Journal of Multivariate Analysis 126 (2014) 114–117

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

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

ARTICLE INFO

Article history:
Received 27 February 2013
Available online 2 January 2014

In memory of Dionyssia Papadatou (2004–2010)

AMS subject classification:
primary 62H20
secondary 60E15

Keywords:
Maximal correlation coefficient
Linear regression
Sarmanov theorem

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.

1. Maximal correlation and linear regression

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

$$\rho(X,Y) := \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}}, \tag{1}$$

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 $0 < \mathrm{Var}[g(W)] < \infty$. Under the present notation, the maximal correlation coefficient is defined as [10,11]

$$R(X,Y) := \sup_{g_1 \in L_2^*(X),\ g_2 \in L_2^*(Y)} \rho(g_1(X), g_2(Y)). \tag{2}$$

Due to results of Sarmanov [21,20], 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,

$$\mathbb{E}(X|Y) = a_1 Y + a_0 \quad \text{(a.s.)}, \qquad \mathbb{E}(Y|X) = b_1 X + b_0 \quad \text{(a.s.)}, \tag{3}$$

then

$$R(X,Y) = |\rho(X,Y)|. \tag{4}$$

☆ Work partially supported by the University of Athens Research Grant 70/4/5637.

E-mail address: npapadat@math.uoa.gr.
URL: http://users.uoa.gr/~npapadat/.

0047-259X/$ – see front matter © 2014 Published by Elsevier Inc.
http://dx.doi.org/10.1016/j.jmva.2013.12.008

The implication (3) $\Rightarrow$ (4) was cited in a number of subsequent works related to maximal correlation of order statistics and records, including Rohatgi and Székely [19], Arnold et al. [2, p. 101], Székely and Gupta [23], David and Nagaraja [6, p. 74], Ahsanullah [1, p. 23] and Barakat [3]. However, as we shall show below, this implication is not valid even in the case of a strictly linear regression, $a_1b_1 \neq 0$. Note that if $R(X,Y) > 0$ then the converse implication, (4) $\Rightarrow$ (3), is valid; see [18, p. 447] and [7].

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

$$\mathbb{E}(X|Y) = \mathbb{E}(Y|X) = 0 \quad \text{(a.s.)} \tag{5}$$

and $R(X,Y) > 0 = |\rho(X,Y)|$ are known for a long time. For instance, P. Bártfai 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 [18, p. 447] and [5]. Furthermore, Székely and Móri [24] 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 et al. [7] 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 $\{(0,\pm1), (\pm1,0)\}$.) Using characterizations of Vershik [26] and Eaton [9], they also showed that for any non-Gaussian spherically symmetric random vector $(U_1,\ldots,U_k)$, with covariance matrix of rank ≥ 2, there exists a pair of uncorrelated linear forms,

$$X = a_1U_1 + \cdots + a_kU_k, \qquad Y = b_1U_1 + \cdots + b_kU_k,$$

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 $a_1b_1 \neq 0$, then several works concerning characterizations of distributions through maximal correlation of order statistics and records – including the papers by Terrell [25], Székely and Móri [24], Nevzorov [16], López-Blázquez and Castaño-Martínez [15], Castaño-Martínez et al. [4], Papadatos and Xifara [17] – would be reduced to trivial consequences of this implication. The same is true for the main result in [7], since it is easily checked that for the partial sums $S_k = X_1 + \cdots + X_k$, based on an i.i.d. sequence with mean μ and finite non-zero variance,

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

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

$$f(x,y) := (1-p)f_\alpha(x,y) + pf_\beta(x,y), \quad (x,y) \in \mathbb{R}^2, \tag{6}$$

where f_ρ be the bivariate standard normal density with correlation coefficient $\rho \in (-1,1)$, that is,

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

Since both f_α and f_β have standard normal marginals, the same is true for f. A straightforward calculation shows that for a random pair (X,Y) with density f,

$$\mathbb{E}(X|Y) = [(1-p)\alpha + p\beta]Y, \qquad \mathbb{E}(Y|X) = [(1-p)\alpha + p\beta]X. \tag{7}$$

If $(1-p)\alpha + p\beta \neq 0$, (7) shows that X and Y have strictly linear regression on each other and, clearly,

$$\rho(X,Y) = \mathbb{E}(XY) = \mathbb{E}[Y\mathbb{E}(X|Y)] = (1-p)\alpha + p\beta.$$

Another simple calculation reveals that

$$\mathbb{E}(X^2Y^2) = \mathbb{E}[Y^2\mathbb{E}(X^2|Y)] = \mathbb{E}\{Y^2[1 + ((1-p)\alpha^2 + p\beta^2)(Y^2 - 1)]\}.$$

