A STATISTICAL BASIS FOR APPROXIMATION AND OPTIMIZATION

By R. N. Rich

Socony Mobil Oil Company, Inc., Princeton, New Jersey

1. Summary. Let T be a compact metric space and let $D = \{t_1, t_2, \cdots\}$ be a countable dense subset. We propose to show that if for $x \varepsilon C(T)$ we define $x_n(t) = E(x(t) \mid x(t_1), \cdots, x(t_n))$ (conditional expectation) then $E|x_n - x|^p \to 0$ where $|\cdot|$ denotes the sup norm and $p \geqq 1$ is such that $E|x|^p < \infty$. Furthermore, if we measure the distance between x_n and x by $\int |x_n(t) - x(t)|^2 \, d\mu(t)$ for some finite measure μ on T, then x_n is (in a least square sense) the optimal prediction for x given $x(t_1), \cdots, x(t_n)$.

We also consider an optimization problem in this same probabilistic setting. Roughly stated, we consider how, given $x(t_1), \cdots, x(t_n)$, one should choose $t \varepsilon T$ so as to maximize $E(x(t))$. The existence of an optimal policy is proved. If we let $S(x)$ denote the supremum of x over T and let $v_n(x)$ denote $x(t)$ where t is the point chosen in accordance with the optimal policy then it is shown that $E|S(x) - v_n(x)| \to 0$ as $n \to \infty$. This last result is obtained under the assumption that $E|x| < \infty$.

2. Conditional expectation.

Definition. A probability space (X, Σ, m) is called inherently regular if for every σ-algebra, $\Sigma' \subset \Sigma$, a regular conditional probability for Σ' exists.

Recall that a regular conditional probability for Σ' is a family $(m_x \mid x \varepsilon X)$ of probabilities on (X, Σ) such that if $A \varepsilon \Sigma$ then $m_x(A)$ as a function of x is Σ'-measurable and for any $A \varepsilon \Sigma, B \varepsilon \Sigma'$ we have $\int_B m_x(A) \, dm(x) = m(A \cap B)$. When such a family exists then the conditional expectation ψ of any integrable random variable φ defined on (X, Σ, m) can be given by $\psi(x) = \int \varphi \, dm_x$. For definitions and a general discussion of these ideas, we refer to [4].

In this section, we assume that (X, Σ, m) is an inherently regular probability space and that Σ' is a sub-σ-algebra of Σ. Moreover, we will denote by $(m_x \mid x \varepsilon X)$ a regular conditional probability for Σ'.

Throughout, B will denote a separable Banach space with norm $|\cdot|$. The dual of B (continuous linear functionals) will be denoted by B^* and the natural pairing between B and B^* by $\langle \cdot, \cdot \rangle$, i.e. $\langle b, \xi \rangle = \xi(b)$ for all $b \varepsilon B$ and $\xi \varepsilon B^*$.

We shall use the theory of vector valued integration as developed in Dunford and Schwartz [1]. Briefly, we recall that if $J: X \to B$ is Σ-measurable (i.e., ξJ is Σ-measurable for all $\xi \varepsilon B^*$) then J is integrable if there exists $b \varepsilon B$ such that $\int \langle J(x), \xi \rangle \, dm(x) = \langle b, \xi \rangle$ for all $\xi \varepsilon B^*$. In this case, we write $\int J \, dm = EJ = b$. Moreover, if J is Σ-measurable and $E|J| < \infty$ then J is integrable. By $L_p(X, \Sigma, m; B)$, we mean the space of all Σ-measurable functions of X into B such that $E|J|^p < \infty$. For these functions, we define $|J|_p = (E|J|^p)^{1/p}$. It follows

Received 29 July 1965.

that with $|\cdot|_p$ as a norm and with the obvious linear operations $L_p(X, \Sigma, m; B)$, for $p \geqq 1$, is a Banach space. Henceforth, since X and m are to remain fixed, we shall denote this space by $L_p(\Sigma; B)$. We shall always assume $p \geqq 1$.

