

# Approximate Bayesian Methods for Multivariate and Conditional Copulae

Clara Grazian and Brunero Liseo

**Abstract** We describe a simple method for making inference on a functional of a multivariate distribution. The method is based on a copula representation of the multivariate distribution, where copula is a flexible probabilistic tool that allows the researcher to model the joint distribution of a random vector in two separate steps: the marginal distributions and a copula function which captures the dependence structure among the vector components. The method is also based on the properties of an approximate Bayesian Monte Carlo algorithm, where the proposed values of the functional of interest are weighted in terms of their empirical likelihood. This method is particularly useful when the likelihood function associated with the working model is too costly to evaluate or when the working model is only partially specified.

## 1 Introduction

Theoretical proposals are now available to model complex situations, thanks to the recent advances in computational methodologies and to the increased power of modern computers. In particular, there are new methods for multivariate analysis, however the goal of modelling complex multivariate structures and estimating them has not yet been reached in a completely satisfactory way.

Copula models have been introduced as probabilistic tools to describe a multivariate random vector via the marginal distributions and a copula function which captures the dependence structure among the vector components, thanks to the Sklar's theorem [1], which states that any  $d$ -dimensional absolutely continuous density can be uniquely represented as

$$f(x_1, \dots, x_d) = f_1(x_1) \dots f_d(x_d) c_{12\dots d}(F_1(x_1), \dots, F_d(x_d)). \quad (1)$$

---

C. Grazian · B. Liseo (✉)  
MEMOTEF, Sapienza Università di Roma, Roma, Italy  
e-mail: brunero.liseo@uniroma1.it

C. Grazian  
e-mail: clara.grazian@uniroma1.it

While it is often straightforward to produce reliable estimates of the marginals, making inference on the dependence structure is more complex. Major areas of application include econometrics, hydrological engineering, biomedical science, signal processing and finance.

In a parametric frequentist approach to copula models, there are no broadly satisfactory methods for the joint estimation of marginal and copula parameters. The most popular method is the so called Inference From the Margins (IFM), where the parameters of the marginal distributions are estimated first, and then pseudo-observations are obtained by pluggin-in the estimates of the marginal parameters. Then inference on the copula parameters is performed using the pseudo-observations: this approach obviously does not account for the uncertainty on the estimation of the marginal parameters. Bayesian alternatives are not yet fully developed, although there are remarkable exceptions ([2, 3], among others).

In this work we consider the general problem of estimating some specific quantities of interest of a generic copula (such as, for example, tail dependence index or Spearman's  $\rho$ ) by adopting an approximate Bayesian approach along the lines of [4]. In particular, we discuss the use of the an approximate Bayesian computation algorithm based on empirical likelihood weights (in the following,  $BC_{EL}$ ), based on the empirical likelihood approximation of the marginal likelihood of the quantity of interest (see [5] for a complete and recent survey on empirical likelihood methods). This approach produces an approximation of the posterior distribution of the quantities of interest, based on an approximation of the likelihood function and on a Monte Carlo approximation of the posterior distribution via simulations. Our approach is general, in the sense that it could be adapted both to parametric and nonparametric modelling of the marginal distributions. Also, the use of empirical likelihood avoids the need of choosing a specific parametric copula model.

## 2 Approximate Bayesian Computation

The idea underlying likelihood-free methods (or approximate Bayesian computation, ABC) is to propose a candidate  $\theta'$  and to generate a data set from the working model with parameter set to  $\theta'$ . If the observed and the simulated data are similar "in some way", then the proposed value is considered a good candidate to have generated the data and becomes part of the sample which will form the approximation to the posterior distribution. Conversely, if the observed and the simulated data are too different, the proposed  $\theta'$  is discarded.

In order to implement a basic version of the ABC algorithm one needs to specify a set of summary statistics to make comparisons, a distance to quantify comparisons and a tolerance level  $\epsilon > 0$ .

