Journal of Modern Applied Statistical Methods

Volume 20 | Issue 2 Article 5

Estimation of Correlation Confidence Intervals Via the Bootstrap: Non Normal Distributions

John Mart V. DelosReyes
Old Dominion University, United States,
https://orcid.org/0000-0002-7230-6506

Miguel A. Padilla
Old Dominion University, United States,
https://orcid.org/0000-0002-7381-3518

Recommended Citation

John Mart V. DelosReyes, Miguel A. Padilla (2021). Estimation of Correlation Confidence Intervals Via the Bootstrap: Non Normal Distributions. Journal of Modern Applied Statistical Methods, 20(2), https://doi.org/10.56801/Jmasm.V20.i2.5

Journal of Modern Applied Statistical Methods
2021, Vol. 20, No. 2,
Doi: 10.56801/Jmasm.V20.i2.5

ISSN 1538 – 9472

Estimation of Correlation Confidence Intervals Via the Bootstrap: Non Normal Distributions

John Mart V. DelosReyes
Old Dominion University, United States

Miguel A. Padilla
Old Dominion University, United States

Confidence intervals (CIs) for the correlation were investigated with interest on how bootstrap CIs perform with non-normal distribution pairings across correlation magnitudes. The bootstrap CIs had acceptable performance if one of the variables was normal or if sample size was greater than 200 if the paired variables were non-normal.

Keywords: Correlation, Bootstrap, Non normal, Simulation.

1. Introduction

The Pearson product-moment correlation (henceforth referred to as the correlation) is a popular and important statistic that captures the linear relationship between two variables. The correlation has been used for over 100 years and has desirable properties that allow it to contribute as a foundational piece to other statistics and statistical models like the t-test, regression, MANOVA, etc. (Hald, 2007). One such useful property is that the correlation has a range of $-1 \le \rho \le 1$. The utility of this range is that it allows for a standardized method of interpreting the linear relationships between a pair of variables.

The correlation is generally well understood when $\rho = 0$, but research about the correlation across its range has been inconsistent. This is due to the correlation having a mutable distribution based on its magnitude. In fact, the correlation only has a symmetric distribution when $\rho = 0$ and is skewed otherwise (See Figure 1). This means that although the breadth of research on $\rho = 0$ is useful, it may not always be applicable to cases where $\rho \ne 0$. One such important situation is in research applications where the presence of an effect is of interest (i.e., $\rho \ne 0$). In addition, precision is important for research applications and can be demonstrated via confidence intervals (CIs). However, the mutable nature of the distribution of the correlation makes CI estimation difficult as well. Attempts have been made to make correlation CIs possible but have yielded mixed results.

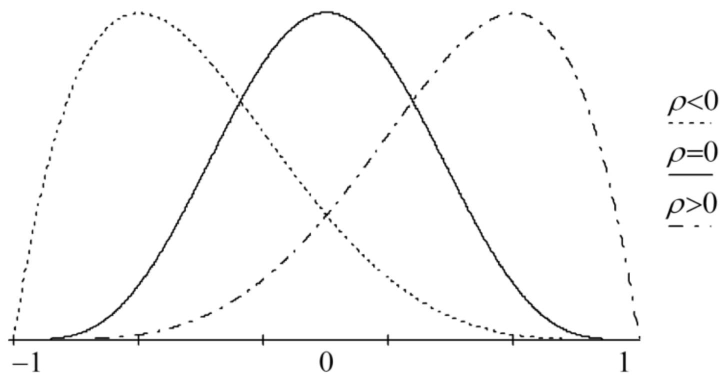

Figure 1. The range restriction effect of the distribution of the correlation.

Fisher (1915) first developed a correlation CI for when $\rho \neq 0$ via the Fisher z-transformation and it was subsequently evaluated by Zeller and Levine (1974) under several simulation conditions. The authors investigated the (a) distribution shape, (b) correlation strength ($\rho = 0$, .32, .71, .95), and (c) sample size ($n = 15$, 50, 100) with $\alpha = .01$ and .05. For distribution shape, the distribution was the same for both variables and the authors investigated the normal, uniform, J, bimodal, and a leptokurtic, but no skewness and kurtosis details were provided for these distributions. Results were based on 3,000 simulation replications. One consistent finding was that the correlation estimates slightly underestimated the true correlation, but this was ultimately negligible when $n > 15$. In addition, the Fisher z-transformation correlation CI was shown to be robust to the mild non-normal distributions investigated (e.g., uniform, J, bimodal, and leptokurtic). These results were consistent for $\alpha = .01$ and .05.

Additional research on the robustness of the Fisher z-transformation CI for the correlation was conducted by Berry and Mielke (2000). The authors investigated (a) distribution shape, (b) correlation strength ($\rho = 0$, .4, .6, .8), and (c) sample size ($n = 10$, 20, 40, 80) with $\alpha = .10$, .05, and .01. For distribution shape, the distribution was the same for both variables and the authors investigated the normal, 3 generalized logistic, and 3 symmetric kappa distributions. The generalized logistic distributions were defined by

$$f(x) = \left(\frac{e^{\theta x}}{\theta}\right)^{1/\theta} \left(1 + \frac{e^{\theta x}}{\theta}\right)^{-(\theta+1)/\theta}, \tag{1}$$

where $\theta = 1$, .1, .01. In this context, $\theta > 1$ results in negative skew and $\theta < 1$ results in positive skew. The symmetric kappa distribution was defined by

$$f(x) = .5\lambda^{-1/\lambda} \left(1 + \frac{|x|^{\lambda}}{\lambda}\right)^{-(\lambda+1)/\lambda}, \tag{2}$$

where $\lambda = 2,\ 3,\ 25$. In this context, $\lambda = 2$ represents a distribution similar to a t distribution with 2 degrees of freedom, $\lambda = 3$ is a heavy tailed distribution, and $\lambda = 25$ is similar to a uniform distribution. The authors provided no skewness and kurtosis details for these distributions. Results were based on 1,000,000 simulation replications. The results showed that Fisher z-transformation CIs had appropriate coverage probability when $\rho = 0$ and $\alpha = .01$ and .05 and for all distribution shapes. However, the coverage probability was consistently underestimated when distribution shapes were non-normal and $\rho \neq 0$. Furthermore, these problems were not remedied with increased sample size but made more severe. These results were consistent for $\alpha = .01,\ .05$, and .10.

The discrepancy in the literature about efficacy of the Fisher z-transformation CI indicates a need for investigation on other correlation CI methods. Additionally, there is interest on the effect non-normal data has on the CI estimates as such data are common in research applications (Blanca, Arnau, Lopez-Montiel, Bono, & Bendayan, 2013). This has led to investigating the viability of the bootstrap for CI estimation as the bootstrap does not have distributional assumptions (Efron & Tibshirani, 1993). Two bootstrap CIs of interest are the Percentile Bootstrap (PB) and the Bias-Corrected and Accelerated (BCa) CIs.

An early application of the bootstrap CI for the correlation was presented by Lunneborg (1985). Lunneborg explored the potential of the bootstrap for estimating correlation CIs using SAT verbal and math scores from a pseudorandom sample of 25 college freshman. In this study, the PB CI, based on 500 bootstrap samples, was compared to the Fisher z-transformation CI with a an undisclosed α. The SAT scores were used because (a) they are real data, (b) the verbal and math scores are known to be bivariate normal, and (c) the verbal-math ρ is large enough such that useful CIs can be estimated for the small pseudorandom sample ($n = 25$). The CIs from both methods were similar under bivariate normality.

