THE AMERICAN STATISTICIAN
2021, VOL. 75, NO. 2, 226–229: Teacher's Corner
https://doi.org/10.1080/00031305.2020.1831956

Taylor & Francis
Taylor & Francis Group

Further Examples Related to Correlations Between Variables and Ranks

Chunming Zhang

Department of Statistics, University of Wisconsin – Madison, Madison, WI

ABSTRACT

Rank statistics $\{R_1,\dots,R_n\}$ of actual variates $\{X_1,\dots,X_n\}$ play an important role in university undergraduate nonparametric statistics courses. This article derives explicit expressions of the correlation coefficients between X_i and R_j for not only $i=j$ but also $i\neq j$, for iid continuous variables $X_1,\dots,X_n$ with a distribution function $F_X(\cdot)$ of X and $n\geq 2$: (a) $\rho_{X_i,R_i}=\sqrt{\frac{n-1}{n+1}}\,\rho_{X,F_X(X)}\in(0,\sqrt{\frac{n-1}{n+1}}]$ for any i, revealing that the correlation can be as close to one as expected, while may also unexpectedly decrease approaching zero for other distributions of X; (b) $\rho_{X_i,R_j}=-\frac{1}{n-1}\rho_{X_i,R_i}\in[-\frac{1}{\sqrt{n^2-1}},0)$ for any $i\neq j$, inferring a negligible negative association with ranks from other data; (c) the partial correlation coefficient between X_i and R_i on X_j for any $i\neq j$ equals $\rho_{(X_i,R_i)\bullet X_j}=\rho_{X_i,R_i}/\sqrt{1-\rho^2_{X_j,R_i}}\in(\rho_{X_i,R_i},\frac{n-1}{\sqrt{n^2-2}}]$, invariably exceeding ρ_{X_i,R_i}. Implications of the results necessitate more relevant interpretation of ranks in sharing information of data.

ARTICLE HISTORY
Received May 2020
Accepted September 2020

KEYWORDS
Nonparametrics; Order-statistics; Pearson correlation coefficient; Rank information; Sample correlation coefficient

1. Introduction

In nonparametric statistics, the notion of "rank" plays a key role in learning utilities of distribution-free methods in analyzing data, when the information underlying their distributions is lacking or unknown. See Richardson (2019, p. 361) and references therein. Ranks (e.g., $\{R_i\}_{i=1}^n$) which are transformed from original data (e.g., $\{X_i\}_{i=1}^n$) are typically interpreted to extract as much numerical information of and relax as much distributional assumptions on $\{X_i\}$ as possible. Indeed, nonparametric methods are developed largely from rank statistics together with order-statistics.

It is thus natural to quantify more precisely the direction and magnitude of the association between variables X_i and ranks R_j, regardless of the scale types (continuous or discrete) and location indices (i or j). Some empirical assessment can be made from simulation studies. We simulate N random samples of observations $\{X_1^{(b)},\dots,X_n^{(b)}\}\overset{\text{iid}}{\sim}X$, $b=1,\dots,N$, from a number of commonly used distributions of X, including uniform, Exponential, Gaussian, Laplace, Weibull, mixture of Gaussians, Student's t, F, and log-normal, and denote the ranks within each sample by $\{R_1^{(b)},\dots,R_n^{(b)}\}$. Figure 1 in Appendix A in the supplementary materials displays boxplots of Pearson product-moment sample correlation coefficients $\widehat{\rho}_1^{(b)}$ for $\{(X_i^{(b)},R_i^{(b)}):i=1,\dots,n\}$, $\widehat{\rho}_2^{(b)}$ for $\{(X_i^{(b)},R_{i+1}^{(b)}):i=1,\dots,n-1\}$ and $\widehat{\rho}_3^{(b)}$ for $\{(X_i^{(b)},R_{i-1}^{(b)}):i=2,\dots,n\}$, respectively, $b=1,\dots,N$, with $N=1000$ and $n=1000$. Evidently, both $\widehat{\rho}_2^{(b)}$ and $\widehat{\rho}_3^{(b)}$ are small in magnitude, centered around zero. In contrast, values of $\widehat{\rho}_1^{(b)}$ are invariably strictly positive, and closer to one in most cases than in the other cases (e.g., Weibull$(1,0.5)$ with shape parameter 0.5, $F_{2,6}$, and log-normal). Nonetheless, the boxplot in Figure 2 (right panel) in the supplementary materials exhibits the near zero tendency of $\widehat{\rho}_1^{(b)}$ for the Weibull$(1,k)$ distribution as k decreases to 0, which seems to be unexpected.

For $N\to\infty$, Stuart (1954, eq. (10)) derived the limit of the sample correlation coefficient between $\{X_i^{(b)}:i=1,\dots,n;b=1,\dots,N\}$ and their ranks $\{R_i^{(b)}:i=1,\dots,n;b=1,\dots,N\}$ to be

$$\left\{\frac{12(n-1)}{\operatorname{var}(X)(n+1)}\right\}^{1/2}[\,\mathrm{E}\{XF_X(X)\}-1/2\,\mathrm{E}(X)],\qquad(1)$$

where $F_X(\cdot)$ is the cumulative distribution function (c.d.f.) of X, and computed (1) for uniform, Gaussian and Gamma distributions. Result (1) agrees with the population correlation coefficient ρ_{X_i,R_i} between X_i and R_i, derived in Gibbons and Chakraborti (2003, eq. (5.10), p. 194) via an alternative approach, which directly employed (without prior justification) the independence property between order-statistics and ranks, though not straightforwardly obvious. Likewise, the tools used in Stuart (1954) and Gibbons and Chakraborti (2003) could be difficult to be extended in dealing with other cases such as $\{\widehat{\rho}_2^{(b)}\}$ and $\{\widehat{\rho}_3^{(b)}\}$ in the Monte Carlo study above, that is, difficult to characterize their population analogues ρ_{X_i,R_j} for $i\neq j$.

Propositions 1 and 2 in this article explicitly evaluate the correlation coefficients, ρ_{X_i,R_j}, between X_i and R_j, for not only $i=j$ but also $i\neq j$, in a different, more elementary, and rigorous way, enabling interpretations with broader perspectives. Students in introductory nonparametric statistics courses can easily follow the derivations presented in this article.

CONTACT Chunming Zhang cmzhang@stat.wisc.edu Department of Statistics, University of Wisconsin – Madison, 1300 University Avenue, Madison, WI 53706.
Supplementary materials for this article are available online. Please go to www.tandfonline.com/r/TAS.

(i) Interestingly, ρ_{X_i,R_j}, for all $1 \le i,j \le n$, are proportional to a common correlation coefficient, $\rho_{X,F_X(X)}$, between X and $F_X(X)$. Particularly, for $i = j$, the deduced form $\rho_{X_i,R_i} = \sqrt{\frac{n-1}{n+1}}\,\rho_{X,F_X(X)} \in (0, \sqrt{\frac{n-1}{n+1}}]$ in (14) is equivalent to (1) as expected, while enhances interpretability. For any $i \ne j$, the deduced $\rho_{X_i,R_j} = -\frac{1}{n-1}\rho_{X_i,R_i} \in [-\frac{1}{\sqrt{n^2-1}}, 0)$ in (15) infers a negligible negative impact on ranks from other observations.

(ii) For the Gaussian distribution, we demonstrate that $\rho_{X,F_X(X)} = 0.977$ is directly connected with the celebrated Stein's identity (Stein 1981), which may partly explain the close proximity to one attained by the uniform distribution.

(iii) For the family of Weibull$(1,k)$ distributions with shape parameter k, the left panel of Figure 2 in the supplementary materials plots the analytic form (21) of $\rho_{X,F_X(X)}$, which confirms the empirical observation in the right panel of Figure 2 in the supplementary materials.

