PSYCHOMETRIKA—VOL. 40, NO. 3
SEPTEMBER, 1975

PROBABILITY SPACES, HILBERT SPACES, AND THE AXIOMS OF TEST THEORY

DONALD W. ZIMMERMAN

CARLETON UNIVERSITY

A branch of probability theory that has been studied extensively in recent years, the theory of conditional expectation, provides just the concepts needed for mathematical derivation of the main results of the classical test theory with minimal assumptions and greatest economy in the proofs. The collection of all random variables with finite variance defined on a given probability space is a Hilbert space; the function that assigns to each random variable its conditional expectation is a linear operator; and the properties of the conditional expectation needed to derive the usual test-theory formulas are general properties of linear operators in Hilbert space. Accordingly, each of the test-theory formulas has a simple geometric interpretation that holds in all Hilbert spaces.

It is generally recognized that the natural setting of mental test theory, like all statistical models, is the theory of probability. Not widely realized, however, is the fact that one particular branch of probability theory that has been studied extensively in recent years, the theory of conditional expectation, provides just the concepts needed for mathematical derivation of the test-theory results with a minimum of effort. The main findings of the classical test theory, as presented by Gulliksen [1950], Guttman [1945], Lord and Novick [1968], and others, can be derived in a very natural way from the concepts of conditional expectation, conditional independence, and related notions. We will amplify this point in the present paper and show that these concepts provide precisely the formalism needed to obtain the classical results with minimal assumptions and with greatest economy in the methods of proof.

In the first part of this paper, we will examine some properties of random variables and their conditional expectations that have been investigated extensively in probability theory [see, for example, Burrill, 1972, Loève, 1963, Rényi, 1970]. The conditional expectation of a random variable defined on a probability space will be regarded as itself a random variable defined on the same probability space. The concepts of conditionally independent random variables and conditionally uncorrelated random variables will be introduced. Using these notions, it is possible to derive the main formulas that are of interest in test theory in a very few steps. Then, by identifying random variables and test scores, these results, which are purely formal consequences of probability theory, can be interpreted as assertions about test reliability, test validity, and so on.

Next, we will show that the properties of conditional expectation that are required to obtain these results are general properties of linear operators in Hilbert space. Consequently, each of the test-theory formulas has a simple geometric interpretation in the space of all random variables with finite variance defined on a given probability space. In this way, the test-theory formulas, which usually are regarded as relations among observed scores, true scores, and error scores, can be embedded first in the theory of conditional expectation and, in turn, in the theory of linear operators in Hilbert space. This approach, we will see, leads to considerable economy in the proofs and yields insight into the structure of the theory.

Sums and Products of Random Variables

Some concepts in probability theory that are important for present purposes, including conditional expectation regarded as a random variable and conditional independence, will be defined and treated in some detail before the main theorems are proved. Since the present development depends in an essential way on addition and multiplication of random variables having the same associated probability space, we recall the following definitions.

A collection of subsets of a set Ω is a σ-algebra if it is closed under countable set operations. Let $\mathfrak{A}$ be a σ-algebra of subsets of Ω, $\mathfrak{A}'$ a σ-algebra of subsets of Φ, and f and g functions from Ω to Φ. Then f is $\mathfrak{A}$-measurable if $\mathfrak{B} = f^{-1}(\mathfrak{A}')$, the σ-algebra induced by f, is contained in $\mathfrak{A}$, and g is $\mathfrak{B}$-measurable if $g^{-1}(\mathfrak{A}')$ is contained in $\mathfrak{B}$.

A probability space $(\Omega, \mathfrak{A}, P)$ consists of a fundamental probability set Ω, a σ-algebra $\mathfrak{A}$ of subsets of Ω, and a non-negative, countably additive set function P, defined on $\mathfrak{A}$, such that $P(\Omega) = 1$. A random point $f : \Omega \to \Phi$ is an $\mathfrak{A}$-measurable function from Ω to another set Φ consisting of elements that are not necessarily numbers, while a random variable $X : \Omega \to R$ is an $\mathfrak{A}$-measurable function from Ω to the real numbers.

Let $X : \Omega \to R$ and $Y : \Omega \to R$ be two random variables with the same associated probability space $(\Omega, \mathfrak{A}, P)$. The random variable $X + Y$ is a function from Ω to R defined by

$$(X + Y)(\omega) = X(\omega) + Y(\omega), \text{ for each } \omega \in \Omega,$$

and the random variable XY is a function from Ω to R defined by

$$(XY)(\omega) = X(\omega)Y(\omega), \text{ for each } \omega \in \Omega.$$

That is, $X + Y$ and XY are defined by pointwise addition and multiplication.

The conditional expectation of X, denoted by M_X, is a random variable $M_X : \Omega \to R$ that depends on X in a way to be made explicit below. On the other hand, the symbol EX denotes the expectation of the random variable X, defined by $EX = \int_\Omega X\, dP$. In some instances which will be evident from context, EX will denote a constant function that takes the value EX at each $\omega \in \Omega$. Similarly, 0 will sometimes represent the function that takes the value 0 at each $\omega \in \Omega$. The symbols Var, σ, Cov, and ρ will be used for variance, standard deviation, covariance, and correlation, respectively.

The Conditional Expectation of a Random Variable with Respect to a Random Point

Let $(\Omega, \mathfrak{A}, P)$ be a probability space. We consider a random point $f : \Omega \to \Phi$, taking values in a set Φ, and a random variable $X : \Omega \to R$, such that the expectation EX exists.

Definition 1.

The *conditional expectation of X with respect to f* is a $\mathfrak{B}$-measurable random variable $M_X : \Omega \to R$, such that

$$E(M_X \mid B) = E(X \mid B), \text{ for all } B \in \mathfrak{B},$$

where $\mathfrak{B} \subseteq \mathfrak{A}$ is the σ-algebra induced by f.

Let $\mathcal{E}_f$ be the equivalence relation on Ω defined by $\omega_1 \mathcal{E}_f \omega_2$ iff $f(\omega_1) = f(\omega_2)$, and let $\Lambda = \Omega/\mathcal{E}_f \subset \mathfrak{B}$ be the partition of Ω induced by $\mathcal{E}_f$. Then, M_X is constant on each element of Λ and

$$M_X(\omega) = E[X \mid \gamma(\omega)], \quad \text{for each} \quad \omega \in \Omega,$$

where γ is the projection of Ω onto Λ—that is, $\gamma(\omega) = f^{-1}(\alpha)$ if $\alpha = f(\omega)$.

