The Annals of Statistics
1979, Vol. 7, No. 3, 653-658

REGRESSION WITH GIVEN MARGINALS[1]

BY RICHARD A. VITALE

Claremont Graduate School

We consider the class of regression functions $\mathfrak{M}(F, G) = \{m(x) = E[Y|X = x], (X, Y) \in \Pi(F, G)\}$ where $\Pi(F, G)$ denotes the set of random vectors with marginal distributions F and G. A characterization of $\mathfrak{M}(F, G)$ is given together with a representation for the projection operator it induces in an appropriate Hilbert space.

1. Introduction.

Let $\Pi(F, G)$ denote the class of random vectors (X, Y) with marginal distributions F and G ($X \sim F$, $Y \sim G$). We shall consider the associated class of regression functions

$$\mathfrak{M}(F, G) = \{m(x) = E[Y|X = x], (X, Y) \in \Pi(F, G)\}.$$

The motivation for looking at this class is similar in spirit to that of isotonic regression (from which we shall, in fact, borrow a result): the extent to which auxiliary information can be incorporated into the regression process. Knowledge of marginal distributions, in particular, is natural in certain types of problems—for instance, a census in which bivariate observations are collected, marginal distributions are known (as from a previous survey), and regression is desired. Alternatively, one may consider the problem of optimal, nonlinear prediction in a stationary time series $\{X_i\}$. If F is the equilibrium distribution of the X_i, then the optimal one-step predictor (squared error loss) is $E[X_{i+1}|X_i = x] \in \mathfrak{M}(F, F)$ (see [3], [5], [6] for related discussions of this problem).

In Section 2, we begin by presenting a characterization of $\mathfrak{M}(F, G)$ for a large class of F and G. Characterizations of the type indicated have been investigated from a variety of points of view and we refer the reader to [7], [9] for other discussions and references. It is fair to state that the common ancestor of all such approaches is the fertile theorem of Hardy, Littlewood and Polya [4, page 49] on the averaging properties of doubly stochastic matrices. In Section 3, we investigate further the structure of $\mathfrak{M}(F, G)$ by considering it as a convex subset of an appropriate Hilbert space and examining the induced projection operator.

2. Characterization of $\mathfrak{M}(F, G)$.

In what follows we shall regard F and G as fixed and satisfying

(A1) F and G are each supported on all of R^1 and are invertible.

(A2) $EY^2 = \int_{-\infty}^{+\infty} y^2 G(dy) < \infty$.

Received January 1977; revised June 1978.

[1]This work was done at the Mathematics Research Center (Madison) and sponsored by the United States Army under Contract No. DAAG29-75-C-0024.

AMS 1970 *subject classifications.* Primary 62J05; secondary 28A65, 46C10, 60G25.

Key words and phrases. Regression, isotonic regression, convex minorant, rearrangement of a function, nonlinear prediction.

The first assumption can be weakened considerably, but we present it to avoid side-issues. The second ensures that $\mathfrak{M}(F, G)$ is a subset of $L_2[(-\infty, +\infty); F]$, the Hilbert space of real-valued functions on R^1 which are square integrable with respect to the measure determined by F (this can be seen directly by noting that $EY^2 = E_X E[Y^2|X] \geqslant E_X(E[Y|X])^2)$.

In characterizing $\mathfrak{M}(F, G)$, we note that if $m(x) = E[Y|X = x] \in \mathfrak{M}(F, G)$, then with the application of marginal probability transformations $U = F(X)$, $V = G(Y)$, we have $m(x) = E[G^{-1}(V)|U = F(x)]$, where U and V are each uniformly distributed on [0, 1]. This is essentially the object of study in [10] and, with only minor modifications, the methods employed there yield the following result.

THEOREM 1. *The following statements are equivalent.*

(i) $m \in \mathfrak{M}(F, G)$.

