

Small-sample likelihood asymptotics for the equi-correlated bivariate normal model

Inferenza asintotica di ordine superiore per il modello normale bivariato equicorrelato

Alessandra R. Brazzale and Valentina Mameli

Abstract We discuss two likelihood-based small-sample confidence intervals for the skewness parameter of the distribution of the maximum (or minimum) of the equi-correlated bivariate normal model. These are compared numerically to their large-sample counterpart, and to an approximate confidence interval whose construction derives from theoretical findings on the intraclass correlation coefficient and Fisher's transformation. The performance of the confidence intervals is analyzed in terms of actual coverage, symmetry of errors, and expected length. A simulation revealed that the considered small-sample procedures perform well even when the sample size is limited. A real-life application to a mono-zygotic twin study is also given.

Abstract *Nel presente lavoro proponiamo intervalli di confidenza asintotici per il parametro di asimmetria della distribuzione del massimo in un modello normale bivariato equicorrelato, con particolare attenzione alla loro applicazione allo studio di gemelli monozigoti.*

Key words: Equi-correlated bivariate normal distribution; modified likelihood ratio; intraclass correlation coefficient; skew-normal distribution.

1 Introduction

The skew-normal distribution, introduced in 1985 by Azzalini [1], represents an asymmetric extension of the normal distribution. Owing to Loperfido [8], who proved that the maximum (or minimum) of a bivariate exchangeable normal random vector is skew-normally distributed with skewness parameter γ (or $-\gamma$) which depends on the correlation coefficient ρ , this distribution can be used to model the

Alessandra R. Brazzale

Department of Statistical Sciences, University of Padua, Italy, e-mail: alessandra.brazzale@unipd.it

Valentina Mameli

Department of Mathematics and Computer Science, University of Cagliari, Italy, e-mail: vmameli@unica.it

maximum (or minimum) of two correlated measurements. This paper emphasizes a present-day application of this model to mono-zygotic twin studies. Here, the need for simple, though reliable procedures for assessing the degree of concordance of a continuous mono-zygotic twin trait was already underlined by [13].

The aim of this paper is to explore the performance of two small-sample confidence intervals for γ which we obtained by applying the higher order frequentist and Bayesian pivots presented in [6] to the three-parameter equi-correlated bivariate normal model. These are compared to their large-sample counterpart, together with a newly derived large-sample confidence interval which borrows from Mameli et al. [9] and uses results from Loperfido [8] and Fisher's transformation. Our findings extend the work by Mameli and Brazzale [10] on higher order likelihood-based inference for γ in the five-parameter bivariate normal model. The extensive numerical work given there revealed that the small-sample frequentist confidence intervals outperform those proposed in [9] with respect to both average length and symmetry of errors, although they behave quite similarly as far as real coverage goes. This is especially true for very limited sample sizes and if ρ is close to -1 . On the other hand, the calculation of the large-sample solution of [9] requires a by far less computational effort.

The paper is organized as follows. Section 2 reviews modern likelihood-based small-sample asymptotics and the equi-correlated bivariate normal model. The application of our results to a dataset of twin measurements is illustrated in Section 3. Section 4 closes the paper with the discussion of the findings from a simulation study.

2 Background theory

Likelihood asymptotics Consider a parametric statistical model with density function $f(y; \theta)$, where $y = (y_1, \dots, y_n)$ is a vector of n observations and $\theta = (\psi, \lambda)$ a k -dimensional parameter, with ψ the scalar parameter of interest and λ a vector of nuisance parameters of dimension $k - 1$. Let $L(\theta) \propto f(y; \theta)$ and $l(\theta) = \log L(\theta)$ denote the likelihood and the log-likelihood functions, respectively. Inference on ψ is commonly based on the signed likelihood root

$$r(\psi) = \text{sign}(\hat{\psi} - \psi) \sqrt{2(l_p(\hat{\psi}) - l_p(\psi))},$$

where $l_p(\psi) = l(\hat{\theta}_\psi)$ denotes the profile log-likelihood and $\hat{\theta}_\psi = (\psi, \hat{\lambda}_\psi)$ represents the constrained maximum likelihood estimate. The signed likelihood root pivot is asymptotically normal up to the order $n^{-1/2}$. As nowadays well known, this approximation may be rather poor if the sample size is limited.

