Journal of Multivariate Analysis 118 (2013) 102–114

Contents lists available at SciVerse ScienceDirect

Journal of Multivariate Analysis

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

A simple method for obtaining the maximal correlation coefficient and related characterizations[☆]

Nickos Papadatos [*], Tatiana Xifara

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

ARTICLE INFO

Article history:
Received 25 October 2012
Available online 6 April 2013

AMS subject classifications:
primary 62H20
62G30
secondary 62E10
60E15

Keywords:
Maximal correlation coefficient
Characterization of exponential distribution
Order statistics
Records
Splitting model

ABSTRACT

We provide a method that enables the simple calculation of the maximal correlation coefficient of a bivariate distribution, under suitable conditions. In particular, the method readily applies to known results on order statistics and records. As an application we provide a new characterization of the exponential distribution: Under a splitting model on independent identically distributed observations, it is the (unique, up to a location-scale transformation) parent distribution that maximizes the correlation coefficient between the records among two different branches of the splitting sequence.

1. Introduction

As is well-known, the *Pearson correlation coefficient* of the random variables (r.v.'s) X and Y is defined as

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

provided that $0<\mathrm{Var}(X)<\infty$ and $0<\mathrm{Var}(Y)<\infty$. It assumes values in the interval $[-1,1]$ and it is a measure of *linear dependence* of X and Y. Although $\rho(X,Y)=0$ for independent X and Y, the converse is not true. Gebelein [10] introduced the *maximal correlation coefficient*,

$$R(X,Y)=\sup\rho\big(g_1(X),g_2(Y)\big),$$

where the supremum is taken over all Borel functions $g_1:\mathbb{R}\to\mathbb{R}$ and $g_2:\mathbb{R}\to\mathbb{R}$ with $0<\mathrm{Var}\,g_1(X)<\infty$ and $0<\mathrm{Var}\,g_2(Y)<\infty$. In contrast to $\rho(X,Y)$, $R(X,Y)$ is defined whenever both X and Y are non-degenerate, assumes values in the interval $[0,1]$ and vanishes if and only if X and Y are independent. The maximal correlation coefficient plays a fundamental role in various areas of statistics; e.g., it is useful in obtaining optimal transformations for regression, Breiman and Friedman [5], and it has applications in the convergence theory of Gibbs sampling algorithms, Liu et al. [15].

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

* Corresponding author at: Section of Statistics and O.R., Department of Mathematics, University of Athens, Panepistemiopolis, 157 84 Athens, Greece.
E-mail addresses: npapadat@math.uoa.gr (N. Papadatos), t.xifara@lancaster.ac.uk (T. Xifara).
URLs: http://users.uoa.gr/~npapadat/ (N. Papadatos), http://www.maths.lancs.ac.uk/~xifara/ (T. Xifara).

0047-259X/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jmva.2013.03.017

However, despite its usefulness, it is often difficult to calculate the maximal correlation coefficient in an explicit form, except in some rare cases. A well-known exception is the result of Gebelein [10] and Lancaster [13] who show that if (X, Y) is bivariate normal then

$$R(X, Y) = |\rho(X, Y)|. \tag{1}$$

Another exception is provided by the surprising result of Dembo et al. [9], and its subsequent extensions given by Bryc et al. [6] and Yu [20]. In its general form the result states that for any independent identically distributed (i.i.d.) non-degenerate r.v.'s $X_1, \ldots, X_n$,

$$R(X_1 + \cdots + X_m, X_{k+1} + \cdots + X_n) = \frac{m-k}{\sqrt{m(n-k)}}, \quad 1 \le k+1 \le m \le n.$$

Finally, we mention an important result of Székely and Móri [18], who showed, using Jacobi polynomials, that if (X, Y) follows a bivariate density of the form

$$f(x, y) = \frac{\Gamma(\alpha+\beta+\gamma)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\gamma)} x^{\alpha-1}(y-x)^{\beta-1}(1-y)^{\gamma-1}, \quad 0 < x < y < 1, \tag{2}$$

where the parameters α, β, γ are positive, then

$$R(X, Y) = \rho(X, Y) = \sqrt{\frac{\alpha\gamma}{(\beta+\alpha)(\beta+\gamma)}}. \tag{3}$$

Observe that for any integers $1 \le i < j \le n$, the density of the pair of order statistics $(U_{i:n}, U_{j:n})$, based on n i.i.d. observations from the standard uniform distribution, is of the form (2) with $\alpha = i, \beta = j - i, \gamma = n + 1 - j$. Actually, (3) extends Terrell's [19] characterization of rectangular distributions through maximal correlation of an ordered pair.

In this article we provide a unified method for obtaining the maximal correlation coefficient when the bivariate distribution has a particular diagonal structure—see next section. The method is very simple (e.g., it readily applies to verify (1) and (3)) and it does not require knowledge of particular sets of orthogonal polynomials. Section 3 presents some notable examples of known characterizations of specific distributions through maximal correlation of ordered data and records. We consider a splitting model based on i.i.d. observations in Section 4. Applying our method it is shown that the records among two different branches of the splitting sequence are maximally correlated if and only if the population distribution is exponential (up to location-scale transformations)—this extends Nevzorov's [17] characterization.

2. The maximal correlation coefficient of bivariate distributions having diagonal structure

Let (X, Y) be an arbitrary random vector with distribution function $F(x, y)$ and assume that both X and Y are non-degenerate. We say that F, similarly the vector (X, Y), has diagonal structure if the following three conditions are satisfied.

A1. We assume that both X and Y have all their moments finite:

$$\mathbb{E}|X|^n < \infty \quad \text{and} \quad \mathbb{E}|Y|^n < \infty \quad \text{for } n = 1, 2, \ldots. \tag{4}$$

It is known that, under (4), there exists a (unique) orthonormal polynomial system (OPS) $\{\phi_n(x) = p_n x^n + \mathrm{Pol}_{n-1}(x),\ p_n > 0, n = 0, 1, \ldots\}$, corresponding to X, and a (unique) OPS $\{\psi_n(y) = q_n y^n + \mathrm{Pol}_{n-1}(y),\ q_n > 0, n = 0, 1, \ldots\}$, corresponding to Y. Here $\phi_0(x) \equiv \psi_0(y) \equiv 1$ and $\mathrm{Pol}_k(t)$ denotes an arbitrary polynomial in t of degree less than or equal to k, which may change from line to line. The orthonormality of the above OPS's means, as usual, that

$$\mathbb{E}[\phi_n(X)\phi_k(X)] = \mathbb{E}[\psi_n(Y)\psi_k(Y)] = \delta_{kn}, \quad k, n = 0, 1, \ldots,$$

where δ_{kn} is Kronecker's delta.

Remark 2.1. For a random variable X we denote by $\nu_X + 1$ the cardinality of its (minimal closed) support, $S(X)$, unifying the cases where $\nu_X < \infty$ and $\nu_X = \infty$. This convention is necessary because the OPS, corresponding to a non-degenerate r.v. X, reduces to the finite set $\{\phi_n(x)\}_{n=0}^{\nu_X}$ if (and only if) its support is concentrated on a finite subset of $\mathbb{R}$, with $\nu_X + 1 \ge 2$ points. This singular case, however, appears in some interesting situations—e.g., see Section 3, regarding the finite population case. In order to fix this problem (and give a unified presentation of the results) we shall proceed as follows. In any case where the support of X is of form $\{x_0, x_1, \ldots, x_{\nu_X}\}$, we shall enlarge the finite set of orthonormal polynomials to $\{\phi_n\}_{n=0}^{\infty}$, keeping $\{\phi_n\}_{n=0}^{\nu_X}$ as above, and defining

$$\phi_n(x) := x^{n-\nu_X-1}(x-x_0)(x-x_1)\cdots(x-x_{\nu_X}), \quad n > \nu_X.$$

Each ϕ_n in the enlarged set is of degree n and has principal coefficient $p_n > 0$. However, since for $n > \nu_X$, $\phi_n(X) = 0$ w.p. 1, the orthonormality assumption has now been relaxed to

$$\mathbb{E}[\phi_n(X)\phi_k(X)] = \delta_{kn}\mathbf{1}_{\{n \le \nu_X\}}, \quad k, n = 0, 1, \ldots,$$

where $\mathbf{1}$ stands for the indicator function. The same conventions will be applied to the OPS of Y, by setting $\psi_n(y) := y^{n-\nu_Y-1}(y-y_0)\cdots(y-y_{\nu_Y})$, whenever $n > \nu_Y$ and $S(Y) = \{y_0, \ldots, y_{\nu_Y}\}$ is finite.

A2. We assume that the OPS $\{\phi_n(x)\}_{n=0}^{\infty}$ is complete in $L^2(X)$, the Hilbert space of all Borel functions $g : \mathbb{R} \to \mathbb{R}$ with $\operatorname{Var} g(X) < \infty$. Clearly, the enlarged OPS of Remark 2.1 is complete if and only if the ordinary OPS is, noting that two functions g_1, g_2 are considered as "equal" if $\mathbb{P}[g_1(X) = g_2(X)] = 1$. Similarly, we assume that the system $\{\psi_n(y)\}_{n=0}^{\infty}$ is complete in $L^2(Y)$.