(ii) $\int_0^x m(F^{-1}(T(u)))\,du \geqslant \int_0^x G^{-1}(u)\,du$ *for all* $x \in [0, 1]$ (*with equality at* $x = 1$) *and all* $T \in \mathfrak{T}$.

(iii) m *lies in the close convex hull* $(L_2[(-\infty, +\infty); F])$ *of functions of the form* $G^{-1} \circ T \circ F$.

Here $\mathfrak{T} = \{T : [0, 1] \to [0, 1]$ one-one, Borel-measurable, measure-preserving$\}$. We note that if $m \circ F^{-1}$ is nondecreasing, then the strongest inequality in (ii) occurs upon taking $T(u) = u$, i.e.,

$$\int_0^x m(F^{-1}(u))\,du \geqslant \int_0^x G^{-1}(u)\,du.$$

The equality condition in (ii) amounts to $\int_{-\infty}^{+\infty} m(x)F(dx) = \int_{-\infty}^{+\infty} yG(dy)$ or $Em(X) = EY$. Finally, for the projection problem it will be useful to note that the mapping $h \in L_2[(-\infty, +\infty); F] \to h \circ F^{-1} \in L_2[[0, 1]; \mu = \text{Lebesgue measure}]$ induces an isometry between the two spaces. The image $\mathfrak{M}_0$ of $\mathfrak{M}(F, G)$ under this mapping (which is, in fact, $\mathfrak{M}(R, G)$, R the uniform distribution on [0, 1]) can be described as follows.

COROLLARY. *The following are equivalent.*

(i) $m_0 \in \mathfrak{M}_0$.

(ii) $\int_0^x m_0(T(u))\,du \geqslant \int_0^x G^{-1}(u)\,du$ *for all* $x \in [0, 1]$ (*with equality at* $x = 1$) *and all* $T \in \mathfrak{T}$.

(iii) m_0 *lies in the closed convex hull* $(L_2[[0, 1]; \mu])$ *of functions of the form* $G^{-1} \circ T$.

PROOF. Change of variables.

REMARK. From (iii), it is evident that for each $T \in \mathfrak{T}$, $m_0 \in \mathfrak{M}_0 \Leftrightarrow m_0 \circ T \in \mathfrak{M}_0$.

3. Projection. Under the assumption $(X, Y) \in \Pi(F, G)$, a natural criterion for judging an estimate $\hat{m}(x)$ of the unknown regression function $m(x)$ is the squared error loss

$$E[m(x) - \hat{m}(x)]^2 = \int_{-\infty}^{+\infty}[m(x) - \hat{m}(x)]^2 F(dx).$$

It is evident that this loss can be reduced (or at least made no larger) by constructing a new estimate $\tilde{\hat{m}}(x)$ which is the projection of $\hat{m}$ onto the convex $\mathfrak{M}(F, G)$. For this reason, it is of interest to investigate the projection operator associated with $\mathfrak{M}(F, G)$ in $L_2[(-\infty, +\infty); F]$: that is, for $h \in L_2[(-\infty, +\infty); F]$, we seek the (unique) element $\tilde{h} \in \mathfrak{M}(F, G)$ which yields

$$\int_{-\infty}^{+\infty}[h(x) - \tilde{h}(x)]^2 F(dx) = \inf_{m \in \mathfrak{M}(F, G)} \int_{-\infty}^{+\infty}[h(x) - m(x)]^2 F(dx)$$

($\tilde{}$ throughout will denote projection in the appropriate space). A feature of this projection is that if a constant is added to h, then $\tilde{h}$ remains the same: this can be seen by expanding

$$\begin{aligned}\int_{-\infty}^{+\infty}[h(x) + c - m(x)]^2 F(dx) = \int_{-\infty}^{+\infty}&[h(x) - m(x)]^2 F(dx)\\ &+ c^2 + 2c\int_{-\infty}^{+\infty} h(x) F(dx)\\ &- 2c\int_{-\infty}^{+\infty} m(x) F(dx)\end{aligned}$$

and noting that the first term alone depends on m since, as we have noted, $\int_{-\infty}^{+\infty} m(x) F(dx) \equiv \int_{-\infty}^{+\infty} y G(dy)$ for $m \in \mathfrak{M}(F, G)$. This being the case, we shall have occasion to invoke the normalization

