The Annals of Statistics
1979, Vol. 7, No. 5, 1121–1126

MONOTONE REGRESSION AND COVARIANCE STRUCTURE

BY GERALD SHEA

The University of Texas at Austin

The monotone regression of a variable X on another variable Y is of particular interest when Y cannot be directly observed. The correlation of X and Y can be tested if at least high and low values of Y can be recognized. If all the components of a random vector have monotone regression on a variable Y, and if they are all uncorrelated given Y, then an inequality due to Chebyshev shows that marginal zero covariances imply that all but at most one of the components are uncorrelated with Y. Cases are examined where marginal uncorrelatedness of attributes implies their independence. Applications to contaminated experiments and to discriminant analysis are noted.

1. Introduction. Lehmann (1966) considered several progressively stronger forms of bivariate dependence. The most general forms were "positive quadrant dependence" where random variables X and Y have $P[X \leqslant x, Y \leqslant y] \geqslant P[X \leqslant x]P[Y \leqslant y]$ for all x, y and "negative quadrant dependence" where the inequality between the probabilities is reversed. Such cases are implied by "regression dependence", say, of X on Y, where $P[X \leqslant x | Y = y]$ is monotone in y. Regression dependence also implies an intuitively appealing relationship between X and Y, namely that the expected value of X given $Y = y$ is monotone in y; we say that X has monotone regression on Y. Monotone regression is an intuitively attractive model for the relationship between an unobservable condition Y and a sign X that reflects changes in Y. This paper generalizes results presented in [8] where the symptoms X_i are regression dependent upon the stage Y of a disease. As noted there, the assumptions that the X_i are conditionally uncorrelated or independent given Y, investigated here in Section 3, are motivated by certain discrimination procedures in statistical diagnosis.

2. Monotone regression. Let $F(x|y) = P[X \leqslant x | Y = y]$, the conditional distribution function of X given $Y = y$. Suppose that X is regression dependent on Y, that is, $F(x|y)$ is monotone in y in the same direction for all x. Lehmann (1966) notes that X and Y are then quadrant dependent; yet at the same time, the identity

$$E[X|Y = y] = \int_0^\infty \{1 - F(x|y)\}\, dx - \int_{-\infty}^0 F(x|y)\, dx$$

shows that $E[X|Y = y]$ is monotone in y.

DEFINITION 2.1. Let G denote the marginal distribution function of Y, and let S denote the support of G. Suppose for all y in S that $E[X|Y = y]$ exists. If $E[X|Y = y]$ is nondecreasing (nonincreasing) for all y in S, we say that the

Received February 1978; revised June 1978.
AMS 1970 *subject classifications.* Primary 62H20; secondary 62H05.
Key words and phrases. Covariance, independence, monotone regression, quadrant dependence, statistical diagnosis.

regression of X on Y is nondecreasing (nonincreasing). If either case holds, X is said to have monotone regression on Y.

REMARK 2.1. As shown in the appendix, monotone regression neither implies nor is implied by quadrant dependence. Also, monotone regression and quadrant dependence together do not imply regression dependence.

LEMMA 2.1. *Suppose that EXY, EX and EY exist. If the regression of X on Y is nondecreasing, $EXY \geqslant (EX)(EY)$; the inequality is reversed if the regression is nonincreasing. Equality holds if and only if $E[X|Y = y] = EX$ for all y in S.*

PROOF.

$$EXY = \int_S yE[X|Y = y]\, dF(y)$$

and

$$(EX)(EY) = \int_S E[X|Y = y]\, dF(y) \int_S y\, dF(y).$$

Therefore the first part is immediate from the Chebyshev inequalities (a) and (b) in the appendix. If S is a singleton set, the equality is trivial; otherwise, the equality follows from part (c) of the Chebyshev lemma given in the Appendix since y is strictly increasing in S.

REMARK 2.2. In certain applications, Y may not be directly observable although low and high cases may be recognizable; for example Y may be the stage of a chronic disease. If X is to be used to discern Y, it may be supposed that the regression of X on Y is monotone; the following theorem and remark allow one to determine whether or not a potential sign X actually presents information about Y.

THEOREM 2.1. *Suppose that the regression of X on Y is monotone and that EXY, EX and EY exist. Then X and Y are uncorrelated if and only if there exists some y_1, y_2 in S, $y_1 \leqslant y_2$ and $0 < G(y_1), G(y_2) < 1$ such that*

$$E[X|Y \leqslant y_1] = E[X|Y > y_2]. \tag{2.1}$$

PROOF. Equality (2.1) implies that

$$(1/P[Y \leqslant y_1])\int_{(-\infty, y_1]} E[X|Y = y]\, dG(y) = (1/P[Y > y_2])\int_{(y_2, \infty)} E[X|Y = y]\, dG(y).$$

Since these are weighted averages of the monotone regression of X on Y, the equality holds if and only if $E[X|Y = y]$ is constant for all y in S. Lemma 2.1 then completes the proof.

