Some counterexamples concerning maximal correlation and linear regression

Nickos Papadatos

Section of Statistics and O.R., Department of Mathematics, University of Athens, Panepistemiopolis, 157 84 Athens, Greece

Abstract

A class of examples concerning the relationship of linear regression and maximal correlation is provided. More precisely, these examples show that if two random variables have (strictly) linear regression on each other, then their maximal correlation is not necessarily equal to their (absolute) correlation.

Key words and phrases: Maximal Correlation Coefficient; Linear Regression; Sarmanov Theorem.

1 Maximal correlation and linear regression

Let \((X,Y)\) be a bivariate random vector such that its Pearson correlation coefficient,

\[
\rho(X,Y) := \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X)} \sqrt{\text{Var}(Y)}},
\]

is well defined. If \(W\) is a non-degenerate random variable then \(L^*_2(W)\) is defined to be the class of measurable functions \(g : \mathbb{R} \to \mathbb{R}\) such that \(0 < \text{Var}[g(W)] < \infty\). Under the present notation, the maximal correlation coefficient is defined as (Gebelein, 1941; Hirschfeld, 1935)

\[
R(X,Y) := \sup_{g_1 \in L^*_2(X), g_2 \in L^*_2(Y)} \rho(g_1(X), g_2(Y)).
\]

Due to results of Sarmanov (1958a, 1958b), it was believed for some time that if both \(X\) and \(Y\) have linear regression on each other, i.e., if for some constants \(a_0, a_1, b_0, b_1\),

\[
\mathbb{E}(X|Y) = a_1 Y + a_0 \text{ (a.s.)}, \quad \mathbb{E}(Y|X) = b_1 X + b_0 \text{ (a.s.)},
\]

then

\[
R(X,Y) = |\rho(X,Y)|.
\]
The implication (3) ⇒ (4) was cited in a number of subsequent works related to maximal correlation of order statistics and records, including Rohatgi and Székely (1992), Arnold, Balakrishnan and Nagaraja (1998, p. 101), Székely and Gupta (1998), David and Nagaraja (2003, p. 74), Ahsanullah (2004, p. 23) and Barakat (2012). However, as we shall show below, this implication is not valid even in the case of a strictly linear regression, \( a_1 b_1 \neq 0 \). Note that if \( R(X, Y) > 0 \) then the converse implication, (4) ⇒ (3), is valid; see Rényi (1959, p. 447) and Dembo, Kagan and Shepp (2001).

Examples of uncorrelated random variables \( X, Y \) with (trivial) linear regression

\[
E(X|Y) = E(Y|X) = 0 \quad \text{(a.s.)}
\]

and \( R(X, Y) > 0 = |\rho(X, Y)| \) are known for a long time. For instance, P. Bártafai has calculated \( R(X, Y) = 1/3 \) for a uniform in the interior of the unit disc. This result was extended by P. Csáki and J. Fischer for the uniform distribution in the domain \(|x|^p + |y|^p < 1 \) (\( p > 0 \)), in which case \( R(X, Y) = (p + 1)^{-1} \); see Rényi (1959, p. 447) and Csáki and Fischer (1963). Furthermore, Székely and Móri (1985) extended this result to the multivariate case and with different exponents. Moreover, in response to a question asked by Sid Browne of Columbia University, Dembo, Kagan and Shepp (2001) constructed a pair \((X, Y)\) satisfying (5) and \( R(X, Y) = 1 \). (Observe that the same is true for the uniform distribution in the four-point domain \( \{(0, \pm 1), (\pm 1, 0)\} \).

Using characterizations of Vershik (1964) and Eaton (1986), they also showed that for any non-Gaussian spherically symmetric random vector \((U_1, \ldots, U_k)\), with covariance matrix of rank \( \geq 2 \), there exists a pair of uncorrelated linear forms,

\[
X = a_1 U_1 + \cdots + a_k U_k, \quad Y = b_1 U_1 + \cdots + b_k U_k,
\]

such that (5) is fulfilled and \( R(X, Y) > |\rho(X, Y)| > 0 \).

