The Annals of Probability
1976, Vol. 4, No. 5, 869-873

CONDITIONS ON THE REGRESSION FUNCTION WHEN BOTH VARIABLES ARE UNIFORMLY DISTRIBUTED[1]

BY R. A. VITALE AND A. C. PIPKIN

Brown University

Consider the class of probability distributions on the unit square with uniform marginal distributions. The class of associated regression functions is characterized by integral inequalities or, equivalently, as forming the closed convex hull of 1-1, measure-preserving mappings of the unit interval onto itself.

1. Let $\mathscr{F}$ be the class of probability distribution functions F on the unit square with uniform marginal distributions functions

$$F(x, 1) = x\,, \qquad F(1, y) = y\,, \qquad\qquad x, y \in [0, 1]\,.$$

We seek to characterize the class $m(\mathscr{F})$ of associated regression functions

$$m(x) = E[Y \mid X = x]\,.$$

We show that condition C, below, characterizes these functions.

Let us explain the main ideas of the theorem in the analogous discrete context. Consider doubly stochastic matrices P. The regression vector for such a matrix is $m = Pv$, where $v = (1, 2, \cdots, n)$. If P is a permutation matrix, then m is a permutation of v. In the general case, since any doubly stochastic matrix is a convex combination of permutation matrices, it follows that m is a convex combination of permutations of v. Now the sum of any k components of v is at least $k(k+1)/2$, and the sum of all of its components is $n(n+1)/2$. The same is therefore true of m. The more difficult part of the theorem is the converse: m is indeed in the convex hull of permutations of v (and hence representable as Pv) if these conditions on sums of components are satisfied.

For the proof of the theorem in its continuous form, we use the following definition of the class of regression functions under consideration (see, for example, Loéve [1]:

DEFINITION. A Borel-measurable function m is in $m(\mathscr{F})$ if, for some pair of random variables X, Y with distribution function in $\mathscr{F}$, $m(X) = E[Y \mid X]$; that is,

$$\int_0^1 m(x)I_A(x)\,dx = \int_0^1 \int_0^1 yI_A(x)F(dx, dy) \tag{1}$$

for all indicator functions I_A of Borel subsets of $[0, 1]$.

Here, as well as later on, we are content to have m uniquely defined almost everywhere (Lebesgue measure).

Received December 15, 1972; revised March 3, 1976.
[1] Supported in part by Grant GJ-710 from the National Science Foundation.
AMS 1970 *subject classifications.* Primary 62J05; Secondary 28A65.
Key words and phrases. Regression function, measure-preserving mapping.

Measure-preserving transformations take the place of the permutations in the analogous discrete case:

DEFINITION. $\mathscr{T} = \{T\colon [0, 1] \to [0, 1]$, invertible, Borel-measurable, measure-preserving$\}$.

Condition C is analogous to the system of bounds on the sums of components:

DEFINITION. $f \in L_2[0, 1]$ satisfies condition C iff

$$\text{(C)} \qquad \int_0^u f(T(x))\, dx \geqq \frac{u^2}{2}, \qquad 0 \leqq u \leqq 1,$$

for all $T \in \mathscr{T}$ and with equality for $u = 1$.

Our main result is the following:

THEOREM. *The following statements are equivalent*:

(i) $m \in m(\mathscr{T})$.
(ii) m *satisfies condition* C.
(iii) $m \in$ *closed convex hull of* $\mathscr{T}$ (L_2 *norm*).

The proof, in Section 2, will proceed by showing that

$$\text{(i)} \to \text{(ii)} \to \text{(iii)} \to \text{(i)}.$$

In Section 3 we show how a distribution F corresponding to a given regression function m can be constructed in certain special cases.

2. Proof of theorem.

(i) $\to$ (ii). A direct application of (1) shows that $0 \leqq m(x) \leqq 1$ a.e. and thus that $m \in L_2[0, 1]$. To verify condition C, we first note that $m \in m(\mathscr{T}) \Rightarrow m \circ T \in m(\mathscr{T})$, $\forall\, T \in \mathscr{T}$. Using (1), we have

$$\begin{aligned}\int_0^1 I_A(x) m(T(x))\, dx &= \int_0^1 I_{T(A)}(x) m(x)\, dx \\ &= \int_0^1 \int_0^1 y I_A(x) G(dx, dy),\end{aligned}$$

where

$$I_A(x) G(dx, dy) = I_{T(A)}(x) F(dx, dy).$$

It is easily verified that $G \in \mathscr{F}$. Hence it is sufficient to show (C) with $T(x) = x$.

