PACIFIC JOURNAL OF MATHEMATICS
Vol. 15, No. 3, 1965

A CHARACTERIZATION OF CONDITIONAL PROBABILITY

MILTON PHILIP OLSON

The conditional probability functions relative to a sub-σ-field $\mathscr{B}$ are shown to constitute a vector-valued measure on the σ-field of the probability space, and it is proved that the conditional expectation relative to $\mathscr{B}$ of an $\mathscr{L}_1$ random variable X is the integral of X with respect to this vector-valued measure. A complete characterization is given of those vector-valued measures which are conditional probabilities. This machinery is illustratively applied to give alternative derivations of results of Moy and Rota on the characterization of conditional expectation operators.

Probabilists often regret that, in general, conditional probabilities do not define probability measures almost everywhere. This defect arises naturally from the fact that conditional probabilities are Radon-Nikodym derivatives of certain set functions; and, hence, they are defined only up to equivalence. It seems advantageous to relinquish the concept of conditional probabilities as point functions, almost everywhere determined, and consider them as they are—elements of a function space. Part of the awkwardness of conditional probabilities then disappears: The conditional probabilities form a vector-valued measure such that conditional expectation of an integrable function is its integral with respect to this conditional probability measure.

Throughout this paper $(\Omega, \mathscr{A}, \mu)$, will be a fixed probability space and $\mathscr{B}$ a fixed sub-σ-field of $\mathscr{A}$. $\mathscr{L}_1(\Omega, \mathscr{A}, \mu)$, or simply $\mathscr{L}_1$, denotes the Banach space of all complex-valued, μ-integrable, $\mathscr{A}$-measurable functions on Ω. $\mathscr{E}^{\mathscr{B}}$ will be the conditional expectation operator in $\mathscr{L}_1$ relative to $\mathscr{B}$. I_A is the indicator of the set A.

In § 1, we show that if $\varphi^{\mathscr{B}}(A) = \mathscr{E}^{\mathscr{B}} I_A$ for $A \in \mathscr{A}$, then $\phi^{\mathscr{B}}$ is a vector-valued measure on $\mathscr{A}$ with values in $\mathscr{L}_1$. Furthermore, we show that if $X \in \mathscr{L}_1$, then

$$(0.1) \qquad \mathscr{E}^{\mathscr{B}} X = \int X d\varphi^{\mathscr{B}} ,$$

the notion of integral here being defined as in [1, p. 323].

Over the last ten years, several characterizations of $\mathscr{E}^{\mathscr{B}}$ have appeared. The papers of Moy [3] and Rota [4] characterized $\mathscr{E}^{\mathscr{B}}$ in

Received March 5, 1964. This paper was prepared with the partial support of the Office of Naval Research, Contract Nonr-222 (43). This paper in whole or in part may be reproduced for any purpose of the United States Government.

terms of its properties as a linear operator; Sidak [5] characterized it by means of its action on $\mathscr{L}_2$. In § 2, our characterization of conditional probability will be completed by identifying those countably additive measures on $\mathscr{A}$ with values in $\mathscr{L}_1$ which are conditional probabilities. For such a measure, a formula like (0.1) will hold, and the corresponding sub-σ-field will be determined.

In § 3, this characterization of conditional probability will be used to obtain the characterization of conditional expectation of Moy [3] and of Rota [4] using a theorem on the representation of operators on a Lebesgue space. In a subsequent publication, further results in [3] and [4] as well as some in [5] will be obtained as applications of our characterization of conditional probabilities.

The author wishes to express his appreciation to Professor Edward W. Barankin for continued inspiration and encouragement in this line of thought, as well as for numerous simplifications and technical improvements in the results.

1. Representation of conditional expectation as an integral with respect to a vector-valued measure. Let the notational conventions introduced above continue to hold. By definition [1, p. 318], a function ψ on the σ-field $\mathscr{A}$ to $\mathscr{L}_1$ is called a vector-valued measure if $\psi(\phi) = 0$ and ψ is countably additive on disjoint sets belonging to $\mathscr{A}$. If $\{A_i\}_{i=1}^{\infty} \subset \mathscr{A}$ such that $A_i \cap A_j = \phi$ when $i \neq j$, then we have from properties of conditional expectations (see [2, p. 347]),

$$\varphi^{\mathscr{B}}\left(\sum_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \varphi^{\mathscr{B}}(A_i) \qquad \text{a.s.} \tag{1.1}$$

Since the partial sums $\sum_{i=1}^{n} \varphi^{\mathscr{B}}(A_i)$ are uniformly bounded a.s. by 1, it follows that their integrals are uniformly continuous with respect to μ. Therefore, be the $\mathscr{L}_r$-convergence theorem [2, p. 163],

$$\lim_{n\to\infty} \sum_{i=1}^{n} \varphi^{\mathscr{B}}(A_i) = \sum_{i=1}^{\infty} \varphi^{\mathscr{B}}(A_i)$$

in $\mathscr{L}_1$-norm. Noting (1.1), countable additivity follows. $\varphi^{\mathscr{B}}(\phi) = 0$ a.s. is immediate. Thus, $\varphi^{\mathscr{B}}$ is a vector-valued measure.

