Electronic Research Archive

https://www.aimspress.com/journal/era

ERA, 33(8): 4893–4916.
DOI: 10.3934/era.2025220
Received: 27 June 2025
Revised: 04 August 2025
Accepted: 08 August 2025
Published: 21 August 2025

Research article

Asymptotic representations for Spearman's footrule correlation coefficient

Liqi Xia[1], **Li Guan**[2] **and Weimin Xu**[3,*]

[1] School of Mathematics, Qilu Normal University, Jinan 250200, China

[2] School of Mathematics, Statistics and Mechanics, Beijing University of Technology, Beijing 100124, China

[3] School of Data Science and Artificial Intelligence, Wenzhou University of Technology, Wenzhou 325000, China

* **Correspondence:** Email: wzxuweimin@163.com.

Abstract: This study addressed the theoretical challenges faced by rank-dependence structures in Spearman's footrule correlation coefficient. Two asymptotic representations were proposed to approximate its null distribution. The first approach simplifies dependencies by substituting empirical distribution functions with their population counterparts. The second employs Hájek projection technique to decompose the initial form into a sum of independent components, establishing rigorous asymptotic normality. Simulation experiments combined with real-world data analyses validated both representations, demonstrating their excellent approximation to the limiting normal distribution under the independence hypothesis.

Keywords: asymptotic representation; Spearman's footrule; rank correlation; correlation coefficient; Hájek projection

1. Introduction

Nonparametric measures of association are pivotal in statistical inference, particularly when data violate parametric assumptions or exhibit complex dependencies. The distribution-free nature of rank-based methods provides robustness against outliers, while the broader nonparametric approach circumvents model misspecification issues. These techniques deliver reliable inference across diverse data structures where parametric methods falter, proving indispensable in genomics, econometrics, and climate science. Such fields routinely feature heavy-tailed distributions, nonlinear associations, and censored observations where conventional parametric approaches fail.

1.1. Spearman's footrule rank correlation coefficient and its advantages

Among these measures, Spearman's footrule rank correlation coefficient [1] has attracted renewed attention as a robust and interpretable rank-based correlation measure [2–4]. By summing the absolute differences between paired ranks, this metric quantifies permutation disarray and provides an intuitive alternative to Euclidean-based measures such as Spearman's rho. Notably, it also admits a tractable population formulation. For continuous variables X and Y with copula C, the population version is defined as:

$$\varphi_C = 1 - 3 \int_{[0,1]^2} |u - v| dC(u, v), \tag{1.1}$$

where u and v represent the marginal distribution functions of X and Y, respectively. Under independence, $\varphi_C = 0$, while perfect agreement or disagreement yields $\varphi_C = 1$ or $\varphi_C = -\frac{1}{2}$ [5].

While historically underappreciated due to a limited recognition of its statistical properties, Spearman's footrule has recently gained traction as a versatile rank correlation measure. Its practical advantages can be organized into four key dimensions:

1) Computational efficiency. By aggregating absolute rank differences, Spearman's footrule achieves linear time complexity ($O(n)$), enabling efficient processing of large datasets. This contrasts sharply with Kendall's tau ($O(n^2)$ for concordance calculations), making the footrule particularly advantageous for high-dimensional applications.
2) Positional sensitivity. The metric directly quantifies the magnitude of rank shifts, offering superior resolution for applications where positional precision matters. For instance, in search engine evaluation, discrepancies in top-ranked results disproportionately impact user experience–a nuance effectively captured by the footrule's displacement-focused design. In contrast, Kendall's tau emphasizes order consistency (pairwise inversions), while Spearman's rho measures linear rank associations, both of which place less weight on specific positional changes.
3) Interpretability. The metric reflects the total displacement of ranks, and its normalized form (see Eq (2.1) in Section 2) provides a clear measure of alignment between rankings, making it accessible to non-expert users.
4) Robustness to outliers. Rooted in rank-based differences (L_1 norm), the footrule demonstrates greater robustness to extreme rank perturbations compared to Spearman's rho (L_2 norm). Although both methods mitigate outlier effects through ranking, the squared differences in rho's formulation amplify noise impacts, making footrule preferable in noise-prone environments.

These advantages have enabled Spearman's footrule to gain widespread adoption across diverse disciplines. In genomics, Kim et al. [6] developed a normalized footrule variant to evaluate microarray reproducibility, addressing challenges posed by low signal-to-noise ratios that often introduce outliers. The metric has also been applied in information retrieval to quantify discrepancies between ranked lists [7, 8], while Iorio et al. [9] and Lin and Ding [10] extended its use to gene expression profiling and bioinformatics analyses. Furthermore, preference learning frameworks have integrated footrule distance into Bayesian Mallows models, enabling aggregation of incomplete rankings and quantification of uncertainties in consensus rankings [11].

In empirical studies, understanding associations between variables often relies on non-parametric measures like Spearman's footrule and Spearman's rho rank correlation coefficient. Both coefficients operate on ranked data, avoiding distributional assumptions, but their objectives diverge. Spearman's footrule quantifies the absolute disparity in rankings, aggregating differences in positional order to assess ranking fidelity. For instance, Spearman's footrule is ideal for evaluating how closely one variable's ranks mirror another's. Conversely, rho measures the strength of monotonic trends by correlating ranked data, effectively serving as a rank-based alternative to Pearson's coefficient. While footrule prioritizes precise alignment in ordinal positions, rho targets broader associative patterns, such as consistent increases or decreases across variables. These distinctions make footrule more suitable for analyzing ranking accuracy, whereas rho is preferred for detecting systematic dependencies. Our study focuses on footrule, leveraging its sensitivity to positional shifts to explore asymptotic properties in ranking data.

1.2. Asymptotic representations in rank-based inference

1.2.1. Theoretical foundations of rank statistics

The inherent dependence structure of ranks has historically complicated theoretical advancements in rank-based statistics. Asymptotic representations serve as crucial technical tools for simplifying inference, exemplified by linear rank statistics (Section 13.1 of the classic monograph in statistics [12]). This class includes pivotal tests like Wilcoxon, van der Waerden, median, and log rank statistics. Theorem 13.5 of van der Vaart [12] establishes an asymptotic representation for linear rank statistics through independent uniformly distributed variables, proved via martingale convergence and Hájek projection. Corollary 13.8 of van der Vaart [12] leverages this representation to demonstrate asymptotic normality and subsequent efficiency proofs.

Beyond linear statistics, Angus [13] used coupling techniques to obtain the asymptotic representation of the rank statistic $B_n = \sum_{k=1}^{n-1} |\pi_k - \pi_{k+1}|$, where $(\pi_1, \pi_2, \ldots, \pi_n)$ is a random permutation of the integers $1, 2, \ldots, n$. This asymptotic representation also includes independent and uniformly distributed random variables and was used to prove its asymptotic normality. Later, this rank statistic was used by Chatterjee [14] to construct the recently popular Chatterjee's rank correlation coefficient. In Shi et al. [15], this asymptotic representation was further used to study the power analysis of Chatterjee's rank correlation coefficient. In addition, Lin and Han [16] and Xia et al. [17] have recently improved this correlation coefficient using asymptotic representations (see their Remark 10 and Section 3, respectively) to establish the relevant asymptotic theory of the statistics and perform hypothesis testing.

1.2.2. Gaps and motivations for Spearman's footrule

Theoretical investigations of Spearman's footrule correlation coefficient have been conducted in several studies. Diaconis and Graham [18] established its asymptotic normality under independence using combinatorial arguments, while Sen and Salama [19] leveraged Markov chain properties and martingale theory to derive similar results, emphasizing significance in permutation-based frameworks. Despite these advances, critical rank-based gaps persist. To address this, we derive two distinct asymptotic representations under independence, composed of independent and identically uniformly distributed random variables independent of the original data distribution. These

representations preserve the distribution-free property of Spearman's footrule-based tests. Our motivations for developing these asymptotic representations align with previous analyses: First, to simplify proofs for the limiting null distribution (Theorem 2.3) using novel approaches distinct from prior literature; second, to enable extension to multivariate Spearman's footrule and establish asymptotic theory (developed separately); third, to facilitate future power analysis and nonparametric confidence interval construction.

1.2.3. Proof strategies

The proof strategies for our two asymptotic representations fundamentally differ from prior methodologies. Specifically, we establish an initial asymptotic representation by replacing empirical distribution functions with their population counterparts, leveraging empirical process theory. This approach circumvents complexities from rank dependencies while directly linking the statistic to its limiting behavior. Building on this foundation, we employ the Hájek projection technique to decompose Spearman's footrule into a linear combination of independent components. This decomposition not only rigorously establishes asymptotic normality but also elucidates rank transformations' structural role in the statistical framework.

2. Asymptotic representations and the limiting null distribution

At the outset, we provide a notation table for clarity and streamlined exposition. This table details Spearman's footrule correlation coefficient along with its first and second representations under the independence assumption of X and Y. It systematically documents their textual locations, theoretical expectations, and variances under independence.

2.1. Asymptotic representations

Consider a bivariate continuous random variable (X, Y) with joint distribution function $P(x, y)$ and marginal distribution functions $F(x)$ and $G(y)$. A size-n sample $\{(X_1, Y_1), \ldots, (X_n, Y_n)\}$ is independently and identically distributed (i.i.d.) from (X, Y). The rank of X_i is defined as $R_i = \sum_{k=1}^{n} \mathbb{I}(X_k \leqslant X_i)$, where $\mathbb{I}(\cdot)$ denotes the indicator function, for $i = 1, ..., n$. Similarly, the rank of Y_i is $S_i = \sum_{k=1}^{n} \mathbb{I}(Y_k \leqslant Y_i)$. Spearman's footrule rank correlation coefficient is then given by

$$\varphi_n := \varphi\left(\{(X_i, Y_i)\}_{i=1}^{i=n}\right) = 1 - \frac{3}{n^2 - 1} \sum_{i=1}^{n} |R_i - S_i|. \tag{2.1}$$

Under the assumption of independence between X and Y, its expectation and variance are as follows:

$$\mathrm{E}\varphi_n = 0, \quad \mathrm{Var}(\varphi_n) = \frac{2n^2 + 7}{5(n + 1)(n - 1)^2},$$

which can be directly calculated using the results from Kleinecke et al. [20].

Although the rank-based nature of φ_n renders the associated tests fully distribution-free (i.e., independent of the underlying data distribution), the inherent dependence among ranks in practical applications complicates the derivation of certain asymptotic theories under the independence

assumption between X and Y. To address this challenge, we introduce two asymptotic representations of φ_n that preserve its theoretical integrity while simplifying the analysis.