(iv) Moreover, for a contaminated sample, from for example, a mixture of Gaussians, containing a proportion of outliers in practical applications, the explicit form of $\rho_{X,F_X(X)}$ and plots in Figure 3 in the supplementary materials depict lower correlations between X_i and R_i from a Gaussian mixture than from a single Gaussian distribution.

(v) An application to the partial correlation coefficient is discussed in (25).

The major derivations of Propositions 1 and 2 appear to be new. The derivations are easy to follow for advanced undergraduate and beginning graduate students and thus will be beneficial in gaining additional insights into and a better understanding of the flexibility and limitations of ranks used in nonparametric statistics. The online supplementary file collects all figures and proofs in the article.

2. Notations, Definitions, and Some Auxiliary Results

The covariance between two random variables X and Y is $\mathrm{cov}(X,Y)$, and the correlation coefficient is $\rho_{X,Y} = \mathrm{cov}(X,Y)/\{\sqrt{\mathrm{var}(X)}\sqrt{\mathrm{var}(Y)}\}$. Two results relevant to succeeding discussions are listed below.

(R1) For X having a location-scale family of distributions, where the c.d.f. is $F_X(x) = F_Z(\frac{x-\mu}{\sigma})$, with a location parameter μ and a scale parameter $\sigma \in (0,\infty)$, it is readily seen that

$$\rho_{X,F_X(X)} = \rho_{Z,F_Z(Z)}, \tag{2}$$

where Z has the c.d.f. $F_Z(\cdot)$.

(R2) For $U \sim \mathrm{Unif}(0,1)$, $F_U(U) = U$, $\mathrm{E}(U) = 1/2$, $\mathrm{var}(U) = 1/12$, and thus

$$\rho_{U,F_U(U)} = 1. \tag{3}$$

For $X_1,\ldots,X_n \overset{\text{iid}}{\sim} X$, where X has the c.d.f. F_X and probability density function (p.d.f.) f_X, with a finite second moment, the variables X_i, order-statistics $X_{(i)}$ and ranks R_i are related according to

$$X_i = X_{(R_i)}, \quad i = 1,\ldots,n, \tag{4}$$

$$X_{(i)} = X_{\Pi_i}, \quad i = 1,\ldots,n, \tag{5}$$

where $\{\Pi_1,\ldots,\Pi_n\}$ is a permutation over $\{1,\ldots,n\}$, illustrated in the diagram below,

$$\boxed{\begin{array}{ccccc} X_1 & \cdots & X_i & \cdots & X_n \\ \| & \cdots & \| & \cdots & \| \\ X_{(R_1)} & \cdots & X_{(R_i)} & \cdots & X_{(R_n)} \end{array}} \text{ namely, } \boxed{\begin{array}{ccccc} X_{\Pi_1} & < \cdots & X_{\Pi_i} & \cdots < & X_{\Pi_n} \\ \| & \cdots & \| & \cdots & \| \\ X_{(1)} & < \cdots & X_{(i)} & \cdots < & X_{(n)} \end{array}}$$

Before proving Propositions 1 and 2, we first list below required results on ranks and order-statistics, among which, basic results (6) and (7) are well-known in nonparametric statistics textbooks (including Daniel 1990; Conover 1999; Higgins 2004; Sprent and Smeeton 2007; Corder and Foreman 2014; Hollander, Wolfe, and Chicken 2014) reviewed in Richardson (2019), results (8)–(12) are nontrivial but more explicit derivations are lacking, and the more advanced statement (13) appears in Lemma 13.1 of van der Vaart (1998) which omits the proof. For concise and complete derivations of Propositions 1 and 2 to be accessible to undergraduate students, Appendix B in the supplementary materials supplies proofs of (8)–(13) using standard uniform random variables, $U_1,\ldots,U_n \overset{\text{iid}}{\sim} \mathrm{Unif}(0,1)$, associated with order-statistics $U_{(1)} \le \cdots \le U_{(n)}$. In the rest of the article, $\mathrm{I}(\cdot)$ denotes an indicator operator.

(R3) The ranks $R_1,\ldots,R_n$ are identically (though not independently) distributed, that is,

$$\begin{aligned} &R_i \sim \mathrm{Unif}\{1,\ldots,n\}, \ \mathrm{E}(R_i) = (n+1)/2, \\ &\mathrm{var}(R_i) = (n^2-1)/12;\ \mathrm{P}(R_i = r, R_j = s) = 1/\{n(n-1)\}, \\ &\text{for } i \ne j \text{ and } r \ne s;\ \mathrm{P}(R_1 = r_1,\ldots,R_n = r_n) = 1/n!, \\ &\text{for any permutation } \{r_1,\ldots,r_n\} \text{ of } \{1,\ldots,n\}. \end{aligned} \tag{6}$$

(R4)

$$\begin{aligned} f_{U_{(1)},\ldots,U_{(n)}}(u_1,\ldots,u_n) &= n!\,\mathrm{I}(0 < u_1 < \cdots < u_n < 1), \\ f_{U_{(k)}}(u) &= \frac{n!}{(k-1)!(n-k)!}u^{k-1}(1-u)^{n-k}\,\mathrm{I}(0 < u < 1), \end{aligned} \tag{7}$$

and satisfying

$$\sum_{k=1}^{n} f_{U_{(k)}}(u) = n, \tag{8}$$

$$\sum_{k=1}^{n} k f_{U_{(k)}}(u) = n(n-1)u + n. \tag{9}$$

(R5)

$$\mathrm{E}\{X_{(k)}\} = \int x \frac{n!}{(k-1)!(n-k)!}\{F_X(x)\}^{k-1} \{1 - F_X(x)\}^{n-k} f_X(x)\,\mathrm{d}x, \tag{10}$$

$$\sum_{k=1}^{n} \mathrm{E}\{X_{(k)}\} = n\,\mathrm{E}(X), \tag{11}$$

$$\sum_{k=1}^{n} k\,\mathrm{E}\{X_{(k)}\} = n(n-1)\,\mathrm{E}\{XF_X(X)\} + n\,\mathrm{E}(X). \tag{12}$$

(R6)

$$(R_1,\ldots,R_n) \text{ is independent of } (X_{(1)},\ldots,X_{(n)}). \tag{13}$$

3. Correlation Coefficient Between Variates and Ranks

3.1. ρ_{X_i,R_i} for $1 \le i \le n$

We first evaluate the correlation coefficient between X_i and R_i. Proposition 1 confirms that ρ_{X_i,R_i} is proportional to $\rho_{X,F_X(X)}$, and bounded below and above by 0 and $\sqrt{(n-1)/(n+1)}$, respectively.

Proposition 1. Let $X_1, \dots, X_n \overset{\text{iid}}{\sim} X$, where X has the c.d.f. F_X and p.d.f. f_X, with mean $\mathrm{E}(X)$ and variance $\mathrm{var}(X) \in (0, \infty)$. Then for $i = 1, \dots, n$ with $n \ge 2$, the correlation coefficient between X_i and R_i is

$$\rho_{X_i,R_i} = \sqrt{\frac{n-1}{n+1}}\, \rho_{X,F_X(X)} \in \Big(0, \sqrt{\frac{n-1}{n+1}}\Big]. \tag{14}$$

As seen from (2) and (3), the upper bound in (14) is achieved for $X \sim \mathrm{Unif}(a, b)$.

3.2. ρ_{X_i,R_j} for $1 \le i \ne j \le n$

For $i \ne j$, Proposition 2 verifies that the correlation coefficient ρ_{X_i,R_j} is negatively proportional to $\rho_{X,F_X(X)}$, and bounded below by $-1/\sqrt{n^2-1}$.