Let $E[X \mid f]$ be a real-valued function that assigns the number $E[X \mid f = \alpha]$ to each point α in the range of f. Then the conditional expectation is a composite function $M_X = E[X \mid f] \circ f$, as in the diagram

$$\begin{tikzcd} \Omega \arrow[r, "f"] \arrow[dr, "M_X"'] & f(\Omega) \arrow[d, "{E[X \mid f]}"] \\ & R \end{tikzcd}$$

According to definition 1, the random point f and the random variables X and M_X all have the same associated probability space $(\Omega, \mathfrak{A}, P)$. Advantages of the notation M_X, instead of the conventional $E[X \mid \mathfrak{B}]$, will appear below, where an operator projecting a linear space of random variables onto a subspace of random variables is considered.

In the test-theory model to be described, the random variable X represents a test score, while the random point f represents selection of an individual or experimental object (see definition 7 below). Accordingly, the function $E[X \mid f]$ assigns to each individual α in the range of f a true score $E[X \mid f](\alpha) = E[X \mid f = \alpha]$. The concept of a true score as the expected value of an observed score is familiar in the classical test theory, as formulated by Guttman [1945], Lord and Novick [1968], and others. The function M_X assigns to each elementary event ω the true score of the individual $\alpha = f(\omega)$. Hence, for a given observed score random variable X, we may consider M_X as a corresponding true score random variable, which is constant when restricted to any individual.

Example 1.

An experiment consists in first tossing a coin and then drawing a ball from one of two urns, depending on the outcome. If the outcome is H, the urn contains a ball labelled 0 and a ball labelled 1, and if the outcome is T, the urn contains a ball labelled 1 and a ball labelled 2. Let the random point f represent the outcome of the coin toss and the random variable X represent the number subsequently drawn, so that $f(\Omega) = \{H, T\}$ and $X(\Omega) = \{0, 1, 2\}$. The conditional expectation M_X assigns to $\omega \varepsilon \Omega$ the number $\frac{1}{2}$ if $f(\omega) = H$ and the number $\frac{3}{2}$ if $f(\omega) = T$. The set Ω and the functions f, X, and M_X are described in Table 1.

Example 2.

Select a point α at random on the interval $[0, 1]$. Then select a point β at random on the interval $[\alpha, \alpha + 1]$. Define $f(\alpha, \beta) = \alpha$ and $X(\alpha, \beta) = \beta$, for each $\omega = (\alpha, \beta) \varepsilon \Omega$. Since $E[X \mid f = \alpha] = \alpha + \frac{1}{2}$, the conditional expectation M_X is the function from Ω to R defined by $M_X(\alpha, \beta) = \alpha + \frac{1}{2}$, for each $\omega = (\alpha, \beta) \varepsilon \Omega$.

Conditioning as an Operator

We can regard M as an operator, which applied to a random variable X with expectation produces another random variable M_X that is constant

TABLE 1

Test Procedure in Example 1

ω	$p(\omega)$	$f(\omega)$	$X(\omega)$	$M_X(\omega)$
ω_1	$\frac{1}{4}$	H	0	$\frac{1}{2}$
ω_2	$\frac{1}{4}$	H	1	$\frac{1}{2}$
ω_3	$\frac{1}{4}$	T	1	3/2
ω_4	$\frac{1}{4}$	T	2	3/2

on each element of Λ. Let X and Y be any random variables with expectations. Then, by the linearity of the expectation,

$$\begin{aligned} M_{X+Y}(\omega) &= E[(X + Y) \mid \gamma(\omega)] \\ &= E[X \mid \gamma(\omega)] + E[Y \mid \gamma(\omega)] \\ &= M_X(\omega) + M_Y(\omega), \text{ for each } \omega \in \Omega, \end{aligned}$$

and therefore

$$M_{X+Y} = M_X + M_Y . \tag{1}$$

Also, $M_{aX} = aM_X$, for $a \in R$, so that M is a linear operator.

Let X and Y be any random variables such that EX and $E(XY)$ exist and such that Y is constant on each element of Λ. Then, again by the linearity property,

$$\begin{aligned} M_{XY}(\omega) &= E[XY \mid \gamma(\omega)] \\ &= E[X \mid \gamma(\omega)]\, Y(\omega) \\ &= M_X(\omega) Y(\omega), \text{ for each } \omega \in \Omega, \end{aligned}$$

and therefore

$$M_{XY} = M_X Y. \tag{2}$$

Similarly, $M_Y = Y$, if Y is constant on each element of Λ. In particular,

$$M_{M_X} = M_X . \tag{3}$$

Since $E(M_X \mid \Omega) = E(X \mid \Omega)$, it follows that

$$EM_X = EX, \tag{4}$$

a fact that will be used often in the sequel. Also, it follows from (1) and (4) that $M_{X-EX} = M_X - EM_X$.

Now, let X and Y be any random variables such that EX, EY, and $E(XY)$ exist. Then, $E(M_X Y) = EM_{M_X Y} = E(M_X M_Y) = EM_{XM_Y} = E(XM_Y)$, by (2) and (4). Also, Cov $(M_X , Y) = E(M_X Y) - (EM_X)(EY) = E(M_X M_Y) - (EM_X)(EM_Y) =$ Cov (M_X , M_Y), and similarly for Cov (X, M_Y), by interchanging X and Y. We have proved

$$\text{Cov}\,(M_X , Y) = \text{Cov}\,(X, M_Y) = \text{Cov}\,(M_X , M_Y). \tag{5}$$

We can now prove easily the following theorem, which is not usually encountered in the theory of conditional expectation, but which will be familiar in the context of test theory. Apart from definitions, all that is needed is the fact that Cov $(M_X , Y) =$ Cov (M_X , M_Y).

Theorem 1.

