ON AN EXTENSION OF THE CONCEPT CONDITIONAL EXPECTATION

H. D. BRUNK¹

The purpose of the present note is in part to extend results of Šidák [3] to cases where the measure is not finite, perhaps not even σ -finite. In particular, the conditional expectation of a random variable in L_1 is obtained without the use of the Radon-Nikodym Theorem, which, indeed, does not apply with the required generality. It was found possible to extend at the same time results of [1] so as to omit the requirement that the measure be totally finite. For this reason the σ -fields discussed in [3] are here replaced by σ -lattices. In this connection it may be observed that $E(X|\mathcal{L})$ is the solution of the regression problem: given $X \in L_2$, choose Y in the class C of random variables in L_2 measurable with respect to a σ -field $\mathcal{L}$ so as to minimize $E(X - Y)^2$; $E(X|\mathcal{L})$ is the projection in L_2 of X on C . Certain problems of maximum likelihood estimation of ordered parameters have solutions which also solve the above regression problem, in which $\mathcal{L}$ is not a σ -field but is a σ -lattice, closed under countable union and countable intersection, but not necessarily under complementation (see references in [1]).

Let $(\Omega, \mathcal{S}, \mu)$ be a measure space: $\mathcal{S}$ is a σ -field of subsets of Ω , and μ is a measure, σ -additive and complete, but not necessarily σ -finite. The symbol $\mathcal{S}$ will also be used to denote the class of all events: an *event* is an equivalence class of sets in $\mathcal{S}$, two sets being equivalent if their symmetric difference has measure 0. The symbol Ω will denote also the equivalence class to which Ω belongs. (Events will often be denoted by descriptions enclosed in brackets.) A *random variable* is an equivalence class of real-valued $\mathcal{S}$ -measurable functions on Ω , two such functions representing the same random variable if they differ on a set of measure 0. If A is an event, A^c will denote the complement of A , and $\chi(A)$ or χ_A will denote the indicator random variable of A . Let L_1 denote the class of integrable random variables, L_2 the class of square-integrable random variables. If $\mathcal{L}$ is a sub- σ -lattice (closed under countable union, countable intersection) of $\mathcal{S}$ containing $\emptyset$ and Ω , let a random variable X be termed $\mathcal{L}$ -measurable if $[X > a] \in \mathcal{L}$ for every real a . A family C of random variables will be called a *convex cone* if $k \geq 0$, $X \in C$, $Y \in C \Rightarrow kX \in C$, $X + Y \in C$; a

Received by the editors February 15, 1962.

¹ This research was partially supported by the Air Force Office of Scientific Research.

conditional σ -lattice if $X_n \in C$, $n = 1, 2, \dots$, $X_n \leq X$ for some random variable $X \Rightarrow \sup_n X_n \in C$, and dually. Let $R(\mathfrak{L})$ denote the class of $\mathfrak{L}$ -measurable random variables; $R(\mathfrak{L})$ is a conditional σ -lattice and convex cone, and contains the constant random variables (cf. Theorem 4). Let $L_1(\mathfrak{L}) = L_1 \cap R(\mathfrak{L})$, $L_2(\mathfrak{L}) = L_2 \cap R(\mathfrak{L})$. The class $L_2(\mathfrak{L})$ is a closed convex cone and sublattice of L_2 (cf. Theorem 5).

Let H be a real Hilbert space (not necessarily separable), and C a closed convex cone in H . Let $X \in H$, and let Y denote the closest point of C to X ; this closest point will be termed the *projection of X on C* . It is characterized by the properties

- (1) $(X - Y, Z) \leq 0$ for all $Z \in C$,
- (2) $(X - Y, Y) = 0$,

where $(\cdot, \cdot)$ denotes inner product. For if there were $Z \in C$ such that $(X - Y, Z) > 0$, on setting $a = (X - Y, Z) / \|Z\|^2$ one would find $Y + aZ$ nearer X than Y ; and if $(X - Y, Y) < 0$ then $(1 - b)Y$ is nearer X than Y , if b is chosen so that $0 < b < \min(1, 2c / \|Y\|^2)$, where $c = -(X - Y, Y)$. Simple calculations show also that if Y_i is the projection of X_i on C , $i = 1, 2$, then

$$(3) \quad \|Y_1 - Y_2\| \leq \|X_1 - X_2\|.$$

It follows that projection on C is a continuous operator (not necessarily linear) in H .

