ON CONDITIONAL EXPECTATION AND QUASI-RINGS

M. V. JOHNS, JR. AND RONALD PYKE

1. Introduction. Let $(\Omega, \mathscr{A}, P)$ denote a complete probability space in which Ω is an arbitrary point set $(\omega \in \Omega)$, $\mathscr{A}$ is a σ-algebra of subsets of Ω $(A \in \mathscr{A})$ and P is a probability measure on $\mathscr{A}$ with respect to which P is complete. Let X, Y, Z, with or without subscripts, denote real-valued $\mathscr{A}$-measurable random variables (r. v.) Let $\mathscr{E}$ denote the space of P-integrable r. v.'s. Define a linear operator E on $\mathscr{E}$ by

$$E \circ X = \int_\Omega X\,dP\,.$$

E is the expectation operator and $E \circ X$ is called the expectation of X. The P-integrability criterion is equivalent to specifying $E \circ |X| < \infty$. Let $\mathscr{F}$, with or without subscripts, denote a complete σ-algebra contained in $\mathscr{A}$, and let $\mathscr{B}_k$ denote the σ-algebra of Borel sets of k-dimensional Euclidean space. For r.v.'s. $i=1, X_i, \cdots, k$, define $\mathscr{B}(X_1, \cdots, X_k) \subset \mathscr{A}$ as the minimal complete σ-algebra containing all inverse images with respect to the vector $(X_1, \cdots, X_k)$ of sets in $\mathscr{B}_k$. For $A \in \mathscr{A}$, let $I_A \in \mathscr{E}$ denote the indicator function of the set A; that is, $I_A(\omega)=1$ or 0 according as $\omega \in A$ or $\omega \notin A$. For $X \in \mathscr{E}$, define the completely-additive set function $Q_X : \mathscr{A} \to R_1$ by $Q_X(A) = E \circ XI_A$.

By the Radon-Nikodym Theorem there exists for $X \in \mathscr{E}$ and $\mathscr{F} \subset \mathscr{A}$, an $\mathscr{F}$-measurable solution $Y \in \mathscr{E}$ to the system of equations

$$(1) \qquad E \circ (X - Y)I_A = 0 \qquad (A \in \mathscr{F})$$

or equivalently

$$Q_X(A) = E \circ YI_A \qquad (A \in \mathscr{F})\,.$$

This solution is unique a. s. (relative to the restriction of P to $\mathscr{F}$). The equivalence class of all such solutions (or any representative thereof) is denoted by $E\{X|\mathscr{F}\}$ and called the conditional expectation of X given $\mathscr{F}$. For $X, Y \in \mathscr{E}$ the notation $E\{X|Y\} = E\{X|\mathscr{B}(Y)\}$ will also be used. This definition of conditional expectation, which is the standard one, makes it necessary when proving theorems about conditional expectations to show at some stage of the proof that a functional equation of the form (1) is valid for all subsets of a specified σ-algebra. That this can be a tedious task is demonstrated by the existing proofs of some of the applications in § 4 of the theorems which are proved below.

Received December 23, 1958. This work was supported by the Office of Naval Research under Contract Nonr-225 (21).

It is the purpose of this note to define conditional expectations in an apparently less restrictive way, by narrowing the class of subsets A for which (1) must hold. It is shown that this definition is, nevertheless, equivalent to that given in the above paragraph. In § 3, some general theorems on conditional expectations are proved using this second definition. The proofs of these theorems are seen to be simpler and shorter than would be possible with conventional techniques. Besides serving to demonstrate the convenience of this second definition, these theorems are important in themselves and several applications of them are given.

The main tool to be used is the concept of a quasi-ring to be introduced and studied in the following section.

2. Conditional expectation given a quasi-ring. Von Neumann [5] defines a half-ring as a family of subsets closed under finite intersections and satisfying a certain finite chain condition. This same concept is termed a semi-ring by Halmos [3]. The related concept of quasi-ring, which is now defined, entails a weaker chain condition. This chain condition, (ii) of Definition 1 below, seems to be much more adaptable than that of von Neumann to problems in conditional expectation, as is demonstrated in § 3.