The total variation of a vector-valued measure need not be finite. Its role in the theory of integration with respect to a vector-valued measure ψ is assumed by a finite positive set function $|||\psi|||$ called the semi-variation of ψ. In general, $|||\psi|||$ is not additive. $|||\psi|||$ is additive if and only if $|||\psi|| = v[\psi]$, the total variation of ψ (see [1, p. 320]).

By definition,

$$|||\varphi^{\mathscr{B}}(A)||| = \sup_{\mathscr{P}, \{\alpha_i\}} \left\| \sum_{i=1}^{n} \alpha_i \varphi^{\mathscr{B}}(A_i) \right\| , \qquad A \in \mathscr{A} \tag{1.2}$$

where $\mathscr{P}$ is the collection of all finite partitions of the set A by elements of $\mathscr{A}$ and the α_i are complex numbers such that $|\alpha_i| \leqq 1$. Recall that

$$(1.3)\qquad \begin{aligned} \|\varphi^{\mathscr{B}}(A)\| &= \int |\varphi^{\mathscr{B}}(A)|\, d\mu \\ &= \int \mathscr{E}^{\mathscr{B}} I_A d\mu = \int_A d\mu = \mu(A)\,, \qquad A \in \mathscr{A}. \end{aligned}$$

Therefore, for any partition $\{A_i\}_{i=1}^n$ of A

$$(1.4)\qquad \begin{aligned} \left\|\sum_{i=1}^n \alpha_i \varphi^{\mathscr{B}}(A_i)\right\| &\leqq \sum_{i=1}^n |\alpha_i|\, \|\varphi^{\mathscr{B}}(A_i)\| \\ &\leqq \sum_{i=1}^n \mu(A_i) = \mu(A) \end{aligned}$$

since the numbers $\{\alpha_i\}_{i=1}^n$ all have moduli less than one. But equality of the first and last members of (1.4) is achieved when $\alpha_1 = \cdots = \alpha_n = 1$, so that $|||\varphi^{\mathscr{B}}||| = \nu(\mu) = \mu$.

Now let $X \in \mathscr{L}_1$; we shall show that X is integrable with respect to $\varphi^{\mathscr{B}}$ (see [1, p. 323]) and that its integral over Ω is $\mathscr{E}^{\mathscr{B}}X$. Let $\{X_n\}_1^\infty$ be a sequence of simple functions converging μ a.s. to X and such that $|X_n| \uparrow X$. It follows that $X_n \to X\ \varphi^{\mathscr{B}}$ a.e. since $|||\varphi^{\mathscr{B}}||| = \mu$. Then [2, p. 348], $\mathscr{E}^{\mathscr{B}}X_n \to \mathscr{E}^{\mathscr{B}}X\ \mu$ a.s. Since

$$(1.5)\qquad |\mathscr{E}^{\mathscr{B}}X_n| \leqq \mathscr{E}^{\mathscr{B}}|X_n| \leqq \mathscr{E}^{\mathscr{B}}|X| \in \mathscr{L}_1\,,$$

the integrals of the $\mathscr{E}^{\mathscr{B}}X_n$ are uniformly continuous and therefore

$$(1.6)\qquad \mathscr{E}^{\mathscr{B}}X_n \longrightarrow \mathscr{E}^{\mathscr{B}}X \qquad \text{in } \mathscr{L}_1\text{-norm.}$$

But if $X_n = \sum_{i=1}^r \alpha_i I_{A_i}$, we have

$$(1.7)\qquad \begin{aligned} \int X_n d\varphi^{\mathscr{B}} &= \sum_{i=1}^r \alpha_i \varphi^{\mathscr{B}}(A_i) \\ &= \sum_{i=1}^r \alpha_i \mathscr{E}^{\mathscr{B}} I_{A_i} = \mathscr{E}^{\mathscr{B}}\Big(\sum_{i=1}^r \alpha_i I_{A_i}\Big) = \mathscr{E}^{\mathscr{B}}X_n\,. \end{aligned}$$

By (1.6) and (1.7), it follows that

$$(1.8)\qquad \lim_{n\to\infty} \int X_n d\varphi^{\mathscr{B}} = \mathscr{E}^{\mathscr{B}}X \qquad \text{in } \mathscr{L}_1\text{-norm.}$$

Since (1.8) holds for any function in $\mathscr{L}_1$, it follows that

$$(1.9)\qquad \lim_{n\to\infty} \int I_A X_n d\varphi^{\mathscr{B}} = \mathscr{E}^{\mathscr{B}} I_A X \quad \text{in } \mathscr{L}_1\text{-norm}$$

for all $A \in \mathscr{A}$.

These last two limit assertions establish, according to definition [1, p. 323], the integrability of X and the formula

$$\int_A X d\varphi^{\mathscr{B}} = \mathscr{E}^{\mathscr{B}} I_A X\,, \qquad A \in \mathscr{A}. \tag{1.10}$$

In particular, (0.1) holds as asserted.

We gather together the results of this section in the form of a theorem.

