CONDITIONAL EXPECTATIONS OF RANDOM VARIABLES WITHOUT EXPECTATIONS[1]

By R. E. Strauch[2]

University of California, Berkeley

1. Introduction and summary. Let $(\Omega, \mathfrak{F}, P)$ be a probability space, and let X be a random variable defined on $(\Omega, \mathfrak{F}, P)$. If $\mathfrak{A}$ is a sub σ-field of $\mathfrak{F}$, then $E(X \mid \mathfrak{A})$ is the a.s. unique $\mathfrak{A}$ measurable function such that, for all $A \in \mathfrak{A}$,

$$\int_A X\, dP = \int_A E(X \mid \mathfrak{A})\, dP, \tag{1}$$

provided EX is defined. ([2], p. 341). If EX is not defined, that is, if $EX^+ = EX^- = \infty$, we may then define $E(X \mid \mathfrak{A}) = E(X^+ \mid \mathfrak{A}) - E(X^- \mid \mathfrak{A})$, provided the difference is defined almost surely ([2], p. 342). We show that this is the only reasonable definition of $E(X \mid \mathfrak{A})$ (Lemma 2), and exhibit several apparent pathologies, akin to the fact that a conditionally convergent series of real numbers may be re-ordered to give any sum.

If X is any random variable with a continuous distribution such that EX is not defined, then we can find an $\mathfrak{A} \subset \mathfrak{F}$ such that $E(X \mid \mathfrak{A}) = 0$ a.s. (Theorem 1), and if Y is any random variable independent of X, we can find an $\mathfrak{A} \subset \mathfrak{F}$ such that $E(X \mid \mathfrak{A}) = Y$ a.s. (Theorem 2). In fact, if $X_1, X_2, \cdots$ is a sequence of independent random variables such that for $n \geqq 2$, EX_n is not defined and X_n has a continuous distribution, we can find a sequence of σ-fields $\mathfrak{A}_1 \subset \mathfrak{A}_2 \subset \cdots \subset \mathfrak{F}$ such that $X_1, \cdots, X_n$ are $\mathfrak{A}_n$ measurable and $E(X_{n+1} \mid \mathfrak{A}_n) = X_n$ a.s. (Theorem 3). The sequence $\{X_n, \mathfrak{F}_n, n = 1, 2, \cdots\}$ is not a martingale however, since for $m > n + 1$, $E(X_m \mid \mathfrak{F}_n)$ is not defined.

We remark that the standard theorem on iterated conditional expectations, ([2], p. 350) which says that if $\mathfrak{A} \subset \mathfrak{B} \subset \mathfrak{F}$ then $E(X \mid \mathfrak{A}) = E(E(X \mid \mathfrak{B}) \mid \mathfrak{A})$ a.s. is valid only if $E(X \mid \mathfrak{A})$ is defined a.s.

2. Preliminaries. Let X, Y, Z, with or without affixes, denote random variables, and let $\mathfrak{A}$, $\mathfrak{B}$, with or without affixes, denote sub σ-fields of $\mathfrak{F}$. $\mathfrak{B}(X_1, \cdots, X_n)$ is the smallest σ-field over which $X_1, \cdots, X_n$ are measurable, and if $\mathfrak{C}$ is any class of sets, $\sigma(\mathfrak{C})$ is the smallest σ-ring containing $\mathfrak{C}$.

Definition 1. Let Γ denote the class of random variables X such that X has a continuous distribution and $EX^+ = EX^- = \infty$.

Definition 2. If $X \in \Gamma$ and $\mathfrak{A} \subset \mathfrak{F}$, then

$$E(X \mid \mathfrak{A}) = E(X^+ \mid \mathfrak{A}) - E(X^- \mid \mathfrak{A}) \tag{2}$$

provided the difference is defined a.s.

Received 19 March 1965.

[1] This research was supported in part by a National Science Foundation Cooperative Graduate Fellowship and in part by the Office of Naval Research, Contract Number NONR-222-43.

[2] Now at the RAND Corp., Santa Monica, Calif.