DEFINITION 1. A collection, $\mathscr{S}$, of subsets of Ω is said to be a *quasi-ring* if and only if

(i) $A, B \in \mathscr{S}$ implies $A \cap B \in \mathscr{S}$;

(ii) $A, B \in \mathscr{S}$ and $A \subset B$ implies that there exists $\{C_j\}_{j=1}^{n} \subset \mathscr{S}$ satisfying $C_i \cap C_k = \phi$ for $i \neq k$ and $B - A = C_1 \cup C_2 \cup \cdots \cup C_n$;

(iii) there exists $\{A_j\}_{j=1}^{\infty} \subset \mathscr{S}$ such that $\Omega = \bigcup_{j=1}^{\infty} A_j$.

In von Neumann's definition of a half-ring, condition (ii) is strengthened by requiring further that $A \cup C_1 \cup \cdots \cup C_j \in \mathscr{S}$ for all $j = 1, 2, \cdots, n$.

Examples of quasi-rings are: any countable class of disjoint sets which include the null set ϕ; in particular, the collection of atoms in an atomic, or discrete, probability space; any algebra or σ-algebra; the class of all left-open, right-closed rectangles in R_n with Lebesgue measure less than or equal to 1. This last example is a quasi-ring which is not a half-ring. Bell makes use of the half-ring analogous to this quasi-ring in his recent paper [1]. A closure property of quasi-rings that will be used in the following sections is given by

LEMMA 1. *If $\mathscr{S}_1$ and $\mathscr{S}_2$ are quasi-rings on a common space Ω then*

$$\mathscr{S} = \mathscr{S}_1 \overset{*}{\cap} \mathscr{S}_2 \equiv \{A \cap B;\ A \in \mathscr{S}_1, B \in \mathscr{S}_2\} \tag{2}$$

is also a quasi-ring. (In common terminology $\mathscr{S}$ is the family of constituents of $\mathscr{S}_1$ and $\mathscr{S}_2$.)

Proof. Clearly $\mathscr{S}$ satisfies (i) of Definition 1. Moreover, let $A_i \in \mathscr{S}_1$ and $B_i \in \mathscr{S}_2 (i = 1, 2)$. If $A_1 \cap B_1 \subset A_2 \cap B_2$, then

$$(3) \qquad \begin{aligned} S &\equiv (A_2 \cap B_2) - (A_1 \cap B_1) \\ &= [(A_2 - A_1) \cap (B_2 \cap B_1)] \cup [(B_2 - B_1) \cap A_2] , \end{aligned}$$

the two terms of the union being disjoint. By hypothesis there exist sequences $\{C_j\}_{j=1}^{n} \in \mathscr{S}_1$, $\{D_k\}_{k=1}^{m} \in \mathscr{S}_2$ satisfying

$$A_2 - A_2 \cap A_1 = \bigcup_{j=1}^{n} C_j, \quad B_2 - B_2 \cap B_1 = \bigcup_{k=1}^{m} D_k$$

and hence by (3), S has the representation

$$S = \bigcup_{j=1}^{x^n} (C_j \cap [B_2 \cap B_1]) \cup \bigcup_{k=1}^{m} (D_k \cap A_2)$$

all terms being disjoint. That $\mathscr{S}$ satisfies condition (iii) is seen by considering the collection of all pairwise intersections between elements of the respective sequences for $\mathscr{S}_1$ and $\mathscr{S}_2$ which satisfy (iii). Q. e. d.

An extension theorem for measures defined on a quasi-ring will now be given. The proof of the theorem is analogous to those of the more classical extension theorems and so will be omitted (e. g., cf. [5]).

For an arbitrary class $\mathscr{C}$ of subsets of Ω let $\sigma(\mathscr{C})$ denote the minimal σ-algebra containing $\mathscr{C}$.

THEOREM 1. *Let μ be a σ-finite completely additive set function defined on a quasi-ring $\mathscr{S}$. There exists a unique completely additive set function μ^* defined on $\sigma(\mathscr{S})$ such that for all $A \in \mathscr{S}$, $\mu^*(A) = \mu(A)$.*

In the event that there exists a finite family satisfying (iii) of Definition 1, the minimal algebra containing $\mathscr{S}$ is the collection of all finite unions of members of $\mathscr{S}$. After extending μ to this minimal algebra, Theorem 1 reduces in this case to a well known extension theorem (cf. Doob [2], p. 605).