THEOREM 1. *Let $(\Omega, \mathscr{A}, \mu)$ be a probability space and $\mathscr{B}$ a sub-σ-field of $\mathscr{A}$. Then the conditional probability functions relative to $\mathscr{B}$, denoted $\varphi^{\mathscr{B}}(A)$ for $A \in \mathscr{A}$, form a vector-valued measure on $\mathscr{A}$ with values in $\mathscr{L}_1(\Omega, \mathscr{A}, \mu)$ having semi-variation μ. Furthermore, if $X \in \mathscr{L}_1$*

$$\mathscr{E}^{\mathscr{B}} X = \int X d\varphi^{\mathscr{B}}\,. \tag{1.11}$$

2. **$\mathscr{L}_1$-valued measures which are conditional probabilities.** In this section, we determine the $\mathscr{L}_1(\Omega, \mathscr{A}, \mu)$-valued measures on $\mathscr{A}$ which are conditional probabilities.

THEOREM 2. *Let ψ be an $\mathscr{L}_1(\Omega, \mathscr{A}, \mu)$-valued measure on $\mathscr{A}$. Then ψ is the vector-valued conditional probability measure relative to some sub-σ-field $\mathscr{B}$ if and only if the following three conditions are satisfied:*

$$\begin{array}{ll} \text{(i)} & \psi(A) \geqq 0 \quad \text{a.s.} \qquad A \in \mathscr{A}\,, \\ \text{(ii)} & \psi(\Omega) = 1 \quad \text{a.s.}, \\ \text{(iii)} & \| I_A \psi(B)\psi(C) \| = \| \psi(A) I_B \psi(C) \| = \| \psi(A)\psi(B) I_C \| \\ & \hfill A, B, C \in \mathscr{A}\,. \end{array} \tag{2.1}$$

If ψ has these three properties, then it is conditional probability relative to the σ-field

$$\mathscr{B} = \{A \in \mathscr{A} \mid \psi(A) = I_A \text{ a.s.}\}\,. \tag{2.2}$$

Proof. Since $0 \leqq \psi(A) \leqq 1$ a.s. for all $A \in \mathscr{A}$, the values of ψ are equivalence classes of μ-essentially bounded functions. Consequently the products shown in (iii) above are all μ-integrable; and, hence, their norms are finite.

The conditions of the theorem are necessary. Let $\varphi^{\mathscr{B}}$ be the conditional probability measure relative to a sub-σ-field $\mathscr{B}$, then $\varphi^{\mathscr{B}}$ is known to satisfy conditions (i) and (ii) above. For any $A, B, C \in \mathscr{A}$,

$$\| I_A \cdot \varphi^{\mathscr{B}}(B) \cdot \varphi^{\mathscr{B}}(C) \| = \int_A \varphi^{\mathscr{B}}(B) \cdot \varphi^{\mathscr{B}}(C) d\mu$$

$$
\begin{aligned}
&= \int_\Omega \mathscr{E}^{\mathscr{B}} \mid I_A \cdot \varphi^{\mathscr{B}}(B) \cdot \varphi^{\mathscr{B}}(C) \mid d\mu \\
&= \int_\Omega (\mathscr{E}^{\mathscr{B}} I_A) \cdot \varphi^{\mathscr{B}}(B) \cdot \varphi^{\mathscr{B}}(C) d\mu \\
&= \int_\Omega \varphi^{\mathscr{B}}(A) \cdot \varphi^{\mathscr{B}}(B) \cdot \varphi^{\mathscr{B}}(C) d\mu \, .
\end{aligned} \tag{2.3}
$$

The right-hand side is symmetric in A, B, and C. Interchanging first A with B and then A with C, (iii) is proved for $\varphi^{\mathscr{B}}$. If $B \in \mathscr{B}$, then I_B is $\mathscr{B}$-measurable and $\varphi^{\mathscr{B}}(B) = I_B$. If $A \in \mathscr{A}$, but $A \notin \mathscr{B}$, then I_A is not $\mathscr{B}$-measurable and $\varphi^{\mathscr{B}}(A) = \mathscr{E}^{\mathscr{B}} I_A \neq I_A$. Thus, $\mathscr{B}$ is completely characterized by (2.2). This completes the proof of necessity.

It remains to prove sufficiency. Let ψ be an $\mathscr{L}_1$-valued measure on $\mathscr{A}$ satisfying conditions (i), (ii), and (iii); and let

$$\mathscr{B} = \{A \in \mathscr{A} \mid \psi(A) = I_A\} \, .$$

By letting $C = \Omega$, so that $\psi(C) = 1$ in (iii), we have

$$\int_A \psi(B) d\mu = \int_B \psi(A) d\mu \qquad A, B \in \mathscr{A} \, . \tag{2.4}$$

This equation can be used to prove that $\mathscr{B}$ is closed under intersections. Let $A, B \in \mathscr{B}$ and $C \in \mathscr{A}$. Then (2.4) implies

$$\int_C \psi(A \cap B) d\mu = \int_\Omega \psi(C) I_{A \cap B} d\mu \, . \tag{2.5}$$

