Modern likelihood inference for the maximum/minimum of a bivariate normal vector

V. Mameli\textsuperscript{a} and A. R. Brazzale\textsuperscript{b}

\textsuperscript{a}Department of Environmental Sciences, Informatics and Statistics, Ca’ Foscari University of Venice, 155 Via Torino, Venezia Mestre, Italy; \textsuperscript{b}Department of Statistical Sciences, University of Padova, 241 Via Cesare Battisti, Padova, Italy

\textbf{ABSTRACT}

We consider the use of modern likelihood asymptotics in the construction of confidence intervals for the parameter which determines the skewness of the distribution of the maximum/minimum of an exchangeable bivariate normal random vector. Simulation studies were conducted to investigate the accuracy of the proposed methods and to compare them to available alternatives. Accuracy is evaluated in terms of both coverage probability and expected length of the interval. We furthermore illustrate the suitability of our proposals by means of two data sets, consisting of, respectively, measurements taken on the brains of 10 mono-zygotic twins and measurements of mineral content of bones in the dominant and non-dominant arms for 25 elderly women.

\textbf{1. Introduction}

Small sample sizes are rather common to researchers in fields such as biology, genetics, medical sciences and psychology. Inference based on the classical first-order normal and $\chi^2$ approximations may then be unreliable. The last four decades have seen the development of so-called higher order likelihood approximations, which require little more effort than is needed for their first-order counterparts while providing highly accurate inferences in small samples. We refer the reader to [1] for a rich collection of realistic examples and case studies, which show how to use the new theory. The aim of this paper is to investigate the feasibility of these modern likelihood-based solutions, whose third-order accuracy makes them well suited for small-sample sizes, for the analysis of bivariate normally distributed continuous data when interest focuses on the maximum (or minimum) value of correlated observations. Our study furthermore deepens the theoretical understanding of the heretofore available solutions, which were developed for large sample sizes.

Indeed, a great variety of studies, especially in the medical area, are based on the analysis of extreme observations. For instance, the extremes coming from the left and right sides of the body are commonplace in comparative studies; see [2] for an illustration about the visual acuity of fellow eyes, or [3,4] and reference therein, for other typical cases. A further instance are twin studies, which since Sir Francis Galton’s [5] seminal paper, have extensively been used for the quantitative ascertainment of genetic and environmental influences. Classical twin designs are a valuable tool in fields such as biomedicine, psychiatry and behavioural sciences, where the number of available observations is far smaller than those typical in modern twin studies. There are several views of how the degree of concordance between twins should be assessed.[6,7] Here, we promote the use of Azzalini’s
skew-normal distribution, which revisits [9] the first contribution to this regard we are aware of, from a modern perspective.

We set off from a theoretical result, due to Loperfido [10] which states that the maximum, or minimum, of two random variables, whose joint distribution is bivariate exchangeable normal with correlation coefficient \( \rho \), is skew-normally distributed with skewness parameter \( \gamma \), or \(-\gamma\), where \( \gamma = \sqrt{(1-\rho)/(1+\rho)} \). Recently, Mameli et al. [11] borrowing from Loperfido [10] and Fisher’s \( z \) transform for \( \rho \), obtained an asymptotic confidence interval for the skewness parameter of the distribution of the maximum/minimum under this framework. Their simulations revealed that the proposed asymptotic confidence interval has similar actual and nominal coverages, though its expected length increases for decreasing sample size and correlation coefficient close to -1. In this paper we explore the performance of confidence intervals for \( \gamma \) obtained from the small-sample solutions recently proposed in [12], and this in terms of both actual coverage and expected length.

The paper organizes as follows. Section 2 reviews modern likelihood-based inference. The skew-normal distribution and Loperfido’s results will be introduced in Section 3. Inference on \( \gamma \) will be discussed in Section 4. Section 5 analyses the twin data collected by Tramo et al. [13] and the dataset presented by Johnson and Wichern [14] using the large- and small-sample solutions of Section 4. The finite-sample properties of confidence intervals will be investigated in Section 6 through simulation. Some concluding remarks are given in Section 7. Table A1 in the appendix provides a list of the most common notations used in the paper together with their definition.

## 2. Likelihood-based inference

### 2.1. First-order theory

Let \( y = (y_1, \ldots, y_n) \) be a sample of size \( n \) with joint log-likelihood function \( l(\theta) = l(\theta; y) \), where \( \theta = (\psi, \lambda) \) is a \( k \)-dimensional parameter, \( \psi \) is the scalar parameter of interest, and \( \lambda \) a vector of nuisance parameters of dimension \( k - 1 \). Under broad regularity conditions, the maximum likelihood estimate of \( \theta \), denoted by \( \hat{\theta} \), may be obtained by solving the score equation \( l_\theta(\hat{\theta}; y) = 0 \), with \( l_\theta(\theta; y) = \partial l(\theta; y)/\partial \theta \). Let \( j(\theta) = \partial^2 l(\theta; y)/\partial \theta \partial \theta^\top \) represent the observed information function for \( \theta \) and \( j(\hat{\theta}) \) the observed Fisher information. Here \( \top \) denotes matrix transposition. The decomposition of the parameter \( \theta \) into \( \psi \) and \( \lambda \) leads to an analogous decomposition of the score vector \( l_\theta(\theta; y) \) and of the observed information function \( j(\theta) \).

The recommended likelihood pivot for making inference on \( \psi \) is the signed likelihood root

\[
r(\psi) = \text{sign}(\hat{\psi} - \psi) \sqrt{2[l_p(\hat{\psi}) - l_p(\psi)]}. \tag{1}
\]

Here \( l_p(\psi) = l(\hat{\theta}_\psi) \), with \( \hat{\theta}_\psi = (\psi, \hat{\lambda}_\psi) \), is the profile log-likelihood, while \( \hat{\lambda}_\psi \) represents the constrained maximum likelihood estimate obtained by maximizing the log-likelihood \( l(\psi, \lambda) \) with respect to \( \lambda \) holding \( \psi \) fixed. The signed likelihood root (1) is asymptotically standard normal up to the order \( n^{-1/2} \), which leads to the first-order \( (1 - \alpha)100\% \) confidence interval for \( \psi \)

\[
\{ \psi : |r(\psi)| \leq z_{1 - \alpha/2} \}, \tag{2}
\]

where \( z_\beta \), with \( \beta \in (0, 1) \), is the \( \beta \)th quantile of the standard normal distribution. The standard normal approximation provides a satisfactory approximation for large sample sizes, but can be highly unreliable for small values of \( n \). The value of \( \psi \) which satisfies Equation (2) can be found numerically by calculating the function \( r(\psi) \) on a grid of points \( \psi \), which are then interpolated using a suitable smoothing function. The numerical issues, which may arise in the interpolation step, can be avoided by excluding the values of \( \psi \) close to the maximum likelihood estimate \( \hat{\psi} \). The details are given in [1, Section 9.3].
2.2. Higher order theory

A nowadays broadly known improvement to the signed likelihood root (1), which was originally introduced by Barndorff-Nielsen,[15] is the modified likelihood root

\[ r^*(\psi) = r(\psi) + \frac{1}{r(\psi)} \log \left( \frac{q(\psi)}{r(\psi)} \right), \]

whose finite-sample distribution may be approximated by the standard normal up to the order \( n^{-3/2} \). The higher order \((1 - \alpha)100\%\) confidence interval for \( \psi \) is hence given by

\[ \{ \psi : |r^*(\psi)| \leq z_{1-\alpha/2} \}. \]

Again, pivot profiling [1, Section 9.3] can be used to identify the upper and lower bounds of the confidence interval. Furthermore, the \( r^*(\psi) \) pivot – like its first-order counterpart \( r(\psi) \) – is invariant under interest-respecting re-parametrizations, that is re-parametrizations of the form \( \tau(\theta) = \tau(\psi, \lambda) = (\zeta, \eta) \) with \( \zeta = \zeta(\psi) \) and \( \eta = \eta(\psi, \lambda) \).

Several expressions for the correction term \( q \) have been proposed, both from the frequentist and the Bayesian perspective. Here, we will focus on the developments by Fraser and Reid.[16] Their formula for \( q \) is based upon the notion of ‘tangent exponential model’ which, at a fixed value of \( y \), denoted \( y_0 \), approximates the true model by a local exponential model with canonical parameter \( \varphi = \varphi(\theta) \), defined as

\[ \varphi^T(\theta) = l_V(\theta; y_0) = \sum_{i=1}^{n} \frac{\partial l(\theta; y)}{\partial y_i} \bigg|_{y=y_0} V_i. \]

Here, \( l_V \) indicates differentiation of the log-likelihood function in the directions given by the \( n \) columns \( V_1, \ldots, V_n \) of the \( n \times k \) matrix \( V \). The matrix \( V \) can be constructed using a vector of pivotal quantities \( z = \{ z_1(y_1, \theta), \ldots, z_n(y_n, \theta) \} \), where each component \( z_i(y_i, \theta) \) has a fixed distribution under the model. The matrix \( V \) is defined through \( z \) by

\[ V = -\left( \frac{\partial z}{\partial y^T} \right)^{-1} \left( \frac{\partial z}{\partial \theta^T} \right) \bigg|_{(y_0, \hat{\theta}_0)}, \]

where \( \hat{\theta}_0 \) is the maximum likelihood estimate at \( y_0 \). The expression of the correction term \( q \) is then

\[ q(\psi) = \frac{\varphi^T(\hat{\theta}) - \varphi(\hat{\psi})}{|\varphi(\hat{\psi})|} \varphi_\lambda(\hat{\theta}_0) \left\{ \frac{|j(\hat{\theta})|}{|j_{\lambda\lambda}(\hat{\theta}_0)|} \right\}^{1/2}, \]

where \( \varphi_\theta(\theta) = \partial \varphi(\theta)/\partial \theta^T \) represents the matrix of partial derivatives of \( \varphi(\theta) \) with respect to \( \theta \), while \( \varphi_\lambda(\theta) = \partial \varphi(\theta)/\partial \lambda^T \) identifies the \( k - 1 \) columns of this matrix which correspond to the nuisance parameter \( \lambda \). Analogously, the matrix \( j_{\lambda\lambda}(\theta) \) is the \( (k - 1) \times (k - 1) \) sub-matrix of the observed information function \( j(\theta) \) with respect to the nuisance parameter \( \lambda \).

The expression of \( q(\psi) \) for the case in which no explicit nuisance parametrization is given can be found in [12].