By restricting the probability m to Σ', we can obtain another space $L_p(\Sigma'; B)$. Since $L_p(\Sigma'; B) \subset L_p(\Sigma; B)$ and since they both are Banach spaces, it follows that $L_p(\Sigma'; B)$ is a closed subspace of $L_p(\Sigma; B)$. We can conclude, by applying a well-known theorem concerning approximations in Banach spaces, that for every $J \varepsilon L_p(\Sigma; B)$ that there exists $P \varepsilon L_p(\Sigma'; B)$ such that $|J - P|_p \leqq |J - K|_p$ for all $K \varepsilon L_p(\Sigma'; B)$. In the case $p = 2$ and B is a Hilbert space, we can find P. To do this, however, we shall need the notion of conditional expectation of a vector valued function.

Definition. For $J \varepsilon L_1(\Sigma; B)$, we define $E(J \mid \Sigma')x = E'J(x) = \int J\, dm_x$. Since $\infty > \int |J|\, dm = \int (\int |J|\, dm_x)\, dm(x)$, it follows that for almost all $x \varepsilon X$, $\int |J| m_x < \infty$ and so by a previous remark $\int J\, dm_x$ exists. We read $E(J \mid \Sigma')$ as the conditional expectation of J given Σ'.

Since the preparation of this work, it has been pointed out to the author by J. L. Doob that a more general definition of conditional expectation has been formulated by Scalora and that Scalora has proved Theorem 2.4, see [5].

Lemma. *$E'\langle J, \xi\rangle = \langle E'J, \xi\rangle$ for almost all $x \varepsilon X$.*

Proof. $E'\langle J, \xi\rangle x = \int \langle J, \xi\rangle\, dm_x = \langle \int J\, dm_x, \xi\rangle = \langle (E'J)x, \xi\rangle$.

Using this lemma and the corresponding facts for conditional expectation in the case of real valued random variables, we can easily prove the following:

Theorem 2.1. *If $J \varepsilon L_1(X, \Sigma)$ then*

(1) *$E'J$ is Σ'-measurable;*

(2) *$\int_F J\, dm = \int_F E'J\, dm$ for all $F \varepsilon \Sigma'$;*

(3) *$E(E'J) = EJ$;*

(4) *$|E'J(x)| \leqq E'|J|(x)$; and*

(5) *if J is Σ'-measurable then $E'J = J$.*

Definition. By a projection of a Banach space B onto a subspace V, we mean an operator (bounded) $P: B \to B$ such that $P^2 = P$ and $P(B) = V$.

Theorem 2.2. *E' is a projection of $L_p(\Sigma; B)$ onto $L_p(\Sigma'; B)$ and $\|E'\| = 1$.*

Proof. First we show that if $J \varepsilon L_p(\Sigma; B)$ then $E'J \varepsilon L_p(\Sigma; B)$. Since $|J| \varepsilon L_p(\Sigma; R)$, we know that $E'|J| \varepsilon L_p(\Sigma; R)$; and since by Theorem 2.1 (Part 4) $|E'J| \leqq E'|J|$, it follows that $E'J \varepsilon L_p(\Sigma; B)$. Now by Theorem 2.1 (1) and (5), it follows that E' maps $L_p(\Sigma; B)$ onto $L_p(\Sigma'; B)$. To complete the proof, we note that $E|E'J|^p \leqq E(E'|J|)^p \leqq E(E'|J|^p) = E|J|^p$ and if J is a constant map $E'J = J$.

We can now prove a theorem which generalizes the principle of least squares.

Theorem 2.3. *If B is a Hilbert space then $L_2(\Sigma; B)$ is a Hilbert space and E' is the orthogonal projection of $L_2(\Sigma; B)$ onto $L_2(\Sigma'; B)$.*

Proof. We let $(\cdot, \cdot)$ denote the inner product in B and we define $[J, K] = \int (J(x), K(x))\, dm(x)$. It is readily verified that $[\cdot, \cdot]$ is an inner product in $L_2(\Sigma; B)$ which induces the norm. We can now apply a theorem (proved in [6]) which asserts that a projection of norm 1 is orthogonal.

