The Annals of Statistics
1975, Vol. 3, No. 5, 1175-1179

A NOTE ON SUBSTITUTION IN CONDITIONAL DISTRIBUTION[1]

BY MICHAEL D. PERLMAN AND MICHAEL J. WICHURA

University of Chicago

The following proposition is sometimes used in distribution theory: for each fixed z suppose that $T(X, z)$ has the distribution Q and is independent of Y; then $T(X, Z(Y))$ has the distribution Q and is independent of Y. An example is presented to show this result is false in general. Additional conditions under which the proposition becomes valid are presented.

Let $(\Omega, \mathscr{F}, P)$ be a probability space. Let $(\mathscr{X}, \mathscr{A})$, $(\mathscr{Y}, \mathscr{B})$, $(\mathscr{Z}, \mathscr{C})$, and $(\mathscr{T}, \mathscr{D})$ be measurable spaces and suppose that $X: \Omega \to \mathscr{X}$, $Y: \Omega \to \mathscr{Y}$, $Z: \mathscr{Y} \to \mathscr{Z}$, and $T: \mathscr{X} \times \mathscr{Z} \to \mathscr{T}$ are respectively $\mathscr{F} - \mathscr{A}$, $\mathscr{F} - \mathscr{B}$, $\mathscr{B} - \mathscr{C}$, and $\mathscr{A} \otimes \mathscr{C} - \mathscr{D}$ measurable. Let Q be a probability measure on $(\mathscr{T}, \mathscr{D})$. The following proposition is commonly used in multivariate distribution theory and elsewhere.

PROPOSITION 1. *Suppose that*

(1) X *is independent of* Y, *and for each* $z \in \mathscr{Z}$ *the random object* $T(X, z)$ *has distribution* Q.

Then

(2) *the random object* $T(X, Z(Y))$ *also has distribution* Q *and is independent of* Y.

EXAMPLE 1 (cf. Rao (1973), 8.b.2). For a well-known application, let $X: p \times p$ and $Y: p \times 1$ be independent, with $X \sim W_p(n, \Sigma)$, $n \geqq p$, Σ p.d. Let $Z(Y) = Y$, and for nonzero $z: p \times 1$ define $T_1(X, z) = z'Xz/z'\Sigma z$, $T_2(X, z) = z'\Sigma^{-1}z/z'X^{-1}z$. Then for all z, $T_1(X, z) \sim \chi_p^2$ and $T_2(X, z) \sim \chi^2_{n-p+1}$, so by Proposition 1, $T_1(X, Y) \sim \chi_p^2$ and $T_2(X, Y) \sim \chi^2_{n-p+1}$ and each is independent of Y. These facts are used to derive the distribution of Hotelling's T^2-statistic. □

Sometimes, however, one needs to weaken the assumption (1) slightly, as in the following.

PROPOSITION 2. *Suppose that*

(3) *for each* $z \in \mathscr{Z}$, *the random object* $T(X, z)$ *has distribution* Q *and is independent of* Y.

Then (2) *holds.*

Received January 1974.

[1] This research was carried out in the Department of Statistics, University of Chicago, under partial support by Research Grant No. NSF GP 32037X from the Division of the Mathematical, Physical and Engineering Sciences of the National Science Foundation.

AMS 1970 *subject classifications.* Primary 62E15; Secondary 62H10.

Key words and phrases. Conditional distribution, independence, substitution, regular conditional probability, Wishart, matrix-variate beta.