Let $\mathfrak{B}'$ denote the class of Borel sets of reals which exclude the origin.

THEOREM 1. *If $X \in L_2$ then there is a unique $Y \in L_2(\mathfrak{L})$ such that*

$$(4) \quad \int XZ d\mu \leq \int YZ d\mu \quad \text{for all } Z \in L_2(L),$$

and

$$(5) \quad B \in Y^{-1}(\mathfrak{B}') \quad \mu(B) < \infty \Rightarrow \int_B X d\mu = \int_B Y d\mu.$$

If also $X \in L_1$ then $Y \in L_1$ and (5) obtains for all $B \in Y^{-1}(\mathfrak{B}')$. Further, if k is a real number such that $X \leq k$ ($X \geq k$) then $Y \leq k$ ($Y \geq k$).

This theorem is proved in [1] for totally finite measures μ , in which case the exclusion of Borel sets containing 0 is not required. That in general (5) need not hold with $\mathfrak{B}'$ replaced by the class $\mathfrak{B}$ of all Borel sets is seen on letting $\mathfrak{L}$ contain only $\emptyset$, Ω , and an event A of finite measure, with $\mu(A^c) = \infty$, and taking for X a random

variable in L_1 such that $\int_A X d\mu > 0$, $\int_A^c X d\mu \neq 0$; then $\int_A^c X d\mu \neq \int_A^c Y d\mu = 0$.

PROOF OF THEOREM 1. Let Y denote the projection of X on $L_2(\mathfrak{L})$. Inequality (4) is then an instance of (1). Let D denote the class of real-valued functions ϕ on the reals such that $\phi(Y) \in L_2(\mathfrak{L})$ and $\int (X - Y)\phi(Y) d\mu = 0$. Using the fact that $a \geq 0$, $m \geq 0$, $Z \in L_2(\mathfrak{L})$ imply that all of the random variables $(mZ - a) \vee 0$, $mZ \wedge a$, $mZ \vee (-a)$, $(mZ + a) \wedge 0$ are in $L_2(\mathfrak{L})$, one finds that D is a lattice containing, for each positive number u and for each positive integer $n > 1/u$, the function ϕ_n defined by $\phi_n(t) = 0$ for $t \leq u - 1/n$, $\phi_n(t) = n(t - u + 1/n)$ for $u - 1/n < t \leq u$, $\phi_n(t) = 1$ for $t > u$. For n sufficiently large, $[\phi_n(Y) \neq 0] \subset [Y > u/2]$, an event of finite measure, so that Lebesgue's Dominated Convergence Theorem may be used to show that $\int_{[Y \geq u]} (X - Y) d\mu = 0$ if $u > 0$. Similarly the integral of $X - Y$ over each of the events $[Y > u]$, $[Y \leq -u]$, $[Y < -u]$ is zero if $u > 0$. The remaining conclusions of the theorem follow.

Another property of projection on a closed convex cone and lattice will be required, valid in a real Hilbert space H which is also a vector lattice satisfying $(a, b) = (a \vee b, a \wedge b)$ and $a \geq 0$, $b \geq 0 \Rightarrow (a, b) \geq 0$. If Y_i denotes the projection of X_i on C , $i = 1, 2$, and if $X_2 \geq X_1$, then $Y_2 \geq Y_1$. For, on setting $Y^* = Y_1 \vee Y_2$, $Y_* = Y_1 \wedge Y_2$ and noting that $Y_1 + Y_2 = Y_* + Y^*$, we find that $(Y^* - Y_1, Y_1 - Y_*) = (Y_2, Y_1) - (Y^*, Y_*) = 0$. It follows that