2.3. Approximations for Bayesian inference

In the Bayesian setting with a prior density \( \pi(\theta) \) for \( \theta \), the analogue of the first-order results of Section 2.1 is the asymptotic normality of the posterior density \( \pi(\theta|y) \) for \( \theta \). The Bayesian counterpart
of the correction term \( q \) in Equation (3), which we will denote by \( q_B \), was obtained by DiCiccio and Martin [17] under the assumption that the nuisance parametrization is given explicitly, and results to

\[
q_B(\psi) = l_p'(\psi) j_p(\hat{\psi})^{-1/2} \left\{ \frac{|j_{\psi\lambda}(\hat{\theta}_\psi)|}{|j_{\lambda\lambda}(\hat{\theta})|} \right\}^{1/2} \frac{\pi(\hat{\theta})}{\pi(\hat{\theta}_\psi)},
\]

where \( l_p'(\psi) = d l_p(\psi)/d\psi \) is the profile score function and \( j_p(\psi) = d^2 l_p(\psi)/d\psi^2 \) the profile observed information function. Posterior quantiles for the parameter \( \psi \) can then be found exploiting the fact that the posterior distribution function

\[
\Pi(\psi_0 | y) = Pr(\psi \leq \psi_0 | y) = 1 - \Phi(r^*_B(\psi_0))
\]

may be approximated to the order \( n^{-3/2} \) by the standard normal distribution function \( \Phi(r^*_B(\psi_0)) \), evaluated at

\[
r^*_B(\psi_0) = r(\psi_0) + \frac{1}{r(\psi_0)} \log \left( \frac{q_B(\psi_0)}{r(\psi_0)} \right).
\]

Again, pivot profiling provides the upper and lower bounds of the (1 - \( \alpha \))100% credible interval for \( \psi \) given by

\[
\{ \psi : |r^*_B(\psi)| \leq z_{1-\alpha/2} \}.
\]

Like for \( q(\psi) \), Fraser et al. [12] provide the expression of the correction term \( q_B(\psi) \) for the case in which the nuisance parametrization is not given explicitly.

### 2.3.1. Matching priors.

Given the prior \( \pi(\theta) \) for \( \theta \), let \( \theta^\pi_{1-\alpha} \) denote the (1 - \( \alpha \))th approximate posterior quantile of \( \theta \) of order \( n^{-s} \), that is, the value of \( \theta \) for which

\[
Pr_{\theta|y}(\theta \leq \theta^\pi_{1-\alpha} | y) = 1 - \alpha + O_p(n^{-s}),
\]

with \( s > 0 \) and \( 0 < \alpha < 1 \). If we also have that

\[
Pr_Y(\theta^\pi_{1-\alpha} \geq \theta | \theta) = 1 - \alpha + O_p(n^{-s}),
\]

with \( \theta^\pi_{1-\alpha} \) the upper bound of a frequentist one-sided (1 - \( \alpha \))100% confidence interval, the prior \( \pi \) is called a probability matching prior to the \( s \)th order of approximation. For such priors, Bayesian and frequentist inference for the parameter \( \theta \) are in perfect agreement up to the order \( s \).

If \( s = 1 \), \( \pi(\theta) \) is called a first-order probability matching prior, while for \( s = 3/2 \) we have a second-order probability matching prior. Welch and Peers [18] showed that the unique first-order probability matching prior, when no nuisance parameter is present, is Jeffrey’s prior.

The same result does not necessarily hold when \( \theta \) includes a nuisance component \( \lambda \). For an orthogonal parametrization,[19] proposed to use the following prior for \( \theta \) in Equation (7),

\[
\pi(\psi, \lambda) \propto i_{\psi\psi}(\psi, \lambda),
\]

where \( i_{\psi\psi}(\psi, \lambda) \) represents the value of the expected Fisher information function corresponding to \( \psi \). The authors call this prior the ‘unique prior’, as it leads to an approximation of the marginal posterior distribution of \( \psi \) accurate to the order \( n^{-3/2} \). When the parametrization \( \theta = (\psi, \eta) \) is not orthogonal, their suggestion is to find an orthogonal parametrization \( (\psi, \eta) \) of the original model for
which the prior can be expressed as Equation (12), and then to re-express the prior in the original parametrization \((\psi, \lambda)\), leading to

\[
\pi(\psi, \lambda) \propto i_{\psi, \lambda}^{1/2}(\psi, \lambda) \left| \frac{\partial \eta}{\partial \lambda} \right|, \tag{13}
\]

with \(i_{\psi, \lambda}(\psi, \lambda) = i_{\psi}(\psi, \lambda) - i_{\lambda}(\psi, \lambda)i_{\lambda}^{-1}(\psi, \lambda)i_{\lambda}(\psi, \lambda)\). Here, the indices \(\psi\) and \(\lambda\) indicate which sub-blocks of the expected Fisher information function to take. Furthermore, \(\left| \frac{\partial \eta}{\partial \lambda} \right|\) represents the Jacobian of the transformation from \((\psi, \eta)\) to \((\psi, \lambda)\).

3. The skew-normal model

The skew-normal distribution was introduced by Azzalini [8] to define a class of asymmetric parametric models which includes the standard normal as a special case. We say that a continuous random variable \(Z \sim SN(\gamma)\), distributes as a skew-normal indexed by the real parameter \(\gamma\), if it has density function

\[
p(z; \gamma) = 2\phi(z) \Phi(\gamma z) \quad \text{with } z \in \mathbb{R}.
\]

Here \(\phi(\cdot)\) and \(\Phi(\cdot)\) denote, respectively, the density and the distribution functions of the standard normal distribution. The class of skew-normal distributions can be widened by including a location parameter \(\mu \in \mathbb{R}\) and a scale parameter \(\sigma > 0\). Thus, if \(X \sim SN(\gamma)\), then \(Y = \mu + \sigma X\) is a skew-normal random variable with parameters \(\mu, \sigma, \gamma\), or, \(Y \sim SN(\mu, \sigma, \gamma)\) for short. Making inference on the skewness parameter is quite challenging, as the expected Fisher information becomes singular as \(\gamma \to 0\); see the solutions proposed by Azzalini and Capitanio [20] and Sartori [21] Functions for manipulating the skew-normal probability distribution and for fitting it to data are given in the R package sn [22]. We refer the reader to [23, 24] for a general treatment of the skew-normal distribution and its extensions.

The focus of this paper is on the distribution of the maximum (or minimum) of an exchangeable bivariate normal random vector. The first contribution to this regard is, to our knowledge [9] who discusses inference for the minimum value with special emphasis on twin studies. We derive our reference model from Corollary 1 of Loperfido [10] which, adapted to our notation, reads as follows.

Let \(X_1\) and \(X_2\) be two random variables whose joint distribution is bivariate normal with common mean \(\mu \in \mathbb{R}\), common variance \(\sigma^2 > 0\) and correlation coefficient \(\rho \in (-1, 1)\). Then for any two real constants \(h\) and \(k \neq -h\), the distribution of \(h \min(X_1, X_2) + k \max(X_1, X_2)\) is

\[
SN\left(\mu(h + k), \sigma \sqrt{h^2 + k^2 + 2\rho hk}, \gamma = \frac{k-h}{|k+h|} \sqrt{1-\rho} \right). \tag{14}
\]

It follows that the distribution of \(\max(X_1, X_2)\) is \(SN(\mu, \sigma, \gamma)\) with \(\gamma = \sqrt{(1-\rho)/(1+\rho)} \geq 0\), whereas the distribution of \(\min(X_1, X_2)\) is \(SN(\mu, \sigma, \gamma)\) with \(\gamma = -\sqrt{(1-\rho)/(1+\rho)} \leq 0\). The special case of \(\rho = 0\) translates into \(\gamma = 1\) and \(\gamma = -1\), respectively.

Note that a first version of Corollary 1 of Loperfido [10] was already derived by Nagaraja [25] in the special case of the bivariate normal model with zero means, unit variances and correlation coefficient \(\rho\), and later re-discovered by Loperfido [26].

4. Inference on the skewness parameter \(\gamma\)

This section outlines known exact and large-sample results for making inference on the skewness parameter \(\gamma = \sqrt{(1-\rho)/(1+\rho)}\) which characterizes the distribution of the maximum between two correlated normal variables, as well as our newly derived small-sample solutions. In particular, we will consider the following three instances: (1) the no-nuisance parameter case, (2) the complete bivariate
normal model, and (3) the equi-correlated bivariate normal model. All methods straightforwardly extend to the case \( \gamma = -\sqrt{(1 - \rho)/(1 + \rho)} \) with the suitable changes.

### 4.1. The bivariate normal model without nuisance parameter

Let \( Y = (X_1, X_2) \) be a bivariate normal random vector with common mean 0, common variance 1 and correlation coefficient \( \rho \in (-1, 1) \). Given an i.i.d. sample \( \{(x_{11}, x_{21}), \ldots, (x_{1n}, x_{2n})\} \) of size \( n \) from \( Y \), Haddad and Provost [27] proposed a range-based exact confidence interval for \( \rho \). The construction of the confidence interval makes use of the two random variables \( D_+ = \sum_{i=1}^{n} (X_{1i} + X_{2i})^2 \) and \( D_- = \sum_{i=1}^{n} (X_{1i} - X_{2i})^2 \). Taking advantage of the independence of \( X_{1i} + X_{2i} \) and \( X_{1i} - X_{2i} \), along with the fact that \( X_{1i} + X_{2i} \sim N(0, 2(1 + \rho)) \) and \( X_{1i} - X_{2i} \sim N(0, 2(1 - \rho)) \), the authors derive the following pivotal quantity

\[
\frac{D_+ (1 - \rho)}{D_- (1 + \rho)} \sim F_{n,n},
\]

where \( F_{n,n} \) is Fisher’s F distribution with \((n, n)\) degrees of freedom. This gives an exact \((1 - \alpha)100\%\) confidence interval for the parameter \( \gamma \) of the form

\[
\left\{ \gamma \in [0, \infty) : \sqrt{\frac{D_-}{D_+}} F_{n,n}(\alpha/2) < \gamma < \sqrt{\frac{D_-}{D_+}} F_{n,n}(1 - \alpha/2) \right\},
\]

where \( F_{n,n}(\beta) \), with \( \beta \in (0, 1) \), represents the \( \beta \)th quantile of the \( F_{n,n} \) distribution.

**4.1.1. Small-sample confidence intervals.**

Reid [28] provides the expression of the canonical parameter \( \varphi \) when interest relies on \( \theta = \rho \), the correlation coefficient of the bivariate normal vector \( Y = (X_1, X_2) \). The reference model in this case is a \((2, 1)\) curved exponential family. Confidence intervals for the parameter \( \gamma \) are derived from the \( r(\rho) \) and \( r^*(\rho) \) pivots thanks to their invariance under interest-respecting re-parametrizations. Credible intervals for \( \gamma \) are obtained from \( r^*_h(\rho) \) by exploiting that \( \gamma = \sqrt{(1 - \rho)/(1 + \rho)} \) is a one-to-one transformation of the random variable \( \rho \).