A3. We assume that the random vector (X, Y) has the *polynomial regression property*. That is,

$$\mathbb{E}(X^n|Y) = A_n Y^n + \mathrm{Pol}_{n-1}(Y), \quad n = 1, 2, \ldots,$$
$$\mathbb{E}(Y^n|X) = B_n X^n + \mathrm{Pol}_{n-1}(X), \quad n = 1, 2, \ldots,$$

where $A_n, B_n \in \mathbb{R}$.

The assumptions A1 and A2 are not very restrictive. For example, they are satisfied whenever both X and Y have finite moment generating functions in a neighborhood of 0; see, for example, [1,12]. However, this is not the case for assumption A3, since it applies to very particular distributions, as the following lemma shows.

Lemma 2.1. *Using the above notation and assuming* A1–A3 *we have that for all* $n, k \in \{1, 2, \ldots\}$,

$$\mathbb{E}[\phi_n(X)\psi_k(Y)] = \delta_{nk}\rho_n, \tag{5}$$

where δ_{nk} *is Kronecker's delta and* $\rho_n := \mathbb{E}[\phi_n(X)\psi_n(Y)] \in [-1, 1]$.

Proof. Set $\nu = \min\{\nu_X, \nu_Y\} \in \{1, 2, \ldots\} \cup \{\infty\}$ (for the definition of ν_X, ν_Y see Remark 2.1). If $n \le \nu$ then $\phi_n(X)$ and $\psi_n(Y)$ are standardized r.v.'s, and we have $\rho_n = \rho(\phi_n(X), \psi_n(Y))$. Therefore, $\rho_n \in [-1, 1]$ in this case. If at least one of ν_X, ν_Y is finite then for every $n > \nu$, $\phi_n(X)\psi_n(Y) = 0$ w.p. 1, so that $\rho_n = 0$ for $n > \nu$. Thus, $\rho_n \in [-1, 1]$ for all n. Now, if $1 \le k < n$ then A3 yields

$$\mathbb{E}[\phi_n(X)\psi_k(Y)] = \mathbb{E}\{\phi_n(X)\mathbb{E}[\psi_k(Y)|X]\} = \mathbb{E}[\phi_n(X)\mathrm{Pol}_k(X)] = 0,$$

because ϕ_n is orthogonal to any polynomial of degree at most $n-1$. Similar arguments apply to the case $1 \le n < k$, and the proof is complete. $\square$

The bivariate distributions satisfying (5) are sometimes called *Lancaster distributions* and the correlations ρ_n form a *Lancaster sequence* with respect to X and Y; see [11,12,14]. Therefore, by Lemma 2.1 we see that assumption A3 forces a distribution to be a Lancaster one. Under certain conditions, the density of a Lancaster distribution, if it exists, has the formal representation (diagonal structure)

$$f(x, y) = f_X(x) f_Y(y)\left(1 + \sum_{n=1}^{\infty} \rho_n \phi_n(x)\psi_n(y)\right),$$

where f_X and f_Y are the marginal densities of X and Y.

If the assumptions A1–A3 are satisfied then we can calculate each ρ_n, and this calculation does not require any knowledge of the polynomial systems $\{\phi_n(x)\}_{n=0}^{\infty}$ and $\{\psi_n(y)\}_{n=0}^{\infty}$. Indeed, we have the following

Lemma 2.2. *Let* $\nu = \min\{\nu_X, \nu_Y\}$ *(see Remark 2.1). Using the above notation and assuming* A1–A3 *we have that for all* $n \in \{1, 2, \ldots\}$,

$$A_n B_n \mathbf{1}_{\{n \le \nu\}} \ge 0, \qquad \rho_n = \operatorname{sign}(A_n)\sqrt{A_n B_n \mathbf{1}_{\{n \le \nu\}}} \quad \text{and} \quad |\rho_n| = \sqrt{A_n B_n \mathbf{1}_{\{n \le \nu\}}}. \tag{6}$$

Proof. Since $\phi_n(X) = p_n X^n + \mathrm{Pol}_{n-1}(X)$ and $\psi_n(Y) = q_n Y^n + \mathrm{Pol}_{n-1}(Y)$ we have

$$\begin{aligned}
\rho_n &= \mathbb{E}\{\psi_n(Y)\mathbb{E}(\phi_n(X)|Y)\} = \mathbb{E}\{\psi_n(Y)[p_n\mathbb{E}(X^n|Y) + \mathrm{Pol}_{n-1}(Y)]\} \\
&= p_n\mathbb{E}[\psi_n(Y)\mathbb{E}(X^n|Y)] + 0 = p_n\mathbb{E}\{\psi_n(Y)[A_nY^n + \mathrm{Pol}_{n-1}(Y)]\} \\
&= p_nA_n\mathbb{E}[\psi_n(Y)Y^n] + 0 = p_nA_n\mathbb{E}\{\psi_n(Y)q_n^{-1}[\psi_n(Y) - \mathrm{Pol}_{n-1}(Y)]\} \\
&= \frac{p_nA_n}{q_n}\mathbb{E}[\psi_n^2(Y)] - 0 = \frac{p_nA_n}{q_n}\mathbf{1}_{\{n \le \nu_Y\}}.
\end{aligned}$$

This shows that ρ_n and $A_n\mathbf{1}_{\{n \le \nu_Y\}}$ have the same sign (in particular, $\rho_n = 0$ for $n > \nu_Y$). Using the same arguments (conditioning on X) it follows that $\rho_n = \frac{q_nB_n}{p_n}\mathbf{1}_{\{n \le \nu_X\}}$; thus, $\rho_n = 0$ for $n > \nu_X$. Therefore, if ν is finite then $\rho_n = 0$ for all $n > \nu$. Finally, $\rho_n^2 = A_nB_n\mathbf{1}_{\{n \le \nu_X\}}\mathbf{1}_{\{n \le \nu_Y\}} = A_nB_n\mathbf{1}_{\{n \le \nu\}}$, and the proof is complete. $\square$

We are now in a position to state and prove our main result.

Theorem 2.1. *If the assumptions* A1–A3 *are satisfied and* $\nu = \min\{\nu_X, \nu_Y\}$ *then*

$$R(X,Y) = \sup_{n\ge 1} |\rho_n| = \sup_{n\ge 1} \sqrt{A_n B_n \mathbf{1}_{\{n\le\nu\}}}. \tag{7}$$

Moreover, if $|\rho_n| < |\rho_{n_0}|$ *for all* $n \ge 1, n \ne n_0$*, then for any* $g_1 \in L^2(X)$ *with* $\operatorname{Var} g_1(X) > 0$ *and for any* $g_2 \in L^2(Y)$ *with* $\operatorname{Var} g_2(Y) > 0$ *we have the inequality*

$$\rho(g_1(X), g_2(Y)) \le |\rho_{n_0}| = \sqrt{A_{n_0} B_{n_0}},$$

with equality if and only if $g_1(x) = a_0 + a_1\phi_{n_0}(x)$ *and* $g_2(y) = b_0 + b_1\psi_{n_0}(y)$ *for some constants* $a_0, b_0, a_1, b_1 \in \mathbb{R}$ *with* $a_1 b_1 \operatorname{sign}(A_{n_0}) > 0$.

Proof. Let $g_1 \in L^2(X)$ and denote by F_X the marginal distribution function of X. By the completeness of $\{\phi_n\}_{n=0}^{\infty}$ it follows that g_1 admits the representation

$$g_1(x) = \sum_{n=0}^{\infty} \alpha_n \phi_n(x), \quad \text{where } \alpha_n = \mathbb{E}[g_1(X)\phi_n(X)] = \int_{\mathbb{R}} g_1(x)\phi_n(x) dF_X(x).$$

Clearly, if ν_X is finite and $n > \nu_X$ then $\alpha_n = 0$, because $\mathbb{P}(\phi_n(X) = 0) = 1$; see Remark 2.1. The constants $\{\alpha_n\}_{n=0}^{\infty}$ are the Fourier coefficients of g_1 with respect to the OPS $\{\phi_n\}_{n=0}^{\infty}$, and the series converges in the $L^2(X)$-sense, i.e.,

$$\lim_N \mathbb{E}\left[g_1(X) - \sum_{n=0}^{N} \alpha_n \phi_n(X)\right]^2 = 0. \tag{8}$$

In particular, $\alpha_0 = \mathbb{E}[g_1(X)]$, and the above limit is usually written as Parseval's identity,

$$\operatorname{Var} g_1(X) = \sum_{n=1}^{\infty} \alpha_n^2$$

(equivalently, $\operatorname{Var} g_1(X) = \sum_{n=1}^{\nu_X} \alpha_n^2$ if $\nu_X < \infty$), since it is easily verified that

$$\mathbb{E}\left[g_1(X) - \sum_{n=0}^{N} \alpha_n \phi_n(X)\right]^2 = \operatorname{Var} g_1(X) - \sum_{n=1}^{N} \alpha_n^2.$$