EXAMPLE 2. Proposition 2 has been applied by Mitra (1970) in his study of the matrix-variate beta distribution. Let $S_1: k \times k$ and $S_2: k \times k$ be independent random matrices, with $S_i \sim W_k(n_i, \Sigma)$ for $i = 1, 2$, $n_1 + n_2 \geqq k$, and Σ p.d. Let $S = S_1 + S_2$ and $U = S^{-\frac{1}{2}} S_1 S^{-\frac{1}{2}}$, where $S^{\frac{1}{2}}$ is a lower triangular square root of S (any other square root of S could be chosen). In his Lemma 3.4 Mitra applies Proposition 2 with $X = (S_1, S_2)$, $Y = S$, $Z(Y) = S^{-\frac{1}{2}}a$, $T_3(X, z) = z'S_1z/z'Sz$, and $Q_3 = \text{Beta}(n_1/2, n_2/2)$, where $a: p \times 1$ is a nonzero fixed vector, and concludes that $a'Ua/a'a \sim Q_3$ and is independent of S. (The independence of $T_3(X, z)$ and $Y = S$ follows from Theorem 2 of Basu (1955), since S is a complete and sufficient statistic for Σ while the distribution of $T_3(X, z)$ does not depend on Σ.) Other applications of Proposition 2 occur in Mitra's Lemmas 3.10 and 3.11, where it is shown that $a'a/a'U^{-1}a \sim \text{Beta}((n_1 - k + 1)/2, n_2/2)$ (provided $n_1 \geqq k$) and $LUL' \sim U$ if $L: k \times k$ is an orthogonal matrix. □

Proposition 2 has a deceptively easy "proof," which runs as follows:

$T(X, z)$ has distribution Q and is independent of Y, for all $z \in \mathscr{Z}$
$\Rightarrow$ Conditional distribution of $T(X, z)$, given $Y = y$, is Q, for all $y \in \mathscr{Y}$ and all $z \in \mathscr{Z}$
$\Rightarrow$ Conditional distribution of $T(X, Z(y))$, given $Y = y$, is Q, for all $y \in \mathscr{Y}$
$\Rightarrow$ Conditional distribution of $T(X, Z(Y))$, given $Y = y$, is Q, for all $y \in \mathscr{Y}$
$\Rightarrow$ $T(X, Z(Y))$ has distribution Q, and is independent of Y.

On close inspection, however, this argument breaks down. In fact Proposition 2 is false, as may be seen from

EXAMPLE 3. Take $X = Y = Z(Y)$ to have a uniform distribution on $[0, 1]$ and take $T(x, z) = zI_{\{z\}}(x)$, where I denotes the indicator function. Then $T(X, z)$ is degenerate at 0 and hence independent of Y for each z, but $T(X, Z(Y)) = X$ is uniformly distributed on $[0, 1]$ and highly dependent on Y. □

The major flaw in the above "proof" occurs in the first step, wherein the correct conclusion is that for each z the conditional distribution of $T(X, z)$ given $Y = y$ is Q for PY^{-1}—almost all y; as the null sets here may depend on z, the second step may not be permissible.

We now present a variety of supplementary conditions under which Proposition 2 is valid. Suppose first that

(4) $\quad Y$ has a discrete distribution.

Then for each $y \in \mathscr{Y}$ and $D \in \mathscr{D}$, (3) implies

$$P\{Y = y, T(X, Z(Y)) \in D\} = P\{Y = y, T(X, Z(y)) \in D\} = P\{Y = y\}Q(D) .$$

Summing over y yields

(5) $$P\{Y \in B, T(X, Z(Y)) \in D\} = P\{Y \in B\}Q(D)$$

for each $B \in \mathscr{B}$, which is equivalent to (2).

In all remaining cases we assume that

(6) X admits a regular conditional probability distribution given Y;

thus we have a family $\{L_y \mid y \in \mathscr{Y}\}$ of probability measures on $(\mathscr{X}, \mathscr{A})$ such that $L_\bullet(A)$ is $\mathscr{B}$-measurable for each $A \in \mathscr{A}$, and

$$P\{X \in A,\ Y \in B\} = \int_B L_y(A) PY^{-1}(dy)$$

for all $A \in \mathscr{A}$, $B \in \mathscr{B}$, where PY^{-1} denotes the probability distribution of Y induced by P. We remark that if $g(x, z)$ is a bounded, $\mathscr{A} \otimes \mathscr{C}$-measurable function then $\int_{\mathscr{X}} g(x, z) L_y(dx)$ is $\mathscr{B} \otimes \mathscr{C}$-measurable in (y, z). This is true if g is a measurable rectangle, and the general case follows by the approximation argument used to prove Fubini's Theorem (Neveu (1965), Proposition III.2.1). In our first theorem we shall also assume that

(7) $\mathscr{T}$ is a topological space, and $\mathscr{D}$ is such that any finite measure ν on $\mathscr{D}$ is uniquely determined by $\{\int_{\mathscr{T}} h\, d\nu \mid h \in H\}$, where H is the collection of bounded real continuous functions on H.