By repeated application of (iii) and various definitions, we have

$$
\begin{aligned}
\int_C \psi(A \cap B) d\mu &= \int_\Omega \psi(C) I_A I_B d\mu \\
&= \int_\Omega \psi(C) \psi(A) I_B d\mu \\
&= \int_\Omega I_C \psi(A) \psi(B) d\mu \\
&= \int_C I_A \cdot I_B d\mu \\
&= \int_C I_{A \cap B} d\mu \, .
\end{aligned} \tag{2.6}
$$

Since this holds for all $C \in \mathscr{A}$, we conclude that $\psi(A \cap B) = I_{A \cap B}$ and that $A \cap B \in \mathscr{B}$. $\mathscr{B}$ is closed under complements: Let $A \in \mathscr{B}$. Then $\psi(A) = I_A$; also $\psi(\Omega) = 1 = I_\Omega$. By the additivity of ψ,

$$1 = \psi(\Omega) = \psi(A) + \psi(A^c) = I_A + \psi(A^c) \, .$$

Therefore $\psi(A^c) = 1 - I_A = I_{A^c}$, and $A^c \in \mathscr{B}$. Thus, $\mathscr{B}$ is a field.

To prove that $\mathscr{B}$ is a σ-field, it suffices to prove that $\mathscr{B}$ is closed under countable disjoint unions. Let $\{A_n\}_{n=1}^{\infty} \subset \mathscr{B}$ be a disjoint family. Then by the countable additivity of ψ,

$$\begin{aligned}\psi\left(\bigcup_{n=1}^{\infty} A_n\right) &= \sum_{n=1}^{\infty} \psi(A_n) \\ &= \sum_{n=1}^{\infty} I_{A_n} \\ &= I_{\bigcup_{n=1}^{\infty} A_n} .\end{aligned} \tag{2.7}$$

So, $\bigcup_{n=1}^{\infty} A_n \in \mathscr{B}$. Therefore $\mathscr{B}$ is a σ-field.

If $B \in \mathscr{B}$ and $A \in \mathscr{A}$, then (2.4) implies

$$\begin{aligned}\int_B \psi(A) d\mu &= \int_A \psi(B) d\mu \\ &= \int_A I_B d\mu \\ &= \mu(A \cap B) .\end{aligned} \tag{2.8}$$

This is the defining relation for conditional probability relative to $\mathscr{B}$ *provided* that $\psi(A)$ is a $\mathscr{B}$-measurable function.

In reality, the statement above is imprecise: $\psi(A)$ is an equivalence class of functions in $\mathscr{L}_1$. It suffices to prove that $\psi(A)$ contains one $\mathscr{B}$-measurable function. (Implicitly, we are using the fact that if $\mathscr{B}$ does not contain the entire collection of null sets in $\mathscr{A}$, these null sets can be adjoined to $\mathscr{B}$ in a harmless way.) From each equivalence class $\psi(A)$, we select a representative. Continuing the imprecision mentioned above, we will also denote this representative by $\psi(A)$, and show that $C_{A,\alpha} = \{\omega \mid \psi(A)(\omega) \leqq \alpha\} \in \mathscr{B}$ for any $A \in \mathscr{A}$ and any real number α. ($C^c_{A,\alpha} = \{\omega \mid \psi(A)(\omega) > \alpha\}$). This will prove that $\psi(A)$ is $\mathscr{B}$-measurable.

Let $B = C^c_{A,\alpha}$ and $C = C_{A,\alpha}$ in (iii) above. Then,

$$\begin{aligned}\int_A \psi(C^c_{A,\alpha})\psi(C_{A,\alpha}) d\mu &= \int_{C_{A,\alpha}} \psi(A)\psi(C^c_{A,\alpha}) d\mu \\ &\leqq \alpha \int_{C_{A,\alpha}} \psi(C^c_{A,\alpha}) d\mu ,\end{aligned} \tag{2.9}$$

and

$$\begin{aligned}\int_A \psi(C^c_{A,\alpha})\psi(C_{A,\alpha}) d\mu &= \int_{C^c_{A,\alpha}} \psi(A)\psi(C_{A,\alpha}) d\mu \\ &\geqq \alpha \int_{C^c_{A,\alpha}} \psi(C_{A,\alpha}) d\mu .\end{aligned} \tag{2.10}$$

Therefore,

$$\begin{aligned}\alpha\int_{C^c_{A,\alpha}}\psi(C_{A,\alpha})d\mu &\leqq \int_A \psi(C^c_{A,\alpha})\psi(C_{A,\alpha})d\mu \\ &\leqq \alpha\int_{C_{A,\alpha}}\psi(C^c_{A,\alpha})d\mu\ .\end{aligned}\tag{2.11}$$

By (2.4), the extremal members of inequality (2.11) are equal, and we have

$$\begin{aligned}\alpha\int_{C^c_{A,\alpha}}\psi(C_{A,\alpha})d\mu &= \int_A \psi(C^c_{A,\alpha})\psi(C_{A,\alpha})d\mu \\ &= \int_{C^c_{A,\alpha}}\psi(A)\psi(C_{A,\alpha})d\mu\ .\end{aligned}\tag{2.12}$$