Let X be any random variable with finite non-zero variance, such that its conditional expectation M_X has non-zero variance. Then

$$\rho^2(X, M_X) = \frac{\text{Var } M_X}{\text{Var } X}. \tag{6}$$

Proof. Using (5), we have Cov (X, M_X) = Cov (M_X, M_X) = Var M_X. Substitution in

$$\rho^2(X, M_X) = \frac{[\text{Cov } (X, M_X)]^2}{\text{Var } X \text{ Var } M_X}$$

completes the proof.

The following theorem gives additional properties of the ratio Var M_X/Var X and the correlation $\rho(X, M_X)$.

Theorem 2.

Let X and Y be any random variables with finite non-zero variance, such that the conditional expectations M_X and M_Y have non-zero variance. Then,

$$\rho(X, M_Y) \leq \rho(X, M_X),$$

that is, the correlation between a random variable X and the conditional expectation of any other random variable does not exceed the correlation between X and its conditional expectation.

Proof. Since Cov (X, M_Y) = Cov (M_X, Y), by (5) and $[\text{Cov } (M_X, M_Y)]^2 \leq$ Var M_X Var M_Y, by the Cauchy–Schwartz inequality, substituting, dividing by Var X Var M_Y, and using (6) gives the above result.

Conditionally Independent Random Variables

We consider now a random point f and two random variables X_1 and X_2, all with the same associated probability space $(\Omega, \mathfrak{A}, P)$. In the sequel, B is an arbitrary element of the σ-algebra $\mathfrak{B}$ induced by f, and D is an arbitrary element of the partition Λ.

Definition 2.

Random variables X_1 and X_2 are *conditionally independent with respect to* f if, for all $D \in \Lambda$, the random variables $X_1 \mid D$ and $X_2 \mid D$ are independent.

If random variables X_1 and X_2 are conditionally independent, the conditional expectation of their product is the product of their conditional expectations. That is,

$$\begin{aligned} M_{X_1X_2}(\omega) &= E[X_1X_2 \mid \gamma(\omega)] \\ &= E[X_1 \mid \gamma(\omega)]E[X_2 \mid \gamma(\omega)] \\ &= M_{X_1}(\omega)M_{X_2}(\omega), \text{ for each } \omega \in \Omega, \end{aligned}$$

and therefore

(7) $$M_{X_1X_2} = M_{X_1}M_{X_2}\,.$$

Definition 3.

Random variables X_1 and X_2 are *conditionally identically distributed with respect to* f if, for all $D \varepsilon \Lambda$, the random variables $X_1 \mid D$ and $X_2 \mid D$ are identically distributed, that is, if the induced conditional probability distributions $P_{X_1|D}$ and $P_{X_2|D}$ are equal. Obviously, if X_1 and X_2 are conditionally identically distributed, then $M_{X_1} = M_{X_2}$, but the converse is not generally true. Independent, identically distributed random variables are familiar in sampling theory. We will see below that conditionally independent, conditionally identically distributed random variables correspond to parallel measurements in test theory.

Example 3.

The experiment in Example 1 is modified as follows. A coin is tossed and then two balls are drawn randomly, with replacement, from one of the two urns, according to the outcome of the coin toss. Let the random variables X_1 and X_2 represent the two numbers obtained. The set Ω and the functions f, X_1, X_2, M_{X_1}, and M_{X_2} are described in Table 2. The dependent random variables X_1 and X_2 are conditionally independent and conditionally identically distributed, since $X_1 \mid f = \alpha_1$ and $X_2 \mid f = \alpha_1$ are independent and

TABLE 2

Test Procedure in Example 3

ω	$p(\omega)$	$f(\omega)$	$X_1(\omega)$	$X_2(\omega)$	$M_{X_1}(\omega)$	$M_{X_2}(\omega)$
ω_1	1/8	H	0	0	½	½
ω_2	1/8	H	0	1	½	½
ω_3	1/8	H	1	0	½	½
ω_4	1/8	H	1	1	½	½
ω_5	1/8	T	1	1	3/2	3/2
ω_6	1/8	T	1	2	3/2	3/2
ω_7	1/8	T	2	1	3/2	3/2
ω_8	1/8	T	2	2	3/2	3/2

identically distributed, and $X_1 \mid f = \alpha_2$ and $X_2 \mid f = \alpha_2$ are independent and identically distributed.

Example 4.

Select a point α at random on the interval $[0, 1]$. Then select two points β_1 and β_2 independently and at random on the interval $[\alpha, \alpha + 1]$. Define $f(\alpha, \beta_1, \beta_2) = \alpha$, $X_1(\alpha, \beta_1, \beta_2) = \beta_1$ and $X_2(\alpha, \beta_1, \beta_2) = \beta_2$, for each $\omega = (\alpha, \beta_1, \beta_2) \in \Omega$. The conditional expectations M_{X_1} and M_{X_2} are the same function from Ω to R, defined by $M_{X_1}(\alpha, \beta_1, \beta_2) = M_{X_2}(\alpha, \beta_1, \beta_2) = \alpha + \frac{1}{2}$, for each $\omega = (\alpha, \beta_1, \beta_2) \in \Omega$. Furthermore, the dependent random variables X_1 and X_2 are conditionally independent and conditionally identically distributed, since for all $\alpha \in [0, 1]$, the random variables $X_1 \mid f = \alpha$ and $X_2 \mid f = \alpha$ are independent and identically distributed.

The following theorem now can be proved easily.

Theorem 3.

Let X_1 and X_2 be conditionally independent, conditionally identically distributed random variables with finite non-zero variance. Then

$$\rho(X_1, X_2) = \frac{\text{Var } M_{X_1}}{\text{Var } X_1}. \tag{8}$$

Proof.

$$\begin{aligned} \text{Cov }(X_1, X_2) &= E(X_1X_2) - (EX_1)(EX_2) \\ &= EM_{X_1X_2} - (EM_{X_1})(EM_{X_2}) \text{ by (4)} \\ &= E(M_{X_1}M_{X_2}) - (EM_{X_1})(EM_{X_2}) \text{ by (7)} \\ &= \text{Cov }(M_{X_1}, M_{X_2}) = \text{Var } M_{X_1} = \text{Var } M_{X_2}. \end{aligned}$$

Substitution in

$$\rho(X_1, X_2) = \frac{\text{Cov }(X_1, X_2)}{\sigma X_1 \sigma X_2}$$

completes the proof.