In contrast, another early simulation study of the bootstrap CI for the correlation was conducted by Rasmussen (1987). Of interest was the impact of a non-normal distribution on the bootstrap CI for testing $\rho = 0$. This was done by investigating (a) distribution shape (normal and lognormal) and (b) sample size ($n = 5,\ 15,\ 30,\ 60$) with $\alpha = .01$ and .05. For distribution shape, the distribution for the variable pairings took the following two forms: normal-normal or normal-lognormal. The PB CI, based on 500 bootstrap samples, was compared to the Fisher z-transformation CI. Results were based on 1,000 simulation replications. However, the results ran counter to Lunneborg's research (1985) as they showed a lack of parity between the Fisher z-transformation and the PB CI. In this case, the PB CIs demonstrated an overall increase in type I error rate and narrower CIs compared to the Fisher z-transformation CI under all conditions, including when the variable pairing was normal-normal. Going from $\alpha = .05$ to $\alpha = .01$ further highlighted this issue. However, Rasmussen noted that the situation did improve with larger sample sizes but was not able to explore this beyond n = 60 due to costs in computational power at the time.

In two recent studies, Padilla and Veprinksy (2012, 2014) developed PB and BCa CIs for the deattenuated correlation. An estimated correlation can become weaker (attenuated) than what may be true in the population due to measurement error (Spearman, 1904). Spearman (1904) developed a correction for this attenuation known as the deattenuated (or disattenuated) correlation (Muchinsky, 1996), but research on this correlation and its corresponding CIs is rare. Padilla and Veprinksy addressed this gap by investigating the bootstrap CIs for the deattenuated correlation under four simulation conditions: the (a) distribution shape, (b) strength of the correlation ($\rho = .10,\ .20,\ .30,\ .40,\ .50$), (c) reliability of both variables in the correlation ($\rho_{ij} = .50,\ .60,\ .70,\ .80,\ .90$), (d) and sample size ($n = 50,\ 100,\ 150,\ 200,\ 250,\ 300$) with $\alpha = .05$. All bootstrap CIs were based on 2,000 bootstrap samples. For distribution shape, both variables had the same distribution from the following distributions investigated:

- normal (skewness = 0, kurtosis = 0)
- uniform (skewness = 0, kurtosis = -1.20)
- triangular (skewness = 0, kurtosis = -0.60)
- beta (skewness = -0.85, kurtosis = 0.22)
- Laplace (skewness = 0, kurtosis = 3)
- Pareto (skewness = 2.81, kurtosis = 14.83)

Results were based on 1,000 simulation replications. Overall, the PB and BCa CIs had good coverage under all simulation conditions with negligible differences between the two CIs. Even so, the BCa CI tended to have slightly better coverage than the PB CI. The one exception was that neither CI performed well with the Pareto distribution. However, the Pareto distribution investigated was skewed and highly peaked (kurtosis = 14.83). Such distributions have range restrictions, and it is well known that distributions with range restrictions attenuate the correlation due to less variability.

In a subsequent study, Bishara and Hittner (2017) investigated several CIs for the correlation. Of interest was the impact of various types and combinations of distributions on the correlation CIs. The following correlation CIs were investigated: the 1) Fisher z-transformation, 2) Spearman rank-order with Fieller's SE, 3) Spearman rank-order with Wright's SE, 4) Box-Cox transformation, 5) ranked inverse normal transformation (RIN), 6) nonparametric bootstrap, 7) nonparametric bootstrap with asymptotic adjustment (AA), 8) nonparametric bootstrap BCa, 9) observed imposed bootstrap, 10) observed imposed bootstrap with AA, and 11) observed imposed bootstrap with BCa. All bootstrap CIs were based on 9,999 bootstrap samples. The performance of the CIs was investigated through a simulation with the following four conditions: (a) distribution shape, (b) distribution pairing, (d) correlation strength ($\rho = 0,\ .5$), and (c) sample size ($n = 10,\ 20,\ 40,\ 80,\ 160$) with $\alpha = .05$. The distributions investigated were a result of a combination of skewness ($\gamma_1 = -4,\ -3,\ -2,\ -1,\ 0,\ 1,\ 2,\ 3,\ 4$) and kurtosis ($\gamma_2 = -1,\ 0,\ 2,\ 4,\ 6,\ 8,\ 10,\ 20,\ 30,\ 40$)

whose feasibility was limited by the lower bound of kurtosis being determined by the squared skewness

$$\gamma_2 \geq \gamma_1^2 - 2 \tag{3}$$

This resulted in 46 skewness and kurtosis combinations being investigated. The distribution pairings investigated either had both variables come from the same distribution or had one variable come from a normal distribution and the other from a non-normal distribution. Overall, 920 simulation scenarios were investigated. Results were based on 10,000 simulation replications.

The primary findings were that the RIN followed by the Spearman rank-order with Fieller's SE CIs had the best performance when data were non-normal. Of the remaining CI methods, only the observed imposed bootstrap with BCa had good enough performance when data were non-normal. However, it tended to exceed 95% coverage by generating somewhat long CIs. The advantage it has is that it keeps the correlation in the scale of the original variables. This is not the case for the RIN and Spearman rank-order with Fieller's SE as both transform the original variables. All the remaining methods did not have good CI coverage when data were non-normal with the Fisher z-transformation CI having the least favorable performance, and the situation was made worse by increasing the sample size when $\rho = .5$.

When the variable pairing included a normal distribution, all CIs generally performed better. However, the transformation methods still outperformed the bootstrap methods in this case. In this situation, the only bootstrap CI methods that were comparable to the transformation methods' performance were the observed imposed bootstrap with AA and observed imposed bootstrap with BCa.

Given previous mixed findings, the goal of the current study is to clearly understand under which conditions the correlation CIs from previous studies are robust. The main conditions of interest are correlation magnitude and distribution of data. The following CI methods were investigated.

1.1 Confidence Interval Estimation

Fisher z-transformation CI. CIs constructed via the Fisher z-transformation (1915) start with

$$z = \left(\frac{1}{2}\right)\ln\left(\frac{1+r}{1-r}\right) \tag{4}$$

and assume that the transformed correlation is a standard normal variate. This allows for a $100(1-\alpha)\%$ CI to be defined as

$$z \pm z_{\alpha/2} \times SE(z), \tag{5}$$

where

$$SE(z) = \frac{1}{\sqrt{n-3}} \tag{6}$$

is the standard error of z.

Spearman Rank-Order with Fieller's SE CI. CIs constructed via the Spearman Rank-Order (1904) have x_{i1} and x_{i2} separately transformed into ascending ranks. From here, the correlation is computed. The CI is then constructed by the Fisher z-transformation defined by equations $4-5$, where

$$SE(z) = \frac{1.03}{\sqrt{n-3}} \tag{7}$$

is Fieller's (1957) standard error.

Ranked Inverse Normal Transformation (RIN) CI. In this method, x_{i1} and x_{i2} are separately transformed through Bliss's (1967) rankit transformation defined as

$$f(x) = \phi^{-1}\left(\frac{x_r - .5}{n}\right), \tag{8}$$

where ϕ^{-1} is the inverse cumulative distribution function and x_r is the ascending rank of each x_i value. The CI is then computed through Fisher's z-transformation with equations $4-6$. The utility of this process is that it will convert the data into an approximately normal distribution.