A key quantity for the determination of the canonical parameter (5) of the approximating tangent full exponential model is the vector \( V = (1 - \hat{\theta})^{-1}(t - \hat{\theta} s, s - \hat{\theta} t) \top \), obtained from the two independent pivots \( Z_1 = (T + S)/(1 + \theta) \) and \( Z_2 = (T - S)/(1 - \theta) \), with \( S = n^{-1} \sum_{i=1}^{n} X_{1i}X_{2i} \) and \( T = (2n)^{-1} \sum_{i=1}^{n} (X_{1i}^2 + X_{2i}^2) \), whose distribution is \( \chi^2_{2}/n \). The canonical parameter takes the form \( \varphi(\theta) = n[(1 - \theta^2)/(1 - \theta^2)]^{-1}\theta(t - \hat{\theta}s) - (s - \hat{\theta} t) \). Later, Reid and Fraser [29] proposed an alternative formulation, \( \bar{\varphi}(\theta) = n\theta/(1 - \theta^2) \), of the canonical parameter. As shown there, both formulations lead to almost the same numerical results as far as the approximation of tail areas is concerned. Turning to the Bayesian world, we may adopt Jeffreys’ prior for \( \rho \), given by

\[
\pi(\rho) \propto \frac{\sqrt{(1 + \rho^2)}}{(1 - \rho^2)}.
\]

As stated in Section 2.3.1, this prior provides a first-order probability matching prior for a scalar parameter in the absence of nuisance parameters.

### 4.2. The complete bivariate normal model

Let \( Y = (X_1, X_2) \) be again a bivariate normal random vector, this time with parameters \( (\mu_1, \mu_2, \sigma_1, \sigma_2, \rho) \) where \( (\mu_1, \sigma_1) \in \mathbb{R} \times \mathbb{R}_+ \) and \( (\mu_2, \sigma_2) \in \mathbb{R} \times \mathbb{R}_+ \) are, respectively, the means and variances of \( X_1 \) and \( X_2 \), and \( \rho \in (-1, 1) \) their correlation. Instructions of how to construct a, yet approximate, confidence interval for \( \rho \) given an i.i.d. sample \( \{(x_{11}, x_{21}), \ldots, (x_{1n}, x_{2n})\} \) of size \( n \)
from \( Y \), can once more be found in [27]. The first step is to standardize the two components \( X_{1i} \) and \( X_{2i} \) into \( X_{1i}^* \) and \( X_{2i}^* \). An approximate confidence interval for the parameter \( \rho \) is obtained by using the fact that \( X_{1i}^* - X_{2i}^* \) and \( X_{1i}^* + X_{2i}^* \) are nearly independent, \( X_{1i}^* + X_{2i}^* \sim N(0, 2(1 + \rho)) \) and \( X_{1i}^* - X_{2i}^* \sim N(0, 2(1 - \rho)) \). The pivot

\[
\frac{D_+ (1 - \rho)}{D_- (1 + \rho)} = \left( \frac{1 + R}{1 - R} \right) \frac{(1 - \rho)}{(1 + \rho)},
\]

where \( D_+ = \sum_{i=1}^n (X_{1i}^* + X_{2i}^*)^2 \) and \( D_- = \sum_{i=1}^n (X_{1i}^* - X_{2i}^*)^2 \) and \( R \) is the sample correlation coefficient defined as

\[
R = \frac{\sum_{i=1}^n (X_{1i} - \bar{X}_1)(X_{2i} - \bar{X}_2)}{\sqrt{\sum_{i=1}^n (X_{1i} - \bar{X}_1)^2} \sqrt{\sum_{i=1}^n (X_{2i} - \bar{X}_2)^2}},
\]

follows approximately an \( F_{n-1,n-1} \) distribution. The corresponding \((1 - \alpha)100\%\) confidence interval for \( \gamma \) becomes

\[
\left\{ \gamma \in [0, \infty) : \sqrt{\frac{1 - R}{1 + R}} F_{n-1,n-1}(\alpha/2) < \gamma < \sqrt{\frac{1 - R}{1 + R}} F_{n-1,n-1}(1 - \alpha/2) \right\}.
\]

A second approximate solution to the inferential problem we are interested in can be found in [11]. Because of the difficulties of obtaining the finite sample distribution of \( R \), inference for \( \rho \) is commonly based on the monotonic transformation \( \frac{1}{2} \ln \left( \frac{(1 + R)/(1 - R)}{1/\sqrt{n - 3}} \right) \), called Fisher’s \( z \)-transform. In particular, the distribution of

\[
Z = \frac{1/2 \ln ((1 + R)/(1 - R)) - 1/2 \ln ((1 + \rho)/(1 - \rho))}{1/\sqrt{n - 3}}
\]

for \( n > 50 \) is approximately standard normal. This turns into an \((1 - \alpha)100\%\) confidence interval for \( \gamma \), which we will call ACI5, of the form

\[
\left\{ \gamma \in [0, \infty) : \exp \left( \frac{-z_{1-\alpha/2}}{\sqrt{n - 3}} \right) \sqrt{\frac{1 - R}{1 + R}} < \gamma < \exp \left( \frac{z_{1-\alpha/2}}{\sqrt{n - 3}} \right) \sqrt{\frac{1 - R}{1 + R}} \right\}.
\]

Note that the upper and lower bounds of both, the confidence interval (20) proposed by Mameli et al. [11] and solution (18) derived from the confidence interval for \( \rho \) of Haddad and Provost, [27] include the multiplying factor \( \sqrt{(1 - R)/(1 + R)} \).

### 4.2.1. Small-sample confidence intervals.

The computation of small-sample confidence intervals for the parameter \( \rho \) from the frequentist point of view is addressed in [30]. Let \( \{(x_{11}, x_{21}), \ldots, (x_{1n}, x_{2n})\} \) be an i.i.d. sample from the bivariate normal random vector \( Y = (X_1, X_2) \). The log-likelihood function

\[
l(\theta) = -n \left( \log (\sigma_1 \sigma_2) + \frac{1}{2} \log (1 - \rho^2) + \frac{\mu_1^2}{2(1 - \rho^2) \sigma_1^2} + \frac{\mu_2^2}{2(1 - \rho^2) \sigma_2^2} - \frac{\mu_1 \mu_2 \rho}{(1 - \rho^2) \sigma_1 \sigma_2} \right) \\
- \frac{1}{2(1 - \rho^2) \sigma_1^2} \sum_{i=1}^n x_{1i}^2 - \frac{1}{2(1 - \rho^2) \sigma_2^2} \sum_{i=1}^n x_{2i}^2 + \frac{\mu_1 \sigma_2 - \mu_2 \sigma_1 \rho}{(1 - \rho^2) \sigma_1 \sigma_2} \sum_{i=1}^n x_{1i} \\
+ \frac{\mu_2 \sigma_1 - \mu_1 \sigma_2 \rho}{(1 - \rho^2) \sigma_1 \sigma_2} \sum_{i=1}^n x_{2i} + \frac{\rho}{(1 - \rho^2) \sigma_1 \sigma_2} \sum_{i=1}^n x_{1i} x_{2i},
\]
with \( \theta = (\rho, \mu_1, \mu_2, \sigma_1, \sigma_2) \), characterizes an exponential family with canonical parameter

\[
\varphi(\theta) = \left( -\frac{1}{(1-\rho^2)\sigma_1^2}, \frac{1}{(1-\rho^2)\sigma_2^2}, \frac{\mu_1\sigma_2 - \mu_2\sigma_1\rho}{(1-\rho^2)\sigma_1\sigma_2}, \frac{\mu_2\sigma_1 - \mu_1\sigma_2\rho}{(1-\rho^2)\sigma_1\sigma_2}, \frac{\rho}{(1-\rho^2)\sigma_1\sigma_2} \right).
\]

Setting \( \psi = \rho \) and \( \lambda = (\mu_1, \mu_2, \sigma_1, \sigma_2) \), \( q \) and \( r^* \) can readily be obtained from Equations (6) and (3), respectively. The computation of the Bayesian credible interval of Section 2.3 requires that we specify a prior for the parameter \( \theta \). The ‘unique prior’ defined by Staicu and Reid [19] may be calculated by referring to the orthogonal re-parametrization

\[
p(x_1, x_2 \mid \psi, \eta) \propto \frac{1}{\eta_4} \exp \left\{ -\frac{1}{2(1-\psi^2)^{1/2}} \eta_4 \left[ \frac{(x_1 - \eta_1)^2}{\eta_3} + \eta_3(x_2 - \eta_2)^2 - 2\psi(x_1 - \eta_1)(x_2 - \eta_2) \right] \right\},
\]

with \( \psi = \rho \) and \( \eta_1 = \mu_1, \eta_2 = \mu_2, \eta_3 = \sigma_1/\sigma_2 \) and \( \eta_4 = \sigma_1\sigma_2(1-\rho^2)^{1/2} \). The matching prior for \( \theta = (\rho, \lambda) \), with \( \lambda = (\mu, \sigma) \), will then be

\[
\pi(\theta) \propto i_{\rho\psi, \lambda}(\theta) \left| \frac{\partial \eta}{\partial \lambda} \right|,
\]

with \( i_{\rho, \psi, \lambda}(\theta) = i_{\rho, \psi}(\theta) - i_{\rho, \lambda}(\theta) i_{\psi, \lambda}^{-1}(\theta) i_{\psi, \rho}(\theta) \). In expression (21) the quantity \( i(\theta) \) denotes the expected Fisher information function, while \( \left| \frac{\partial \eta}{\partial \lambda} \right| \) represents the Jacobian of the transformation from the orthogonal parametrization \((\rho, \eta)\) to the original parametrization \((\rho, \lambda)\). Note that [31,32] used the same parametrization but with \( \mu_1 = \mu_2 = 0 \), which implies that Fisher’s expected information function only includes the parameters \( \psi, \eta_3 \) and \( \eta_4 \).

As in the previous section, confidence intervals for the parameter \( \gamma \) are readily derived from the \( r(\rho) \) and \( r^*(\rho) \) pivots thanks to their invariance under interest-respecting re-parametrizations, while credible intervals for \( \gamma \) are obtained from \( r^*_B(\rho) \) by using the biunivocal transformation, \( \sqrt{(1-\rho)/(1+\rho)} \), of the random variable \( \rho \).