DEFINITION 2. Let $X \in \mathscr{E}$ and $\mathscr{S} \subset \mathscr{A}$ where $\mathscr{S}$ is a quasi-ring. The class (or any representative thereof) of all $\sigma(\mathscr{S})$-measurable $Y \in \mathscr{E}$ satisfying the system of equations

$$(4) \qquad E \circ (X - Y) I_A = 0 \qquad (A \in \mathscr{S})$$

will be denoted by $E\{X \mid \mathscr{S}\}$, and called the conditional expectation of X given $\mathscr{S}$.

As a corollary to Theorem 1, one immediately obtains

THEOREM 2. *For $X \in \mathscr{E}$ and $\mathscr{S} \subset \mathscr{A}$*

$$E\{X \mid \mathscr{S}\} = E\{X \mid \sigma(\mathscr{S})\} \qquad \text{a. s.}$$

3. Some general theorems on conditional expectation. The following definition will be used:

DEFINITION 3. Quasi-rings $\mathscr{S}_1$ and $\mathscr{S}_2$ are said to be *conditionally independent* given a quasi-ring $\mathscr{S}$ (to be abbreviated as c. i. $\mid \mathscr{S}$) if and only if for all $A \in \mathscr{S}_1, B \in \mathscr{S}_2$,

$$(5) \qquad E\{I_A I_B \mid \mathscr{S}\} = E\{I_A \mid \mathscr{S}\} E\{I_B \mid \mathscr{S}\} \qquad \text{a. s.}$$

X and Y are said to be c. i. $\mid \mathscr{S}$ if and only if $\mathscr{B}(X)$ and $\mathscr{B}(Y)$ are c. i. $\mid \mathscr{S}$ (cf. Loève [4], p. 351).

The obvious notational changes are made in defining conditional independence given a r. v. If $\mathscr{S}_1$ and $\mathscr{S}_2$ are c. i. $\mid \{\phi, \Omega\}$, they are of course, independent in the usual stochastic sense. The above definition of conditional independence is closely related to that for σ-algebras given in Loève [4], as is shown by the next lemma. For well known properties of conditional expectations used in the following proofs, the reader is referred to [4].

LEMMA 2. *For* $\sigma(\mathscr{S}_1)$ *and* $\sigma(\mathscr{S}_2)$ *to be* c. i. $\mid \sigma(\mathscr{S})$ *it is necessary and sufficient that* $\mathscr{S}_1$ *and* $\mathscr{S}_2$ *be* c. i. $\mid \mathscr{S}$.

Proof. The necessity of the condition is immediate. The proof of sufficiency is by transfinite induction. Let $\mathscr{L}_1$ denote the class of all countable unions of elements of $\mathscr{S}_1$. For all ordinals α less than or equal to the first uncountable ordinal, α_0 say, define recursively $\mathscr{L}_\alpha$ as the set of countable unions of differences of elements of $\mathscr{T}_\alpha \equiv \bigcup_{\beta<\alpha} \mathscr{L}_\beta$. It is well known that $\sigma(\mathscr{S}_1) = \mathscr{T}_{\alpha_0}$. By hypothesis the equality (5) holds for all $A \in \mathscr{S}_1$ and $B \in \mathscr{S}_2$. Since $\mathscr{S}_1$ is closed under finite intersections, any countable union of elements in $\mathscr{S}_1$, and hence by definition any element of $\mathscr{L}_1$ may be represented as a disjoint union of elements in $\mathscr{S}_1$. Therefore, since conditional expectations have (a. s.) the linear and limit properties of integrals, it follows that (5) holds for all $A \in \mathscr{L}_1$ Clearly $\mathscr{L}_1$ is also closed under finite intersections. For induction purposes, assume that for any ordinal $\alpha < \alpha_0$, $\mathscr{T}_\alpha$ satisfies (5) and is closed under finite intersections. It is clear that (5) holds for differences of elements in $\mathscr{T}_\alpha$. For if $C, D \in \mathscr{T}_\alpha$, $C - D = C - (C \cap D)$, and since by assumption $C \cap D \in \mathscr{T}_\alpha$, (5) follows by writing $I_{C-D} = I_C - I_{C\cap D}$. Moreover, countable unions of elements of $\mathscr{T}_\alpha$ may be shown to satisfy (5) in the same way as was used above for $\mathscr{L}_1$. Therefore (5) is satisfied for all elements of $\mathscr{L}_{\alpha+1}$ and hence of $\mathscr{T}_{\alpha+1}$. From the identity $(A - B) \cap (C - D) = (A \cap C) - (B \cup D)$, it follows that $\mathscr{L}_{\alpha+1}$ and hence