Therefore, the assumption $\operatorname{Var} g_1(X) > 0$ implies that $\alpha_n \ne 0$ for at least one $n \ge 1$. Similarly, for any $g_2 \in L^2(Y)$ we have

$$\operatorname{Var} g_2(Y) = \sum_{n=1}^{\infty} \beta_n^2, \quad \text{where } \beta_n = \mathbb{E}[g_2(Y)\psi_n(Y)] = \int_{\mathbb{R}} g_2(y)\psi_n(y) dF_Y(y),$$

where F_Y is the marginal distribution of Y, $\{\beta_n\}_{n=0}^{\infty}$ are the Fourier coefficients of g_2 with respect to the OPS $\{\psi_n\}_{n=0}^{\infty}$ ($\beta_n = 0$ if $\nu_Y < \infty$ and $n > \nu_Y$) and, as for X,

$$\lim_N \mathbb{E}\left[g_2(Y) - \sum_{n=0}^{N} \beta_n \psi_n(Y)\right]^2 = \operatorname{Var} g_2(Y) - \lim_N \sum_{n=1}^{N} \beta_n^2 = 0. \tag{9}$$

Using the above we can show that

$$\mathbb{E}[g_1(X)\psi_n(Y)] = \alpha_n \rho_n \quad \text{and} \quad \mathbb{E}[g_2(Y)\phi_n(X)] = \beta_n \rho_n, \quad n = 1, 2, \ldots. \tag{10}$$

Indeed, for any $N \ge n$ we have

$$\mathbb{E}[g_1(X)\psi_n(Y)] = \mathbb{E}\left\{\left[g_1(X) - \sum_{k=0}^{N} \alpha_k \phi_k(X)\right]\psi_n(Y)\right\} + \sum_{k=0}^{N} \alpha_k \mathbb{E}[\phi_k(X)\psi_n(Y)].$$

Now $N \ge n$, $\phi_0(x) \equiv 1$, $\mathbb{E}[\psi_n(Y)] = 0$, $\mathbb{E}[\psi_n^2(Y)] = \mathbf{1}_{\{n\le\nu_Y\}}$ and $\mathbb{E}[\phi_k(X)\psi_n(Y)] = \delta_{kn}\rho_n$ for $k \ge 1$. Thus, in view of (8) and by the Cauchy–Schwarz inequality,

$$\begin{aligned} 0 \le \left(\mathbb{E}[g_1(X)\psi_n(Y)] - \alpha_n\rho_n\right)^2 &= \left(\mathbb{E}\left\{\left[g_1(X) - \sum_{k=0}^{N} \alpha_k\phi_k(X)\right]\psi_n(Y)\right\}\right)^2 \\ &\le \mathbb{E}\left[g_1(X) - \sum_{k=0}^{N} \alpha_k\phi_k(X)\right]^2 \mathbb{E}[\psi_n^2(Y)] \to 0, \quad \text{as } N \to \infty. \end{aligned}$$

Therefore, since $(\mathbb{E}[g_1(X)\psi_n(Y)] - \alpha_n\rho_n)^2$ does not depend on N, we conclude the first identity in (10). The remainder of (10) follows similarly. From (10) we obtain

$$\mathbb{E}\left[\left(g_1(X) - \sum_{n=0}^{N}\alpha_n\phi_n(X)\right)\left(g_2(Y) - \sum_{n=0}^{N}\beta_n\psi_n(Y)\right)\right] = \mathrm{Cov}[g_1(X), g_2(Y)] - \sum_{n=1}^{N}\rho_n\alpha_n\beta_n.$$

Thus, squaring the above identity and applying the Cauchy–Schwarz inequality to the resulting squared expectation we conclude, in view of (8) and (9), that

$$\mathrm{Cov}[g_1(X), g_2(Y)] = \sum_{n=1}^{\infty}\rho_n\alpha_n\beta_n = \sum_{n=1}^{\nu}\rho_n\alpha_n\beta_n. \tag{11}$$

(Recall that $\nu = \min\{\nu_X, \nu_Y\}$; for the definition of ν_X, ν_Y see Remark 2.1.) Therefore, combining the above we obtain the expression

$$\rho(g_1(X), g_2(Y)) = \frac{\sum_{n=1}^{\infty}\rho_n\alpha_n\beta_n}{\sqrt{\sum_{n=1}^{\infty}\alpha_n^2}\sqrt{\sum_{n=1}^{\infty}\beta_n^2}} = \frac{\sum_{n=1}^{\nu}\rho_n\alpha_n\beta_n}{\sqrt{\sum_{n=1}^{\nu_X}\alpha_n^2}\sqrt{\sum_{n=1}^{\nu_Y}\beta_n^2}}. \tag{12}$$

Observe that, in view of (11),

$$\begin{aligned}
\Big(\mathrm{Cov}[g_1(X), g_2(Y)]\Big)^2 &= \left|\sum_{n=1}^{\infty}\rho_n\alpha_n\beta_n\right|^2 \le \left(\sum_{n=1}^{\infty}|\rho_n|\,|\alpha_n|\,|\beta_n|\right)^2 \\
&= \left(\sum_{n=1}^{\infty}(\sqrt{|\rho_n|}|\alpha_n|)(\sqrt{|\rho_n|}|\beta_n|)\right)^2 \\
&\le \left(\sum_{n=1}^{\infty}|\rho_n|\alpha_n^2\right)\left(\sum_{n=1}^{\infty}|\rho_n|\beta_n^2\right) \\
&\le \left(\big(\sup_{n\ge 1}|\rho_n|\big)\sum_{n=1}^{\infty}\alpha_n^2\right)\left(\left(\sup_{n\ge 1}|\rho_n|\right)\sum_{n=1}^{\infty}\beta_n^2\right) \\
&= \big(\sup_{n\ge 1}\rho_n^2\big)\left(\sum_{n=1}^{\infty}\alpha_n^2\right)\left(\sum_{n=1}^{\infty}\beta_n^2\right).
\end{aligned}$$

The above inequality, combined with (12), shows that

$$R(X, Y) \le \sup_{n\ge 1}|\rho_n| = R, \quad \text{say.}$$

On the other hand, for any $\epsilon > 0$ we can find an index n_0 (with $n_0 \le \nu$ if ν is finite) such that $|\rho_{n_0}| > R - \epsilon$, and thus, $|\rho(\phi_{n_0}(X), \psi_{n_0}(Y))| = |\rho_{n_0}| > R - \epsilon$. Therefore,

$$\begin{aligned}
R(X, Y) &= \sup\rho(g_1(X), g_2(Y)) \\
&\ge \max\{\rho(\phi_{n_0}(X), \psi_{n_0}(Y)), \rho(-\phi_{n_0}(X), \psi_{n_0}(Y))\} \\
&= \max\{\rho_{n_0}, -\rho_{n_0}\} = |\rho_{n_0}| > R - \epsilon.
\end{aligned}$$

Since the inequality $R(X, Y) > R - \epsilon$ holds for all $\epsilon > 0$ it follows that $R(X, Y) \ge R$, and thus, $R(X, Y) = R$. It is clear that if the sequence $\{|\rho_n|\}_{n=1}^{\infty}$ has a unique maximum, say $|\rho_{n_0}| > 0$, then it is necessary that $n_0 \le \nu$ if ν is finite. Working as above, it is easily shown that

$$\Big(\mathrm{Cov}[g_1(X), g_2(Y)]\Big)^2 \le \rho_{n_0}^2\left(\sum_{n=1}^{\infty}\alpha_n^2\right)\left(\sum_{n=1}^{\infty}\beta_n^2\right) = \rho_{n_0}^2\,\mathrm{Var}\,g_1(X)\mathrm{Var}\,g_2(Y),$$

with equality if and only if $\alpha_n = \beta_n = 0$ for all $n \ge 1$, $n \ne n_0$. Combining this with the fact that ρ_{n_0} $(=\rho(\phi_{n_0}(X), \psi_{n_0}(Y)))$ has the sign of A_{n_0}, completes the proof. $\square$

3. Examples providing known characterizations via maximal correlation

The following known results are immediate applications of Theorem 2.1.

The bivariate normal case. Assumptions A1–A3 are easily checked for a bivariate normal. Indeed, if (X,Y) is bivariate normal with $\mathbb{E}(X)=\mu_1$, $\mathbb{E}(Y)=\mu_2$, $\mathrm{Var}(X)=\sigma_1^2>0$, $\mathrm{Var}(Y)=\sigma_2^2>0$ and $\rho(X,Y)=\rho\in[-1,1]$ then it is well-known that $(X|Y=y)\sim\mathcal{N}(\mu_1+\rho\frac{\sigma_1}{\sigma_2}(y-\mu_2),(1-\rho^2)\sigma_1^2)$. It follows that

$$(X|Y=y)\stackrel{\mathrm{d}}{=}\mu_1+\rho\frac{\sigma_1}{\sigma_2}(y-\mu_2)+\sigma_1\sqrt{1-\rho^2}Z,$$

where $Z\sim\mathcal{N}(0,1)$ and $\stackrel{\mathrm{d}}{=}$ denotes equality in distribution. Therefore,

$$\mathbb{E}[X^n|Y=y]=\mathbb{E}\left[\mu_1+\rho\frac{\sigma_1}{\sigma_2}(y-\mu_2)+\sigma_1\sqrt{1-\rho^2}Z\right]^n=\rho^n\frac{\sigma_1^n}{\sigma_2^n}y^n+\mathrm{Pol}_{n-1}(y).$$