### 4.3. The equi-correlated bivariate normal model

Let \((x_{11}, x_{21}), \ldots, (x_{1n}, x_{2n})\) be an i.i.d. sample from an equi-correlated bivariate normal vector \( Y = (X_1, X_2) \), where \( X_1 \) and \( X_2 \) have common mean \( \mu \in \mathbb{R} \), common variance \( \sigma^2 > 0 \) and correlation coefficient \( \rho \in (-1, 1) \). Maximum likelihood estimation for \( \theta = (\rho, \mu, \sigma) \) yields

\[
\hat{\mu} = \frac{1}{2n} \sum_{i=1}^{n} (x_{1i} + x_{2i}), \quad \hat{\sigma} = \sqrt{\frac{1}{2n} \sum_{i=1}^{n} \left[ (x_{1i} - \hat{\mu})^2 + (x_{2i} - \hat{\mu})^2 \right]},
\]

and

\[
\hat{\rho} = \frac{1}{n\sigma^2} \sum_{i=1}^{n} (x_{1i} - \hat{\mu})(x_{2i} - \hat{\mu}).
\]

Fisher [33] called \( \rho \) the intraclass correlation coefficient and its maximum likelihood estimate \( \hat{\rho} \) the intraclass correlation.

#### 4.3.1. Small-sample confidence intervals.

A first confidence interval for \( \gamma \), which is accurate up to the second order, can be derived by using an argument equivalent to the one adopted by Mameli et al. [11] for the complete bivariate normal model,
and exploits the fact that Fisher’s transformation of \( \hat{\rho} \) is again a normalizing and variance stabilizing transformation as it is the case for the well-known transformation of the sample correlation. That is, the pivot

\[
\sqrt{n} \left[ \frac{1}{2} \log \left( \frac{1 + \hat{\rho}}{1 - \hat{\rho}} \right) - \frac{1}{2} \log \left( \frac{1 + \rho}{1 - \rho} \right) + \frac{1}{2n} \right],
\]

is asymptotically standard normal up to the order \( n^{-1} \); see [34]. Expression (22) leads straightforwardly to the following asymptotic confidence interval for \( \gamma \):

\[
\left\{ \gamma \in [0, \infty) : \exp \left( \frac{-z_{1-\alpha/2}}{\sqrt{n}} - \frac{1}{2n} \right) \sqrt{\frac{(1 - \hat{\rho})}{(1 + \hat{\rho})}} < \gamma < \exp \left( \frac{z_{1-\alpha/2}}{\sqrt{n}} - \frac{1}{2n} \right) \sqrt{\frac{(1 - \hat{\rho})}{(1 + \hat{\rho})}} \right\},
\]

which we will indicate by ACI3.

Further small-sample confidence intervals may be derived by setting, as in Section 4.2, \( \psi = \rho \) and \( \lambda = (\mu, \sigma) \) and using the \( r^x(\rho) \) pivot from Equations (6) and (3), respectively. The log-likelihood function

\[
\ell(\theta) = -n \left[ 2 \log \sigma + \frac{1}{2} \log (1 - \rho^2) + \frac{\mu^2}{(1 + \rho)\sigma^2} + \frac{\sum_{i=1}^n (x_{1i} + x_{2i})}{(1 + \rho)\sigma^2} + 2\rho \frac{\sum_{i=1}^n x_{1i}x_{2i} - \sum_{i=1}^n (x_{1i}^2 + x_{2i}^2)}{2(1 - \rho^2)\sigma^2} \right],
\]

now characterizes an exponential family with canonical parameter

\[
\varphi(\theta) = \left( \frac{\mu}{\sigma^2(1 + \rho)}, \frac{\rho}{\sigma^2(1 - \rho^2)}, -\frac{1}{2(1 - \rho^2)\sigma^2} \right).
\]

In order to obtain credible confidence intervals from Equation (8) with frequentist coverage, we assume as matching prior \( \pi(\theta) \), the ‘unique prior’ defined in [19]. This requires an orthogonal re-parametrization of the model of the form

\[
p(x_1, x_2 \mid \rho, \eta) \propto \frac{1}{\eta_2} \exp \left\{ -\frac{1}{2(1 - \rho^2)^{1/2} \eta_2} \left[ (x_1 - \eta_1)^2 + (x_2 - \eta_1)^2 - 2\rho (x_1 - \eta_1)(x_2 - \eta_1) \right] \right\},
\]

with \( \eta_1 = \mu \) and \( \eta_2 = \sigma^2(1 - \rho^2)^{1/2} \). The matching prior for \( \theta = (\rho, \lambda) \), with \( \lambda = (\mu, \sigma) \), will then be

\[
\pi(\theta) \propto i_{r^x, \lambda}^{1/2}(\theta) \left| \frac{\partial \eta}{\partial \lambda} \right|,
\]

with \( i_{r^x, \lambda}(\theta) = i_{\rho \mu}(\theta) - i_{\rho \lambda}(\theta) i_{\mu \lambda}^{-1}(\theta) i_{\mu \rho}(\theta) \). As before, the quantity \( i(\theta) \) indicates the expected Fisher information function, while \( \left| \partial \eta / \partial \lambda \right| \) denotes the Jacobian of the transformation from \( (\rho, \eta) \) to \( (\rho, \lambda) \).

Once more, confidence and credible intervals for \( \gamma \) are obtained via pivot profiling [1] from the pivots \( r(\rho), r^x(\rho), \) and \( r^x_2(\rho) \).

5. Two real data examples

5.1. ‘Corpus callosum’ surface area in mono-zygotic twins

This first example considers the twin study conducted by Tramo et al. [13] which focuses on similarities of the brains of mono-zygotic twin pairs such as in forebrain volume, cortical surface area, and callosal area. The data, as available on StatLib, include the measurements on the brains of 10
Table 1. Tramo et al. [13] corpus callosum surface area data. Lower (LB) and upper (UB) bounds of 95% confidence intervals for the parameter $\gamma$. Pivots used, with corresponding confidence intervals: 1st – likelihood root $r$ (2); 3rd – modified likelihood root $r^*$ (4); Bayes – Bayesian modified likelihood root $r^*_B$ (9); HP – [27] approximate pivot (17); ACI5 – [11] approximate pivot (19).

| Method     | LB   | UB   | Length |
|------------|------|------|--------|
| 1st        | 0.121| 0.435| 0.314  |
| 3rd        | 0.119| 0.493| 0.374  |
| Bayes      | 0.114| 0.518| 0.404  |
| HP         | 0.114| 0.460| 0.346  |
| ACI5       | 0.109| 0.481| 0.372  |

Figure 1. Tramo et al. [13] corpus callosum surface area data. Pivot functions for the parameter $\gamma$ obtained from: likelihood root $r$ (solid), modified likelihood root $r^*$ (bold); Bayesian modified likelihood root $r^*_B$ (long-dashes); expression (17) by Haddad and Provost [27] (dotted); approximate pivot (19) by Mameli et al. [11] (dashed). Black horizontal lines: 2.5% and 97.5% normal quantiles; grey horizontal lines: 2.5% and 97.5% quantiles of the $F_{9,9}$ distribution.

pairs of mono-zygotic twins, five male and five female. Our variable of interest is the corpus callosum surface area. The purpose of the analysis is to model the distribution of the maximum between the area in the first twin and in the second twin. The order of twins is determined by birth, though this does not affect the method since the two variables are supposed to be exchangeable. The paired measurements exhibit a strong positive linear relationship with correlation coefficient amounting to $R = 0.90$. To assure that all conditions of Corollary 1 of Loperfido [10] hold, we first standardize the pairs of observations $\{(x_{11}, x_{21}), \ldots, (x_{1n}, x_{2n})\}$ as on page 7. The bivariate Shapiro–Wilk normality test ($W = 0.97$, p-value = 0.86) supports the hypothesis of bivariate normality of the standardized data. Maximum likelihood yields $\hat{\gamma} = 0.229$.

The five 95% confidence intervals for $\gamma$, computed using the methods outlined in Section 4.2, are given in Table 1. The interval based on the third-order Bayesian solution $r^*_B$ is wider than the confidence intervals obtained from the first-order pivot $r$, the higher order frequentist pivot $r^*$, the large sample (HP) confidence interval by Haddad and Provost [27] and the ACI5 confidence interval by Mameli et al. [11]. Figure 1 shows how to compute the lower and upper bounds numerically. The intervals based on $r$ (1st), $r^*$ (3rd), $r^*_B$ (Bayes) and ACI5 can be read off from the intersections of the corresponding pivots with the horizontal black lines, which represent the 2.5% and the 97.5% quantiles of the standard normal distribution. The lower and upper bounds of the HP confidence interval are computed similarly, but this time by referring to the horizontal grey lines, which represent the 2.5%, and the 97.5% quantiles of the $F$ distribution with $(9,9)$ degrees of freedom.
Table 2. Johnson and Wichern [14] mineral content measurements of the ulna data. Lower (LB) and upper (UB) bounds of 95% confidence intervals for the parameter $\gamma$. Pivots used, with corresponding confidence intervals: 1st – likelihood root $r$ (2); 3rd – modified likelihood root $r^*$ (4); Bayes – Bayesian modified likelihood root $r^*_B$ (9); ACI$_3$ – approximate pivot (22).

| Method  | LB    | UB    | Length |
|---------|-------|-------|--------|
| 1st     | −0.596| −0.269| 0.327  |
| 3rd     | −0.588| −0.261| 0.327  |
| Bayes   | −0.573| −0.253| 0.320  |
| ACI$_3$ | −0.581| −0.265| 0.316  |

Figure 2. Johnson and Wichern [14] mineral content measurements of the ulna data. Pivot functions for the parameter $\gamma$ obtained from: likelihood root $r$ (solid), modified likelihood root $r^*$ (bold); Bayesian modified likelihood root $r^*_B$ (long-dashes); approximate pivot (22) (dashed). Horizontal lines: 2.5% and 97.5% normal quantiles.

5.2. Mineral content measurements in the ulna of elderly women

This second example considers the data available in [14, p.24], which consist of mineral content measurements of three bones, among these the ulna, in the dominant and non-dominant arm for 25 elderly women. The purpose of the analysis now is to model the distribution of the minimum between the contra-lateral measurements of the ulna. To take account of the bilateral symmetries in the human body, we constrain the model parameters to $\mu_1 = \mu_2 = \mu$ and $\sigma_1 = \sigma_2 = \sigma$. The bivariate Shapiro–Wilk normality test confirms the hypothesis of bivariate normality ($W = 0.9708$, $p$-value = 0.6645). Furthermore, as expected, the two contra-lateral measurements present a strong positive linear relationship with intraclass correlation $\hat{\rho} = 0.724$. The maximum likelihood estimate of $-\gamma$ is $-\hat{\gamma} = -0.4004$.

