A Discrete Analogue of Terrell’s Characterization of Rectangular Distributions

Nickos Papadatos*

National and Kapodistrian University of Athens, Department of Mathematics, Section of Statistics and Operations Research, Panepistemiopolis, 15784 Athens, Greece

Received June 2, 2022; revised May 8, 2023; accepted May 21, 2023

Abstract—Terrell [18] showed that the Pearson coefficient of correlation of an ordered pair from a random sample of size two is at most one-half, and the equality is attained only for rectangular (uniform over some interval) distributions. In the present note it is proved that the same is true for the discrete case, in the sense that the correlation coefficient attains its maximal value only for discrete rectangular (uniform over some finite lattice) distributions.

DOI: 10.3103/S1066530723020035

Keywords: discrete rectangular distribution, order statistics, Hahn polynomials, Pearson coefficient of correlation

1. INTRODUCTION: A BRIEF HISTORY AND THE MAIN RESULT

For independent and identically distributed random variables $X_1$, $X_2$ from a probability distribution function $F$, the corresponding order statistics will be denoted by $X_{1:2} \leq X_{2:2}$, that is, $X_{1:2} = \min\{X_1, X_2\}$, $X_{2:2} = \max\{X_1, X_2\}$, and the Pearson coefficient of correlation by

$$\rho_{12} := \frac{\text{Cov}(X_{1:2}, X_{2:2})}{\sqrt{\text{Var}X_{1:2}} \sqrt{\text{Var}X_{2:2}}}. $$

For a random pair $(X, Y)$, the correlation coefficient, $\rho(X, Y)$, is well defined (and belongs to the interval $[-1, 1]$) if and only if both $X, Y$ are non-degenerate with finite second moment. Thus, for $\rho_{12}$ to be well defined it is necessary and sufficient that $F$ is non-degenerate and possesses finite second moment (so that $0 < \text{Var}X_{j:2} < \infty$, $j = 1, 2$).

It is known for a long time that $\rho_{12} > 0$; this follows immediately if we take expectations to $X_{1:2}X_{2:2} = X_1X_2$ and $X_{1:2} + X_{2:2} = X_1 + X_2$, yielding $\text{Cov}(X_{1:2}, X_{2:2}) = (\mathbb{E}X_{2:2} - \mathbb{E}X_1)^2 > 0$, since $\mathbb{E}X_1 < \mathbb{E}X_{2:2}$ for any non-degenerate $F$. At the time of 1980’s, David Scott and Robert Bartoszyński, in connection with a problem in cell division, proposed to Terrell the conjecture that $\rho_{12}$ is never greater then one-half. This was proved true.

Theorem 1.1 [18]. If $F$ is a continuous distribution with finite variance then $\rho_{12} \leq 1/2$, with equality if and only if $F$ is a rectangular distribution.

Terrell’s result is based on the expansion of a function in a series of Legendre polynomials, it is quite complicated, and imposes the unnecessary restriction that $F$ is continuous. Székely and Móri (1985) interpreted Terrell’s result as a maximal correlation problem—see Gebelein [4], Rényi [16]—that is,

$$R(X_{1:2}, X_{2:2}) := \sup_{g_1, g_2} \rho(g_1(X_{1:2}), g_2(X_{2:2})), $$

where the supremum is taken over non-constant functions $g_1 \in L^2(X_{1:2})$, $g_2 \in L^2(X_{2:2})$. In this way, Székely and Móri improved Terrell’s result in four directions. First, they removed the restriction that $F$ is continuous; second, they simplified Terrell’s proof; third, they showed that the correlation coefficient of $g_1(X_{1:2})$ and $g_2(X_{2:2})$ is less than $1/2$, for any distribution and any functions $g_1$ and $g_2$; fourth, and

*E-mail: npapadat@math.uoa.gr, users.uoa.gr/~npapadat/
most important, they extended these results to any sample size \( n \geq 2 \), obtaining the inequality (denote by \( X_{1:n} \leq \cdots \leq X_{n:n} \) the order statistics from a random sample of size \( n \) from \( F \))

\[
\rho(X_{i:n}, X_{j:n}) \leq \sqrt{\frac{i(n+1-j)}{j(n+1-i)}}, \quad 1 \leq i < j \leq n, \quad n \geq 2,
\]

which is valid for all non-degenerate distribution functions \( F \) with \( \text{Var}X_{1:n} + \text{Var}X_{j:n} < \infty \). The equality in (1.1), for a single value of \((i, j, n)\), characterizes the rectangular distributions.

Subsequently, Nevzorov [12]—see also López-Blázquez [7]—obtained a similar inequality for upper records, in which the equality characterizes the location-scale family of the standard exponential random variable. Later on, López-Blázquez and Castaño-Martínez [8] proved that the TSM inequality (1.1) is also valid when the order statistics are based on a without-replacement sample of size \( n \), taken from a finite population \( \mathcal{P} = \{x_1, \ldots, x_N\} \) with \( N > n \) distinct elements.

It is shown in Papadatos and Xifara [14] that all presenting results are based, essentially, in the following polynomial regression property (PRP):

\[
\mathbb{E}(X^k|Y) = A_k Y^k + P_{k-1}(Y), \quad \mathbb{E}(X^k|X) = B_k X^k + Q_{k-1}(X), \quad k = 1, 2, \ldots
\]

where \( P_{k-1} \) and \( Q_{k-1} \) are polynomials of degree at most \( k - 1 \). If a random pair \((X, Y)\) satisfies (1.2) then, under mild conditions, its maximal correlation, \( R(X, Y) \), equals to \( \sup_{k \geq 1} \sqrt{A_k B_k} \), so one simply has to calculate the principal coefficients \( A_k \) and \( B_k \) in (1.2), and choose the value of \( k = k_0 \) that maximizes the products \( A_k B_k \). If this value of \( k_0 \) is unique, then the equality in the inequality

\[
\rho(g_1(Y), g_2(Y)) \leq R(X, Y)
\]

is attained if and only if \( \mathbb{P}[g_1(X) = c_1 + \lambda_1 \phi_{k_0}(X)] = 1 \), \( \mathbb{P}[g_2(Y) = c_2 + \lambda_2 \psi_{k_0}(Y)] = 1 \), and \( \lambda_1 \lambda_2 > 0 \), where \( \{\phi_k\}_{k=0}^{\infty} \) is the complete, orthonormal polynomial system in \( L^2(X) \) (with the convention that each \( \phi_k \) has positive principal coefficient) and \( \{\psi_k\}_{k=0}^{\infty} \) the corresponding system in \( L^2(Y) \). Let us denote by \( U_{1:n} < \cdots < U_{n:n} \) the order statistics either from uniform in the interval \((0, 1)\), or from the discrete uniform in \( \{1, \ldots, N\} \) (the latter in the without-replacement case). It is a simple exercise to verify that the pair \((U_{i:n}, U_{j:n})\) possesses the PRP (1.2), and it is easy to calculate \( A_k \) and \( B_k \) in the without-replacement case this is slightly more complicated); also, it is plain to check the PRP for upper records \((W_n, W_m)\) \((n < m)\) from the standard exponential, and to obtain the constants \( A_k \) and \( B_k \) in (1.2). Then, the PRP method yields all the presenting results at once and, in the discrete case, it provides the exact characterization of those populations \( \mathcal{P} \) that attain the upper bound in the TSM inequality (1.1), even if ties are allowed; see Section 3 in Papadatos and Xifara [14]. In all of the foregoing results, the sequence \( A_k B_k \) is uniquely maximized by its first term, hence, \( R = \sqrt{A_1 B_1} \), and the equality is attained by linear functions \( g_1, g_2 \); equivalently, \( R(X, Y) = \|\rho(X, Y)\| \). While the equality \( |\rho| = R \) often appears to problems regarding maximal correlation under PRP, it is not always true; see Papadatos [13].