$\mathscr{T}_{\alpha+1}$ is closed under finite intersection. It therefore follows by transfinite induction that (5) holds for all $A \in \sigma(\mathscr{S}_1)$ and $B \in \mathscr{S}_2$. The lemma follows by a repetition of the above argument for $\mathscr{S}_2$.

It is remarked that if there exists a conditional probability distribution relative to $\sigma(\mathscr{S})$ in the sense of Doob [2], the conditional expectations of (5) may be considered as integrals with respect to the distribution. In this case one might be tempted to view Lemma 2 as a simple extension of measures, and hence as a corollary to Theorem 1. Closer examination shows this to be a false supposition.

LEMMA 3. *For* $X, Y \in \mathscr{E}$, *let* X *and* Y *be* c. i. $| \mathscr{F}$. *Then if* $XY \in \mathscr{E}$

$$E\{XY \mid \mathscr{F}\} = E\{X \mid \mathscr{F}\}E\{Y \mid \mathscr{F}\} \qquad \text{a. s.}$$

Proof. This result follows from (5) upon approximating X and Y by simple functions in the usual way. The assumption that $XY \in \mathscr{E}$ is certainly not a necessary one but has been postulated in keeping with Definition 2.

The main theorem of this paper is

THEOREM 3. *Let* $X \in \mathscr{E}$ *and* $\mathscr{F}_i \subset \mathscr{A}(i = 1, 2)$ *be given. If* $\mathscr{B}(X)$ *and* $\mathscr{F}_2$ *are* c. i. $| \mathscr{F}_1$ *then*

$$E\{X \mid \mathscr{F}_1 \overset{*}{\cap} \mathscr{F}_2\} = E\{X \mid \mathscr{F}_1\} \qquad \text{a. s.} \tag{6}$$

Proof. Define $\mathscr{S} = \mathscr{F}_1 \overset{*}{\cap} \mathscr{F}_2$. $\mathscr{S}$ is a quasi-ring by Lemma 1. From Theorem 2, (4), and the fact that $E\{X \mid \mathscr{F}_1\}$ is $\sigma(\mathscr{S})$-measurable, it follows that to prove (6) it suffices to show that

$$E \circ XI_S = E \circ E\{X \mid \mathscr{F}_1\}I_S \qquad \text{a. s.}$$

for all $S \in \mathscr{S}$. Let $S = A \cap B$ for $A \in \mathscr{F}_1, B \in \mathscr{F}_2$. Then

$$\begin{aligned} E \circ XI_{A\cap B} &= E \circ E\{XI_B \mid \mathscr{F}_1\}I_A && \text{a. s.} \\ &= E \circ E\{X \mid \mathscr{F}_1\}E\{I_B \mid \mathscr{F}_1\}I_A && \text{a. s.} \end{aligned}$$

since X and I_B are c. i. $| \mathscr{F}_1$. Therefore

$$\begin{aligned} E \circ XI_{A\cap B} &= E \circ E\{I_B E\{X \mid \mathscr{F}_1\} \mid \mathscr{F}_1\}I_A && \text{a. s.} \\ &= E \circ E\{X \mid \mathscr{F}_1\}I_{A\cap B} && \text{a. s.} \end{aligned}$$

by (1). *Q.* e. d.