Table 2 shows the four 95% confidence intervals for the parameter $-\gamma$ obtained by using the methods outlined in Section 4.3. The ACI$_3$ confidence interval of equation (23) is smaller than the confidence intervals obtained from the first-order pivot $r$ and the higher order frequentist and Bayesian pivots $r^*$ and $r^*_B$, respectively. Figure 2 plots the pivot functions $r$ (1st), $r^*$ (3rd), $r^*_B$ (Bayes) and ACI$_3$ for the ulna data. As before, intervals can be read off from the intersections of the corresponding pivots with the horizontal lines which represent the 2.5% and the 97.5% quantiles of the standard normal distribution.

6. Numerical assessment

We designed three simulations studies to assess and compare the finite-sample properties of the methods discussed in this paper. The summary statistics used are: empirical coverage (CP), upper error
probability (UE), that is, the percentage of the true parameter values falling above the upper bound, lower error probability (LE), that is, the percentage of the true parameter values falling below the lower bound, and average length (AL) of the confidence intervals considered in Section 4. All simulations were run using the numerical computing environment R.[35] For every occurrence, 10,000 replicates were generated for the four sample sizes $n = 5, 10, 15, 20$. The simulation error amounts to ±0.004. We only report the summary statistics for the smallest sample size $n = 5$; the omitted results are available as supplementary material.

6.1. Considered scenarios

6.1.1. Simulation 1
We consider a bivariate normal model with zero means, unit variances and unknown correlation $\rho$, which takes values from $-0.9$ to $+0.9$, with step size 0.1. The purpose is to compare the behaviour of the higher order pivot $r^*$ with its first-order counterpart $r$, the Bayesian small-sample solution $r_B^*$ and the exact method (14), which apply when no nuisance parameter is present, and this for small-sample sizes.

6.1.2. Simulation 2
We investigate the finite-sample performance of the confidence intervals obtained when nuisance parameters are present, again while emphasizing small-sample sizes. The pivots used are the higher order solution $r^*$, its first-order counterpart $r$, the Bayesian competitor $r_B^*$, and the large-sample solutions (18) and (20). We used $\mu_1 = \mu_2 = 7, \sigma_1 = \sigma_2 = 0.9$, while again $\rho \in \{-0.9, -0.8, \ldots, 0.8, 0.9\}$. Note that the simulation set-up borrows from the twin data example of Section 5.1, for which the maximum likelihood estimate is $\hat{\theta} = (0.900, 7.061, 6.924, 0.905, 0.872)$.

6.1.3. Simulation 3
We explore the finite-sample performance of the different confidence intervals which are available for the equi-correlated bivariate normal model, again while emphasizing on small-sample sizes. The pivots used are the higher order solution $r^*$, its first-order counterpart $r$, the Bayesian competitor $r_B^*$, and the approximate solution (22). We set $\mu = 0.7, \sigma = 0.1$, while again $\rho \in \{-0.9, -0.8, \ldots, 0.8, 0.9\}$. Note that the simulation set-up borrows from the real data example of Section 5.2, for which the maximum likelihood estimate is $\hat{\theta} = (0.724, 0.699, 0.103)$.

6.2. Discussion

6.2.1. Simulation 1
Figure 3 shows the actual coverage of the nominal 95% confidence intervals for $\gamma$ derived from Equation (15) (exact), and by using the pivots $r$ (1st), $r^*$ (3rd), and $r_B^*$ (Bayes) for the no nuisance parameter case. The higher order likelihood pivots $r^*$ and, to a somewhat lesser extent, $r_B^*$ outperform their first-order counterpart $r$, even for the very limited sample sizes considered in the four scenarios of Simulation 1. The differences among the four pivots fade out as the sample size increases.

Table 3 summarizes the performance of the nominal 95% confidence intervals for $\gamma$ derived from the four pivots considered in Simulation 1. Being there no nuisance parameter present, the corresponding true coverage probabilities are very close to the nominal level, as we may have expected. Inspection of the upper and lower error probabilities reveals that the $r^*$ and $r_B^*$ pivots also improve in terms of symmetry over $r$. The exact method produces confidence intervals for $\gamma$ which are, on average, wider than the confidence intervals obtained from the first-order solution $r$ and the higher order pivots $r^*$ and $r_B^*$. The average length of all four confidence intervals is larger for negative values of $\rho$, and increases when the correlation tends to $-1$. 
Figure 3. Simulation 1: bivariate normal with means 0 and variances 1. Empirical coverage of nominal two-sided 95% confidence intervals for $\gamma$ for varying values of $\rho$ and sample size $n$. From top left to bottom right: $n = 5$, $n = 10$, $n = 15$, $n = 20$. Pivots used: likelihood root $r$ (1st), modified likelihood root $r^*$ (3rd); Bayesian modified likelihood root $r^*_B$ (Bayes); expression (14) by Haddad and Provost [27] (exact). Based on 10,000 replicates. Legend: –○– 1st order; –△– 3rd order; –■– Bayes; –●– exact; — nominal.

6.2.2. Simulation 2.

Figure 4 reports the actual coverage of the nominal 95% confidence intervals for $\gamma$ obtained from expressions (17) (HP) and (19) (ACI5), and by using the pivots $r$ (1st), $r^*$ (3rd) and $r^*_B$ (Bayes). In terms of real coverage, $r^*$ again outperforms its first-order counterpart $r$. It also outperforms the large-sample (HP) proposal by Haddad and Provost [27] and, surprisingly, the Bayesian solution $r^*_B$. The most accurate method is the large-sample confidence interval developed by Mameli et al.,[11] although the differences fade out for increasing sample size.

Table 4 summarizes the performance of the nominal 95% confidence intervals for $\gamma$ derived from the five methods considered in Simulation 2. The results reveal that $r^*$ is more accurate than $r$, especially when the sample size is small, because of both an, on average, larger width and its capability of correctly centering the confidence intervals. The $r^*_B$ pivot consistently overestimates the real coverage, while guaranteeing symmetry on the tails, because of the, on average, longer confidence intervals it produces. The ACI5 and HP methods lead to confidence intervals for $\gamma$ which are remarkably asymmetric. Their better performance with respect to, respectively, $r^*$ and $r$ may be explained by the, on average, larger widths of the corresponding confidence intervals. For all five methods considered, the expected length becomes larger for negative values of $\rho$, especially when $\rho$ is close to $-1$. This is in agreement with,[11] who noted the same behaviour for their ACI5 confidence interval.
Table 3. Summary statistics for Simulation 1: bivariate normal with means 0 and variances 1. Empirical coverage (CP), upper (UE) and lower (LE) error probability and average length (AL) of nominal two-sided 95% confidence intervals for $\gamma$, for varying values of $\rho$ and sample size $n = 5$. Pivots used: likelihood root $r$ (1st), modified likelihood root $r^*$ (3rd); Bayesian modified likelihood root $r^*_B$ (Bayes); expression (14) by Haddad and Provost [27] (exact). Based on 10,000 replicates; simulation error: $\pm 0.004$.

| $\rho$ | Method | CP    | UE    | LE    | AL    |
|--------|--------|-------|-------|-------|-------|
| (a)    |        |       |       |       |       |
| -0.9   | 1st    | 0.924 | 0.039 | 0.037 | 6.714 |
|        | 3rd    | 0.954 | 0.024 | 0.022 | 6.556 |
|        | Bayes  | 0.953 | 0.024 | 0.022 | 6.604 |
|        | exact  | 0.950 | 0.024 | 0.026 | 11.329 |
| -0.8   | 1st    | 0.938 | 0.038 | 0.024 | 4.753 |
|        | 3rd    | 0.952 | 0.023 | 0.026 | 4.600 |
|        | Bayes  | 0.947 | 0.024 | 0.029 | 4.587 |
|        | exact  | 0.951 | 0.023 | 0.026 | 7.803 |
| -0.7   | 1st    | 0.939 | 0.038 | 0.023 | 3.861 |
|        | 3rd    | 0.948 | 0.024 | 0.029 | 3.702 |
|        | Bayes  | 0.940 | 0.024 | 0.036 | 3.648 |
|        | exact  | 0.949 | 0.025 | 0.026 | 6.144 |
| -0.6   | 1st    | 0.934 | 0.040 | 0.026 | 3.345 |
|        | 3rd    | 0.948 | 0.024 | 0.028 | 3.198 |
|        | Bayes  | 0.943 | 0.023 | 0.034 | 3.130 |
|        | exact  | 0.952 | 0.025 | 0.024 | 5.244 |
| -0.5   | 1st    | 0.930 | 0.038 | 0.033 | 2.946 |
|        | 3rd    | 0.945 | 0.024 | 0.031 | 2.807 |
|        | Bayes  | 0.942 | 0.025 | 0.033 | 2.727 |
|        | exact  | 0.947 | 0.025 | 0.028 | 4.483 |
| -0.4   | 1st    | 0.932 | 0.037 | 0.031 | 2.659 |
|        | 3rd    | 0.950 | 0.022 | 0.027 | 2.531 |
|        | Bayes  | 0.945 | 0.024 | 0.030 | 2.441 |
|        | exact  | 0.953 | 0.024 | 0.023 | 3.948 |
| -0.3   | 1st    | 0.922 | 0.042 | 0.035 | 2.426 |
|        | 3rd    | 0.942 | 0.027 | 0.031 | 2.314 |
|        | Bayes  | 0.934 | 0.032 | 0.033 | 2.234 |
|        | exact  | 0.948 | 0.026 | 0.027 | 3.552 |
| -0.2   | 1st    | 0.926 | 0.039 | 0.034 | 2.241 |
|        | 3rd    | 0.949 | 0.026 | 0.025 | 2.141 |
|        | Bayes  | 0.940 | 0.030 | 0.030 | 2.055 |
|        | exact  | 0.952 | 0.025 | 0.023 | 3.189 |
| -0.1   | 1st    | 0.927 | 0.037 | 0.036 | 2.087 |
|        | 3rd    | 0.951 | 0.023 | 0.026 | 1.995 |
|        | Bayes  | 0.937 | 0.031 | 0.032 | 1.911 |
|        | exact  | 0.953 | 0.023 | 0.024 | 2.854 |
| 0      | 1st    | 0.925 | 0.037 | 0.038 | 1.927 |
|        | 3rd    | 0.949 | 0.024 | 0.027 | 1.842 |
| (b)    |        |       |       |       |       |
|        | Bayes  | 0.937 | 0.031 | 0.032 | 1.758 |
|        | exact  | 0.952 | 0.023 | 0.025 | 2.579 |
| 0.1    | 1st    | 0.927 | 0.038 | 0.035 | 1.806 |
|        | 3rd    | 0.951 | 0.025 | 0.024 | 1.732 |
|        | Bayes  | 0.941 | 0.031 | 0.029 | 1.647 |
|        | exact  | 0.953 | 0.024 | 0.022 | 2.346 |
| 0.2    | 1st    | 0.925 | 0.033 | 0.042 | 1.695 |
|        | 3rd    | 0.948 | 0.025 | 0.026 | 1.636 |
|        | Bayes  | 0.936 | 0.030 | 0.034 | 1.550 |
|        | exact  | 0.949 | 0.024 | 0.028 | 2.110 |
| 0.3    | 1st    | 0.925 | 0.036 | 0.040 | 1.577 |
|        | 3rd    | 0.948 | 0.027 | 0.024 | 1.526 |
|        | Bayes  | 0.938 | 0.031 | 0.030 | 1.439 |
|        | exact  | 0.950 | 0.026 | 0.024 | 1.903 |
| 0.4    | 1st    | 0.929 | 0.032 | 0.039 | 1.460 |
|        | 3rd    | 0.948 | 0.028 | 0.024 | 1.421 |
|        | Bayes  | 0.942 | 0.032 | 0.026 | 1.347 |
|        | exact  | 0.952 | 0.025 | 0.024 | 1.701 |
Table 3. Continued.