In the present note we shall prove the following discrete analogue of Theorem 1.1. First, we provide a useful definition (see Balakrishnan et al. [2]).

**Definition 1.1** (Uniform distribution on \( \mathcal{P} = \{x_1, \ldots, x_N\} \) in the possible presence of ties). Let \( k_1 \) of the \( x_i \)’s be equal to \( y_1 \), \( k_2 \) of the \( x_i \)’s be equal to \( y_2 \), and so on, where \( k_1 + \cdots + k_m = N \) and, with no loss of generality, \( y_1 < \cdots < y_m \). We say that \( X \) is uniform on \( \mathcal{P} = \{x_1, \ldots, x_N\} \) if \( \mathbb{P}(X = y_j) = k_j/N, \ j = 1, \ldots, m \).

The main result is the following.

**Theorem 1.2.** Let \( X_1, X_2 \) be independent random variables with uniform distribution on \( \mathcal{P} = \{x_1, \ldots, x_N\} \), where \( N \geq 2 \) and \( x_1 \leq \cdots \leq x_N \) with \( x_1 < x_N \). Then, under the notation of Theorem 1.1,

\[
\rho_{12} \leq \frac{1 - N^{-2}}{2 + N^{-2}},
\]

and the equality is attained if and only if \( \mathcal{P} \) is a discrete lattice, that is, \( x_{i+1} - x_i = \lambda > 0 \) (constant), \( i = 1, \ldots, N - 1 \).
A straightforward computation shows that the upper bound in (1.3) equals to $\rho_{12}$ when $P = P_0 := \{1, \ldots, N\}$. If $U_{1:2} \leq U_{2:2}$ are the order statistics from $P_0$, it can be checked that the random pair $(U_{1:2}, U_{2:2})$ does not posses the PRP (1.2); in particular, $U_{1:2}$ does not have linear regression on $U_{2:2}$. Indeed, for $x, y \in P_0$ with $x \leq y$, $\mathbb{P}(U_{1:2} = x|U_{2:2} = y) = 2/(2y - 1)$ if $x < y$, and $\mathbb{P}(U_{1:2} = y|U_{2:2} = y) = 1/(2y - 1)$. Hence, one finds $\mathbb{E}(U_{1:2}|U_{2:2} = y) = y^2/(2y - 1)$, and this regression (with $y$ restricted to $P_0$) is, clearly, nonlinear, unless $N = 2$. Therefore, (1.2) fails, and this confirms the essentially different nature of the present problem, compared to the preceding ones. Notice that the case $N = 2$ is trivial, since $\rho_{12} \equiv 1/3$ for every choice of $x_1, x_2$, with $x_1 < x_2$ (the coefficient of correlation is location-scale invariant).

Results for some particular interesting cases where PRP is not satisfied are provided by Castraño-Martínez et al. [3] and López-Blázquez and Salamanca-Miño [10]. The present problem, however, does not fall to the “non-diagonal” case which has been studied by the second article, and it is not a maximal correlation problem, either; see Remark 3.4.

It is important to note that Theorem 1.2 seems to be already proved by López-Blázquez and Salamanca-Miño [9], see their Theorem 3.2. A proof is not included there, but it is claimed that arguments similar to those presented in Lemmas 2.1 and 2.2 are sufficient to derive Theorem 3.2, in the same way that their Theorem 2.3 (Terrell’s result) is proved. The main tool used by the authors, in order to simplify considerably Terrell’s proof, is the inequality $\rho_{12} \leq \rho_{12}'$ with equality if and only if $\sigma_1^2 = \sigma_2^2$, where $\sigma_j^2 = \text{Var}X_j$, $\sigma_j^2 = \text{Var}X_{j:2}$ $(j = 1, 2)$,

$$
\rho_{12} = \rho_{12}(F) = \frac{\sigma_{12}}{(\sigma_1^2 \sigma_2^2)^{1/2}}, \quad \rho_{12}' = \rho_{12}'(F) = \frac{\sigma^2}{(\sigma_1^2 \sigma_2^2)^{1/2}} - 1.
$$

Then, the authors assert that, in order to maximize $\rho_{12}(F)$ over all distributions $F \in \mathcal{F}$, it suffices to consider only those distributions $F \in \mathcal{F}_0 \subset \mathcal{F}$ that satisfy $\sigma_1^2 = \sigma_2^2$. This would be true if we a-priori knew that for each $F \in \mathcal{F}$ there is $F_0 \in \mathcal{F}_0$ with $\rho_{12}(F) \leq \rho_{12}'(F_0) = \rho_{12}(F_0)$. Although this is true in our case (with $F_0$ the uniform over some interval), it does not hold in general. Indeed, it is easily seen that other pairs of positive, symmetric, bounded, functionals, like

$$
w = \frac{5}{6} \left( \frac{\sigma^2}{\sigma_1^2 + 4\sigma_2^2} + \frac{\sigma^2}{4\sigma_1^2 + \sigma_2^2} \right), \quad w' = \frac{5}{6} \left( \frac{\sigma^2}{\sigma_1^2 + 4\sigma_2^2} + \frac{\sigma^2}{4\sigma_1^2 + \sigma_2^2} \right) + \left( 1 - \frac{2(\sigma_1^2 \sigma_2^2)^{1/2}}{\sigma_1^2 + \sigma_2^2} \right)
$$

satisfy $w \leq w'$ with equality if and only if $\sigma_1^2 = \sigma_2^2$, the restricted maximum (under $\sigma_1^2 = \sigma_2^2$) is equal to 1/2, and the maximum is attained uniquely by rectangular distributions; exactly like $\rho_{12}$ and $\rho_{12}'$, above. However, the unrestricted supremum of $w$ is greater than one-half, as the Bernoulli distribution with $p = 9/10$, and some absolutely continuous random variables, show.