This is the case, for example, if $\mathscr{T}$ is a metric space and $\mathscr{D}$ its Borel σ-algebra, or if $\mathscr{T}$ is a compact space and $\mathscr{D}$ its Baire σ-algebra (Neveu (1965), Proposition II.7.2).

THEOREM 1. *Suppose* (6) *and* (7) *hold. Then* (3) *implies* (2) *if either*

(8a) *there exists a first countable topology on* Z *such that* $T(x, \bullet)$ *is continuous on* $\mathscr{Z}$ *for all* $x \in \mathscr{X}$, *and*

(8b) *there exists a σ-finite measure* μ *on* $(\mathscr{Z}, \mathscr{C})$ *such that the complement of any μ-null set is dense in* $\mathscr{Z}$,

or

(9a) $\mathscr{X}$ *and* $\mathscr{Y}$ *are topological spaces, with* $\mathscr{Y}$ *second countable*; $\mathscr{A}$ *and* $\mathscr{B}$ *are their Baire and Borel σ-algebras, respectively, and*

(9b) $T(\bullet, z)$ *is continuous on* $\mathscr{X}$ *for each* $z \in \mathscr{Z}$, *and*

(9c) $y_n \to y$ *in* $\mathscr{Y}$ *implies* $L_{y_n} \to L_y$ *weakly, in the sense that* $\int_{\mathscr{X}} g\, dL_{y_n} \to \int_{\mathscr{X}} g\, dL_y$ *for all bounded real continuous functions* g *on* $\mathscr{X}$.

PROOF. It suffices to show that for each $h \in H$,

$$E\{h[T(X, Z(Y))] \mid Y\} = \int_{\mathscr{T}} h\, dQ\,, \tag{10}$$

since this implies

$$E\{I_B(Y) h[T(X, Z(Y))]\} = P\{Y \in B\} \int_{\mathscr{T}} h\, dQ \tag{11}$$

for each $B \in \mathscr{B}$ which, by (7), implies (5). Define

$$\psi(y, z) = \int_{\mathscr{X}} h[T(x, z)]L_y(dx) , \qquad J = \int_{\mathscr{T}} h\, dQ .$$

By an argument of Bahadur and Bickel (1968, page 378, lines 4–15) the function $y \to \psi(y, Z(y))$ is a version of the conditional expectation of $h[T(X, Z(Y))]$ given Y, so to prove (10) it suffices to show that

(12) $$\psi(y, Z(y)) = J \qquad \text{for almost all} \quad y \in \mathscr{Y} .$$

For each z, $y \to \psi(y, z)$ is a version of $E\{h[T(X, z)] \mid Y\}$, so (3) implies

(13) for each $z \in \mathscr{Z}$, $\psi(y, z) = J$ for almost all $y \in \mathscr{Y}$, where the exceptional set may depend on z.

Now assume (8). By (13), an application of Fubini's theorem to the $\mathscr{B} \otimes \mathscr{C}$-measurable set $\{(y, z) \mid \psi(y, z) = J\}$ gives the existence of a PY^{-1}-null set Λ such that $y \notin \Lambda$ implies $\psi(y, z) = J$ for μ-almost all $z \in \mathscr{Z}$. But (8a) and the Lebesgue dominated convergence theorem imply that $\psi(y, \cdot)$ is continuous on $\mathscr{Z}$, and hence (8b) yields

(14) $$\text{for all} \quad y \notin \Lambda , \quad \psi(y, z) = J \quad \text{for } \textit{all} \quad z \in \mathscr{Z} ,$$

which implies (12).

Next assume (9). Then $\psi(\cdot, z)$ is continuous on $\mathscr{Y}$ for each z, so (13) implies (14) with Λ taken to be the complement of the support of PY^{-1}, i.e., Λ is the (countable) union of all open, PY^{-1}-null sets in $\mathscr{Y}$. □

THEOREM 2. *Suppose* (6) *holds. Then* (3) *implies* (2) *if*

(15a) $\mathscr{Y}$ *is a second countable topological space, with* $\mathscr{B}$ *its Borel* σ-*algebra, and*

(15b) $y_n \to y$ *in* $\mathscr{Y}$ *implies* $L_{y_n} \to L_y$ *in the setwise sense that* $L_{y_n}(A) \to L_y(A)$ *for each* $A \in \mathscr{A}$.