That is,

$$\mathbb{E}(X^n|Y)=A_nY^n+\mathrm{Pol}_{n-1}(Y),\quad \text{where } A_n=\rho^n\frac{\sigma_1^n}{\sigma_2^n},\ n=1,2,\ldots.$$

Similarly, $\mathbb{E}(Y^n|X)=B_nX^n+\mathrm{Pol}_{n-1}(X)$, where $B_n=\rho^n\frac{\sigma_2^n}{\sigma_1^n}$ for all $n\ge1$. Thus, A3 is satisfied, while A1 and A2 are well-known for the normal law (the moment generating function is finite). Since $\nu=\infty$, it follows from (6) that $|\rho_n|=\sqrt{A_nB_n}=|\rho|^n$, $\rho_n=\mathrm{sign}(\rho^n)|\rho|^n=\rho^n$, and, by (7), $R(X,Y)=\sup_{n\ge1}|\rho_n|=\max_{n\ge1}|\rho|^n=|\rho|$. Moreover, in the particular case where $0<|\rho|<1$, the equality in

$$|\rho(g_1(X),g_2(Y))|\le|\rho|$$

is attained if and only if both g_1,g_2 are linear. It is worth noting that (11) takes the simple form (holding for any $\rho\in[-1,1]$)

$$\mathrm{Cov}[g_1(X),g_2(Y)]=\sum_{n=1}^{\infty}\frac{\rho^n\sigma_1^n\sigma_2^n}{n!}\mathbb{E}[g_1^{(n)}(X)]\mathbb{E}[g_2^{(n)}(Y)],\tag{13}$$

provided that $g_1,g_2\in C^\infty$, $g_1(X)\in L^2(X)$, $g_2(Y)\in L^2(Y)$, and that $\mathbb{E}|g_1^{(n)}(X)|<\infty$ and $\mathbb{E}|g_2^{(n)}(Y)|<\infty$ for all n, where $g_i^{(n)}$ denotes the n-th derivative of g_i, $i=1,2$. Of course, one can apply (13) to the case $X=Y$. Then, $\mu_1=\mu_2=\mu$, say, $\rho=1$, and $\sigma_1=\sigma_2=\sigma$, say, and (13) yields the generalized Stein identity for the $\mathcal{N}(\mu,\sigma^2)$ distribution (see [1]):

$$\mathrm{Cov}[g_1(X),g_2(X)]=\sum_{n=1}^{\infty}\frac{(\sigma^2)^n}{n!}\mathbb{E}[g_1^{(n)}(X)]\mathbb{E}[g_2^{(n)}(X)].$$

Characterization of rectangular distributions via maximal correlation of order statistics. Terrell [19], using Legendre polynomials, proved that if $X_{1:2}\le X_{2:2}$ are the order statistics of two i.i.d. observations from a distribution with finite variance then

$$\rho(X_{1:2},X_{2:2})\le\frac{1}{2},$$

and the equality characterizes the rectangular (uniform over some non-degenerate interval) distributions. However, Theorem 2.1 applies immediately here. Indeed, if $\mathcal{U}(a,b)$ denotes the uniform distribution over (a,b) and $U_1,U_2\sim\mathcal{U}(0,1)$ then it is obvious that the order statistics of U_1,U_2, $U_{1:2}\le U_{2:2}$, satisfy the following:

$$\begin{aligned}U_{1:2}|U_{2:2}\sim\mathcal{U}(0,U_{2:2})\Rightarrow\mathbb{E}\big(U_{1:2}^n|U_{2:2}\big)&=\int_0^{U_{2:2}}t^n\frac{1}{U_{2:2}}dt=\frac{1}{n+1}U_{2:2}^n,\\ U_{2:2}|U_{1:2}\sim\mathcal{U}(U_{1:2},1)\Rightarrow\mathbb{E}\big(U_{2:2}^n|U_{1:2}\big)&=\int_{U_{1:2}}^1t^n\frac{1}{1-U_{1:2}}dt\\&=\frac{1}{n+1}(1+U_{1:2}+\cdots+U_{1:2}^n).\end{aligned}$$

Thus, $A_n=B_n=\frac{1}{n+1}$ and $|\rho_n|=\frac{1}{n+1}$. Therefore, $\max_{n\ge1}|\rho_n|=|\rho_1|=\frac{1}{2}$. It follows from Theorem 2.1 that $\rho(g(U_{1:2}),g(U_{2:2}))\le\frac{1}{2}$, with equality if and only if g is linear. Since for order statistics $X_{1:2}\le X_{2:2}$ from an arbitrary distribution F

$$(X_{1:2},X_{2:2})\stackrel{\mathrm{d}}{=}(g(U_{1:2}),g(U_{2:2})),\quad\text{where } g(u)=\inf\{x:F(x)\ge u\},\ 0<u<1,$$

(the above g is usually denoted as F^{-1}), Terrell's result follows. The above argument can easily be extended to provide the characterization of Székely and Móri [18], concerning the order statistics $X_{1:n} \le \cdots \le X_{n:n}$ of a random sample $X_1, \ldots, X_n$. They show, using Jacobi polynomials, that for any integers $1 \le i < j \le n$,

$$\rho(X_{i:n}, X_{j:n}) \le \sqrt{\frac{i(n+1-j)}{j(n+1-i)}},$$

with equality if and only if the random sample arises from a rectangular distribution. Indeed, set $g(u) = F^{-1}(u) = \inf\{x : F(x) \ge u\}$, $0 < u < 1$, where F is the common distribution function of the i.i.d. r.v.'s $X_1, \ldots, X_n$, and consider the order statistics $U_{1:n} \le \cdots \le U_{n:n}$ of a random sample $U_1, \ldots, U_n$ from $\mathcal{U}(0, 1)$. Then, $(X_{i:n}, X_{j:n}) \stackrel{\mathrm{d}}{=} (g(U_{i:n}), g(U_{j:n}))$. Thus, $\rho(X_{i:n}, X_{j:n}) = \rho(g(U_{i:n}), g(U_{j:n}))$, which is well defined whenever $0 < \operatorname{Var} X_{i:n} + \operatorname{Var} X_{j:n} < \infty$. Since for any $s \in (0, 1)$, $(U_{i:n}|U_{j:n} = s) \stackrel{\mathrm{d}}{=} \widetilde{U}_{i:j-1}$, where $\widetilde{U}_{i:m}$ is the i-th order statistic of a sample with size m from $\mathcal{U}(0, s)$, we have

$$\widetilde{U}_{i:j-1} \stackrel{\mathrm{d}}{=} sU_{i:j-1} \Rightarrow \mathbb{E}[U_{i:n}^k | U_{j:n} = s] = \mathbb{E}[(sU_{i:j-1})^k] = s^k \mathbb{E}(U_{i:j-1}^k).$$

Now,

$$\begin{aligned}\mathbb{E}\big(U_{i:j-1}^k\big) &= \int_0^1 u^k \frac{1}{B(i, j-i)} u^{i-1}(1-u)^{j-i-1} du \\ &= \frac{B(k+i, j-i)}{B(i, j-i)} = \frac{(k+i-1)!(j-1)!}{(k+j-1)!(i-1)!}.\end{aligned}$$

In addition, for any $t \in (0, 1)$ we have $(U_{j:n}|U_{i:n} = t) \stackrel{\mathrm{d}}{=} \widetilde{U}_{j-i:n-i}$, where $\widetilde{U}_{j-i:n-i}$ is the $(j-i)$-th order statistic of a sample with size $n-i$ from $\mathcal{U}(t, 1)$. Clearly, if $\widetilde{U} \sim \mathcal{U}(t, 1)$ then $\widetilde{U} \stackrel{\mathrm{d}}{=} t+(1-t)U$ where $U \sim \mathcal{U}(0, 1)$. So, $(U_{j:n}|U_{i:n} = t) \stackrel{\mathrm{d}}{=} t+(1-t)U_{j-i:n-i}$ and since $U_{j-i:n-i} \stackrel{\mathrm{d}}{=} 1 - U_{n+1-j:n-i}$, we get $(U_{j:n}|U_{i:n} = t) \stackrel{\mathrm{d}}{=} 1 - U_{n+1-j:n-i} + tU_{n+1-j:n-i}$. Therefore,

$$\begin{aligned}\mathbb{E}\big[U_{j:n}^k | U_{i:n} = t\big] &= \mathbb{E}\big(1 - U_{n+1-j:n-i} + tU_{n+1-j:n-i}\big)^k = t^k \mathbb{E}\big(U_{n+1-j:n-i}^k\big) + \mathrm{Pol}_{k-1}(t) \\ &= \frac{(n+k-j)!(n-i)!}{(n+k-i)!(n-j)!} t^k + \mathrm{Pol}_{k-1}(t).\end{aligned}$$

Thus, A3 is satisfied with $A_k = [i]_k/[j]_k$ (where $[\alpha]_k := \alpha(\alpha+1)\cdots(\alpha+k-1)$), and $B_k = [n+1-j]_k/[n+1-i]_k$. Clearly, $\nu = \infty$, and hence,

$$\rho_k^2 = A_k B_k = \frac{[i]_k [n+1-j]_k}{[j]_k [n+1-i]_k}.$$

This is a strictly decreasing sequence in k, and Theorem 2.1 yields the inequality

$$\rho(X_{i:n}, X_{j:n}) \le \sqrt{\rho_1^2} = \sqrt{\frac{i(n+1-j)}{j(n+1-i)}},$$

with equality if and only if $g(u)$ $(=F^{-1}(u)) = \alpha u + \beta$ for some $\alpha > 0$ and $\beta \in \mathbb{R}$, i.e., $X \sim \mathcal{U}(\beta, \beta+\alpha)$, $\alpha > 0$.