\section{2. PROOF OF THEOREM 1.2}

We shall apply the ordinary method of Terrell [18] to the case where $U$ follows a discrete uniform in $P_0 = \{1, \ldots, N\}$. In this case, the orthonormal polynomial system is provided by Hahn polynomials, $\{\psi_k\}_{k=0}^{N-1}$, where

$$
\psi_k(x) = N^{1/2} \binom{N + k}{2k + 1}^{-1/2} \binom{2k}{k}^{-1/2} \sum_{j=0}^{k} (-1)^{k-j} \binom{k+j}{j} \binom{N-1-j}{k-j} \binom{x-1}{j}.
$$

The following two properties will be used in the sequel; cf. [8]. First,

$$
\psi_k(x) = A_k x^k + B_k x^{k-1} + \cdots,
$$

where

$$
A_k = \frac{\sqrt{N}}{k!} \binom{2k}{k}^{1/2} \binom{N+k}{2k+1}^{-1/2},
$$

$$
B_k = -\frac{k(N+1)\sqrt{N}}{k!} \binom{2k-1}{k} \binom{N+k}{2k+1}^{-1/2} \binom{2k}{k}^{-1/2}.
$$

\begin{equation}
(2.1)
\end{equation}
Second, the polynomials $\psi_k$ are orthonormal in $L^2(U) = \{ g : P_0 \to \mathbb{R} \}$, that is,

$$\mathbb{E} \psi_k(U) \psi_m(U) = \frac{1}{N} \sum_{x=1}^{N} \psi_k(x) \psi_m(x) = \delta_{k,m} = \begin{cases} 1, & \text{if } k = m \\ 0, & \text{if } k \neq m, \end{cases}$$

and this implies that any function $g : P_0 \to \mathbb{R}$ has a unique expansion in Hahn polynomial series, namely,

$$g(x) = \sum_{k=0}^{N-1} \delta_k \psi_k(x), \quad \text{where } \delta_k = \mathbb{E} \psi_k(U) g(U) \quad (k = 0, \ldots, N - 1). \tag{2.2}$$

The constants $\{\delta_k\}_{k=0}^{N-1}$ are the Fourier coefficients of $g$.

An arbitrary $\mathcal{P} = \{x_1, \ldots, x_N\}$ (with $x_1 \leq \cdots \leq x_N$) is the image of some nondecreasing $g : P_0 \to \mathbb{R}$, namely, $g(i) = x_i, \ i = 1, \ldots, N$. In other words, if $X$ is uniformly distributed over $\mathcal{P}$ (in the sense of Definition 1.1) and $U$ is uniformly distributed over $P_0$ then $X$ and $g(U)$ have the same distribution. Therefore, $(g(U_1), g(U_2))$ and $(X_1, X_2)$ have the same distribution, where $X_1, X_2$ are independent copies of $X$, and $U_1, U_2$ are independent copies of $U$. Since $g$ is non-decreasing, it holds $(\min\{g(U_1), g(U_2)\}, \max\{g(U_1), g(U_2)\}) = (g(\min\{U_1, U_2\}), g(\max\{U_1, U_2\})) = (g(U_1), g(U_2))$. This shows that the random pairs $(X_1, X_2)$ and $(g(U_1), g(U_2))$ have the same distribution; hence, $\rho_{12} = \rho(g(U_1), g(U_2))$. Expanding $g$ according to (2.2), we will be able to express $\rho_{12}$ as a function of $\delta_1, \ldots, \delta_{N-1}$, and, next, we shall maximize the resulting function. The calculations, below, do not a-priori impose any monotonicity restrictions on $g$; the only assumption is that $g$ is non-constant (otherwise, the variance of $X$ is zero and $\rho_{12}$ is undefined).

**Lemma 2.1** (Hahn representation). Expand an arbitrary $g : P_0 \to \mathbb{R}$ in a Hahn polynomial series as in (2.2), and let $\sigma_1^2 = \text{Var}g(U_1), \sigma_2^2 = \text{Var}g(U_2), \sigma_{12} = \text{Cov}(g(U_1), g(U_2))$. Then,

\begin{align*}
\sigma_1^2 &= \sum_{k=1}^{N-1} \delta_k^2 - \frac{2}{N} \sum_{k=1}^{N-2} \lambda_k \delta_k \delta_{k+1}, \\
\sigma_2^2 &= \sum_{k=1}^{N-1} \delta_k^2 - \frac{2}{N} \sum_{k=1}^{N-2} \lambda_k \delta_k \delta_{k+1}, \\
\sigma_{12} &= \frac{\delta_1^2}{3} \left(1 - \frac{1}{N^2}\right),
\end{align*}

where, in (2.3) and (2.4),

$$\lambda_k = (k + 1) \sqrt{\frac{N^2 - (k + 1)^2}{2(k + 1)(2k + 3)}}, \quad k = 1, \ldots, N - 1. \tag{2.6}$$

**Proof.** If $\delta_1 = \cdots = \delta_{N-1} = 0$ then $g \equiv g_0$ (constant) and the result is trivial. Assume that $\beta := \delta_1^2 + \cdots + \delta_{N-1}^2 > 0$, and set $h(x) := (g(x) - \delta_0)/\sqrt{\beta}$, so that $h(x) = \sum_{k=1}^{N-1} \delta_k \psi_k(x)$ where $\delta_k = \delta_k/\sqrt{\beta}, \ k = 1, \ldots, N - 1$, are the Fourier coefficients of $h$. By construction, $\sum_{k=1}^{N-1} \delta_k^2 = 1$, $\sigma_1^2 = \beta \text{Var}h(U_1), \sigma_2^2 = \beta \text{Var}h(U_2), \sigma_{12} = \beta \text{Cov}(h(U_1), h(U_2))$. The orthonormality of $\psi_k$ shows that the random variable $Y := h(U)$ is standardized: $\mathbb{E}Y = 0, \mathbb{E}Y^2 = 1$.