However, in the author’s opinion, it is important to definitely know that (3) does not imply (4) even in the non-trivial linear regression case. Indeed, if this implication were valid in the particular case where \( a_1 b_1 \neq 0 \), then several works concerning characterizations of distributions through maximal correlation of order statistics and records – including the papers by Terrell (1983), Székely and Móri (1985), Nevzorov (1992), López-Blázquez and Castaño-Martínez (2006), Castaño-Martínez, López-Blázquez and Salamanca-Miño (2007), Papadatos and Xifara (2012) – would be reduced to trivial consequences of this implication. The same is true for the main result in Dembo, Kagan and Shepp (2001), since it is easily checked that for the partial sums \( S_k = X_1 + \cdots + X_k \), based on an iid sequence with mean \( \mu \) and finite non-zero variance,

\[
E(S_{n+m}|S_n) = S_n + m\mu \quad \text{(a.s.)}, \quad E(S_n|S_{n+m}) = \frac{n}{n+m}S_{n+m} \quad \text{(a.s.)}.
\]

The purpose of the present note is to present examples of random vectors \((X, Y)\), with \( X \) and \( Y \) possessing strictly linear regression on each other, and such that \( R(X, Y) > |\rho(X, Y)| > 0 \). The proposed examples, contained in the next section, are as elementary as possible.


2 Counterexamples

Normal marginals. Consider the standardized Hermite polynomials

\[ h_n(t) := \frac{(-1)^n}{\sqrt{n!}} e^{t^2/2} \frac{d^n}{dt^n}(e^{-t^2/2}), \quad n = 0, 1, \ldots, \]

which form an orthonormal system with respect to the standard normal density, \( \phi(t) := e^{-t^2/2}/\sqrt{2\pi} \). That is, if \( Z \sim N(0, 1) \) then

\[ \mathbb{E}[h_n(Z)h_m(Z)] = \int_{-\infty}^{\infty} h_n(t)h_m(t)\phi(t)dt = \delta_{n,m}, \]

where \( \delta_{n,m} \) is Kronecker’s delta. The first three standardized Hermite polynomials are given by

\[ h_0(t) = 1, \quad h_1(t) = t, \quad h_2(t) = \frac{t^2 - 1}{\sqrt{2}}. \]

Moreover, the following properties of \( h_n \) are well known – see, e.g., Chihara (1978):

\[ h'_{n+1}(t) = \sqrt{n+1}h_n(t), \quad h_{n+2}(t) = \frac{t}{\sqrt{n+2}}h_{n+1}(t) - \frac{n+1}{n+2}h_n(t), \quad n = 0, 1, \ldots. \] (6)

Using (6) and induction on \( n \), it can be shown that for all \( n = 0, 1, \ldots, \)

\[ \int_{-\infty}^{\infty} \phi(t)h_n(\rho z + \sqrt{1-\rho^2}t)dt = \rho^n h_n(z). \] (7)

Let \( f_\rho \) be the bivariate standard normal density with correlation coefficient \( \rho \in (-1, 1) \), that is,

\[ f_\rho(x, y) = \frac{1}{2\pi \sqrt{1-\rho^2}} \exp \left( -\frac{1}{2(1-\rho^2)} (x^2 - 2\rho xy + y^2) \right), \quad (x, y) \in \mathbb{R}^2. \]

Fix \( p \in (0, 1) \), \( \alpha, \beta \in (-1, 1) \), and define the (symmetric) mixture density

\[ f(x, y) := (1-p)f_\alpha(x, y) + pf_\beta(x, y), \quad (x, y) \in \mathbb{R}^2. \] (8)

Since both \( f_\alpha \) and \( f_\beta \) have standard normal marginals, the same is true for \( f \). If a random pair \((X,Y)\) has density \( f \), then a version of the conditional density of \( X \), given \( Y \), is easily found to be

\[ f_{X|Y}(x|y) = \frac{f(x, y)}{\phi(y)} = \frac{1-p}{\sqrt{1-\alpha^2}} \phi \left( \frac{x - \alpha y}{\sqrt{1-\alpha^2}} \right) + \frac{p}{\sqrt{1-\beta^2}} \phi \left( \frac{x - \beta y}{\sqrt{1-\beta^2}} \right). \]

Hence, by virtue of (7),

\[ \int_{-\infty}^{\infty} h_n(x)f_{X|Y}(x|y)dx = [(1-p)\alpha^n + p\beta^n]h_n(y), \]
and, similarly,
\[ \int_{-\infty}^{\infty} h_n(y) f_{Y|X}(y|x) \, dy = [(1-p)\alpha^n + p\beta^n]h_n(x). \]