Bootstrap CI for the Correlation. The bootstrap for a pair of variables x and y can be outlined in three steps. Suppose the observed data is $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$, where $\mathbf{x}_i = (x_i, y_i)$ is the pair of variables. First, obtain the bth bootstrap sample with replacement from X; i.e., $\mathbf{X}^{(b)} = (\mathbf{x}_1^{(b)}, \mathbf{x}_2^{(b)}, \ldots, \mathbf{x}_n^{(b)})$. Second, the bth estimate of the correlation from $\mathbf{X}^{(b)}$ is computed as

$$r_{xy}^{(b)} = \frac{C_{xy}^{(b)}}{S_x^{(b)} S_y^{(b)}} \tag{9}$$

and stored where $C_{xy}^{(b)}$, $S_x^{(b)}$, and $S_y^{(b)}$ are the covariance and corresponding SDs for the bth sample, respectively. Third, compile the stored estimates $r_{xy}^{(1)}, r_{xy}^{(2)}, \ldots, r_{xy}^{(B)}$ to create the empirical sampling distribution (ESD) of r_{xy} for $b = 1, 2, \ldots, B$ bootstrap samples. The ESD can then be summarized to obtain statistical quantities for inference about r_{xy}.

The quantities of interest here are the PB and BCa CIs. The PB CI is estimated by obtaining the $\alpha/2$ and $1-\alpha/2$ percentiles from the r_{xy} ESD where α is the significance levels. The BCa CI follows the same process as the PB CI, but the bounds are adjusted for the bias and skewness (or acceleration) of the r_{xy} ESD. See Efron and Tibshirani (1993) for details.

2. Methodology

2.1 Data Generation

A Monte Carlo simulation was used to investigate and compare the properties of the correlation CI methods under different simulation conditions. This simulation was structured in a 6 (corr. magnitude) × 11 (sample size) × 23 (distribution pairings) simulation design for a total of 1518 conditions. For each simulation condition, 1,000 data replicants were obtained. Data were simulated in three steps as follows. First, generate normal and non-normal data according to Headrick (2002) as follows

$$\begin{bmatrix} x_i \\ x_j \end{bmatrix} \sim \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho_{ij} \\ \rho_{ij} & 1 \end{bmatrix} \right), \tag{10}$$

where

$$x_i = c_{0i} + c_{1i} z_i + c_{2i} z_i^2 + c_{3i} z_i^3 + c_{4i} z_i^4 + c_{5i} z_i^5, \tag{11}$$

$$x_j = c_{0j} + c_{1j} z_j + c_{2j} z_j^2 + c_{3j} z_j^3 + c_{4j} z_j^4 + c_{5j} z_j^5, \tag{12}$$

$$\begin{bmatrix} z_i \\ z_j \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho_{ij}^* \\ \rho_{ij}^* & 1 \end{bmatrix} \right), \tag{13}$$

$c_0, ..., c_5$ are constants, and ρ_{ij}^* is the intermediate correlation. Second, estimate the correlation CIs for each data replication in equation 10. Third, determine if the CIs contain the population correlation (ρ). The following simulation conditions were investigated.

2.2 Conditions

Sample Size (n). Sample size was included because CI estimation is impacted by sample size. The following sample sizes were investigated: n = 20, 30, 40, 50, 100, 150, 200, 250, 300, 350, 400 as they draw parity with the previous research on the correlation CIs (Bishara & Hittner, 2017; Padilla & Veprinsky, 2012, 2014). The granularity in the lower end of the sample sizes is due to previous research suggesting that bootstrap CIs cease to function if sample size is too small (Rasmussen, 1987).

Correlation Magnitude (ρ). Correlation magnitude was considered because the distribution of the correlation changes when $\rho \neq 0$ and becomes more skewed as it approaches ±1 (see Figure 1). Cohen (1988) advises that correlation coefficients equal to .10, .30, and .50 represent small, moderate, and strong correlations, respectively. The present investigation utilized correlation coefficients ranging from .00 to .50 in increments of .10 to reflect Cohen's standards.

Distribution Pairings (x_{ij}). The degree of non-normality in the distribution pairings was investigated because the Fisher method assumes multivariate normality while the bootstrap methods do not. Additionally, non-normal data is common in applied settings and there is a lack of consensus in the literature on how to approach the correlation when non-normality occurs. The selected distributions fall into the general categories of symmetric and non-symmetric and were investigated in previous research (Bishara & Hittner, 2017; Padilla & Veprinsky, 2012, 2014). The symmetric distributions were as follows: normal, triangular, uniform, and Laplace. The non-symmetric distribution were as follows: beta ($\alpha = 4$, $\beta = 1.25$), beta ($\alpha = 4$, $\beta = 1.5$), chi-square ($df = 16$), chi-square ($df = 4$), chi-square ($df = 3$), chi-square ($df = 2$), chi-square ($df = 1$), and Pareto. Table 1 gives the constants used to generate these distributions and Figure 2 displays these distributions.

Investigation of the distributions also considered the pairwise nature of the variables involved in the correlation. This dictated four main types of distribution pairings. In the first pairing, the variables had the same symmetric distribution (e.g., both variables were uniform). Similarly, in the second pairing, the variables had the same non-symmetric distribution (e.g., both variables were Pareto). In the third pairing, one variable was always normal and the other was symmetric (e.g., one variable was normal and the other Laplace). In the third pairing, a normal-normal paring was not included. In the fourth pairing, one variable was always normal and the other was non-symmetric (e.g., one variable was normal and the other Pareto).

2.3 Criteria for Evaluating CIs

The $100(1-\alpha)\%$ CI coverage was assessed using Bradley's (1978) criterion. The criterion is defined as $1-1.5\alpha \le 1-\alpha^* \le 1-0.5\alpha$, where α* is the true probability of Type I error. CI coverage is defined as the proportion of CIs that contain the true population correlation ρ. As such, acceptable coverage for $\alpha = .05$ is given by $[.925, .975]$. Bootstrap based CIs were estimated from a total of 2,000 bootstrap samples.

Table 1. Constants for Headrick's (2002) fifth-order polynomial transformation method

Distribution	Skew	Kurtosis	c_0	c_1	c_2	c_3	c_4	c_5
Normal	.000	.000	.000	1.000	.000	.000	.000	.000
Triangular	.000	-.600	.000	1.081	.000	-.029	.000	-.002
Uniform	.000	-1.200	.000	1.347	.000	-.140	.000	.002
Laplace	.000	3.000	.000	0.728	.000	.096	.000	-.002
Beta (a=4, b=1.25)	-.848	.221	.199	1.071	-.229	-.041	.010	.001
Beta (a=4, b=1.5)	-.694	.069	.163	1.089	-.187	-.044	.008	.001
Chi-Square (df=16)	.710	.750	-.117	.976	.117	.004	.000	.000
Chi-Square (df=4)	1.410	3.000	-.228	.901	.232	.015	-.001	.000
Chi-Square (df=3)	1.630	4.000	-.259	.867	.265	.021	-.002	.000
Chi-Square (df=2)	2.000	6.000	-.308	.801	.319	.034	-.004	.000
Chi-Square (df=1)	2.830	12.000	-.398	.621	.417	.068	-.006	.000
Pareto	2.811	14.828	-.346	.712	.347	.028	.000	.004