We omit the trivial proof that equality holds in (C) at $u = 1$. For $u \in (0, 1)$, let $A = [0, u]$ in (1). Then

$$\begin{aligned}\int_0^u m(x)\, dx &= E\{I_A(X) E[Y \mid X]\} = E[Y I_A(X)] \\ &= P[X \leqq u] E[Y \mid X \leqq u] = u E[Y \mid X \leqq u].\end{aligned}$$

Now, conditioned on the event $X \leqq u$, Y is stochastically at least as large as X: for $0 \leqq y \leqq u$,

$$P(Y \leqq y \mid X \leqq u) = \frac{F(u, y)}{u} \leqq \frac{F(1, y)}{u} = \frac{y}{u} = P(X \leqq y \mid X \leqq u).$$

Hence

$$E[Y \mid X \leqq u] \geqq E[X \mid X \leqq u] = \frac{u}{2},$$

and it follows that

$$\int_0^u m(x)\,dx \geqq \frac{u^2}{2}.$$

(ii) → (iii). We use the following criterion (see, for example, Luenberger [2]): If $(m, \phi) \geqq a \; \forall \phi \in L_2[0, 1]$, where a satisfies

$$(T, \phi) = \int_0^1 T(x)\phi(x)\,dx \geqq a \quad \forall T \in \mathscr{T},$$

then $m \in$ closed convex hull of $\mathscr{T}$.

For ϕ a step function, there is some $T_\phi \in \mathscr{T}$ such that $\phi \circ T_\phi$ is a monotone nonincreasing step function,

$$\phi(T_\phi(x)) = c_0 + \sum_{k=1}^{s} c_k I_{[0, x_k]}(x) \quad \text{a.e.},$$

where

$$c_k > 0 \quad (k \geqq 1) \qquad \text{and} \qquad 0 < x_1 < \cdots < x_s < 1.$$

Then

$$\int_0^1 m(x)\phi(x)\,dx = \int_0^1 m(T_\phi(x))\phi(T_\phi(x))\,dx = c_0 J(1) + \sum_{k=1}^{s} c_k J(x_k),$$

where

$$J(x) = \int_0^x m(T_\phi(u))\,du.$$

By (C), $J(1) = \frac{1}{2}$ and $J(x) \geqq x^2/2$. Hence

$$\int_0^1 m(x)\phi(x)\,dx \geqq \frac{c_0}{2} + \sum_{k=1}^{s} \frac{c_k x_k^2}{2}.$$

But the right-hand member is equal to the integral of $x\phi \circ T_\phi$, so

$$\int_0^1 m(x)\phi(x)\,dx \geqq \int_0^1 x\phi(T_\phi(x))\,dx = \int_0^1 T_\phi^{-1}(x)\phi(x)\,dx.$$

Hence $(T_\phi^{-1}, \phi) \geqq a \Rightarrow (m, \phi) \geqq a$.

For ϕ not a step function, we use a sequence of step functions approaching ϕ in norm to obtain the desired result. The straightforward but lengthy details are omitted.

(iii) → (i). Since $m \in$ closed convex hull of $\mathscr{T}$, there is a sequence $\{m_n\}$ converging to m (L_2 and a.e.) in which each m_n is a convex combination of functions in $\mathscr{T}$:

$$m_n(x) = \sum_{k=1}^{N(n)} \theta_{nk} T_{nk}(x), \qquad T_{nk} \in \mathscr{T},$$

with positive weights θ_{nk} whose sum is unity. Let G_n be defined by

$$I_A(x) G_n(dx, dy) = \sum_{k=1}^{N(n)} \theta_{nk} I_{T_{nk}(A)}(x) F^*(dx, dy),$$

where F^* assigns uniform measure to the diagonal $y = x$. It is straightforward to verify that $G_n \in \mathscr{F}$ with associated regression function m_n:

$$\int_0^1 I_A(x) m_n(x)\,dx = \int_0^1 \int_0^1 y I_A(x) G_n(dx, dy).$$

As $n \to \infty$, the left-hand side tends to the corresponding member of (1). As for the right-hand side, we extract a subsequence of the G_n converging weakly to some G. Via standard arguments, it is seen that $G \in \mathscr{F}$ and that the limit of the right-hand side is

$$\int_0^1 \int_0^1 y I_A(x) G(dx, dy) .$$

Comparing with (1), we conclude that $m \in m(\mathscr{F})$.