(A3) $\int_{-\infty}^{+\infty} h(x) F(dx) = \int_{-\infty}^{+\infty} y G(dy)$

and, equivalently, for $l = h \circ F^{-1}$

(A3)′ $\int_0^1 l(u)\, du = \int_0^1 G^{-1}(u)\, du$.

We now investigate the projection operator, isolating the main aspects of the argument in two lemmas. Some notation will prove to be convenient: let $I(x) = \int_0^x G^{-1}(u)\, du$, and let capitalization generally indicate integration, e.g., $L(x) = \int_0^x l(u)\, du$. If $A(x) \in C[0, 1]$, then denote by $A^*(x)$ the convex minorant of A (i.e., the greatest convex function less than or equal to A).

LEMMA. *Let $l \in L_2[[0, 1]; \mu]$ be nondecreasing* (a.e.) *and satisfy* (A3)′. *The projection $\tilde{l}$ of l onto $\mathfrak{M}_0$ satisfies*

$$\tilde{L}(x) = \int_0^x \tilde{l}(u)\, du = L(x) - (L - I)^*(x).$$

PROOF. The proof will be given first for step functions and then extended.

(I) For a fixed integer $N \geq 1$, suppose that l is of the form

$$l(u) = \sum_{j=0}^{N-1} l_j I_{[x_j, x_{j+1}]}(u), \qquad x_j = \frac{j}{N}, l_j \leq l_{j+1}.$$

We argue first that it is enough to restrict attention to candidates for projection which are similarly nondecreasing step functions: given $n \in \mathfrak{M}_0$, we apply the Cauchy-Schwarz inequality to get

$$\int_0^1[l(u) - n(u)]^2\, du = \sum_{j=0}^{N-1}\int_{x_j}^{x_{j+1}}[l_j - n(u)]^2\, du \geq \sum_{j=0}^{N-1}\frac{1}{N}(l_j - n_j)^2$$

where $n_j = N\int_{x_j}^{x_{j+1}} n(u)\,du$. The lower bound is attained for $n(u)$ identically constant on subintervals. Moreover, it can be further reduced by rearranging the n_j to be nondecreasing ([4], Theorem 368). If $n_j^{(T)}$ are the rearranged values, then we have

$$\int_0^1[l(u) - n(u)]^2\,du \geqslant \int_0^1[l(u) - n^{(T)}(u)]^2\,du$$

where $n^{(T)}(u) = \sum_{j=0}^{N-1} n_j^{(T)} I_{[x_j, x_{j+1}]}(u)$. We now show that $n^{(T)}(u) \in \mathfrak{M}_0$. Since $n^{(T)}(u)$ is nondecreasing (a.e.), by the remark after Theorem 1, it is enough to show that $N^{(T)}(x) = \int_0^x n^{(T)}(u)\,du \geqslant I(x)$ with equality at $x = 1$. The latter condition follows from the normalization (A3)$'$. Since $I(x)$ is convex and $N^{(T)}(x)$ is piece-wise linear, it is enough to verify the inequality constraints at the nodes $\{x_j\}$. We have $N^{(T)}(x_k) = \int_0^{x_k} n^{(T)}(u)\,du = (1/N)\sum_{j=0}^{k-1} n_j^{(T)}$, which is the integral of $n(u)$ over k of the subintervals. Equivalently, it is equal to $\int_0^{x_k} n(T(u))\,du$ for some T which appropriately permutes the subintervals. By (ii) of the corollary, this is bounded from below by $I(x_k)$.

