mathcal{M}_\infty$ . The rotundity of the latter space (hypothesis (i)) implies that  $g_{\infty,k} = h_\infty$  a. e., and the limit is independent of  $k$ .

Repeating this argument, it follows that  $g_{n,k} \rightarrow h_\infty$  in  $\sigma(L^p, M)$  and, by the  $(H_M)$ -property, this convergence is also in norm. But by (b),  $h_n \rightarrow h_\infty$  in norm. So for any  $k \geq 1$ ,

$$\rho(h_n - g_{n,k}) \leq \rho(h_n - h_\infty) + \rho(h_\infty - g_{n,k}) \rightarrow 0, \tag{11}$$

as  $n \rightarrow \infty$ . Hence (8) is true for large enough  $n$ .

III. With the additional condition that  $P$  has the finite subset property,  $h_n \rightarrow h_\infty$  a. e. as  $n \rightarrow \infty$ . [This always holds if  $P$  is  $\sigma$ -finite or localizable.]

*Proof of III.* The partition  $(A_1, \dots, A_k)$  will now be chosen suitably to prove the result. Let  $0 < \varepsilon < 1$  and  $\delta > 0$  be given. Define

$$A_1 = \{ \omega : |h_n - h_{n+1}|(\omega) > \delta \}, \text{ and for } 2 \leq i \leq k, \text{ let}$$

$$A_i = \{ \omega : |h_n - h_{n+j}|(\omega) \leq \delta, 1 \leq j \leq i-1, |h_n - h_{n+i}|(\omega) > \delta \}.$$

Finally, let  $A_{k+1} = \Omega - \bigcup_{i=1}^k A_i$ . Then  $A_i \in \mathcal{B}_{n+i+1}$  and  $A_i$  are disjoint. On  $A_{k+1}$  one has

$$\sup_{1 \leq i \leq k} h_{n+i} - \inf_{1 \leq j \leq k} h_{n+j} \leq 2\delta. \tag{12}$$

Let  $0 < k_1, k_2 \leq k$  and  $B \in \Sigma$  be any set with  $P(B) < \infty$  and  $\rho'(\chi_B) < \infty$ . Then

$$P \left[ \left\{ \omega : \sup_{1 \leq i \leq k_1} h_{n+i}(\omega) - \inf_{1 \leq j \leq k_2} h_{n+j}(\omega) > 2\delta \right\} \cap B \right] \leq P \left[ \left( \bigcup_{i=1}^k A_i \right) \cap B \right],$$

$$= \sum_{i=1}^k \int_{\Omega} \chi_{A_i} \chi_B dP, \text{ since } A_i \text{ are disjoint,}$$

$$\leq \frac{1}{\delta} \int_{\Omega} \left( \sum_{i=1}^k |h_n - h_{n+i}| \chi_{A_i} \right) \chi_B dP, \text{ since } |h_n - h_{n+i}| \geq \delta \text{ on } A_i,$$

$$= \frac{1}{\delta} \int_{\Omega} \left| \sum_{i=1}^k (h_n - h_{n+i}) \chi_{A_i} \right| \chi_B dP, \text{ since } A_i \text{ are disjoint,}$$

$$\leq \frac{1}{\delta} \int_{\Omega} |h_n - g_{n,k}| \chi_B dP,$$

$$\leq \frac{1}{\delta} \rho(h_n - g_{n,k}) \cdot \rho'(\chi_B), \text{ by H\"{o}lder's inequality,}$$

$$\leq \frac{\varepsilon}{\delta} \rho'(\chi_B), \text{ by Step II, for } n \geq n_\varepsilon.$$

Now letting  $k \rightarrow \infty$ , and then  $k_1, k_2 \rightarrow \infty$  in the above, it follows that

$$P\left[\left\{\omega: \sup_{i \geq n} h_i(\omega) - \inf_{j \geq n} h_j(\omega) \geq 2\delta\right\} \cap B\right] \leq \frac{\varepsilon}{\delta} \rho'(\chi_B). \quad (13)$$