Using $\mathbb{E}(X^2) = \mathbb{E}(Y^2) = 1$, $\mathbb{E}(X^4) = \mathbb{E}(Y^4) = 3$, $\mathrm{Var}(X^2) = \mathrm{Var}(Y^2) = 2$, we get

$$\rho(X^2,Y^2) = (1-p)\alpha^2 + p\beta^2.$$

Hence, if the parameter vector $(p,\alpha,\beta) \in (0,1) \times (-1,1)^2$ satisfies $0 < |(1-p)\alpha + p\beta| < |(1-p)\alpha^2 + p\beta^2|$ then

$$R(X,Y) \geq |\rho(X^2,Y^2)| = |(1-p)\alpha^2 + p\beta^2| > |(1-p)\alpha + p\beta| = |\rho(X,Y)| > 0.$$

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 $R(X,Y) \geq \rho(X^2,Y^2) = 5/16$.

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, $\mathrm{supp}(f_i) \subseteq [\alpha_i, \omega_i]$, $-\infty < \alpha_i < \omega_i < \infty$ $(i = 1, 2)$. 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 ϕ_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

$$1 < \sup_{\alpha_1 \le x \le \omega_1} |\phi_n(x)| = c_n < \infty, \qquad 1 < \sup_{\alpha_2 \le y \le \omega_2} |\psi_n(y)| = d_n < \infty, \quad n = 1, 2, \ldots.$$

Consider an arbitrary real sequence $\{\rho_n\}_{n=1}^{\infty}$ such that

$$\sum_{n=1}^{\infty} |\rho_n| c_n d_n \le 1, \tag{8}$$

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 $0 < \lambda \le (\sum_{n=1}^{N} n c_n d_n)^{-1}$. Then, the function

$$f(x, y) := f_1(x)f_2(y)\left(1 + \sum_{n=1}^{\infty} \rho_n \phi_n(x)\psi_n(y)\right), \quad (x, y) \in [\alpha_1, \omega_1] \times [\alpha_2, \omega_2], \tag{9}$$

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 (8), the series in (9) 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

$$f_{X|Y}(x|y) = f_1(x)\left(1 + \sum_{n=1}^{\infty} \rho_n \phi_n(x)\psi_n(y)\right), \quad \alpha_1 \le x \le \omega_1 \text{ (for each } y \in \mathrm{supp}(f_2)),$$

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

Due to the orthonormality of the polynomials it follows that for all $n \ge 1$,

$$\mathbb{E}(\phi_n(X)|Y) = \rho_n \psi_n(Y) \quad \text{(a.s.)}, \qquad \mathbb{E}(\psi_n(Y)|X) = \rho_n \phi_n(X) \quad \text{(a.s.)}. \tag{10}$$

Clearly, if $\rho_1 \ne 0$, (10) with $n = 1$ shows that X and Y have strictly linear regression on each other. From (10) we conclude that $\rho(\phi_n(X), \psi_n(Y)) = \rho_n$ for all $n \ge 1$ and, therefore, $\rho(X, Y) = \rho(\phi_1(X), \psi_1(Y)) = \rho_1$ and $R(X, Y) \ge \sup_{n \ge 1} |\rho_n|$. Since the choice of $\{\rho_n\}_{n=1}^{\infty}$ is quite arbitrary (see (8)), it follows that

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

3. Concluding remarks

(a) Kingman [12] proved that there exist random vectors (X, Y) with $\mathbb{E}(X|Y) = a_1 Y$, $\mathbb{E}(Y|X) = b_1 X$ and $-1 < a_1 b_1 < 0$; of course, in such cases, $\mathbb{E}(X^2 + Y^2) = \infty$ so that $\rho(X, Y)$ is not defined. An open question is to find examples of L^2 random vectors (X, Y) with non-zero linear regressions, whose marginal densities have one sided unbounded supports, and $R(X, Y) > |\rho(X, Y)| > 0$. The question about exponential marginals was posed by Milan Stehlik of Johannes Kepler University.

(b) It is obvious that the construction (9) 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 [15] and Castaño-Martínez et al. [4].

(c) Distributions with densities of the form (9) are known as Lancaster distributions; Lancaster [14], Koudou [13], Diaconis and Griffiths [8]. They can be viewed as extensions of the Sarmanov-type distribution ($\rho_n = 0$ for $n \ge 2$) which, assuming standard uniform marginals, generalizes the so called Farlie–Gumbel–Morgenstern family.

(d) It is of some interest to observe that the mixture density (6) admits a series representation of the form (9). Indeed, let $\{h_n\}_{n=0}^{\infty}$ be the orthonormal system of (standardized) Hermite polynomials corresponding to the (univariate) standard normal density $\phi(t) = e^{-t^2/2}/\sqrt{2\pi}$. Sarmanov [22] showed that the series

$$g(x, y) := \phi(x)\phi(y)\left(1 + \sum_{n=1}^{\infty} \rho_n h_n(x) h_n(y)\right), \quad (x, y) \in \mathbb{R}^2,$$

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 ρ_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 (6).