Inference on ψ may be improved by adjusting the signed likelihood root as in [2]. The modified signed likelihood root

$$r^* = r + \frac{1}{r} \log \left(\frac{q}{r} \right) \quad (1)$$

is asymptotically standard normal up to the order $n^{-3/2}$. In the Bayesian framework, a similar expression

$$r_B^* = r + \frac{1}{r} \log \left(\frac{q_B}{r} \right), \quad (2)$$

provides a third order approximation

$$\Pi(\psi_0 | y) = \Pr(\psi \leq \psi_0 | y) \doteq 1 - \Phi(r_B^*),$$

for posterior quantiles for the parameter ψ . The correction term q_B in (2) depends on both the underlying model and the elicited prior for the parameter θ , whereas the frequentist counterpart q in (1) relies only on the model; see expressions (4) and (3) below and [11] for an extensive review and list of reference. In addition, inference based on r and r^* is invariant under interest-respecting re-parametrizations.

The equi-correlated bivariate normal model Let $(x_1, y_1), \dots, (x_n, y_n)$ be an independent sample from an equi-correlated bivariate normal vector (X, Y) , where X and Y have common mean μ , common variance σ^2 and correlation coefficient ρ . The corresponding log-likelihood function for $\theta = (\rho, \mu, \sigma)$ is

$$l(\theta) = -n \left(\log[(1 - \rho^2)^{1/2} \sigma^2] + \frac{\mu^2}{(1 + \rho)\sigma^2} \right) + \frac{n\mu(\bar{x} + \bar{y})}{(1 + \rho)\sigma^2} + \frac{2\rho \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i^2 - \sum_{i=1}^n y_i^2}{2(1 - \rho^2)\sigma^2},$$

which characterizes a full exponential family with canonical parameter $\varphi(\theta) = (\mu/[\sigma^2(1 + \rho)], \rho/[\sigma^2(1 - \rho^2)], -1/[2(1 - \rho^2)\sigma^2])$. Maximum likelihood yields the estimate $\hat{\theta}$, whose components are

$$\hat{\mu} = \frac{1}{2n} \sum_{i=1}^n (x_i + y_i), \quad \hat{\sigma} = \sqrt{\frac{1}{2n} \sum_{i=1}^n [(x_i - \hat{\mu})^2 + (y_i - \hat{\mu})^2]}, \quad \hat{\rho} = \frac{1}{n\hat{\sigma}^2} \sum_{i=1}^n (x_i - \hat{\mu})(y_i - \hat{\mu}).$$

Fisher [5] called ρ the intraclass correlation coefficient and its maximum likelihood estimate $\hat{\rho}$ the intraclass correlation.

Confidence intervals for γ Our interest lies on $\gamma = \sqrt{(1 - \rho)/(1 + \rho)}$, the parameter which characterizes the skewness of the distribution of the maximum between the two correlated normal variates X and Y .

Likelihood-based confidence intervals are obtained via pivot profiling (see [3]) from the pivots r , r^* , and r_B^* . The correction term in (1) becomes

$$q = \frac{|\varphi(\hat{\theta}) - \varphi(\hat{\theta}_\psi)| \quad \varphi_\lambda(\hat{\theta}_\psi)|}{|\varphi_\theta(\hat{\theta})|} \left\{ \frac{|j(\hat{\theta})|}{|j_{\lambda\lambda}(\hat{\theta}_\psi)|} \right\}^{\frac{1}{2}}, \quad (3)$$

where $\varphi_\theta(\theta)$ is the matrix of partial derivatives of $\varphi(\theta)$ with respect to θ , while $\varphi_\lambda(\theta)$ represents the matrix of partial derivatives of $\varphi(\theta)$ with respect to λ . Similarly, $j_{\lambda\lambda}(\theta)$ is the $(k-1) \times (k-1)$ sub-matrix of the observed information function $j(\theta)$ with respect to λ ; see [6]. The Bayesian analogue

$$q_B = l'_p(\psi) j_p(\psi)^{-\frac{1}{2}} \left\{ \frac{|j_{\lambda\lambda}(\hat{\theta}_\psi)|}{|j_{\lambda\lambda}(\hat{\theta})|} \right\}^{\frac{1}{2}} \frac{\pi(\hat{\theta})}{\pi(\hat{\theta}_\psi)}, \quad (4)$$

where $l'_p(\psi) = dl_p(\psi)/d\psi$ is the profile score function and $j_p(\psi) = -d^2l_p(\psi)/d\psi^2$ the profile observed information function, was given in [4]. Here, we assume for $\pi(\theta)$ the “unique prior”, defined in [14], which requires an orthogonal re-parametrization of the model of the form

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

with $\psi = \rho$, $\eta_1 = \mu$ and $\eta_2 = \sigma^2(1 - \rho^2)^{1/2}$.