We remark that if H is a Hilbert space and P is the orthogonal projection of H onto a subspace M then for any $h \varepsilon H$ the element of M closest to h is $P(H)$. Hence, this last theorem answers the question posed earlier. We proceed to establish a convergence theorem.

DEFINITION. If $\Sigma_n (n \geqq 1)$ is a sequence of σ-algebras contained in Σ such that (i) $\Sigma_i \subset \Sigma_{i+1}$ for all $i \geqq 1$ and (ii) Σ is the minimal σ-algebra containing the algebra $\cup \Sigma_n$ then we write $\Sigma_n \rightarrow \Sigma$.

THEOREM 2.4. *If $J \varepsilon L_p(\Sigma; B)$ and if $\Sigma_n \rightarrow \Sigma$ then $E(J \mid \Sigma_n) \rightarrow J$.*

PROOF. Since Σ is minimal over the algebra $\cup \Sigma_n$ we know ([1], p. 167) that we can find a sequence J_i of $\cup \Sigma_n$ simple functions such that $J_i \rightarrow J$. We can assume that $J_i \varepsilon L_p(\Sigma_i ; B)$. Now denoting $E(\cdot \mid \Sigma_n)$ by $E_n(\cdot)$ and the identity map of $L_p(\Sigma; B)$ into itself by I, we have

$$|(I - E_n)J|_p = |(I - E_n)(J - J_n) + J_n)|_p \leqq |(I - E_n)(J - J_n)|_p + |(I - E_n)J_n| \leqq \|I - E_n\| \cdot |J - J_n|_p .$$

Note that $(I - E_n)J_n = 0$ by Theorem 2.1 (5). Since $\|I - E_n\| \leqq 2$ and $|J - J_n|_p \rightarrow 0$ the theorem is proved.

3. Probability in Banach spaces. We now turn to the study of probabilities in Banach spaces. We let X denote a separable Banach space with the norm of $x \varepsilon X$ denoted by $N(x)$ and D a countable determining set for X. That is, D is a countable subset of X^* such that $N(x) = \sup \{|\xi(x)| : \xi \varepsilon D\}$. That such a set D exists is proved in [2].

DEFINITION. We let Σ denote the σ-algebra of X generated by the open sets of X.

THEOREM 3.1. *Let Σ_D be the smallest σ-algebra making all the functionals in D measurable. Then $\Sigma = \Sigma_D$.*

PROOF. Clearly $\Sigma_D \subset \Sigma$. For $a \varepsilon X$ we define $N_a(x) = N(x - a)$. Since D is a determining set for X, we have $N_a(x) = \sup \{|\xi(x) - \xi(a)| : \xi \varepsilon D\}$ and since each $\xi \varepsilon D$ is Σ_D-measurable it follows that N_a is Σ_D-measurable. Therefore, Σ_D contains all the open spheres of X. Now X is second countable and since the open spheres form a base for the topology of X it follows that every open set is a countable union of spheres. Therefore, Σ_D containing all the open sets must equal Σ.

THEOREM 3.2. *If m is a probability on (X, Σ) then (X, Σ, m) is inherently regular.*

PROOF. Since X is a separable metric space, it can be homeomorphically imbedded in R^∞ (a countable product of the reals). For a proof see [3], p. 125. Moreover, since X is complete in its metric, it follows [3], p. 207 that its image in R^∞ is a G_δ and hence is a Borel set. Now we can apply a theorem of Doob (proved in [4], p. 361) the essential content of which is that Borel sets in R^∞ are inherently regular for any probability defined on their Borel subsets.

We now proceed to prove a lemma which we will use later. We assume that a probability m has been defined on (X, Σ).