| ρ   | Method | CP   | UE   | LE   | AL   |
|-----|--------|------|------|------|------|
| 0.5 | 1st    | 0.927| 0.032| 0.040| 1.350|
|     | 3rd    | 0.943| 0.031| 0.026| 1.324|
|     | Bayes  | 0.940| 0.033| 0.027| 1.257|
|     | exact  | 0.947| 0.026| 0.026| 1.499|
| 0.6 | 1st    | 0.928| 0.031| 0.041| 1.236|
|     | 3rd    | 0.940| 0.035| 0.025| 1.209|
|     | Bayes  | 0.937| 0.037| 0.025| 1.157|
|     | exact  | 0.948| 0.027| 0.025| 1.313|
| 0.7 | 1st    | 0.935| 0.023| 0.042| 1.061|
|     | 3rd    | 0.945| 0.029| 0.026| 1.054|
|     | Bayes  | 0.935| 0.038| 0.026| 1.039|
|     | exact  | 0.950| 0.026| 0.024| 1.097|
| 0.8 | 1st    | 0.936| 0.025| 0.038| 0.865|
|     | 3rd    | 0.948| 0.027| 0.024| 0.864|
|     | Bayes  | 0.946| 0.029| 0.024| 0.905|
|     | exact  | 0.951| 0.026| 0.023| 0.874|
| 0.9 | 1st    | 0.925| 0.035| 0.040| 0.554|
|     | 3rd    | 0.954| 0.022| 0.024| 0.563|
|     | Bayes  | 0.953| 0.023| 0.024| 0.652|
|     | exact  | 0.952| 0.026| 0.023| 0.598|

Figure 4. Simulation 2: bivariate normal with means 7 and variances 0.9. Empirical coverage of nominal equi-tailed 95% confidence intervals for γ for varying values of ρ and sample size n. From top left to bottom right: n = 5, n = 10, n = 15, n = 20. Pivots used: likelihood root r (1st), modified likelihood root r* (3rd); Bayesian modified likelihood root r^B (Bayes); expression (17) by Haddad and Provost [27] (HP) and expression (19) by Mameli et al. [11] (ACI5). Based on 10,000 replicates. Legend: ○ – 1st order; △ – 3rd order; ■ – Bayes; ● – HP; ○ – ACI5; – nominal.
Table 4. Summary statistics for Simulation 2: bivariate normal with means 7 and variances 0.9. Empirical coverage (CP), upper (UE) and lower (LE) error probability and average length (AL) of nominal two-sided 95% confidence intervals for $\gamma$, for varying values of $\rho$ and sample size $n = 5$. Pivots used: likelihood root $r$ (1st), modified likelihood root $r^*$ (3rd); Bayesian modified likelihood root $r^*_B$ (Bayes); expression (17) by Haddad and Provost [27] (HP); expression (19) by Mameli et al. [11] (ACI5). Based on 10,000 replicates; simulation error: ±0.004.

| $\rho$ | Method  | CP      | UE      | LE      | AL     |
|--------|---------|---------|---------|---------|--------|
|        | (a)     |         |         |         |        |
|        | 1st     | 0.842   | 0.118   | 0.041   | 11.313 |
|        | 3rd     | 0.938   | 0.035   | 0.026   | 12.490 |
|        | Bayes   | 0.970   | 0.016   | 0.014   | 14.316 |
|        | HP      | 0.904   | 0.027   | 0.024   | 18.031 |
|        | ACI5    | 0.956   | 0.034   | 0.010   | 24.343 |
|        | (b)     |         |         |         |        |
| −0.9   | 1st     | 0.849   | 0.107   | 0.044   | 8.565  |
|        | 3rd     | 0.940   | 0.035   | 0.025   | 9.071  |
|        | Bayes   | 0.968   | 0.017   | 0.015   | 10.617 |
|        | HP      | 0.909   | 0.066   | 0.025   | 12.216 |
|        | ACI5    | 0.953   | 0.034   | 0.013   | 16.492 |
| −0.8   | 1st     | 0.847   | 0.102   | 0.051   | 7.001  |
|        | 3rd     | 0.938   | 0.034   | 0.028   | 7.743  |
|        | Bayes   | 0.968   | 0.018   | 0.014   | 9.189  |
|        | HP      | 0.906   | 0.064   | 0.030   | 9.559  |
|        | ACI5    | 0.955   | 0.031   | 0.014   | 12.906 |
| −0.7   | 1st     | 0.843   | 0.101   | 0.056   | 5.865  |
|        | 3rd     | 0.939   | 0.034   | 0.027   | 6.644  |
|        | Bayes   | 0.969   | 0.017   | 0.014   | 7.967  |
|        | HP      | 0.906   | 0.062   | 0.031   | 7.785  |
|        | ACI5    | 0.954   | 0.031   | 0.015   | 10.510 |
| −0.6   | 1st     | 0.839   | 0.100   | 0.061   | 5.078  |
|        | 3rd     | 0.940   | 0.033   | 0.028   | 5.883  |
|        | Bayes   | 0.969   | 0.017   | 0.014   | 7.096  |
|        | HP      | 0.906   | 0.059   | 0.035   | 6.647  |
|        | ACI5    | 0.954   | 0.029   | 0.016   | 8.974  |
| −0.5   | 1st     | 0.840   | 0.096   | 0.064   | 4.439  |
|        | 3rd     | 0.937   | 0.034   | 0.029   | 5.206  |
|        | Bayes   | 0.967   | 0.017   | 0.015   | 6.317  |
|        | HP      | 0.905   | 0.058   | 0.038   | 5.807  |
|        | ACI5    | 0.950   | 0.031   | 0.020   | 7.840  |
| −0.4   | 1st     | 0.845   | 0.089   | 0.067   | 3.892  |
|        | 3rd     | 0.939   | 0.032   | 0.030   | 4.637  |
|        | Bayes   | 0.969   | 0.016   | 0.015   | 5.650  |
|        | HP      | 0.906   | 0.053   | 0.040   | 5.044  |
|        | ACI5    | 0.954   | 0.027   | 0.020   | 6.810  |
| −0.3   | 1st     | 0.835   | 0.088   | 0.077   | 3.426  |
|        | 3rd     | 0.938   | 0.030   | 0.032   | 4.143  |
|        | Bayes   | 0.969   | 0.013   | 0.018   | 5.066  |
|        | HP      | 0.903   | 0.050   | 0.047   | 4.417  |
|        | ACI5    | 0.951   | 0.024   | 0.025   | 5.963  |
| −0.2   | 1st     | 0.842   | 0.082   | 0.076   | 3.055  |
|        | 3rd     | 0.937   | 0.032   | 0.031   | 3.726  |
|        | Bayes   | 0.969   | 0.014   | 0.017   | 4.569  |
|        | HP      | 0.905   | 0.048   | 0.047   | 3.956  |
|        | ACI5    | 0.951   | 0.025   | 0.024   | 5.341  |
| 0      | 1st     | 0.833   | 0.083   | 0.084   | 2.762  |
|        | 3rd     | 0.937   | 0.031   | 0.032   | 3.405  |
|        | Bayes   | 0.969   | 0.015   | 0.015   | 4.186  |
| 0.1    | 1st     | 0.893   | 0.023   | 0.024   | 4.808  |
|        | 3rd     | 0.938   | 0.034   | 0.028   | 3.073  |
|        | Bayes   | 0.969   | 0.016   | 0.015   | 3.784  |
|        | HP      | 0.905   | 0.048   | 0.047   | 3.197  |
|        | ACI5    | 0.953   | 0.024   | 0.023   | 4.316  |

(continued)
Table 4. Continued.