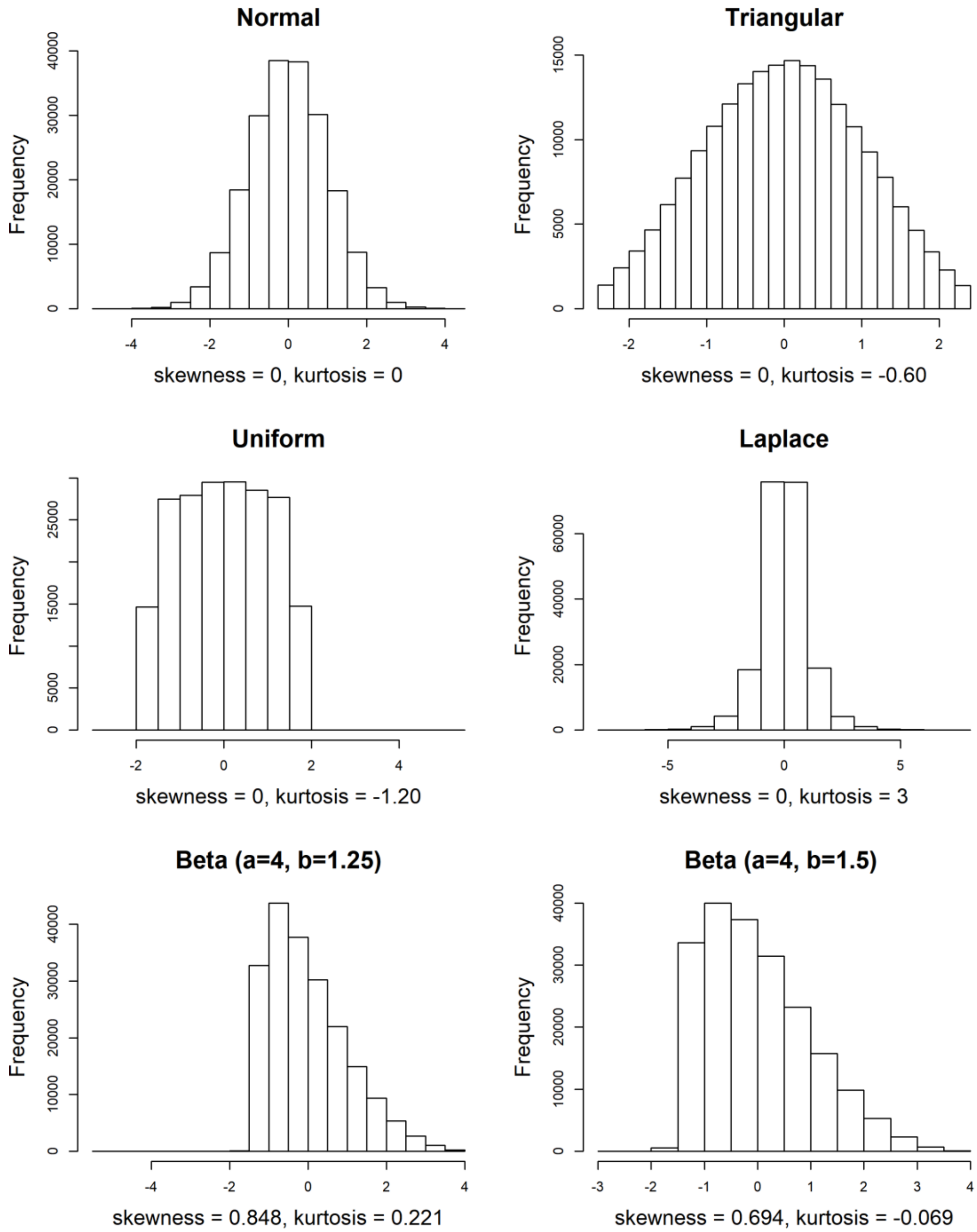

Figure 2. Distributions considered for this study.

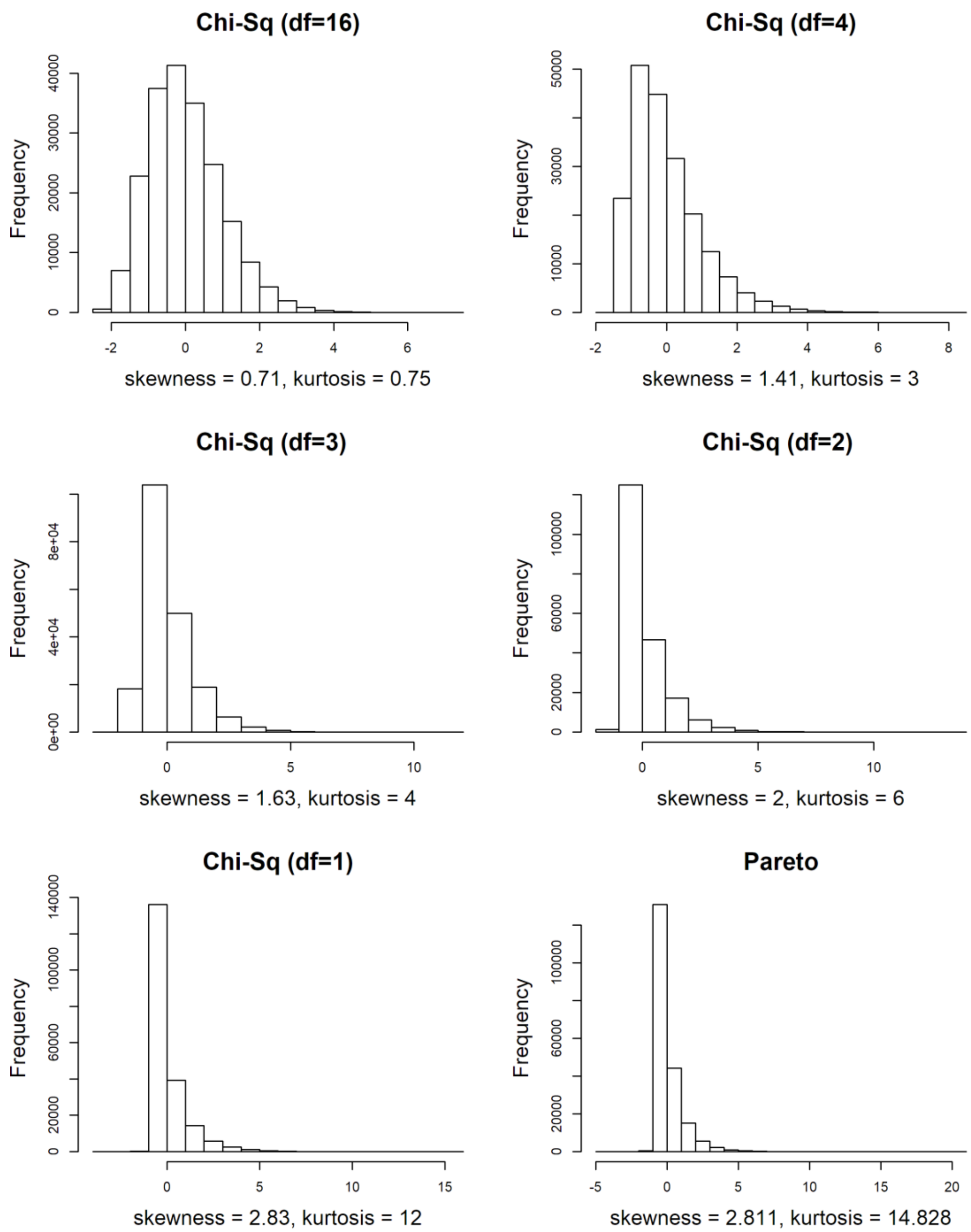

Figure 2. Continued. Distributions considered for this study.