Or,

$$\int_{C^c_{A,\alpha}}(\psi(A)-\alpha)\psi(C_{A,\alpha})d\mu = 0\ .\tag{2.13}$$

But on $C^c_{A,\alpha}$, $\psi(A)-\alpha>0$ μ a.s. Thus, since $\psi(C_{A,\alpha})\geqq 0$ μ a.s., $\psi(C_{A,\alpha})=0$ μ a.s. on $C^c_{A,\alpha}$. Using (iii) again with $A=B=\Omega$ and $C=C_{A,\alpha}$, we have

$$\int_\Omega \psi(C_{A,\alpha})d\mu = \int_{C_{A,\alpha}} 1 d\mu = \int_{C_{A,\alpha}} I_{C_{A,\alpha}} d\mu = \mu(C_{A,\alpha})\ .\tag{2.14}$$

But

$$\begin{aligned}\int_\Omega \psi(C_{A,\alpha})d\mu &= \int_{C_{A,\alpha}}\psi(C_{A,\alpha})d\mu + \int_{C^c_{A,\alpha}}\psi(C_{A,\alpha})d\mu \\ &= \int_{C_{A,\alpha}}\psi(C_{A,\alpha})d\mu\ .\end{aligned}\tag{2.15}$$

Comparing (2.14) and (2.15),

$$\int_{C_{A,\alpha}}\psi(C_{A,\alpha})d\mu = \int_{C_{A,\alpha}} I_{C_{A,\alpha}} d\mu\ .\tag{2.16}$$

On $C^c_{A,\alpha}$, $\psi(C_{A,\alpha})=I_{C_{A,\alpha}}=0$ μ a.s. On $C_{A,\alpha}$, $0\leqq\psi(C_{A,\alpha})\leqq 1$ by (i) and $I_{C_{A,\alpha}}=1$, so $I_{C_{A,\alpha}}-\psi(C_{A,\alpha})\geqq 0$ μ a.s. But (2.16) implies

$$\int_\Omega (I_{C_{A,\alpha}}-\psi(C_{A,\alpha}))d\mu = 0\ .\tag{2.17}$$

Thus, $\psi(C_{A,\alpha})=I_{C_{A,\alpha}}$ μ a.s. and $C_{A,\alpha}\in\mathscr{B}$ for all $A\in\mathscr{A}$ and α real. This concludes the proof of Theorem 2.

3. Characterization of conditional expectation in $\mathscr{L}_p, 1\leqq p<\infty$. Throughout this section, T will be a continuous linear map of $\mathscr{L}_p$ into $\mathscr{L}_p$. As usual, it will be convenient to use the same symbol for a function and its equivalence class. Thus, the statement "X is bounded by $M > 0$" means that the equivalence class of X consists of functions whose μ-essential supremum is not greater than M.

In [3], among more general results, Moy established the following which we now rephrase and state.

Moy's result: Let T be a continuous linear transformation of $\mathscr{L}_1$ into $\mathscr{L}_1$ having the following four properties:

(i) if X is bounded, then TX is bounded;

(ii) if X and Y are bounded, then

$$T(X\cdot TY) = TX\cdot TY\ ;$$

(iii)
$$T1 = 1\ ;$$

(iv) T is a contraction, that is, for every $X \in \mathscr{L}_1$

$$\| TX \| \leqq \| X \|\ .$$

Then, we have for every $X \in \mathscr{L}_1$,

$$TX = \mathscr{E}^{\mathscr{B}} X\ ,$$

where $\mathscr{B}$ is the σ-field of all sets $A \in \mathscr{A}$ such that $T(I_A \cdot X) = I_A \cdot TX$ for all bounded X.

In [4], Rota made the following definition which we will use below.

DEFINITION 1. An averaging operator T in $\mathscr{L}_p(\Omega, \mathscr{A}, \mu)$ (where p is a fixed real number $1 \leqq p \leqq \infty$) is a linear operator in $\mathscr{L}_p(\Omega, \mathscr{A}, \mu)$ with the following three properties:

(i) T is a contraction operator:

$$\| TX \|_{\mathscr{L}_p} \leqq \| X \|_{\mathscr{L}_p} \qquad \text{for } X \in \mathscr{L}_p\ ; \tag{3.1}$$

(ii) if Y is of class $\mathscr{L}_p$ and X is an essentially bounded function on $(\Omega, \mathscr{A}, \mu)$, then the function $(TX)\cdot(TY)$ is of class $\mathscr{L}_p$ and

$$T(X\cdot TY) = (TX)\cdot(TY)\ ; \tag{3.2}$$

(3.3) (iii)
$$T1 = 1\ .$$

We note that the operator T in Moy's result is an averaging operator, following a standard argument of approximation of a function in $\mathscr{L}_1$ by a sequence of bounded measurable functions. In the case $1 < p < \infty$, Rota [4] proved the following result which we rephrase and state.

Rota's Result: Let T be an averaging operator in $\mathscr{L}_p(\Omega, \mathscr{A}, \mu)$ (for fixed p). Then there exists a unique sub-σ-field $\mathscr{B} \subset \mathscr{A}$ such that $TX = \mathscr{E}^{\mathscr{B}}X$ for $X \in \mathscr{L}_p$.