$$\begin{aligned} (X_2 - Y^*, Y_1 - Y_*) &= (X_2 - X_1, Y_1 - Y_*) + (X_1 - Y_1, Y_1 - Y_*) \\ &= (X_2 - X_1, Y_1 - Y_*) - (X_1 - Y_1, Y_*) \geq 0 \end{aligned}$$

since $X_2 - X_1 \geq 0$, $Y_1 - Y_* \geq 0$, and $Y_* \in C$. Hence

$$\begin{aligned} \|X_2 - Y_2\|^2 &= \|X_2 - Y^*\|^2 + 2(X_2 - Y^*, Y_1 - Y_*) + \|Y_1 - Y_*\|^2 \\ &\geq \|X_2 - Y^*\|^2 + \|Y_1 - Y_*\|^2 \end{aligned}$$

which implies

$$Y^* = Y_2 \quad \text{and} \quad Y_* = Y_1.$$

THEOREM 2. If $X \in L_1$ then there is a unique $Y \in L_1(\mathfrak{L})$ such that

$$(6) \quad \int X Z d\mu \leq \int Y Z d\mu \text{ for all bounded } Z \in L_2(\mathfrak{L})$$

and

$$(7) \quad \int_B X d\mu = \int_B Y d\mu \text{ for all } B \in Y^{-1}(\mathfrak{B}').$$

PROOF. For positive integers M, N , define $X_{N,M} = N$ if $X > N$, $X_{N,M} = X$ if $1/N < X \leq N$, $X_{N,M} = 0$ if $-1/M \leq X \leq 1/N$, $X_{N,M} = X$ if $-M \leq X \leq -1/M$, $X_{N,M} = -M$ if $X < -M$. Let $Y_{N,M}$ denote the projection on $L_2(\mathcal{L})$ of $X_{N,M}$. From the remarks preceding the theorem it follows that $Y_N = \lim_M Y_{N,M}$ and $Y = \lim_N Y_N$ exist as extended real-valued random variables. Let $E_{N,M} = [Y_{N,M} > 0]$, $F_{N,M} = [Y_{N,M} < 0]$. Then by Theorem 1,

$$\int_{E_{N,M}} |Y_{N,M}| d\mu = \int_{E_{N,M}} Y_{N,M} d\mu = \int_{E_{N,M}} X_{N,M} d\mu \leq \int_{E_{N,M}} |X| d\mu,$$

and similarly

$$\int_{F_{N,M}} |Y_{N,M}| d\mu \leq \int_{F_{N,M}} |X| d\mu,$$

so that

$$\int |Y_{N,M}| d\mu \leq \int |X| d\mu.$$

Using the monotone convergence theorem, one finds that $Y \in L_1(\mathcal{L})$. Limit arguments lead to (6) and (7). Let $\phi_n(t) = [t \wedge n] \vee (-n)$; then $Z \in L_1(\mathcal{L}) \Rightarrow \phi_n(Z) \in L_2(\mathcal{L})$, and $\phi_n(Z)$ is bounded. If both Y_1 and Y_2 satisfy (6) and (7) then $\int (X - Y_i) \phi_n(Y_i) d\mu = 0$ and $\int (X - Y_i) \phi_n(Y_{3-i}) d\mu \leq 0$, $i = 1, 2$, hence $\int (Y_2 - Y_1) [\phi_n(Y_2) - \phi_n(Y_1)] d\mu \leq 0$ for every n , which in turn implies $Y_1 = Y_2$.

Remark on Theorem 2. If $\mathcal{L}$ is indeed a σ -field of events, rather than just a σ -lattice, then $Z \in L_2(\mathcal{L})$ implies $-Z \in L_2(\mathcal{L})$, so that Theorem 2 yields the conditional expectation, $E(X|\mathcal{L})$. We note that no use is made here of the Radon-Nikodym Theorem, which indeed need not hold in $(\Omega, \mathcal{S}, \mu)$; Segal shows in [2] that the Radon-Nikodym Theorem holds only in "localizable" measure spaces. While linearity fails for general σ -lattices $\mathcal{L}$, no better notation for the random variable Y in Theorem 2 occurs to the author than $E(X|\mathcal{L})$.