3. A construction. Given m satisfying condition C, it is sometimes possible to show that $m \in m(\mathscr{F})$ by direct construction of a distribution with regression function m. We show how to do this for cases in which $m \in C^1[0, 1]$ and $0 < m'(x) < 1$. Specifically, we construct a singular distribution concentrated on $y = x$ and $y = g(x)$, where $g(x)$ is yet to be specified. Let $F(y \mid X = x)$ be a conditional distribution function with atoms at $y = x$ and $y = g(x)$, and let

$$F(x, y) = \int_0^x F(y \mid X = u)\, du .$$

We take the sum of the two atoms in $F(y \mid X = x)$ to be unity in order to produce the required uniform marginal distribution in x and require that they yield the correct conditional expectation $m(x)$. These conditions determine the two atoms as

$$\frac{g(x) - m(x)}{g(x) - x} \quad (\text{at} \quad y = x) \qquad \text{and} \qquad \frac{m(x) - x}{g(x) - x} \quad (\text{at} \quad y = g(x)) .$$

At points where $g(x) = x$, we let $F(y \mid X = x)$ have a single atom of weight 1 at $y = x$.

The function g is determined from the condition $F(1, y) = y$. We describe the solution without indicating how it was derived. First, we introduce the auxiliary function

$$I(u) = \int_0^u (m(x) - x)\, dx .$$

From the assumed properties of m, it follows that I is positive on (0, 1) and strictly concave downward, with $I(0) = I(1) = 0$. Then the equation $I(y) = I(x)$ has a unique nonidentical solution $y = g(x)$ for all x except the point x_0 where $m(x_0) = x_0$; there we define $g(x_0) = x_0$.

The following properties of g will be used:

(i) $g(g(x)) = x,\ g(0) = 1,\ g(1) = 0.$

(ii) $x < x_0 \Rightarrow x < m(x) < m(g(x)) < g(x),$
$x > x_0 \Rightarrow x > m(x) > m(g(x)) > g(x).$

(iii) $g'(x) = \dfrac{I'(x)}{I'(g(x))} = \dfrac{m(x) - x}{m(g(x)) - g(x)} < 0\ (x \neq x_0).$

We now verify that $F(1, y) = y$. Suppose that $y < x_0$ (the verification for $y > x_0$ is similar). Then

$$\begin{aligned} F(y \mid X = u) &= \frac{g(u) - m(u)}{g(u) - u} && (u \leqq y) , \\ &= 0 && (y < u < g(y)) , \\ &= \frac{m(u) - u}{g(u) - u} && (g(y) \leqq u) . \end{aligned}$$

Hence $F(1, y) = S_1 + S_2$, where

$$S_1 = \int_0^y \frac{g(u) - m(u)}{g(u) - u} du , \qquad S_2 = \int_{g(y)}^1 \frac{m(u) - u}{g(u) - u} du .$$

By making the change of variable $u = g(v)$ in S_2 and using the previously listed properties of g, we obtain

$$S_2 = \int_y^0 \frac{m(g(v)) - g(v)}{v - g(v)} \cdot \frac{m(v) - v}{m(g(v)) - g(v)} dv = \int_0^y \frac{m(v) - v}{g(v) - v} dv .$$

It then follows immediately that $S_1 + S_2 = y$.

Acknowledgment. The authors thank Professor A. Marshall for a suggestion that has been incorporated into the proof of the theorem.

REFERENCES

[1] Loéve, M. (1965). *Probability Theory.* Van Nostrand, Princeton.
[2] Luenberger, D. (1969). *Optimization by Vector Space Methods.* Wiley, New York.

Mathematics Research Center
Madison, Wisconsin 53706

Division of Applied Mathematics
Brown University
Providence, Rhode Island 02912

Note added in proof. We thank Professor R. Sine for suggesting the relevance of J. V. Ryff (1965), "Orbits of L^1-functions under doubly stochastic transformations," *Trans. Amer. Math. Soc.* **117** 92–100; and J. R. Brown (1966), "Approximation theorems for Markov operators," *Pacific J. Math.* **16** 13–23.