In the proof, Rota defines $\mathscr{B}$ as the smallest σ-field containing all the σ-fields of inverse images of the Borel sets by bounded functions Y fixed under T. It is clear that this procedure will yield a σ-field $\mathscr{B}$ which is completed with respect to μ-null sets, just as Moy's construction of $\mathscr{B}$ does. The above statement must be interpreted in this light. Comparing the recipe for construction of $\mathscr{B}$ in the two results, we see that Rota's criterion for membership in $\mathscr{B}$ is, on the surface, weaker than Moy's. On the other hand, if $TI_A = I_A$, then $TI_A \cdot X = I_A \cdot X$ for X bounded and $T(TI_A) \cdot X) = T(I_A \cdot X)$. By the smoothing property (3.2) of averaging operators, $T((TI_A) \cdot X) = TI_A \cdot TX = T(I_A \cdot X)$. Hence $I_A \cdot TX = T(I_A \cdot X)$ for all X bounded, and the two recipes define the same family of sets.

A theorem can now be stated which contains both Moy's result and Rota's. This general theorem will be proved using a theorem on the representation of operators on a Lebesgue space and the characterization of conditional probabilities given in § 2.

THEOREM 3. *Let T be an averaging operator in $\mathscr{L}_p(\Omega, \mathscr{A}, \mu)$ (where p is a fixed real number $1 \leqq p < \infty$). Then there exists a unique sub-σ-field $\mathscr{B} \subset \mathscr{A}$ completed with respect to the μ-null sets of $\mathscr{A}$ such that $TX = \mathscr{E}^{\mathscr{B}}X$ for $X \in \mathscr{L}_p$. Furthermore,*

$$\mathscr{B} = \{A \in \mathscr{A} \mid TI_A = I_A\} .$$

Proof. The first part of the proof consists in showing that $T^*X = TX$ for all X bounded. A brief explanation of the meaning of this "self-adjointness" relation is given for the sake of completeness: As is well known for $1 \leqq p < \infty$, $\mathscr{L}_p{}^*$ is isometrically isomorphic to $\mathscr{L}_q$, where q is determined by the relation $1 = 1/p + 1/q$. Every linear operator $T: \mathscr{L}_p \to \mathscr{L}_p$ determines a mapping $T^*: \mathscr{L}_q \to \mathscr{L}_q$. Since $\mathscr{L}_\infty \subset \mathscr{L}_p \cap \mathscr{L}_q$ as sets for $1 \leqq p < \infty$, it makes sense to compare the action of T and T^* on bounded functions. To avoid confusion, the norm will be shown explicitly: thus, $\|\cdot\|_{\mathscr{L}_p}$.

The fact that $\| TX \|_{\mathscr{L}_p} \leqq \| X \|_{\mathscr{L}_p}$ and $\| T1 \|_{\mathscr{L}_p} = \| 1 \|_{\mathscr{L}_p} = 1$ implies

$$\| T \|_{\mathscr{B}(\mathscr{L}_p)} = 1 , \tag{3.4}$$

where $\|\cdot\|_{\mathscr{B}(\mathscr{L}_p)}$ is operator norm in $\mathscr{B}(\mathscr{L}_p)$, the bounded linear operators on $\mathscr{L}_p$.

Rota [3, p. 58] has shown that for $1 < p < \infty$ the averaging operator hypotheses imply that $T^*1 = 1$. In the case $p = 1$, let $Y_0 = T^*1$. Then,

$$\int Y_0 d\mu = \int T^*1 \cdot 1 d\mu = \int 1 \cdot T1 d\mu = 1 . \tag{3.5}$$

Further, since $T^*: \mathscr{L}_\infty \to \mathscr{L}_\infty$, $T^*1 = Y_0 \in \mathscr{L}_\infty$; and

$$\begin{aligned} \| Y_0 \|_{\mathscr{L}_\infty} &= \sup_{\substack{X \in \mathscr{L}_1 \\ \|X\|_{\mathscr{L}_1} \leqq 1}} \left| \int Y_0 \cdot X d\mu \right| \\ &= \sup_{\substack{X \in \mathscr{L}_1 \\ \|X\|_{\mathscr{L}_1} \leqq 1}} \left| \int T^*1 \cdot X d\mu \right| \\ &= \sup_{\substack{X \in \mathscr{L}_1 \\ \|X\|_{\mathscr{L}_1} \leqq 1}} \left| \int T X d\mu \right| \leqq \| T \|_{\mathscr{B}(\mathscr{L}_1)} = 1 . \end{aligned} \tag{3.6}$$

This means $-1 \leqq Y_0 \leqq +1$ μ a.s.; and, taken together with (3.5), this means $Y_0 = 1$ μ a.s. Hence, we have for $1 \leqq p < \infty$

$$T^*1 = 1 . \tag{3.7}$$

Since $\int 1 \cdot TX d\mu = \int T^*1 \cdot X d\mu$, this gives immediately

$$\int TX d\mu = \int X d\mu , \qquad X \in \mathscr{L}_p . \tag{3.8}$$

And so, finally, for any bounded X and $A \in \mathscr{A}$, we may calculate

$$\begin{aligned} \int_A TX d\mu &= \int I_A \cdot TX d\mu = \int T(I_A \cdot TX) d\mu && \text{(by 3.8)} \\ &= \int (TI_A) \cdot (TX) d\mu && \text{(by 3.2).} \end{aligned} \tag{3.9}$$