Lemma. *Let Σ' be a σ-algebra contained in Σ and let Σ' be the minimal σ-algebra containing a given countable algebra Π. Then if $(m_x \mid x \in X)$ is a regular conditional probability for Σ' there exists a set $M \in \Sigma'$ such that $m(M) = 0$ and $x \notin M$ implies $m_x(F) = \chi_F(x)$ for all $F \in \Sigma'$. Here χ_F denotes the characteristic function of F.*

Proof. For $F \in \Pi$, we define $F^0 = \{x \mid m_x(F) \neq \chi_F(x)\}$. It is easily seen that $F^0 \in \Sigma'$ and that $m(F^0) = 0$. Let $M = \cup \{F^0 : F \in \Pi\}$. Clearly $m(M) = 0$ and for $x \notin M$ and any $F \in \Pi$ we have $m_x(F) = \chi_F(x)$. Both $m_x(F)$ and $\chi_F(x)$, for fixed $x \in X$, define measures on Σ' and since they agree on a ring which generates Σ' they must agree on all of Σ'. This completes the proof.

Theorem 3.3. *Let O be a continuous linear map of X into a separable Banach space Y. Let Π denote the σ-algebra for Y generated by its open sets and let $\Sigma' = O^{-1}\Pi$. Denote by I the identity map of X into itself and suppose that $EN < \infty$. Then for almost all $x \in X$, we have $O(x) = O((E'I)x)$.*

Proof. Clearly Π is generated by a countable ring (Y is second countable) and so Σ' is also. We let M be the set guaranteed by the preceding lemma.

We assume $x \notin M$. Let $y = O(x)$ and let $V = O^{-1}(y)$. We know by the linearity and continuity of O that V is a closed convex subset of X. If $\bar{x} = (E'I)x \notin V$ then there exists $\xi \in X^*$ and a real number α such that $\xi(\bar{x}) < \alpha \leqq \xi(z)$ for all $z \in V$, ([1], p. 417).

Since $x \in V$ and $x \notin M$, we have that $m_x(V) = 1$ and so $\xi(\bar{x}) = \xi(\int I\, dm_x) = \int \xi(z)\, dm_x(z) = \int_V \xi(z)\, dm_x(z) \geqq \alpha$. This contradiction completes the proof.

This theorem may be given the following interpretation: if we consider $O(x)$ to represent an observation and $(E'I)(x)$ to represent a prediction for x given $O(x)$ then the theorem asserts that the predicted value for x will be consistent with the observation.

Definition. If (Y, Π) is a measurable space with Π a σ-algebra and if $\varphi_1, \cdots, \varphi_n$ are Σ-measurable maps of X into Y, we define, for any $J \in L_1(\Sigma; B)$, $E(J \mid \varphi_1, \cdots, \varphi_n)$ to be the conditional expectation of J given the smallest σ-algebra making $\varphi_1, \cdots, \varphi_n$ measurable.

4. Applications to function spaces. We let T be a compact metric space with $D = \{t_1, t_2, \cdots\}$ a countable dense subset and we let $X = C(T)$ the space of all continuous real valued functions on T. For $x \in X$, we define the norm of x by $|x| = \sup\{|x(t)| : t \in T\}$. It is known ([3], p. 245) that X is a separable Banach space. For $t \in T$, we define $t^*(x) = x(t)$. We let I be the identity map of X into itself. We assume that (X, Σ, m) is a probability space where Σ, of course, is the σ-algebra generated by open subsets of X.

Theorem 4.1. *If $E|x|^p < \infty$ and if $P_n = E(I \mid t_1^*, \cdots, t_n^*)$ then $E|P_n x - x|^p \to 0$ as $n \to \infty$.*

Proof. First note that $E|x|^p < \infty$ is equivalent to $I \in L_p(X, \Sigma, m; X)$ and so the conditional expectations P_n of I are well defined. Also $\{t^* : t \in D\}$ is a determining set for X. The conclusion follows from Theorem 2.4 and Theorem 3.1.