3. Results

The major impact on CI coverage was a non-symmetric distribution paired with itself. Additionally, sample size generally had a noticeable stabilizing effect on CI coverage with minimal trend changes in CI coverage when $n > 200$. As such, the results are presented in terms of $n \leq 200$ for when a non-symmetric distribution is paired with itself. Results for the Spearman and RIN CIs are not presented as they had consistent acceptable coverage. The following results are displayed in Figures 3 to 6 and summarized in Table 2 for $n = 20 - 200$.

3.1 Sample Size of 20

The Fisher z-transformation CI had acceptable coverage when $|\text{skewness}| \leq 1.41$. However, the Fisher z-transformation CI broke down as skewness increased. The PB CI did not have acceptable coverage. The BCa had acceptable coverage when $|\text{skewness}| \leq 0.848$.

3.2 Sample Size of 30

The Fisher z-transformation CI had acceptable coverage when $|\text{skewness}| \leq 0.848$. Like before, the Fisher z-transformation CI broke down as skewness increased. The PB CI was able to maintain acceptable coverage for $|\text{skewness}| \leq 0.71$. As before, the BCa had acceptable coverage when $|\text{skewness}| \leq 0.848$.

3.3 Sample Size of 40

The Fisher z-transformation CI had acceptable coverage when $|\text{skewness}| \leq 0.848$ but broke down as skewness increased. The PB CI had acceptable coverage when $|\text{skewness}| \leq 1.41$. The BCa had acceptable coverage when $|\text{skewness}| \leq 0.848$.

3.4 Sample Size of 50

The Fisher z-transformation CI had acceptable coverage when $|\text{skewness}| \leq 0.848$ but broke down as skewness increased. The PB CI had acceptable coverage when $|\text{skewness}| \leq 1.63$. The BCa had acceptable coverage when $|\text{skewness}| \leq 0.848$.

3.5 Sample Size of 100

The Fisher z-transformation CI had acceptable coverage when $|\text{skewness}| \leq 1.41$ but broke down as skewness increased. The PB CI had acceptable coverage when $|\text{skewness}| \leq 2$. The BCa had acceptable coverage when $|\text{skewness}| \leq 1.41$.

3.6 Sample Size of 150

The Fisher z-transformation CI had acceptable coverage when $|\text{skewness}| \leq 0.848$ but broke down as skewness increased. The PB CI was generally able to maintain acceptable coverage when $|\text{skewness}| \leq 2.83$. The exceptions for the PB CI were when

$|\text{skewness}| = 1.63$ or when $|\text{skewness}| = 2.811$. The BCa CI had acceptable coverage when $|\text{skewness}| \le 1.41$.

3.7 Sample Size of 200 or More

The Fisher z-transformation CI had acceptable coverage when $|\text{skewness}| \le 0.848$ but broke down as skewness increased. The PB CI had acceptable coverage when $|\text{skewness}| \le 2.83$. The BCa CI had acceptable coverage when $|\text{skewness}| \le 2$.

3.8 Summary of Results for Sample Size of 20 to 200

In general, the Fisher z-transformation had acceptable coverage when $|\text{skewness}| \le 0.848$ for $n = 30 - 200$. The only exceptions are when $|\text{skewness}| \le 1.41$ for $n = 20$ and 100. The PB CI did not have acceptable coverage when $n = 20$. However, it had acceptable coverage with subsequent sample sizes of $n = 30 - 200$ for $|\text{skewness}| \le 0.710$ to $|\text{skewness}| \le 2.83$; i.e., increasing the sample size allowed for more deviation from normality. Finally, the BCa CI had acceptable coverage when $|\text{skewness}| \le 0.848$ for $n = 20 - 100$. However, increasing the sample size to $n = 100$ and 200 attained acceptable coverage for $|\text{skewness}| \le 1.41$ and $|\text{skewness}| \le 2$, respectively.

Table 2. Summary of Acceptable Coverage for Paired Non-Normal Distributions with Sample Size ≤ 200.

Method	n	Skewness	Kurtosis	Distribution
Fisher z-transformation	20	≤ |1.410|	≤ |3.000|	Chi-Square(df = 4)
	30	≤ |0.848|	≤ |0.221|	Beta(α = 4, β = 1.25)
	40	≤ |0.848|	≤ |0.221|	Beta(α = 4, β = 1.25)
	50	≤ |0.848|	≤ |0.221|	Beta(α = 4, β = 1.25)
	100	≤ |1.410|	≤ |3.000|	Chi-Square(df = 4)
	150	≤ |0.848|	≤ |0.221|	Beta(α = 4, β = 1.25)
	200	≤ |0.848|	≤ |0.221|	Beta(α = 4, β = 1.25)
PB	20	NA	NA	NA
	30	≤ |0.710|	≤ |0.750|	Chi-Square(df = 16)
	40	≤ |1.410|	≤ |3.000|	Chi-Square(df = 4)
	50	≤ |1.630|	≤ |4.000|	Chi-Square(df = 3)
	100	≤ |2.000|	≤ |6.000|	Chi-Square(df = 2)
	150	≤ |2.830|	≤ |12.00|	Chi-Square(df = 1)
	200	≤ |2.830|	≤ |12.00|	Chi-Square(df = 1)
BCa	20	≤ |0.848|	≤ |0.221|	Beta(α = 4, β = 1.25)
	30	≤ |0.848|	≤ |0.221|	Beta(α = 4, β = 1.25)
	40	≤ |0.848|	≤ |0.221|	Beta(α = 4, β = 1.25)
	50	≤ |0.848|	≤ |0.221|	Beta(α = 4, β = 1.25)
	100	≤ |1.410|	≤ |3.000|	Chi-Square(df = 4)
	150	≤ |1.410|	≤ |3.000|	Chi-Square(df = 4)
	200	≤ |2.000|	≤ |6.000|	Chi-Square(df = 2)