Recall the empirical distribution functions $F_n(x) = \frac{1}{n}\sum_{k=1}^{n}\mathbb{I}(X_k \leqslant x)$ and $G_n(y) = \frac{1}{n}\sum_{k=1}^{n}\mathbb{I}(Y_k \leqslant y)$ of X and Y for any $x \in R$ and $y \in R$. Through intuitive and straightforward calculation, φ_n can be rewritten as

$$\begin{aligned}
\varphi_n &= 1 - \frac{3}{n^2-1}\sum_{i=1}^{n}|R_i - S_i| \\
&= \frac{3}{n^2-1}\left(\frac{n^2-1}{3} - \sum_{i=1}^{n}|R_i - S_i|\right) \\
&= \frac{3}{n^2-1}\left(\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}|R_i - S_j| - \sum_{i=1}^{n}|R_i - S_i|\right) \\
&= \frac{3n^2}{n^2-1}\left(\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}|F_n(X_i) - G_n(Y_i)| - \frac{1}{n}\sum_{i=1}^{n}|F_n(X_i) - G_n(Y_i)|\right).
\end{aligned}$$

A natural inclination is to replace these two empirical functions with their population counterparts F and G, but there are still remaining terms that need to be addressed. Notably, $F(X)$ and $G(Y)$ follow a uniform distribution over the interval $[0, 1]$. This ultimately leads to the following asymptotic representations and theorems. Appendix A.1 presents all proof details. It should first be clarified that although the proposed asymptotic representations could, in some sense, also be viewed as stochastic representations, the subsequent theorems primarily treat them within an asymptotic framework.

Theorem 2.1 (**The first asymptotic representation**). *Under the assumption of independence between X and Y, φ_n is asymptotically identically distributed with the following form,*

$$\varphi_n' = \frac{3n^2}{n^2-1}\left(\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}|U_i - V_j| - \frac{1}{n}\sum_{i=1}^{n}|U_i - V_i|\right), \tag{2.2}$$

where, $U_1, ..., U_n$ and $V_1, ..., V_n$ are i.i.d. random variables from uniform distribution $U(0, 1)$, and U_i and V_i are also independent for $i = 1, \cdots, n$. Additionally,

$$\mathrm{E}\varphi_n' = 0, \quad \mathrm{Var}(\varphi_n') = \frac{2n^2}{5(n+1)^2(n-1)}.$$

To further simplify the expression (2.2), we now apply Hájek projection to φ_n', leading to the following theorem.

Theorem 2.2 (**The second asymptotic representation**). *Under the assumption of independence between X and Y, φ_n''s Hájek asymptotic representation is as follows:*

$$\varphi_n'' = \frac{3}{n+1}\sum_{i=1}^{n}\left(\frac{2}{3} - |U_i - V_i| - U_i(1-U_i) - V_i(1-V_i)\right), \tag{2.3}$$

with expectation and variance,

$$\mathrm{E}\varphi_n'' = 0, \quad \mathrm{Var}(\varphi_n'') = \frac{2n}{5(n+1)^2}.$$

2.2. The limiting null distribution and Berry-Essen bound

Based on the proofs of Theorems 2.1 and 2.2, it can be readily derived that the difference between Spearman's footrule correlation coefficient and φ_n'' is $O_p(1/\sqrt{n})$, i.e., $\varphi_n - \varphi_n'' = O_p(1/\sqrt{n})$. This represents a relatively fast convergence rate. One significant application of the asymptotic representations developed in this study is to establish the asymptotic normality of φ_n under the independence condition between X and Y. By utilizing Theorem 2.2 in conjunction with Theorem 2.1, the limiting null distribution of φ_n can be readily obtained.

Theorem 2.3 (**The limiting null distribution**). *Under the assumption of independence between X and Y, $\sqrt{n}\varphi_n$, $\sqrt{n}\varphi_n'$, and $\sqrt{n}\varphi_n''$ converge weakly to the same normal distribution with mean 0 and variance* $2/5$.

Remark 1. *Existing literature presents multiple approaches for deriving the limiting null distribution of $\sqrt{n}\varphi_n$. Diaconis and Graham [18] established normality using Hoeffding's combinatorial central limit theorem [21]. Sen and Salama [19] incorporated martingale techniques to study asymptotic normality. Shi et al. [22] derived convergence rates using combinatorial central limit theorem and Cramér-type moderate deviations [23]. Shi et al. [24] obtained alternative convergence rates via Edgeworth expansion [25]. Both approaches naturally yield the limiting distribution. Our methodology demonstrably differs from these techniques. It provides a foundation for further theoretical investigations of Spearman's footrule.*

Although existing literature in Remark 1 employs various methods to establish the convergence of the standardized Spearman's footrule to the standard normal distribution, the approach proposed in this paper also arrives at the same conclusion. The following theorem utilizes the form of the Berry-Esseen bound to demonstrate this convergence.

Theorem 2.4 (**Berry-Essen bound**). *Under the assumption of independence between X and Y, for any $x \in R$,*

$$\sup_{x\in R}\left|P(\varphi_n/\sqrt{\mathrm{Var}(\varphi_n)} \leqslant x) - \Phi(x)\right| \leqslant \frac{C}{\sqrt{n}},$$

where C is a constant independent of n, and Φ is the cumulative distribution function of the standard normal distribution.

3. Simulation studies

This section evaluates the performance of our two asymptotic representations using Monte Carlo simulations. The assessment focuses on two aspects. First, Section 3.1 examines estimated means and variances. These are calculated for Spearman's footrule (φ_n) and its two representations (φ_n', φ_n''). Second, Section 3.2 investigates approximation quality. This includes pairwise comparisons between φ_n, φ_n', and φ_n''. It also assesses their approximation to the normal limit distribution.

For the calculation of φ_n, let X and Y be drawn from the standard normal distribution and the standard uniform distribution, respectively. For φ_n' and φ_n'', their calculations are performed by generating random numbers from the standard uniform distribution according to Eqs (2.2) and (2.3). Appendix A.2 additionally provides supplementary simulations under different data distributions and in terms of testing performance.

3.1. Simulation of estimated mean, variance, bias, and root mean square error

For the proposed asymptotic representations, which act as estimators for the population parameter φ_C (introduced in the first paragraph of Section 1.1), it is natural to evaluate their performance in simulation studies by examining metrics such as estimated mean (EM), estimated variance (EV), bias(Bias), and root mean square error (RMSE). The definitions of bias and RMSE are given by

$$\text{Bias} = \frac{1}{10000}\sum_{k=1}^{10000}\left(\hat{\theta}_k - \theta\right) \qquad \text{and} \qquad \text{RMSE} = \sqrt{\frac{1}{10000}\sum_{k=1}^{10000}\left(\hat{\theta}_k - \theta\right)^2},$$

respectively, where $\hat{\theta}$ takes one from $\{\varphi_n, \varphi_n', \varphi_n''\}$. Additionally, according to the consistency of φ_n (Nelsen [5]), φ_n converges to population φ_C in probability, and under the assumption that X and Y are independent, $\varphi_C = 0$. Further, since both φ_n' and φ_n'' are asymptotically equivalent to φ_n (as established in Theorems 2.2 and 2.3), φ_n' and φ_n'' also converge to 0 in probability. Consequently, the true values of φ_n, φ_n', and φ_n'' in the definitions of Bias and RMSE are all 0, meaning $\theta = 0$. The sample sizes throughout this subsection are set to $n = 10, 20, 30, \cdots, 100$, with 10,000 simulation repetitions.

The quantitative results are summarized in Table 1. Analysis of these results reveals that the EM and EV closely approximate their true counterparts (true variances are detailed in Table 2). With respect to RMSE, all methods exhibit a decreasing trend in RMSE as sample size increases. Notably, the RMSE values of the two proposed asymptotic representations are consistently smaller than those of φ_n. This indicates that both our asymptotic representations and the original Spearman's footrule (φ_n) serve as valid estimators, with our proposed methods demonstrating superior performance compared to φ_n.

Table 1. The EM, EV, Bias, and RMSE of φ_n, φ_n', and φ_n''.

		$n = 10$	$n = 20$	$n = 30$	$n = 40$	$n = 50$	$n = 60$	$n = 70$	$n = 80$	$n = 90$	$n = 100$
φ_n	EM	–0.00038	0.00051	–0.00013	–0.00170	0.00019	–0.00144	–0.00094	–0.00014	–0.00008	0.00066
	EV	0.04652	0.02079	0.01389	0.01018	0.00804	0.00673	0.00589	0.00504	0.00449	0.00394
	Bias	–0.00038	0.00051	–0.00013	-0.00170	0.00019	–0.00144	–0.00094	–0.00014	–0.00008	0.00066
	RMSE	0.21568	0.14418	0.11783	0.10088	0.08966	0.08202	0.07673	0.07102	0.06701	0.06280
φ_n'	EM	0.00225	0.00213	–0.00020	0.00001	–0.00098	0.00074	–0.00045	–0.00040	0.00052	–0.00033
	EV	0.03677	0.01965	0.01290	0.00984	0.00779	0.00657	0.00572	0.00497	0.00448	0.00402
	Bias	0.00225	0.00213	–0.00020	0.00001	–0.00098	0.00074	–0.00045	–0.00040	0.00052	–0.00033
	RMSE	0.19177	0.14017	0.11357	0.09921	0.08824	0.08106	0.07561	0.07049	0.06693	0.06340
φ_n''	EM	0.00122	0.00083	–0.00014	0.00204	0.00083	–0.00030	0.00032	0.00032	0.00083	0.00047
	EV	0.03237	0.01781	0.01252	0.00971	0.00786	0.00645	0.00552	0.00497	0.00428	0.00389
	Bias	0.00122	0.00083	–0.00014	0.00204	0.00083	–0.00030	0.00032	0.00032	0.00083	0.00047
	RMSE	0.17991	0.13346	0.11190	0.09858	0.08865	0.08030	0.07426	0.07048	0.06542	0.06238

Regarding bias evaluation, although the bias values remain relatively close to the theoretical zero value, a more detailed visualization of their trends is necessary. To achieve this, we conducted additional simulations with an increased sample size of $n = 1000$ while maintaining other experimental parameters. The resulting bias and RMSE trends are presented in Figure 1. Analysis of the figure reveals that the biases of φ_n, φ_n', and φ_n'' all approach zero as the sample size grows, though with some fluctuations. For RMSE visualization, φ_n exhibits marginally higher values only at small sample sizes. As n increases, all three methods demonstrate nearly identical convergence patterns toward zero, indicating comparable asymptotic performance when sample sizes are sufficiently large.