The basic ABC may be inefficient, because it proposes values of  $\theta$  from its prior distribution, therefore, ABC algorithms are often linked with other methods, for instance, with Markov Chain Monte Carlo (MCMC) or Sequential Monte Carlo (SMC) methods. In this work, we will focus on a different ABC approach, described

in Algorithm 1 where  $M$  simulations from the prior are generated; this method avoids the most expensive step in computational time, that is the generation of new data sets. A detailed description of the method is in [4]; it represents a re-sampling scheme where the proposed values are re-sampled with weights proportional to their empirical likelihood values.

---

```

for  $i = 1$  to  $M$  do
  repeat
    Generate  $\theta_i$  from the prior distribution  $\pi(\theta)$ 
    Set the weight for  $\theta_i$  as  $\omega_i = L_{EL}(\theta_i; \mathbf{x})$ .
  end for
  for  $i = 1$  to  $M$  do
    Draw, with replacement, a value  $\theta_i$  from the previous set of  $M$  values using weights
     $\omega_i, i = 1, \dots, M$ .
  end for

```

---

**Algorithm 1:**  $BC_{EL}$  algorithm

### 3 The Proposed Approach

We propose to adapt the  $BC_{EL}$  algorithm to a situation where the statistical model is only partially specified and the main goal is the estimation of a finite dimensional quantity of interest, i.e. a situation where the complete structure of the mutual dependence is considered a nuisance parameter and it is kept as general as possible. While the main interest of [4] was the approximation of the full posterior distribution of the parameters of the model, here we use the empirical likelihood (EL) approach to avoid a parametric definition of the model for the observed data and focus only on a particular functional of the distribution, which summarizes the correlation among the variables.

We assume that a data set is available in the form of a size  $n \times d$  matrix  $\mathbf{X}$ , where  $n$  is the sample size and  $d$  is the number of variables, that is

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1d} \\ x_{21} & x_{22} & \dots & x_{2d} \\ \dots & \dots & x_{ij} & \dots \\ x_{n1} & x_{n2} & \dots & x_{nd} \end{pmatrix}.$$

In the following,  $X_{[ \cdot, j ]}$  will denote the  $j$ -th column (variable) and  $X_{[ i, \cdot ]}$  the  $i$ -th row of  $\mathbf{X}$ , respectively. For each  $j = 1, \dots, d$ , we consider the available data information in  $X_{[ \cdot, j ]}$  to produce an estimate of the marginal CDF of  $X_{[ \cdot, j ]}$ . Let  $\lambda_j = (\lambda_j^{(1)}, \lambda_j^{(2)}, \dots, \lambda_j^{(B)})$ ,  $j = 1, 2, \dots, d$ , be the posterior sample of size  $B$  obtained

for the distribution of  $X_{[.,j]}$ . Notice that the vector  $\lambda_j$  can be either a sample from the posterior distribution of the parameters of the model we have adopted for  $X_{[.,j]}$  or a posterior sample of  $CDF$ 's in a nonparametric set-up.

Then we use a copula representation for estimating the multivariate dependence structure of the random vector  $\mathbf{X}$ ,

$$H(x_1, \dots, x_d) = C_\theta(F_1(x_1), F_2(x_2), \dots, F_d(x_d)) \tag{2}$$

where  $\theta$  is the parameter related to the copula function.

Estimating the copula  $C_\theta(\cdot)$  can be managed either using some parametric model for the copula (such as Clayton, Gaussian, Skew-t, Gumbel, etc.) or using a nonparametric approach. In this paper, we take a nonparametric route (in many situations it is difficult to prefer a model instead of another) and we concentrate on some specific function of  $C_\theta(\cdot)$ , say  $\varphi = T(\mathbf{F})$ , for example the Spearman's measure of association  $\rho$  between two components of  $\mathbf{X}$ , say  $X_h$  and  $X_j$ , which is defined as the correlation coefficient among the transformed values  $U_i = F_i(x_i)$ ,  $i = j, h$ :

$$\rho = 12 \int_0^1 \int_0^1 (C(u_j, u_h) - u_h u_j) du_j du_h. \tag{3}$$

Its sampling counterpart  $\rho_n$  is the correlation among ranks  $\mathbf{R}$  and  $\mathbf{S}$  of the data observed for the two variables of interest and it can be written as:

$$\rho_n = \frac{1}{n} \sum_{i=1}^n \left( \frac{12}{n^2 - 1} R_i S_i \right) - 3 \frac{n + 1}{n - 1}. \tag{4}$$

If interest lies only in a functional of the copula, instead of in its entire structure, we use Algorithm 2 to produce an approximation of the posterior distribution of the functional itself  $\varphi = T(\mathbf{F})$ .