Conditionally Uncorrelated Random Variables

We continue to consider a random point f and two random variables X_1 and X_2, all with the same associated probability space $(\Omega, \mathfrak{A}, P)$.

Definition 4.

Random variables X_1 and X_2 are *conditionally uncorrelated with respect to f* if, for all $D \in \Lambda$, the random variables $X_1 \mid D$ and $X_2 \mid D$ are uncorrelated—that is, if $\text{Cov }(X_1 \mid D, X_2 \mid D) = 0$.

Conditionally independent random variables are conditionally uncorrelated, but the converse is not generally true. Random variables X_1 and X_2 are conditionally uncorrelated if and only if $M_{X_1X_2} = M_{X_1}M_{X_2}$.

The conclusion in Theorem 3 now can be obtained from a weaker hypothesis in the following way. Note that the proof used only these facts: $EX = EM_X$, which is true for any random variable with expectation, $\text{Var } X_1 = \text{Var } X_2$, $M_{X_1X_2} = M_{X_1}M_{X_2}$, which is true if X_1 and X_2 are conditionally uncorrelated, and $\text{Var } M_{X_1} = \text{Var } M_{X_2} = \text{Cov } (M_{X_1}, M_{X_2})$, which is true if $M_{X_1} = M_{X_2} + C$ for a constant function C, or, equivalently, if $M_{X_1-EX_1} = M_{X_2-EX_2}$. Explicitly, the last equivalence is proved as follows. If $M_{X_1-EX_1} = M_{X_2-EX_2}$, then $M_{X_1} - EM_{X_1} = M_{X_2} - EM_{X_2}$ and $M_{X_1} = M_{X_2} + C$, for a constant function $C = EM_{X_1} - EM_{X_2}$. Conversely, if $M_{X_1} = M_{X_2} + C$, for some constant function C, then $M_{X_1} - EM_{X_1} = M_{X_2} + C - E(M_{X_2} + C)$ and $M_{X_1} - EM_{X_1} = M_{X_2} - EM_{X_2}$ and $M_{X_1} - EX_1 = M_{X_2-EX_2}$. Therefore, if X_1 and X_2 are conditionally uncorrelated with $\text{Var } X_1 = \text{Var } X_2$ and $M_{X_1-EX_1} = M_{X_2-EX_2}$, the same conclusion is obtained as in Theorem 3.

Sums of Conditionally Independent Random Variables

The notion of the reliability of lengthened tests has been prominent in test theory. The following theorem expresses a correlation between a sum of random variables and the conditional expectation of the sum in terms of the correlation between each random variable and its conditional expectation.

Theorem 4.

Let $X_1, X_2, \cdots, X_n$ be conditionally independent, conditionally identically distributed random variables with finite non-zero variance, and let $S_n = X_1 + X_2 + \cdots + X_n$. Then,

$$\rho^2(S_n, M_{S_n}) = \frac{n\rho^2(X_1, M_{X_1})}{1 + (n-1)\rho^2(X_1, M_{X_1})}. \tag{9}$$

Proof. Since $X_1, X_2, \cdots, X_n$ are conditionally independent and conditionally identically distributed, we have $\text{Cov } (X_i, X_j) = \text{Cov } (M_{X_i}, M_{X_j}) = \text{Var } M_{X_1}$, for $i, j = 1, 2, \cdots, n$, as in the proof of Theorem 3,

$$\text{Var } S_n = \sum_{i=1}^{n} \text{Var } X_i + \sum_{i \neq j} \text{Cov } (X_i, X_j) = n \text{ Var } X_1 + n(n-1) \text{ Var } M_{x_1},$$

and $\text{Var } M_{S_n} = n^2 \text{ Var } M_{X_1}$. Since $\rho^2(S_n, M_{S_n}) = \text{Var } M_{S_n}/\text{Var } S_n$ and $\rho^2(X_1, M_{X_1}) = \text{Var } M_{X_1}/\text{Var } X_1$, by (6), substitution establishes the above result.

The same result can be obtained under the weaker hypothesis that $X_1, X_2, \cdots, X_n$ are conditionally uncorrelated with $\text{Var } X_1 = \cdots = \text{Var } X_n$ and $M_{X_1-EX_1} = \cdots = M_{X_n-EX_n}$. Also, using (8), a similar argument leads to the following result: let $X_1, X_2, \cdots, X_n, X_{n+1}, \cdots, X_{2n}$ be conditionally uncorrelated with $\text{Var } X_1 = \cdots = \text{Var } X_{2n}$ and $M_{X_1 - EX_1} = M_{X_{2n} - EX_{2n}}$, and let $S_n = X_1 + X_2 + \cdots + X_n$ and $S_n' = X_{n+1} + X_{n+2} + \cdots + X_{2n}$. Then,

$$\rho(S_n, S_n') = \frac{n\rho(X_1, X_2)}{1 + (n - 1)\rho(X_1, X_2)}. \tag{10}$$

Example 5.

Select a point α at random on the interval $[0, 1]$. Then select n points $\beta_1, \beta_2, \cdots, \beta_n$ independently and at random on the interval $[\alpha, \alpha + 1]$. Define $f(\omega) = \alpha$, $X_i(\omega) = \beta_i$, for each $\omega = (\alpha, \beta_1, \beta_2, \cdots, \beta_n) \in \Omega$, and $S_n = \sum_{i=1}^n X_i$, $i = 1, 2, \cdots, N$. Then, $\text{Var } M_{X_i} = 1/12$, $\text{Var } X_i = 1/6$, $\rho^2(X_i, M_{X_i}) = \frac{1}{2}$, by (6), and

$$\rho^2(S_n, M_{S_n}) = \frac{n\rho^2(X_i, M_{X_i})}{1 + (n - 1)\rho^2(X_i, M_{X_i})} = \frac{n}{n + 1}, \text{ by (9)}.$$

The Complement of the Conditional Expectation and the Conditional Variance

For a given probability space $(\Omega, \mathfrak{A}, P)$, a random point f, and a random variable X, we can define another random variable that is related to the conditional expectation, as follows.

Definition 5.