If  $h_0 = \overline{\lim}_n h_n - \underline{\lim}_n h_n \geq 0$ , then  $\{\omega: h_0(\omega) \geq 2\delta\}$  is contained in the set  $\{\cdot\}$  above and so (13) implies:

$$P[\{\omega: h_0(\omega) \geq 2\delta\} \cap B] \leq \frac{\varepsilon}{\delta} \rho'(\chi_B). \quad (14)$$

Letting  $\varepsilon \rightarrow 0$  and then  $\delta \rightarrow 0$ , (14) yields that  $[\omega: h_0(\omega) > 0]$  is a locally P-null set, since B with the given property is otherwise unrestricted. However, by ([15], p. 257), when P has the finite subset property, every local P-null set is a P-null set. In this application, if  $\rho'(\chi_A) < \infty$  for every  $P(A) < \infty$ , then a direct application of ([15], Thm. 2 on p. 257) is possible, and an additional (easy) argument is needed in the general case (with  $\rho = \rho''$ ). Thus  $h_n \rightarrow h_\infty$  a. e., in any case.

This completes the proof of the Theorem.

*Remark.* — In [12] the prediction sequences were multiplied by a bounded function in the proof corresponding to Step III above and this has complicated the argument unnecessarily. As seen here in the work of [12], the bounded function should be taken as unity. Unlike in [12], the  $L^p$  spaces admitted in the present work need not even be reflexive, let alone uniformly rotund (see the next section). The  $\mathcal{M}_n$  are determined by the observed process  $(f_1, \dots, f_n)$ , and thus  $h_n = \Phi_n(f_1, \dots, f_n)$ , where  $\Phi_n$  is a Borel function of  $n$  arguments. In case  $\Phi_n$  is restricted to be a linear function and  $\mathcal{M}_n$  is a linear span of  $f_1, \dots, f_n$ , then the corresponding problem is usually called the linear prediction (or approximation). This is not considered in this section. In the present context its norm convergence is a consequence but the pointwise convergence is not. The latter is usually quite difficult (and may even be false), the Fourier series analysis being a case in point.

#### 4. SOME COMPLEMENTS

A brief comparison of the above results with known work on concrete spaces will be included here. This will also show the generality and limitations of the present methods.

1. If  $\mathcal{X}$  is uniformly rotund and smooth, then the hypotheses of Theorem 2.1 are automatically satisfied (cf. [4] [5]). When  $\mathcal{X} = L^p(\Omega, \Sigma, P)$ ,

$1 < p < \infty$ ,  $P(\Omega) < \infty$  and  $\mathcal{X} = L^\Phi(\Omega, \Sigma, P)$  [a uniformly rotund Orlicz space], then not only Theorem 2.1, but the hypothesis of Theorem 3.3 is also satisfied. These cases were initially considered in [1] and [12] using the uniform rotundity crucially. The needed smoothness properties for the Orlicz spaces are automatic when they are uniformly rotund (see [11]).

2. If  $\mathcal{X} = L^\Phi(\Omega, \Sigma, P)$ ,  $P(\Omega) < \infty$ , is an Orlicz space and  $\Phi(x) = |x| \log |x|$  for  $|x| \geq 2$ , and appropriately defined on  $[0,2]$  (cf. [9], p. 24), then the hypotheses of Theorems 2.1 and 3.3 are satisfied. The fact that  $(H_M)$  holds here is not easy to show, but this was proved in [13]. Since this  $\mathcal{X}$  is not reflexive, it follows that the results of this paper, which subsume this case, are more general than the previously known cases. If  $\mathcal{X} = L^2(\Omega, \Sigma, P)$  and  $P(\Omega) < \infty$ , then  $\{h_n, \mathcal{B}_n, n \geq 1\}$  is simply a martingale and the result of Theorem 3.3, in this case, is a consequence of the standard martingale convergence theorem.

3. Using the special nature of  $L^\Phi(\Omega, \Sigma, P)$  when  $P(\Omega) < \infty$ , it was possible to prove a more general result in [13] on « $\Phi$ -mean convergence» instead of the norm convergence to which it reduces when  $\Phi$  satisfies a growth condition. This is done even for vector valued processes  $\{f_1, f_2, \dots\}$  and the corresponding  $h_n \rightarrow h_\infty$  a. e. as  $n \rightarrow \infty$  was then proved. However, the present abstract treatment, besides clarifying the structure of the problem, includes many other cases. An account of the theory of Orlicz spaces may be found in [9] and the above needed extensions in [11].