We also obtained a further confidence interval for γ , which is accurate up to the second order. Its derivation is equivalent to the one discussed in [9], 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

$$\sqrt{n} \left[\frac{1}{2} \log \left(\frac{1+\hat{\rho}}{1-\hat{\rho}} \right) - \frac{1}{2} \log \left(\frac{1+\rho}{1-\rho} \right) + \frac{1}{2n} \right], \quad (5)$$

is asymptotically standard normal up to the order n^{-1} ; see [7]. Expression (5) leads straightforwardly to an asymptotic confidence interval for γ .

3 Real data example

Here, we consider the twin study conducted by Tramo et al. [15]. This study gives, among others, the head circumference of ten pairs of mono-zygotic twins for which bivariate normality is supported by the corresponding Shapiro-Wilk test (p -value = 0.198). We are interested in the parameter γ which characterizes the skewness of the distribution of the maximum between the head circumference in the first twin and in the second twin. Maximum likelihood yields $\hat{\gamma} = 0.270$. The interval based on the pivot (5), here *ACI*, is shorter than the likelihood-based confidence intervals obtained from r (1st), r^* (3rd), and r_B^* (Bayes); see Table 1.

Table 1 Lower (LB) and upper (UB) bounds of 95% confidence intervals for γ .

Method	LB	UB	Length
1st	0.142	0.511	0.369
3rd	0.129	0.499	0.370
Bayes	0.118	0.470	0.353
ACI	0.170	0.408	0.238

The same analysis carried out on a second variable of the same dataset, this time the *corpus callosum* surface area of the twins, led to identical conclusions and inspired the simulation study described below.

4 Simulation study

We conducted a simulation study to assess the behaviour of nominal 95% confidence intervals based upon the small-sample pivots r^* and r_B^* , and to compare them with those obtained from the large-sample counterpart r and the second order pivot (5). The accuracy of the confidence intervals is evaluated in terms of empirical coverage (CP), upper (UE) and lower (LE) error probability, and average length (AL). We consider 10,000 bivariate normal random samples with $\mu = 7$ and $\sigma = 0.9$ for the four sample sizes $n = 5, 10, 15, 20$, while the values of ρ range from a moderate to a strong correlation.

Figure 1 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 (5), although the Bayesian solution r_B^* somewhat overestimates the

