

## APPROXIMATE BAYESIAN COMPUTATION FOR COPULA ESTIMATION

Clara Grazian

*Université Dauphine, Paris, France*

*Dipartimento di Scienze Statistiche, Sapienza Università di Roma, Roma, Italia*

*CREST, Paris, France*

Brunero Liseo<sup>1</sup>

*Dipartimento di metodi e Modelli per l'Economia, il territorio e la Finanza, Sapienza Università di Roma, Roma, Italia*

### 1. INTRODUCTION

Copula models are nowadays widely used in multivariate data analysis. Major areas of application include econometrics (Huynh *et al.*, 2015), geophysics (Scholzel and Friederichs, 2008), quantum mechanics (Resconi and Licata, 2015) climate prediction (Scheffzik *et al.*, 2013) genetics (He *et al.*, 2012), actuarial science and finance (Cherubini *et al.* (2004), among the others). A 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.

From a statistical perspective, whereas it is generally simple to produce reliable estimates of the parameters of the marginal distributions of the data, the problem of estimating the dependence structure, however it is modelled, is crucial and often complex, especially in high dimensional situations. On the other hand, dependence is one of the most fundamental features in (applied) statistics, economics and probability. A huge list of important applications can be found in the recent monograph by Joe (2014).

In a 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 Margin (IFM) method, where the parameters of the marginal distributions are estimated first, and then pseudo data are obtained by plugging in the estimates of the marginal parameters. Then inference on the copula parameters is performed using the pseudo-data: this approach obviously does not account for the uncertainty on the estimation of the marginal parameters. Bayesian alternative are not yet fully developed, although

---

<sup>1</sup> Corresponding Author. E-mail: brunero.liseo@uniroma1.it

Min and Czado (2010), Craiu and Sabeti (2012), Smith (2013) and Wu *et al.* (2014) are remarkable exceptions.

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 Mengersen *et al.* (2013). In particular, we discuss the use of the  $BC_{el}$  algorithm, based on the empirical likelihood approximation of the marginal likelihood of the quantity of interest. Our approach is approximate in two aspects:

- i. elicitation of the prior distribution is required only on the quantity of interest. Its prior distribution is combined with the empirical likelihood in order to produce an approximation to the “true” posterior distribution.
- ii. we do not use the “true” likelihood function, but rather an approximation based on empirical likelihood theory (Owen, 2010). Hopefully, this will reduce the potential bias for incorrect distributional assumptions.

Note, however, that the word “true” in the above list should be better spelled as “true-under-the-assumed-model”. In situations where a true model is too hard to specify, or too complex to deal with, the empirical likelihood can be an extremely valuable tool.

Our approach can be adapted both to parametric and nonparametric modelling of the marginal distributions. The method described in this paper is in the spirit of Hoff (2007), but it is based on a different kind of approximation; the results, although from a different perspective, can be also interpreted in the light of Schennach (2005), where a Bayesian nonparametric interpretation of a tilted version of the empirical likelihood is provided.

## 2. PRELIMINARIES: COPULAE AND EMPIRICAL LIKELIHOOD

A copula model is a way of representing the joint distribution of a random vector  $\mathbf{X} = (X_1, \dots, X_m)$ . Given an  $m$ -variate cumulative distribution function (CDF)  $\mathbf{F}$ , it is possible to show (Sklar, 1959) that there always exists an  $m$ -variate function  $C : [0, 1]^m \rightarrow [0, 1]$ , such that  $\mathbf{F}(x_1, \dots, x_m) = C(F_1(x_1), \dots, F_m(x_m))$ , where  $F_j$  is the marginal CDF of  $X_j$ . In other terms, the copula function  $C$  is a CDF with uniform margins on  $[0, 1]$ : it binds together the univariate CDF's  $F_1, F_2, \dots, F_m$  in order to produce the  $m$ -variate CDF  $\mathbf{F}$ . The copula function  $C$  does not depend on the marginal distributions of  $\mathbf{F}$ , but rather it accounts for potential dependence among the components of the random vector  $\mathbf{X}$ .

For each pair of components of  $\mathbf{X}$ , say  $X_i$  and  $X_j$ , let us assume that they have continuous CDF's  $F_i$  and  $F_j$ . It is well known that both the transformed variables  $U_i = F_i(X_i)$  and  $U_j = F_j(X_j)$  have uniform marginal distributions. A semiparametric copula model consists of a parametric model for the joint distribution of  $(U_i, U_j)$  and no assumptions on the marginal CDF's. A nonparametric copula is assumed when the joint distribution of  $(U_i, U_j)$  depends on an infinite dimensional parameter. In this paper we will allow the marginal distributions  $F_j$ 's to follow either a parametric or a non parametric model. For the copula function we will not make any parametric assumption. Rather, we will limit our goal to

the estimation of a particular function of interest of the copula  $C$ . A discussion on the classical approaches to semiparametric estimation of copula models can be found in Genest *et al.* (1995).