COROLLARY 3.1. *Let* $X \in \mathscr{E}$ *and let* X *and* Z *be* c. i. $| Y$. *Then*

$$E\{X \mid Y, Z\} = E\{X \mid Y\} \qquad \text{a. s.} \tag{7}$$

It is of interest to state this result under the stronger but more common assumption of independence, viz.,

COROLLARY 3.2. *For* $X \in \mathscr{E}$, *let the random vector* (X, Y) *be independent of* Z. *Then* (7) *holds.*

Proof. This is *a* consequence of the fact that (X, Y) being independent of Z implies that X and Z are c. i. $\mid Y$. To see this, consider

$$\begin{aligned} E\{I_{A\cap B} \mid Y\} &= E\{E\{I_{A\cap B} \mid Y, X\} \mid Y\} = E\{I_A E\{I_B \mid Y, X\} \mid Y\} \quad \text{a. s.} \\ &= E\{I_A \mid Y\}E\{I_B\} = E\{I_A \mid Y\}E\{I_B \mid Y\} \quad \text{a. s.} \end{aligned}$$

where $A \in \mathscr{B}(X)$, $B \in \mathscr{B}(Z)$.

It should be noted that Corollaries 3.1 and 3.2 remain valid if the random variables Y and Z are replaced by random functions since the proofs depend only on the properties of the corresponding σ-algebras.

Before stating a generalization of Theorem 3, we prove the following lemma:

LEMMA 4. *If* $\mathscr{F}_2$ *and* $\mathscr{F}_3$ *are* c. i. $\mid \mathscr{F}_1$, *then* $\mathscr{F}_1 \overset{*}{\cap} \mathscr{F}_2$ *and* $\mathscr{F}_1 \overset{*}{\cap} \mathscr{F}_3$ *are* c. i. $\mid \mathscr{F}_1$.

Proof. Let $A_i \in \mathscr{F}_i (i = 1, 2, 3)$ and $B_1 \in \mathscr{F}_1$. Then

$$\begin{aligned} E\{I_{A_1\cap A_2} I_{B_1\cap A_3} \mid \mathscr{F}_1\} &= I_{A_1} I_{B_1} E\{I_{A_2} I_{A_3} \mid \mathscr{F}_1\} \quad \text{a. s.} \\ &= I_{A_1} E\{I_{A_2} \mid \mathscr{F}_1\} I_{B_1} E\{I_{A_3} \mid \mathscr{F}_1\} \quad \text{a. s.} \\ &= E\{I_{A_1\cap A_2} \mid \mathscr{F}_1\} E\{I_{B_1\cap A_3} \mid \mathscr{F}_1\} \quad \text{a. s.} \end{aligned}$$

by hypothesis and lemma follows.

THEOREM 4. *Let* $Y \in \mathscr{E}$ *and* $\mathscr{F}_i \subset \mathscr{A}$ $(i = 1, 2, 3)$ *be given. If* $\mathscr{B}(Y) \subset \sigma(\mathscr{F}_1 \cup \mathscr{F}_2)$ *and if* $\mathscr{F}_2$ *and* $\mathscr{F}_3$ *are* c. i. $\mid \mathscr{F}_1$, *then*

$$E\{Y \mid \mathscr{F}_1 \overset{*}{\cap} \mathscr{F}_3\} = E\{Y \mid \mathscr{F}_1\} \quad \text{a. s.} \tag{8}$$

Proof. By Lemma 4 it follows that $\mathscr{F}_1 \overset{*}{\cap} \mathscr{F}_2$ and $\mathscr{F}_3$ are c. i.$\mid \mathscr{F}_1$. Therefore, (8) becomes a consequence of Theorem 3 since $\mathscr{F}_1 \overset{*}{\cap} \mathscr{F}_2$ and $\mathscr{F}_3$ being c. i. $\mid \mathscr{F}_1$ implies that $\mathscr{B}(Y)$ and $\mathscr{F}_3$ are c. i. $\mid \mathscr{F}_1$.

Of particular importance is the following special case of the above theorem:

COROLLARY 4.1. *Let* $g: R_2 \to R_1$ *be a* $\mathscr{B}_2$*-measurable function, and* r. v.'s X, Y, Z *be such that* $g(X, Y) \in \mathscr{E}$, *and either* X *and* Z *are* c. i. $\mid Y$ *or the vector* (X, Y) *is independent of* Z. *Then*

$$E\{g(X, Y) \mid Y, Z\} = E\{g(X, Y) \mid Y\} \qquad \text{a. s.}$$

As before, this result remains valid if the random variables X, Y and Z are replaced by random functions.