| ρ  | Method | CP    | UE    | LE    | AL    |
|----|--------|-------|-------|-------|-------|
| 0.2| 1st    | 0.843 | 0.070 | 0.087 | 2.149 |
|    | 3rd    | 0.941 | 0.028 | 0.031 | 2.721 |
|    | Bayes  | 0.970 | 0.014 | 0.016 | 3.360 |
|    | HP     | 0.906 | 0.041 | 0.053 | 2.766 |
|    | ACI5   | 0.954 | 0.020 | 0.026 | 3.735 |
| 0.3| 1st    | 0.843 | 0.068 | 0.089 | 1.898 |
|    | 3rd    | 0.940 | 0.029 | 0.030 | 2.433 |
|    | Bayes  | 0.969 | 0.016 | 0.016 | 3.010 |
|    | HP     | 0.906 | 0.041 | 0.053 | 2.444 |
|    | ACI5   | 0.953 | 0.021 | 0.026 | 3.300 |
| 0.4| 1st    | 0.845 | 0.062 | 0.094 | 1.657 |
|    | 3rd    | 0.944 | 0.027 | 0.029 | 2.156 |
|    | Bayes  | 0.972 | 0.014 | 0.014 | 2.674 |
|    | HP     | 0.908 | 0.037 | 0.055 | 2.133 |
|    | ACI5   | 0.956 | 0.019 | 0.026 | 2.880 |
| 0.5| 1st    | 0.838 | 0.058 | 0.104 | 1.412 |
|    | 3rd    | 0.936 | 0.026 | 0.038 | 1.879 |
|    | Bayes  | 0.970 | 0.012 | 0.018 | 2.337 |
|    | HP     | 0.905 | 0.033 | 0.062 | 1.813 |
|    | ACI5   | 0.953 | 0.015 | 0.033 | 2.448 |
| 0.6| 1st    | 0.841 | 0.056 | 0.104 | 1.217 |
|    | 3rd    | 0.942 | 0.026 | 0.033 | 1.652 |
|    | Bayes  | 0.971 | 0.013 | 0.016 | 2.061 |
|    | HP     | 0.909 | 0.031 | 0.060 | 1.559 |
|    | ACI5   | 0.956 | 0.014 | 0.030 | 2.105 |
| 0.7| 1st    | 0.850 | 0.052 | 0.098 | 1.012 |
|    | 3rd    | 0.939 | 0.029 | 0.032 | 1.415 |
|    | Bayes  | 0.968 | 0.016 | 0.016 | 1.772 |
|    | HP     | 0.909 | 0.031 | 0.060 | 1.290 |
|    | ACI5   | 0.955 | 0.015 | 0.030 | 1.741 |
| 0.8| 1st    | 0.849 | 0.045 | 0.105 | 0.794 |
|    | 3rd    | 0.939 | 0.027 | 0.034 | 1.160 |
|    | Bayes  | 0.968 | 0.014 | 0.018 | 1.459 |
|    | HP     | 0.911 | 0.026 | 0.063 | 0.995 |
|    | ACI5   | 0.955 | 0.012 | 0.033 | 1.344 |
| 0.9| 1st    | 0.852 | 0.041 | 0.106 | 0.569 |
|    | 3rd    | 0.939 | 0.029 | 0.032 | 0.894 |
|    | Bayes  | 0.968 | 0.017 | 0.016 | 1.125 |
|    | HP     | 0.913 | 0.024 | 0.064 | 0.675 |
|    | ACI5   | 0.956 | 0.012 | 0.032 | 0.912 |

### 6.2.3. Simulation 3

Figure 5 shows the actual coverage of the nominal 95% confidence intervals for $\gamma$ by using the pivots $r$ (1st), $r^*$ (3rd), $r^*_B$ (Bayes) and (22) (ACI5) in the equi-correlated bivariate normal model. In terms of real coverage, $r^*$ again outperforms its first order counterpart $r$, the Bayesian solution $r^*_B$ and the approximate solution (23).

Table 5 reveals that the small-sample pivots $r^*$ and $r^*_B$ exhibit more reliable coverage than the confidence intervals obtained from their large-sample counterpart $r$ and the pivot (22), although the Bayesian solution $r^*_B$ somewhat overestimates the nominal level. Moreover, $r$ and, to a somewhat lesser extent (22), exhibit, for small-sample sizes, an unsatisfactory behaviour as far as the symmetry of errors is concerned, although the ACI5 confidence intervals seem to improve as soon as $n \geq 10$ (see the Supplementary Material). In addition, $r^*_B$ and Equation (22) produce confidence intervals with, respectively, larger and shorter expected widths as compared with the intervals obtained from $r^*$ and $r$. The finite-sample differences among the four pivots vanish as the sample size increases. For all four methods, the expected length becomes wider for negative values of $\rho$, especially when $\rho$ approaches $-1$. 


7. Concluding remarks

In this paper we investigate the behaviour of modern likelihood-based small-sample procedures to compute confidence intervals for the parameter of skewness which characterizes the distribution of the maximum/minimum of a bivariate normal and exchangeable random vector. This distribution may, for example, be used for assessing the degree of concordance of a continuous mono-zygotic twin trait when interest focuses on the pairwise maximum or minimum, as in Section 5.1. Or, it may represent the reference model for making inference on contra-lateral measurements, as in Section 5.2.

Extensive numerical investigation revealed that the higher order frequentist pivot $r^*$ is highly accurate, especially for the rather small-sample sizes which may be encountered, and for the challenging situation where $\rho$ is close to $-1$. This is in agreement with the findings by Sun and Wong [30], though their contribution focuses on $\rho$ and does not consider the custom-tailored statistics of Section 4.2. When no nuisance parameter is present, $r^*$ yields confidence intervals which, for practical purposes, may be considered exact. Among the four alternatives available in the presence of nuisance parameters, the only real competitor to $r^*$, in terms of both real coverage and required computational efforts, is the ACI3 confidence interval, though it leads to, on average, longer confidence intervals which counterbalance the lack of symmetry on the tails. The potential applicability of the ACI3 method to studies on twins was already put forward in [11]. The construction of confidence sets for $\gamma$ in the three-parameter bivariate normal model shows that the small-sample frequentist confidence interval based on $r^*$ outperforms its three alternatives with respect to both average length and symmetry of errors. The ACI3 confidence interval requires a minor computational effort than the likelihood-based.
Table 5. Summary statistics for Simulation 3: bivariate normal with means 0.7 and variances 0.1. Empirical coverage (CP), upper (UE) and lower (LE) error probability and average length (AL) of nominal two-sided 95% confidence intervals for γ, for varying values of ρ and sample size n = 5. Pivots used: likelihood root r (1st), modified likelihood root r* (3rd); Bayesian modified likelihood root r* Bayes; expression (22) (ACI 3). Based on 10,000 replicates; simulation error: ±0.004.

| ρ   | Method | CP   | UE   | LE   | AL    |
|-----|--------|------|------|------|-------|
| (a) |        |      |      |      |       |
| −0.9| 1st    | 0.916| 0.066| 0.018| 10.378|
|     | 3rd    | 0.951| 0.028| 0.021| 10.671|
|     | Bayes  | 0.972| 0.015| 0.013| 12.069|
|     | ACI 3 | 0.904| 0.058| 0.038| 10.431|
| −0.8| 1st    | 0.913| 0.067| 0.020| 8.075 |
|     | 3rd    | 0.951| 0.026| 0.023| 8.208 |
|     | Bayes  | 0.972| 0.014| 0.014| 9.445 |
|     | ACI 3 | 0.904| 0.056| 0.040| 7.105 |
| −0.7| 1st    | 0.911| 0.066| 0.023| 6.641 |
|     | 3rd    | 0.949| 0.026| 0.025| 6.722 |
|     | Bayes  | 0.972| 0.013| 0.015| 7.809 |
|     | ACI 3 | 0.901| 0.058| 0.041| 5.620 |
| −0.6| 1st    | 0.912| 0.068| 0.020| 5.717 |
|     | 3rd    | 0.949| 0.028| 0.023| 5.776 |
|     | Bayes  | 0.973| 0.014| 0.013| 6.748 |
|     | ACI 3 | 0.903| 0.058| 0.039| 4.794 |
| −0.5| 1st    | 0.913| 0.066| 0.021| 4.996 |
|     | 3rd    | 0.952| 0.024| 0.024| 5.041 |
|     | Bayes  | 0.973| 0.013| 0.014| 5.912 |
|     | ACI 3 | 0.906| 0.056| 0.038| 4.142 |
| −0.4| 1st    | 0.905| 0.074| 0.021| 4.485 |
|     | 3rd    | 0.944| 0.031| 0.024| 4.528 |
|     | Bayes  | 0.970| 0.017| 0.013| 5.314 |
|     | ACI 3 | 0.895| 0.064| 0.040| 3.731 |
| −0.3| 1st    | 0.909| 0.070| 0.021| 3.927 |
|     | 3rd    | 0.947| 0.028| 0.024| 3.962 |
|     | Bayes  | 0.973| 0.014| 0.012| 4.656 |
|     | ACI 3 | 0.899| 0.060| 0.041| 3.264 |
| −0.2| 1st    | 0.913| 0.066| 0.021| 3.532 |
|     | 3rd    | 0.950| 0.028| 0.022| 3.566 |
|     | Bayes  | 0.971| 0.015| 0.014| 4.194 |
|     | ACI 3 | 0.904| 0.057| 0.038| 2.942 |
| −0.1| 1st    | 0.912| 0.068| 0.020| 3.216 |
|     | 3rd    | 0.950| 0.028| 0.022| 3.247 |
|     | Bayes  | 0.974| 0.014| 0.012| 3.819 |
|     | ACI 3 | 0.903| 0.060| 0.037| 2.675 |
| 0    | 1st    | 0.910| 0.070| 0.021| 2.870 |
|     | 3rd    | 0.949| 0.028| 0.022| 2.898 |
| (b)  | Bayes  | 0.971| 0.015| 0.014| 3.409 |
|     | ACI 3 | 0.900| 0.061| 0.039| 2.381 |
| 0.1  | 1st    | 0.909| 0.069| 0.021| 2.615 |
|     | 3rd    | 0.945| 0.030| 0.025| 2.640 |
|     | Bayes  | 0.971| 0.014| 0.014| 3.106 |
|     | ACI 3 | 0.900| 0.060| 0.040| 2.175 |
| 0.2  | 1st    | 0.910| 0.067| 0.023| 2.330 |
|     | 3rd    | 0.947| 0.028| 0.025| 2.353 |
|     | Bayes  | 0.972| 0.013| 0.016| 2.769 |
|     | ACI 3 | 0.900| 0.059| 0.041| 1.939 |
| 0.3  | 1st    | 0.912| 0.066| 0.021| 2.097 |
|     | 3rd    | 0.949| 0.026| 0.024| 2.118 |
|     | Bayes  | 0.973| 0.013| 0.014| 2.493 |
|     | ACI 3 | 0.902| 0.058| 0.040| 1.746 |
| 0.4  | 1st    | 0.914| 0.066| 0.020| 1.884 |
|     | 3rd    | 0.951| 0.027| 0.022| 1.904 |
|     | Bayes  | 0.974| 0.014| 0.012| 2.240 |
|     | ACI 3 | 0.906| 0.058| 0.036| 1.571 |
Table 5. Continued.

| $\rho$ | Method | CP    | UE    | LE    | AL    |
|-------|--------|-------|-------|-------|-------|
| 0.5   | 1st    | 0.907 | 0.073 | 0.020 | 1.684 |
|       | 3rd    | 0.944 | 0.033 | 0.023 | 1.702 |
|       | Bayes  | 0.970 | 0.017 | 0.013 | 2.003 |
|       | ACI3   | 0.895 | 0.065 | 0.040 | 1.405 |
| 0.6   | 1st    | 0.909 | 0.069 | 0.022 | 1.436 |
|       | 3rd    | 0.947 | 0.027 | 0.026 | 1.453 |
|       | Bayes  | 0.972 | 0.014 | 0.014 | 1.710 |
|       | ACI3   | 0.897 | 0.061 | 0.042 | 1.198 |
| 0.7   | 1st    | 0.909 | 0.070 | 0.021 | 1.212 |
|       | 3rd    | 0.945 | 0.031 | 0.024 | 1.228 |
|       | Bayes  | 0.973 | 0.014 | 0.013 | 1.446 |
|       | ACI3   | 0.898 | 0.063 | 0.039 | 1.008 |
| 0.8   | 1st    | 0.908 | 0.073 | 0.019 | 0.969 |
|       | 3rd    | 0.947 | 0.030 | 0.022 | 0.986 |
|       | Bayes  | 0.971 | 0.016 | 0.013 | 1.162 |
|       | ACI3   | 0.897 | 0.064 | 0.040 | 0.801 |
| 0.9   | 1st    | 0.914 | 0.066 | 0.020 | 0.686 |
|       | 3rd    | 0.949 | 0.028 | 0.023 | 0.705 |
|       | Bayes  | 0.973 | 0.014 | 0.013 | 0.831 |
|       | ACI3   | 0.905 | 0.057 | 0.038 | 0.548 |

However, despite the fact that its construction borrows from [11], it performs poorly in terms of both coverage probability and symmetry of error rates. This is especially true for small-sample sizes.