REMARK 2.3. Lehmann (1966) has shown that if X and Y are quadrant dependent, then X and Y are independent if and only if they are uncorrelated. The theorem demonstrates that if the physical relationship between X and Y can be expressed by monotone regression and quadrant dependence, then equal means for extreme cases of Y indicates that X offers no information about Y.

3. Multivariate results. Sometimes a number of variates of attributes are observed in order to learn about an underlying process that is itself unobservable. For example $X_1, X_2, \cdots, X_K$ may be K distinct signs of a progressive structural weakness or biological disease denoted by Y. Such signs are of value because they simultaneously reflect the progress of Y; we assume that $X_1, \cdots, X_K$ all have monotone regression on Y. Further if there are no spurious correlations between pairs (X_i, X_j) other than common dependence on Y, we assume that (X_i, X_j) are conditionally uncorrelated given Y.

In some applications it is possible to separate cases where $Y = 0$ from those with $Y > 0$ although it is not possible to identify positive Y values. For instance, if Y is the stage of a disease, $Y = 0$ corresponds to a state of health and $Y > 0$ to a diseased state. Therefore a necessary condition for conditional uncorrelatedness is that X_i and X_j be uncorrelated given the healthy state.

THEOREM 3.1. *If $X_1, \cdots, X_K$ all have monotone regression on Y, if the conditional and marginal expectations of all products X_iX_j exist, and if all X_i, X_j are pairwise uncorrelated given Y, then $X_1, \cdots, X_K$ are uncorrelated if and only if at most one X_i is correlated with Y.*

PROOF. Chebyshev's lemma and Lemma 2.1 show that if X_i and X_j are both correlated with Y, then their marginal covariances are nonzero.

REMARK 3.1. Jogdeo (1968) showed that if X_1, X_2, X_3 have

$$P[X_1 \leqslant x_1, X_2 \leqslant x_2, X_3 \leqslant x_3] \geqslant \Pi_{i=1}^3 P[X_i \leqslant x_i], \tag{3.1}$$

$$EX_iX_j = (EX_i)(EX_j) \qquad \text{for all} \quad i \neq j, \tag{3.2}$$

and

$$E[X_iX_j|X_k] = E[X_i|X_k]E[X_j|X_k] \tag{3.3}$$

for some k, $1 \leqslant k \leqslant 3$, and $i, j \neq k$, $i \neq j$, then X_1, X_2 and X_3 are independent. In some cases, there may be no apparent reason for assuming (3.3) although each X_i has monotone regression on an underlying condition Y. By Theorem 3.1, (3.2) shows that at most one, say $X_k(Y)$, is correlated with Y. If *a priori* Y is the only physical link among X_1, X_2, X_3, then

$$E[X_iX_j|X_k(Y)] = E[X_i|X_k(Y)]E[X_j|X_k(Y)] \tag{3.3'}$$

may be assumed. The following theorem gives another set of conditions where uncorrelatedness implies independence.

THEOREM 3.2. *Suppose that $X_1, \cdots, X_K$ are all conditionally independent given Y, and each has monotone regression and is quadrant dependent on Y. Then $X_1, \cdots, X_K$ are independent if they are uncorrelated.*

PROOF. By Theorem 3.1, at most one X_i, say X_1, is correlated with Y. Therefore quadrant dependence implies that $X_2, X_3, \cdots, X_K$ are all independent of Y.

Finally the distribution function of $X_1, \cdots, X_K$

$$\begin{aligned} F(x_1, \cdots, x_K) &= E[F(x_1, \cdots, x_K | Y)] \\ &= E[\Pi_{i=1}^K F(x_i | Y)] \\ &= \Pi_{i=2}^K F(x_i) E[F(x_1 | Y)] \\ &= \Pi_{i=1}^K F(x_i). \end{aligned}$$

REMARK 3.2. Jogdeo (1977) considered an experiment where independent responses were contaminated by an outside variate. If each observation has monotone regression and is quadrant dependent on the possible contaminator, then Theorem 3.2 shows that if the observations are uncorrelated, they are uncontaminated.