Table 2. Notation summary table for Spearman's footrule correlation coefficient and its first and second asymptotic representations.

	Notation	Location	Expectation	Variance
Spearman's footrule correlation coefficient	φ_n	Eq (2.1)	0	$\frac{2n^2+7}{5(n+1)(n-1)^2}$
The first asymptotic representation	φ_n'	Eq (2.2)	0	$\frac{2n^2}{5(n+1)^2(n-1)}$
The second asymptotic representation	φ_n''	Eq (2.3)	0	$\frac{2n}{5(n+1)^2}$

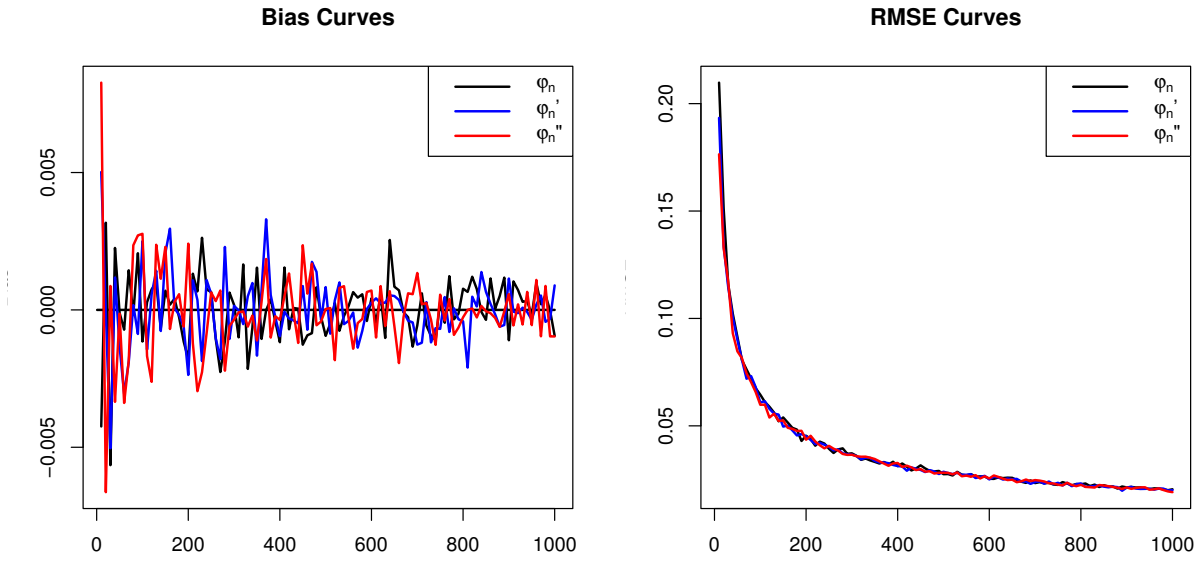

Figure 1. The bias and root mean square error (RMSE) curves of φ_n, φ_n', and φ_n''.

3.2. Simulation of the normal limiting distribution

In this subsection, we simulate the asymptotic behaviors of $\sqrt{n}\varphi_n$, $\sqrt{n}\varphi_n'$, and $\sqrt{n}\varphi_n''$ through three approaches. For the first two methods, we estimate their empirical density functions and cumulative distribution functions (CDFs) via simulations. Specifically, we use four sample sizes: $n = 10, 20, 30$, and 100, with 100,000 simulation repetitions each. For the third approach, we apply the Kolmogorov-Smirnov (KS) two-sample test to assess the distributional identity between pairs of the proposed methods and between each method and the normal distribution [26]. A total of six pairwise comparisons are conducted: $\varphi_n - N(0, 0.4)$, $\varphi_n' - N(0, 0.4)$, and $\varphi_n'' - N(0, 0.4)$, as well as $\varphi_n - \varphi_n'$, $\varphi_n - \varphi_n''$, and $\varphi_n' - \varphi_n''$, where $N(0, 0.4)$ denotes the limiting null distribution shared by $\sqrt{n}\varphi_n$, $\sqrt{n}\varphi_n'$ and $\sqrt{n}\varphi_n''$. The sample sizes for this analysis range from $n = 10$ to $n = 100$ in increments of 10, with 1000 simulation repetitions per setting. The KS test is implemented using the "ks.test" function from R's base library. Empirical density function and CDF curves are displayed in Figures 2 and 3, respectively, while all KS test p-values are reported in Table 3.

Figures 2 and 3 demonstrate that even at $n = 30$, the distributions of $\sqrt{n}\varphi_n'$, $\sqrt{n}\varphi_n''$, and $\sqrt{n}\varphi_n$ closely approximate the theoretical limiting distribution (normal with mean 0 and variance 2/5). Among these, $\sqrt{n}\varphi_n'$ provides the best approximation, followed by $\sqrt{n}\varphi_n''$, with $\sqrt{n}\varphi_n$ being the worst. This ordering reflects their respective convergence rates to the limiting distribution. Notably, at extremely small

sample sizes ($n = 10$), $\sqrt{n}\varphi_n$ performs poorly as shown in the first subfigures. This occurs due to rank permutations inherent in φ_n's structure. Despite numerous distinct permutations (10!), φ_n exhibits significant value repetitions. Even with 100,000 simulation repetitions, relatively few distinct values emerge (dozens). This sparsity causes the non-smooth kernel density curve for φ_n in Figure 2 and the stepwise empirical CDF in Figure 3 at $n = 10$. As sample size increases, this phenomenon diminishes. All methods exhibit similar asymptotic behavior converging to the theoretical normal distribution.

Table 3. The p-values of KS test for six combinations.

	$n = 10$	$n = 20$	$n = 30$	$n = 40$	$n = 50$	$n = 60$	$n = 70$	$n = 80$	$n = 90$	$n = 100$
$\varphi_n - N(0, 0.4)$	0.00003	0.03328	0.08691	0.18112	0.18112	0.34100	0.10828	0.50036	0.36998	0.02246
$\varphi'_n - N(0, 0.4)$	0.31358	0.31358	0.14834	0.93558	0.79439	0.31358	0.60992	0.31358	0.64756	0.26338
$\varphi''_n - N(0, 0.4)$	0.26338	0.68523	0.75910	0.40047	0.46577	0.85929	0.64756	0.09710	0.95406	0.43243
$\varphi_n - \varphi'_n$	0.00103	0.04282	0.43243	0.24060	0.21933	0.46577	0.75910	0.98826	0.85929	0.31358
$\varphi_n - \varphi''_n$	0.00000	0.09710	0.07762	0.21933	0.03328	0.53605	0.43243	0.40047	0.34100	0.53605
$\varphi'_n - \varphi''_n$	0.31358	0.72255	0.72255	0.82796	0.34100	0.13383	0.53605	0.36998	0.72255	0.68523

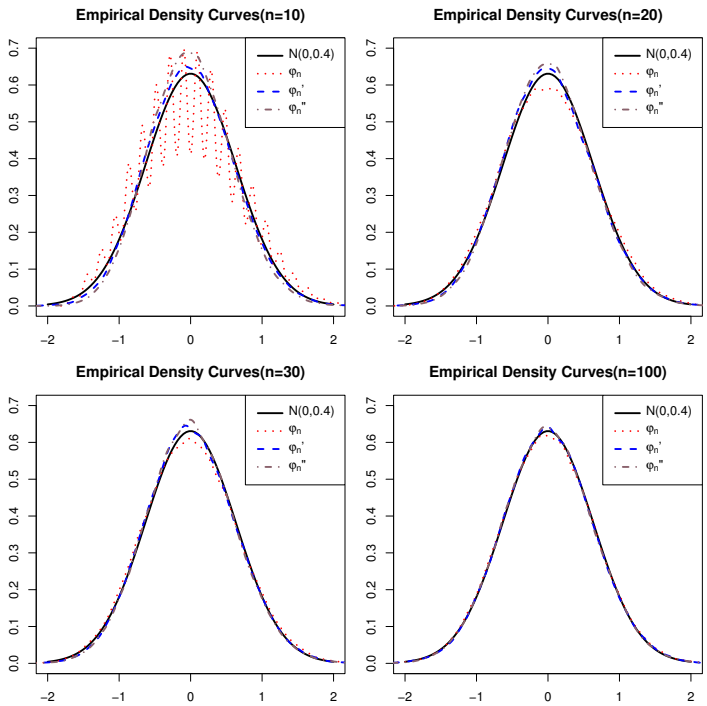

Figure 2. Empirical density curves of $\sqrt{n}\varphi_n$, $\sqrt{n}\varphi'_n$, and $\sqrt{n}\varphi''_n$, estimated using kernel density estimation with a Gaussian kernel, where the solid line represents a normal curve with a mean of 0 and a variance of 0.4.

The p-values reported in Table 3 reveal that, with the exception of the small-sample case ($n = 10$), all pairwise comparisons between the methods demonstrate approximate distributional equivalence. Furthermore, the two proposed asymptotic representations and the original Spearman's footrule correlation coefficient all exhibit close approximation to normal distributions. For the $n = 10$

scenario, the aforementioned conclusions hold, provided that φ_n is excluded from the comparison. However, the KS test for φ_n yields statistically significant p-values approaching zero, indicating strong evidence that the distribution of φ_n differs from both the other methods and the normal distribution. This observation aligns consistently with prior analyses.

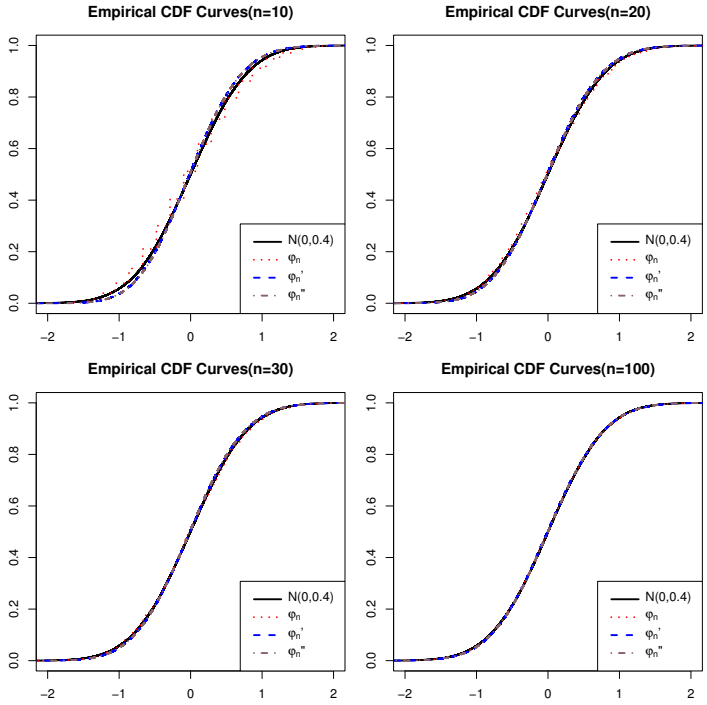

Figure 3. Empirical CDF curves of $\sqrt{n}\varphi_n$, $\sqrt{n}\varphi_n'$, and $\sqrt{n}\varphi_n''$. The solid line represents a normal cumulative distribution function curve with a mean of 0 and a variance of 0.4.