Proposition 2. Let $X_1, \dots, X_n \overset{\text{iid}}{\sim} X$, where X has the c.d.f. F_X and p.d.f. f_X, with mean $\mathrm{E}(X)$ and variance $\mathrm{var}(X) \in (0, \infty)$. Then for $1 \le i \ne j \le n$ with $n \ge 2$, the correlation coefficient between X_i and R_j is

$$\rho_{X_i,R_j} = -\frac{1}{\sqrt{n^2-1}}\rho_{X,F_X(X)} \in \Big[-\frac{1}{\sqrt{n^2-1}}, 0\Big). \tag{15}$$

As seen from (2) and (3), the lower bound in (15) is attained for $X \sim \mathrm{Unif}(a, b)$.

3.3. The Common Quantity $\rho_{X,F_X(X)}$ in Propositions 1 and 2

Recall that ρ_{X_i,R_j} in Propositions 1–2, for any $1 \le i, j \le n$, are proportional to $\rho_{X,F_X(X)}$,

$$\rho_{X,F_X(X)} = \frac{\mathrm{E}\{XF_X(X)\} - \mathrm{E}(X)/2}{\sqrt{\mathrm{var}(X)}\sqrt{1/12}}, \tag{16}$$

where $F_X(X) \sim \mathrm{Unif}(0, 1)$. It is thus natural to compute $\rho_{X,F_X(X)}$ for some commonly used distributions of X. Examples 1–6 explicitly evaluate $\rho_{X,F_X(X)}$, which are also supported by centers of boxplots in Figure 1 in the supplementary materials.

Example 1. For the uniform distribution $\mathrm{Unif}(a, b)$ with $-\infty < a < b < \infty$,

$$\rho_{X,F_X(X)} = \rho_{X,X} = 1 \tag{17}$$

agreeing with a direct calculation using (16).

Example 2. For the Exponential distribution $\mathrm{Exp}(\lambda)$ with $0 < \lambda < \infty$,

$$\rho_{X,F_X(X)} = \sqrt{3}/2 \approx 0.866. \tag{18}$$

Example 3. For the Gaussian distribution $\mathbb{N}(\mu, \sigma^2)$ with $\mu \in \mathbb{R}$ and $\sigma \in (0, \infty)$,

$$\rho_{X,F_X(X)} = \sqrt{3/\pi} \approx 0.977. \tag{19}$$

Example 4. For the Laplace distribution $\mathrm{Laplace}(\mu, \sigma)$ with $\mu \in \mathbb{R}$ and $\sigma \in (0, \infty)$,

$$\rho_{X,F_X(X)} = 3\sqrt{6}/8 \approx 0.9186. \tag{20}$$

Example 5. For the Weibull distribution $\mathrm{Weibull}(\lambda, k)$ with the scale parameter $\lambda \in (0, \infty)$ and shape parameter $k \in (0, \infty)$,

$$\rho_{X,F_X(X)} = \frac{\sqrt{3}(1 - 1/2^{1/k})}{\sqrt{(2^{2/k}/\sqrt{\pi})\Gamma(1/2 + 1/k)/\Gamma(1 + 1/k) - 1}}, \tag{21}$$

which is graphed in Figure 2 (left panel) in the supplementary materials as k varies, where $\Gamma(\cdot)$ denotes the Gamma function. Particularly, $\rho_{X,F_X(X)} = 3\sqrt{3}/(4\sqrt{5}) \approx 0.5809$ for $k = 0.5$. Contrary to many other distributions, $\rho_{X,F_X(X)}$ decreases approaching zero, at the rate $O((1/k)^{1/4}/2^{1/k})$, as k drops from 0.5 to 0.

Example 6. For the mixture distribution of $\mathbb{N}(\mu_1, \sigma_1^2)$ and $\mathbb{N}(\mu_2, \sigma_2^2)$ with proportions p and $1 - p$, where $p \in (0, 1)$, $\mu_1 \in \mathbb{R}, \mu_2 \in \mathbb{R}, \sigma_1 \in (0, \infty)$, and $\sigma_2 \in (0, \infty)$, we can compute $\rho_{X,F_X(X)}$ as in (16), where

$$\begin{aligned}\mathrm{cov}\{X, F_X(X)\} &\\ = p(1-p)\Big[\mu_2\Big\{\Phi\Big(\tfrac{\mu_2-\mu_1}{\sqrt{\sigma_1^2+\sigma_2^2}}\Big) - \tfrac{1}{2}\Big\} + \mu_1\Big\{\Phi\Big(\tfrac{\mu_1-\mu_2}{\sqrt{\sigma_1^2+\sigma_2^2}}\Big) - \tfrac{1}{2}\Big\}\Big] &\\ + \tfrac{p^2\sigma_1+(1-p)^2\sigma_2}{2\sqrt{\pi}} + p(1-p)\sqrt{\sigma_1^2+\sigma_2^2}\,\phi\Big(\tfrac{\mu_1-\mu_2}{\sqrt{\sigma_1^2+\sigma_2^2}}\Big), &\end{aligned} \tag{22}$$

and

$$\mathrm{var}(X) = p(1-p)(\mu_1 - \mu_2)^2 + \{p\sigma_1^2 + (1-p)\sigma_2^2\}. \tag{23}$$

Particularly, if $\mu_1 = \mu_2$, then

$$\rho_{X,F_X(X)} = \sqrt{3/\pi}\,\frac{p^2\sigma_1 + (1-p)^2\sigma_2 + p(1-p)\sqrt{2}\sqrt{\sigma_1^2+\sigma_2^2}}{\sqrt{p\sigma_1^2 + (1-p)\sigma_2^2}}; \tag{24}$$

moreover, if $\sigma_1 = \sigma_2$, then (24) reduces to (19) for a single Gaussian distribution. For two special cases, (i) $\mu_1 = \mu_2$ and $\sigma_1 = k\sigma_2$, and (ii) $\mu_1 - \mu_2 = k\sigma$ and $\sigma_1 = \sigma_2 = \sigma$, plots in Figure 3 in the supplementary materials using $p = 0.8$ and $p = 0.1$ indicate that $\rho_{X,F_X(X)}$ can be as low as 0.6 in case (i) and 0.57 in case (ii) as k varies.

4. Discussion

Propositions 1 and 2 have implications useful for some other aspects. For example, the partial correlation coefficient between X_i and R_i on X_j, for $1 \le i \ne j \le n$ with $n \ge 2$, can be computed from ρ_{X_i,R_i} and ρ_{X_i,R_j} according to $\rho_{(X_i,R_i)\bullet X_j} = \frac{\rho_{X_i,R_i} - \rho_{X_i,X_j}\rho_{X_j,R_i}}{\sqrt{1-\rho_{X_i,X_j}^2}\sqrt{1-\rho_{X_j,R_i}^2}} = \frac{\rho_{X_i,R_i}}{\sqrt{1-\rho_{X_j,R_i}^2}}$, which combined with (14) and (15) implies

$$\rho_{X_i,R_i} < \rho_{(X_i,R_i)\bullet X_j} \le \frac{n-1}{\sqrt{n^2-2}}. \tag{25}$$

Again, as seen from (2) and (3), the upper bound in (25) is achieved for $X \sim \mathrm{Unif}(a, b)$.

It may also be helpful to discuss results with discrete variables, where the distribution of ranks in (6) may not hold due to ties in ranks R_i. As an illustration, Figure 4 in the supplementary materials displays (in a way similar to Figure 1 in the supplementary materials) Pearson sample correlation coefficients using commonly used discrete distributions. In all examples, X_i and R_i are highly correlated, whereas X_i and R_j in the case of $i \neq j$ are nearly uncorrelated. Rigorous derivations are beyond the scope of the current article, and we hope to present in future work.