We now write $Y_1 = h(U_1), Y_2 = h(U_2), Z_1 = h(U_1^2), Z_2 = h(U_2^2)$, with $U_1, U_2$ being independent uniform from $P_0$; notice that the inequality $Z_1 \leq Z_2$ may fail, since $h$ has not been assumed monotonic. According to the argument above, $Y_1, Y_2$ are independent and standardized. Taking expectations to the obvious identities $Z_1 + Z_2 = Y_1 + Y_2$ and $Z_1 Z_2 = Y_1 Y_2$ we see that $\mathbb{E}Z_1 = -\mathbb{E}Z_2$ and $\mathbb{E}Z_1 Z_2 = 0$. Hence, $\text{Cov}(Z_1, Z_2) = (\mathbb{E}Z_2)^2$; equivalently, $\sigma_{12} = \beta (\mathbb{E}Z_2)^2$. The relation $2 = \text{Var}(Y_1 + Y_2) = \text{Var}(Z_1 + Z_2) = \text{Var}Z_1 + \text{Var}Z_2 + 2(\mathbb{E}Z_2)^2$ shows that $\text{Var}Z_1 = 2 - \text{Var}Z_2 - 2(\mathbb{E}Z_2)^2 = 2 - (\mathbb{E}Z_2)^2 - \mathbb{E}Z_2^2$ and $\sigma_1^2 = \beta (2 - (\mathbb{E}Z_2)^2 - \mathbb{E}Z_2^2)$. Since $\text{Var}Z_2 = \mathbb{E}Z_2^2 - (\mathbb{E}Z_2)^2$ implies $\sigma_2^2 = \beta (\mathbb{E}Z_2^2 - (\mathbb{E}Z_2)^2)$, all quantities $\sigma_1^2, \sigma_2^2, \sigma_{12}$ (and $\rho_{12}$), are expressed in terms of $\mathbb{E}Z_2 = \mathbb{E}h(U_2), \mathbb{E}Z_2^2 = \mathbb{E}h(U_2^2)$.
The probability mass function of \( U_{2:2} \) is given by \( \mathbb{P}(U_{2:2} = j) = (2j - 1)/N^2, \ j = 1, \ldots, N, \) and therefore, \( \mathbb{E}Z_2 = (2/N)\mathbb{E}Uh(U) \) because \( \mathbb{E}h(U) = \mathbb{E}Y = 0. \) Since each \( \psi_k(U) \) \((k \geq 2)\) is orthogonal to any polynomial of degree at most one, hence to \( U, \) we have

\[
\mathbb{E}Uh(U) = d_1\mathbb{E}U\psi_1(U) + \sum_{k=2}^{N-1} d_k\mathbb{E}U\psi_k(U) = d_1\mathbb{E}U\psi_1(U) = \frac{d_1}{A_1}\mathbb{E}\psi_1(U)(A_1U + B_1) = \frac{d_1}{A_1},
\]

because \( A_1U + B_1 = \psi_1(U) \) and \( \psi_1 \) is orthogonal to constants. Substituting the value of \( A_1 \) from (2.1), we obtain \( \mathbb{E}Z_2 = 2d_1/(NA_1) = (1-N^{-2})^{1/2}d_1/\sqrt{3}, \) and the relations \( \sigma_{12} = \beta(\mathbb{E}Z_2)^2 \) and \( d_1 = \delta_1/\sqrt{\beta} \) imply (2.5).

The computation of the second moment of \( Z_2 \) is more involved. Write

\[
\mathbb{E}Z_2^2 = \frac{1}{N^2} \sum_{j=1}^{N} (2j-1)h(j)^2 = \frac{1}{N} \mathbb{E}(2U-1)h(U)^2 = \frac{2}{N} \mathbb{E}Uh(U)^2 - \frac{1}{N} \tag{2.7}
\]

because \( \mathbb{E}Y^2 = \mathbb{E}h(U)^2 = 1. \) It remains to compute

\[
\mathbb{E}Uh(U)^2 = \sum_{k,m=1}^{N-1} d_kd_m\mathbb{E}U\psi_k(U)\psi_m(U) = \sum_{k=1}^{N-1} d_k^2\mathbb{E}U\psi_k(U)^2 + 2\sum_{k=1}^{N-2} d_kd_{k+1}\mathbb{E}U\psi_k(U)\psi_{k+1}(U);
\]

all other terms (with \( m > k \)) vanish, due to the orthogonality of \( \psi_m(U) \) to the polynomial \( U\psi_k(U) \) (of degree \( k+1 \)) when \( m \geq k + 2. \) We proceed to compute \( \mathbb{E}U\psi_k(U)^2 \) and \( \mathbb{E}U\psi_k(U)\psi_{k+1}(U). \) Write \( \psi_k(x) = A_kx^k + B_kx^{k-1} + P_{k-2}(x), \) where \( P_{k-2} \) is a polynomial of degree at most \( k-2 \) \((P_{-1} \equiv 0). \) Then,

\[
\mathbb{E}U\psi_k(U)^2 = \mathbb{E}U\psi_k(U)(A_kU^k + B_kU^{k-1} + P_{k-2}(U)) = A_k\mathbb{E}U^{k+1}\psi_k(U) + B_k\mathbb{E}U^k\psi_k(U),
\]

because \( \psi_k(U) \) is orthogonal to the polynomial \( UP_{k-2}(U). \) Next, we calculate

\[
\mathbb{E}U^{k+1}\psi_k(U) = \frac{1}{A_{k+1}}\mathbb{E}\psi_k(U) \left( \psi_{k+1}(U) - B_{k+1}U^k - P_{k-1}(U) \right) = -\frac{B_{k+1}}{A_{k+1}} \mathbb{E}U^k\psi_k(U),
\]

where the last equality follows from the orthogonality of \( \psi_k \) to both \( \psi_{k+1} \) and \( P_{k-1}. \) Therefore,

\[
\mathbb{E}U\psi_k(U)^2 = A_k\mathbb{E}U^{k+1}\psi_k(U) + B_k\mathbb{E}U^k\psi_k(U) = \left( B_k - \frac{A_kB_{k+1}}{A_{k+1}} \right) \mathbb{E}U^k\psi_k(U).
\]

The calculation of \( \mathbb{E}U^k\psi_k(U) \) is easy:

\[
\mathbb{E}U^k\psi_k(U) = \frac{1}{A_k}\mathbb{E}\psi_k(U)(A_kU^k + B_{k-1}U^{k-1} + \cdots) = \frac{1}{A_k} \mathbb{E}\psi_k(U)^2 = \frac{1}{A_k}.
\]

Hence, using the fact that \( B_k/A_k = -k(N+1)/2, \) see (2.1), we obtain \( \mathbb{E}U\psi_k(U)^2 = B_k/A_k - B_{k+1}/A_{k+1} = (N+1)/2; \) surprisingly, this expectation is independent of \( k. \) Proceeding similarly, and in view of (2.1), we have \( \mathbb{E}U\psi_k(U)\psi_{k+1}(U) = \mathbb{E}U(A_kU^k + B_kU^{k-1} + \cdots)\psi_{k+1}(U) = A_k\mathbb{E}U^{k+1}\psi_{k+1}(U) = A_k/A_{k+1} = \lambda_k/2, \) with \( \lambda_k \) as in (2.6). Combining the preceding formulae (recall that \( \sum_{k=1}^{N-1} d_k^2 = 1 \)) we obtain \( \mathbb{E}Uh(U)^2 = (N+1)/2 + \sum_{k=1}^{N-2} \lambda_kd_kd_{k+1}. \) Then, from (2.7), \( \mathbb{E}Z_2^2 = 1 + (2/N)\sum_{k=2}^{N-1} \lambda_kd_kd_{k+1}. \) Moreover, since \( \mathbb{E}(Z_2)^2 = (1-N^{-2})d_1^2/3 \) we have \( \text{Var}Z_2 = 1 - (1-N^{-2})d_1^2/3 + (2/N)\sum_{k=2}^{N-1} \lambda_kd_kd_{k+1}, \) and the relation \( \sigma_2^2 = \beta \text{Var}Z_2 \) yields (2.4), because \( \beta d_1^2 = \sigma_2^2 \) and \( \beta d_kd_{k+1} = \delta_{k,k+1} \) (recall that \( \beta = \sigma_2^2 + \cdots + \sigma_{N-1}^2 \)). Similarly, substituting the preceding expressions for \( \mathbb{E}(Z_2)^2 \) and \( \mathbb{E}Z_2^2 \) to the relation \( \text{Var}Z_1 = 2 - \mathbb{E}(Z_2)^2 - \mathbb{E}Z_2^2 \) we obtain \( \text{Var}Z_1 = 1 - (1-N^{-2})d_1^2/3 - (2/N)\sum_{k=2}^{N-1} \lambda_kd_kd_{k+1}, \) and since \( \sigma_2^2 = \beta \text{Var}Z_1, \) (2.3) follows.