It follows that, almost surely,
\[ \mathbb{E}[h_n(X)|Y] = [(1-p)\alpha^n + p\beta^n]h_n(Y), \quad \mathbb{E}[h_n(Y)|X] = [(1-p)\alpha^n + p\beta^n]h_n(X). \quad (9) \]

Clearly, if \((1-p)\alpha + p\beta \neq 0\), \((9)\) with \(n=1\) shows that \(X\) and \(Y\) have strictly linear regression on each other. Moreover, for all \(n \geq 1\) we have
\[
\rho(h_n(X), h_n(Y)) = \mathbb{E}[h_n(X)h_n(Y)] = \mathbb{E}[h_n(Y)\mathbb{E}(h_n(X)|Y)] = (1-p)\alpha^n + p\beta^n.
\]

Hence, if the parameter vector \((p, \alpha, \beta) \in (0,1) \times (-1,1)^2\) satisfies \(0 < |(1-p)\alpha + p\beta| < \sup_{n\geq2} |(1-p)\alpha^n + p\beta^n|\) then
\[
R(X, Y) \geq \sup_{n\geq1} |(1-p)\alpha^n + p\beta^n| > |(1-p)\alpha + p\beta| = |\rho(X, Y)| > 0.
\]

For example, the particular choice \((p, \alpha, \beta) = (1/2, -1/4, 3/4)\) leads to \(\mathbb{E}(X|Y) = Y/4\) (a.s.), \(\mathbb{E}(Y|X) = X/4\) (a.s.), \(\rho(X, Y) = 1/4\) and
\[
R(X, Y) \geq \rho(X^2, Y^2) = \frac{5}{16}.
\]

**Arbitrary marginals with bounded supports.** Let \(f_1\) and \(f_2\) be two univariate probability densities (with respect to Lebesgue measure on \(\mathbb{R}\)) with bounded supports, \(\text{supp}(f_i) \subseteq [\alpha_i, \omega_i], -\infty < \alpha_i < \omega_i < \infty \quad (i = 1, 2)\). Since \(f_1\) has finite moments of any order, it is well known that there exists an orthonormal polynomial system \(\{\phi_n(x)\}_{n=0}^{\infty}\), corresponding to \(f_1\). That is,
\[
\int_{-\infty}^{\infty} \phi_n(x)\phi_m(x)f_1(x) \, dx = \delta_{n,m},
\]
where \(\delta_{n,m}\) is Kronecker’s delta. Clearly, the support of \(f_1\) contains infinitely many points and, therefore, each \(\phi_n\) is of degree \(n\). By the same reasoning, there exists an orthonormal polynomial system \(\{\psi_n(y)\}_{n=0}^{\infty}\), corresponding to \(f_2\). Since every polynomial is uniformly bounded in any finite interval, we can find constants \(c_n, d_n\) such that
\[
1 < \sup_{\alpha_1 \leq x \leq \omega_1} |\phi_n(x)| = c_n < \infty, \quad 1 < \sup_{\alpha_2 \leq y \leq \omega_2} |\psi_n(y)| = d_n < \infty, \quad n = 1, 2, \ldots.
\]

Consider an arbitrary real sequence \(\{\rho_n\}_{n=1}^{\infty}\) such that
\[
\sum_{n=1}^{\infty} |\rho_n|c_n d_n \leq 1, \quad (10)
\]
e.g., \( \rho_n = 6(\pi^2 n^2 c_n d_n)^{-1} \) \((n = 1, 2, \ldots)\) or \( \rho_n = \lambda n \) \((n = 1, \ldots, N)\) and \( \rho_n = 0 \), otherwise, where \(0 < \lambda \leq (\sum_{n=1}^{N} n c_n d_n)^{-1}\). Then, the function

\[
f(x, y) := f_1(x)f_2(y) \left( 1 + \sum_{n=1}^{\infty} \rho_n \phi_n(x)\psi_n(y) \right), \quad (x, y) \in [\alpha_1, \omega_1] \times [\alpha_2, \omega_2], \tag{11}
\]

and \( f := 0 \) outside \([\alpha_1, \omega_1] \times [\alpha_2, \omega_2]\), is a bivariate probability density with marginal densities \( f_1, f_2 \); this is so because, due to (10), the series in (11) converges, for each \((x, y)\) in the domain of definition, to a value greater than or equal to \(-1\). (Actually, the series converges uniformly and absolutely in \([\alpha_1, \omega_1] \times [\alpha_2, \omega_2]\).) Therefore, \( f(x, y) \) is nonnegative. Next, it is easily checked that its integral over \(\mathbb{R}^2\) equals 1, due to the orthonormality of the polynomials. Finally, it is obvious that the marginal densities of \( f \) are \( f_1, f_2 \).