Supplementary Materials

The online supplementary file collects all figures and proofs in the article.

Acknowledgments

The author thanks the editor, the associate editor, and two anonymous reviewers for helpful comments.

Funding

The work was partially supported by U.S. National Science Foundation grants DMS-2013486 and DMS-1712418, and provided by the University of Wisconsin-Madison Office of the Vice Chancellor for Research and Graduate Education with funding from the Wisconsin Alumni Research Foundation.

References

Conover, W. J. (1999), *Practical Nonparametric Statistics* (3rd ed.), New York: Wiley. [227]

Corder, G. W., and Foreman, D. I. (2014), *Nonparametric Statistics: A Step-by-Step Approach* (2nd ed.), Hoboken, NJ: Wiley. [227]

Daniel, W. W. (1990), *Applied Nonparametric Statistics* (2nd ed.), Pacific Grove, CA: Duxbury. [227]

Gibbons, J. D., and Chakraborti, S. (2003), *Nonparametric Statistical Inference* (4th ed.), New York: Marcel Dekker, Inc. [226]

Higgins, J. J. (2004), *An Introduction to Modern Nonparametric Statistics*, Belmont, CA: Brooks/Cole. [227]

Hollander, M., Wolfe, D. A., and Chicken, E. (2014), *Nonparametric Statistical Methods* (3rd ed.), Hoboken, NJ: Wiley. [227]

Richardson, A. (2019), "A Comparative Review of Nonparametric Statistics Textbooks," *The American Statistician*, 73, 360–366. [226,227]

Sprent, P., and Smeeton, N. C. (2007), *Applied Nonparametric Statistical Methods* (4th ed.), Boca Raton, FL: Chapman & Hall/CRC Press. [227]

Stein, C. M. (1981), "Estimation of the Mean of a Multivariate Normal Distribution," *The Annals of Statistics*, 9, 1135–1151. [227]

Stuart, A. (1954), "The Correlation Between Variate-Values and Ranks in Samples From a Continuous Distribution," *British Journal of Statistical Psychology*, 7, 37–44. [226]

van der Vaart, A. W. (1998), *Asymptotic Statistics*, Cambridge: Cambridge University Press. [227]

Appendix A: Figures in the paper

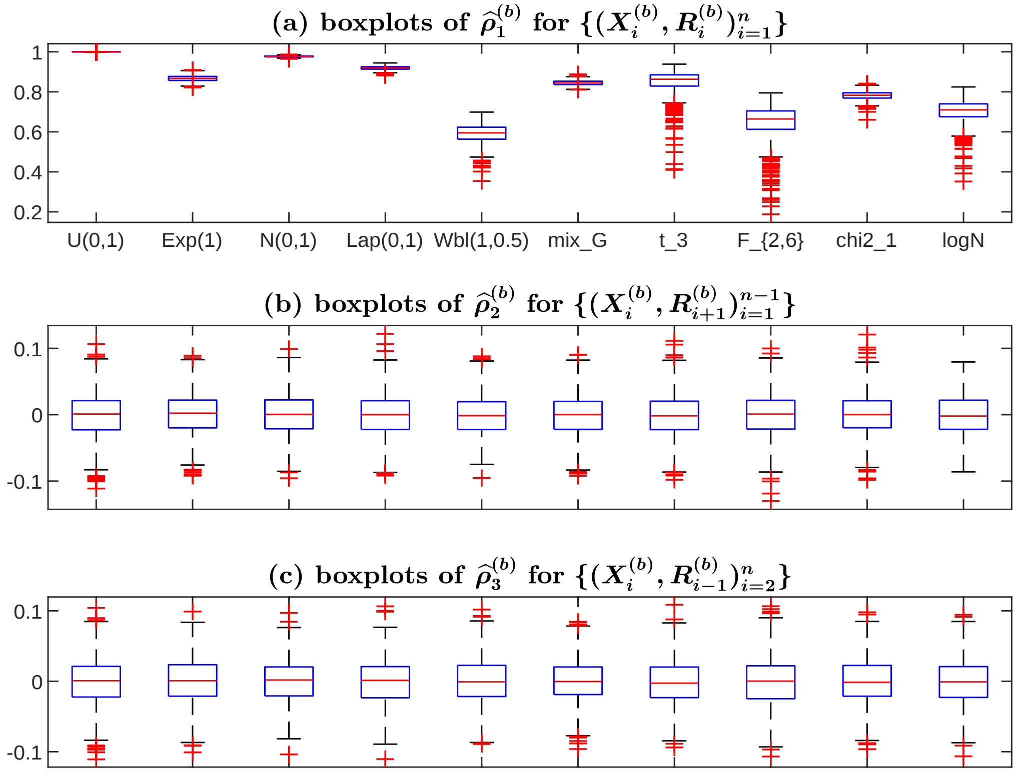

Figure 1: **(Simulation studies for continuous variables)** *Boxplots of* $\widehat{\rho}_1^{(b)}$ *for* $\{(X_i^{(b)}, R_i^{(b)})\}_{i=1}^n$ *in the top panel,* $\widehat{\rho}_2^{(b)}$ *for* $\{(X_i^{(b)}, R_{i+1}^{(b)})\}_{i=1}^{n-1}$ *in the middle panel, and* $\widehat{\rho}_3^{(b)}$ *for* $\{(X_i^{(b)}, R_{i-1}^{(b)})\}_{i=2}^n$ *in the bottom panel,* $b = 1, \ldots, N$*, with* $N = 1000$ *and* $n = 1000$*. Choices of the distribution of* X, Unif$(0,1)$, Exp(1), $\mathbb{N}(0,1)$, Laplace$(0,1)$, Weibull$(1,0.5)$, *mixture Gaussians* $0.8\mathbb{N}(0,1^2)+0.2\mathbb{N}(0,4^2)$, t_3, $F_{2,6}$, χ_1^2, *log-normal, from left to right, are indicated below the boxplot.*

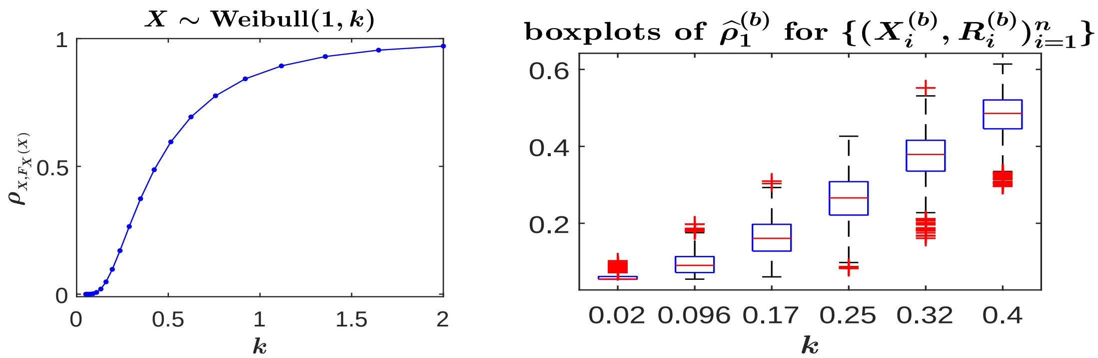

Figure 2: **($X \sim$ Weibull$(1, k)$)** *Left panel: Plot of* (21) *versus* $k > 0$. *Right panel: Boxplots of* $\widehat{\rho}_1^{(b)}$ *for* $\{(X_i^{(b)}, R_i^{(b)})\}_{i=1}^n$, $b = 1, \ldots, N$, *with* $N = 1000$ *and* $n = 1000$, *for* $X \sim \text{Weibull}(1, k)$, *with choices of* k *indicated below the boxplot.*