Empirical likelihood has been introduced by Owen: Owen (2010) is a complete and recent survey; it is a way of producing a nonparametric likelihood for a quantity of interest in an otherwise unspecified statistical model. It is particularly useful when a true likelihood is not readily available either because it is too expensive to evaluate or when the model is not completely specified. Assume that our dataset is composed of  $n$  independent replicates  $(x_1, \dots, x_m)$  of some random vector  $\mathbf{X}$  with distribution  $\mathbf{F}$  and corresponding density  $f$ . Rather than defining the usual likelihood function in terms of  $f$ , the empirical likelihood is constructed with respect to a given quantity of interest, say  $\varphi$ , expressed as a functional of  $\mathbf{F}$ , i.e.  $\varphi(\mathbf{F})$ , and then a sort of profile likelihood of  $\varphi$  is computed in a nonparametric way. More precisely, consider a given set of generalized moment conditions of the form

$$E_{\mathbf{F}}(h(X, \varphi)) = 0, \quad (1)$$

where  $h(\cdot)$  is a known function, and  $\varphi$  is the quantity of interest. The resulting empirical likelihood is defined as

$$L_{EL}(\varphi; \mathbf{x}) = \max_{\mathbf{p}} \prod_{i=1}^n p_i,$$

where the maximum is searched over the set of vectors  $\mathbf{p}$  such that  $0 \leq p_i \leq 1$ ,  $\sum_{i=1}^n p_i = 1$ , and

$$\sum_{i=1}^n h(\mathbf{x}_i, \varphi) p_i = 0.$$

Whereas the first two conditions are obvious and independent of  $\varphi$ , the third one induces a profiling of the information towards the quantity of interest, through a sort of unbiasedness condition.

### 3. ABC AND EL

Approximate Bayesian computation has now become an essential tool for the analysis of complex stochastic models, in the case where the likelihood function is unavailable in closed form or it is too expensive to be repeatedly evaluated (Marin *et al.*, 2012). It can be considered as a class of popular algorithms that achieves posterior simulation by avoiding the computation of the likelihood function. A crucial condition for the use of ABC algorithms is that it must be relatively easy to generate new pseudo-observations from the working model, for a fixed value of the parameter vector. In its simplest form, the ABC algorithm “proposes” a (pseudo)-randomly drawn parameter value  $\theta^*$  from the prior distribution and a new data set is generated, conditionally on  $\theta^*$ ; then the value is accepted only if the new data are “similar enough” to the actual observed data. It can be proved that the set of accepted values represents a sample from an approximation of the posterior distribution of  $\theta$  (Sisson and Fan, 2011). However, it is often highly

inefficient to propose values from the prior distribution, since it is generally much more diffuse than the posterior distribution. Many more sophisticated computational strategies are available in order to avoid generating values from the prior distribution, see Marin *et al.* (2012) for example; here we will not discuss these issues and we rather concentrate on a different ABC approach, which can avoid the most expensive step in computational time, that is the proposal of new data sets. This method has been proposed by Mengersen *et al.* (2013) and it represents a re-sampling scheme where the proposed values are re-sampled with weights proportional to their empirical likelihood. In practice, the algorithm belongs to the family of “sampling importance re-sampling” - SIR, (Rubin, 1988) - methods for models in which the “true likelihood” evaluation is out of reach and the “true” weights are approximated by their empirical likelihood.