4. Application to a real dataset

The Boston dataset, first introduced by Harrison Jr and Rubinfeld [27], contains 506 observations of 1970s Boston suburbs. Originally developed to analyze air pollution's impact on housing values through hedonic pricing models, it became a landmark study linking environmental economics to property valuation. Available in MASS package of R software (via data(Boston)), this dataset remains a standard reference in urban economics for studying environmental factors while controlling housing price determinants.

In this dataset, the relationship between median home value (medv) and average room count (rm) represents a fundamental economic hypothesis in real estate: physical attributes intrinsically determine property valuation. Harrison Jr and Rubinfeld [27] prioritized rm as a control variable because room quantity directly proxies living space–a primary driver of housing demand. Testing this pairing validates core microeconomic theory that housing characteristics command market premiums. Practically, quantifying this correlation helps appraisers establish baseline valuations, informs homebuyer decisions, and reveals suburbanization patterns where larger homes command disproportionate value premiums. The analysis also contextualizes the original study's core finding:

after controlling for environmental disamenities, structural features dominate price determination.

To examine the dependency between the two variables, we employed Spearman's footrule correlation coefficient (φ_n), yielding a value of 0.484, which indicates a statistically strong relationship between them. This empirically demonstrates that structural characteristics dominate housing valuation even after controlling for environmental factors. To further validate our proposed approaches, we implemented hypothesis testing using the φ_n statistic and evaluated its test performance approximation through two asymptotic representations. Critical values for all three methods were established via permutation approach with 1000 replicates at significance levels of $\alpha = 0.01$ and $\alpha = 0.001$. Additionally, to generate independent random datasets, we performed 1000 random samples at varying sample sizes ($n = 10, 15, 20, 25, 30, 35, 40, 45, 50$) from the Boston dataset's 506 observations. All rejection rates are presented in Table 4, demonstrating that for sample sizes exceeding 30, our test achieves statistical power equivalent to the original φ_n method at both significance levels.

Table 4. The rejection rates of φ_n, φ'_n, and φ''_n at significance levels of $\alpha = 0.01$ and $\alpha = 0.001$.

		$n = 10$	$n = 15$	$n = 20$	$n = 25$	$n = 30$	$n = 35$	$n = 40$	$n = 45$	$n = 50$
$\alpha = 0.01$	φ_n	0.370	0.598	0.782	0.883	0.937	0.966	0.986	0.993	0.998
	φ'_n	0.530	0.674	0.822	0.903	0.942	0.975	0.986	0.996	0.998
	φ''_n	0.561	0.703	0.839	0.904	0.951	0.974	0.987	0.997	0.998
$\alpha = 0.001$	φ_n	0.169	0.388	0.566	0.732	0.827	0.889	0.933	0.972	0.986
	φ'_n	0.293	0.473	0.623	0.760	0.850	0.898	0.933	0.969	0.990
	φ''_n	0.329	0.500	0.635	0.767	0.862	0.917	0.941	0.975	0.989

5. Conclusions

Spearman's footrule, while robust to distributional assumptions, encounters theoretical challenges due to rank dependencies. Two asymptotic representations have been developed to address this issue under independence assumptions. The first representation simplifies the statistic by incorporating population distribution functions, reducing structural complexity. The second approach employs Hájek projection to decompose the footrule into independent components, thereby reinforcing asymptotic normality and enhancing theoretical tractability. These methodologies collectively improve the analytical framework for Spearman's footrule while preserving its distributional robustness.

Our proposed asymptotic representations for Spearman's footrule share a fundamental objective with Diaconis and Graham [18] and Sen and Salama [19]: establishing the asymptotic normality of the sum of absolute rank differences $\sum_{i=1}^{n} |R_i - S_i|$ under the assumption of independence between X and Y. The key distinction lies in the technical strategies employed. While Diaconis and Graham [18] directly invoked the combinatorial central limit theorem (arguably the most direct proof), and Sen and Salama [19] utilized a Markov chain construction followed by a martingale approach (a more intricate technique), our methodology indirectly achieves this via asymptotic representations derived using empirical process theory and Hájek projection. Crucially, the second representation manifests as a sum of i.i.d. random variables. This representation-based foundation facilitates extensions beyond

asymptotic normality–such as deriving the Berry-Esseen bound (Theorem 2.4), enabling multivariate generalizations, power analysis for tests, and confidence interval construction, which are not readily attainable via the combinatorial or martingale approaches of the prior works.

A. Appendix

A.1. Appendix 1

Lemma A.1. *Given that U_1, V_1, and V_2 are independently and identically distributed from the uniform distribution $U(0,1)$, through simple integral calculation, the following facts can be easily deduced:* $\mathrm{E}|U_1 - V_1| = \frac{1}{3}$, $\mathrm{E}(|U_1 - V_1||U_1) = \frac{1}{2} - U_1(1-U_1)$. $\mathrm{E}(U_1(1-U_1)) = \frac{1}{6}$, $\mathrm{Var}(|U_1 - V_1|) = \frac{1}{18}$, $\mathrm{Var}(U_1(1-U_1)) = \frac{1}{180}$, $\mathrm{Cov}(|U_1 - V_1|, U_1(1-U_1)) = -\frac{1}{180}$, $\mathrm{Cov}(|U_1 - V_1|, |U_1 - V_2|) = \frac{1}{180}$.

Lemma A.2 (Lemma 19.24 in van der Vaart [12])**.** *Suppose that $\mathcal{F}$ is a P–Donsker class of measurable functions, and f_n is a sequence of random functions that take their values in $\mathcal{F}$ such that $\int (f_n(x) - f(x))^2 \, dP(x)$ converges in probability to 0 for some $f \in L_2(P)$. Then $\mathbb{G}_n(f_n - f) \xrightarrow{P} 0$ and hence $\mathbb{G}_n f_n \rightsquigarrow \mathbb{G}_P f$.*

Proof of Theorem 2.1. Let $(X_1, Y_1), \ldots, (X_n, Y_n) \in \mathcal{Z} = \mathbb{R} \times \mathbb{R}$ be a random sample from a probability distribution P defined on a measurable space $(\mathcal{Z}, \mathcal{A})$. We denote two empirical distributions as $\mathbb{P}_n = n^{-1} \sum_{i=1}^n \delta_{(X_i,Y_i)}$ and $\mathbb{P}'_n = n^{-2} \sum_{i=1}^n \sum_{j=1}^n \delta_{(X_i,Y_j)}$, where $\delta_{(x,y)}$ represents the probability distribution degenerate at the point (x, y). For a given measurable function $f : \mathcal{Z} \mapsto \mathbb{R}$, we use $\mathbb{P}_n f$ and $\mathbb{P}'_n f$ to denote the expectations of f under the empirical measures $\mathbb{P}_n$ and $\mathbb{P}'_n$, respectively. Similarly, Pf represents the expectation of f under $P(x, y) = F(x)G(y)$. Thus, the expressions are given by:

$$\mathbb{P}_n f = \frac{1}{n} \sum_{i=1}^n f(X_i, Y_i), \mathbb{P}'_n f = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n f\left(X_i, Y_j\right), \quad Pf = \int f dP. \tag{A.1}$$

We choose

$$f(x, y) = |F(x) - G(y)| \text{ and } f_n(x, y) = |F_n(x) - G_n(y)|, \tag{A.2}$$

where $F_n(x) = \frac{1}{n} \sum_{k=1}^n \mathbb{I}(X_k \leqslant x)$ and $G_n(y) = \frac{1}{n} \sum_{k=1}^n \mathbb{I}(Y_k \leqslant y)$.

Let $\mathcal{F}_0$ be the collection of cumulative distribution functions for all univariate continuous variables and $\mathcal{F} = \{|F(x) - G(y)| \in \mathbb{R} : F, G \in \mathcal{F}_0 \text{ and } x, y \in \mathbb{R}\}$. Example 19.6 of van der Vaart [12] illustrates that $\mathcal{F}_0$ is a P-Donsker class. Thus, for all $(x, y) \in \mathbb{R}^2$ and $F, G \in \mathcal{F}_0$, according to the information provided on page 19 of Kosorok [28], along with the fact that $|a - b| = \max\{a, b\} - \min\{a, b\}$, $\mathcal{F}$ is also a P-Donsker class.

By the law of large numbers, for every x and y, it is apparent that $\mathrm{Sup}_{x \in R} |F_n(x) - F(x)| \xrightarrow{a.s.} 0$ and $\mathrm{Sup}_{y \in R} |G_n(y) - G(y)| \xrightarrow{a.s.} 0$ hold, thus resulting in $\mathrm{Sup}_{x \in R, y \in R} |f_n(x, y) - f(x, y)| \xrightarrow{a.s.} 0$, hence for some $f \in L_2(P)$, $\int (f_n(x) - f)^2 dP \xrightarrow{p} 0$ follows. Then, define

$$\mathbb{G}_n := \sqrt{n}(\mathbb{P}_n - P), \quad \mathbb{G}'_n := \sqrt{n}(\mathbb{P}'_n - P).$$

The empirical process evaluated at f is

$$\mathbb{G}_n f = \sqrt{n}(\mathbb{P}_n f - Pf), \ \ \mathbb{G}'_n f = \sqrt{n}(\mathbb{P}'_n f - Pf).$$

Thus, based on previous analysis, by directly applying Lemma A.2 (see also Lemma 19.24 in van der Vaart [12]), one has

$$\mathbb{G}'_n (f_n - f) = o_p(1), \ \ \mathbb{G}_n (f_n - f) = o_p(1),$$

i.e.,

$$(\mathbb{P}'_n - P)(f_n - f) = o_p(n^{-1/2}), \ \ (\mathbb{P}_n - P)(f_n - f) = o_p(n^{-1/2}).$$