But

$$\begin{aligned} \int_A T^*X d\mu &= \int (T^*X) \cdot I_A d\mu = \int X \cdot (TI_A) d\mu \\ &= \int T(X \cdot TI_A) d\mu && \text{(by 3.8)} \\ &= \int (TX) \cdot (TI_A) d\mu && \text{(by 3.3).} \end{aligned} \tag{3.10}$$

Therefore, since $\int_A T^*X d\mu = \int_A TX d\mu$ for all $A \in \mathscr{A}$, we can conclude

$$T^*X = TX \qquad \text{for all } X \text{ bounded.} \tag{3.11}$$

In the next part of the proof, the principal instrument will be a theorem on the representation of operators on a Lebesgue space whose general form Dunford and Schwartz [1, p. 540] ascribe to Kantorovich and Vulich:

Result of Kantorovich and Vulich: Let $(\Omega, \mathscr{A}, \mu)$ be a measure space and let T be a continuous linear map of the Banach space $\mathfrak{X}$ into $\mathscr{L}_1(\Omega, \mathscr{A}, \mu)$. Then there is a uniquely determined function $\psi(\cdot)$ on $\mathscr{A}$ to $\mathfrak{X}^*$ such that

(i) for each $X \in \mathfrak{X}$, the set function $\psi(\cdot)[X]$ is μ-continuous and countably additive on $\mathscr{A}$;

(ii) for every $X \in \mathfrak{X}$, we have

$$TX = \frac{d[\psi(\cdot)X]}{d\mu} \tag{3.12}$$

(iii) the norm of T satisfies the relations

$$\sup_{A \in \mathscr{A}} \| \psi(A) \|_{\mathfrak{X}^*} \leqq \| T \|_{op} \leqq 4 \sup_{A \in \mathscr{A}} \| \psi(A) \|_{\mathfrak{X}^*} . \tag{3.13}$$

Conversely, if the function $\psi(\cdot)$ on $\mathscr{A}$ to $\mathfrak{X}^*$ satisfies (i), then (ii) defines an operator T on $\mathfrak{X}$ to $\mathscr{L}_1$ whose norm satisfies (iii).

Furthermore, T is weakly compact if and only if $\psi(\cdot)$ is countably additive on $\mathscr{A}$ in the strong topology of $\mathfrak{X}^*$.

The proof of this important theorem can be found in [1, p. 498]. To apply the result above to the operator T, we choose to consider T as a continuous linear map of $\mathscr{L}_p(1 \leqq p < \infty, p$ fixed) into $\mathscr{L}_1$. Thus, $\mathfrak{X}$ is $\mathscr{L}_p$ and $\mathfrak{X}^*$ can be identified with $\mathscr{L}_q$. If $\psi(\cdot)$ is the function given in the result above, then for any $X \in \mathscr{L}_1$ and $A \in \mathscr{A}$, we have

$$\begin{aligned} \int_A TX d\mu &= \int_A \frac{d\psi(\cdot)[X]}{d\mu} d\mu = \psi(A)[X] \\ &= \int \psi(A) \cdot X d\mu . \end{aligned} \tag{3.14}$$

On the other hand, $T = T^*$ on bounded functions, so

$$\begin{aligned} \int_A TX d\mu &= \int I_A \cdot (TX) d\mu \\ &= \int (T^* I_A) \cdot X d\mu = \int (TI_A) \cdot X d\mu . \end{aligned} \tag{3.15}$$

Hence, for each $A \in \mathscr{A}$, putting $X = I_B$, we get

$$\int_B TI_A d\mu = \int_B \psi(A) d\mu \qquad B \in \mathscr{A}. \tag{3.16}$$

And therefore,

$$TI_A = \psi(A) , \qquad A \in \mathscr{A}. \tag{3.17}$$

Rota shows [4, p. 58] that an averaging operator in $\mathscr{L}_p$, $1 < p < \infty$

maps bounded measurable functions into bounded measurable functions. For $p = 1$, $\psi(A) \in \mathscr{L}_\infty$; and, hence $TI_A = \psi(A)$ is a bounded measurable function for $1 \leqq p < \infty$. $\psi(\phi) = TI_\phi = 0$. A simple domination argument shows that if $\{A_j\}_{j=1}^\infty$ is a disjoint family, $\sum_{j=1}^n I_{A_j} \to I_{\cup_{j=1}^\infty A_j}$ in $\mathscr{L}_p$-norm. Therefore

$$T\Big(\sum_{j=1}^n I_{A_j}\Big) = \sum_{j=1}^n \psi(A_j) \longrightarrow \psi\Big(\bigcup_{j=1}^\infty A_j\Big) \quad \text{in } \mathscr{L}_p\text{-norm.} \tag{3.18}$$