The *complement of the conditional expectation of X with respect to f* is a random variable $(X - M_X): \Omega \rightarrow R$, defined by

$$(X - M_X)(\omega) = X(\omega) - M_X(\omega), \quad \text{for each } \omega \in \Omega,$$

that is, $X - M_X$ is just the pointwise difference of X and M_X. We will see that this concept corresponds to an error score in test theory.

We can consider $1 - M$ as an operator which applied to a random variable X with expectation produces another random variable $X - M_X$. Here 1 denotes the identity operator defined by $1(X) = X$, and $1 - M$ is defined by $(1 - M)(X) = 1(X) - M(X) = X - M_X$. We also write $(1 - M)_X$ for $(1 - M)(X)$. For any random variable X with expectation, $M_{X - M_X} = 0$, by (1) and (3), and $EM_{X - M_X} = E(X - M_X) = 0$, by (4). More generally, it follows from definition 1 that

$$E[(X - M_X) \mid B] = 0, \quad \text{for all } B \in \mathfrak{B}.$$

Also, it is easy to see that $(1 - M)_{X+Y} = (1 - M)_X + (1 - M)_Y = (X - M_X) + (Y - M_Y)$, for any X and Y with expectations, and that $(1 - M)_{XY} = (X - M_X)Y$, for any Y that is constant on all $D \in \Lambda$.

Let X and Y be any random variables such that EX, EY, and $E(XY)$ exist. Then, $E[(X - M_X)Y] = E(XY) - E(M_XY) = E(XY) - E(XM_Y) = E[X(Y - M_Y)]$. Also, $E[(X - M_X)(Y - M_Y)] = E(XY) - E(M_XM_Y) =$

$E(XY) - E(M_X Y)$. Since $E(X - M_X) = E(Y - M_Y) = 0$, we have proved

$$\text{(11)} \qquad \text{Cov}\,(X - M_X, Y) = \text{Cov}\,(X, Y - M_Y) = \text{Cov}\,(X - M_X, Y - M_Y).$$

In particular, $\text{Cov}\,(X-M_X, X) = \text{Cov}\,(X-M_X, X-M_X) = \text{Var}\,(X-M_X)$. Also, $\text{Cov}\,(X - M_X, M_Y) = \text{Cov}\,(X, M_Y - M_{M_Y}) = \text{Cov}\,(X, 0) = 0$. Hence, the covariance between a random variable in the range of M and a random variable in the range of $1 - M$ is 0. In particular,

$$\text{Cov}\,(X - M_X, M_X) = 0.$$

Since $X = M_X + (X - M_X)$ and $\text{Var}\,X = \text{Var}\,M_X + \text{Var}\,(X - M_X) + 2\,\text{Cov}\,(M_X, X - M_X)$, it follows that

$$\text{(12)} \qquad \text{Var}\,X = \text{Var}\,M_X + \text{Var}\,(X - M_X),$$

that is, the variance of a random variable is the sum of the variance of its conditional expectation and the variance of the complement of its conditional expectation. It is apparent that $\text{Var}\,M_X/\text{Var}\,X$ and $\text{Var}\,(X - M_X)/\text{Var}\,X$ are numbers between 0 and 1. The following theorem is analogous to Theorem 1.

Theorem 5.

Let X be a random variable with finite non-zero variance, such that the complement of its conditional expectation $X - M_X$ has non-zero variance. Then,

$$\text{(13)} \qquad \rho^2(X, X - M_X) = \frac{\text{Var}\,(X - M_X)}{\text{Var}\,X}.$$

Proof. Since $\text{Cov}\,(X, X - M_X) = \text{Var}\,(X - M_X)$, substitution in

$$\rho^2(X, X - M_X) = \frac{[\text{Cov}\,(X, X - M_X)]^2}{\text{Var}\,X\,\text{Var}\,(X - M_X)}$$

gives this result. From (6), (12), and (13), we conclude that

$$\text{(14)} \qquad \rho^2(X, M_X) + \rho^2(X, X - M_X) = 1.$$

The following theorem is analogous to Theorem 2.

Theorem 6.

Let X and Y be any random variables with finite non-zero variance, such that $X - M_X$ and $Y - M_Y$ have non-zero variance. Then,

$$\text{(15)} \qquad \rho(X, Y - M_Y) \leq \rho(X, X - M_X),$$

that is, the correlation between a random variable X and the complement of its conditional expectation is not less than the correlation between X and the complement of the conditional expectation of any other random variable.

Proof. Since Cov $(X, Y - M_Y) =$ Cov $(X - M_X, Y - M_Y)$, by (11), and $[\text{Cov}(X - M_X, Y - M_X)]^2 \leq \text{Var}(X - M_X) \text{Var}(Y - M_Y)$, by the Cauchy-Schwartz inequality, substituting, dividing by Var X Var $(Y - M_Y)$, and using (13) gives this result.

The following theorem establishes an important property of the random variable $X - M_X$.

Theorem 7.

Let X and Y be any random variables with finite variance. Then,

$$\text{Var}(X - M_X) \leq \text{Var}(X - M_Y), \tag{16}$$

that is, the variance of the complement of the conditional expectation of X does not exceed the variance of the difference between X and the conditional expectation of any other random variable.

Proof. Since $X - M_Y = (X - M_X) + M_{X-Y}$ and since Cov $(X - M_X, M_{X-Y}) = 0$, we obtain Var $(X - M_Y) =$ Var $(X - M_X) +$ Var M_{X-Y}.

We now show the relation of these concepts to another sort of decomposition of Var X that is well-known in probability theory.

Definition 6.

The *conditional variance of X with respect to f* is a random variable $V_X : \Omega \to R$ defined by

$$V_X(\omega) = \text{Var}[X \mid \gamma(\omega)], \quad \text{for each } \omega \in \Omega.$$

If Var $[X \mid f]$ is a function that assigns the number Var $[X \mid f = \alpha]$ to each point α in the range of f, then $V_X =$ Var $[X \mid f] \circ f$, as in the case of the conditional expectation. The random variables V_X and M_X are related by $V_X = M_{X^2} - M_X^2$. Writing Var $M_X = EM_X^2 - (EM_X)^2$ and $EV_X = EM_{X^2} - EM_X^2$ and adding the two equations establishes that

$$\text{Var}\, X = \text{Var}\, M_X + EV_X\,; \tag{17}$$

that is, the unconditional variance of a random variable is the sum of the variance of its conditional expectation and the expectation of its conditional variance.