**Lemma 2.2.** We define the rational functions \( R_0(x) \equiv 1 \) and, recurrently,

\[
R_k(x) := 1 - \frac{k^2(1-k^2/x)}{(4k^2-1)R_{k-1}(x)}, \quad k = 1, 2, \ldots
\]
Lemma 2.3. For $N \geq 2$, there exist nonnegative numbers $\{\alpha_k\}_{k=1}^{N-1}$ and $\{\beta_k\}_{k=1}^{N-1}$ satisfying the following four properties:

(i) $\alpha_1 = \frac{1}{3} \left( 2 + \frac{1}{N^2} \right)$, $\beta_{N-1} = 0$;
(ii) $\alpha_k > \beta_{k-1} > 0$, $k = 2, \ldots, N - 1$;
(iii) $\alpha_k + \beta_{k-1} = 1$, $k = 2, \ldots, N - 1$;
(iv) $\alpha_k \beta_k = \frac{(k+1)^2}{(2k+1)(2k+3)} \left( 1 - \frac{(k+1)^2}{N^2} \right)$, $k = 1, \ldots, N - 1$.

**Proof.** Define $\alpha_k := R_k(N^2)$ and $\beta_k := 1 - R_{k+1}(N^2)$, $k = 1, \ldots, N - 1$, where the functions $R_k$ are defined in Lemma 2.2. Then, (i) and (iii) are obvious, and (ii) follows from (2.8), since $\alpha_k = R_k(N^2) > (k + 1)/(2k + 1) > 1/2$, because $N^2 > k^2$, and, by the same reasoning, $\beta_k \leq 0$ unless $k = N - 1$. Finally, (iv) is trivial for $k = N - 1$ (since $\beta_{N-1} = 0$), and for $k = 1, \ldots, N - 2$,

$$
\alpha_k \beta_k = \alpha_k(1 - \alpha_{k+1}) = \frac{(k+1)^2(1 - (k+1)^2/N^2)}{4(k+1)^2 - 1} = \frac{(k+1)^2}{(2k+1)(2k+3)} \left( 1 - \frac{(k+1)^2}{N^2} \right)
$$

due to the recurrent relation of Lemma 2.2. Hence, Lemma 2.3 is proved.

**Lemma 2.4** (Terrell–Hahn representation). Consider the numbers $\{\alpha_k\}_{k=1}^{N-1}$ and $\{\beta_k\}_{k=1}^{N-1}$ as in Lemma 2.3. Then, with the convention $\delta_N := 0$,

$$
\sigma_1^2 = \sum_{k=1}^{N-1} \left( \sqrt{\alpha_k} \delta_k - \sqrt{\beta_k} \delta_{k+1} \right)^2, \quad \sigma_2^2 = \sum_{k=1}^{N-1} \left( \sqrt{\alpha_k} \delta_k + \sqrt{\beta_k} \delta_{k+1} \right)^2,
$$

where $\sigma_1^2 = \text{Var}(U_{1:2})$ and $\sigma_2^2 = \text{Var}(U_{2:2})$ are as in Lemma 2.1.

**Proof.** Let $s_i^2$ be the first sum. Expanding the squares we have

$$
s_1^2 = \sum_{k=1}^{N-1} \alpha_k \delta_k^2 + \sum_{k=1}^{N-1} \beta_k \delta_{k+1}^2 + 2 \sum_{k=1}^{N-1} \sqrt{\alpha_k} \beta_k \delta_k \delta_{k+1} = \alpha_1 \delta_1^2 + \sum_{k=2}^{N-1} (\alpha_k + \beta_{k-1}) \delta_k^2 - 2 \sum_{k=1}^{N-2} \sqrt{\alpha_k} \beta_k \delta_k \delta_{k+1},
$$

MATHEMATICAL METHODS OF STATISTICS Vol. 32 No. 2 2023
and Lemma 2.3 implies that

\[
\sigma_{1}^{2} = \frac{1}{3} \left( 2 + \frac{1}{N^2} \right) \delta_{1}^{2} + \sum_{k=2}^{N-1} \delta_{k}^{2} - \frac{2}{N} \sum_{k=1}^{N-2} \lambda_{k} \delta_{k} \delta_{k+1}
\]

with \( \lambda_{k} \) as in (2.6). Observing that \((2 + N^{-2})/3 = 1 - (1 - N^{-2})/3\), \( \sigma_{1}^{2} \) reduces to the expression of \( \sigma_{2}^{2} \), given in (2.3). The derivation of the expression for \( \sigma_{2}^{2} \) is completely similar.

**Theorem 2.1.** For \( N \geq 2 \) and any non-constant \( g : \mathcal{P}_{0} = \{1, \ldots, N\} \rightarrow \mathbb{R} \),

\[
\rho(g(U_{1:2}), g(U_{2:2})) \leq \frac{1 - N^{-2}}{2 + N^{-2}},
\]

with equality if and only if \( g \) is linear.

**Proof.** Let \( g(U) = \sum_{k=0}^{N-1} \delta_{k} \psi_{k}(U) \) be the Hahn representation of \( g \), with \( \delta_{2}^{2} + \cdots + \delta_{N-1}^{2} > 0 \) (otherwise, \( g \) is constant). If \( \sigma_{i}^{2} = \text{Var}(g(U_{i:2}) \ (i = 1, 2) \) and \( \sigma_{12} = \text{Cov}(g(U_{1:2}), g(U_{2:2})) \) then, from Lemma 2.4 (and with the convention \( \delta_{N} = 0 \)),

\[
\sigma_{1}^{2} \sigma_{2}^{2} = \left\{ \sum_{k=1}^{N-1} (\sqrt{\alpha_{k}} \delta_{k} - \sqrt{\beta_{k}} \delta_{k+1})^{2} \right\} \left\{ \sum_{k=1}^{N-1} (\sqrt{\alpha_{k}} \delta_{k} + \sqrt{\beta_{k}} \delta_{k+1})^{2} \right\}
\]