LEMMA 1. *$E(X \mid \mathcal{A})$ is defined and finite* a.s. *if and only if there exists $\{A_n, n = 1, 2, \cdots\} \subset \mathcal{A}$ such that $\bigcup_{n=1}^{\infty} A_n = \Omega$ and $\int_{A_n} |X|\, dP < \infty$ for all n.*

PROOF. $E(X \mid \mathcal{A})$ is defined and finite a.s. if and only if both terms in the right hand side of (2) are defined and finite a.s. This happens if and only if the indefinite integrals of X^+ and X^-, and hence of $|X|$ are σ-finite on $\mathcal{A}$.

DEFINITION 3. A class of sets $\mathcal{C}$ is a semiring if $A, B \varepsilon \mathcal{C}$ implies $A \cap B \varepsilon \mathcal{C}$ and $A - B = \bigcup_{i=1}^{n} C_i$, with $C_i \varepsilon \mathcal{C}$ and $C_i \cap C_j = \emptyset$ for $i \neq j$, $1 \leqq i, j \leqq n$.

If $\mathcal{C}$ is a semiring, then $r(\mathcal{C})$, the collection of finite disjoint unions of elements of $\mathcal{C}$, is a ring, and $\sigma(\mathcal{C}) = \sigma(r(\mathcal{C}))$. Also $\sigma(\mathcal{C})$ is the smallest monotone class containing $r(\mathcal{C})$. (See [1], p. 22, problem 6; p. 26, problems 3 and 5; and p. 27. Our definition of semiring is slightly different from that of Halmos, but the proofs are the same.)

LEMMA 2. *Let $X \varepsilon \Gamma$, and suppose $\mathcal{C} \subset \mathcal{F}$ is a semiring such that $\sigma(\mathcal{C})$ is a σ-field, and for each $A \varepsilon \mathcal{C}$, $\int_A |X|\, dP < \infty$. Then $E(X \mid \sigma(\mathcal{C}))$ exists and is finite* a.s. *and if Z is any $\sigma(\mathcal{C})$ measurable random variable such that for all $A \varepsilon \mathcal{C}$, $\int_A Z\, dP = \int_A X\, dP$, then $Z = E(X \mid \sigma(\mathcal{C}))$* a.s.

PROOF. Since $\sigma(\mathcal{C})$ is a field, Ω is covered by a countable subset of $\mathcal{C}$, and the first statement follows from Lemma 1. To prove the second, assume $Z \neq E(X \mid \sigma(\mathcal{C}))$ a.s. Then we can find a $B \varepsilon r(\mathcal{C})$ such that $ZI_B \neq E(X \mid \sigma(\mathcal{C}))I_B$ a.s. But the class of sets A for which $\int_A ZI_B\, dP = \int_A XI_B\, dP$ contains $r(\mathcal{C})$, and is a monotone class, since $\int |XI_B|\, dP < \infty$ and $\int |ZI_B|\, dP < \infty$. Hence $ZI_B = E(XI_B \mid \sigma(\mathcal{C})) = E(X \mid \sigma(\mathcal{C}))I_B$ a.s., which is a contradiction.

In particular, if $X \varepsilon \Gamma$ and $\mathcal{A} \subset \mathcal{F}$ such that $E(X \mid \mathcal{A})$ is defined and finite a.s., then $\mathcal{C} = \{A \mid A \varepsilon \mathcal{A}, \int_A |X|\, dP < \infty\}$ is a ring and $\mathcal{A} = \sigma(\mathcal{C})$. Hence $E(X \mid \mathcal{A})$ is the a.s. unique $\mathcal{A}$ measurable random variable which satisfies (1) or all $A \varepsilon \mathcal{C}$.

3. The results. Theorem 1 is a corollary of Theorem 2, but is included as a separate theorem as the proof is somewhat simpler.

THEOREM 1. *If $X \epsilon \Gamma$, then there exists an $\mathcal{A} \subset \mathcal{B}(X)$ such that $E(X \mid \mathcal{A}) = 0$* a.s.