We remark that if Σ' is a σ-algebra contained in Σ (as we are assuming) and if $(m_x \mid x \in X)$ is a regular conditional probability for Σ' then $E(I \mid \Sigma')x$ is the function $\bar{x}(t) = \int z(t)\, dm_x(z)$. Note that using this last equation, we can define $\bar{x}$ without using the notion of vector valued integration. More precisely, if $E|x| < \infty$ then $E|x| = \int (\int |z|\, dm_x(z))\, dm(x)$ and so $\int |z|\, dm_x(z)$ is finite for almost all $x \varepsilon X$. Since $|z(t)| \leqq |z|$, it follows that $\int z(t)\, dm_x(z)$ is defined for almost all $x \varepsilon X$. We now show that the function $J(x) = \bar{x}$ is, in a certain sense, an optimal prediction (given Σ') for the function $I(x) = x$.

We denote by Π the σ-algebra generated by the open subsets of T and we let μ denote a finite measure on (T, Π) which we take, for convenience, to be normalized (i.e., $\mu(T) = 1$). We let $H = L_2(T, \Pi, \mu; R)$ and we denote the inner product in H by $(\cdot, \cdot)$ and the norm by $|\cdot|_2$. We remark that $X \subset H$. Note that we still denote the norm in X by $|\cdot|$.

Lemma. *Let $h \varepsilon H$ and let $h'(x, t) = x(t)\cdot h(t)$. Then h' is $\Sigma \times \Pi$ measurable.*

Proof. If h is continuous on T then h' is continuous on $X \times T$ and so is $\Sigma \times \Pi$ measurable. If $h_n \to h$ pointwise (a.e) on T then $h_n' \to h'$ pointwise (a.e) on $X \times T$ and so the lemma follows from the fact that every $h \varepsilon H$ is the pointwise (a.e) limit of continuous functions ([1], p. 170).

Theorem 4.2. *Assume $E|x|^2 < \infty$ and let $J\colon X \to H$ be defined by $J(x) = \bar{x}$. Then $J \varepsilon L_2(\Sigma'; H)$ and if $K \varepsilon L_2(\Sigma'; H)$ it follows that $E \int |x(t) - Kx(t)|^2\, d\mu(t) \geqq E \int |x(t) - \bar{x}(t)|^2\, d\mu(t)$, that is, $E(|x - K_x|_2)^2 \geqq E(|x - \bar{x}|_2)^2$.*

Proof. Let $I\colon X \to H$ be the inclusion map. We shall show that $(E'I)x = \bar{x}$. The theorem will then follow from 2.3. Let $y = (E'I)x$. Then for all $h \varepsilon H$ we have, by the definition of vector valued integral and by an interchange in the order of integration (justified by the previous lemma), the following equality:

$$(y, h) = \int (z, h)\, dm_x(z) = \int (\int z(t)h(t)\, d\mu(t))\, dm_x(z)$$
$$= \int (\int z(t)h(t)\, dm_x(z))\, d\mu(t) = \int \bar{x}(t)h(t)\, d\mu(t) = (\bar{x}, h).$$

Since this holds for all h, it follows that $y = \bar{x}$.

Theorem 4.3. *If $E|x| < \infty$ and if $P_n = E(I \mid t_1^*, \cdots, t_n^*)$ then for almost all $x \varepsilon X$, $(P_n x)t_i = x(t_i)$ for $i = 1, \cdots, n$.*

Proof. The theorem follows from 3.3 by letting $O(x) = (x(t_1), \cdots, x(t_n))$.

We now turn to the study of an optimization problem. We shall first give an intuitive formulation of the problem. Let X, Σ, m be as before and let O be a continuous linear map of X into Y. An element $x \varepsilon X$ is to be chosen in accordance with the probability m, and we are then to be told the value $O(x)$. On the basis of this information, we are to choose a point $t \varepsilon T$; and after having chosen this point, we receive as payoff the value $x(t)$. What policy shall we adopt so as to maximize our expected payoff? A policy is essentially a function ψ from Y into T or equivalently a function φ from X into T such that φ is constant on the sets $O^{-1}(y)$ for all $y \varepsilon Y$. The expected payoff for a policy φ would be, of course $E(x(\varphi x))$. This motivates the following definitions:

Definition. For $x \varepsilon X$, we define $V(x) = O^{-1}(Ox)$.