---

```

for  $i = 1$  to  $M$  do
  repeat
    Generate  $\theta_i$  from the prior distribution  $\pi(\theta)$ 
    Set the weight for  $\theta_i$  as  $\omega_i = L_{EL}(\theta_i; \text{data})$ .
  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 (Mengersen *et al.*, 2013)

#### 4. THE PROPOSED APPROACH

In this paper we propose to adapt the  $BC_{EL}$  algorithm of Mengersen *et al.* (2013) 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. In practice this represents the prototypical semiparametric set-up, where one is mainly interested in some meaningful characteristic of the population, although the statistical model may contain nuisance parameters which are often introduced in order to produce more flexible models that might better fit the data at hand. In order to make robust inference on the quantity of interest, a reasonable model should account for the uncertainty on the nuisance parameters, in some way. Even if some of these additional parameters are not particularly important in terms of estimation - they often lack of a precise physical meaning - their estimates can dramatically affect inferences on the parameter of interest. In these circumstances it might be more reasonable and *robust* to partially specify the model and adopt a semiparametric approach.

Oh and Patton (2013) consider, in a frequentist perspective, a Simulated Method of Moments estimation for copula models. Their paper is very close

in spirit to what we are proposing, although their main goal is the analysis of partially specified models rather than models with an intractable likelihood.

#### 4.1. The algorithm in full detail

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

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

In the following,  $X_{[:,j]}$  will denote the  $j$ -th column (variable) and  $X_{[i,:]}$  the  $i$ -th row of  $\mathbf{X}$ , respectively. For each  $j = 1, \dots, m$ , we consider the available data information in  $X_{[:,j]}$  to produce an estimate of the marginal CDF of  $X_{[:,j]}$ . Let  $\boldsymbol{\lambda}_j = (\lambda_j^{(1)}, \lambda_j^{(2)}, \dots, \lambda_j^{(S)})'$ ,  $j = 1, 2, \dots, m$  be the posterior sample obtained from *some* Bayesian inference method for the distribution of  $X_{[:,j]}$ . Notice that the vector  $\boldsymbol{\lambda}_j$  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  $\mathbf{X}$ ,

$$\mathbf{F}(x_1, \dots, x_m) = C_\theta(F_1(x_1), F_2(x_2), \dots, F_m(x_m)),$$

where  $\theta$  is the parameter related to the copula function. Since we are assuming that one has already *estimated* the marginal  $F_j(x_j)$ 's,  $j = 1, \dots, m$ , one now needs to consider the copula  $C_\theta(\cdot)$  only. This step can be managed either using some parametric model for the copula (such as Clayton, Gaussian, Skew-t, Gumbel, etc.) or using a nonparametric approach.

Parametric copulae in Bayesian inference have been already investigated in several papers. Here we should mention Hoff (2007), Silva and Lopes (2008), Min and Czado (2010), Smith *et al.* (2012) and Craiu and Sabeti (2012). In this paper, we take a nonparametric route and we concentrate on some specific function of  $C_\theta(\cdot)$ , say  $\varphi = T(C_\theta)$ . This is particularly useful and meaningful in those situations where there is no theoretical or empirical evidence that a given copula should be preferred and we are mainly interested in some specific synthetic measure of the multivariate dependence, like for example, the upper tail dependence index between two components of  $\mathbf{X}$ , that is

$$\chi = \lim_{u \rightarrow 1} P(U_j > u | U_h > u) \approx \lim_{u \rightarrow 1} \left[ 2 - \frac{\log P(U_j < u, U_h < u)}{\log P(U_h < u)} \right]$$

where  $U_i = F_i(x_i)$ ,  $i = j, h$ . Another popular quantity, which we will consider in the final section is 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  $U_i = F_i(x_i)$ ,  $i = j, h$  or, in a copula language, as

$$\begin{aligned}\rho &= 12 \int_0^1 \int_0^1 (C(u_j, u_h) - u_h u_j) du_j du_h \\ &= 12 \int_0^1 \int_0^1 C(u_j, u_h) du_j du_h - 3.\end{aligned}\quad (2)$$

