Small-sample inference on measures of concordance for the Gaussian bivariate copula with emphasis on Gini's gamma index

Inferenza asintotica per piccoli campioni per l'indice gamma di Gini nella copula bivariata Gaussiana

Mameli, V. and Brazzale, A. R.

Abstract Likelihood-based small-sample procedures to compute confidence intervals for measures of concordance in a Gaussian bivariate copula are presented, with special emphasis on Gini's gamma index.

Abstract *Nel presente lavoro proponiamo l'uso di metodi asintotici per piccoli campioni basati sulla funzione di verosimiglianza per derivare intervalli di confidenza per l'indice di Gini in una copula bivariata normale.*

Key words: Equi-correlated bivariate model, Gaussian copula, Gini's index, modified signed likelihood root, signed likelihood root

1 Introduction

Copula functions are useful tools to construct bivariate distributions as well as multivariate ones; see [11]. They are key tools in numerous fields of application as diverse as Biology, Genetics, and medical research to model dependent random variables. In these fields small sample sizes are rather common. However, as it is well known, inference based on the classical first order approximations may produce unreliable results when sample sizes are small. Higher order approximations then provide appreciably better solutions; see [2] and reference therein. Our research will address the construction of small-sample confidence intervals based on the findings of [4] for measures of concordance of a bivariate Gaussian copula from both the frequen-

Valentina Mameli
Department of Environmental Sciences, Informatics and Statistics, University Ca' Foscari of Venice, Italy, e-mail: mameli.valentina@virgilio.it

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

tist and the Bayesian point of view.

More specifically, the problem we are going to tackle is defined as follows. Let $F(x,y)$ be a joint cumulative distribution function with marginal cumulative distributions $F_1(x)$ and $F_2(y)$. For all real (x,y), the bivariate normal copula C is defined as

$$C(F_1(x),F_2(y);\rho) = \Phi_2\left(\Phi^{-1}(F_1(x)),\Phi^{-1}(F_2(y));\rho\right), \quad \rho \in (-1,1), \tag{1}$$

where $\Phi_2(\cdot,\cdot;\rho)$ and $\Phi(\cdot)$ are respectively the distribution function of the bivariate standard normal distribution with correlation parameter ρ and the distribution function of the univariate standard normal distribution. The Gaussian copula reduces to the bivariate standard normal distribution when $F_1(x) = \Phi(x)$ and $F_2(y) = \Phi(y)$, i.e. when both margins are standard normal.
Among the most important measures of concordance for the bivariate normal copula we focus our attention on the construction of small-sample confidence intervals for Gini's gamma index [5], whose expression can be found in closed form and depends on the correlation coefficient ρ (see [7]):

$$\gamma(\rho) = \frac{4}{\pi}\arcsin\left(\frac{1}{4}\left(\sqrt{(1+\rho)(3+\rho)} - \sqrt{(1-\rho)(3-\rho)}\right)\right). \tag{2}$$

Although here we concentrate on Gini's index, our proposal directly applies to other measure of concordance such as Kendall's tau, Spearman's rho, and Blomqvist's beta. Note that for the bivariate Gaussian copula, Kendall's tau and Blomqvist's beta coincide ([7]). Note also that all these measures are completely determined by the copula, i.e. they are independent of the margins ([10]).

The paper is organized as follows. Section 2 reviews likelihood-based small-sample asymptotics. Section 3 investigates the performance of our method through a simulation study.

2 Background theory

Given a parametric statistical model with density function $f(y;\theta)$, let $y = (y_1,\dots,y_n)$ be a vector of n independent observations and $\theta = (\psi,\lambda)$ a k-dimensional parameter, where ψ is the scalar interest parameter and λ is a $k-1$-dimensional nuisance parameter. Let $L(\theta) \propto f(y;\theta)$ and $l(\theta) = \log L(\theta)$ denote the likelihood and the log-likelihood functions, respectively. Inference for the parameter ψ can be 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 $n^{-1/2}$. It provides satisfactory approximations for large sample sizes, but can be rather inaccurate for small ones. Improvements can be obtained by using the modified signed likelihood root ([1])

$$r^*(\psi) = r(\psi) + \frac{1}{r(\psi)} \log\left(\frac{q(\psi)}{r(\psi)}\right), \tag{3}$$

which is asymptotically standard normal up to the order $n^{-3/2}$. From the Bayesian point of view, a similar expression, which is asymptotically standard normal up to the order $n^{-3/2}$, is given by

$$r_B^*(\psi) = r(\psi) + \frac{1}{r(\psi)} \log\left(\frac{q_B(\psi)}{r(\psi)}\right). \tag{4}$$

For the expressions of the correction terms q_B in (4) and q in (3) we refer the reader to [4], [8], [2] and reference therein.