THEOREM 3. *If $\mathcal{L}$ is a sub- σ -field of $\mathcal{S}$ then $E(\cdot|\mathcal{L})$ is the unique continuous extension to L_1 of projection in L_2 on $L_2(\mathcal{L})$.*

For the case $\mu(\Omega) < \infty$ this theorem is given by Šidák [3]. For the more general situation in which $\mathcal{L}$ is an arbitrary sub- σ -lattice of $\mathcal{S}$ containing $\emptyset$ and Ω , the present author does not know whether or not the conclusion obtains. However, it is not difficult to show that if X_n is a monotone sequence of elements of L_1 converging in L_1 to X

then $E(X_n | \mathcal{L})$ converges in L_1 to $E(X | \mathcal{L})$.

PROOF OF THEOREM 3. Let $X_n \in L_1$, $n=0, 1, 2, \dots$, $\int |X_n - X_0| d\mu \rightarrow 0$. Set $Y_n = E(X_n | \mathcal{L})$, $n=0, 1, 2, \dots$. In the present situation $E(\cdot | \mathcal{L})$ is a linear operator as well as isotone. Also $Y_0 - Y_n \in L_1(\mathcal{L})$, $n=1, 2, \dots$, so that there is a σ -finite event $A \in \mathcal{L}$ such that $Y_0 - Y_n = 0$ on A^c , $n=1, 2, \dots$. Then $\int |Y_0 - Y_n| d\mu = \int_A |Y_0 - Y_n| d\mu \leq \int_A E(|X_0 - X_n| | \mathcal{L}) d\mu = \int_A |X_0 - X_n| d\mu \rightarrow 0$ as $n \rightarrow \infty$. Thus Y_n converges in L_1 to Y_0 . The proof is completed by the remark that two continuous transformations from L_1 to L_1 coinciding on $L_1 \cap L_2$ coincide also on L_1 .

THEOREM 4. *Necessary and sufficient conditions that a family F of random variables be $R(\mathcal{L})$ for some σ -lattice $\mathcal{L}$ containing $\emptyset$ and Ω are:*

- (8) F is a conditional σ -lattice;
- (9) each constant random variable is in F ;
- (10) F is a convex cone.

PROOF. One readily verifies that if $\mathcal{L}$ is a sub- σ -lattice of $\mathcal{S}$ containing $\emptyset$ and Ω then $F = R(\mathcal{L})$ has the stated properties. Conversely, suppose F has these properties. Let $\mathcal{L}$ denote the class of events A for which there exist real a and $X \in F$ such that $A = [X > a]$. Property (9) implies $\emptyset \in \mathcal{L}$, $\Omega \in \mathcal{L}$. For fixed $X \in F$, let D denote the class of real-valued Borel functions ϕ of a real variable such that $\phi(X) \in F$. One verifies that D is a lattice under the usual partial ordering of real valued functions, and contains the function $\phi(t) = 0$ for $t \leq a$, $\phi(t) = 1$ for $t > a$, for every real a . If a is real and X is in F , then the indicator random variable of the event $[X > a]$ is also in F . Property (8) then implies that $\mathcal{L}$ is a σ -lattice. It follows from the definition of $\mathcal{L}$ that $F \subset R(\mathcal{L})$. The reverse inclusion is shown using approximation by simple random variables, completing the proof of the theorem.

THEOREM 5. *Necessary and sufficient conditions that a family $M \subset L_2$ be $L_2(\mathcal{L})$ for some σ -lattice $\mathcal{L}$ containing $\emptyset$ and Ω are:*

- (11) M is a lattice closed in L_2 ;
- (12) a real, $X \in M$, $A = [X > a]$, $\mu(A) < \infty \Rightarrow \chi(A) \in M$;
- (12) a real, $X \in M$, $A = [X \geq a]$, $\mu(A^c) < \infty \Rightarrow -\chi(A^c) \in M$;
- (13) M is a convex cone.