PROOF. For $a > 0$, we define

$$g_a(x) = \int_{\min(x,0)}^{a} t\, dF(t)$$

where F is the distribution function of X. For fixed a, g_a is a continuous monotone function of x, and $g(-\infty) = -\infty$. Let $I(0, a) = \{x \mid x \leqq a \text{ and } g_a(x) \geqq 0\}$ for $a > 0$, and let $I(0, 0) = \emptyset$. Then $I(0, a)$ is an interval about 0, and $\int_{I(0,a)} t\, dF(t) = 0$. For $a < b$, $I(0, a) \subset I(0, b)$, and we let $I(a, b) = I(0, b) - I(0, a)$. Then $\mathcal{C} = \{X^{-1}I(a, b) \mid 0 \leqq a < b\}$ is a semiring. For $A \varepsilon \mathcal{C}$, $A = X^{-1}I(a, b)$, $\int_A |X|\, dP < \infty$, and

$$\int_A X\, dP = \int_{I(0,b)} x\, dF(x) - \int_{I(0,a)} x\, dF(x) = 0.$$

Now set $\mathcal{A} = \sigma(\mathcal{C})$. Since $\Omega = \bigcup_{n=1}^{\infty} X^{-1}I(0, n) \varepsilon \mathcal{C}$, $\mathcal{A}$ is a σ-field, and by Lemma 2, $E(X \mid \mathcal{A}) = 0$ a.s.

COROLLARY. *If $X \varepsilon \Gamma$ and α is any finite constant, then there exists $\mathcal{A} \subset \mathcal{F}$ such that $E(X \mid \mathcal{A}) = \alpha$* a.s.

PROOF. Find $\mathcal{A}$ such that $E(X - \alpha \mid \mathcal{A}) = 0$ a.s.

THEOREM 2. *Let $X \varepsilon \Gamma$, and let Y be any random variable independent of X. Then there exists an $\mathcal{A}$ such that $\mathcal{B}(Y) \subset \mathcal{A} \subset \mathcal{B}(X, Y)$ and $E(X \mid \mathcal{A}) = Y$ a.s.*

PROOF. For fixed $a > 0$, the function

$$g_a(x, y) = \int_{\min(x,y)}^{y+a} (t - y)\, dF(t)$$

is a continuous hence Borel measurable function of x and y. For each y, $g_a(\cdot, y)$ is a continuous, monotone function of x, and $g_a(-\infty, y) = -\infty$. For $a > 0$, $y_1 < y_2$, let

$$I(0, a, y_1, y_2) = \{(x, y) \mid y_1 \leqq y < y_2 ; x \leqq y + a; g_a(x, y) \geqq 0\},$$

and let $I(0, 0, y_1, y_2) = \varnothing$.

For fixed y, the section $I_y(0, a, y_1, y_2)$ is an interval (empty if $y \not\varepsilon [y_1, y_2]$), and

$$\int_{I_y(0,a,y_1,y_2)} (x - y)\, dF(x) = 0.$$

For $a < b$, $I(0, a, y_1, y_2) \subset I(0, b, y_1, y_2)$, and we define

$$I(a, b, y_1, y_2) = I(0, b, y_1, y_2) - I(0, a, y_1, y_2),$$

$$\mathcal{C} = \{(X, Y)^{-1} I(a, b, y_1, y_2) \mid 0 \leqq a < b, y_1 < y_2\}.$$

Then $\mathcal{C}$ is a semiring, and $\mathcal{A} = \sigma(\mathcal{C})$ is a σ-field. For any $A \varepsilon \mathcal{C}$, $\int_A |X|\, dP < \infty$ and $\int_A X\, dP = \int_A Y\, dP$. But Y is $\mathcal{A}$ measurable, hence by Lemma 2, $E(X \mid \mathcal{A}) = Y$ a.s.

This is quite different from the situation when EX is defined. If $E(X \mid \mathcal{A}) = Y$ a.s. and EX is defined, then $E(X \mid \mathcal{B}(Y)) = Y$ a.s. and thus if Y is independent of X, then $Y = EX$ a.s. In our case, however, $E(X \mid \mathcal{B}(Y))$ is undefined, since $E(X^+ \mid \mathcal{B}(Y)) = E(X^- \mid \mathcal{B}(Y)) = \infty$ a.s.