The same arguments apply to the case where (X, Y) has a density as in (2). Then, it is easily shown that for any fixed x and y in $(0, 1)$,

$$(X|Y = y) \stackrel{\mathrm{d}}{=} yB_{\alpha,\beta} \quad \text{and} \quad (Y|X = x) \stackrel{\mathrm{d}}{=} x + (1-x)B_{\beta,\gamma} \stackrel{\mathrm{d}}{=} 1 - B_{\gamma,\beta} + xB_{\gamma,\beta},$$

where $B_{r,s}$ denotes a Beta r.v. with parameters $r > 0$ and $s > 0$. It follows that

$$\mathbb{E}(X^n|Y) = A_n Y^n \quad \text{and} \quad \mathbb{E}(Y^n|X) = B_n X^n + \mathrm{Pol}_{n-1}(X)$$

with

$$A_n = \mathbb{E}\big(B_{\alpha,\beta}^n\big) = \frac{[\alpha]_n}{[\alpha+\beta]_n} \quad \text{and} \quad B_n = \mathbb{E}\big(B_{\gamma,\beta}^n\big) = \frac{[\gamma]_n}{[\beta+\gamma]_n}.$$

Since $\rho_n^2 = A_n B_n = \frac{[\alpha]_n[\gamma]_n}{[\alpha+\beta]_n[\beta+\gamma]_n}$ is strictly decreasing in n, Theorem 2.1 shows that $R(X, Y) = |\rho_1| = \rho_1 = \rho(X, Y)$, which is identical to (3).

Nevzorov's characterization of exponential distribution. Nevzorov [17] proved that for any $n, m \in \{1, 2, \ldots\}$,

$$\rho(R_n, R_{n+m}) \le \sqrt{\frac{n}{n+m}},$$

where R_i is the i-th (upper) record from a continuous distribution F with finite variance. Here $R_1 = X_1$ is the first observed random variable in the i.i.d. sequence $\{X_i\}_{i=1}^{\infty}$. Moreover, equality characterizes the location-scale family of the standard exponential distribution.

Theorem 2.1 obtains Nevzorov's result immediately. Indeed, if W_i denotes the i-th record from $\mathcal{E}\mathrm{xp}(1)$ (with density $f(x) = e^{-x}, x > 0$) then

$$(W_n, W_{n+m}) \stackrel{\mathrm{d}}{=} (E_1 + \cdots + E_n, E_1 + \cdots + E_{n+m}), \quad n, m \in \{1, 2, \ldots\},$$

where $\{E_i\}_{i=1}^{\infty}$ is an i.i.d. sequence from $\mathcal{E}\mathrm{xp}(1)$—see, e.g., [2]. Setting $X = E_1 + \cdots + E_n$ and $Y = E_1 + \cdots + E_{n+m}$, the joint density of (X, Y) is

$$f_{X,Y}(x, y) = \frac{1}{\Gamma(n)\Gamma(m)} x^{n-1}(y-x)^{m-1}e^{-y}, \quad 0 < x < y < \infty,$$

and the conditional densities are

$$f_{X|Y}(x|y) = \frac{\Gamma(n+m)}{\Gamma(n)\Gamma(m)} x^{n-1}(y-x)^{m-1}y^{-(n+m-1)}, \quad x \in (0, y),$$

and

$$f_{Y|X}(y|x) = \frac{1}{\Gamma(m)}(y-x)^{m-1}e^{-(y-x)}, \quad y \in (x, \infty).$$

It follows that

$$\mathbb{E}(X^k|Y = y) = \frac{(k+n-1)!(n+m-1)!}{(k+n+m-1)!(n-1)!} y^k$$

and

$$\mathbb{E}(Y^k|X = x) = x^k + \frac{1}{\Gamma(m)} \sum_{i=1}^{k} \binom{k}{i} \Gamma(i+m) x^{k-i}.$$

Thus, A3 is satisfied with $A_k = \frac{(k+n-1)!(n+m-1)!}{(k+n+m-1)!(n-1)!}$ and $B_k = 1$, so that

$$\rho_k^2 = A_k B_k = \frac{(k+n-1)!(n+m-1)!}{(k+n+m-1)!(n-1)!} = \frac{[n]_k}{[n+m]_k}.$$

Since this is a strictly decreasing sequence in k, Theorem 2.1 yields the inequality

$$\rho(R_n, R_{n+m}) = \rho(g(W_n), g(W_{n+m})) \le \sqrt{\rho_1^2} = \sqrt{\frac{n}{n+m}},$$

where $g(u) = F^{-1}(1 - e^{-u})$, $u > 0$. The equality holds if and only if g is increasing and linear. That is, if and only if F is the distribution function of $\alpha E + \beta$ where $\alpha > 0$, $\beta \in \mathbb{R}$ and $E \sim \mathcal{E}\mathrm{xp}(1)$.

López-Blázquez and Castaño-Martínez' result on maximal correlation of order statistics from a finite population. Let $U_{1:n}^{(N)} < U_{2:n}^{(N)} < \cdots < U_{n:n}^{(N)}$ be the order statistics corresponding to a simple random sample, $U_1^{(N)}, \ldots, U_n^{(N)}$, taken without replacement from the finite ordered population $\Pi_N = \{1, 2, \ldots, N\}$, where $2 \le n < N$. Since $\mathbb{P}(U_{i:n}^{(N)} = k) = \binom{k-1}{i-1}\binom{N-k}{n-i}\binom{N}{n}^{-1}$ for $k \in \{i, i+1, \ldots, N-(n-i)\}$ (and 0 otherwise), and this defines a probability mass function with support $A_{i:n}^{(N)} := \{i, i+1, \ldots, N-(n-i)\}$, we conclude the identity

$$\sum_{k=i}^{N-(n-i)} \binom{k-1}{i-1}\binom{N-k}{n-i} = \binom{N}{n}, \quad 1 \le i \le n \le N. \tag{14}$$

Setting $[\alpha]_m = \alpha(\alpha+1)\cdots(\alpha+m-1)$ (with $[\alpha]_0 = 1$ for all $\alpha \in \mathbb{R}$) we can derive, with the help of (14), a simple expression for the ascending moments of $U_{i:n}^{(N)}$:

$$\mathbb{E}\left\{[U_{i:n}^{(N)}]_m\right\} = [N+1]_m \frac{[i]_m}{[n+1]_m}, \quad m = 1, 2, \ldots. \tag{15}$$

We also mention the following obvious relations, holding for all $1 \le i < j \le n$:

$$(U_{i:n}^{(N)}, U_{j:n}^{(N)}) \stackrel{\mathrm{d}}{=} (N+1-U_{n+1-i:n}^{(N)}, N+1-U_{n+1-j:n}^{(N)}), \tag{16}$$

$$(U_{i:n}^{(N)}|U_{j:n}^{(N)} = s) \stackrel{\mathrm{d}}{=} U_{i:j-1}^{(s-1)}, \quad s \in \{j, j+1, \ldots, N-(n-j)\}, \tag{17}$$

$$(U_{j:n}^{(N)}|U_{i:n}^{(N)} = k) \stackrel{\mathrm{d}}{=} k + U_{j-i:n-i}^{(N-k)}, \quad k \in \{i, i+1, \ldots, N-(n-i)\}. \tag{18}$$

Now, applying (15) and (17) we have

$$\mathbb{E}\big\{[U_{i:n}^{(N)}]_m|U_{j:n}^{(N)}=s\big\}=[s]_m\frac{[i]_m}{[j]_m},\quad m=1,2,\ldots. \tag{19}$$

Let $(X,Y)=(U_{i:n}^{(N)},U_{j:n}^{(N)})$ and observe that $\nu_X=\nu_Y=N-n\geq 1$; see Remark 2.1. Relation (19) shows that

$$\mathbb{E}\big([X]_m|Y\big)=\frac{[i]_m}{[j]_m}[Y]_m=\frac{[i]_m}{[j]_m}Y^m+\mathrm{Pol}_{m-1}(Y),\quad m=1,2,\ldots,$$

which implies, using induction on m, that

$$\mathbb{E}(X^m|Y)=\frac{[i]_m}{[j]_m}Y^m+\mathrm{Pol}_{m-1}(Y),\quad m=1,2,\ldots. \tag{20}$$