It should be remarked that many of the foregoing results may be obtained by elementary means for cases where the random variables involved possess joint probability density functions with respect to some dominating measure. In many applications, however, the existence of such density functions cannot be postulated.

4. Applications. As a first application of the results of § 3, the following theorem shows the equivalence of certain characterizations of conditional indepedence :

THEOREM 5. *For* r. v.'s X, Y, Z, *the following statements are equivalent* :

(a) Z *and* X *are* c. i. $\mid Y$

(b) Z-Y *and* X-Y *are* c. i. $\mid Y$

(c) $P\{Z \leqq z \mid Y, X\} = P\{Z \leqq z \mid Y\}$ a. s. *for all* $z \in R_1$.

Proof. (Note first the standard definition $P\{A \mid \mathscr{F}\} \equiv E\{I_A \mid \mathscr{F}\}$ which has been presupposed in (c).) Lemma 4 shows that $(a) \rightarrow (b)$. Since $\mathscr{B}(Z) \subset \mathscr{B}(Y, Z - Y)$ and $\mathscr{B}(Y, X) = \mathscr{B}(Y, X - Y)$ it follows from Theorem 4 that (b) $\rightarrow$ (c). (c) implies that $E\{I_A \mid Y, X\} = E\{I_A \mid Y\}$ for all A of the form $\{z_1 < Z \leqq z_2\}$ with $z_1, z_2 \in R_1$. The collection of all such inverse images forms a quasi-ring, $\mathscr{S}$, say, such that $\sigma(\mathscr{S}) = \mathscr{B}(Z)$. It follows then that for $A \in \mathscr{S}$, $B \in \mathscr{B}(X)$,

$$E\{I_A I_B \mid Y\} = E\{I_B E\{I_A \mid Y, X\} \mid Y\} = E\{I_A \mid Y\} E\{I_B \mid Y\} \quad \text{a. s.}$$

and (a) follows by Lemma 2. Q. e. d.

The equivalence of (a) and (c) has been proved in a different form by Doob ([2], pp. 83–85) for the more general case in which Z and X are allowed to be finite-dimensional random vectors. It should be pointed out that the restriction to one-dimensional r. v.'s was solely for presentation purposes throughout this paper, and that all of the above results carry through when the conditioning r. v.'s are replaced by arbitrary families of r.v.'s. This is true simply because all results involving r. v.'s have been stated in terms of their induced σ-algebras. Roughly speaking, in this more general context, the implication (c)$\rightarrow$(a) of Theorem 5 states that for a Markov process the past and future are c. i. given the present.

A second application is in proving the statement that a stochastic process $\{X_t : t \in T\}$ with independent increments is a Markov process. Indeed this statement is a simple corollary of Theorem 4. For $t_1 < t_2 <$

$\cdots < t_n$, consider

$$P\{X_{t_n} \leqq x \mid X_{t_1}, X_{t_2}, \cdots, X_{t_{n-1}}\} = P\{(x_{t_n} - X_{t_{n-1}}) + X_{t_{n-1}} \leqq x \mid X_{t_{n-1}}, (X_t, \cdots, X_{t_{n-2}})\} \quad \text{a. s.}$$
$$= P\{X_{t_n} \leqq x \mid X_{t_{n-1}}\} \quad \text{a. s.}$$

The last equality is a consequence of the remark following Corollary 4.1, since $X_{t_n} - X_{t_{n-1}}$ and $(X_{t_1}, \cdots, X_{t_{n-2}})$ are independent. A proof of this fact, using only the standard theorems of conditional expectation, is lengthy and rather unattractive (cf., Doob [2], p. 85).

REFERENCES

1. C. B. Bell, *On the structure of stochastic independence*, Illinois J. Math., **2** (1958), 415-424.
2. J. L. Doob, *Stochastic processes*, J. Wiley and Sons, 1953.
3. P. Halmos, *Measure theory*, Van Nostrand, 1950.
4. M. Loève, *Probability theory*, Van Nostrand, 1955.
5. J. von Neumann, *Functional operators*, Vol. 1, Ann. of Math. Studies No. 21, Princeton University Press, 1950.

STANFORD UNIVERSITY