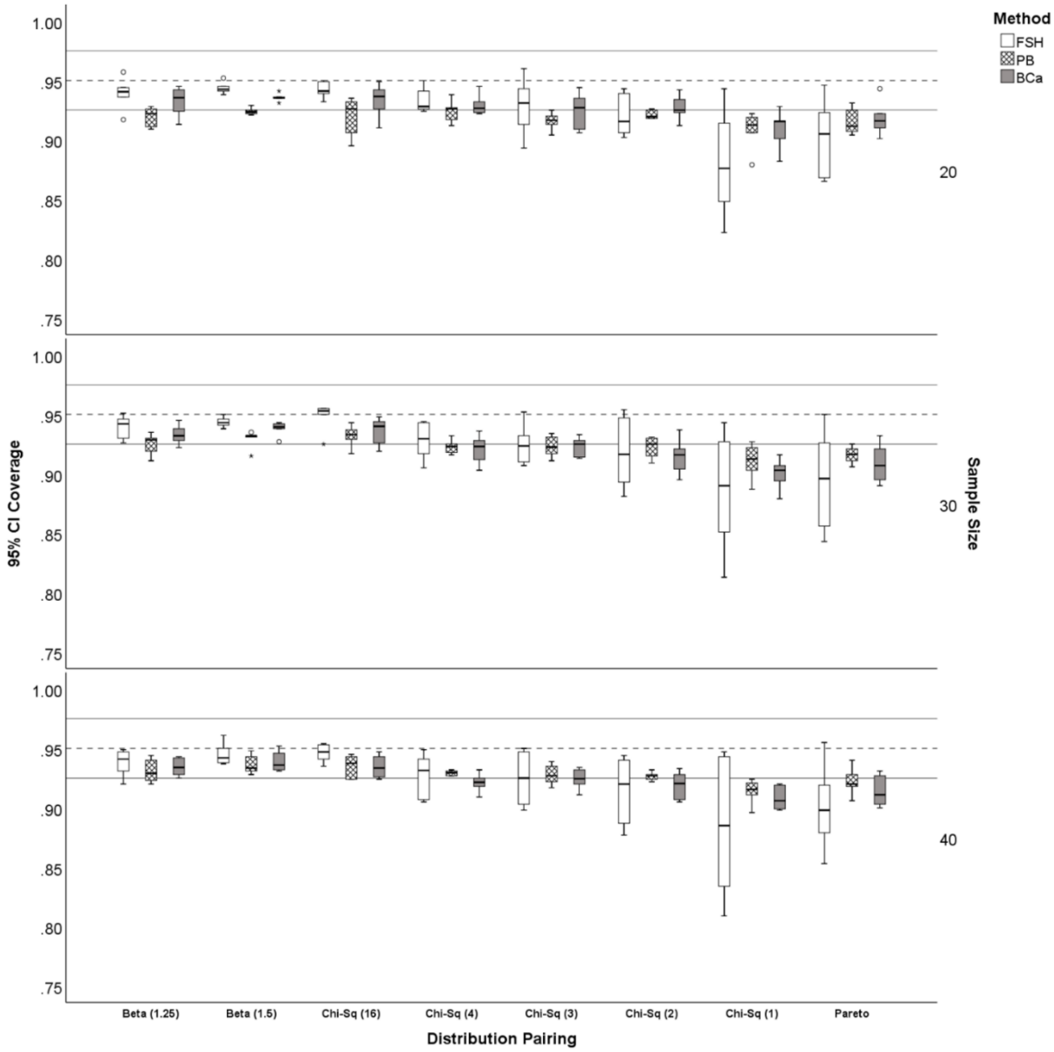

Figure 3. Distribution of 95% CI coverage for non-symmetric with non-symmetric distribution pairings by sample size of $20-40$. Fisher z-transformation (FSH), percentile bootstrap (PB), bias-corrected and accelerated bootstrap (BCa). Bootstrap methods (PB and BCa) were based on 2,000 bootstrap samples. The dashed line is at .95 and the solid lines are at the acceptable coverage of $[.925, .975]$.

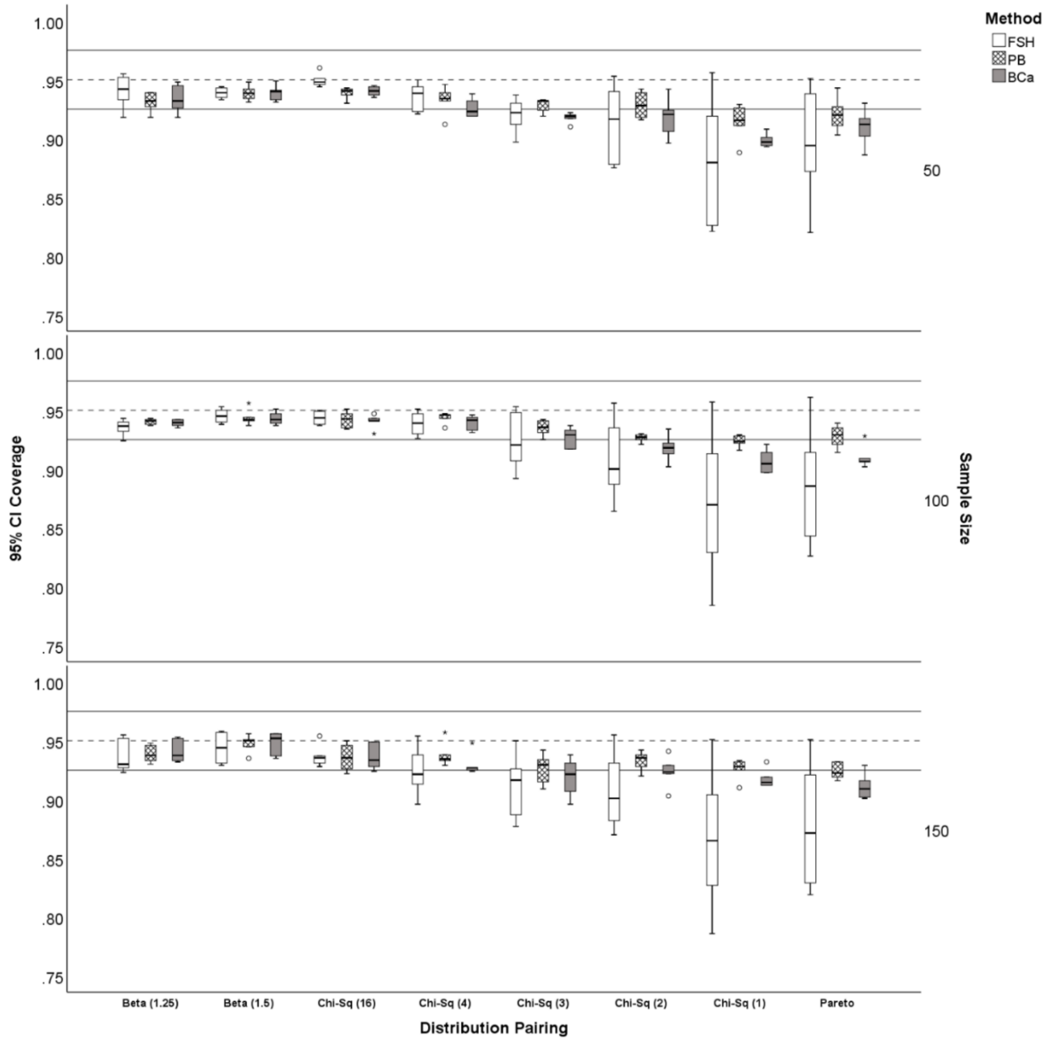

Figure 4. Distribution of 95% CI coverage for non-symmetric with non-symmetric distribution pairings by sample size of $50-150$. Fisher z-transformation (FSH), percentile bootstrap (PB), bias-corrected and accelerated bootstrap (BCa). Bootstrap methods (PB and BCa) were based on 2,000 bootstrap samples. The dashed line is at .95 and the solid lines are at the acceptable coverage of $[.925, .975]$.