We now have a discrete problem to solve:

$$\text{minimize } \sum_{j=0}^{N-1}(l_j - n_j)^2$$

subject to (a) the n_j are nondecreasing, and

(b) $\sum_{j=0}^{k-1} n_j \geqslant I(x_k)$ $k = 1, \cdots, N-1$ with equality at $k = N$.

With only constraint (b), the problem is treated in [1], pages 46–51, as a generalized isotonic regression. Letting L and $\tilde{L}$ denote the partial sum vectors of l and the solution vector $\tilde{l}$ respectively and setting $I = (0, I(x_1), I(x_2), \cdots, I(x_N))$, we have

$$\tilde{L} = L - (L - I)^*$$

where * here denotes the convex minorant of a vector. A straightforward argument shows that $\Delta_k^2(L - I)^* \leqslant \Delta_k^2(L - I)$ (Δ_k^2 denoting a second difference). Hence

$$\Delta_k^2\tilde{L} = \Delta_k^2[L - (L - I)^*] = \Delta_k^2 L - \Delta_k^2(L - I)^* \geqslant \Delta_k^2 I \geqslant 0.$$

It follows that $\tilde{L}$ is convex and that $\tilde{l}$ is nondecreasing. Thus (a) is satisfied automatically.

Translating the solution of the discrete problem into step function terms, we get $\tilde{L}(x) = L(x) - (L - I)^*(x)$.

(II) If $l(u)$ is not a step function, then for each $N \geqslant 1$, approximate $l(u)$ by

$$l_N(u) = \sum_{j=0}^{N-1}\left[N\int_{x_j}^{x_{j+1}} l(u)\,du\right] I_{[x_j, x_{j+1}]}(u).$$

By (I), we have

$$\tilde{L}_N(x) = L_N(x) - (L_N - I)^*(x). \tag{1}$$

Now as $N \to \infty$, $l_N \to l$ and $\tilde{l}_N \to \tilde{l}$ in $L_2[[0, 1]; \mu]$. Since $(L(x) - L_N(x))^2 \leqslant x\int_0^x (l_N(u) - l(u))^2\,du \leqslant \int_0^1[l_N(u) - l(u)]^2\,du$, we conclude that $L_N(x) \to L(x)$. Similarly, $\tilde{L}_N(x) \to \tilde{L}(x)$. Further, since $L_N \to L$ *uniformly* and * operates continuously in the uniform norm, $(L_N - 1)^* \to (L - I)^*$. Taking limits ($N \to \infty$) in (1) yields the lemma.

If l is not monotone, then some additional preparation is required to obtain its projection on $\mathfrak{M}_0$. For $l \in L_2[[0, 1]; \mu]$, define $l_\uparrow \in L_2[[0, 1]; \mu]$ to be the increasing rearrangement of l. There exists a measure-preserving transformation $S : [0, 1] \to [0, 1]$, *not necessarily one-one*, such that $l = l_\uparrow \circ S$ ([8]).

LEMMA. *Let $l \in L_2[[0, 1]; \mu]$ and satisfy* (A3)′. *Then, if $\tilde{l}$ and $\tilde{l}_\uparrow$ are the projections of l and $l_\uparrow$ respectively onto $\mathfrak{M}_0$,*

$$\tilde{l} = \tilde{l}_\uparrow \circ S.$$

REMARK. The construction for $\tilde{l}_\uparrow$ has been given in the previous lemma.

PROOF. We shall make use of a result of Brown [2], page 23, to the effect that S is the limit of invertible measure preserving maps S_n in the weak operator topology. That is, for any $g \in L_2[[0, 1]; \mu]$,

$$g \circ S_n \to g \circ S.$$

Accordingly, we have

$$l = l_\uparrow \circ S = \lim l_\uparrow \circ S_n,$$

and, by the continuity of the projection operator,

$$\tilde{l} = \lim \widetilde{l_\uparrow \circ S_n}.$$

A simple argument now shows that $\widetilde{l_\uparrow \circ S_n} = \tilde{l}_\uparrow \circ S_n$: we have (by a change of variables)

$$\inf_{m_0 \in \mathfrak{M}_0} \int_0^1 \left[(l_\uparrow \circ S_n)(u) - m_0(u) \right]^2 du$$

$$= \inf_{m_0 \in \mathfrak{M}_0} \int \left[l_\uparrow(u) - (m_0 \circ S_n^{-1})(u) \right]^2 du.$$