3 Confidence intervals for Gini's gamma index

In this section we consider the case where marginals are standard normals. By using the results found for the correlation coefficient ρ in the equi-correlated bivariate normal (see [3]) and the results for the bivariate normal (see [13] and [6]), combined with the invariance property of r and r^* under interest-respecting re-parametrizations, we derive confidence intervals for Gini's gamma index. The method also applies to any other measure of concordance for the bivariate Gaussian normal copula. Confidence limits based on the normal approximations of r, r^* and r_B^* are established via pivot profiling; see [2].

3.1 Numerical assesment

In this section we provide a simulation study designed to assess the performance of nominal 95% confidence intervals based upon the small-sample pivots r^* and r_B^*, and to compare these with those obtained from the large-sample counterpart r. The accuracy of the confidence intervals is evaluated in terms of empirical coverage (CP), and upper (UE) and lower (LE) error probability.

Simulation 1 We explore the finite-sample performance of the confidence intervals for Gini's gamma index in the equi-correlated bivariate normal model, while emphasizing on small sample sizes. We specialize to the case $\mu = 0.7$, $\sigma = 0.1$, whith $\rho \in \{-0.9, -0.8, \ldots, 0.8, 0.9\}$.

Simulation 2 We investigate the finite-sample performance of confidence intervals for Gini's gamma index for the complete bivariate normal model, again while emphasizing small sample sizes. We use $\mu_1 = 1$, $\mu_2 = 2$, $\sigma_1 = 0.9$, $\sigma_2 = 0.1$, while $\rho \in \{-0.9, -0.8, \ldots, 0.8, 0.9\}$.

Figures 1 and 2 plot the empirical coverage of the 95% confidence intervals for Simulations 1 and 2, respectively.

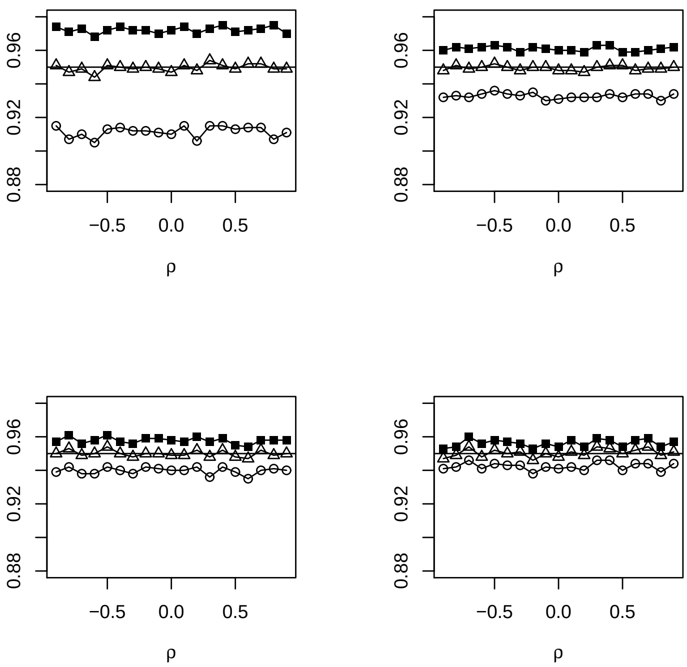

Fig. 1 Simulation 1. Empirical coverage of 95% confidence intervals for γ for varying values of ρ and sample sizes n. From top left to bottom right: $n = 5$, $n = 10$, $n = 15$, $n = 20$. Legend: –○– 1st order; –△– 3rd order; –■– Bayes; — nominal.