Similarly, set $i'=n+1-j$, $j'=n+1-i$ (so that $1\leq i'<j'\leq n$) and write $U_{i'}$ instead of $U_{i':n}^{(N)}$, $U_{j'}$ instead of $U_{j':n}^{(N)}$. Now, applying relations (16) and (19),

$$\begin{aligned}\mathbb{E}\big([Y]_m|X=k\big)&=\mathbb{E}\left\{[N+1-U_{i'}]_m|U_{j'}=N+1-k\right\}\\&=\mathbb{E}\left\{(-1)^m[U_{i'}]_m+\mathrm{Pol}_{m-1}(U_{i'})\mid U_{j'}=N+1-k\right\}\\&=(-1)^m\mathbb{E}\left\{[U_{i'}]_m\mid U_{j'}=N+1-k\right\}+\mathrm{Pol}_{m-1}(N+1-k)\\&=(-1)^m[N+1-k]_m\frac{[i']_m}{[j']_m}+\mathrm{Pol}_{m-1}(k)\\&=[k]_m\frac{[i']_m}{[j']_m}+\mathrm{Pol}_{m-1}(k)=[k]_m\frac{[n+1-j]_m}{[n+1-i]_m}+\mathrm{Pol}_{m-1}(k).\end{aligned}$$

It follows that $\mathbb{E}([Y]_m|X)=\frac{[n+1-j]_m}{[n+1-i]_m}[X]_m+\mathrm{Pol}_{m-1}(X)=\frac{[n+1-j]_m}{[n+1-i]_m}X^m+\mathrm{Pol}_{m-1}(X)$ and, using induction on m,

$$\mathbb{E}(Y^m|X)=\frac{[n+1-j]_m}{[n+1-i]_m}X^m+\mathrm{Pol}_{m-1}(X),\quad m=1,2,\ldots. \tag{21}$$

Clearly, (20) and (21) show that A3 is satisfied for (X,Y). Moreover, we have found that $A_m=[i]_m/[j]_m$ and $B_m=[n+1-j]_m/[n+1-i]_m$, both of which do not depend on N. Since $A_m>0$ and $\rho_m=\sqrt{A_mB_m\mathbf{1}_{\{m\leq N-n\}}}$ is strictly decreasing in $m\in\{1,\ldots,N-n,N-n+1\}$, Theorem 2.1 yields the inequality

$$\rho\big(g_1(U_{i:n}^{(N)}),g_2(U_{j:n}^{(N)})\big)\leq\sqrt{\rho_1^2}=\sqrt{\frac{i(n+1-j)}{j(n+1-i)}}.$$

The equality holds if and only if both g_1 and g_2 are (non-constant and) linear and with the same monotonicity. More precisely, the restriction of g_1 in the set $A_{i:n}^{(N)}$ has to be non-constant and linear and the restriction of g_2 in the set $A_{j:n}^{(N)}$ has to be non-constant, linear and with the same monotonicity as g_1. Note that both sets $A_{i:n}^{(N)}$ and $A_{j:n}^{(N)}$ contain at least two points if and only if $N\geq n+1$.

Lemma 2.1 of Balakrishnan et al. [3] asserts that for the non-decreasing function $g:\{1,2,\ldots,N\}\to\{x_1\leq x_2\leq\cdots\leq x_N\}:=\widetilde{\Pi}_N$ with $g(i)=x_i$, $i=1,2,\ldots,N$,

$$\big(g(U_{i:n}^{(N)}),g(U_{j:n}^{(N)})\big)\stackrel{\mathrm{d}}{=}(X_{i:n},X_{j:n}),\quad 1\leq i<j\leq n,$$

where $X_{1:n}\leq X_{2:n}\leq\cdots\leq X_{n:n}$ are the order statistics corresponding to a simple random sample drawn (without replacement) from the finite population $\widetilde{\Pi}_N$. Suppose that $\rho(X_{i:n},X_{j:n})$ is well-defined or, equivalently, that the elements of $\widetilde{\Pi}_N$ satisfy $x_i<x_{N-(n-i)}$ and $x_j<x_{N-(n-j)}$ (otherwise, at least one of $X_{i:n},X_{j:n}$ would be degenerate). Then we conclude that

$$\rho(X_{i:n},X_{j:n})\leq\sqrt{\frac{i(n+1-j)}{j(n+1-i)}},\quad 1\leq i<j\leq n<N. \tag{22}$$

The equality, for fixed i,j,n,N, characterizes those finite populations $\widetilde{\Pi}_N$ for which the sets $\{x_i,x_{i+1},\ldots,x_{N-(n-i)}\}$ and $\{x_j,x_{j+1},\ldots,x_{N-(n-j)}\}$, which may or may not have common points, consist of consecutive terms of two (possibly different) strictly increasing arithmetic progresses. That is, a population of size N with elements $x_1\leq x_2\leq\cdots\leq x_N$ satisfying $x_i<x_{N-(n-i)}$ and $x_j<x_{N-(n-j)}$ attains the equality in (22) if and only if there exist constants $a_1>0$, $b_1\in\mathbb{R}$, $a_2>0$ and $b_2\in\mathbb{R}$ such that

$$x_k=\begin{cases}a_1k+b_1,&\text{for }k=i,i+1,\ldots,N-(n-i),\\a_2k+b_2,&\text{for }k=j,j+1,\ldots,N-(n-j),\\\text{arbitrary},&\text{otherwise.}\end{cases}$$

López-Blázquez and Castaño-Martínez [16], using Hahn polynomials, have obtained a corresponding inequality for the correlation ratio, which implies inequality (22). Their arguments, however, apply to populations $\widetilde{\Pi}_N$ having N distinct elements. We also refer to Theorem 2.1 and Corollary 2.1 in [7], noting that the characterization result stated in Corollary 2.1 of this article is incomplete, unless the sets $A_{i:n}^{(N)}$ and $A_{j:n}^{(N)}$ have at least two common points, i.e., $N \ge n + (j - i) + 1$.

4. Records from a splitting model and a Nevzorov-type characterization of the exponential distribution

Assume that in a particular country and for a specific athletic event, the consecutive performances of the athletes are described by an i.i.d. sequence $\{X_i\}_{i=1}^{\infty}$. Here and elsewhere in this section, the common distribution of each X_i will be assumed to be continuous, i.e., with no atoms—absolute continuity is not required. As the time goes on, the common practice is that some data regarding the sequence of national records, i.e., the sequence $\{R_i\}_{i=1}^{\infty}$, are recorded, in contrast to the original performances of the athletes, X_i, which are usually lost or forgotten. The above considerations give rise to the classical record model, based on an i.i.d. sequence, which is well-developed in the literature; see [2]. Under this classical model the observed sequence $\{R_i\}_{i=1}^{n}$ of the first n upper national records is defined as $R_1 = X_1$ and $R_i = X_{T(i)}$, $i = 2, \ldots, n$, where $T(i) = \min\{j \in \{1, 2, \ldots\} : X_j > R_{i-1}\}$.

Suppose now that, after the appearance of the n-th national record, the initial country is divided into, say, two new countries (branches), and assume that the athletes in each country are of the same strength as they were before the division. Then, the subsequent national records in each branch will take into account the current (common) national record, R_n, and the subsequent sequence of their individual records will be of the form $(R'_{n+n_1}, R''_{n+n_2})$, with $n_1, n_2 \in \{1, 2, \ldots\}$. Clearly,

$$R'_{n+n_1} \stackrel{\mathrm{d}}{=} R_{n+n_1} \quad \text{and} \quad R''_{n+n_2} \stackrel{\mathrm{d}}{=} R_{n+n_2} \tag{23}$$

where R_{n+m} is the $(n+m)$-th record from the initial sequence, but as n_1 and n_2 become large, the r.v.'s R'_{n+n_1} and R''_{n+n_2} should tend towards independence.

Thus, the actual definition of the splitting record sequence is equivalent to the following model: Let $\{X_1, X'_1, X''_1, X_2, X'_2, X''_2, \ldots\}$ be an i.i.d. sequence of r.v.'s. Define the n-th upper record R_n as before (based on the X_i's), then set $R'_n = R''_n := R_n$ and $T'(n) = T''(n) := T(n)$. For $i = 1, 2, \ldots$ define the subsequent record times and record values by

$$T'(n+i) = \min\{j \in \{1, 2, \ldots\} : X'_j > R'_{n+i-1}\}, \qquad R'_{n+i} = X'_{T'(n+i)}, \quad \text{and}$$
$$T''(n+i) = \min\{j \in \{1, 2, \ldots\} : X''_j > R''_{n+i-1}\}, \qquad R''_{n+i} = X''_{T''(n+i)}.$$

Clearly, it is of some interest to study the correlation behavior of the marginal records under this model, since large correlation among these variables entails good prediction of one branch to the other. It is not surprising that, similarly to the classic case, the splitting record sequence satisfies several interesting properties. In particular, in what follows we shall make use of the following lemma.

Lemma 4.1. (a) *Let $\{(W'_{n+n_1}, W''_{n+n_2})\}_{n_1,n_2=1}^{\infty}$ be the splitting record sequence based on the i.i.d. sequence $\{E_i, E'_i, E''_i\}_{i=1}^{\infty}$ from the standard exponential distribution, $\mathcal{E}\mathrm{xp}(1)$. Then for each $n_1, n_2 \in \{1, 2, \ldots\}$,*

$$(W'_{n+n_1}, W''_{n+n_2}) \stackrel{\mathrm{d}}{=} (E_1 + \cdots + E_n + E'_1 + \cdots + E'_{n_1}, E_1 + \cdots + E_n + E''_1 + \cdots + E''_{n_2}). \tag{24}$$

(b) *Let $\{(R'_{n+n_1}, R''_{n+n_2})\}_{n_1,n_2=1}^{\infty}$ be the splitting record sequence based on the i.i.d. sequence $\{X_i, X'_i, X''_i\}_{i=1}^{\infty}$ from a non-atomic (continuous) distribution function F. Then, for each $n_1, n_2 \in \{1, 2, \ldots\}$,*

$$(R'_{n+n_1}, R''_{n+n_2}) \stackrel{\mathrm{d}}{=} (g(W'_{n+n_1}), g(W''_{n+n_2})), \tag{25}$$

where $g(u) = F^{-1}(1 - e^{-u})$, $u > 0$, with $F^{-1}(y) = \inf\{x : F(x) \ge y\}$, $y \in (0, 1)$.