4. As regards the vector valued case of [13], the following remark is in order, and it owes to some correspondence with J. J. Uhl: In ([13], Proposition 2.4), one must suppose the B-space to be reflexive so that the given six statements are equivalent. If only the B-space is assumed to have the Lebesgue-Radon-Nikodým property (as stated there), then one has only the following implications :

(1)  $\Leftrightarrow$  (2)  $\Leftarrow$  (3)  $\Rightarrow$  (4)  $\Leftrightarrow$  (6)  $\Rightarrow$  (2) and (3)  $\Leftrightarrow$  (5). But (2)  $\Rightarrow$  (3) is not necessarily true, contrary to a remark in [13]. In fact (2)  $\Leftrightarrow$  (3) if and only if the B-space there is reflexive. However, for the rest of the work, this additional implication is not needed and only a certain subset of the (easier) implications, noted above, are used. Thus none of the results of [13] is affected.

5. Finally, the next *all*-important problem is the explicit methods of (or algorithms for) the calculation of the predictors  $h_n$  for specific spaces. Except for a non-linear integral equation given in [13], for certain Orlicz spaces, very little is known about this question. The work of [6] and an example in [2] indicate that (infinite dimensional) programming methods

may be useful here. Also some general procedures of finding the closest elements in infinite dimensional spaces have been described in [3]. Both the methods of [3] and [6] appear useful in the considerations of the above-mentioned problem.

## REFERENCES

- [1] T. ANDÔ and I. AMEMIYA, Almost everywhere convergence of prediction sequences in  $L^p$ ,  $1 < p < \infty$ , *Z. Wahrscheinlichkeitstheorie verw. Geb.*, t. 4, 1965, p. 113-120.
- [2] R. C. BUCK, Linear spaces and approximation theory, in *On Numerical Approximation* (Ed. R. E. Langer). Univ. Wisconsin Press, Madison, 1959, p. 11-23.
- [3] R. C. BUCK, Application of duality in approximation theory, in *Approximation of Functions* (Ed. H. L. Garabedian), Elsevier, New York, 1965, p. 27-42.
- [4] D. F. CUDIA, Rotundity, *Proc. Symp. Pure Math.*, Vol. 7, *Amer. Math. Soc.*, 1963, p. 73-97.
- [5] M. M. DAY, *Normed Linear Spaces*, Springer-Verlag, Berlin, 1958.
- [6] R. J. DUFFIN and L. A. KARLOVITZ, Formulation of linear programs in analysis I: Approximation theory, *S. I. A. M. J. Appl. Math.*, t. 16, 1968, p. 662-675.
- [7] N. DUNFORD and J. T. SCHWARTZ, *Linear Operators, Part I: General Theory*, Interscience, New York, 1958.
- [8] K. FAN and I. GLICKSBERG, Some geometric properties of the spheres in a normed linear space, *Duke Math. J.*, t. 25, 1958, p. 553-568.
- [9] M. A. KRASNOSEL'SKIĭ and Ya. B. RUTICKIĭ, *Convex Functions and Orlicz Spaces*, P. Noordhoff Ltd., Gronigen, 1961.
- [10] W. A. J. LUXEMBURG and A. C. ZAAANEN, Compactness of integral operators in Banach function spaces, *Math. Ann.*, t. 149, 1963, p. 150-180.
- [11] M. M. RAO, Smoothness of Orlicz spaces, *Indag. Math.*, t. 27, 1965, p. 671-690.
- [12] M. M. RAO, Notes on pointwise convergence of closed martingales, *Indag. Math.*, t. 29, 1967, p. 170-176.
- [13] M. M. RAO, Abstract non linear prediction and operator martingales, *J. Multivariate Anal.*, t. 1, 1971, p. 129-157.
- [14] V. L. ŠMULIAN, Sur la structure de la sphère unitaire dans l'espace de Banach, *Mat. Sb. (N. S.)*, t. 9 (51), 1941, p. 545-561.
- [15] A. C. ZAAANEN, *Integration*, North Holland Publishing Co., Amsterdam, 1967.

(Manuscrit reçu le 21 février 1972).