We now describe the algorithm in a pseudo-language:

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

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

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

[2:] Given a prior distribution  $\pi(\varphi)$  for the quantity of interest  $\varphi$ ,  
for  $b = 1, \dots, B$ ,

1. draw  $\varphi^{(b)} \sim \pi(\varphi)$ ;
2. compute  $EL(\varphi^{(b)}; \mathbf{u}^{(s)}) = \omega_{bs}; s = 1, \dots, S$ .
3. take the average weight  $\omega_b = S^{-1} \sum_{s=1}^S \omega_{bs}$

**end for**

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

### Algorithm 2: ABCOP algorithm

The final output of the above algorithm is then a posterior sample drawn from an approximation of the posterior distribution of the quantity of interest  $\varphi$ . There are several critical issues both in the practical implementation of the method and in its theoretical properties. First, the empirical likelihood is based on moment conditions of the form (1). In practical applications these conditions might hold only asymptotically. This is the case, for example, of the Spearman's  $\rho$ , which we discuss in the next session. Its sample counterpart  $\rho_n$  is only an asymptotically unbiased estimator of  $\rho$  so the moment condition is strictly valid only for large samples. Also, prior information is only provided for the marginal distributions and for  $\varphi$ : this, of course, has advantages and, on the other hand, poses theoretical issues. The main advantage is the ease of elicitation: one need not to elicit unnecessary aspects of the prior distribution. This is mainly in the spirit of the partially specified models, quite popular in the econometric literature. Another obvious advantage of the proposed approach is the implied robustness of

the method, with respect to different prior opinions about non-essential aspects of the dependence structure. The most important disadvantage of the method is its inefficiency when compared to a parametric copula, under the assumption that the parametric copula is the true model. The practical implementation of the algorithm is quite simple in  $\mathbf{R}$ ; it uses some functions contained in the suite `gmm`: see for example Chaussé (2010).

From a computational perspective the above algorithm 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  $X_{[\cdot,j]}$ 's might not require huge values of iterations  $S$ , still it might be very expensive to run  $S$  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  $b = 1, \dots, B$ , a randomly selected (among the  $S$  rows) row  $\lambda^s$  is used to transform the actual data into pseudo-data lying in  $[0, 1]^m$ . With this modification the above algorithm gets transformed into Algorithm 3.

## 5. A SIMPLE ILLUSTRATION: SPEARMAN'S $\rho$

We first illustrate the method in a simple situation, with  $m = 2$ , and assuming that the two marginal distributions of the data are known: without loss of generality we can then assume that they are both uniform in  $[0, 1]$ ; in this case there are no practical differences between Algorithm 2 and Algorithm 3.

The Spearman's  $\rho$  measure of dependence has been defined in (2). Starting from a sample of size  $n$  from a bivariate distribution, say  $(x_i, y_i)$ ,  $i = 1, \dots, n$ , the sampling counterpart of  $\rho$ , say  $\rho_n$ , is nothing but the correlation among ranks and it can be written as

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

where

$$R_i = \text{rank}(x_i) = \sum_{k=1}^n I(x_k \leq x_i), \quad S_i = \text{rank}(y_i) = \sum_{k=1}^n I(y_k \leq y_i), \quad i = 1, \dots, n.$$

Since we assume that the marginal distributions are known, pseudo-data coincide with the actual data, and we work with a single  $n \times 2$  matrix  $\mathbf{U}$  whose generic element is given by  $u_{ij} = x_{ij}$  with  $u_{ij} \in [0, 1]$ ,  $i = 1, \dots, n$ ,  $j = 1, 2$ . Then we take the ranks  $(R_i, S_i)$  of the original values and compute  $\rho_n$ . Also we are able to evaluate the empirical likelihood of  $\rho$  for a given value of  $\rho_n$  as  $\max_{p_i} EL(\rho; \rho_n) = \prod_{i=1}^n np_i(\rho)$  under the constraints  $\sum_{i=1}^n p_i = 1$ ,  $0 \leq p_i \leq 1$ ,  $i = 1, \dots, n$  and