Definition. A function φ from X to T is called admissible if (i) φ is constant on each $V(x)$ and (ii) $x(\varphi x)$ is Σ-measurable in x. Remark. If φ is admissible and if $E|x| < \infty$ then $Ex(\varphi x) < \infty$.

We now assume $E|x| < \infty$ and we let Σ' denote the σ-algebra induced by O. We let $(m_x \mid x \in X)$ denote a regular conditional probability for Σ'. By earlier results, we know there exists a set $M \in \Sigma'$ such that (i) $m(M) = 0$, (ii) $x \notin M$ implies $m_x(F) = \chi_F(x)$ for all $F \in \Sigma$, and (iii) $x \notin M$ implies $O(x) = O(\bar{x})$, i.e. $\bar{x} \in V(x)$. For the remainder of this section, M will denote such a set.

DEFINITION. If f is any real valued function, we denote by $S(f)$ the supremum of $f(x)$ for all x in the domain of f.

We will use this notation only when the domain of f is understood. In particular if $x \in X$ then $S(x) = \sup \{x(t) : t \in T\}$. We remark that on X, S is Σ-measurable.

DEFINITION. A function θ of X into T is called Σ-measurable if $\theta^{-1}(B) \in \Sigma$ where B is any set in the σ-algebra of T generated by the open subsets of T.

We remark that an admissible function of X into T need not be Σ-measurable.

LEMMA. *There exists a map θ of X into T such that θ is Σ-measurable and $x(\theta x) = Sx$.*

PROOF. Let K denote a compact subset of the real line. For $x \in C(K)$ define $\varphi(x) = \inf \{t : t \in K \text{ and } x(t) = S(x)\}$. We first show that φ is lower semicontinuous. To do this, let a be a real number and let $A = \{x : x \in C(K) \text{ and } \varphi(x) \leqq a\}$. If $\{x_n\}$ is a sequence in A and $x_n \to x$, we can assume $\varphi(x_n) \to t \in K$. Since $\varphi(x_n) \leqq a$ for all n, it follows that $t \leqq a$. Moreover, $x_n(\varphi x_n) \to x(t)$ and it is easily verified that $S(x_n) \to S(x)$. By the definition of φ, we know that $x_n(\varphi x_n) = S(x_n)$ and so $S(x) = x(t)$ from which we can conclude that $\varphi(x) \leqq t \leqq a$ and so A is closed.

To prove the lemma, we let K denote the Cantor discontinuum and let δ denote a continuous function from K onto T (see [3], p. 166). We define the function $\delta^*: X \to C(K)$ by $\delta^*(x) = x\delta$ for all $x \in X$ and we define $\theta(x) = \delta(\varphi(\delta^* x))$. Since each of the maps $\delta, \varphi, \delta^*$ is either continuous or semicontinuous, it follows that θ is measurable. That $x(\theta x) = Sx$ is easily verified.

LEMMA 4.1. *Let φ be admissible. Then $x \notin M$ implies $\int z(\varphi z)\, dm_x(z) = \bar{x}(\varphi \bar{x})$.*

PROOF. If $x \notin M$ then $m_x(V(x)) = 1$ and $\bar{x} \in V(x)$. Therefore, $\int z(\varphi z)\, dm_x(z) = \int_{V(x)} z(\varphi z)\, dm_x(z) = \int_{V(x)} z(\varphi \bar{x})\, dm_x(z) = \bar{x}(\varphi \bar{x})$.

DEFINITION. As before, we let I denote the identity map of X into itself and we let $J = E(I \mid \Sigma')$.

Recall that $J(x) = \bar{x}$.

LEMMA. *For all $x \notin M$, we have $J^2 x = Jx$.*

PROOF. First we note that J is Σ'-measurable and so J is constant on the sets $V(x)$. Moreover, by Theorem 3.3 we have $J(x) \in V(x)$ for all $x \notin M$. Since $x \in V(x)$, this proves the lemma.