From (12) and (17) it follows that Var $(X - M_X) = EV_X$. However, note that

$$\begin{aligned} V_{X-M_X}(\omega) &= \text{Var}[(X - M_X) \mid \gamma(\omega)] \\ &= \text{Var}[X \mid \gamma(\omega)] = V_X(\omega), \text{ for each } \omega \in \Omega, \end{aligned}$$

so that $V_{X-M_X} = V_X$.

Test Procedures and Reliability

We now examine the foregoing results in the context of test theory. The conclusions obtained in the above theorems are purely formal consequences of probability theory and are not specific to any model. The reader will have observed, however, that these conclusions correspond to familiar formulas in test theory. This is not surprising, because test-theory models have been constructed in such a way that the hypotheses of these theorems have been generally satisfied. In the next section the nature of this correspondence will be stated more explicitly and it will be shown that the theorems about conditional expectation, as well as the test theory formulas, are special cases of properties of linear operators in Hilbert space.

We will now identify random variables and test scores, so that the previous theorems become assertions about test reliability, test validity, and so on.

Definition 7.

A *test procedure* is a 5-tuple $T = (\Omega, \mathfrak{A}, P, f, X)$, where Ω is a set of *outcomes*, $\mathfrak{A}$ is a collection of subsets of Ω representing *observable events*, $f : \Omega \rightarrow \Phi$ is an assignment of *individuals or experimental objects*, and $X : \Omega \rightarrow R$ is an assignment of *scores*, such that $(\Omega, \mathfrak{A}, P)$ is a probability space, f is a random point, and X is a random variable.

Every test procedure determines a collection of conditional random variables $\{X \mid f = \alpha\}$, $\alpha \in f(\Omega)$, which can be regarded as assignments of scores to particular individuals or objects. That is, the distribution induced by $X \mid f = \alpha$ represents the various scores that might be obtained by an individual α on repeated trials or replications of a test.

Otherwise expressed, an individual is first selected from a population, and then a measurement is taken, resulting in a score for that individual. The set Ω represents all conceivable outcomes of the entire procedure. We will see that this scheme is general enough to describe situations in which a collection of individuals are tested at the same time.

Example 6.

Table 1 describes a test procedure $T = (\Omega, \mathfrak{A}, P, f, X)$, as explained in Example 1. Here $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$, $\mathfrak{A}$ is the set of all subsets of Ω, the function P assigns to each set in $\mathfrak{A}$ $\frac{1}{4}$ times the number of elements in the set, and the functions f and X are as described in the table. It is easy to verify that $(\Omega, \mathfrak{A}, P)$ is a probability space, f is a random point, and X is a random variable.

Example 7.

Example 2 describes a test procedure $T = (\Omega, \mathfrak{A}, P, f, X)$, where $\Omega =$

$\{(\alpha, \beta) \varepsilon R^2 \mid 0 \leq \alpha \leq 1, \alpha \leq \beta \leq \alpha + 1\}$, $\mathfrak{A}$ is the σ-algebra of Borel sets, P is Lebesgue measure, and the functions f and X are as described.

Example 8.

Let $T = (\Omega, \mathfrak{A}, P, f, X)$ be a test procedure such that f is constant on Ω, that is, $f(\Omega)$ is a one-point set $\{\alpha\}$. Then X and $X \mid f = \alpha$ are the same function from Ω to R, and $M_X(\omega) = EX$, for all $\omega \varepsilon \Omega$. The procedure can be interpreted as taking measurements that are subject to random variability on a single individual.

Example 9.

Let $T = (\Omega, \mathfrak{A}, P, f, X)$ be a test procedure such that f is one-to-one. Then $E[X \mid \gamma(\omega)] = X(\omega)$, for all $\omega \varepsilon \Omega$, and $M_X = X$. The procedure can be interpreted as taking perfectly reliable measurements on randomly selected individuals.

Example 10.

Let $T = (\Omega, \mathfrak{A}, P, f, X_1, X_2, \cdots, X_n)$ be a vector-valued test procedure, such that $X_1, X_2, \cdots, X_n$ are conditionally independent and conditionally identically distributed with respect to f. Then, for every α in the range of f, $X_1 \mid f = \alpha, X_2 \mid f = \alpha, \cdots, X_n \mid f = \alpha$ is a finite sequence of independent, identically distributed random variables, or, in other words, a random sample of size n (see also Example 5).

Example 11.

We show that a stochastic process determines a test procedure. Let $\{X_\alpha, \alpha \varepsilon \Phi\}$ be a stochastic process with associated probability space $(\Omega, \mathfrak{A}, P)$ and parameter set Φ, let $\mathfrak{A}'$ be a σ-algebra of subsets of Φ, and let P' be a probability measure defined on $\mathfrak{A}'$. Construct the product probability space $(\Phi \times \Omega, \mathfrak{A}' \times \mathfrak{A}, P' \times P)$ and define a random variable $X : \Phi \times \Omega \to R$ by

$$X(\alpha, \omega) = X_\alpha(\omega), \quad \text{for each } (\alpha, \omega) \varepsilon \Phi \times \Omega,$$

and a random point $f : \Phi \times \Omega \to \Phi$ by

$$f(\alpha, \omega) = \alpha, \quad \text{for each } (\alpha, \omega) \varepsilon \Phi \times \Omega.$$

Then $(\Phi \times \Omega, \mathfrak{A}' \times \mathfrak{A}, P' \times P, f, X)$ is a test procedure, and $M_X = m_X \circ f$, where m_X is the mean-value function of the stochastic process. In particular, if $\Phi = \{\alpha_1, \alpha_2, \cdots, \alpha_k\}$ is a finite set of individuals or experimental objects, with $p(\alpha_i) = 1/k, i = 1, 2, \cdots, k$, the random vector $(X_{\alpha_1}, X_{\alpha_2}, \cdots, X_{\alpha_k})$ can be interpreted as assignment of scores to all individuals at the same time, and the corresponding random variable X represents variation of scores over all individuals.