$$\sum_{i=1}^n p_i \left( \frac{12R_i S_i}{n^2 - 1} - 3 \frac{n+1}{n-1} - \rho \right) = 0.$$

[1:] For  $j = 1, \dots, m$ , produce a posterior sample for the parameters of the marginal distributions of the  $X_{[\cdot, j]}$ 's, say  $\boldsymbol{\lambda}_j = \lambda_j^{(1)}, \lambda_j^{(2)}, \dots, \lambda_j^{(S)}$ ,  $j = 1, \dots, m$ . Store them into a  $S \times k$  matrix  $\boldsymbol{\lambda} = (\boldsymbol{\lambda}_1, \dots, \boldsymbol{\lambda}_j, \dots, \boldsymbol{\lambda}_m)$  where  $k$  is the sum of the dimensions of the parameter spaces of the marginal distributions.

[2:] Given a prior distribution  $\pi(\varphi)$  for the quantity of interest  $\varphi$ ,  
**for**  $b = 1, \dots, B$ ,

1. draw a random uniform integer  $t(b)$  in  $\{1, 2, \dots, S\}$ .
2. use the  $t(b)$ -th row of  $\boldsymbol{\lambda}$  to create a matrix of uniformly distributed *pseudo*-data

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

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

3. draw  $\varphi^{(b)} \sim \pi(\varphi)$ ;
4. compute

$$EL(\varphi^{(b)}; \mathbf{u}^{(t(b))}) = \omega_b;$$

**end for**

[3.] store the values  $(\varphi^{(b)}, \omega_b)$ ,  $b = 1, \dots, B$ .

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

---

**Algorithm 3:** Modified ABCOP algorithm

---

From general results on empirical likelihood (Owen, 2010), one has

$$EL(\rho; \rho_n) = \prod_{i=1}^n \left( 1 + \eta g(R_i, S_i; \rho) \right)^{-1}$$

where  $\eta$  is the Lagrange multiplier which can be explicitly obtained from

$$\sum_{i=1}^n \frac{g(R_i, S_i; \rho)}{1 + \eta g(R_i, S_i; \rho)} = 0,$$

where

$$g(R_i, S_i; \rho) = \frac{12R_i S_i}{n^2 - 1} - 3 \frac{n+1}{n-1} - \rho.$$

We can then use Algorithm 2, with  $S = 1$ , to produce a posterior sample for the quantity of interest  $\rho$ .



### 5.1. A small scale simulation

As an illustration we have simulated 1,000 samples of size  $n = 100$  from a bivariate Clayton's Copula, whose expression is

$$C(u, v) = \begin{cases} uv & \theta = 0 \\ (u^{-\theta} + v^{-\theta} - 1)^{-1/\theta} & \theta > 0 \end{cases}.$$

For comparative purposes we have also implemented the nonparametric frequentist procedure described in Genest and Favre (2007), where a confidence interval for the Spearman's  $\rho$  is constructed based on the asymptotic sampling distribution of  $\rho_n$ . In the simulation  $\theta$  has been set equal to 1.076, which implies  $\rho \approx 0.5$ . Figure 1 compares the frequentist behavior of confidence procedure and our proposal. The first three box-plots show the sampling (over the 1,000 generated samples) distribution of i) the lower limit of the equal-tail confidence interval with nominal coverage set at 0.95, ii) the point estimate  $\rho_n$ , and iii) the upper limit of the equal-tail confidence interval with nominal coverage set at 0.95. The last three box-plots show the sampling distribution of some specific quantiles (namely the 2.5th, the median and the 97.5th percentiles) of the approximated posterior distribution. Samples were generated by fixing  $\theta = 1.076$ . The prior distribution for  $\rho$  has been taken  $\text{Unif}(-1, 1)$ . Computations were done in **R**, using libraries `copula` and `gmm`.

One can see that our procedure produces more precise estimates in terms of intervals. The empirical estimate and the posterior median behave very similarly. The average length of the confidence interval is 0.820 while the average length of the equal-tail 0.95 credible set is 0.784. In 626 out of 1,000 simulation, the Approximate Bayesian interval was shorter than the classical confidence interval.

### 5.2. Simulated non uniform data

Here we show an example of bivariate data with non uniform marginal distribution. Data were generated from a Clayton copula with  $\theta = 1.076$  ( $\rho \approx 0.5$ ), and the two marginal distributions were transformed into an exponential distribution with mean 1/3 (for  $X_1$ ) and a Gaussian distribution with mean 3 and variance 1 (for  $X_2$ ). Figure 2 shows the scatterplot of raw and transformed data. In this particular case the observed value for  $\rho_n$  was 0.568.