It is important to note that the approximation might hold only asymptotically: for example, if the sample version of the Spearman's  $\rho$  is used to approximate the posterior distribution of  $\rho$ , one has to consider that the sample version is only asymptotically unbiased. One advantage of the proposed Algorithm is that prior information is only provided for the marginal distributions and for  $\varphi$ ; so the prior elicitation is easier: it is not necessary to define a prior distribution for the entire copula function.

Moreover, the method is robust with respect to different prior opinions about non-essential aspects of the dependence structure and with respect to the copula definition. The most important disadvantage of the method is its inefficiency when compared to a parametric copula, as usual in nonparametric or semiparametric setting; however this is true only under the assumption that the parametric copula is the true model.

From a computational perspective Algorithm 2 is quite demanding, since one needs to run a  $BC_{EL}$  algorithm for each row of the posterior sample from the marginals. Even though the estimation of the marginal densities of the  $X_{[.,j]}$ 's might not

[1:] For  $b = 1, \dots, B$ , use the  $s$ -th row of the posterior simulation  $\lambda_1^{(b)}, \lambda_2^{(b)}, \dots, \lambda_d^{(b)}$  to create a matrix of uniformly distributed *pseudo*-observations

$$\mathbf{u}^{(b)} = \begin{pmatrix} u_{11}^{(b)} & u_{12}^{(b)} & \dots & u_{1d}^{(b)} \\ u_{21}^{(b)} & u_{22}^{(b)} & \dots & u_{2d}^{(b)} \\ \dots & \dots & u_{ij}^{(b)} & \dots \\ u_{n1}^{(b)} & u_{n2}^{(b)} & \dots & u_{nd}^{(b)} \end{pmatrix}$$

with  $u_{ij}^{(b)} = F_j(x_{ij}; \lambda_j^{(b)})$ .

[2:] Given a prior distribution  $\pi(\varphi)$  for the quantity of interest  $\varphi$ ,

**for**  $m = 1, \dots, M$ ,

1. draw  $\varphi^{(m)} \sim \pi(\varphi)$ ;
2. compute  $EL(\varphi^{(m)}; \mathbf{u}^{(b)}) = \omega_{mb}; b = 1, \dots, B$ .
3. take the average weight  $\omega_m = B^{-1} \sum_{b=1}^B \omega_{mb}$

**end for**

[3:] re-sample - with replacement - from  $\{(\varphi^{(b)}, \omega_b), b = 1, \dots, B\}$ .

**Algorithm 2:** ABCOP algorithm

require a huge number of iterations  $B$ , still it might be very expensive to run  $B$  different  $BC_{EL}$  algorithms. To avoid this computational burden, we propose to modify the above algorithm by simply performing a single run of the  $BC_{EL}$  algorithm, where, for each iteration  $m = 1, \dots, M$ , a randomly selected (among the  $B$  rows) row  $\lambda^b$  is used to transform the actual data into pseudo-observations lying in  $[0, 1]^d$ .

## 4 An Example: Spearman's $\rho$

The definition of the Spearman's  $\rho$  given in (4) can be interpreted as an average distance between the copula  $C$  and the independence copula  $\Pi(u, v) = uv$ . Thus, in a  $d$ -dimensional setting the multivariate  $\rho$  becomes

$$\begin{aligned} \rho &= \frac{\int_{[0,1]^d} C(\mathbf{u})d\mathbf{u} - \int_{[0,1]^d} \Pi(\mathbf{u})d\mathbf{u}}{\int_{[0,1]^d} M(\mathbf{u})d\mathbf{u} - \int_{[0,1]^d} \Pi(\mathbf{u})d\mathbf{u}} \\ &= \frac{d+1}{2^d - (d+1)} \left\{ 2^d \int_{[0,1]^d} C(\mathbf{u})d\mathbf{u} - 1 \right\}. \end{aligned} \tag{5}$$