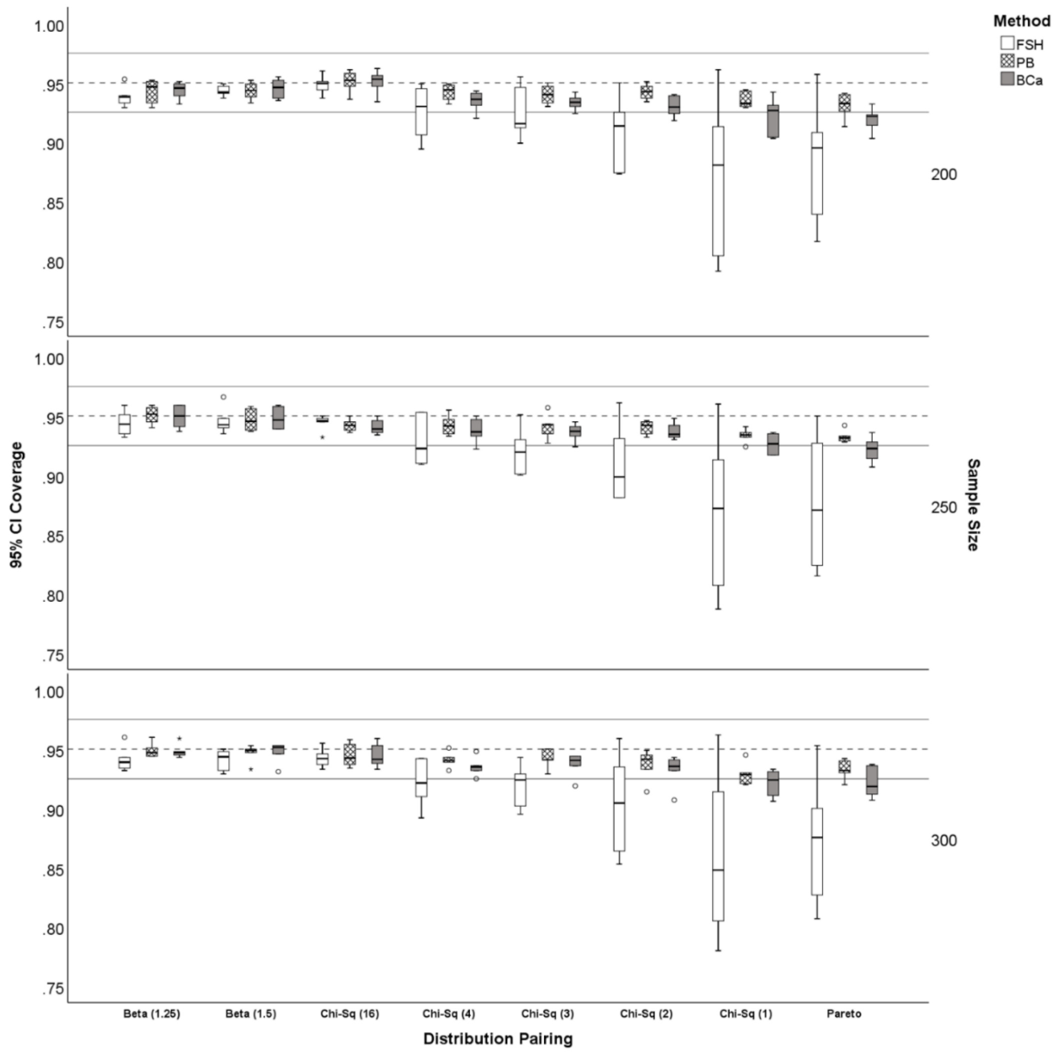

Figure 5. Distribution of 95% CI coverage for non-symmetric with non-symmetric distribution pairings by sample size of $200-300$. Fisher z-transformation (FSH), percentile bootstrap (PB), bias-corrected and accelerated bootstrap (BCa). Bootstrap methods (PB and BCa) were based on 2,000 bootstrap samples. The dashed line is at .95 and the solid lines are at the acceptable coverage of $[.925, .975]$.

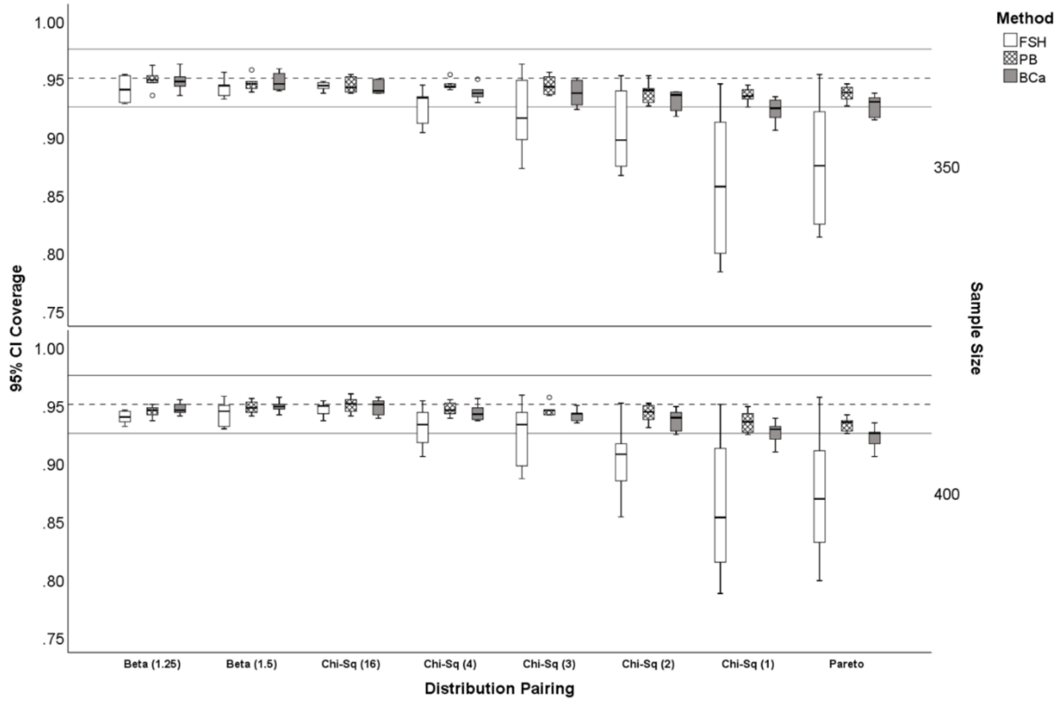

Figure 6. Distribution of 95% CI coverage for non-symmetric with non-symmetric distribution pairings by sample size of $350-400$. Fisher z-transformation (FSH), percentile bootstrap (PB), bias-corrected and accelerated bootstrap (BCa). Bootstrap methods (PB and BCa) were based on 2,000 bootstrap samples. The dashed line is at .95 and the solid lines are at the acceptable coverage of $[.925, .975]$.

4. Conclusion

Developing and researching correlation CIs has been a challenge. Much of this challenge stems from the mutable nature of the distribution for the correlation. Additionally, further complications may arise from the distribution of each variable in the correlation. Multiple studies in the past have attempted to determine how robust the correlation is to these conditions but have yielded mixed results (Berry & Mielke, 2000; Lunneborg, 1985; Padilla & Veprinsky, 2012, 2014; Rasmussen, 1987; Zeller & Levine, 1974). To address this discrepancy, the goal of the current study was to expand upon previous research by exploring the granularity of multiple sample sizes, correlation magnitudes, and distribution pairings to determine at what point a correlation CI would no longer be stable. This is key for applied research where it is essential to recognize which CI will yield optimal results. One example is the call to use effect sizes to help with replication concerns in the behavioral/social sciences (Collaboration, 2015). In such cases, providing a CI benefits applied research by providing precision information about the effect size (e.g., the correlation). Regardless, the current study has some findings consistent with previous research and further expands on qualities that were not previously explored.

The current study had similar findings with the study conducted by Bishara and Hittner (2017) with regards to the Spearman and RIN CIs. The two CIs had consistently good coverage probability across all conditions for both studies. However, caution is advised when using the Spearman and RIN CIs as they both use irreversible transformations of the correlation. As such, some reservation should be used when using and interpreting the Spearman and RIN CIs.

The current study also shared similar findings to Bishara and Hittner (2017) with regards to the Fisher z-transformation CI. However, those findings were not discussed in detail in their research. Generally, the Fisher z-transformation CI had excellent coverage probability so long as one of the paired variables was normal. If this was not the case, the coverage probability performance broke down. This indicates that the Fisher z-transformation CI is sensitive to non-normality. In the present study, this is most apparent when a distribution had $|\text{skewness}| \geq 1.41$. The sensitivity to skew is also made apparent by noticing that the correlation sampling distribution becomes more skewed the closer the correlation gets to one or negative one (see Figure 1). It is also worth noting that in these cases, increases in sample size did not alleviate the situation but instead made it worse. It is therefore unwise to use the Fisher z-transformation CI if the pair of correlated variables are non-normal.