Figure 3 shows the histogram of the  $BC_{EL}$  posterior sample for  $\rho$  obtained from Algorithm 2. One can notice that the posterior mass is practically entirely on the right of zero, and the posterior mean is 0.56, very close to the observed  $\rho_n$ .

### 5.3. An alternative estimator

From a purely pragmatic perspective, it might be tempting to follow an unconventional and "hybrid" route, which we now describe. For each  $s = 1, \dots, S$ ,

1. Provide an estimate of  $\varphi$ , using  $\hat{\varphi}^{(s)}$  from the *plugged-in* model

$$p(\mathbf{x}; \text{marginals}, \varphi, \lambda_1^{(s)}, \dots, \lambda_m^{(s)}) \quad (4)$$

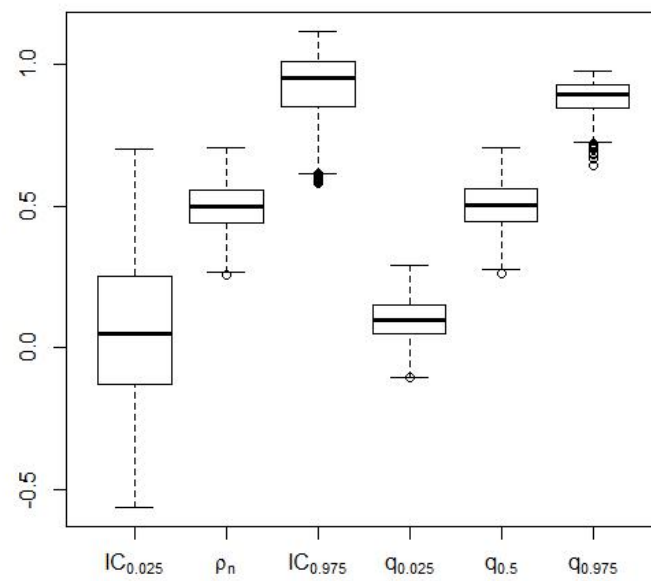


Figure 1 – 1,000 simulations from a Clayton copula: sample size is 100; the true value of  $\rho$  is equal to 0.5. The six box-plots respectively shows the simulated sampling distribution of: i) the lower limit of the equal-tail confidence interval with nominal coverage set at 0.95, ii) the point estimate  $\rho_n$ , iii) the upper limit of the equal-tail confidence interval with nominal coverage set at 0.95. iv) the 2.5-th v) the 50-th and vi) the 95-th quantile of the approximate posterior distribution.

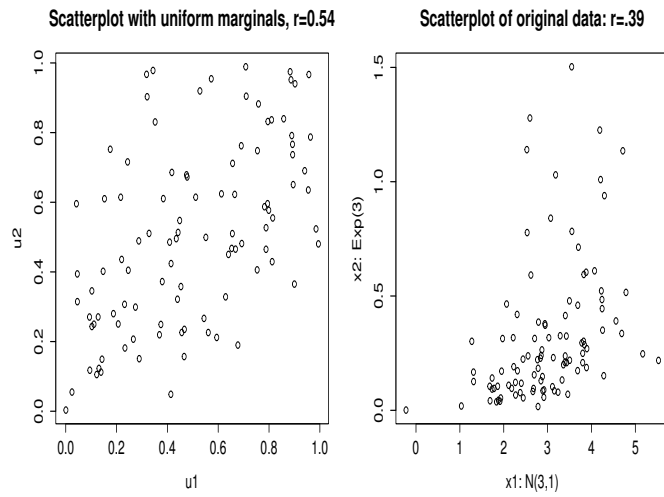


Figure 2 – Scatterplot of the simulated data and pseudo-data:  $X_1 \sim \text{Exp}(3)$ ;  $X_2 \sim N(3, 1)$

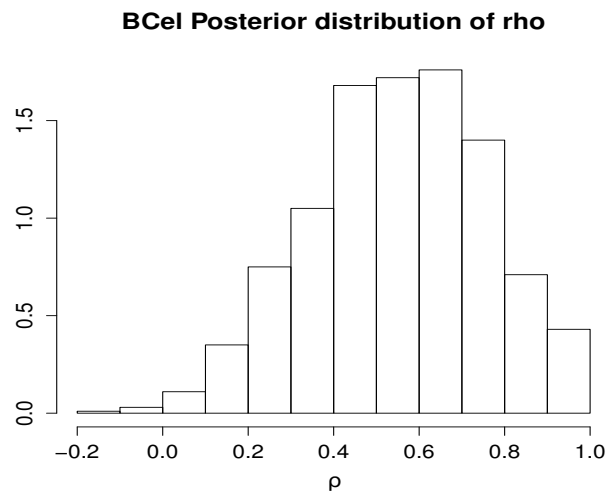


Figure 3 – Histogram of the posterior sample of values of  $\rho$ , using Algorithm 2.

In particular, one could use a sort of maximum likelihood estimate of  $\varphi$ , assuming that the sampling distribution is given by (4).

2. Use the distribution of the  $\hat{\varphi}^{(s)}$ 's as a surrogate of the posterior distribution of  $\varphi$ .

This approach is a further approximation in many ways. First, the distribution of  $\hat{\varphi}^{(s)}$ 's in step 2 of the above procedure could not properly be treated as a posterior distribution, since we have not introduced any prior distribution on  $\varphi$ . Second, the distribution in step 2 is not a distribution on  $\varphi$ : rather, it can be interpreted as the posterior distribution of the following quantity

$$\hat{\varphi}(\Lambda) = \operatorname{argmax}_{\varphi} p(\mathbf{x} | \text{marginals}, \varphi, \Lambda). \quad (5)$$

Notice that

$$\begin{aligned} \mathbb{E}^{\Lambda}(\hat{\varphi}(\Lambda)) &\neq \operatorname{argmax}_{\varphi} \mathbb{E}^{\Lambda}(p(\mathbf{x} | \text{marginals}, \varphi, \Lambda)) \\ &= \operatorname{argmax}_{\varphi} IL(\varphi; \mathbf{x}) = \hat{\varphi}^{(IL)}, \end{aligned}$$

where the above expectation is taken with respect to the posterior distribution of the marginal parameters  $\Lambda$ , based on the ‘‘marginal’’ samples and suitable prior information, and  $IL$  represents the ‘‘correct’’ integrated likelihood,

$$IL(\varphi; \mathbf{x}) = \int_{\Lambda} p(\mathbf{x}; \boldsymbol{\lambda}, \varphi) \pi(\boldsymbol{\lambda} | \varphi) d\boldsymbol{\lambda}.$$

Also,  $\operatorname{Var}(\hat{\varphi}(\Lambda))$  under-reports the variability of the estimator, since

$$MSE = \operatorname{Var}(\hat{\varphi}(\Lambda)) + (\mathbb{E}^{\Lambda}(\hat{\varphi}(\Lambda)) - \hat{\varphi}^{(IL)})^2$$

However, in practical applications this method works better than the IFM approach, described in §1. Figure 4 shows the behavior of this method with the data used in Figure 2. One can notice a slight bias towards larger values of  $\rho$  and an incorrect report of uncertainty.

## 6. EXAMPLE: SPEARMAN’S $\rho$ FOR STUDENT-T LOG-RETURNS

We now analyze a real data-set containing the log-returns FTSE-MIB of two Italian banks, Monte dei Paschi di Siena (BMPS) and Banco Popolare (BP), by assuming that the log-returns for each bank may be described by a GARCH(1,1) model with Student- $t$  innovations for the log-returns  $\{y_t\}$  from 01/07/2013 to 30/06/2014 (only weekdays) available on the web page <https://it.finance.yahoo.com>.