THEOREM 4.4. *Let θ be a Σ-measurable function of X into T such that $x(\theta x) = S(x)$ and let $\bar{\theta} = \theta J$. Then $\bar{\theta}$ is admissible and if φ is any admissible function then $E(x(\bar{\theta} x)) \geqq E(x(\varphi x))$.*

PROOF. Since J is Σ'-measurable, it follows that $\bar{\theta}$ is Σ'-measurable. Therefore, $\bar{\theta}$ is constant on the sets $V(x)$. Moreover, the map of X into $X \times T$ defined by $x \to (x, \bar{\theta} x)$ is measurable and since the map of $X \times T$ into the reals is defined by $(x, t) \to x(t)$ is continuous it follows that $x(\bar{\theta}x)$ is Σ-measurable, in fact $x(\bar{\theta}x)$ is even Σ'-measurable. Now let $x \varepsilon M$. Then since $J^2x = Jx$ we have that $\bar{\theta}(\bar{x}) = \theta(\bar{x})$. Therefore, $\int z(\bar{\theta}z)\, dm_x(z) = \bar{x}(\bar{\theta}\bar{x}) = \bar{x}(\theta\bar{x}) \geqq \bar{x}(\varphi\bar{x}) = \int z(\varphi z)\, dm_x(z)$. Now $E(x(\bar{\theta}x)) = \int \bar{x}(\theta\bar{x})\, dm(x)$ and similarly $E(x(\varphi x)) = \int \bar{x}(\varphi\bar{x})\, dm(x)$ and so the conclusion follows.

We now proceed to prove a convergence theorem. To do this, it will be convenient to change the notation slightly. We will denote the norm of $x \varepsilon X$ by $N(x)$, i.e. $N(x) = \sup \{|x(t)|: t \varepsilon T\}$. We will use $|\cdot|$ only to denote absolute value. As before, we let I denote the identity map of X onto itself and we let θ be a measurable map of X into T such that $x(\theta x) = Sx$.

LEMMA. *Let $K_n \to K$ in $L_1(X, \Sigma; X)$. Then $SK_n \to SK$ in $L_1(X, \Sigma; R)$. Here R denotes the reals.*

PROOF. It is easily proved that $|SK_nx - SKx| \leqq N((K_n - K)x)$ and so the lemma follows.

THEOREM 4.5. *If $EN < \infty$ then $E|Sx - x(\theta_nx)| \to 0$ as $n \to \infty$ where $\theta_n = \theta E(I \mid t_1^*, \cdots, t_n^*)$.*

PROOF. Let $J_n = E(I \mid t_1^*, \cdots, t_n^*)$. It follows from Theorem 4.1 that $J_n \to I$ in $L_1(X, \Sigma; X)$. Now by Lemma 4.1 and Theorem 4.4, we have that (1) $\int z(\theta_nz)\, dm_x(z) = SJ_nx$ for almost all $x \varepsilon X$. Since $S(x) \geqq x(t)$ for all $t \varepsilon T$, it follows that $E|S(x) - x(\theta_nx)| = ES - Ex(\theta_nx)$. In view of (1), we have $Ex(\theta_nx) = ESJ_n$ and since $ES - ESJ_n \leqq E|S - SJ_n|$ the conclusion follows from the preceding lemma.

REFERENCES

[1] DUNFORD, N. and SCHWARTZ, J. (1958). *Linear Operators*. Interscience, New York.

[2] HILLE, E. and PHILIPS, R. (1957). *Functional Analysis and Semi-Groups*. Amer. Math. Soc., Providence.

[3] KELLY, J. (1955). *General Topology*. Van Nostrand, New York.

[4] LOÈVE, M. (1955). *Probability Theory*. Van Nostrand, New York.

[5] SCALORA, F. (1961). Abstract martingale convergence theorems. *Pacific J. Math.* **11** 347–374.

[6] ZAANEN, A. (1953). *Linear Analysis*. North Holland, Amsterdam.