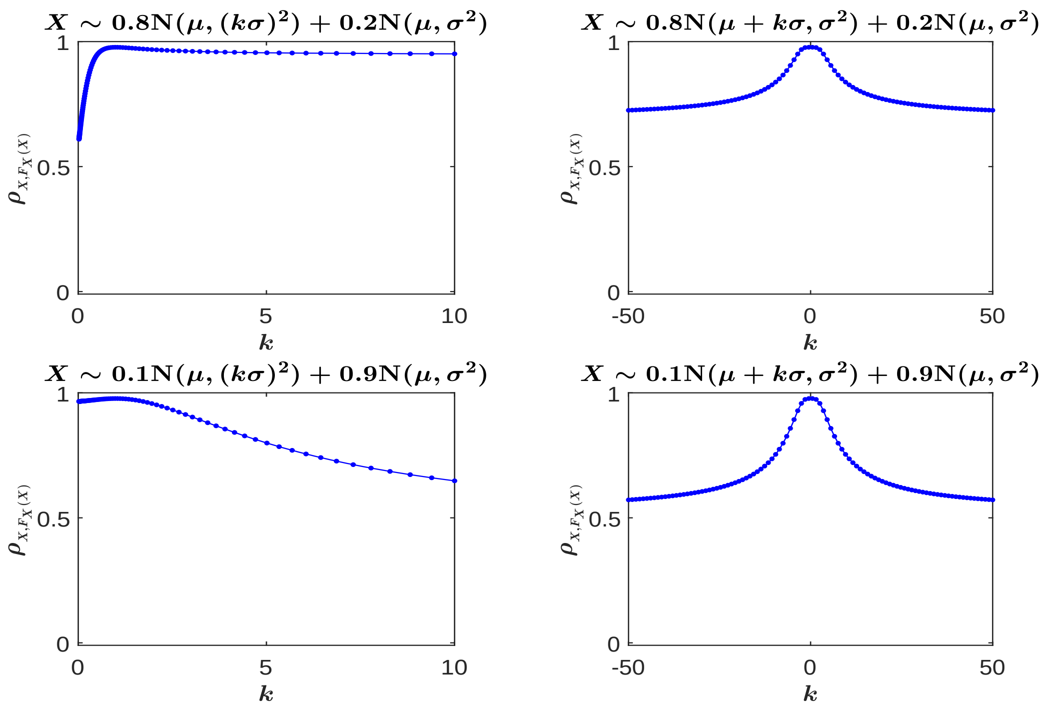

Figure 3: **($X \sim$ mixture Gaussian distribution)** *Left panels: plot of* $\rho_{X,F_X(X)}$ *versus* k, *where* $X \sim p\mathbb{N}(\mu, (k\,\sigma)^2) + (1-p)\mathbb{N}(\mu, \sigma^2)$ *with* $p = 0.8$ *and* $p = 0.1$. *Right panels: plot of* $\rho_{X,F_X(X)}$ *versus* k, *where* $X \sim p\mathbb{N}(\mu + k\,\sigma, \sigma^2) + (1-p)\mathbb{N}(\mu, \sigma^2)$ *with* $p = 0.8$ *and* $p = 0.1$.

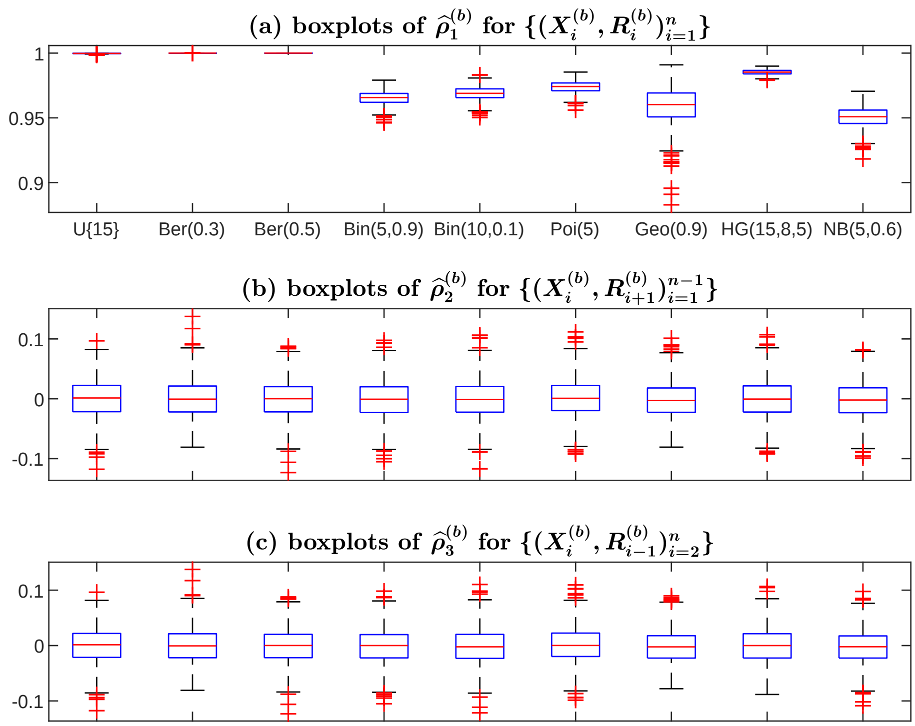

Figure 4: **(Simulation studies for discrete variables)** *Boxplots of* $\widehat{\rho}_1^{(b)}$ *for* $\{(X_i^{(b)}, R_i^{(b)})\}_{i=1}^n$ *in the top panel,* $\widehat{\rho}_2^{(b)}$ *for* $\{(X_i^{(b)}, R_{i+1}^{(b)})\}_{i=1}^{n-1}$ *in the middle panel, and* $\widehat{\rho}_3^{(b)}$ *for* $\{(X_i^{(b)}, R_{i-1}^{(b)})\}_{i=2}^n$ *in the bottom panel,* $b = 1, \ldots, N$*, with* $N = 1000$ *and* $n = 1000$*. Choices of the distribution of* X*, discrete uniform distribution on integers* $\{1, \ldots, 15\}$*, Bernoulli with success probability* 0.3*, Bernoulli with success probability* 0.5*, Binomial with parameters* $(5, 0.9)$*, Binomial with parameters* $(10, 0.1)$*, Poisson with parameter* 5*, Geometric with parameter* 0.9*, Hyper-Geometric with parameters* $(15, 8, 5)$*, Negative-Binomial with parameters* $(5, 0.6)$*, from left to right, are indicated below the boxplot.*

Appendix B: Proofs in the paper

Lemma 1 *Suppose that $F(x)$ and $G(x)$ are similarly ordered on the real line. Then for a random variable X, it follows that* $\operatorname{cov}\{F(X), G(X)\} \ge 0$.

Proof: Our proof is motivated from the Tchebychef's inequality (p. 43 and p. 168 of Hardy *et al.* (1988)), which states that for similarly ordered functions $F(x)$ and $G(x)$ on the interval $\mathcal{I}$, it holds that $|\mathcal{I}| \int_{\mathcal{I}} F(x)G(x)\,\mathrm{d}x \ge \int_{\mathcal{I}} F(x)\,\mathrm{d}x \int_{\mathcal{I}} G(y)\,\mathrm{d}y$.