Further expanding these two expressions leads to the following forms:

$$\mathbb{P}'_n f_n - \mathbb{P}'_n f - Pf_n + Pf = o_p(n^{-1/2}), \ \ \mathbb{P}_n f_n - \mathbb{P}_n f - Pf_n + Pf = o_p(n^{-1/2}).$$

Note that the combination of Eqs (A.1) and (A.2) can yield the following forms:

$$\mathbb{P}'_n f_n = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left|F_n(X_i) - G_n(Y_j)\right|, \ \ \mathbb{P}_n f_n = \frac{1}{n}\sum_{i=1}^{n}\left|F_n(X_i) - G_n(Y_i)\right|.$$

$$\mathbb{P}'_n f = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left|F(X_i) - G(Y_j)\right| = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left|U_i - V_j\right|.$$

$$\mathbb{P}_n f = \frac{1}{n}\sum_{i=1}^{n}\left|F(X_i) - G(Y_i)\right| = \frac{1}{n}\sum_{i=1}^{n}\left|U_i - V_i\right|.$$

$$Pf_n = \mathrm{E}\left|F_n(X_i) - G_n(Y_i)\right|, \ \ Pf = \mathrm{E}\left|F(X_i) - G(Y_i)\right| = \mathrm{E}\left|U_i - V_i\right|.$$

Thus,

$$\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left|F_n(X_i) - G_n(Y_j)\right| = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left|U_i - V_j\right| + Pf_n - Pf + o_p(n^{-1/2}),$$

$$\frac{1}{n}\sum_{i=1}^{n}\left|F_n(X_i) - G_n(Y_i)\right| = \frac{1}{n}\sum_{i=1}^{n}\left|U_i - V_i\right| + Pf_n - Pf + o_p(n^{-1/2}).$$

Combining the above equations, we obtain

$$\begin{aligned}
\varphi_n &= 1 - \frac{3}{n^2-1}\sum_{i=1}^{n}|R_i - S_i| \\
&= \frac{3}{n^2-1}\left(\frac{n^2-1}{3} - \sum_{i=1}^{n}|R_i - S_i|\right) \\
&= \frac{3}{n^2-1}\left(\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}|R_i - S_j| - \sum_{i=1}^{n}|R_i - S_i|\right) \\
&= \frac{3n^2}{n^2-1}\left(\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}|F_n(X_i) - G_n(Y_i)| - \frac{1}{n}\sum_{i=1}^{n}|F_n(X_i) - G_n(Y_i)|\right)
\end{aligned}$$

$$
\begin{aligned}
&= \frac{3n^2}{n^2-1}\left(\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left|U_i - V_j\right| - \frac{1}{n}\sum_{i=1}^{n}|U_i - V_i| + o_p(n^{-1/2})\right) + o_p(n^{-1/2}) \\
&= \varphi_n' + o_p(n^{-1/2}).
\end{aligned}
$$

Next, relying on the facts stated in Lemma A.1, we deal with the expectation and variance of φ_n'. Let $C_1 = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left|U_i - V_j\right|$ and $C_2 = \frac{1}{n}\sum_{i=1}^{n}|U_i - V_i|$, it is easy to calculate that $\mathrm{E}\varphi_n' = 0$.

Furthermore, routine calculation yields

$$
\begin{aligned}
\mathrm{Var}(C_1) &= \frac{1}{n^4}\left\{n^2\mathrm{Var}(|U_1 - V_1|) + 2\times n^2\times(n-1)\mathrm{Cov}(|U_1 - V_1|, |U_1 - V_2|)\right\} \\
&= \frac{1}{n^4}\left\{n^2\times\frac{1}{18} + 2n^2(n-1)\times\frac{1}{180}\right\} \\
&= \frac{n+4}{90n^2}. \\
\mathrm{Var}(C_2) &= \frac{1}{n^2}\times n\mathrm{Var}(|U_1 - V_1|) = \frac{1}{18n}. \\
\mathrm{Cov}(C_1, C_2) &= \frac{1}{n^3}\left(\sum_{i=1}^{n}\sum_{j=1}^{n}\left|U_i - V_j\right|, \sum_{i=1}^{n}|U_i - V_i|\right) \\
&= \frac{1}{n^3}\left\{n\mathrm{Var}(|U_1 - V_1|) + 2(n-1)\mathrm{Cov}(|U_1 - V_1|, |U_1 - V_2|)\times n\right\} \\
&= \frac{1}{n^3}\left\{n\times\frac{1}{18} + 2n(n-1)\times\frac{1}{180}\right\} \\
&= \frac{n+4}{90n^2}.
\end{aligned}
$$

Ultimately, it is derived that

$$
\begin{aligned}
\mathrm{Var}(\varphi_n') &= \left(\frac{3n^2}{n^2-1}\right)^2\left[\mathrm{Var}(C_1) + \mathrm{Var}(C_2) - 2\mathrm{Cov}(C_1, C_2)\right] \\
&= \left(\frac{3n^2}{n^2-1}\right)^2\left(\frac{n+4}{90n^2} + \frac{1}{18n} - 2\times\frac{n+4}{90n^2}\right) \\
&= \frac{2n^2}{5(n+1)^2(n-1)}.
\end{aligned}
$$

The proof of Theorem 2.1 is now complete. □

Proof of Theorem 2.2. Rewrite φ_n' as

$$
\begin{aligned}
\varphi_n' &= \frac{3}{n^2-1}\left(\sum_{i=1}^{n}\sum_{j=1}^{n}|U_i - V_j| - n\sum_{i=1}^{n}|U_i - V_i|\right) \\
&= \frac{3}{n^2-1}\left(\sum_{i\neq j}^{n}|U_i - V_j| - (n-1)\sum_{i=1}^{n}|U_i - V_i|\right) \\
&= \frac{3}{n^2-1}\left(\frac{n(n-1)}{2}T_1 - (n-1)T_2\right),
\end{aligned}
\tag{A.3}
$$

where $T_1 = \dfrac{2}{n(n-1)} \sum_{i \neq j}^{n} |U_i - V_j|$, and $T_2 = \sum_{i=1}^{n} |U_i - V_i|$.

Since T_2 is already a sum of independent and identically distributed terms, its Hájek projection remains itself. Thus, we only need to calculate the Hájek representation for T_1.

In fact, T_1 is a U-statistic that can be expressed as

$$T_1 = \frac{2}{n(n-1)} \sum_{i \neq j}^{n} |U_i - V_j| = \frac{2}{n(n-1)} \sum_{i<j}^{n} h\left((U_i, V_i)^\top, \left(U_j, V_j\right)^\top\right),$$

where the symmetric kernel function is taken as

$$h\left((u_1, v_1)^\top, (u_2, v_2)^\top\right) = |u_1 - v_2| + |u_2 - v_1|.$$

It is evident that the variance of T_1 exists. Let $\theta = \mathrm{E}\left[h\left((U_i, V_i)^\top, \left(U_j, V_j\right)^\top\right)\right]$, and $h_1\left((u, v)^\top\right) = \mathrm{E}\left[h\left((u, v)^\top, (U_2, V_2)^\top\right)\right] - \theta$. According to Lemma A.1 and through simple derivation, we have $\theta = E(|U_1 - V_2| + |U_2 - V_1|) = \dfrac{2}{3}$ and $h_1\left((u, v)^\top\right) = \dfrac{1}{3} - u(1-u) - v(1-v)$. The projection of $T_1 - \dfrac{2}{3}$ is then given by

$$\widetilde{T}_1 := \frac{2}{n} \sum_{i=1}^{n} \left[\frac{1}{3} - U_i(1-U_i) - V_i(1-V_i)\right].$$

Then, by applying Lemma 12.3 from van der Vaart [12], we obtain

$$T_1 - \frac{2}{3} = \widetilde{T}_1 + O_p(\frac{1}{n}),$$

i.e.,

$$T_1 = \frac{2}{n} \sum_{i=1}^{n} \left[\frac{1}{3} - U_i(1-U_i) - V_i(1-V_i)\right] + \frac{2}{3} + O_p(\frac{1}{n}).$$

Substituting this result into Eq (A.3), we get

$$\begin{aligned} \varphi_n' &= \frac{3}{n+1} \sum_{i=1}^{n} \left(\frac{2}{3} - |U_i - V_i| - U_i(1-U_i) - V_i(1-V_i)\right) + O_p(\frac{1}{n}) \\ &= \varphi_n'' + O_p(\frac{1}{n}). \end{aligned}$$

Additionally, utilizing the results from Lemma A.1, it is easy to derive that

$$\mathrm{E}\varphi_n'' = 0,$$

$$\begin{aligned} \mathrm{Var}(\varphi_n'') &= \left(\frac{3}{n+1}\right)^2 \times n\mathrm{Var}\left(\frac{2}{3} - |U_i - V_i| - U_i(1-U_i) - V_i(1-V_i)\right) \\ &= n\left(\frac{3}{n+1}\right)^2 \times [\mathrm{Var}(|U_1 - V_1|) + \mathrm{Var}(U_1(1-U_1)) \times 2 + 2\mathrm{Cov}(|U_1 - V_1|, U_1(1-U_1))] \end{aligned}$$

$$= n\left(\frac{3}{n+1}\right)^2 \times \left(\frac{1}{18} + \frac{1}{180} \times 2 + 2 \times \left(-\frac{1}{180}\right) \times 2\right)$$
$$= \frac{2n}{5(n+1)^2}.$$

Thus, this proof is complete. □

Proof of Theorem 2.3. φ_n'' can be expressed as the sum of independently and identically distributed random variables with existing second moment, so its asymptotic normality can be easily obtained through the ordinary central limit theorem. By further utilizing the asymptotic representations of Theorems 2.1 and 2.2, the asymptotic normality of φ_n' and φ_n is also apparent. □

Proof of Theorem 2.4. Within equation φ_n'', it is evident that $\frac{2}{3} - |U_i - V_i| - U_i(1-U_i) - V_i(1-V_i), i = 1 \cdots, n$ are i.i.d. random variables with finite second moments. Applying the classical Berry-Esseen theorem [29] to φ_n'' yields

$$\sup_{x \in R} \left| P(\varphi_n'' / \sqrt{\mathrm{Var}(\varphi_n'')} \leqslant x) - \Phi(x) \right| \leqslant \frac{C}{\sqrt{n}}.$$

for any $x \in R$, and C is a constant independent of n, and Φ is the cumulative distribution function of the standard normal distribution. Furthermore, by the lemma on page 228 of Serfling [30], for any sequence of positive constants $\{a_n\}$, one has

$$\begin{aligned} &\sup_{x \in R} \left| P(\varphi_n / \sqrt{\mathrm{Var}(\varphi_n)} \leqslant x) - \Phi(x) \right| \\ &= \sup_{x \in R} \left| P(\varphi_n / \sqrt{\mathrm{Var}(\varphi_n)} \leqslant x) - \varphi_n'' / \sqrt{\mathrm{Var}(\varphi_n'')} + \varphi_n'' / \sqrt{\mathrm{Var}(\varphi_n'')} - \Phi(x) \right| \\ &= P(\left| \varphi_n / \sqrt{\mathrm{Var}(\varphi_n)} - \varphi_n'' / \sqrt{\mathrm{Var}(\varphi_n'')} \right| > a_n) + O(a_n) + O(\frac{1}{\sqrt{n}}). \end{aligned} \tag{A.4}$$