PROOF. It suffices to show that (10) holds when $h = I_D$, for each $D \in \mathscr{D}$, for this directly yields (5). Now (3) again implies (13), where now

$$\psi(y, z) = \int_{\mathscr{X}} I_D[T(x, z)]L_y(dx) , \qquad J = Q(D) ,$$

and (15b) implies that $\psi(\cdot, z)$ is continuous on $\mathscr{Y}$ for each z, so (14) holds with Λ the complement of the support of PY^{-1}. This implies (12), and therefore (10). □

Some remarks are in order concerning these theorems.

1. All topological assumptions concerning $(\mathscr{X}, \mathscr{A})$, $(\mathscr{Y}, \mathscr{B})$, $(\mathscr{Z}, \mathscr{C})$, and $(\mathscr{T}, \mathscr{D})$ are satisfied if they are in fact metric spaces with the associated Borel σ-algebras, and $\mathscr{Y}$ is separable.

2. Inspection of the proof of Theorem 1 shows that in (8a), continuity of $T(x, \cdot)$ on $\mathscr{Z}$ could be replaced by the weaker hypothesis that for almost all $y \in \mathscr{Y}$, $T(x, \cdot)$ is continuous at $Z(y)$ for L_y-almost all x. However, this hypothesis

requires detailed knowledge of the L_y's, about which virtually nothing except their existence may be known. Example 3 shows that "$T(x, \cdot)$ is continuous at $Z(y)$ for L_y-almost all x" cannot be replaced by "$T(x, \cdot)$ is continuous at $Z(y)$ for PX^{-1}-almost all x" two sentences above.

3. A typical measure μ satisfying (8b) is Lebesgue measure on a Euclidean space. In this case, the continuity assumption in (8a) can be weakened. Thus if $\mathscr{X}$ is an interval on the real line, it is enough to assume that $T(x, \cdot)$ is either right or left continuous at each point in $\mathscr{X}$.

4. By Scheffe's theorem (Rao (1973), 2c. 4(xv)), the hypothesis of setwise convergence in (15b) holds if the L_y's all have densities with respect to some σ-finite measure on $(\mathscr{X}, \mathscr{A})$ and these densities vary continuously with y. As can be seen from Example 2, however, detailed knowledge about the L_y's may not be available; thus condition (8) is probably more useful than either (9) or (15).

Finally, it is of interest to compare Proposition 2 to the oft-used

PROPOSITION 3. *Suppose* (6) *holds. If*

(16) *for almost all* $y \in \mathscr{Y}$, *the conditional distribution of* $T(X, z)$ *given* $Y = y$ (i.e., $L_y T(\cdot, z)^{-1}$) *equals* Q *for all* $z \in \mathscr{X}$,

then (2) *holds.*

The validity of Proposition 3 is immediate, since (16) implies that (14) holds, where $\phi(y, z)$ and J are defined as in the proof of Theorem 2. At first sight, (3) and (16) seem to be equivalent. As Example 3 shows, however, (3) is in fact strictly weaker than (16)—indeed, too weak, unless buttressed by additional assumptions such as (8), (9), or (15). Also, notice that Proposition 3 implies Proposition 1, for if X is independent of Y then we may take $L_y = PX^{-1}$ for all y.

REFERENCES

[1] BAHADUR, R. R. and BICKEL, P. J. (1968). Substitution in conditional expectation. *Ann. Math. Statist.* **39** 377–378.

[2] BASU, D. (1955). On statistics independent of a complete sufficient statistic. *Sankhya* **15** 377–380.

[3] MITRA, S. K. (1970). A density-free approach to the matrix variate beta distribution. *Sankhya Ser. A.* **32** 81–88.

[4] NEVEU, J. (1965). *Mathematical Foundations of the Calculus of Probabilities.* Holden-Day, San Francisco.

[5] RAO, C. R. (1973). *Linear Statistical Inference and its Applications* (2nd ed.). Wiley, New York.

DEPARTMENT OF STATISTICS
THE UNIVERSITY OF CHICAGO
1118 EAST 58TH STREET
CHICAGO, ILLINOIS 60637