We expand this inequality as follows. Let $F_X(x)$ be the C.D.F. of X. The fact $\{F(x) - F(y)\}\{G(x) - G(y)\} \ge 0$ for any $x, y \in \mathbb{R}$ implies that $\iint \{F(x) - F(y)\}\{G(x) - G(y)\}\,\mathrm{d}F_X(x)\,\mathrm{d}F_X(y) \ge 0$, in which the double integral can be re-written as

$$\begin{aligned} &= \int F(x)G(x)\,\mathrm{d}F_X(x) \int \mathrm{d}F_X(y) - \int F(x)\,\mathrm{d}F_X(x) \int G(y)\,\mathrm{d}F_X(y) \\ &\quad - \int F(y)\,\mathrm{d}F_X(y) \int G(x)\,\mathrm{d}F_X(x) + \int F(y)G(y)\,\mathrm{d}F_X(y) \int \mathrm{d}F_X(x) \\ &= 2\,\mathrm{E}\{F(X)G(X)\} - 2\,\mathrm{E}\{F(X)\}\,\mathrm{E}\{G(X)\} = 2\operatorname{cov}(F(X), G(X)). \end{aligned}$$

This completes the proof. ■

Proofs of (8) and (9). Result (8) $\sum_{k=1}^{n} f_{U_{(k)}}(u) = n \sum_{k=1}^{n} \frac{(n-1)!}{(k-1)!(n-k)!} u^{k-1}(1-u)^{n-k}$ follows from applying (7) and the Binomial formula.

Similarly, $(k-1) f_{U_{(k)}}(u) = n(n-1)u \frac{(n-2)!}{(k-2)!(n-k)!} u^{k-2}(1-u)^{n-k}$ gives

$$\sum_{k=1}^{n} (k-1) f_{U_{(k)}}(u) = n(n-1)u. \tag{B.1}$$

Thus (9) follows from (B.1) and (8). ■

Proofs of (10), (11) and (12). For $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} F_X$, $X_k = F_X^{-1}(U_k)$, and thus $X_{(k)} = F_X^{-1}(U_{(k)})$. Applying (7) gives

$$\mathrm{E}\{X_{(k)}\} = \mathrm{E}\{F_X^{-1}(U_{(k)})\} = \int_0^1 F_X^{-1}(u) f_{U_{(k)}}(u)\,\mathrm{d}u.$$

By the change of variables $x = F_X^{-1}(u)$, i.e., $u = F_X(x)$, (10) is proved.

Note that

$$\sum_{k=1}^{n} \mathrm{E}\{X_{(k)}\} = \sum_{k=1}^{n} \mathrm{E}(X_k) = n\,\mathrm{E}(X),$$

which verifies (11). Using (9),

$$\begin{aligned}
\sum_{k=1}^{n} k\,\mathrm{E}\{X_{(k)}\} &= \sum_{k=1}^{n} k\,\mathrm{E}\{F_X^{-1}(U_{(k)})\} \\
&= \int_0^1 F_X^{-1}(u) \sum_{k=1}^{n} k f_{U_{(k)}}(u)\,\mathrm{d}u \\
&= \int_0^1 F_X^{-1}(u)\{n(n-1)u + n\}\,\mathrm{d}u \\
&= \int x\{n(n-1)F_X(x) + n\}\,\mathrm{d}F_X(x),
\end{aligned}$$

which proves (12). ■

Proof of (13)**.** Recall that for $X \sim F_X(\cdot)$, $(X_{(1)}, \ldots, X_{(n)}) \overset{\mathrm{D}}{=} (F_X^{-1}(U_{(1)}), \ldots, F_X^{-1}(U_{(n)}))$. It thus suffices to show that for $U_1, \ldots, U_n \overset{\text{i.i.d.}}{\sim} \mathrm{Unif}(0,1)$, the vector of ranks $(R_1, \ldots, R_n)$ is independent of the vector of order-statistics $(U_{(1)}, \ldots, U_{(n)})$.

To show this, let S_n denote the set of all $n!$ permutations of $\{1, \ldots, n\}$. For any $\{r_1, \ldots, r_n\} \in S_n$, and $0 < u_1 < \cdots < u_n < 1$, consider

$$\begin{aligned}
& \lim_{\delta \to 0_+} \mathrm{P}(R_1 = r_1, \ldots, R_n = r_n \mid U_{(1)} \in u_1 \pm \delta/2, \ldots, U_{(n)} \in u_n \pm \delta/2) \\
= & \frac{\lim_{\delta \to 0_+} \mathrm{P}(R_1 = r_1, \ldots, R_n = r_n, U_{(1)} \in u_1 \pm \delta/2, \ldots, U_{(n)} \in u_n \pm \delta/2)/\delta^n}{\lim_{\delta \to 0_+} \mathrm{P}(U_{(1)} \in u_1 \pm \delta/2, \ldots, U_{(n)} \in u_n \pm \delta/2)/\delta^n} \\
= & I_1/I_2,
\end{aligned}$$

where $I_2 = n!$ as in (7). By (4) and (5),

$$\begin{aligned}
I_1 &= \lim_{\delta \to 0_+} \mathrm{P}(\cup_{\{\pi_1,\ldots,\pi_n\} \in S_n}\{U_{\pi_1} < \cdots < U_{\pi_n}, \\
&\quad R_1 = r_1, \ldots, R_n = r_n, U_{(1)} \in u_1 \pm \delta/2, \ldots, U_{(n)} \in u_n \pm \delta/2\})/\delta^n \\
&= \sum_{\{\pi_1,\ldots,\pi_n\} \in S_n} \lim_{\delta \to 0_+} \mathrm{P}(\{U_{\pi_1} < \cdots < U_{\pi_n}, \\
&\quad R_1 = r_1, \ldots, R_n = r_n, U_{\pi_1} \in u_1 \pm \delta/2, \ldots, U_{\pi_n} \in u_n \pm \delta/2\})/\delta^n \\
&= \sum_{\{\pi_1,\ldots,\pi_n\} \in S_n} \lim_{\delta \to 0_+} \mathrm{P}(\{\pi_{r_1} = 1, \ldots, \pi_{r_n} = n, \\
&\quad U_{\pi_1} \in u_1 \pm \delta/2, \ldots, U_{\pi_n} \in u_n \pm \delta/2\})/\delta^n, \\
&= \sum_{\{\pi_1,\ldots,\pi_n\} \in S_n} \mathrm{I}(\pi_{r_1} = 1, \ldots, \pi_{r_n} = n) \\
&\quad \lim_{\delta \to 0_+} \mathrm{P}(U_{\pi_1} \in u_1 \pm \delta/2, \ldots, U_{\pi_n} \in u_n \pm \delta/2)/\delta^n \\
&= \sum_{\{\pi_1,\ldots,\pi_n\} \in S_n} \mathrm{I}(\pi_{r_1} = 1, \ldots, \pi_{r_n} = n) \\
&= 1.
\end{aligned}$$

Thus, $I_1/I_2 = 1/n!$, i.e., the conditional distribution of $(R_1, \ldots, R_n)$ given $(X_{(1)}, \ldots, X_{(n)})$ is identical to the unconditional distribution (6) of $(R_1, \ldots, R_n)$. ■

Proof of Proposition 1. From (4), (13), (6) and (12), we can write

$$\begin{aligned}
\mathrm{E}(X_i R_i) &= \mathrm{E}\{X_{(R_i)} R_i\} \\
&= \sum_{k=1}^{n} \mathrm{E}\{X_{(k)} k \,\mathrm{I}(R_i = k)\} \\
&= \sum_{k=1}^{n} k\, \mathrm{E}\{X_{(k)}\}\, \mathrm{P}(R_i = k) \\
&= \frac{1}{n} \sum_{k=1}^{n} k\, \mathrm{E}\{X_{(k)}\} \\
&= (n-1)\, \mathrm{E}\{X F_X(X)\} + \mathrm{E}(X).
\end{aligned}$$

This, combined with (6) and the fact of $F_X(X) \sim \mathrm{Unif}(0,1)$, gives

$$\begin{aligned}
\mathrm{cov}(X_i, R_i) &= \mathrm{E}(X_i R_i) - \mathrm{E}(X_i)\, \mathrm{E}(R_i) \\
&= (n-1)\, \mathrm{E}\{X F_X(X)\} + \mathrm{E}(X) - \mathrm{E}(X)(n+1)/2 \\
&= (n-1)[\, \mathrm{E}\{X F_X(X)\} - 1/2\, \mathrm{E}(X)] \\
&= (n-1)\mathrm{cov}(X, F_X(X)).
\end{aligned}$$

Also, note that the function $F_X(x)$ is monotone increasing in x. Applying Lemma 1 in Appendix B, we conclude that $\mathrm{cov}(X, F_X(X)) \geq 0$. Moreover, $\mathrm{P}\{(X - \mathrm{E}(X))F_X(X) > 0\} > 0$ indicates $\mathrm{cov}(X, F_X(X)) > 0$, and in turn $\mathrm{cov}(X_i, R_i) > 0$.