Additional analysis from the proofs of Theorems 2.1 and 2.2 establishes that $\varphi_n - \varphi_n'' = O_p(1/\sqrt{n})$. Combining this with the variances of φ_n and φ_n'', in Eq (A.4), we only need to select $a_n = O(\frac{1}{\sqrt{n}})$ to ensure

$$\sup_{x \in R} \left| P(\varphi_n / \sqrt{\mathrm{Var}(\varphi_n)} \leqslant x) - \Phi(x) \right| \leqslant \frac{C}{\sqrt{n}}.$$

This completes the proof. □

A.2. Appendix 2

This appendix provides additional simulations for Section 3 of the main paper. Specifically, Section A.2.1 investigates the proposed asymptotic representations for mean and variance simulations under heavy-tailed and skewed data; Section A.2.2 further supplements the simulations in Section 3.2 of the main paper under heavy-tailed and skewed data; and Section A.2.3 examines simulations related to testing performance.

A.2.1. Additional simulation of estimated mean, variance, bias, and root mean square error

For the simulations in this subsection, we similarly employ EM, EV, Bias, and RMSE. Data generation follows two scenarios: 1) Heavy-tailed distribution: $X \sim N(0,1), Y \sim t(3)$, where $t(3)$ denotes a t-distribution with 3 degrees of freedom; and 2) skewed distribution: $X \sim N(0,1)$, $Y \sim \chi^2(3)$, where $\chi^2(3)$ represents a chi-square distribution with 3 degrees of freedom. All other settings remain identical to Section 3.1. All quantitative results are presented in Table A1, with bias and RMSE visualizations shown in Figures A1 and A2.

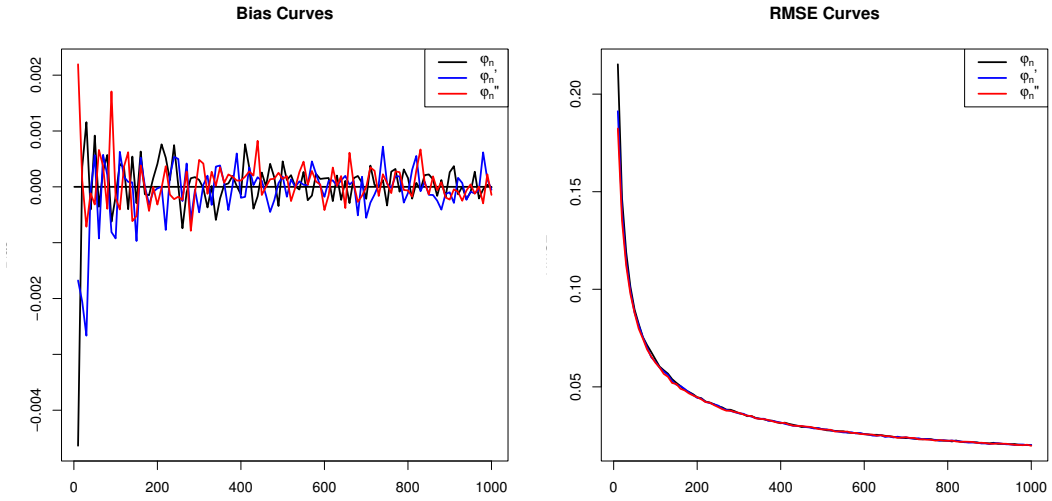

Figure A1. The bias and root mean square error (RMSE) curves of φ_n, φ_n', and φ_n'' under heavy-tailed distribution.

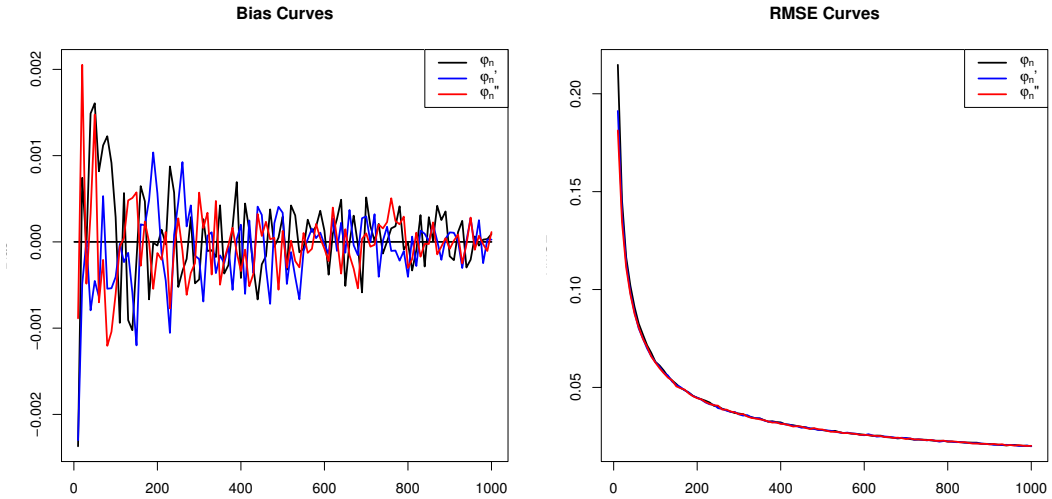

Figure A2. The bias and root mean square error (RMSE) curves of φ_n, φ_n', and φ_n'' under skewed distribution.

Table A1. The EM, EV, Bias, and RMSE of φ_n, φ_n', and φ_n'' under heavy-tailed and skewed distribution.

		$n = 10$	$n = 20$	$n = 30$	$n = 40$	$n = 50$	$n = 60$	$n = 70$	$n = 80$	$n = 90$	$n = 100$
Heavy-tailed distribution											
φ_n	EM	0.00028	–0.00134	–0.00317	–0.00147	0.00137	0.00177	0.00076	–0.00002	–0.00017	0.00139
	EV	0.04646	0.02114	0.01407	0.01032	0.00836	0.00674	0.00572	0.00506	0.00442	0.00403
	Bias	0.00028	–0.00134	–0.00317	–0.00147	0.00137	0.00177	0.00076	–0.00002	–0.00017	0.00139
	RMSE	0.21554	0.14540	0.11865	0.10159	0.09143	0.08209	0.07562	0.07111	0.06646	0.06351
φ_n'	EM	0.00200	–0.00019	0.00063	0.00062	–0.00019	0.00080	0.00026	0.00058	–0.00009	0.00010
	EV	0.03628	0.01881	0.01329	0.00986	0.00781	0.00671	0.00567	0.00495	0.00442	0.00389
	Bias	0.00200	-0.00019	0.00063	0.00062	–0.00019	0.00080	0.00026	0.00058	-0.00009	0.00010
	RMSE	0.19047	0.13716	0.11526	0.09929	0.08835	0.08193	0.07530	0.07035	0.06651	0.06235
φ_n''	EM	0.00103	0.00052	–0.00046	–0.00076	0.00039	–0.00087	0.00035	0.00002	–0.00010	0.00021
	EV	0.03340	0.01791	0.01250	0.00971	0.00767	0.00644	0.00558	0.00489	0.00445	0.00386
	Bias	0.00103	0.00052	–0.00046	–0.00076	0.00039	–0.00087	0.00035	0.00002	–0.00010	0.00021
	RMSE	0.18274	0.13383	0.11181	0.09855	0.08756	0.08023	0.07473	0.06994	0.06672	0.06216
Skewed distribution											
φ_n	EM	–0.00030	–0.00031	–0.00045	–0.00010	0.00115	–0.00072	0.00248	–0.00111	0.00014	–0.00139
	EV	0.04708	0.02099	0.01391	0.01027	0.00817	0.00675	0.00596	0.00512	0.00454	0.00401
	Bias	–0.00030	–0.00031	–0.00045	–0.00010	0.00115	–0.00072	0.00248	–0.00111	0.00014	–0.00139
	RMSE	0.21698	0.14488	0.11794	0.10135	0.09037	0.08213	0.07721	0.07158	0.06739	0.06331
φ_n'	EM	–0.00131	0.00198	–0.00138	–0.00084	0.00313	0.00036	–0.00008	–0.00068	0.00120	0.00005
	EV	0.03660	0.01889	0.01295	0.00971	0.00786	0.00653	0.00576	0.00489	0.00445	0.00398
	Bias	–0.00131	0.00198	-0.00138	–0.00084	0.00313	0.00036	–0.00008	–0.00068	0.00120	0.00005
	RMSE	0.19131	0.13746	0.11379	0.09854	0.08869	0.08083	0.07588	0.06990	0.06670	0.06312
φ_n''	EM	–0.00007	0.00096	0.00115	0.00000	0.00133	0.00099	–0.00005	0.00036	–0.00012	0.00054
	EV	0.03331	0.01839	0.01249	0.00966	0.00777	0.00636	0.00550	0.00485	0.00439	0.00394
	Bias	–0.00007	0.00096	0.00115	0.00000	0.00133	0.00099	–0.00005	0.00036	-0.00012	0.00054
	RMSE	0.18250	0.13560	0.11176	0.09829	0.08817	0.07974	0.07417	0.06963	0.06628	0.06276

The data in the table demonstrate that, under both heavy-tailed and skewed data, the estimated means and variances approach their true values significantly as sample size increases. Together, the quantitative and visual results indicate that bias decreases with growing sample size, albeit in a fluctuating manner, while all RMSEs gradually diminish. Notably, the RMSEs of our proposed asymptotic representations are smaller than those of the original φ_n, eventually converging to similar levels. These findings align with previous analyses and demonstrate the robustness of our method across different data distributions.