More nuance into the qualities of the PB and BCa CIs was also found. As it stands, the PB and BCa CIs have issues maintaining acceptable coverage when both variables are non-normal and when sample sizes are small. However, the PB was shown to have acceptable coverage when the paired variables had $|\text{skewness}| \leq 2.83$ and when sample size was $n \geq 200$. In addition, the BCa was shown to have acceptable coverage when the paired variables had $|\text{skewness}| \leq 2$ and when sample size was $n \geq 200$. This quality was not captured by Bishara and Hittner's (2017) study as $n = 160$ was the largest sample size they investigated. Additionally, the required sample size for acceptable coverage is lowered to $n \geq 100$ if one of the paired variables is normal. This is also consistent regardless of correlation magnitude. These finding agree with previous research that suggest that the bootstrap works better with increased sample size (Rasmussen, 1987). Even so, it is worth noting that the BCa CI performed better than the PB CI when sample sizes were smaller, but the PB CI outperformed the BCa CI as sample sizes were larger.

Despite the findings gathered in the current study, there are still refinements that could be made to further advance the literature. Like the Bishara and Hittner (2017) study, the current study had a condition where a normal variable was paired with another variable. In the current study, such pairings typically achieved acceptable coverage. In future research, it may be of interest to explore if using a non-normal symmetric variable paired with another variable will yield acceptable coverage. If this is the case, it would afford researchers more flexibility in utilizing bootstrap CIs. It may also be of interest to combine the bootstrap with the Fisher z-transformation.

The Fisher z-transformation CI had excellent coverage performance when at least one variable was normal. Additionally, increased sample size tended to improve coverage performance for the bootstrap CIs. Combining these properties of the Fisher z-transformation and bootstrap may yield CIs that have acceptable coverage performance even when both paired variables are highly skewed (i.e., $|\text{skewness}| \geq 1.41$).

Further improvements can be made by expanding the pool of conditions explored. Given that the distribution of the correlation becomes more skewed as it gets closer to positive or negative one (see Figure 1), it may be fruitful to investigate the bootstrap CI coverage when the correlation is greater than .50 (i.e., $\rho > .50$). Also, the bootstrap CIs generally had acceptable coverage when $n = 200$ with improved performance as sample size increased. Therefore, it may be of interest to explore when exactly coverage is maximized (i.e., reaches a point of diminishing returns) and how larger sample sizes impact the correlation when it is greater than .50. As such, further exploration of the bootstrap may yield more promising results.

In summary, the correlation CIs investigated in the current study generally had acceptable coverage probability performance but there are some considerations to keep in my mind. The RIN and Spearman CIs both had consistently good performance across all conditions but risk misinterpretation as they involve irreversible transformations of the correlation; which is not an issue for the other CIs. The Fisher z-transformation CI had excellent performance when at least one of the paired variables was normal regardless of the sample size investigated. However, the performance of the Fisher z-transformation CI was shown to break down when the paired variables had $|\text{skewness}| \geq 1.41$ and increasing the sample size made the performance worse. The PB was shown to have acceptable coverage when the paired variables had $|\text{skewness}| \leq 2.83$ and when sample size was $n \geq 200$. The BCa was shown to have acceptable coverage when the paired variables had $|\text{skewness}| \leq 2$ and when sample size was $n \geq 200$. The sample size needed for acceptable coverage for the PB CI and BCa CI reduced to $n \geq 100$ each if one of the paired variables was normal. Additionally, the BCa CI had better performance than the PB CI when sample sizes were smaller and the PB CI had better performance than the BCa CI when sample sizes were larger. Given these findings, one can confidently use the Fisher z-transformation CI with $n \geq 20$ in the following two situations: when one of the paired variables is normal or if the paired variables have $|\text{skewness}| \leq .848$. If the paired variables are non-normal (i.e., $|\text{skewness}| \geq .848$), the PB and BCa CIs generally performed well with $n \geq 200$, but the PB is recommended as it had better performance for more extreme non-normal paired variables (i.e., $|\text{skewness}| \geq 2$).

References

Berry, K. J., & Mielke, P. W. (2000). A monte carlo investigation of the fisher z transformation for normal and nonnormal distributions. Psychological Reports, 87(3), 1101-1114. doi:10.2466/pr0.2000.87.3f.1101

Bishara, A. J., & Hittner, J. B. (2017). Confidence intervals for correlations when data are not normal. Behav Res, 49, 294-309. doi:10.3758/s13428-016-0702-8

Blanca, M. J., Arnau, J., Lopez-Montiel, D., Bono, R., & Bendayan, R. (2013). Skewness and kurtosis in real data samples. Methodology, 9(2), 78-84. doi:10.1027/1614-2241/a000057

Bradley, J. V. (1978). Robustness? British Journal of Mathematical and Statistical Psychology, 31(2), 144-152. doi:10.1111/j.2044-8317.1978.tb00581.x

Cohen, J. (1988). Statistical Power Analysis for the Behavioral Sciences: Lawrence Erlbaum Associates.

Collaboration. (2015). Estimating the reproducibility of psychological science. Science, 349(6521). doi:10.1126/science.aac4716

Efron, B., & Tibshirani, R. (1993). An introduction to the bootstrap. New York: Chapman & Hall.

Fieller, E. C., Hartley, H. O., & Pearson, E. S. (1957). Tests for rank correlation coefficients: I. Biometrika, 44(470-481). doi:10.2307/2332878

Fisher, R. A. (1915). Frequency distribution of the values of the correlation coefficient in the samples from an infinitely large population. Biometrika, 10(4), 507-521. doi:10.2307/2331838

Hald, A. (2007). A history of parametric statistical inference from Bernoulli to Fisher, 1713-1935. New York: Springer.

Headrick, T. C. (2002). Fast fifth-order polynomial transforms for generating univariate and multivariate nonnormal distributions. Computational Statistics & Data Analysis, 40(4), 685-711. doi:10.1016/S0167-9473(02)00072-5

Lunneborg, C. E. (1985). Estimating the correlation coefficient: The bootstrap approach. Psychological Bulletin, 98(1), 209-215. doi:10.1037/0033-2909.98.1.209

Muchinsky, P. M. (1996). The correction for attenuation. Education and Psychological Measurement, 56(1), 63-75. doi:10.1177/0013164496056001004

Padilla, M. A., & Veprinsky, A. (2012). Correlation attenuation due to measurement error: A new approach using the bootstrap procedure. Educational and Psychological Measurement, 72(5), 827-846. doi:10.1177/0013164412443963

Padilla, M. A., & Veprinsky, A. (2014). Bootstrapped deattenuated correlation: Nonnormal distributions. Educational and Psychological Measurement, 74(5), 823-830. doi:10.1177/0013164414531780

Rasmussen, J. L. (1987). Estimating the correlation coefficient: Bootstrap and parametric approaches. Psychological Bulletin, 101(1), 136-139. doi:10.1037/0033-2909.101.1.136

Spearman, C. (1904). The proof and measurement of association between two things. The American Journal of Psychology, 15(1), 72-101. doi:10.2307/1412159

Zeller, R. A., & Levine, Z. H. (1974). The effects of violating the normality assumption underlying r. Sociological Methods & Research, 2(4), 511-519. doi:10.1177/004912417400200406