Utilizing (6) and $F_X(X) \sim \mathrm{Unif}(0,1)$ again gives

$$\begin{aligned}
\rho_{X_i, R_i} &= \frac{\mathrm{cov}(X_i, R_i)}{\sqrt{\mathrm{var}(X)}\sqrt{(n+1)(n-1)/12}} \\
&= \sqrt{\frac{n-1}{n+1}} \frac{\mathrm{cov}(X, F_X(X))}{\sqrt{\mathrm{var}(X)}\sqrt{1/12}} \\
&= \sqrt{\frac{n-1}{n+1}} \rho_{X, F_X(X)}. \qquad ■
\end{aligned}$$

Proof of Proposition 2. For $1 \leq i \neq j \leq n$, using (4) and (13),

$$\begin{aligned}
\mathrm{E}(X_i R_j) &= \mathrm{E}\{X_{(R_i)} R_j\} \\
&= \sum_{k=1}^{n} \mathrm{E}\{X_{(k)} R_j \,\mathrm{I}(R_i = k)\} \\
&= \sum_{k=1}^{n} \mathrm{E}\{X_{(k)}\}\, \mathrm{E}\{R_j \,\mathrm{I}(R_i = k)\}.
\end{aligned} \tag{B.2}$$

For $\mathrm{E}\{R_j\,\mathrm{I}(R_i = k)\}$ in (B.2), we obtain from (6)

$$\begin{aligned}
\mathrm{E}\{R_j\,\mathrm{I}(R_i = k)\} &= \sum\sum_{1\le r_2\ne r_1\le n} r_2\,\mathrm{I}(r_1 = k)\,\mathrm{P}(R_j = r_2, R_i = r_1) \\
&= \sum_{r_2: r_2\ne k} r_2\,\mathrm{P}(R_j = r_2, R_i = k) \\
&= \frac{1}{n(n-1)} \sum_{r_2: r_2\ne k} r_2 = \frac{1}{n(n-1)}(1 + \cdots + n - k) \\
&= \frac{(n+1)}{2(n-1)} - k\frac{1}{n(n-1)}. \qquad \text{(B.3)}
\end{aligned}$$

Putting (B.3) into (B.2), we obtain

$$\begin{aligned}
\mathrm{E}(X_i R_j) &= \sum_{k=1}^{n} \mathrm{E}\{X_{(k)}\}\Big\{\frac{(n+1)}{2(n-1)} - k\frac{1}{n(n-1)}\Big\} \\
&= \frac{(n+1)}{2(n-1)} \sum_{k=1}^{n} \mathrm{E}\{X_{(k)}\} - \frac{1}{n(n-1)} \sum_{k=1}^{n} k\,\mathrm{E}\{X_{(k)}\} \\
&= \frac{(n+1)}{2(n-1)} n\,\mathrm{E}(X) - \frac{1}{n(n-1)}[n(n-1)\,\mathrm{E}\{XF_X(X)\} + n\,\mathrm{E}(X)] \\
&= \frac{n+2}{2}\,\mathrm{E}(X) - \mathrm{E}\{XF_X(X)\},
\end{aligned}$$

where (11) and (12) are used. It follows that

$$\begin{aligned}
\mathrm{cov}(X_i, R_j) &= \mathrm{E}(X_i R_j) - \mathrm{E}(X_i)\,\mathrm{E}(R_j) \\
&= \frac{n+2}{2}\,\mathrm{E}(X) - \mathrm{E}\{XF_X(X)\} - \frac{n+1}{2}\,\mathrm{E}(X) \\
&= \mathrm{E}(X)/2 - \mathrm{E}\{XF_X(X)\} \\
&= -[\,\mathrm{E}\{XF_X(X)\} - \mathrm{E}(X)\,\mathrm{E}\{F_X(X)\}] \\
&= -\mathrm{cov}(X, F_X(X)). \qquad \text{(B.4)}
\end{aligned}$$

Combining (6) and (B.4) leads to

$$\begin{aligned}
\rho_{X_i, R_j} &= \frac{\mathrm{cov}(X_i, R_j)}{\sqrt{\mathrm{var}(X)}\sqrt{(n^2-1)/12}} \\
&= \frac{-1}{\sqrt{n^2-1}} \frac{\mathrm{cov}(X, F_X(X))}{\sqrt{\mathrm{var}(X)}\sqrt{1/12}} \\
&= -\frac{1}{\sqrt{n^2-1}}\rho_{X, F_X(X)}. \qquad \blacksquare
\end{aligned}$$

Detailed derivations in Section 3.3.

Example 1: From (2), it suffices to consider $X \sim \mathrm{Unif}(0, 1)$, with $F_X(x) = x$. It is immediate to obtain (17). ■

Example 2: From (2), it suffices to consider $X \sim \text{Exp}(1)$. Using $f_X(x) = e^{-x}\,\text{I}(x > 0)$, $F_X(x) = 1 - e^{-x}$, $\text{E}(X) = 1$, and $\text{var}(X) = 1$, we get

$$\text{E}\{XF_X(X)\} = \int_0^{\infty} x(1 - e^{-x})e^{-x}\,\text{d}x = 3/4,$$

and thus (18). ∎

Example 3: From (2), it suffices to consider $X \sim \mathbb{N}(0,1)$. Recall $f_X(x) = \phi(x) = \exp(-x^2/2)/\sqrt{2\pi}$, $F_X(x) = \Phi(x)$, $\text{E}(X) = 0$, and $\text{var}(X) = 1$. By the Stein identity Stein (1981), $\text{E}\{XF_X(X)\} = \text{E}\{Z\Phi(Z)\} = \text{E}\{\Phi'(Z)\} = \text{E}\{\phi(Z)\}$, with $Z \sim \mathbb{N}(0,1)$, where

$$\text{E}\{\phi(Z)\} = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-x^2}\,\text{d}x = 1/(2\sqrt{\pi}). \tag{B.5}$$