A.2.2. Additional simulations of asymptotic behavior for the proposed methods

For this subsection, we adopt three methods to simulate the asymptotic behavior of φ_n, φ_n', and φ_n''. The first two approaches estimate their empirical density functions and cumulative distribution functions (CDFs), utilizing the heavy-tailed and skewed data generation methods from Section A.2.1. Simulation results under these two distributions are presented in Figures A3–A6. The third method involves conducting 1000 repetitions of the two-sample Kolmogorov-Smirnov (KS) test, calculating the proportion of rejecting the null hypothesis for six combinations under two significance levels ($\alpha = 0.01$ and $\alpha = 0.05$). The simulation results are summarized in Table A2. All other settings remain consistent with Section 3.2 of the main paper.

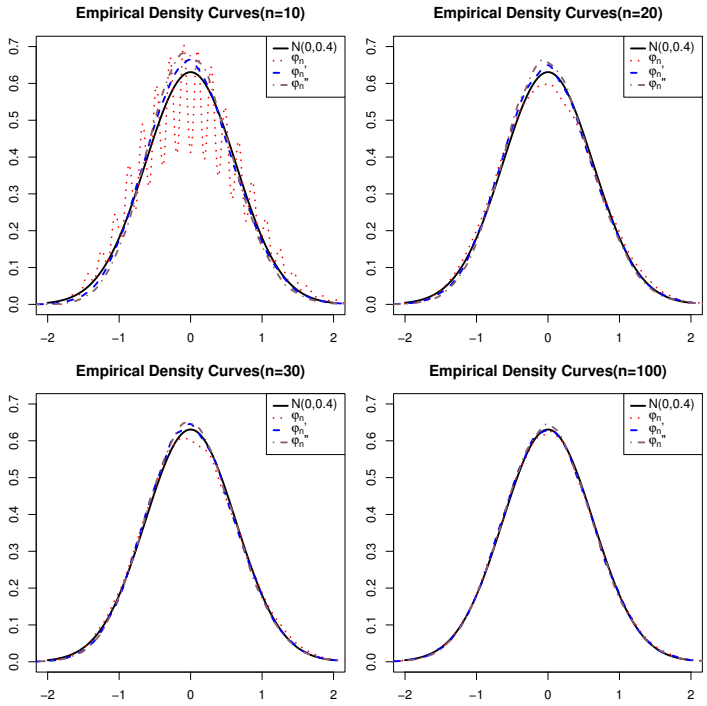

Figure A3. Empirical density curves of $\sqrt{n}\varphi_n$, $\sqrt{n}\varphi_n'$, and $\sqrt{n}\varphi_n''$ under heavy-tailed distribution.

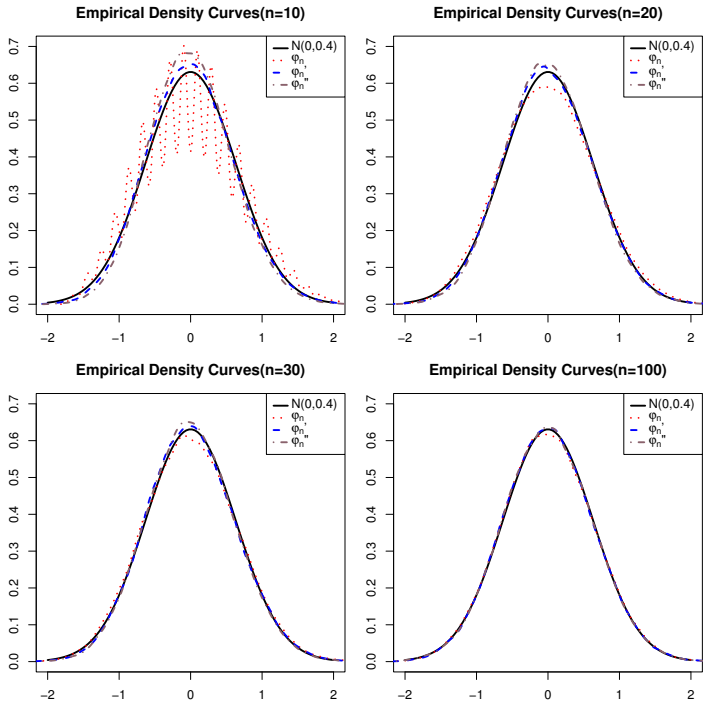

Figure A4. Empirical density curves of $\sqrt{n}\varphi_n$, $\sqrt{n}\varphi_n'$, and $\sqrt{n}\varphi_n''$ under skewed distribution.

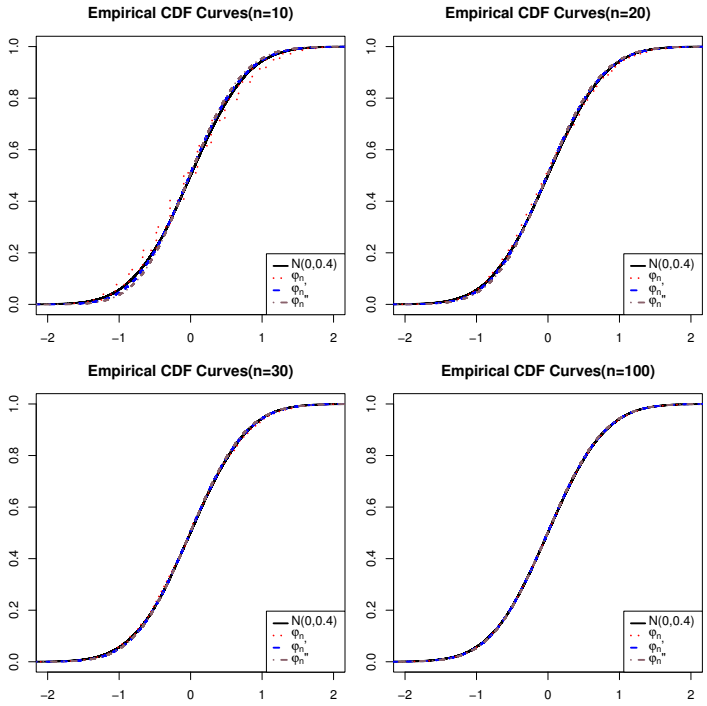

Figure A5. Empirical CDF curves of $\sqrt{n}\varphi_n$, $\sqrt{n}\varphi_n'$, and $\sqrt{n}\varphi_n''$ under heavy-tailed distribution.

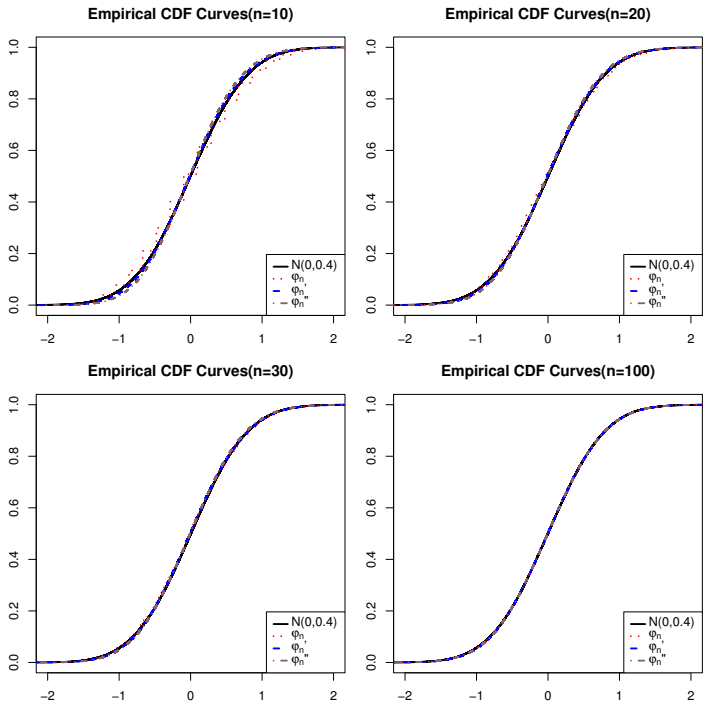

Figure A6. Empirical CDF curves of $\sqrt{n}\varphi_n$, $\sqrt{n}\varphi_n'$, and $\sqrt{n}\varphi_n''$ under skewed distribution.

The simulation figures demonstrate that results under heavy-tailed and skewed data distributions largely align with those from standard distributions. Once the sample size exceeds $n = 30$, all methods approach the normal distribution $N(0, 0.4)$, though the proposed asymptotic representations exhibit better approximation, with φ_n' performing optimally. For the KS test rejection rates in Table A2, combinations involving φ_n show significant rejection at very small sample sizes ($n = 10$). For our proposed methods, only the $\varphi_n'' - N(0, 0.4)$ combination exhibits slight rejection. This phenomenon diminishes as sample sizes increase ($n \geqslant 20$). Combinations involving φ_n retain some rejection capability at $n = 20$ and $n = 30$, but all combinations show no rejection once n exceeds 40. This indicates that, with increasing sample size, the distributions of our proposed asymptotic representations converge to the same normal distribution as the original Spearman's footrule correlation coefficient.

Table A2. Rejection rates of KS test at various significance levels for six combinations.

	$n = 10$	$n = 20$	$n = 30$	$n = 40$	$n = 50$	$n = 60$	$n = 70$	$n = 80$	$n = 90$	$n = 100$
					$\alpha = 0.01$					
$\varphi_n - N(0, 0.4)$	0.923	0.107	0.035	0.018	0.016	0.012	0.016	0.014	0.014	0.007
$\varphi_n' - N(0, 0.4)$	0.024	0.018	0.012	0.006	0.010	0.012	0.012	0.005	0.012	0.010
$\varphi_n'' - N(0, 0.4)$	0.038	0.014	0.017	0.016	0.005	0.010	0.011	0.009	0.013	0.005
$\varphi_n - \varphi_n'$	0.981	0.133	0.042	0.021	0.023	0.020	0.010	0.009	0.009	0.008
$\varphi_n - \varphi_n''$	1.000	0.175	0.041	0.022	0.016	0.016	0.021	0.011	0.009	0.011
$\varphi_n' - \varphi_n''$	0.017	0.013	0.012	0.008	0.005	0.007	0.008	0.007	0.010	0.009
					$\alpha = 0.05$					
$\varphi_n - N(0, 0.4)$	1.000	0.308	0.132	0.092	0.076	0.059	0.063	0.058	0.054	0.048
$\varphi_n' - N(0, 0.4)$	0.078	0.070	0.070	0.051	0.051	0.053	0.061	0.054	0.055	0.052
$\varphi_n'' - N(0, 0.4)$	0.162	0.078	0.069	0.068	0.059	0.050	0.058	0.060	0.061	0.033
$\varphi_n - \varphi_n'$	1.000	0.377	0.146	0.093	0.075	0.074	0.058	0.060	0.070	0.049
$\varphi_n - \varphi_n''$	1.000	0.489	0.162	0.127	0.098	0.064	0.065	0.063	0.065	0.053
$\varphi_n' - \varphi_n''$	0.075	0.061	0.057	0.063	0.058	0.043	0.049	0.054	0.050	0.057

A.2.3. Simulated test size and power

To evaluate the performance of the two proposed methods in terms of test power, this subsection uses their determined critical values to simulate how closely they approximate the performance of the original test statistic. The critical values are determined using the permutation method with 1000 permutations. Four distinct models are simulated: the first model generates independent X and Y data to compute the empirical size under the null hypothesis, and the remaining three models generate data under normal distribution, heavy-tailed distribution, and mixture distribution to compute the empirical power. The sample size ranges from $n = 10$ to 100, with 1000 simulation replicates and a significance level of $\alpha = 0.05$. The models are specified as follows:

(a) (Independent) $X \sim N(0, 1)$, $Y \sim U(0, 1)$, and X is independent of Y.
(b) (Linear) $X \sim N(0, 1)$, $Y = 0.5X + 0.9\epsilon$, $\epsilon \sim N(0, 1)$, and ϵ is independent of X.
(c) (Heavy-tailed) $Z \sim N(0, 1)$, $\epsilon_1 \sim Cauchy(0, 1)$, $\epsilon_2 \sim Cauchy(0, 1)$, $X = Z + 0.6\epsilon_1$, $Y = Z + 0.6\epsilon_2$, with ϵ_1, ϵ_2, and Z mutually independent.