The proof of Lemma 4.1 is simple and is left to the reader—cf. [2]. With the help of this lemma, Theorem 2.1 yields the following characterization.

Theorem 4.1. *If $(R'_{n+n_1}, R''_{n+n_2})$ are splitting records based on an i.i.d. sequence $\{X_i, X'_i, X''_i\}_{i=1}^{\infty}$ from a non-atomic distribution F with $\mathbb{E}(R'_{n+n_1})^2 < \infty$ and $\mathbb{E}(R''_{n+n_2})^2 < \infty$ then*

$$\rho(R'_{n+n_1}, R''_{n+n_2}) \le \frac{n}{\sqrt{n+n_1}\sqrt{n+n_2}}.$$

The equality holds if and only if F is the distribution function of $\alpha E + \beta$ for some $\alpha > 0$ and $\beta \in \mathbb{R}$, where $E \sim \mathcal{E}\mathrm{xp}(1)$.

Proof. Set $X = E_1 + \cdots + E_n + E'_1 + \cdots + E'_{n_1}$ and $Y = E_1 + \cdots + E_n + E''_1 + \cdots + E''_{n_2}$ with $(E_1, \ldots, E''_{n_2})$ being a vector of $n + n_1 + n_2$ i.i.d. standard exponential r.v.'s. It can be shown (see the proof of Theorem 4.2) that for all $k \in \{1, 2, \ldots\}$,

$$\mathbb{E}(X^k|Y) = \frac{[n]_k}{[n+n_2]_k} Y^k + \mathrm{Pol}_{k-1}(Y), \qquad \mathbb{E}(Y^k|X) = \frac{[n]_k}{[n+n_1]_k} X^k + \mathrm{Pol}_{k-1}(X).$$

That is, the random vector (X, Y) has the polynomial regression property with $A_k = [n]_k/[n+n_2]_k$ and $B_k = [n]_k/[n+n_1]_k$. Clearly, $\rho_k^2 = ([n]_k)^2/([n+n_1]_k[n+n_2]_k)$ is strictly decreasing in k. In view of Lemma 4.1, Theorem 2.1 shows that, with $g(u) = F^{-1}(1 - e^{-u})$,

$$\begin{aligned}\rho(R'_{n+n_1}, R''_{n+n_2}) &= \rho\big(g(W'_{n+n_1}), g(W''_{n+n_2})\big)\\ &= \rho(g(X), g(Y)) \le \sqrt{\rho_1^2} = \frac{n}{\sqrt{n+n_1}\sqrt{n+n_2}},\end{aligned}$$

where the equality holds if and only if $g : (0, \infty) \to \mathbb{R}$ is linear. This, together with the fact that g is assumed to be strictly increasing, completes the proof. $\square$

Provided that every component is representative as a sum on independent gamma r.v.'s with the same scale parameter, say $1/\lambda$, Theorem 4.1 and Nevzorov's [17] characterization reflect the polynomial regression property of a specific class of multivariate gamma random vectors. Recall that a random variable X follows a gamma distribution with parameters $\alpha > 0$ and $\lambda > 0$ if its density is given by

$$f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, \quad x > 0.$$

This is denoted by $X \sim \Gamma(\alpha; \lambda)$, while the notation $X \sim \Gamma(0; \lambda)$ (for some $\lambda > 0$) means that X is degenerate and takes the value zero w.p. 1. In any case, $\mathbb{E}X = \alpha/\lambda$ and $\operatorname{Var} X = \alpha/\lambda^2$. Under the above notation one can easily verify the following result, which contains both Theorem 4.1 and Nevzorov's characterization as particular cases. In fact, Theorem 4.2 obtains the maximal correlation of Cherian's bivariate gamma distribution—see [8], [4, pp. 322–325].

Theorem 4.2. *Let $X_i \sim \Gamma(\alpha_i; \lambda)$ $(i = 0, 1, 2)$ be independent r.v.'s with $\lambda > 0$, $\alpha_i \ge 0$ $(i = 0, 1, 2)$ and $\alpha_0 + \alpha_i > 0$ $(i = 1, 2)$. Then the random vector $(X, Y) = (X_0 + X_1, X_0 + X_2)$ follows a bivariate distribution with gamma marginals, namely $X \sim \Gamma(\alpha_0 + \alpha_1; \lambda)$ and $Y \sim \Gamma(\alpha_0 + \alpha_2; \lambda)$. Moreover, (X, Y) satisfies the polynomial regression property. More precisely, for all $n \in \{1, 2, \ldots\}$,*

$$\mathbb{E}(X^n|Y) = \sum_{j=0}^{n} \binom{n}{j} \frac{[\alpha_0]_j[\alpha_1]_{n-j}}{\lambda^{n-j}[\alpha_0+\alpha_2]_j} Y^j, \qquad \mathbb{E}(Y^n|X) = \sum_{j=0}^{n} \binom{n}{j} \frac{[\alpha_0]_j[\alpha_2]_{n-j}}{\lambda^{n-j}[\alpha_0+\alpha_1]_j} X^j,$$

where $[\alpha]_0 \equiv 1$ for all $\alpha \in \mathbb{R}$ and $[\alpha]_k = \alpha(\alpha+1)\cdots(\alpha+k-1)$ for $k \in \{1, 2, \ldots\}$. Finally, for any $g_1 \in L^2(X)$ with $\operatorname{Var} g_1(X) > 0$ and for any $g_2 \in L^2(Y)$ with $\operatorname{Var} g_2(Y) > 0$ we have the inequality

$$\rho(g_1(X), g_2(Y)) \le \frac{\alpha_0}{\sqrt{\alpha_0+\alpha_1}\sqrt{\alpha_0+\alpha_2}}.$$

Provided that $\alpha_1 + \alpha_2 > 0$, the equality holds if and only if either (i) *$\alpha_0 = 0$ and g_1, g_2 are arbitrary or* (ii) *$\alpha_0 > 0$ and both g_1, g_2 are non-constant, linear and with the same monotonicity.*

Proof. Cases $\alpha_0 = 0$ and $\alpha_1 = \alpha_2 = 0$ are simple (X, Y are independent and $X = Y$ w.p. 1, respectively). Both cases $\alpha_0 > 0, \alpha_1 = 0, \alpha_2 > 0$ and $\alpha_0 > 0, \alpha_1 > 0, \alpha_2 = 0$ are similar to Nevzorov's case and can be shown as in Section 3. Assume now that $\alpha_i > 0$ for $i = 0, 1, 2$. Then, it is easily shown that the conditional density of X given $Y = y$ (for any fixed $y > 0$) is

$$f_{X|Y}(x|y) = ce^{-\lambda x} \int_0^{\min\{x,y\}} w^{\alpha_0-1}(x-w)^{\alpha_1-1}(y-w)^{\alpha_2-1} e^{\lambda w} dw, \quad x > 0,$$

where

$$c = c(\alpha_0, \alpha_1, \alpha_2; \lambda; y) = \frac{\lambda^{\alpha_1}\Gamma(\alpha_0+\alpha_2)}{y^{\alpha_0+\alpha_2-1}\Gamma(\alpha_0)\Gamma(\alpha_1)\Gamma(\alpha_2)}.$$

Despite the fact that this conditional density is not given in a closed form, we can calculate $\mathbb{E}(X^n|Y = y)$ using Tonelli's theorem. Indeed, consider the nonnegative functions $\theta(w) = w^{\alpha_0-1}e^{\lambda w}$ $(w > 0)$ and $h(x, y, w) = (x-w)^{\alpha_1-1}(y-w)^{\alpha_2-1}\mathbf{1}_{\{w<\min\{x,y\}\}}$ $(x, y, w > 0)$. Then,

$$\begin{aligned}\mathbb{E}(X^n|Y = y) &= c\left\{\int_0^y x^n e^{-\lambda x} \int_0^x \theta(w)h(x,y,w)dwdx + \int_y^\infty x^n e^{-\lambda x} \int_0^y \theta(w)h(x,y,w)dwdx\right\}\\ &= c\left\{\int_0^y \theta(w) \int_w^y x^n e^{-\lambda x} h(x,y,w)dxdw + \int_0^y \theta(w) \int_y^\infty x^n e^{-\lambda x} h(x,y,w)dxdw\right\}\end{aligned}$$

$$= c\int_0^y \theta(w)\int_w^\infty x^n e^{-\lambda x} h(x,y,w)dxdw$$
$$= c\int_0^y w^{\alpha_0-1}(y-w)^{\alpha_2-1}\left\{\int_0^\infty (x+w)^n e^{-\lambda x}x^{\alpha_1-1}dx\right\}dw.$$