Assume now that the random vector \((X, Y)\) has density \( f \). Then \( X \) has density \( f_1 \) and \( Y \) has density \( f_2 \). Moreover, versions of the conditional densities are given by

\[
f_{X|Y}(x|y) = f_1(x) \left( 1 + \sum_{n=1}^{\infty} \rho_n \phi_n(x)\psi_n(y) \right), \quad \alpha_1 \leq x \leq \alpha_2 \quad \text{(for each } y \in \text{supp}(f_2)),
\]

\[
f_{Y|X}(y|x) = f_2(y) \left( 1 + \sum_{n=1}^{\infty} \rho_n \phi_n(x)\psi_n(y) \right), \quad \alpha_2 \leq y \leq \omega_2 \quad \text{(for each } x \in \text{supp}(f_1)).
\]

Due to the orthonormality of the polynomials it follows that for all \( n \geq 1 \),

\[
E(\phi_n(X)|Y) = \rho_n \psi_n(Y) \quad \text{(a.s.)}, \quad E(\psi_n(Y)|X) = \rho_n \phi_n(X) \quad \text{(a.s.).} \tag{12}
\]

Clearly, if \( \rho_1 \neq 0 \), (12) with \( n = 1 \) shows that \( X \) and \( Y \) have strictly linear regression on each other. From (12) we conclude that \( \rho(\phi_n(X), \psi_n(Y)) = \rho_n \) for all \( n \geq 1 \) and, therefore, \( \rho(X, Y) = \rho(\phi_1(X), \psi_1(Y)) = \rho_1 \) and \( R(X, Y) \geq \sup_{n \geq 1} |\rho_n| \). Since the choice of \( \{\rho_n\}_{n=1}^{\infty} \) is quite arbitrary (see (10)), it follows that

\[
R(X, Y) > |\rho(X, Y)| = |\rho_1| > 0 \quad \text{whenever } 0 < |\rho_1| < \sup_{n \geq 2} |\rho_n|.
\]

**Remark.** (a) It is obvious that the construction (11) can be adapted to the discrete (lattice) case where \((X, Y) \in \{1, \ldots, N\}^2\), covering the characterizations (for finite populations) treated by López-Blázquez and Castaño-Martínez (2006) and Castaño-Martínez, López-Blázquez and Salamanca-Miño (2007).

(b) Distributions with densities of the form (11) are known as Lancaster distributions; see, e.g., Koudou (1998) or Diaconis and Griffiths (2012). They can be viewed as extensions of the Sarmanov-type distribution \((\rho_n = 0 \text{ for } n \geq 2)\) which, assuming standard uniform marginals, generalizes the so called Farlie-Gumbel-Morgenstern family.

(c) It is of some interest to observe that the mixture density (8) admits a series representation of the form (11). Indeed, Sarmanov (1966) showed that the series

\[
g(x, y) := \phi(x)\phi(y) \left( 1 + \sum_{n=1}^{\infty} \rho_nh_n(x)h_n(y) \right), \quad (x, y) \in \mathbb{R}^2,
\]
represents a bivariate density if and only if \( \rho_n = \mathbb{E} U^n, \ n = 1, 2, \ldots, \) where \( U \) is a random variable with \( \mathbb{P}(|U| < 1) = 1. \) The special choice of a two-valued \( U \) with \( \mathbb{P}(U = \beta) = p = 1 - \mathbb{P}(U = \alpha) \) leads to \( \rho_n = \mathbb{E} U^n = (1 - p)\alpha^n + p\beta^n. \) Substituting these values of \( \rho_n \) in the series representation of \( g, \) above, and in view of Mehler’s identity (tetrachoric series) of the bivariate normal density,

\[
f_{\rho}(x, y) = \phi(x)\phi(y) \left( 1 + \sum_{n=1}^{\infty} \rho^n h_n(x)h_n(y) \right), \quad (x, y) \in \mathbb{R}^2, \ -1 < \rho < 1,
\]

we conclude that \( g = (1 - p)f_\alpha + pf_\beta, \) as in (8).

Acknowledgement. I would like to thank H.N. Nagaraja and H.M. Barakat for bringing to my attention the papers by Rohatgi and Székely (1992) and Sarmanov (1958a, 1958b).

REFERENCES