The multivariate extension of the empirical copula is

$$\hat{C}_n(\mathbf{u}) = \frac{1}{n} \sum_{i=1}^n \prod_{j=1}^d \mathbb{I}_{\{\hat{U}_{ijn} \leq u_i\}} \quad \text{for } \mathbf{u} = (u_1, u_2, \dots, u_n) \in [0, 1]^d \tag{6}$$

where  $\hat{U}_{ijn} = \hat{F}(X_{ij})$  for  $i = 1, \dots, d$  and  $\hat{F}(\cdot)$  is the empirical marginal distribution function. Therefore, a nonparametric estimator of  $\rho$  is

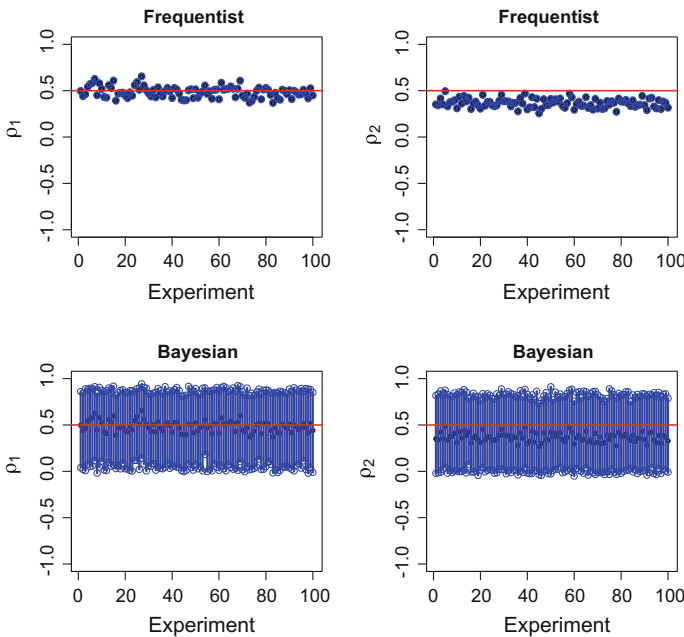
$$\hat{\rho}_{1n} = h(d) \left\{ 2^d \int_{[0,1]^d} \hat{C}_n(\mathbf{u}) d\mathbf{u} - 1 \right\} = h(d) \left\{ \frac{2^d}{n} \sum_{i=1}^n \prod_{j=1}^d (1 - \hat{U}_{ijn}) - 1 \right\} \tag{7}$$

where  $h(d) = (d + 1)/(2^d - d - 1)$ . An alternative estimator is

$$\hat{\rho}_{2n} = h(d) \left\{ 2^d \int_{[0,1]^d} \Pi(\mathbf{u}) d\hat{C}(\mathbf{u}) - 1 \right\} = h(d) \left\{ \frac{2^d}{n} \sum_{i=1}^n \prod_{j=1}^d \hat{U}_{ijn} - 1 \right\} \tag{8}$$

Asymptotic properties of these estimators are assessed in [6].

Once an estimator of the multivariate version of  $\rho$  is available, it is possible to apply the procedure presented in Sect. 3. On the other hand, the variance of the proposed estimators can be explicitly computed only in few cases, for example in the case of the independence copula. [6] proposes to estimate it in a nonparametric way via a bootstrap method. Nevertheless, in practice this method tends to underestimate the variance, as it is shown in Fig. 1, where the frequentist procedure for a fixed  $n$



**Fig. 1** 100 simulations from a Clayton copula: sample size is 100; the true value of  $\rho$  is equal to 0.5 (red line). The results for the frequentist procedure are available above, the ones for the Bayesian procedure are available below. The black lines are the point estimates of  $\rho_1$ , the blue lines represent the lower and the upper bounds of the intervals of level 0.95

leads to a coverage of about 10 % (coverage of 0 % for the interval of  $\hat{\rho}_2$ ), while the proposed Bayesian method has the expected coverage, even if the average length is necessarily greater, about 0.822, i.e. the intervals contain almost half of the parameter space.