We will say that X_1 and X_2 are *parallel* if they are conditionally independent and conditionally identically distributed (see definitions 2 and 3). In other words, X_1 and X_2 are parallel if there exists a vector-valued test procedure $T = (\Omega, \mathfrak{A}, P, f, X_1, X_2)$, such that X_1 and X_2 are conditionally independent and conditionally identically distributed with respect to f. Test procedures $T_1 = (\Omega_1, \mathfrak{A}_1, P_1, f_1, X_1)$ and $T_2 = (\Omega_2, \mathfrak{A}_2, P_2, f_2, X_2)$ will be called *equivalent* if (f_1, X_1) and (f_2, X_2) are identically distributed, that is, if the joint probability distributions induced on $\Phi \times R$ are equal. If $T = (\Omega, \mathfrak{A}, P, f, X_1, X_2)$ is such that X_1 and X_2 are parallel, then

$$T_1 = (\Omega, \mathfrak{A}, P, f, X_1) \quad \text{and} \quad T_2 = (\Omega, \mathfrak{A}, P, f, X_2)$$

are equivalent, but the converse is not generally true. Extension of these ideas to more than two random variables is obvious.

The concept of parallelism as used in test theory implies that "it makes no difference which test is used" as far as the conditional distributions of X_1 and X_2 are concerned. By the above definition, if $\{X_1, X_2, \cdots, X_n\}$ are parallel, then all ordered subsets $(X_{i_1}, X_{i_2}, \cdots, X_{i_m})$, $m \leq n$, containing the same number of elements are identically distributed random vectors. Hence, the concept of parallelism implies that the random variables are *exchangeable* or *symmetrically dependent* (see Feller, [1966], Loève, [1963], Rényi, [1970], for further discussion of the concept of exchangeable random variables). Note that Theorems 1 and 3 imply the following: the correlation between two exchangeable random variables equals the square of the correlation between either random variable and its conditional expectation.

If we require only that X_1 and X_2 be conditionally uncorrelated, instead of conditionally independent, and that Var X_1 = Var X_2 and $M_{X_1 - EX_1} = M_{X_2 - EX_2}$, instead of $P_{X_1|f} = P_{X_2|f}$, then we have a weaker concept of parallelism that leads to the same formulas, as mentioned above. The meaning of the weaker concept of parallelism will become apparent in the next section, when orthogonal random variables are considered. If random variables X_1 and X_2 are conditionally uncorrelated, then Cov $(X_1 - M_{X_1}, X_2 - M_{X_2}) = 0$, but the converse is not generally true. However, if Cov $(X_1 - M_{X_1}, X_2 - M_{X_2}$ $= 0$, then $E(X_1X_2) = E(M_{X_1}M_{X_2})$ and Cov (X_1, X_2) = Cov (M_{X_1}, M_{X_2}). Therefore, the conclusion of Theorem 3 can be obtained under a still weaker hypothesis: if X_1 and X_2 are such that Var X_1 = Var X_2, $M_{X_1 - EX_1} = M_{X_2 - EX_2}$, and Cov $(X_1 - M_{X_1}, X_2 - M_{X_2}) = 0$, then $\rho(X_1, X_2)$ = Var M_{X_1}/Var X_1, and similarly for Theorem 4.

Definition 8.

The *reliability coefficient* of a test procedure $T = (\Omega, \mathfrak{A}, P, f, X)$, defined if X has finite non-zero variance, is the ratio Var M_X/Var X, where M_X is the conditional expectation of X with respect to f.

Equivalent test procedures have the same reliability coefficient. If

Var $M_X = 0$, the reliability coefficient is defined and is 0. Theorem 1 shows that the reliability coefficient is also given by $\rho^2(X, M_X)$ if M_X has non-zero variance, and Theorem 3 shows that it is given by $\rho(X_1, X_2)$, where X_1 and X_2 are parallel and have the same distribution as X. Similarly, the *validity coefficient* of a vector-valued test procedure $T = (\Omega, \mathfrak{A}, P, f, X, Y)$ is the absolute value of $\rho(X, Y)$, where Y is a random variable representing a criterion.

The true scores of the classical test theory can be regarded as expected values of the conditional random variables $\{X \mid f = \alpha\}$, that is, as values taken by a function $t = E[X \mid f]$ defined on the range of f by $t(\alpha) = E[X \mid f = \alpha]$, for each $\alpha \in f(\Omega)$. The function t is related to the conditional expectation by $M_X = t \circ f$. The error scores of individuals are, similarly, a collection of random variables $\{e_\alpha\}$, defined by $e_\alpha = [X \mid f = \alpha] - t(\alpha)$ and are related to the complement of the conditional expectation by $[(X - M_X) \mid f = \alpha] = e_\alpha$. An additional assumption sometimes made in test theory is that, for arbitrary X and Y, Cov $(X - M_X, Y - M_Y) = 0$. None of the above results depend on this assumption, although in Theorem 3 a zero correlation between $X_1 - M_{X_1}$ and $X_2 - M_{X_2}$ is implied by the hypothesis, and similarly for Theorem 4.

Hilbert Space Interpretation

In previous sections of this paper, some of the main formulas in test theory were derived from properties of the conditional expectation of random variables. In this section we will show that these same properties of the conditional expectation that were needed to obtain the above results are in fact general properties of linear operators in Hilbert space. Consequently, the test-theory formulas have a simple geometric interpretation.

Let $(\Omega, \mathfrak{A}, P)$ be a probability space, and let $\mathfrak{B} \subseteq \mathfrak{A}$ be the σ-algebra induced by a random point $f : \Omega \to \Phi$. The collection of all random variables with finite variance defined on $(\Omega, \mathfrak{A}, P)$ up to equivalence with respect to P, denoted by $L_2(\mathfrak{A})$, is a Hilbert space under the inner product $(X_1, X_2) = E(X_1X_2)$ with corresponding norm $||X|| = (EX^2)^{1/2}$ [see, for example, Bachman and Narici, 1966, Burrill, 1972, Halmos, 1951]. The collection of all random variables with finite variance defined on $(\Omega, \mathfrak{B}, P)$, denoted by $L_2(\mathfrak{B})$, is a closed subspace of $L_2(\mathfrak{A})$.