Thus, ψ is a vector-valued measure on $\mathscr{A}$ to the bounded measurable functions on $\mathscr{A}$, which is countably additive in the $\mathscr{L}_p$-topology $(1 \leqq p < \infty)$. A fortiori, ψ is $\mathscr{L}_1$-countably additive. Since $TI_\Omega = \psi(\Omega) = 1$, condition (ii) in (2.1) is satisfied. Furthermore, for $A \in \mathscr{A}$

$$\mu(A) = \int (T^*I)\cdot I_A d\mu = \int \psi(A) d\mu\ . \tag{3.19}$$

But,

$$\begin{aligned}\int \psi(A) d\mu &\leqq \int |\psi(A)|\, d\mu \leqq \int |TI_A|\, d\mu \\ &\leqq \|TI_A\|_{\mathscr{L}_1} \leqq \|T\|_{\mathscr{B}(\mathscr{L}_1)} \cdot \|I_A\|_{\mathscr{L}_1} \leqq \mu(A)\ ,\end{aligned} \tag{3.20}$$

since $\|T\|_{\mathscr{B}(\mathscr{L}_p)} = 1, 1 \leqq p < \infty$. Therefore, equality holds throughout (3.20), so

$$\int [|\psi(A)| - \psi(A)] d\mu = 0\ . \tag{3.21}$$

Because $|\psi(A)| - \psi(A) \geqq 0$, we have

$$\psi(A) = |\psi(A)| \geqq 0\ . \tag{3.22}$$

Thus, condition (ii) in (2.1) is satisfied.

The smoothing property (3.2) implies T is idempotent: $T^2X = T(1\cdot TX) = T1\cdot TX = TX$. Now let $A, B, C \in \mathscr{A}$, using (3.17), (3.2), (3.8), the fact that T maps bounded measurable functions into bounded measurable functions, and idempotence, we have

$$\begin{aligned}\int_A \psi(B)\psi(C) d\mu &= \int_A (TI_B)\cdot(TI_C) d\mu \\ &= \int I_A \cdot T(I_B\cdot(TI_C)) d\mu \\ &= \int T(I_A\cdot T(I_B\cdot(TI_C)) d\mu \\ &= \int (TI_A)\cdot T(I_B\cdot(TI_C)) d\mu\end{aligned} \tag{3.23}$$

$$= \int (TI_A)\cdot(TI_B)\cdot(T^2 I_C)d\mu$$
$$= \int (TI_A)\cdot(TI_B)\cdot(TI_C)d\mu \, .$$

Clearly, the result is the same if A is interchanged with B or A is interchanged with C. Thus, condition (iii) in (2.1) is satisfied; and $\psi(\cdot)$ is a conditional probability in $\mathscr{L}_1$, by Theorem 2. The integral with respect to $\psi(\cdot)$ of $X \in \mathscr{L}_1$ is $\mathscr{E}^{\mathscr{B}}X$, where $\mathscr{B}$ is the sub-σ-field defined by

$$\mathscr{B} = \left\{ A \in \mathscr{A} \mid \psi(A) = \int I_A d\psi = I_A \right\} .$$

Thus, $\mathscr{B}$ consists of those sets in $\mathscr{A}$ whose indicator functions are invariant under $\mathscr{E}^{\mathscr{B}}$. It is well known [2, p. 348] that $\mathscr{E}^{\mathscr{B}}$ defines an $\mathscr{L}_p$-continuous linear transformation ($1 \leqq p < \infty$). Then,

$$\mathscr{E}^{\mathscr{B}} I_A = \psi(A) = TI_A \, , \tag{3.24}$$

and so the restriction of $\mathscr{E}^{\mathscr{B}}$ to $\mathscr{L}_p$ agrees with the $\mathscr{L}_p$-continuous operator T on a generating set; hence, it agrees on all of $\mathscr{L}_p$. We conclude that T is $\mathscr{E}^{\mathscr{B}}$ on $\mathscr{L}_p$. This, with the fact then that $\mathscr{B} = \{A \in \mathscr{A} \mid TI_A = I_A\}$, finishes the proof of Theorem 3.

This theorem shows that the set of averaging operators in $\mathscr{L}_r$ and $\mathscr{L}_s$ are in one-to-one correspondence $1 \leqq r, s < \infty$. Moreover, every averaging operator T on $\mathscr{L}_r$ has a unique $\mathscr{L}_s$-continuous extension $1 \leqq s \leqq r < \infty$. In a subsequent publication, the author intends to relate these facts to the work of Sidak [5].

References

1. N. Dunford and J. Schwartz, *Linear Operators Part I: General Theory*, New York, 1958.

2. M. Loève, *Probability Theory*, 3rd ed., Princeton, 1963.

3. S. C. Moy, *Characterizations of conditional expectation as a transformation on function spaces*, Pacific J. Math. **4** (1954), 47–63.

4. G. C. Rota, *On the representation of averaging operators*, Rendiconti del Seminario Matematico della Universita di Padova, Vol. XXX, 1960, pp. 52–64.

5. Z. Sidak, *On relations between strict-sense and wide-sense conditional expectations*, Theory of Probability and Its Applications, Vol. II, 1957, pp. 267–272.

UNIVERSITY OF CALIFORNIA, BERKELEY