The GARCH(1,1) model for Student- $t$  innovation may be rewritten via data augmentation, following Geweke (1993):

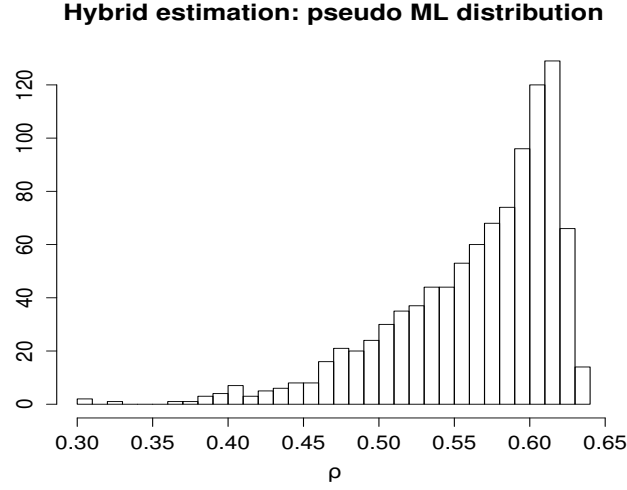


Figure 4 – Hybrid method: “posterior” distribution of  $\hat{\varphi}(\lambda)$

$$\begin{aligned}
 y_t &= \varepsilon_t \sqrt{\frac{\nu - 2}{\nu}} \omega_t h_t \quad t = 1, \dots, T \\
 \varepsilon_t &\sim \mathcal{N}(0, 1) \\
 \omega_t &\sim IG\left(\frac{\nu}{2}, \frac{\nu}{2}\right) \\
 h_t &= \alpha_0 + \alpha_1 y_{t-1}^2 + \beta h_{t-1} \quad t = 1, \dots, T
 \end{aligned}$$

where  $\alpha_0 > 0$ ,  $\alpha_1, \beta \geq 0$  and  $\nu > 2$ ,  $\mathcal{N}(0, 1)$  denotes the standard normal distribution and  $IG(a, b)$  denotes the inverted gamma distribution with shape parameter  $a$  and scale parameter  $b$ . Figure 5 shows the scatterplot of the log-returns and the transformed version of them, using, as a point estimate, the posterior mean of each parameter.

For each bank, the posterior distribution of the model parameters  $(\alpha_0, \alpha_1, \beta, \nu)$  may be approximated by using the **R** package `bayesGARCH` (Ardia and Hoogerheide, 2010). Once a sample from the approximated distribution is simulated for each parameter and for each bank, Algorithm 3 is applied as follows:

for  $m = 1, \dots, M$

- 1: Simulate a value  $\rho^{(m)} \sim Unif(-1, 1)$ .
- 2: Sample two integer values  $b_j^{(m)}$  ( $j = 1, 2$ ) in  $\{1, \dots, S\}$ , where  $S$  is the number of posterior simulations.
- 3: Consider the  $b_j^{(m)}$ -th row of the MCMC output for the parameters of the  $j$ -th marginal (i.e.  $\alpha_{0j}, \alpha_{1j}, \beta_j, \nu_j$ ), for  $j = 1, 2$ ).

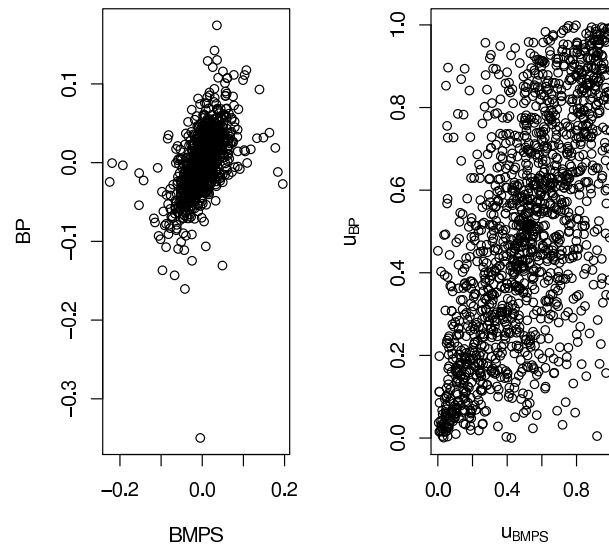


Figure 5 – Scatterplot of the log-returns of the investments of Monte dei Paschi di Siena (BMPS) and Banco Popolare (BP) on the left and of the transformed data on the right.

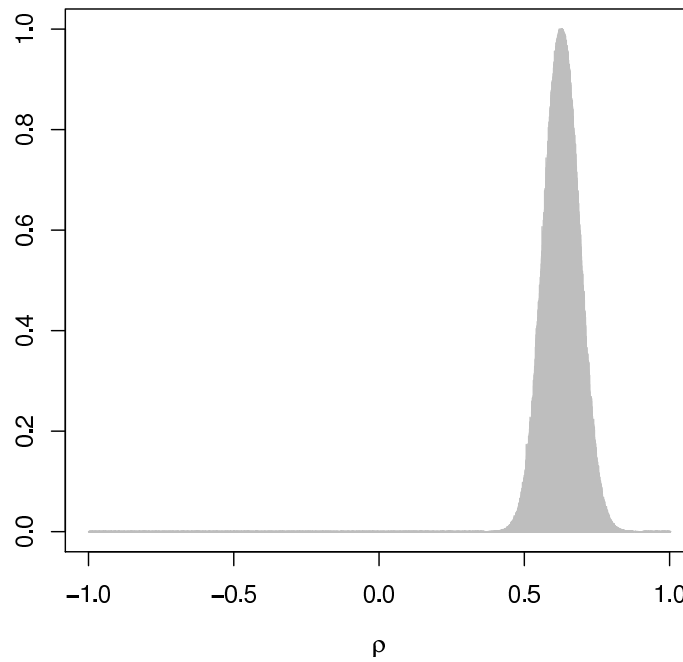


Figure 6 – Approximation of the posterior distribution of the Spearman's  $\rho$  for the log-returns of the investments of two Italian institutes based on 10,000 simulations.

- 4: Compute pseudo-data  $u_{ij}^{(m)}$  for  $i = 1, \dots, T$  and  $j = 1, 2$  as

$$u_{ij}^{(m)} = \mathbf{F}_{\nu_j} \left( y_i; \frac{\nu_j^{(m)} - 2}{\nu_j^{(m)}} h_{ij}^{(m)} \right)$$

where  $\mathbf{F}_{\nu}(x, d)$  is the CDF of a Student- $t$  distribution with  $\nu$  degrees of freedom and scale parameter  $d$ .

- 5: Compute the estimated sample Spearman's  $\rho_n^{(m)}$  as in (3) and the weight relative to the simulated  $\rho^{(m)}$  as  $\omega^{(m)} = EL(\rho_n^{(m)}; \mathbf{u}_1^{(m)}, \mathbf{u}_2^{(m)})$  as in Owen (2010).

The output of Algorithm 3 relative to the log-returns of Monte dei Paschi di Siena and Banco Popolare are shown in Figure 6: the estimated posterior mean of  $\rho$  is 0.614.

#### ACKNOWLEDGEMENTS

The Authors are grateful to two anonymous referees for their helpful suggestions and to N. Lunardon for providing some  $\mathbf{R}$  functions, very useful for our simulation. C. Grazian's research has been sponsored by Università ItaloFrancese (Université FrancoItalienne). B. Liseo's research has been sponsored by grant n. C26A13BBCB, Sapienza Università di Roma, 2013.

#### REFERENCES

- D. ARDIA, L. HOOGERHEIDE (2010). *Bayesian Estimation of the GARCH(1,1) Model with Student-t Innovations*. The R Journal, 2, no. 2, pp. 41–47.
- P. CHAUSSÉ (2010). *Computing Generalized Method of Moments and Generalized Empirical Likelihood with R*. Journal of Statistical Software, 11, no. 34, pp. 1–35.
- U. CHERUBINI, E. LUCIANO, W. VECCHIATO (2004). *Copula Methods in Finance*. John Wiley & Sons, New York, San Francisco, Calif.
- V. R. CRAIU, A. SABETI (2012). *In mixed company: Bayesian inference for bivariate conditional copula models with discrete and continuous outcomes*. J. Multivariate Anal., 110, pp. 106–120.
- C. GENEST, A.-C. FAVRE (2007). *Everything you always wanted to know about copula modeling but were afraid to ask*. Journal of Hydrologic Engineering, pp. 347–368.