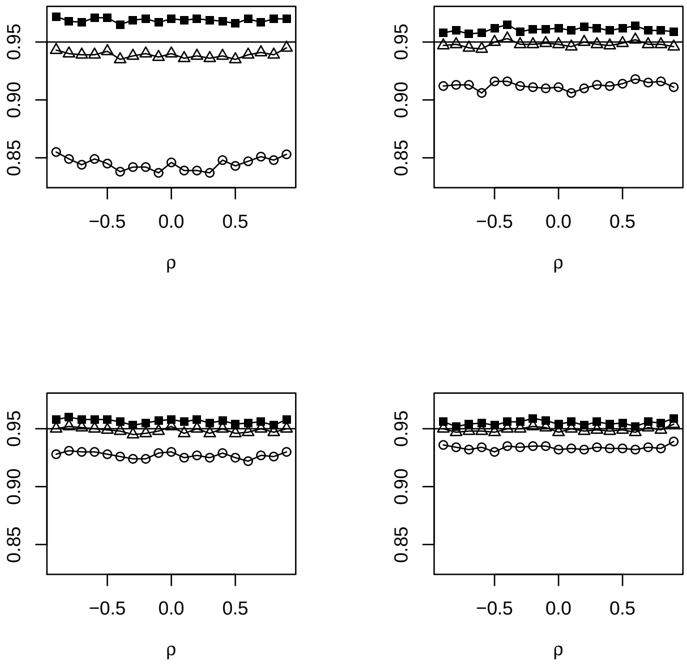

Fig. 2 Simulation 2. Empirical coverage of 95% confidence intervals for γ for varying values of ρ and sample sizes n. From top left to bottom right: $n = 5$, $n = 10$, $n = 15$, $n = 20$. Legend: –○– 1st order; –△– 3rd order; –■– Bayes; — nominal.

4 Discussion and conclusions

Figures 1 and 2 both reveal that the small-sample pivots r^* and r^*_B exhibit more reliable coverage than the confidence intervals obtained from their large-sample counterpart r, although the Bayesian solution r^*_B somewhat overestimates the nominal level. The higher order solutions guarantee symmetry on the tails. The differences among the pivots vanish as the sample size increases. These findings are in agreement with the results of a simulation study for the case of no-nuisance parameters which, because of space constraints, is not reported here.
As previously stated, we can use the proposed method for the estimation of any other measure of concordance for the bivariate Gaussian normal copula. Although in our simulation study we specialize to the case that both margins are standard normal, due to the invariance with respect to the two marginal distributions, the current method does not depend on the specification of the marginals. Finally, note that

our proposal can also be generalized to handle any type of copula for which the measures of concordance can be found in closed form.

References

1. Barndorff-Nielsen, O. (1983). On a formula for the distribution of the maximum likelihood estimator. *Biometrika*, **70**, 343–365.
2. Brazzale, A. R., Davison, A. C. and Reid, N. (2007). *Applied Asymtpotics: Case Studies in Small-Sample Statistics.* Cambridge University Press.
3. Brazzale, A. R. and Mameli, V. (2014). Small-sample likelihood asymptotics for the equi-correlated bivariate normal model. *Proceedings of the XLVII Meeting of the Italian Statistical Society*, 1–6. (ISBN:978-88-8467-874-4)
4. Fraser, D. A. S., Reid, N. and Wu, J. (1999). A simple general formula for tail probabilities for frequenstist and Bayesian inference. *Biometrika*, **86**, 249–264.
5. Gini, C. (1914). *L'ammontare e la composizione della ricchezza delle nazioni.* F. Bocca, Torino .
6. 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.
7. Meyer, C. (2013). The Bivariate Normal Copula. *Communications in Statistics - Theory and Methods*, **42**, 2402–2422.
8. Reid, N. (2003). Asymptotics and the theory of inference. *Annals of Statistics*, **31**, 1695–1731.
9. Reid, N. and Fraser, D. A. S. (2010). Mean log-likelihood and higher-order approximations. *Biometrika*, **97**, 159–170.
10. Schmid, F., Schmidt, R., Blumentritt, T., Gaißer, S. and Ruppert, M. (2010). Copula-based measure of multivariate association. *Copula Theory and Its Applications*. In P.Jaworski et al. (eds), Lecture Notes in Statistics, **198**, Springer Berlin Heidelberg, 209–236.
11. Sklar, A. (1959). Fonctions de repartition a n dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Universite de Paris*, **8**, 229–231.
12. Staicu, A. M. and Reid, N. (2008). On probability matching priors. *The Canadian Journal of Statistics*, **36**, 613–622.
13. Sun Y., Wong A. C. M. (2007). Interval estimation for the normal correlation coefficient. *Statistics & Probability Letters*, **77**, 1652–1661.