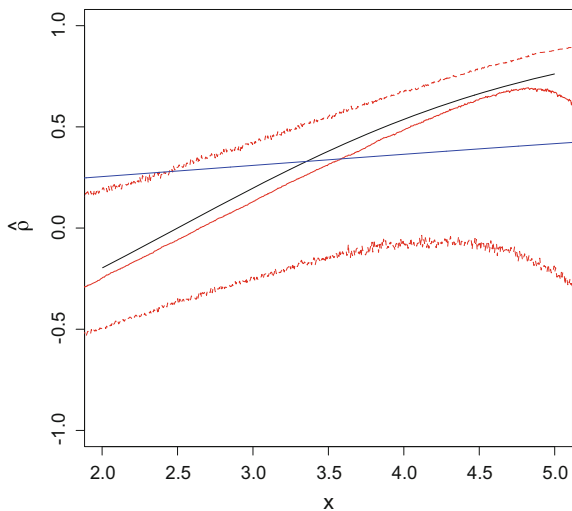
### 5 Further Research

Algorithm 2 produces an approximation of the posterior distribution of any summary of the multivariate dependence, once a multivariate estimator is available, as in the case of the Spearman’s  $\rho$ . In some cases the analysis may be focused on measures of dependence as functions of some available conditioning variables. In the case of two response variables  $X_1$  and  $X_2$ , both depending on the same covariate  $Z$ , the observations  $(x_{1i}, x_{2i}, z_i)$  follow a distribution  $F_{X_1, X_2 | Z}(\cdot | z)$ . [7] proposes the following estimator for the Spearman’s  $\rho$ .

$$\hat{\rho}_n(x) = 12 \sum_{i=1}^n w_{ni}(x, h_n)(1 - \hat{U}_{i1})(1 - \hat{U}_{i2}) \tag{9}$$

where  $\hat{U}_{i,j} = \sum_{i'=1}^n w_{i'}(x, h_n)\mathbb{I}(U_{i'j} \leq u_{ij})$  for  $j = 1, 2$ ,  $U_{ij} = F_j(x_{ij})$  and  $w_{ij}(x, h_n)$  are appropriately chosen weights depending on  $x_{ij}$  and a bandwidth  $h_n$ , for example kernel-based weights as the Nadaraya-Watson. Unfortunately, estimator (9) is based on an estimator of the conditional copula, given in [7], which is biased. A first simulation study implemented for 10,000 simulations of the function  $\rho(x)$  (see Fig. 2) shows that, while the estimator (9) is not able to capture the true function

**Fig. 2** Simulations from the conditional Clayton copula based on 10,000 ABC simulations of  $\rho(x)$  and 100 data points: true function  $\rho(x)$  in black, Bayesian estimates in red (median, 0.05 and 0.95 credible bands), frequentist estimate in blue



(it underestimates the dependence among values), the Bayesian estimate obtained via Algorithm 2 can recover it, even if the variance increases as the value of the covariate increases. Further research will be focused on trying to understand why this happens and on producing more stable estimates.

## References

1. Sklar M (1959) Fonctions de répartition à  $n$  dimensions et leurs marges. *Publ Inst Statist Univ Paris* 8:229–231
2. Craiu VR, Sabeti A (2012) In mixed company: Bayesian inference for bivariate conditional copula models with discrete and continuous outcomes. *J Multivar Anal* 110:106–120
3. Mengersen K, Pudlo P, Robert CP (2013) Bayesian computation via empirical likelihood. *Proc Natl Acad Sci* 110(4):1321–1326
4. Min A, Czado C (2010) Bayesian inference for multivariate copulas using pair-copula constructions. *J Financ Econometrics* 8(4):511–546
5. Owen AB (2010) *Empirical likelihood*. Chapman & Hall/CRC Press, New York
6. Schmid F, Schmidt R (2007) Multivariate extensions of Spearman's Rho and related statistics. *Stat Probab Lett* 77(4):407–416
7. Gijbels I, Veraverbeke N, Omelka M (2011) Conditional copulas, association measures and their applications. *Comput Stat Data Anal* 55(5):1919–1932