- C. GENEST, K. GHOUDI, L. RIVEST (1995). *A semiparametric estimation procedure of dependence parameters in multivariate families of distributions*. Biometrika, 82, no. 3, pp. 543–552.

- J. GEWEKE (1993). *Bayesian treatment of the independent Student-t linear model*. J. Appl. Econometr., S1, no. 8, pp. S19–S40.
- J. HE, H. LI, A. C. EDMONDSON, D. L. RADER, M. LI (2012). *A Gaussian copula approach for the analysis of secondary phenotypes in case-control genetic association studies*. Biostat., 13, no. 3, pp. 497–508.
- P. HOFF (2007). *Extending the rank likelihood for semiparametric copula estimation*. The Annals of Applied Statistics, 1, no. 1, pp. 265–283.
- V. N. HUYNH, V. KREINOVICH, S. SRIBOONCHITTA (2015). *Modeling Dependence in Econometrics*. Springer, New York.
- H. JOE (2014). *Dependence modelling with copulas, Monographs on Statistics and Applied Probability, 134*. Chapman & Hall–CRC Press, London.
- J. MARIN, P. PUDLO, C. ROBERT, R. RYDER (2012). *Approximate Bayesian computational methods*. Statistics and Computing, 6, no. 22, pp. 1167–1180.
- K. MENGENSEN, P. PUDLO, C. ROBERT (2013). *Bayesian computation via empirical likelihood*. Proc. of the National Academy of Sciences, 4, no. 110, pp. 1321–1326.
- A. MIN, C. CZADO (2010). *Bayesian Inference for Multivariate Copulas using Pair-copula Constructions*. Journal of Financial Econometrics, 4, no. 8, pp. 511–546.
- D. OH, A. PATTON (2013). *Simulated Method of Moments Estimation for Copula-Based Multivariate Models*. J. Amer. Stat. Assoc., 502, no. 108, pp. 689–700.
- A. OWEN (2010). *Empirical Likelihood*. Chapman & Hall/CRC Press, New York (USA).
- G. RESCONI, I. LICATA (2015). *Entropy and Copula Theory in Quantum Mechanics*. In *Proceedings of the 1st Int. Electron. Conf. Entropy Appl., 3–21 November 2014*, Sciforum Electronic Conference Series.
- D. RUBIN (1988). *Using the SIR algorithm to simulate posterior distributions (with discussion)*. In *Bayesian Statistics 3 J.M. Bernardo, M.H. DeGroot, D.V. Lindley, and A.F.M. Smith, Eds.*, Oxford University Press, Oxford (UK), Bayesian Statistics, pp. 395–402.
- R. SCHEFZIK, T. L. THORARINSDOTTIR, T. GNEITING (2013). *Uncertainty Quantification in Complex Simulation Models Using Ensemble Copula Coupling*. Statist. Sci., 28, no. 4, pp. 616–640.
- S. SCHENNACH (2005). *Bayesian exponentially tilted empirical likelihood*. Biometrika, 1, no. 92, pp. 31–46.
- C. SCHOLZEL, P. FRIEDERICHS (2008). *Multivariate non-normally distributed random variables in climate research: introduction to the copula approach*. Non-linear Processes in Geophysics, 15, pp. 761–772.



- R. D. S. SILVA, H. F. LOPES (2008). *Copula, marginal distributions and model selection: a Bayesian note*. Stat. Comput., 18, no. 3, pp. 313–320.
- S. SISSON, Y. FAN (2011). *Likelihood-free MCMC*. In *Handbook of Markov chain Monte Carlo*, Chapman & Hall/CRC, London, Handb. Mod. Stat. Methods, pp. 313–335.
- M. SKLAR (1959). *Fonctions de répartition à  $n$  dimensions et leurs marges*. Publ. Inst. Statist. Univ. Paris, 8, pp. 229–231.
- M. SMITH (2013). *Bayesian Approaches to Copula Modelling*. In *Hierarchical Models and MCMC: A Tribute to Adrian Smith, P. Damien, P. Dellaportas, N. Polson, and D. Stephens (Eds)*, Oxford University Press, Oxford (UK), Bayesian Statistics, pp. 395–402.
- M. SMITH, Q. GAN, R. KOHN (2012). *Modelling dependence using skew- $t$  copulas: Bayesian inference and applications*. Journal of Applied Econometrics, 3, no. 27, pp. 500–522.
- J. WU, X. WANG, S. WALKER (2014). *Bayesian Nonparametric Inference for a Multivariate Copula Function*. Methodol. Comput. Appl. Probab., 1, no. 16, pp. 747–763.

## SUMMARY

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 and it is based on the properties of an Approximate Bayesian Monte Carlo algorithm, where the proposed values of the functional of interest are weighed in terms of their empirical likelihood. This method is particularly useful when the “true” likelihood function associated with the working model is too costly to evaluate or when the working model is only partially specified.

**Keywords:** ABC algorithms; Multivariate distributions; Partially specified models.