If $\mu(\Omega) < \infty$, (12) may be replaced by (9) with F replaced by M , for then (11), (9) and (13) together imply (12). The special case of Theorem 5 in which $\mu(\Omega) < \infty$, " σ -lattice" is replaced by " σ -field" and "lattice" is replaced by "vector lattice" in (11) is due to Šidák [3].

PROOF OF THEOREM 5. The necessity is a consequence of well-known properties of L_2 and $R(\mathfrak{L})$. To prove the sufficiency, let $\mathfrak{L}^+$ denote the class of events A for which there exist $a \geq 0$ and $X \in M$ such that $A = [X > a]$, and $\mathfrak{L}^-$ the class of events B such that there exist $b \geq 0$ and $Y \in M$ such that $B = [Y \geq -b]$. It follows from (12) that $\chi(A) \in M$ if $a > 0$ and $-\chi(B^c) \in M$ if $b > 0$. From (11) and (13) it follows that if C is a countable union of events in $\mathfrak{L}^+$ then C is also a countable union of events $[X > a]$ for which $a > 0$, since $[X > 0] = \bigcup_n [X > 1/n]$. Further, if $a_i > 0$, and if $A_i = [X_i > a_i] \in \mathfrak{L}^+$, $i = 1, 2$, then $A_1 \cup A_2 = [(X_1/a_1) \vee (X_2/a_2) > 1] \in \mathfrak{L}^+$ by (11). It follows that if C is a countable union of events in $\mathfrak{L}^+$ then there are events $A_i \in \mathfrak{L}^+$ such that $\chi(A_i) \in M$, $A_i \subset A_{i+1}$, $i = 1, 2, \dots$, and such that $C = \bigcup_{i=1}^{\infty} A_i$. Define $c_n = \min_{1 \leq j \leq n} (1/j [\mu(A_j) - \mu(A_{j-1})]^{1/2})$ ($A_0 = \emptyset$) and

$$Z_n = \sum_{\nu=1}^n c_{\nu} [\chi(A_{\nu}) - \chi(A_{\nu-1})] = \sum_{\nu=1}^n \chi(A_{\nu}) (c_{\nu} - c_{\nu+1}) + \chi(A_n) c_{n+1},$$

$$Z = \sum_{\nu=1}^{\infty} c_{\nu} [\chi(A_{\nu}) - \chi(A_{\nu-1})].$$

Property (13) implies that $Z_n \in M$, $n = 1, 2, \dots$. Also

$$\int (Z - Z_n)^2 d\mu = \sum_{\nu=n+1}^{\infty} c_{\nu}^2 [\mu(A_{\nu}) - \mu(A_{\nu-1})] \rightarrow 0$$

as $n \rightarrow \infty$, so that by (11), $Z \in M$, and $C = [Z > 0] \in \mathfrak{L}^+$. Thus $\mathfrak{L}^+$ is closed under formation of countable unions. We note that if $A = [X > a]$, $X \in M$, $a > 0$, then $\chi(A) \in M$ by (12), so that also $A = [\chi(A) > 0]$; thus for every $A \in \mathfrak{L}^+$ there exists $X \in M$ such that $A = [X > 0]$. If $A_i \in \mathfrak{L}^+$, $A_i = [X_i > 0]$, $X_i \in M$, $i = 1, 2$, then $A_1 \cap A_2 = [X_1 \wedge X_2 > 0] \in \mathfrak{L}^+$ by (11). Thus $\mathfrak{L}^+$ is closed under the formation of finite intersections. If C is a countable intersection of events in $\mathfrak{L}^+$, there is therefore a sequence $\{C_n\}$ of events in $\mathfrak{L}^+$ such that $C_n \supset C_{n+1}$, $n = 1, 2, \dots$, and $C = \bigcap_{n=1}^{\infty} C_n$. If $C_1 = [X_1 > 0]$, $X_1 \in M$, set $A_i = [X_1 > 1/i]$, so that $A_i \in \mathfrak{L}^+$, $A_i \subset A_{i+1}$, $\bigcup_{i=1}^{\infty} A_i = C_1$, and $\chi(A_i) \in M$, $i = 1, 2, \dots$. If $C_n = [X_n > 0]$, $X_n \in M$, $X_n \geq 0$, then $A_i C_n = [\chi(A_i) \wedge X_n > 0]$; also $\mu(A_i) = \mu([X_1 > 1/i])$ so that (12) and (13) imply $\chi(A_i C_n) \in M$, $i = 1, 2, \dots$, $n = 1, 2, \dots$. Further, for fixed i , $\bigwedge_{n=1}^k \chi(A_i C_n)$ converges in L_2 as $k \rightarrow \infty$ to $\bigwedge_{n=1}^{\infty} \chi(A_i C_n) \in M$ by (11), so that $\bigcap_{n=1}^{\infty} A_i C_n = [\bigwedge_{n=1}^{\infty} \chi(A_i C_n) > 0] \in \mathfrak{L}^+$. Hence $C = \bigcup_{i=1}^{\infty} \bigcap_{n=1}^{\infty} A_i C_n \in \mathfrak{L}^+$. Thus $\mathfrak{L}^+$ is closed under formation of countable intersections. Similarly $\mathfrak{L}^-$ is a σ -lattice. Further, the following examination of cases demonstrates that $A \in \mathfrak{L}^+$, $B \in \mathfrak{L}^- \Rightarrow AB \in \mathfrak{L}^+$,