The unique m_0 for which the infima are attained satisfies $m_0 = \widetilde{l_\uparrow \circ S_n}$ (from the left hand side) and $m_0 = \tilde{l}_\uparrow \circ S_n$ (from the right hand side).

We conclude that

$$\tilde{l} = \lim \tilde{l}_\uparrow \circ S_n = \tilde{l}_\uparrow \circ S.$$

We can now state our result.

THEOREM 2. *Let $h \in L_2[(-\infty, +\infty); F]$ and satisfy* (A3). *Let $(h \circ F^{-1})_\uparrow$ be the increasing rearrangement of $h \circ F^{-1}$ with $h \circ F^{-1} = (h \circ F^{-1})_\uparrow \circ S$. Then the projection $\tilde{h}$ of h onto $\mathfrak{M}(F, G)$ is given by*

$$\tilde{h} = \widetilde{(h \circ F^{-1})}_\uparrow \circ S \circ F$$

where $\widetilde{(h \circ F^{-1})}_\uparrow$ satisfies

$$\int_0^x \widetilde{(h \circ F^{-1})}_\uparrow(u)\, du = J_1(x) - J_2^*(x)$$

and $J_1(x) = \int_0^x (h \circ F^{-1})_\uparrow(u)\, du$, $J_2(x) = J_1(x) - \int_0^x G^{-1}(u)\, du$.

PROOF. Together with the indicated isometry between $L_2[[0, 1]; \mu]$ and $L_2[(-\infty, +\infty); F]$, the statement combines the two lemmas.

4. Remarks. Despite the rather formidable analytical representation of the projection operator induced by $\mathfrak{M}(F, G)$, computational techniques have proved to be accessible. In particular, discretized versions of * and $\uparrow$, together with the extraction of the measure-preserving transformation S, are reasonably straightforward (see [1] for descriptions of some relevant algorithms).

REFERENCES

[1] Barlow, R. E., Bartholomew, D. J., Bremner, J. M. and Brunk, H. D. (1972). *Statistical Inference Under Order Restrictions*. Wiley, New York.

[2] Brown, J. R. (1966). Approximation theorems for Markov operators. *Pacific J. Math.* **16** 13–23.

[3] Grenander, U., McClure, D. E. and Vitale, R. A. (1969). Prediction. *DAM and CCIS Report*. Brown University.

[4] Hardy, G. H., Littlewood, J. E. and Pólya, G. (1967). *Inequalities, 2nd ed.* Cambridge Univ. Press.

[5] Jaglom, A. M. (1971). Examples of optimal nonlinear extrapolation of stationary random processes. *Selected Transl. Math. Statist. Probability* **9** 273–298.

[6] Masani, P. and Wiener, N. (1959). Non-linear prediction. In *Probability and Statistics*. (U. Grenander ed.), Wiley, New York 190–212.

[7] Ryff, J. V. (1965). Orbits of L^1-functions under doubly stochastic transformations. *Trans. Amer. Math. Soc.* **117** 92–100.

[8] Ryff, J. V. (1970). Measure preserving transformations and rearrangements. *J. Math. Anal. Appl.* **31** 449–458.

[9] Strassen, V. (1964). The existence of probability measures with given marginals. *Ann. Math. Statist.* **36** 423–439.

[10] Vitale, R. A. and Pipkin, A. C. (1976). Conditions on the regression function when both variables are uniformly distributed. *Ann. Probability* **4** 869–873.

Department of Mathematics
Claremont Graduate School
Claremont, California 91711