We can easily extend this result to a sequence of independent random variables to obtain

THEOREM 3. *Let $X_1, X_2, \cdots$ be a sequence of independent random variables such that, for $n \geqq 2$, $X_n \varepsilon \Gamma$. Then there exists a sequence of σ-fields $\mathcal{A}_1, \mathcal{A}_2, \cdots$ such that $\mathcal{B}(X_1, \cdots, X_n) \subset \mathcal{A}_n \subset \mathcal{B}(X_1, \cdots, X_{n+1})$ and $E(X_{n+1} \mid \mathcal{A}_n) = X_n$ a.s.*

PROOF. According to Theorem 2, for each n we can find $\mathcal{B}_n$ such that $\mathcal{B}(X_n) \subset \mathcal{B}_n \subset \mathcal{B}(X_n, X_{n+1})$ and $E(X_{n+1} \mid \mathcal{B}_n) = X_n$ a.s. Let $\mathcal{A}_n = \sigma(\mathcal{B}_n \cup \mathcal{B}(X_1, \cdots, X_{n-1}))$. Then $\mathcal{B}(X_1, \cdots, X_n) \subset \mathcal{A}_n \subset \mathcal{B}(X_1, \cdots, X_{n+1})$ and

$$\begin{aligned} E(X_{n+1} \mid \mathcal{A}_n) &= E(X_{n+1}^+ \mid \mathcal{A}_n) - E(X_{n+1}^- \mid \mathcal{A}_n) && \text{a.s.} \\ &= E(X_{n+1}^+ \mid \mathcal{B}_n) - E(X_{n+1}^- \mid \mathcal{B}_n) && \text{a.s.} \\ &= E(X_{n+1} \mid \mathcal{B}_n) && \text{a.s.} \\ &= X_n && \text{a.s.} \end{aligned}$$

The second equality follows from the fact that EX_{n+1}^+ and EX_{n+1}^- are defined, and X_{n+1}^+ and X_{n+1}^- are independent of $\mathcal{B}(X_1, \cdots, X_{n-1})$.

If we consider random variables with discrete distributions, the theorems may fail, as is shown by the following example. Let $\Omega = \{1, 2, \cdots\}$, let $\mathfrak{F}$ be all subsets of Ω, and let $P\{j\} = 2^{-j}$. Let $X(j) = 2^j$ if j is odd and $-\pi 2^j$ if j is even. Let A be any set for which $\int_A |X| \, dP < \infty$. Then A is finite, and $\int_A X \, dP = m - n\pi \neq 0$ for some integers m and n. Thus there does not exist an $\mathfrak{A} \subset \mathfrak{F}$ such that $E(X \mid \mathfrak{A}) = 0$ a.s.

Slightly weaker results do hold in general however. We can no longer start with a given probability space and random variables on it, and construct conditioning σ-fields within the σ-fields generated by the random variables. We can, however, start with given distributions, and construct a probability space, random variables on it with the required distributions, and conditioning σ-fields such that the resulting conditional expectations have the desired properties. For example Theorem 2 would become

THEOREM 2'. *Let F and G be probability distribution functions such that the mean of F is not defined. Then there exists a probability space $(\Omega, \mathfrak{F}, P)$, random variables X with distribution function F and Y with distribution function G, and a σ-field $\mathfrak{A} \subset \mathfrak{F}$ such that $\mathfrak{B}(Y) \subset \mathfrak{A}$ and $E(X \mid \mathfrak{A}) = Y$* a.s.

The method of proof is as follows. We construct a probability space $(\Omega, \mathfrak{F}, P)$ and random variables X, Y, and Z such that X and Y have the required distribution functions, Y and Z are independent, Z has a continuous distribution, and $X = f(Z)$, where f is a non-decreasing function such that $f(0) = 0$.

We then define, for $a > 0$

$$g_a(z, y) = \int_{\min(z,0)}^{a} (f(t) - y) \, dF'(t)$$

where F' is the distribution function of Z. The remainder of the proof proceeds as in Theorem 2.

REFERENCES

[1] HALMOS, PAUL R. (1950). *Measure Theory.* Van Nostrand, Princeton.
[2] LOÈVE, MICHEL (1963). *Probability Theory* (3rd ed.). Van Nostrand, Princeton.