In the case of finite probability spaces, if Ω contains n elements and if Λ contains m elements, then $L_2(\mathfrak{A})$ is an n-dimensional Euclidean space, and $L_2(\mathfrak{B})$ is an m-dimensional subspace. The inner product of random variables centered at expectation is the covariance, and the norm of a random variable centered at expectation is the standard deviation—that is, $(X - EX, Y - EY) = \text{Cov}\,(X, Y)$ and $||X - EX|| = \sigma X$.

For any random variable $X \in L_2(\mathfrak{A})$, the conditional expectation M_X is just the projection of X onto $L_2(\mathfrak{B})$, and the complement of the conditional expectation, $X - M_X$, is the projection of X onto the orthogonal complement of $L_2(\mathfrak{B})$. The operator M that assigns to each X its conditional expectation M_X and the operator $1 - M$ that assigns to each X the complement of the conditional expectation $X - M_X$ are orthogonal projections. Every random variable $X \varepsilon L_2(\mathfrak{A})$ is a unique sum of a random variable M_X in the subspace $L_2(\mathfrak{B})$ and a random variable $X - M_X$ in the orthogonal complement of $L_2(\mathfrak{B})$, and $||X||^2 = ||M_X||^2 + ||X - M_X||^2$. In particular, $||X - EX||^2 = ||M_{X-EX}||^2 + ||X - M_X||^2$, or Var X = Var M_X + Var $(X - M_X)$.

The relations $M_{M_X} = M_X$ and $(1 - M)_{X-M_X} = X - M_X$ correspond to the fact that a projection operator is idempotent and $M_{X-M_X} = 0$ and $(1 - M)_{M_X} = 0$ to the fact that the orthogonal complement of a subspace is projected onto the zero vector. The relation Cov $(M_X , Y - M_Y) = 0$, or, in other words, that the true score on any test is uncorrelated with the error score on any test, is just the fact that any vector in $L_2(\mathfrak{B})$ is orthogonal to any vector in the orthogonal complement of $L_2(\mathfrak{B})$.

The equality Cov (X, M_X) = Var M_X corresponds to $(X - EX, M_{X-EX}) = ||M_{X-EX}||^2$ and Cov $(X, X - M_X)$ = Var $(X - M_X)$ to $(X - EX, X - M_X) = ||X - M_X||^2$, which are properties of all orthogonal projections. The equality Cov (M_X , Y) = Cov (X, M_Y) corresponds to $(M_{X-EX} , Y - EY) = (X - EX, M_{Y-EY})$ and implies that the operator M is self-adjoint, and the equality Cov $(X - M_X , Y)$ = Cov $(X, Y - M_Y)$ corresponds to $(X - M_X , Y - EY) = (X - EX, Y - M_Y)$ and implies that the operator $1 - M$ is self-adjoint, again properties of all orthogonal projections.

Theorem 1 indicates that the cosine of the angle between $X - EX$ and M_{X-EX} equals $||M_{X-EX}||/||X - EX||$. Similarly, Theorem 5 indicates that the cosine of the angle between $X - EX$ and $X - M_X$ (or the sine of the angle between $X - EX$ and M_{X-EX}) equals $||X - M_X||/||X - EX||$, and the relation $\rho^2(X, M_X) + \rho^2(X, X - M_X) = 1$ is just the identity $\sin^2 + \cos^2 = 1$. Various formulas in test theory that relate variances and correlations of observed scores, true scores, and error scores can be obtained by substitutions in this and other trigonometric identities. Since these variance ratios equal correlation coefficients, we see that $||M_{X-EX}||/||X-EX|| \leq 1$ and $||X - M_X||/||X - EX|| \leq 1$, or the norm of each of the operators M and $1 - M$ is 1, another property of orthogonal projections.

Theorem 2 indicates that the cosine of the angle between $X - EX$ and $M(Y - EY)$ does not exceed the cosine of the angle between $X - EX$ and $M(X - EX)$, and Theorem 6 indicates that the cosine of the angle between $X - EX$ and $(1 - M)(Y - EY)$ does not exceed the cosine of the angle between $X - EX$ and $(1 - M)(X - EX)$. Both expressions correspond to the fact that the angle between a vector and its orthogonal projection onto a subspace is the minimal angle between that vector and all vectors in the subspace. The conclusion of Theorem 7 can be written

$$||(X - EX) - M_{X-EX}|| \leqq ||(X - EX) - M_{Y-EY}||$$

and corresponds to the fact that the distance between a vector and its orthogonal projection onto a subspace is the minimal distance between that vector and all vectors in the subspace.

Formulas relating random variables and their conditional expectations derived in previous sections of this paper are special cases of relations that hold in all Hilbert spaces. Hence, the test-theory results are, in turn, special cases of these general properties of linear operators in Hilbert space. Because of this correspondence, the formulas can be derived directly from these properties. For example, knowing that $L_2(\mathfrak{A})$ is a Hilbert space and that M is an orthogonal projection, one can conclude immediately that $M_{M_X} = M_X$ and that Cov (M_X , Y) = Cov (X, M_Y), since an orthogonal projection is idempotent and self-adjoint. Viewing the formulas in this setting may prove helpful in organizing the theory and perhaps suggesting new relationships in the context of test theory.

REFERENCES

Bachman, G. and Narici, L. *Functional analysis*. New York: Academic Press, 1966.

Burrill, C. W. *Measure, integration, and probability*. New York: McGraw-Hill, 1972.

Feller, W. *An introduction to probability theory and its applications* (Vol. II). New York: Wiley, 1966.

Gulliksen, H. *Theory of mental tests*. New York: Wiley, 1950.

Guttman, L. A basis for analyzing test-retest reliability. *Psychometrika*, 1945, **10**, 255–282.

Halmos, P. *Introduction to Hilbert space*. New York: Chelsea, 1951.

Loève, M. *Probability theory*. Princeton, N. J.: Van Nostrand, 1963.

Lord, F. M. and Novick, M. *Statistical theories of mental test scores*. Reading, Mass.: Addison-Wesley, 1968.

Rényi, A. *Foundations of probability*. San Francisco: Holden-Day, 1970.

Manuscript received 8/13/74
Revised manuscript received 12/9/74