Acknowledgments

I would like to thank H.N. Nagaraja and H.M. Barakat for bringing to my attention the papers by Rohatgi and Székely [19] and Sarmanov [21,20].

References

[1] M. Ahsanullah, Record Values—Theory and Applications, Univ. Press Amer. Inc., New York, 2004.
[2] B.C. Arnold, N. Balakrishnan, H.N. Nagaraja, Records, Wiley, New York, 1998.
[3] H.M. Barakat, The maximal correlation for the generalized order statistics and dual generalized order statistics, Arab. J. Math. 1 (2012) 149–158.
[4] A. Castaño-Martínez, F. López-Blázquez, B. Salamanca-Miño, Maximal correlation between order statistics, in: M. Ahsanullah, M. Raqab (Eds.), Recent Developments in Ordered Random Variables, Nova Science Publishers, 2007, pp. 55–68.
[5] P. Csáki, J. Fischer, On the general notion of maximal correlation, Magyar Tud. Akad. Mat. Kutató Int. Közl., Publ. Math. Inst., Hungar. Acad. Sci. 8 (1963) 27–51.
[6] H.A. David, H.N. Nagaraja, Order Statistics, Wiley, New York, 2003.
[7] A. Dembo, A. Kagan, L.A. Shepp, Remarks on the maximum correlation coefficient, Bernoulli 7 (2001) 343–350.
[8] P. Diaconis, R. Griffiths, Exchangeable pairs of Bernoulli random variables, Krawtchouck polynomials, and Ehrenfest urns, Aust. N. Z. J. Stat. 54 (2012) 81–101.
[9] M. Eaton, A characterization of spherical distributions, J. Multivariate Anal. 20 (1986) 272–276.
[10] H. Gebelein, Das statistische problem der Korrelation als variation und eigenwertproblem und sein Zusammenhang mit der ausgleichrechnung, Z. Angew. Math. Mech. 21 (1941) 364–379.
[11] H.O. Hirschfeld, A connection between correlation and contingency, Proc. Cambridge Philos. Soc. 31 (1935) 520–524.
[12] J.F.C. Kingman, Random variables with ansymmetrical linear regressions, Math. Proc. Cambridge Philos. Soc. 98 (1985) 355–365.
[13] A.E. Koudou, Lancaster bivariate probability distributions with Poisson, negative binomial and gamma margins, TEST 7 (1998) 95–110.
[14] H.O. Lancaster, Some properties of the bivariate normal distribution considered in the form of a contingency table, Biometrika 44 (1957) 289–292.
[15] F. López-Blázquez, A. Castaño-Martínez, Upper and lower bounds for the correlation ratio of order statistics from a sample without replacement, J. Statist. Plann. Inference 136 (2006) 43–52.
[16] V.B. Nevzorov, A characterization of exponential distributions by correlations between records, Math. Methods Statist. 1 (1992) 49–54.
[17] N. Papadatos, T. Xifara, A simple method for obtaining the maximal correlation coefficient and related characterizations, J. Multivariate Anal. 118 (2013) 102–114.
[18] A. Rényi, On measures of dependence, Acta Math. Acad. Sci. Hungar. 10 (1959) 441–451.
[19] V.K. Rohatgi, G.J. Székely, On the background of some correlation inequalities, J. Stat. Comput. Simul. 40 (1992) 260–262.
[20] O.V. Sarmanov, Maximum correlation coefficient (non-symmetric case), Dokl. Akad. Nauk SSSR 121 (1958) 52–55 (in Russian); Mathematical Statistics and Probability, Vol. 2, Amer. Math. Soc., 1962, pp. 207–210. Also in Selected Translations.
[21] O.V. Sarmanov, Maximum correlation coefficient (symmetric case), Dokl. Akad. Nauk SSSR 120 (1958) 715–718 (in Russian); Mathematical Statistics and Probability, Vol. 4, Amer. Math. Soc., 1963, pp. 271–275. Also in Selected Translations.
[22] O.V. Sarmanov, Generalized normal correlation and two-dimensional Fréchet classes, Dokl. Akad. Nauk SSSR 168 (1966) 32–35 (in Russian); Sov. Math. Dokl. 7 (1966) 596–599. English translation.
[23] G.J. Székely, A.K. Gupta, On a paper of V.B. Nevzorov, Math. Methods Statist. 7 (1998) 122.
[24] G.J. Székely, T.F. Móri, An extremal property of rectangular distributions, Statist. Probab. Lett. 3 (1985) 107–109.
[25] G.R. Terrell, A characterization of rectangular distributions, Ann. Probab. 11 (1983) 823–826.
[26] A.M. Vershik, Some characteristic properties of Gaussian stochastic processes, Theory Probab. Appl. 9 (1964) 390–394.