The straightforward extension of our proposal is to the skew-normal model. For instance, Loperfido’s theorem provides the reference models for mono-zygotic twin studies for which information on the pair $(X_1, X_2)$ is missing, and only their maximum (or minimum) value is recorded. This may happen because of practical reasons; see [9] for a rather early treatment. As pointed out there, because healthy mono-zygotic twins share an identical genetic mark-up, time of onset for a particular event in the first twin – such as getting a cold or developing leukaemia – is likely to closely follow in the second twin, so that only the smaller or larger record may be kept. The same holds true for clinical comparative studies. Furthermore, working with the maximum (or minimum) of two correlated measurements can be, at times, more reliable than the study of the original values, especially if the measurements of the smaller (or larger) values are more accurate.

A further possible line of research is the investigation of small-sample asymptotic theory in the non-standard setting of finite skew-normal mixtures (see, for instance,[36,37]).

Supplementary Material

The Supplementary Material contains additional results for the simulation studies of Section 6, and the R code used in Examples 5.1 (twins.R) and 5.2 (longitudinal.R). As permission to use the data of Example 5.1 must be required from the Authors of [13], the code in twins.R reproduces the analysis using simulated data from a bivariate normal model with vector of parameters $\hat{\theta} = (0.900, 7.061, 6.924, 0.905, 0.872)$.

Acknowledgments

The data on twins used in Section 5.1 were obtained from the electronic software repository StatLib. We would like to express our gratitude to [13] who have submitted the data set on twins in this archive and have given us the authorization to use them. The authors furthermore acknowledge the useful comments by the Associate Editor and by an anonymous Referee, which led to the improved version of the paper.

Disclosure statement

No potential conflict of interest was reported by the authors.
Funding

This research was partially supported by the University of Padova grant no. CPDA101912.

References

[1] Brazzale AR, Davison AC, Reid N. Applied asymptotics: case studies in small-sample statistics. Cambridge (UK): Cambridge University Press; 2007.
[2] Viana MAG, Lee H-M. Correlation analysis of ordered symmetrically dependent observations and their concomitants of order statistics. Can J Stat. 2006;34:327–340.
[3] Loperfido N. Modeling maxima of longitudinal contralateral observations. Test. 2008;17:370–380.
[4] Sheikhi A, Tata M. The exact joint distribution of concomitants of order statistics and their order statistics under normality. REVSTAT Statist J. 2013;11:121–134.
[5] Galton F. The history of twins, as a criterion of the relative powers of nature and nurture. J Anthropological Inst Great Britain Ireland. 1876;5:391–406.
[6] Kraemer HC. What is the ‘right’ statistical measure of twin concordance (or diagnostic reliability and validity)? Arch Gen Psychiatry. 1997;54:1121–1124.
[7] Lyons MJ, Faraone SV, Tsuang MT, Goldberg J, Eaves LJ, Meyer JM, True WR, Eisen SA. Another view on the ‘right’ statistical measure of twin concordance. Arch Gen Psychiatry. 1997;54:1126–1128.
[8] Azzalini A. A class of distributions which includes the normal ones. Scand J Stat. 1985;12:171–178.
[9] Roberts C. A correlation model useful in the study of twins. J Amer Statist Assoc. 1966;61:1184–1190.
[10] Loperfido N. A note on skew-elliptical distributions and linear functions of order statistics. Stat Probab Lett. 2008;78:3184–3186.
[11] Mameli V, Musio M, Saleau E, Biggeri A. Large sample confidence intervals for the skewness parameter of the skew-normal distribution based on Fisher’s transformation. J Appl Stat. 2012;39:1693–1702.
[12] Fraser DAS, Reid N, Wu J. A simple general formula for tail probabilities for frequentist and Bayesian inference. Biometrika. 1999;86:249–264.
[13] Tramo MJ, Loftus WC, Stukel TA, Green RL, Weaver JB, Gazzaniga MS. Brain size, head size, and IQ in monozygotic twins. Neurology. 1998;50:1246–1252.
[14] Johnson RA, Wichern DW. Applied multivariate statistical analysis. 5th ed. Englewood Cliffs, NJ: Prentice-Hall; 2002.
[15] Barndorff-Nielsen O. On a formula for the distribution of the maximum likelihood estimator. Biometrika. 1983;70:343–365.
[16] Fraser DAS, Reid N. Ancillary and third order significance. Utilitas Math. 1995;47:33–53.
[17] DiCiccio TJ, Martin MA. Approximations of tail probabilities for a class of smooth functions with applications to Bayesian and conditional inference. Biometrika. 1991;78:891–902.
[18] Welch BL, Peers HW. On formulæ for confidence points based on integrals of weighted likelihoods. J R Statist Soc Ser B. 1963;25:318–329.
[19] Staicu A-M, Reid NM. On probability matching priors. Can J Stat. 2008;36:613–622.
[20] Azzalini A, Capitanio A. Statistical applications of the multivariate skew-normal distributions. J R Statist Soc Ser B. 1999;61:579–602.
[21] Sartori N. Bias prevention of maximum likelihood estimates for scalar skew normal and skew t distributions. J Statist Plan Inference. 2006;136:4259–4275.
[22] Azzalini A. R package SN: The skew-normal and skew-t distributions (version 0.4-18). Available from: http://azzalini.stat.unipd.it/SN.
[23] Genton MG. Skew-elliptical distributions and their applications: a journey beyond normality. Boca Raton (FL): Chapman & Hall / CRC; 2004.
[24] Azzalini A, Capitanio A. The skew-normal and related families. Cambridge (UK): Cambridge University Press; 2014.
[25] Nagaraja HN. A note on linear functions of ordered correlated normal random variables. Biometrika. 1982;69:284–285.
[26] Loperfido N. Statistical implications of selectively reported inferential results. Stat Probab Lett. 2002;56:13–22.
[27] Haddad JN, Provost SB. Approximations to the distribution of the sample correlation coefficient. World Acad Sci, Eng Technol. 2011;52:910–915.
[28] Reid N. Asymptotics and the theory of inference. Ann Stat. 2003;31:1695–1731.
[29] Reid N, Fraser DAS. Mean log-likelihood and higher-order approximations. Biometrika. 2010;97:159–170.
[30] Sun Y, Wong ACM. Interval estimation for the normal correlation coefficient. Stat Probab Lett. 2007;77:1652–1661.
[31] Gosh M. Objective priors: an introduction for frequentists. Statist Sci.. 2011;26:187–202.
[32] Gosh M, Mukherjee B, Santra U, Kim D. Bayesian and likelihood-based inference for the bivariate normal correlation coefficient. J Statist Plan Inference. 2010;140:1410–1416.
[33] Fisher RA. On the ‘probable error’ of a coefficient of correlation deduced from a small sample. Metron. 1921;1:3–32.
[34] Konishi S. Normalizing and variance stabilizing transformation for intraclass correlations. Ann Instit Statist Math. 1985;37:87–94.
[35] R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria; 2013. Available from: http://www.R-project.org.
[36] Lin TI. Maximum likelihood estimation for multivariate skew normal mixture models. J Multivariate Anal. 2009;100:257–265.
[37] Lin TI, Lee JC, Yen SY. Finite mixture modelling using the skew normal distribution. Statist Sin. 2007;17:909–927.

**Appendix**

| Symbol | Description                                      | Definition                      |
|--------|--------------------------------------------------|---------------------------------|
| $\rho$ | Correlation coefficient of the bivariate normal distribution | –                               |
| $\gamma$ | Skewness parameter of the skew-normal distribution | $\gamma = \pm \sqrt{(1 - \rho)/(1 + \rho)}$ |
| $\psi$ | Scalar interest parameter | –                               |
| $\lambda$ | $(k - 1)$-dimensional nuisance parameter | –                               |
| $\theta$ | $k$-dimensional parameter | $\theta = (\psi, \lambda)$ |
| $l$ | Log-likelihood function | –                               |
| $\pi$ | Prior | –                               |
| $j$ | Observed information function | –                               |
| $i$ | Expected information function | –                               |
| $\hat{\psi}$ | Maximum likelihood estimate of $\psi$ | –                               |
| $\hat{\lambda}$ | Maximum likelihood estimate of $\lambda$ | –                               |
| $\hat{\theta}$ | Maximum likelihood estimate of $\theta$ | –                               |
| $\hat{\lambda}_\psi$ | Constrained maximum likelihood estimate of $\lambda$ | $\hat{\lambda}_\psi = \arg\max_\lambda l(\hat{\psi}, \lambda)$ |
| $\hat{\theta}_\psi$ | Constrained maximum likelihood estimate of $\theta$ | $\hat{\theta}_\psi = (\hat{\psi}, \hat{\lambda}_\psi)$ |
| $l_\rho$ | Profile log-likelihood | $l_\rho(\psi) = l(\hat{\theta}_\psi)$ |
| $j_\rho$ | Profile observed information function | $j_\rho(\psi) = d^2 l_\rho(\psi)/d\psi^2$ |
| $r$ | Signed likelihood root | (1)                             |
| $\ast_r$ | Modified likelihood root | (3)                             |
| $\psi$ | Canonical parameter | (5)                             |
| $q$ | Frequentist correction term | (6)                             |
| $q_B$ | Bayesian correction term | (7)                             |
| $\ast_B$ | Bayesian modified likelihood root | (8)                             |