$A \cup B \in \mathfrak{L}^-$. CASE 1: $A = [\chi(A) = 1]$, $\chi(A) \in M$, $B = [-\chi(B^c) = 0]$, $-\chi(B^c) \in M$. Set $X = \chi(A) - 2\chi(B^c)$, then $AB = [X > 0] = [X = 1] \in \mathfrak{L}^+$ since $X \in M$. Set $Y = 2\chi(A) - \chi(B^c)$, then $A \cup B = [Y \geq 0] = [Y > -1] \in \mathfrak{L}^-$. CASE 2: $A = \bigcup_{i=1}^{\infty} A_i$, $A_i \subset A_{i+1}$, $A_i = [\chi(A_i) = 1]$, $\chi(A_i) \in M$, $i = 1, 2, \dots$; $B = \bigcap_{i=1}^{\infty} B_i$, $B_i \supset B_{i+1}$, $B_i = [-\chi(B_i^c) = 0]$, $-\chi(B_i^c) \in M$, $i = 1, 2, \dots$. Then $AB = \bigcup_{i=1}^{\infty} \bigcap_{j=1}^{\infty} A_i B_j \in \mathfrak{L}^+$ and $A \cup B = \bigcap_{j=1}^{\infty} \bigcup_{i=1}^{\infty} (A_i \cup B_j) \in \mathfrak{L}^-$. Thus $A \in \mathfrak{L}^+$, $B \in \mathfrak{L}^- \Rightarrow AB \in \mathfrak{L}^+$, $A \cup B \in \mathfrak{L}^-$. Set $\mathfrak{L} = \mathfrak{L}^+ \cup \mathfrak{L}^-$, then $\mathfrak{L}$ is a σ -lattice and $A \in \mathfrak{L}$ if and only if there exist real a and $X \in M$ such that $A = [X > a]$. Thus if $X \in M$ then $X \in L_2$ and $[X > a] \in \mathfrak{L}$ for every real a , so that $X \in L_2(L)$. Hence $M \subset L_2(\mathfrak{L})$, and the reverse inclusion can be shown using approximation by simple random variables, completing the proof of the theorem.

Finally, it is clear that the results of this paper make it possible to extend the results of [1] so as to omit the restriction $\mu(\Omega) < \infty$.

REFERENCES

1. H. D. Brunk, *Best fit to a random variable by a random variable measurable with respect to a σ -lattice*, Pacific J. Math. **11** (1961), 785-802.
2. I. J. Segal, *Equivalences of measure spaces*, Amer. J. Math. **73** (1951), 275-313.
3. Zbyněk Šidák, *On relations between strict sense and wide sense conditional expectations*, Teor. Verojatnost. i Primenen. **2** (1957), 283-288.

UNIVERSITY OF MISSOURI AND
UNIVERSITY OF CALIFORNIA, RIVERSIDE