Thus, using (16) gives (19). ∎

Example 4: From (2), it suffices to consider $\mu = 0$ and $\sigma = 1$. Recalling $f_X(x) = 2^{-1}e^{-|x|} = 2^{-1}e^{x}\,\text{I}(x \le 0) + 2^{-1}e^{-x}\,\text{I}(x > 0)$, $F_X(x) = 2^{-1}e^{x}\,\text{I}(x \le 0) + (1 - 2^{-1}e^{-x})\,\text{I}(x > 0)$, $\text{E}(X) = 0$, and $\text{var}(X) = 2$, we obtain

$$\begin{aligned}\text{E}\{XF_X(X)\} &= \int_{-\infty}^{0} x\Big(\frac{1}{2}e^{x}\Big)\frac{1}{2}e^{x}\,\text{d}x + \int_0^{+\infty} x\Big(1 - \frac{1}{2}e^{-x}\Big)\frac{1}{2}e^{-x}\,\text{d}x \\ &= 3/8,\end{aligned}$$

and thus (20). ∎

Example 5: From (2), it suffices to consider $\lambda = 1$. We use $f_X(x) = kx^{k-1}e^{-x^k}\,\text{I}(x \ge 0)$, $F_X(x) = 1 - e^{-x^k}$, $\text{E}(X) = \Gamma(1 + 1/k)$, and $\text{var}(X) = \Gamma(1 + 2/k) - \Gamma^2(1 + 1/k)$, to compute

$$\begin{aligned}\text{E}\{XF_X(X)\} &= \int_0^{\infty} x(1 - e^{-x^k})kx^{k-1}e^{-x^k}\,\text{d}x \\ &= (1 - 1/2^{1/k+1})\Gamma(1 + 1/k),\end{aligned}$$

and use (16) to get

$$\begin{aligned}\rho_{X,F_X(X)} &= \frac{(1 - 1/2^{1/k+1})\Gamma(1 + 1/k) - (1/2)\Gamma(1 + 1/k)}{\sqrt{\Gamma(1 + 2/k) - \Gamma^2(1 + 1/k)}\sqrt{1/12}} \\ &= \frac{1/2 - 1/2^{1/k+1}}{\sqrt{\Gamma(1 + 2/k)/\Gamma^2(1 + 1/k) - 1}\sqrt{1/12}},\end{aligned}$$

i.e., (21). ∎

Example 6: In this case, direct calculations give $F_X(x) = p\,\Phi(\frac{x-\mu_1}{\sigma_1}) + (1-p)\Phi(\frac{x-\mu_2}{\sigma_2})$, $f_X(x) = p\frac{1}{\sigma_1}\phi(\frac{x-\mu_1}{\sigma_1}) + (1-p)\frac{1}{\sigma_2}\phi(\frac{x-\mu_2}{\sigma_2})$, $\mathrm{E}(X) = p\mu_1 + (1-p)\mu_2$, $\mathrm{E}(X^2) = \{p\mu_1^2 + (1-p)\mu_2^2\} + \{p\sigma_1^2 + (1-p)\sigma_2^2\}$, and (23). Accordingly,

$$\begin{aligned}
\mathrm{E}\{XF_X(X)\} &= p^2 \int x\,\Phi\Big(\frac{x-\mu_1}{\sigma_1}\Big)\frac{1}{\sigma_1}\phi\Big(\frac{x-\mu_1}{\sigma_1}\Big)\,\mathrm{d}x \\
&\quad +(1-p)^2 \int x\,\Phi\Big(\frac{x-\mu_2}{\sigma_2}\Big)\frac{1}{\sigma_2}\phi\Big(\frac{x-\mu_2}{\sigma_2}\Big)\,\mathrm{d}x \\
&\quad +p(1-p) \int x\,\Phi\Big(\frac{x-\mu_1}{\sigma_1}\Big)\frac{1}{\sigma_2}\phi\Big(\frac{x-\mu_2}{\sigma_2}\Big)\,\mathrm{d}x \\
&\quad +p(1-p) \int x\,\Phi\Big(\frac{x-\mu_2}{\sigma_2}\Big)\frac{1}{\sigma_1}\phi\Big(\frac{x-\mu_1}{\sigma_1}\Big)\,\mathrm{d}x \\
&= I_1 + I_2 + I_3 + I_4,
\end{aligned} \tag{B.6}$$

where

$$\begin{aligned}
I_1 &= p^2\,\mathrm{E}\{(\mu_1 + \sigma_1 Z)\Phi(Z)\} \\
&= p^2[\mu_1\,\mathrm{E}\{\Phi(Z)\} + \sigma_1\,\mathrm{E}\{Z\Phi(Z)\}] \\
&= p^2\Big(\mu_1 \times \frac{1}{2} + \sigma_1\frac{1}{2\sqrt{\pi}}\Big),
\end{aligned}$$

in which (B.5) is used, and similarly,

$$I_2 = (1-p)^2\Big(\mu_2 \times \frac{1}{2} + \sigma_2\frac{1}{2\sqrt{\pi}}\Big).$$

In (B.6),

$$\begin{aligned}
I_3 &= p(1-p)\,\mathrm{E}\Big\{(\mu_2 + \sigma_2 Z)\Phi\Big(\frac{\mu_2-\mu_1}{\sigma_1} + \frac{\sigma_2}{\sigma_1}Z\Big)\Big\} \\
&= p(1-p)\Big[\mu_2\,\mathrm{E}\Big\{\Phi\Big(\frac{\mu_2-\mu_1}{\sigma_1} + \frac{\sigma_2}{\sigma_1}Z\Big)\Big\} + \sigma_2\,\mathrm{E}\Big\{Z\Phi\Big(\frac{\mu_2-\mu_1}{\sigma_1} + \frac{\sigma_2}{\sigma_1}Z\Big)\Big\}\Big] \\
&= p(1-p)\Big\{\mu_2\,\Phi\Big(\frac{\mu_2-\mu_1}{\sqrt{\sigma_1^2+\sigma_2^2}}\Big) + \frac{\sigma_2^2}{\sqrt{\sigma_1^2+\sigma_2^2}}\phi\Big(\frac{\mu_1-\mu_2}{\sqrt{\sigma_1^2+\sigma_2^2}}\Big)\Big\}
\end{aligned}$$

is obtained by calculus, and similarly,

$$I_4 = p(1-p)\Big\{\mu_1\,\Phi\Big(\frac{\mu_1-\mu_2}{\sqrt{\sigma_1^2+\sigma_2^2}}\Big) + \frac{\sigma_1^2}{\sqrt{\sigma_1^2+\sigma_2^2}}\phi\Big(\frac{\mu_2-\mu_1}{\sqrt{\sigma_1^2+\sigma_2^2}}\Big)\Big\}.$$

Hence,

$$\begin{aligned}
&\mathrm{E}\{XF_X(X)\} \\
&= p\mu_1/2 + (1-p)\mu_2/2 \\
&\quad +p(1-p)\Big[\mu_2\Big\{\Phi\Big(\frac{\mu_2-\mu_1}{\sqrt{\sigma_1^2+\sigma_2^2}}\Big) - \frac{1}{2}\Big\} + \mu_1\Big\{\Phi\Big(\frac{\mu_1-\mu_2}{\sqrt{\sigma_1^2+\sigma_2^2}}\Big) - \frac{1}{2}\Big\}\Big] \\
&\quad +\frac{p^2\sigma_1 + (1-p)^2\sigma_2}{2\sqrt{\pi}} + p(1-p)\sqrt{\sigma_1^2+\sigma_2^2}\,\phi\Big(\frac{\mu_1-\mu_2}{\sqrt{\sigma_1^2+\sigma_2^2}}\Big),
\end{aligned}$$

which yields (22).

If $\mu_1 = \mu_2 = \mu$, then

$$\begin{aligned} \mathrm{E}\{XF_X(X)\} &= \mu \times \frac{1}{2} + \frac{1}{2\sqrt{\pi}}\left\{p^2\sigma_1 + (1-p)^2\sigma_2 + p(1-p)\sqrt{\sigma_1^2+\sigma_2^2}\sqrt{2}\right\}, \\ \mathrm{cov}\{X, F_X(X)\} &= \frac{1}{2\sqrt{\pi}}\left\{p^2\sigma_1 + (1-p)^2\sigma_2 + p(1-p)\sqrt{\sigma_1^2+\sigma_2^2}\sqrt{2}\right\}, \end{aligned}$$

which gives (24). ■

References

Hardy, G. H., Littlewood, J. E., and Pólya, G. (1988), *Inequalities* (2nd ed.), Cambridge, UK: Cambridge University Press.