\[
\geq \left\{ \sum_{k=1}^{N-1} (\sqrt{\alpha_{k}} \delta_{k} - \sqrt{\beta_{k}} \delta_{k+1}) \left( \sqrt{\alpha_{k}} \delta_{k} + \sqrt{\beta_{k}} \delta_{k+1} \right) \right\}^{2}
\]

\[
= \left\{ \sum_{k=1}^{N-1} (\alpha_{k} \delta_{k}^{2} - \beta_{k} \delta_{k+1}^{2}) \right\}^{2} = \left\{ \alpha_{1} \delta_{1}^{2} + \sum_{k=2}^{N-1} (\alpha_{k} - \beta_{k-1}) \delta_{k}^{2} \right\}^{2} \geq \alpha_{1} \delta_{1}^{4},
\]

where the second inequality holds true because \( \alpha_{k} - \beta_{k-1} > 0 \) and \( \alpha_{1} > 0 \), see Lemma 2.3(ii), while the first one is a simple application of the well-known Cauchy–Schwarz inequality. On substituting the value of \( \alpha_{1} \) from Lemma 2.3(i), the previous argument implies the inequality

\[
(\sigma_{1}^{2} \sigma_{2}^{2})^{1/2} \geq \frac{1}{3} \left( 2 + \frac{1}{N^2} \right) \delta_{1}^{2},
\]

in which the equality holds if and only if \( \delta_{k} = 0 \) for all \( k \geq 2 \), i.e., when \( g \) is linear. From (2.5) of Lemma 2.1, \( \sigma_{1}^{2} = 3 \sigma_{12}/(1 - N^{-2}) \); thus, \( (\sigma_{1}^{2} \sigma_{2}^{2})^{1/2} \geq \sigma_{12}(2 + N^{-1})/(1 - N^{-2}) \), and this is equivalent to the desired inequality.

Applying Theorem 2.1 to the particular case where \( g \) is given by \( g(i) = x_{i}, \ i = 1, \ldots, N \) (which is nondecreasing and non-constant), we conclude the result of Theorem 1.2.

3. CONCLUDING REMARKS

**Remark 3.1** (An improvement to the covariance/expectation bound for \( n = 2 \)). Since the variance of \( X = g(U) \) equals to \( \sigma^{2} = \delta_{2}^{2} + \cdots + \delta_{N-1}^{2} \geq \delta_{1}^{2} \), an immediate consequence of (2.5) is the Bessel-type inequality \( \text{Cov}(X_{1:2}, X_{2:2}) \leq (1 - N^{-2})/3 \); this is similar to the bound \( \text{Cov}(X_{1:2}, X_{2:2}) \leq \sigma^{2}(1 - 2N^{-1})/(3(1 - N^{-1})) \), obtained by Balakrishnan et al. [2] in the without-replacement case. Notice that the equality in any of these two bounds is attained only for lattice populations. Both bounds converge to \( \sigma^{2}/3 \) as \( N \rightarrow \infty \), and the maximizing population (provided \( \mu = \mathbb{E}X \) and \( \sigma^{2} \) are held fixed), converges weakly to the uniform \( U(\mu - \sigma \sqrt{3}, \mu + \sigma \sqrt{3}) \) distribution. Papathanasiou [15] proved that the inequality \( \text{Cov}(X_{1:2}, X_{2:2}) \leq \sigma^{2}/3 \) holds true for random samples \( X_{1}, X_{2} \) from any distribution with finite variance \( \sigma^{2} \), and the equality characterizes the location-family of rectangular distributions with fixed variance. Later on, Balakrishnan and Balasubramanian [1] observed that, in view of the obvious identity \( \text{Cov}(X_{1:2}, X_{2:2}) = (\mathbb{E}X_{2:2} - \mu)^{2} \), the supremum of this covariance equals to the smallest upper bound for \( \mathbb{E}X_{2:2} - \mu \) in terms of \( \sigma \). The form of the latter bound, called HDG bound, is long known for \( \mathbb{E}X_{n:n} \) (any \( n \geq 2 \)), and it is proved in two classical papers (appeared in consecutive pages of the same
Clearly, this is equivalent to Papathanasiou’s covariance bound. The improvement on the HDG bound for uniform populations (with \( N \) elements, in the sense of Definition 1.1) is an immediate consequence of the discrete covariance bound. More precisely,

\[
\mathbb{E}X_{2,2} = \mu + (\mathbb{E}X_{2,2} - \mu) = \mu + \text{Cov}(X_{1,2}, X_{2,2})^{1/2} \leq \mu + (1 - N^{-2})^{1/2} \frac{\sigma}{\sqrt{3}},
\]

and the equality characterizes the lattice population with mean \( \mu \) and variance \( \sigma^2 \), that is, \( x_k = \mu + \sigma \sqrt{3}(2k - N - 1)/\sqrt{N^2 - 1}, k = 1, \ldots, N \). These bounds are in accordance to Theorem 3.1 of [9].

**Remark 3.2** (Connection to the continuous case). One may define the functional \( \rho_{12}(\cdot) \) with domain \( \mathcal{F} := \{ F : F \text{ is a distribution function with finite non-zero variance} \} \) as

\[
\rho_{12}(F) := \rho(X_{1,2}, X_{2,2}), \quad X_1, X_2 \text{ independent with distribution } F,
\]

where \( X_{1,2} \leq X_{2,2} \) are the order statistics. For \( N \geq 2 \), we denote by \( \mathcal{F}_N := \{ P : P = \{ x_1, \ldots, x_N \} \text{ is any uniform population with at least two distinct elements} \} \) (uniform in the sense of Definition 1.1). Then it can be verified that \( \mathcal{F}_0 := \cup_{N \geq 2} \mathcal{F}_N \) is dense in \( \mathcal{F} \) with respect to \( \mathcal{W}_2 \), the Wasserstein distance of order two, defined as \( \mathcal{W}_2(X, Y) = \inf \mathbb{E}(X - Y)^2 \), where the infimum is taken over all couplings of \( X \) and \( Y \). Under the generated topology, the functional \( \rho_{12} \) is continuous, and Theorem 2.1 shows that \( \sup_{F \in \mathcal{F}_0} \rho_{12}(F) = 1/2 \); hence, the same is true for \( \sup_{F \in \mathcal{F}} \), and this implies the bound in Terrell’s result, Theorem 1.1.