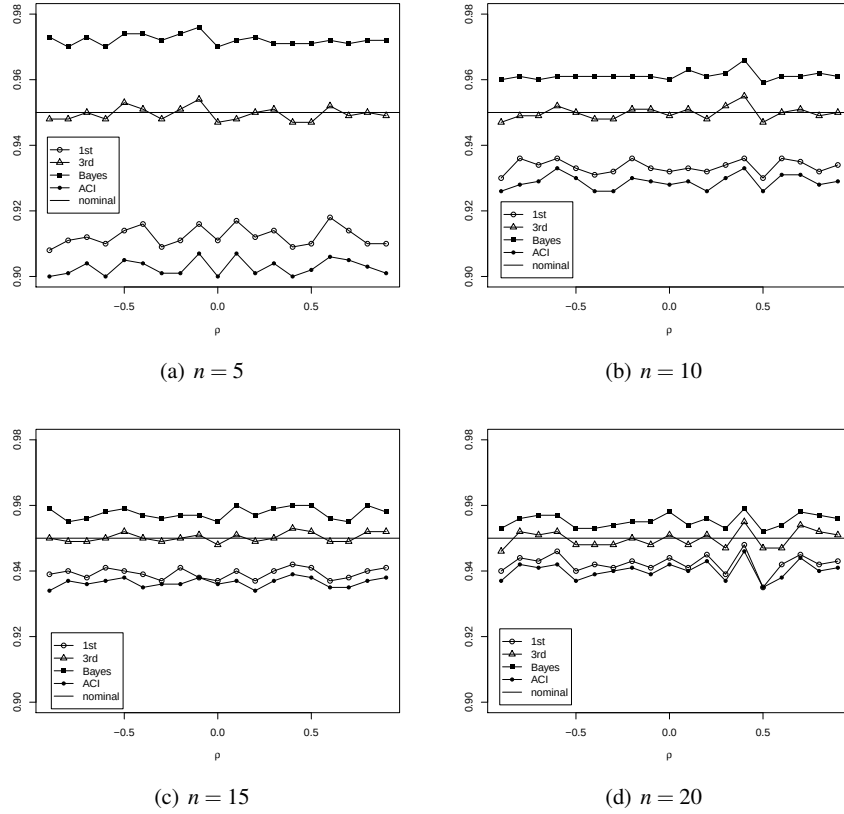


Fig. 1 Empirical coverage of the 95% confidence intervals for γ for varying values of ρ . Legend: $\circ$ — 1st order; $\triangle$ — 3rd order; $\blacksquare$ — Bayes; $\bullet$ — ACI; — nominal.

nominal level. Unreported results indicate that r and, to a somewhat lesser extent (5), exhibit, for small sample sizes, an unsatisfactory behaviour as far as the symmetry of errors is concerned, although the confidence intervals based on (5) seem to improve as soon as $n \geq 10$. In addition, r_B^* and (5) produce confidence intervals with, respectively, larger and shorter expected width 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 ρ , especially when ρ approaches -1 .

Acknowledgements The data on twins used in Section 3 were obtained from the electronic software repository *StatLib*. We would like to thank the authors of [15] for uploading the dataset and authorizing us to use them. All calculations were carried out using the numerical computing environment R (www.r-project.org). This research was partially supported by *Università di Padova – Progetto di Ricerca di Ateneo 2010* grant no. CPDA101912. The authors acknowledge helpful suggestions from an anonymous referee.

References

1. Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, **12**, 171–178.
2. Barndorff-Nielsen, O. (1983). On a formula for the distribution of the maximum likelihood estimator. *Biometrika*, **70**, 343–365.
3. Brazzale, A. R., Davison, A. C. and Reid, N. (2007). *Applied Asymptotics: Case Studies in Small-Sample Statistics*. Cambridge University Press.
4. DiCiccio, T. J. and Martin, A. (1991). Approximations of tail probabilities for a class of smooth functions with applications to Bayesian and conditional inference. *Biometrika*, **78**, 891–902.
5. Fisher, R. A. (1921). On the “probable error” of a coefficient of correlation deduced from a small sample. *Metron*, **1**, 3–32.
6. Fraser, D. A. S., Reid, N. and Wu, J. (1999). A simple general formula for tail probabilities for frequentist and Bayesian inference. *Biometrika*, **86**, 249–264.
7. Konishi, S. (1985). Normalizing and variance stabilizing transformation for intraclass correlations. *Annals of the Institute of Statistical Mathematics*, **37**, 87–94.
8. Loperfido, N. (2002). Statistical implications of selectively reported inferential results. *Statistics & Probability Letters*, **56**, 13–22.
9. Mameli, V., Musio, M., Saleau, E. and Biggeri, A. (2012). Large sample confidence intervals for the skewness parameter of the skew-normal distribution based on Fisher’s transformation. *Journal of Applied Statistics*, **39**, 1693–1702.
10. Mameli, V. and Brazzale, A. R. (2013). Modern likelihood inference for the parameter of skewness: An application to mono-zygotic twin studies. Working Paper Series, N. 10, December 2013, Department of Statistical Sciences, University of Padova. <http://www.stat.unipd.it/~brazzale> (main Publications – Submitted).
11. Reid, N. (2003). Asymptotics and the theory of inference. *Annals of Statistics*, **31**, 1695–1731.
12. Reid, N. and Fraser, D. A. S. (2010). Mean log-likelihood and higher-order approximations. *Biometrika*, **97**, 159–170.
13. Smith, C. (1974). Concordance in twins: methods and interpretation. *The American Journal of Human Genetics*, **26**, 454–466.
14. Staicu, A. M. and Reid, N. (2008). On probability matching priors. *The Canadian Journal of Statistics*, **36**, 613–622.
15. Tramo, M. J., Loftus, W. C., Green, R. L., Stukel, T. A., Weaver, J. B. and Gazzaniga, M. S. (1998). Brain size, head size, and IQ in monozygotic twins. *Neurology*, **50**, 1246–1252.