Now, expanding $(x+w)^n$ according to Newton's formula and using $\int_0^\infty x^{j+\alpha_1-1}e^{-\lambda x}dx = \Gamma(\alpha_1+j)/\lambda^{\alpha_1+j}$ $(j = 0, 1, \ldots, n)$, the inner integral is equal to

$$\int_0^\infty (x+w)^n e^{-\lambda x}x^{\alpha_1-1}dx = \frac{1}{\lambda^{\alpha_1}}\sum_{j=0}^n \binom{n}{j}\frac{\Gamma(\alpha_1+j)}{\lambda^j}w^{n-j}.$$

Substituting this expression to the double integral, we obtain

$$\mathbb{E}(X^n|Y=y) = \frac{c}{\lambda^{\alpha_1}}\sum_{j=0}^n \binom{n}{j}\frac{\Gamma(\alpha_1+j)}{\lambda^j}\int_0^y w^{\alpha_0+(n-j)-1}(y-w)^{\alpha_2-1}dw$$
$$= \frac{c}{\lambda^{\alpha_1}}\sum_{j=0}^n \binom{n}{j}\frac{\Gamma(\alpha_1+j)}{\lambda^j}\frac{\Gamma(\alpha_2)\Gamma(\alpha_0+n-j)}{\Gamma(\alpha_0+\alpha_2+n-j)}y^{\alpha_0+\alpha_2+(n-j)-1}$$
$$= \frac{\Gamma(\alpha_0+\alpha_2)}{\Gamma(\alpha_0)\Gamma(\alpha_1)}\sum_{j=0}^n \binom{n}{j}\frac{\Gamma(\alpha_0+j)\Gamma(\alpha_1+n-j)}{\lambda^{n-j}\Gamma(\alpha_0+\alpha_2+j)}y^j.$$

Therefore, X has polynomial regression on Y and, similarly, Y has polynomial regression on X. It follows that (X, Y) satisfies conditions A1–A3 and, moreover,

$$\rho_n = \mathrm{sign}(A_n)\sqrt{A_nB_n} = \frac{[\alpha_0]_n}{\sqrt{[\alpha_0+\alpha_1]_n}\sqrt{[\alpha_0+\alpha_2]_n}}.$$

Since $|\rho_n| = \rho_n$ is strictly decreasing in n, a final application of Theorem 2.1 completes the proof. □

Theorem 4.2 includes Nevzorov's [17] characterization because, taking $\lambda = 1, \alpha_0 = n, \alpha_1 = 0, \alpha_2 = m$ and $g_1(u) = g_2(u) = F^{-1}(1-e^{-u}), u > 0$, we have that, under the standard record model, $(R_n, R_{n+m}) \stackrel{d}{=} (g(W_n), g(W_{n+m})) \stackrel{d}{=} (g(X), g(Y))$. Here (W_n, W_{n+m}) are the corresponding upper records from the standard exponential distribution. Clearly, the theorem also includes the result on splitting record models of Theorem 4.1—the only difference being that, due to Lemma 4.1, one has now to put $\alpha_1 = n_1$ (rather than $\alpha_1 = 0$) and $\alpha_2 = n_2$ (rather than $\alpha_2 = m$).

Provided that $g_1, g_2 \in C^\infty(0,\infty)$, $g_1(X) \in L^2(X)$, $g_2(Y) \in L^2(Y)$, and assuming that $\mathbb{E}|X^n g_1^{(n)}(X)| < \infty$ and $\mathbb{E}|Y^n g_2^{(n)}(Y)| < \infty$ for all n, where $g_i^{(n)}$ denotes the n-th derivative of g_i, $i = 1, 2$, it is of some interest to note that (11) yields the covariance identity

$$\mathrm{Cov}[g_1(X), g_2(Y)] = \sum_{n=1}^\infty \frac{[\alpha_0]_n}{n![\alpha_0+\alpha_1]_n[\alpha_0+\alpha_2]_n}\mathbb{E}\big(X^n g_1^{(n)}(X)\big)\mathbb{E}\big(Y^n g_2^{(n)}(Y)\big). \tag{26}$$

Of course one can apply (26) to the case $\alpha_1 = \alpha_2 = 0$, $\alpha_0 > 0$. Then, $X = Y \sim \Gamma(\alpha_0; \lambda)$ and we reobtain the generalized Stein-type identity for the $\Gamma(\alpha_0; \lambda)$ distribution (see [1]):

$$\mathrm{Cov}[g_1(X), g_2(X)] = \sum_{n=1}^\infty \frac{1}{n![\alpha_0]_n}\mathbb{E}\big(X^n g_1^{(n)}(X)\big)\mathbb{E}\big(X^n g_2^{(n)}(X)\big). \tag{27}$$

Similarly, we can apply (26) to the classical record setup from the standard exponential (setting $\lambda = 1, \alpha_0 = n, \alpha_1 = 0$ and $\alpha_2 = m$). Then we get

$$\mathrm{Cov}[g_1(W_n), g_2(W_{n+m})] = \sum_{k=1}^\infty \frac{1}{k![n+m]_k}\mathbb{E}\big(W_n^k g_1^{(k)}(W_n)\big)\mathbb{E}\big(W_{n+m}^k g_2^{(k)}(W_{n+m})\big). \tag{28}$$

5. Conclusions

The simplicity of the proposed method depends heavily on the polynomial regression property, A3, which is satisfied by all bivariate distributions discussed in the present article. Incidentally, in all of our cases we concluded that $R = |\rho(X, Y)| = \sqrt{A_1B_1}$, and some times it is asserted that this is the typical situation whenever A3 is merely satisfied for $n = 1$ (i.e., when both variables have linear regression). However, this is not true, e.g., when (X, Y) is uniformly distributed on the interior of the unit disc (then $A_1 = B_1 = \rho(X, Y) = 0$); see, also, [9].

Castaño-Martínez et al. [7] develop a correlation model for partial minima (or maxima) rather than records. Their Section 3 indicates that many difficulties can enter to the correlation problem when A3 fails. It appears that, in such cases, one has to calculate the values of $\rho_{n,k} := \mathbb{E}[\phi_n(X)\psi_k(Y)]$ for all n and k. This is not an easy task in general, in contrast to the present simplified situation, where knowledge of the values A_n and B_n in A3 suffices for the calculation of the maximal correlation coefficient.

Acknowledgment

We are grateful to an anonymous referee for pointing out a serious error in the previous version of our main result, Theorem 2.1, regarding the finite support case.

References

[1] G. Afendras, N. Papadatos, V. Papathanasiou, An extended Stein-type covariance identity for the Pearson family, with applications to lower variance bounds, Bernoulli 17 (2011) 507–529.
[2] B.C. Arnold, N. Balakrishnan, H.N. Nagaraja, Records, Wiley, New York, 1998.
[3] N. Balakrishnan, C. Charalambides, N. Papadatos, Bounds on expectation of order statistics from a finite population, J. Statist. Plann. Inference 113 (2003) 569–588.
[4] N. Balakrishnan, C.D. Lai, Continuous Bivariate Distributions, second ed., Springer, New York, 2009.
[5] L. Breiman, J.H. Friedman, Estimating optimal transformations for multiple regression and correlation, J. Amer. Statist. Assoc. 80 (1985) 590–598.
[6] W. Bryc, A. Dembo, A. Kagan, On the maximum correlation coefficient, Theory Probab. Appl. 49 (2005) 132–138.
[7] 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.
[8] K.C. Cherian, A bivariate correlated gamma-type distribution function, J. Indian Math. Soc. 5 (1941) 133–144.
[9] A. Dembo, A. Kagan, L.A. Shepp, Remarks on the maximum correlation coefficient, Bernoulli 7 (2001) 343–350.
[10] H. Gebelein, Das statistische problem der korrelation als variation und eigenwertproblem und sein zusammenhang mit der ausgleichrechnung, ZAMM Z. Angew. Math. Mech. 21 (1941) 364–379.
[11] A.E. Koudou, Probabilités de Lancaster, Expo. Math. 14 (1996) 247–275.
[12] A.E. Koudou, Lancaster bivariate probability distributions with Poisson, negative binomial and gamma margins, TEST 7 (1998) 95–110.
[13] H.O. Lancaster, Some properties of the bivariate normal distribution considered in the form of a contingency table, Biometrika 44 (1957) 289–292.
[14] H.O. Lancaster, The Chi-Squared Distribution, Wiley, New York, 1969.
[15] J.S. Liu, W.H. Wong, A. Kong, Covariance structure of the Gibbs sampler with applications to the comparisons of estimators and augmentation schemes, Biometrika 81 (1994) 27–40.
[16] 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.
[17] V.B. Nevzorov, A characterization of exponential distributions by correlations between records, Math. Methods Statist. 1 (1992) 49–54.
[18] G.J. Székely, T.F. Móri, An extremal property of rectangular distributions, Statist. Probab. Lett. 3 (1985) 107–109.
[19] G.R. Terrell, A characterization of rectangular distributions, Ann. Probab. 11 (1983) 823–826.
[20] Y. Yu, On the maximal correlation coefficient, Statist. Probab. Lett. 78 (2008) 1072–1075.