**Remark 3.3** (An equivalent minimization problem). To simplify notation, set \( \gamma_k = 2\lambda_k/N \) (\( k = 1, \ldots, N - 1 \)) with \( \lambda_k \) as in (2.6), and \( \lambda = \alpha_1 \) as in Lemma 2.3(i); that is,

\[
\lambda = \frac{1}{3} \left( 2 + \frac{1}{N^2} \right), \quad \gamma_k = \frac{2(k + 1)}{N} \sqrt{\frac{N^2 - (k + 1)^2}{(2k + 1)(2k + 3)}}, \quad k = 1, \ldots, N - 1.
\]

According to Lemma 2.1, we have

\[
\left( \frac{1}{\rho_{12}} \right)^2 = \frac{9N^4}{(N^2 - 1)^2} \left( \lambda \delta_1^2 + \frac{\delta_1^2 + \cdots + \delta_{N-1}^2}{\delta_1^2} \right)^2 - \frac{(\gamma_1 \delta_1 \delta_2 + \gamma_2 \delta_2 \delta_3 + \cdots + \gamma_{N-2} \delta_{N-2} \delta_{N-1})^2}{\delta_1^2},
\]

where this expression should be treated as \( +\infty \) for \( \delta_1 = 0 \). The main difficulty in the proof of Theorem 2.1 was to minimize the above ratio w.r.t. \((\delta_1, \ldots, \delta_{N-1}) \in \mathbb{R}^{N-1} \setminus \{0\} \), and the main result showed that the minimizing points are coincide with the axis \( \{ (\delta_1, 0, \ldots, 0), \delta_1 \neq 0 \} \). The ratio is scale invariant, so, assuming \( \delta_1 \neq 0 \) (since \( \delta_1 = 0 \) makes the ratio infinite, and, certainly, cannot be a minimizer) and setting \( x_k = \delta_k/\delta_1 \) and \( n = N - 1 \) (and removing the constant scalar) we are led to

\[
f(x_2, \ldots, x_n) := \left( \lambda + x_2^2 + \cdots + x_n^2 \right)^2 - (\gamma_1 x_2 + \gamma_2 x_2 x_3 + \cdots + \gamma_{n-1} x_{n-1} x_n)^2.
\]

Consequently, it remains to minimize the (four degree polynomial) function \( f \) and to prove that its global minimum is uniquely attained at \( \textbf{0} \), so that \( f \geq f(\textbf{0}) = \lambda^2 \). Despite the fact that this (unrestricted) minimization problem looks like a simple exercise, this is not so; the minimization points (if exist) depend on the parameters \( \lambda, \gamma_1, \ldots, \gamma_{n-1} \). To see this, set \( n = 3, \lambda_0 = 11/16 \) (since \( \lambda = \lambda_0 \) when \( N = 4 \) \( n = 3 \)), and write \( f(x_2, x_3; \gamma_1, \gamma_2) = (\lambda_0 + x_2^2 + x_3^2)^2 - (\gamma_1 x_2 + \gamma_2 x_2 x_3)^2 \). It can be checked that the function \( f(x_2, x_3; 4/3, 2/3) \), though positive, is globally minimized at the points \((x_2, x_3)\) with

\[
x_2 = \pm \left( 66\sqrt{6} - 81 \right)^{1/2} / 16, \quad x_3 = (3\sqrt{6} - 5) / 16
\]

not at \( \textbf{0} \). Also, \( f(\cdot; 1, 2) \) is not bounded below; the minimum value of \( f(\cdot; 1, 8/5) \) is negative; \( f(\cdot; 1, 1) \) attains its global minimum at \( \textbf{0} \). These examples (in the simplest case \( n = 3 \)) indicate that the multi-parameter case is too complicated, to enable an explicit solution. In the contrary, the proposed Terrell-type method of proof succeed in obtaining Theorems 1.2, 2.1. Motivated from these examples, a natural question of mathematical nature would ask for (necessary and) sufficient conditions on the arbitrary positive parameters \( \lambda, \gamma_1, \ldots, \gamma_{n-1} \), guaranteeing that the minimum of \( f(x_2, \ldots, x_n) \), above, is uniquely attained at \( \textbf{0} \).
Remark 3.4 (Connection to maximal correlation). It is important to make clear that the assertion of
Theorem 2.1 cannot be reduced to a maximal correlation problem (unless \( N = 2 \)), in contrast to all of
the existing results mentioned in Section 1. To see this, it is useful to recall Rényi’s [16] characterization
of the maximal correlation of \((X, Y)\), namely,

\[
R(X, Y)^2 = \sup_f \mathbb{E} \left[ (\mathbb{E} f(X) | Y)^2 \right],
\]

where the supremum is taken over \( f \) with \( \mathbb{E} f(X) = 0, \mathbb{E} f(X)^2 = 1 \). Applying Rényi’s result to \((X, Y) = (U_{1:2}, U_{2:2})\), and writing \( f(k) = x_k, k = 1, \ldots, N \), it is seen that

\[
R^2 := R(U_{1:2}, U_{2:2})^2 = \frac{1}{N^2} \max \sum_{j=1}^{N} \frac{1}{2j - 1} \left\{ 2(x_1 + \cdots + x_{j-1}) + x_j \right\}^2,
\]

where the maximum is taken over \( x_1, \ldots, x_N \) satisfying

\[
\sum_{k=1}^{N} (2N - 2k + 1)x_k = 0, \quad \sum_{k=1}^{N} (2N - 2k + 1)x_k^2 = N^2.
\]

For \( N = 3 \) one finds a maximization point explicitly, namely,

\[
x_1^* = -\sqrt{\frac{2}{5} + \frac{13}{10\sqrt{19}}}, \quad x_2^* = \sqrt{1 - \frac{2}{\sqrt{19}}}, \quad x_3^* = \sqrt{4 - \frac{1}{2\sqrt{19}}},
\]

yielding \( R = (2 + \sqrt{19})/15 > (1 - N^{-2})/(2 + N^{-2}) = 8/19 \). Setting \( f(k) = x_k^* \) and \( g(Y) = (1/R)\mathbb{E}(f(X)|Y) \), that is, \( g(k) = y_k^* \) with

\[
y_1^* = -\sqrt{4 - \frac{1}{2\sqrt{19}}}, \quad y_2^* = -\sqrt{1 - \frac{2}{\sqrt{19}}}, \quad y_3^* = \sqrt{\frac{2}{5} + \frac{13}{10\sqrt{19}}},
\]

it is easy to check that the pair \((f, g)\) is maximally correlated: \( \mathbb{E} f(U_{1:2}) = \mathbb{E} g(U_{2:2}) = 0, \mathbb{E} f(U_{1:2})^2 = \mathbb{E} g(U_{2:2})^2 = 1 \), and \( \rho(f(U_{1:2}), g(U_{2:2})) = \mathbb{E} f(U_{1:2}) g(U_{2:2}) = (2 + \sqrt{19})/15 \).