(d) (Mixture) $X \sim N(0,2)$, $E \sim Ber(0.25)$, $Z \sim N(0,2)$, $Y = (1-E)Z+EX$, with X, E, and Z mutually independent.

All simulation results are presented in Figure A7. As shown in Figure A7(a), the empirical size of our two proposed asymptotic representations approaches the significance level of 0.05 for sample sizes $n > 30$, and their performance converges to that of the original Spearman's footrule correlation coefficient. Similarly, Figure A7(b)–(d) demonstrate that regardless of whether X and Y exhibit a linear relationship or follow a heavy-tailed distribution or a mixture distribution, their empirical power increasingly aligns with that of the footrule φ_n as n increases.

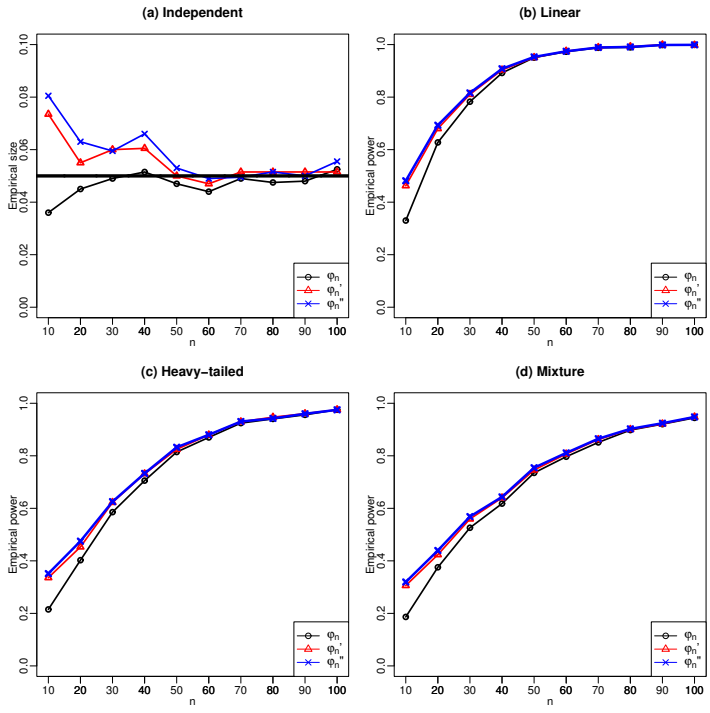

Figure A7. The empirical size and power of tests for φ_n, φ_n', and φ_n'' under various distributions.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

This work was supported in part by the Zhejiang Provincial Natural Science Foundation of China under Grant LZ23A010003.

Conflict of interest

The authors declare there is no conflict of interest.

References

1. C. Spearman, 'Footrule' for measuring correlation, *Br. J. Med. Psychol.*, **2** (1906), 89–108. https://doi.org/10.1111/j.2044-8295.1906.tb00174.x
2. D. K. Bukovšek, B. Mojškerc, On the exact region determined by Spearman's footrule and Gini's gamma, *J. Comput. Appl. Math.*, **410** (2022), 114212. https://doi.org/10.1016/j.cam.2022.114212
3. C. Chen, W. Xu, W. Zhang, H. Zhu, J. Dai, Asymptotic properties of Spearman's footrule and Gini's gamma in bivariate normal model, *J. Franklin Inst.*, **360** (2023), 9812–9843. https://doi.org/10.1016/j.jfranklin.2023.07.024
4. A. Pérez, M. Prieto-Alaiz, F. Chamizo, E. Liebscher, M. Úbeda-Flores, Nonparametric estimation of the multivariate Spearman's footrule: A further discussion, *Fuzzy Set Syst.*, **467** (2023), 108489. https://doi.org/10.1016/j.fss.2023.02.010
5. R. B. Nelsen, *An Introduction to Copulas*, 2^{nd} edition, Springer, 2006. https://doi.org/10.1007/0-387-28678-0
6. B. S. Kim, S. Y. Rha, G. B. Cho, H. C. Chung, Spearman's footrule as a measure of cDNA microarray reproducibility, *Genomics*, **84** (2004), 441–448. https://doi.org/10.1016/j.ygeno.2004.02.015
7. R. Fagin, R. Kumar, D. Sivakumar, Comparing top k lists, *SIAM J. Discrete Math.*, **17** (2003), 134–160. https://doi.org/10.1137/S0895480102412856
8. S. Mikki, Comparing Google Scholar and ISI Web of Science for earth sciences, *Scientometrics*, **82** (2010), 321–331. https://doi.org/10.1007/s11192-009-0038-6
9. F. Iorio, R. Tagliaferri, D. di Bernardo, Identifying network of drug mode of action by gene expression profiling, *J. Comput. Biol.*, **16** (2009), 241–251. https://doi.org/10.1089/cmb.2008.10TT
10. S. Lin, J. Ding, Integration of ranked lists via cross entropy Monte Carlo with applications to mRNA and microRNA studies, *Biometrics*, **65** (2009), 9–18. https://doi.org/10.1111/j.1541-0420.2008.01044.x
11. V. Vitelli, Ø. Sørensen, M. Crispino, A. Frigessi, E. Arjas, Probabilistic preference learning with the Mallows rank model, *J. Mach. Learn. Res.*, **18** (2018), 1–49.
12. A. W. van der Vaart, *Asymptotic Statistics*, Cambridge university press, 1998. https://doi.org/10.1017/CBO9780511802256
13. J. E. Angus, A coupling proof of the asymptotic normality of the permutation oscillation, *Probab. Eng. Inf. Sci.*, **9** (1995), 615–621. https://doi.org/10.1017/S0269964800004095
14. S. Chatterjee, A new coefficient of correlation, *J. Am. Stat. Assoc.*, **116** (2021), 2009–2022. https://doi.org/10.1080/01621459.2020.1758115
15. H. Shi, M. Drton, F. Han, On the power of Chatterjee's rank correlation, *Biometrika*, **109** (2022), 317–333. https://doi.org/10.1093/biomet/asab028
16. Z. Lin, F. Han, On boosting the power of Chatterjee's rank correlation, *Biometrika*, **110** (2023), 283–299. https://doi.org/10.1093/biomet/asac048

17. L. Xia, R. Cao, J. Du, X. Chen, The improved correlation coefficient of Chatterjee, *J. Nonparam. Stat.*, **37** (2025), 265–281. https://doi.org/10.1080/10485252.2024.2373242

18. P. Diaconis, R. L. Graham, Spearman's footrule as a measure of disarray, *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, **39** (1977), 262–268. https://doi.org/10.1111/j.2517-6161.1977.tb01624.x

19. P. K. Sen, I. A. Salama, The Spearman footrule and a Markov chain property, *Stat. Probab. Lett.*, **1** (1983), 285–289. https://doi.org/10.1016/0167-7152(83)90046-9

20. D. C. Kleinecke, H. K. Ury, L. F. Wagner, *Spearman's Footrule—An Alternative Rank Statistic*, 1962. Available from: https://apps.dtic.mil/sti/html/tr/AD0403502/.

21. W. Hoeffding, A combinatorial central limit theorem, *Ann. Math. Stat.*, **22** (1951), 558–566. https://doi.org/10.1214/aoms/1177729545

22. X. Shi, M. Xu, J. Du, Max-sum test based on Spearman's footrule for high-dimensional independence tests, *Comput. Stat. Data Anal.*, **185** (2023), 107768. https://doi.org/10.1016/j.csda.2023.107768

23. L. H. Y. Chen, X. Fang, Q. Shao, From Stein identities to moderate deviations, *Ann. Probab.*, **41** (2013), 262–293. https://doi.org/10.1214/12-AOP746

24. X. Shi, W. Zhang, J. Du, E. Kwessi, Testing independence based on Spearman's footrule in high dimensions, *Commun. Stat. Theory Methods*, **54** (2025), 2360–2377. https://doi.org/10.1080/03610926.2024.2369313

25. C. G. Small, *Expansions and Asymptotics for Statistics*, Chapman and Hall/CRC, 2010. https://doi.org/10.1201/9781420011029

26. G. Schröer, D. Trenkler, Exact and randomization distributions of Kolmogorov-Smirnov tests two or three samples, *Comput. Stat. Data Anal.*, **20** (1995), 185–202. https://doi.org/10.1016/0167-9473(94)00040-P

27. D. Harrison Jr, D. L. Rubinfeld, Hedonic housing prices and the demand for clean air, *J. Environ. Econ. Manage.*, **5** (1978), 81–102. https://doi.org/10.1016/0095-0696(78)90006-2

28. M. R. Kosorok, *Introduction to Empirical Processes and Semiparametric Inference*, Springer, 2008. https://doi.org/10.1007/978-0-387-74978-5

29. A. C. Berry, The accuracy of the Gaussian approximation to the sum of independent variates, *Trans. Am. Math. Soc.*, **49** (1941), 122–136. https://doi.org/10.1090/S0002-9947-1941-0003498-3

30. R. J. Serfling, *Approximation Theorems of Mathematical Statistics*, John Wiley & Sons, 1980.

AIMS Press