REMARK 3.3. Discrimination procedures based on an assumption of independent attributes or clusters of attributes within each population have been reported to be relatively successful [2, 7]. This may be due to the use of marginal distributions that fit the data well [1] or to the variability of multivariate estimates drawn from small samples [8]. In fact, if the variates are independent, the product of marginal estimates will have a lower convex loss than a multivariate estimate. In cases where the procedure attempts to discriminate among intervals on a continuum, such a high versus low risk, the concepts of monotone regression, quadrant dependence, and conditional independence may be applicable. In such cases, Theorem 3.2 allows correlation to be used to determine independence as in [7].

APPENDIX

A1. Quadrant dependence does not imply monotone regression. In Table 1, $P[X \leqslant x, Y \leqslant y] \geqslant P[X \leqslant x]P[Y \leqslant y]$ but $E[X | Y = y]$ is not monotone in y.

TABLE 1

Positively quadrant dependent joint probabilities.

X	Y = 1	Y = 2	Y = 3
1	.10	.02	.30
0	.30	.08	.20

A2. Monotone regression does not imply quadrant dependence. In Table 2, $E[X | Y = y]$ is trivially monotone, but $P[X \leqslant 0, Y \leqslant 1] < P[X \leqslant 0]P[Y \leqslant 1]$ while $P[X \leqslant 1, Y \leqslant 1] > P[X \leqslant 0]P[Y \leqslant 1]$.

TABLE 2

Joint probabilities with monotone regression of X on Y.

X	Y = 1	Y = 2
2	.02	.06
1	.45	.40
0	.03	.04

A3. Monotone regression and quadrant dependence do not imply regression dependence. In Table 3, $P[X \leqslant x, Y \leqslant y] \geqslant P[X \leqslant x]P[Y \leqslant y]$ and $E[X|Y = y]$ is trivially monotone but $P[X \leqslant 0|Y = y]$ is strictly increasing in y while $P[X \leqslant 1|Y = y]$ is strictly decreasing in y.

TABLE 3
Positively quadrant dependent joint probabilities with monotone regression of X on Y.

X	Y: 1	2
2	0	.5
1	.2	0
0	.1	.2

A4. A version of an inequality due to Chevyshev: suppose that u and v are monotone real-valued functions defined on a totally ordered set R and that u, v, and uv are integrable with respect to some positive measure μ such that $\mu(R) < \infty$.

(a) If u and v are both nonincreasing or both nondecreasing in R, then

$$\mu(R)\int_R uv\, d\mu \geqslant \int_R u\, d\mu \int_R v\, d\mu. \tag{A.1}$$

(b) If one of u and v is nonincreasing and the other is nondecreasing, then the inequality in (A.1) is reversed.

(c) Equality holds in (A.1) if and only if u or v is constant a.e. (μ). The proof is based on the identity due to Franklin (1885):

$$\int_R\left[\int_R\{[u(x) - u(y)][v(x) - v(y)]\}\,\mu(dx)\right]\mu(dy) = 2\left[\mu(R)\int_R uv\, d\mu - \int_R u\, d\mu \int_R v\, d\mu\right].$$

Acknowledgment. The author wishes to thank the referee for several improvements in this article, in particular for counterexample A.3.

REFERENCES

[1] Bishop, C. R. and Warner, H. R. (1969). A mathematical approach to medical diagnosis: application to polycythemic states utilizing clinical findings with values continuously distributed. *Comp. and Biomed. Res.* **2** 486–493.

[2] Croft, D. J. and Machol, R. E. (1974). Mathematical methods in medical diagnosis. *Ann. Biomed. Eng.* **2** 69–89.

[3] Franklin, F. (1885). Proof of a theorem of Tschebyscheff's on definite integrals. *Amer. J. Math.* **7** 377–379.

[4] Jogdeo, K. (1968). Characterizations of independence in certain families of bivariate and multivariate distributions. *Ann. Math. Statist.* **39** 443–441.

[5] Jogdeo, K. (1977). Association and probability inequalities. *Ann. Statist.* **5** 495–504.

[6] Lehmann, E. L. (1966). Some concepts of dependence. *Ann. Math. Statist.* **37** 1137–53.

[7] NORUSIS, M. J. and JACQUEZ, J. A. (1975). Diagnosis II. Diagnostic models based on attribute clusters: a proposal and comparisons. *Comp. and Biomed. Res.* **8** 173–188.
[8] SHEA, G. (1978). An analysis of the Bayes' procedure for diagnosing multistage diseases. *Comp. and Biomed. Res.* **11** 65–75.

DEPARTMENT OF MATHEMATICS
THE UNIVERSITY OF TEXAS AT AUSTIN
AUSTIN, TEXAS 78712