The situation is similar for all \( N \), namely, it can be shown that \( \rho_N := (1 - N^{-2})/(2 + N^{-2}) < R_N < 1/2, N \geq 3 \), where \( R_N \) is the maximal correlation of an ordered pair from the uniform population, \( \mathcal{P}_0 \), of size \( N \). Clearly, for \( N \geq 4 \) it would be quite surprising if \( R_N \) could be expressed in a closed form. An advanced, effective, algorithm producing numerical values of \( R_N \), is recently proposed in [11]. Since the maximal correlation is attainable in our case (the corresponding operator is compact; see [16], the TSM inequality (1.1) implies that \( R_N < 1/2 \). Certainly, \( R_N \geq \rho_N \). In order to prove that \( R_N \geq \rho_N \), it suffices to find a pair of functions with coefficient of correlation greater than \( \rho_N \). Testing small perturbations of linear functions, we were led to define \( f_0(x) := x - 3x^2/N^3, g_0(y) := y + 3y^2/N^3 \) (both strictly increasing). Set \( \rho_0 = \rho(f_0(U_{1:2}), g_0(U_{2:2})) = \sigma_1^2 = \text{Var}f_0(U_{1:2}), \sigma_2^2 = \text{Var}g_0(U_{2:2}) \). For \( N \geq 3 \) we computed (with the help of Mathematica)

\[
\frac{60^2 \pi^2 \sigma_2^2 N^{14} (2N^2 + 1)^2}{(N^2 - 1)^2 (N^2 - 4)^2} (\rho_0^2 - \rho_N^2) = -666 - 1332N + 846N^2 + 2952N^3 + 1077N^4
\]
\[
- 828N^5 - 934N^6 - 792N^7 - 485N^8 + 34N^{10} + 20N^{12}.
\]

The substitution \( N = n + 3 (n \geq 0) \) shows that the right-hand side, \( w(n) \), is strictly positive:

\[
w(n) = 7010100 + 35183016n + 72768816n^2 + 86119956n^3 + 66523137n^4 + 35823456n^5
\]
\[
+ 13910474n^6 + 3946848n^7 + 815185n^8 + 119820n^9 + 11914n^{10} + 720n^{11} + 20n^{12}.
\]

Hence, \( R_N \geq \rho_0 > \rho_N \) for every \( N \geq 3 \).
The main result of the present note, although quite specialized, gives rise to a number of questions that, at least to author’s view, are of some interest. Let $U_{1:n} \leq \cdots \leq U_{n:n}$ be the ordered sample from the uniform population $P_0 = \{1, \ldots, N\}$. For $1 \leq i < j \leq n$ we define
\[ R_N = R_N(i, j : n) := \sup_{f,g} \rho(f(U_{i:n}), g(U_{j:n})), \quad R'_N = R'_N(i, j : n) := \sup_g \rho(g(U_{i:n}), g(U_{j:n})), \]
where the suprema are taken over non-constant $f, g$ (with domain $P_0$). The first one is the well-known maximal correlation; the second appeared in Theorem 2.1 for $i = 1, j = 2, n = 2$. Moreover, define
\[ \rho_N = \rho_N(i, j : n) := \rho(U_{i:n}, U_{j:n}), \quad R''_N = R''_N(i, j : n) := \sup_{g \neq f} \rho(g(U_{i:n}), g(U_{j:n})), \]
where the supremum is taken over nondecreasing non-constant $g$. It is obvious that for all $i, j, n, N, 0 < \rho_N \leq R''_N \leq R'_N \leq R_N < [i(n + 1 - j)]^{1/2}[j(n + 1 - i)]^{-1/2}$; see (1.1). It would be of interest to calculate these quantities for all choices of $i, j, n, N$, and, especially, to obtain the population $P_0$ (i.e.,
the function $g$ in $R''_N$) that dominates the coefficient of correlation of order statistics from any other population $P = \{x_1, \ldots, x_N\}$. In the continuous case (as well as the without-replacement one), the extremal population is the (lattice) uniform population; however, the computations in [11] show that the lattice population in not extremal in general.

ACKNOWLEDGMENTS

The author would like to cordially thank an anonymous Referee for careful reading of the manuscript, and for the correction of some mistakes and typos. Thanks are also due to Fernando López-Blázquez, Universidad de Sevilla, for useful discussions.

CONFLICT OF INTEREST

The author declares that he has no conflicts of interest.

REFERENCES

1. N. Balakrishnan and K. Balasubramanian, “Equivalence of Hartley–David–Gumbel and Papathanasiou bounds and some further remarks,” Statist. Probab. Lett. 16, 39–41 (1993).
2. N. Balakrishnan, C. Charalambides, and N. Papadatos, “Bounds on expectation of order statistics from a finite population,” J. Statist. Plann. Inference 113, 569–588 (2003).
3. A. Castaño-Martínez, F. López-Blázquez, and B. Salamanca-Miño, Maximal Correlation between Order Statistics, in: Recent Developments in Ordered Random Variables, Ed. by M. Ahsanullah and M. Raqab (Nova Science Publishers, 2007), p. 55–68.
4. H. Gebelein, “Das Statistische Problem der Korrelation als Variation und Eigenwertproblem und sein Zusammenhang mit der Ausgleichrechnung,” Angew. Math. Mech. 21, 364–379 (1941).
5. E. J. Gumbel, “The maxima of the mean largest value and of the range,” Ann. Math. Statist. 25, 76–84 (1954).
6. H. O. Hartley and H. A. David, “Universal bounds for mean range and extreme observations,” Ann. Math. Statist. 25, 85–99 (1954).
7. F. López-Blázquez, “Caracterización de Distribuciones mediante el Valor Esperado de Estadísticos Ordenados y Records,” PhD. Thesis, Universidad de Sevilla (1990).
8. F. López-Blázquez and A. Castaño-Martínez, “Upper and lower bounds for the correlation ratio of order statistics from a sample without replacement,” J. Statist. Plann. Inference 136, 43–52 (2006).
9. F. López-Blázquez and B. Salamanca-Miño, “On Terrerl’s characterization of uniform distribution,” Statist. Papers 40, 335–342 (1999).
10. F. López-Blázquez and B. Salamanca-Miño, “Maximal correlation in a non-diagonal case,” J. Multivar. Anal. 131, 265–278 (2014).
11. F. López-Blázquez and B. Salamanca-Miño, “Automatic differentiation and maximal correlation of order statistics from discrete parents,” Comput. Statist. 36, 2889–2915 (2021).
12. V. B. Nevzorov, “A characterization of exponential distributions by correlations between records,” Math. Meth. Statist. 1, 49–54 (1992).
13. N. Papadatos, “Some counterexamples concerning maximal correlation and linear regression,” J. Multivariate Anal. 126, 114–117 (2014).
14. N. Papadatos, and T. Xifara, “A simple method for obtaining the maximal correlation coefficient and related characterizations,” J. Multivariate Anal. 118, 102–114 (2013).
15. V. Papathanasiou, “Some characterizations of distributions based on order statistics,” Statist. Probab. Lett. 9, 145–147 (1990).
16. A. Rényi, “On measures of dependence,” Acta Math. Acad. Sci. Hungar. 10, 441–451 (1959).
17. G. J. Székely and T. F. Móri, “An extremal property of rectangular distributions,” Statist. Probab. Lett. 3, 107–109 (1985).
18. G. R. Terrell, “A characterization of rectangular distributions,” Ann. Probab. 11(3), 823–826 (1983).