

Approximate Bayesian Inference for Copula Models

Inferenza bayesiana approssimata per copule

Brunero Liseo

Abstract We propose a method for making inference on a functional of a multivariate distribution. The method is based on a copula representation and on the properties of an Approximate Bayesian algorithm based on empirical likelihood.

Abstract *Si propone un metodo per la stima di un funzionale di interesse relativo ad un modello multivariato. Il metodo si basa sulla rappresentazione in termini di copula e sfrutta le proprietà delle tecniche computazionali bayesiane approssimate che utilizzano metodi di verosimiglianza empirica.*

Key words: ABC methods, Bayesian inference, copula, empirical likelihood.

1 Introduction

Copula models are nowadays widely used in multivariate data analysis. A copula is a flexible tool that allows the researcher to model the joint distribution in two separate steps: the marginal distributions and a copula function which captures the dependence structure of the joint distribution.

Whereas it is 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 defined, is crucial and complex, especially in high dimensional problems. On the other hand, dependence is one of the most fundamental features in (applied) statistics, economics and probability.

In the classical approach to copula estimation, there are no broadly satisfactory methods for the joint estimation of marginal and dependence parameters. Bayesian alternative are not yet fully developed, although Min and Czado (2010), Smith et al. (2012) and Wu et al. (2013) are remarkable instances.

Brunero Liseo

MEMOTEF, Sapienza Università di Roma, via del Castro Laurenziano, 9 Roma 00161, e-mail: brunero.liseo@uniroma1.it

In this work we consider the general problem of estimating some specific quantity of interest of a generic copula (such as, for example, tail dependence index or Spearman's ρ) by adopting an approximate Bayesian approach along the lines of Mengersen et al. (2013). In particular, we discuss 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 is required only on the parameter of interest, and its prior is combined with the empirical likelihood; ii) we do not use the “true” likelihood function, but rather an approximation based on empirical likelihood theory (Owen, 2010). This hopefully reduces the potential for incorrect distributional assumptions. Our approach can be applied both to parametric and nonparametric modelling of the marginal distributions. The method described in this paper, although using a different kind of approximation, is in the spirit of Hoff (2007), and it is similar, although from a different perspective, to the analysis in Schennach (2005).

2 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 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 . The copula function C is a CDF with uniform margins on $[0, 1]$: it binds together the univariate CDF F_1, F_2 , and F_m 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 , assume they have continuous CDF's F_1 and F_2 . It is well known that both the transformed variables $U_1 = F_1(X_1)$ and $U_2 = F_2(X_2)$ have uniform marginal distributions. A semiparametric copula model consists of a parametric model for the joint distribution of (U_1, U_2) and no assumptions on the marginal CDF's. A nonparametric copula is assumed when the joint distribution of (U_1, U_2) depends on an infinite dimensional parameter. In this paper we will allow the marginal distributions F_j 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 .

Empirical likelihood 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 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 parameter of interest, say ϕ , expressed as a functional of F , and then a sort of

profile likelihood is computed in a nonparametric manner. Consider a given set of generalized moment conditions of the form $E_F(h(X, \varphi)) = 0$, where $h(\cdot)$ is a known function, and φ is the quantity of interest. The resulting empirical likelihood is defined as $L_{EL}(\varphi; \mathbf{x}) = \max_{\mathbf{p}} \prod_{i=1}^n p_i$, for all $\mathbf{p}$ such that $0 \leq p_i \leq 1$, $\sum_{i=1}^n p_i = 1$, and $\sum_{i=1}^n h(x_i, \varphi) p_i = 0$. See Owen (2010) for a general discussion of empirical likelihood methods.

3 ABC and EL

Approximate Bayesian computation has now become an essential tool for the analysis of complex stochastic models when the likelihood function is unavailable or expensive to evaluate. 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 value from the prior and the value is accepted only if it produces new data which are similar enough to the actual data. Many more sophisticated computational strategies are available in order to avoid generating values from the prior; 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 is a sort of re-sampling scheme where the proposed values are re-sampled with weights proportional to their empirical likelihood. In practice, the algorithm is a sort of sampling importance re-sampling method for models in which the “true likelihood” evaluation is out of reach and the “true” weights are approximated through their empirical likelihood.

Algorithm 1 BC_{EL} algorithm

```

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

```

4 The proposed approach

Here 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 our main goal is the estimation of a finite dimensional parameter of interest. In practice this is 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 with the aim to creating more flexible models that might better fit the data at hand. In order to make inference on the quantity of interest, one must deal with them, in some way. Even if some of the additional parameters are not particularly important in terms of estimation - they often lack of a precise physical meaning - their estimates can dramatically affect the inference 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.

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. 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 use the information in $X_{[:,j]}$ to estimate the marginal CDF of $X_{[:,j]}$. Let $\lambda_j = (\lambda_j^{(1)}, \lambda_j^{(2)}, \dots, \lambda_j^{(S)})$, $j = 1, 2, \dots, m$ be the posterior sample obtained from *some* Bayesian inference method for the distribution of X_j . λ_j can be either a sample from the posterior distribution of the parameters of the model 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 $(X_1, \dots, X_m)$,

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

where θ is the parameter related to the copula function. Since we have already *estimated* the $F_j(x_j)$'s, we now need to estimate $C_\theta(\cdot)$. This can be done 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 follow a nonparametric route and we concentrate on some specific function of $C_\theta(\cdot)$, say $\psi = T(F)$. This is particularly useful when there is no theoretical or empirical evidence that a given copula should be used and we are mainly interested in some specific synthetic measure of the multivariate dependence.

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 ϕ . There are several critical issues both in the practical implementation of the method and in its theoretical properties. First, the constraints might hold only asymptotically: for example the sample version of the Spearman's ρ is only asymptotically unbiased. Also, prior information is only provided for the marginal distributions and for ϕ :

Algorithm 2 ABCOP algorithm

[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 φ ,
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 $\omega_b = S^{-1} \sum_{s=1}^S \omega_{bs}$
4. re-sample - with replacement - from $(\varphi^{(b)}, \omega_b), b = 1, \dots, B$.

end for

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 proposed 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 **R**; it use some functions contained in the suite `gmm`: see for example Chaussé (2010).

5 A simple illustration: Spearman's ρ

To illustrate the method, for the sake of brevity, we only consider a simple example with $m = 2$. The Spearman's ρ measure of dependence is defined by $\rho = 12 \int_0^1 \int_0^1 C(u, v) du dv - 3$. Its sampling counterpart ρ_n is nothing but the correlation among ranks and it can be written as

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

First we produce a size S posterior sample from the two marginal distributions of X_1 and X_2 , say $(\lambda_j^1, \lambda_j^S)$, for $j = 1, 2$. For each fixed $s = 1, \dots, S$ let $\lambda^{(s)} = (\lambda_1^{(s)}, \lambda_2^{(s)})$, and compute the $n \times 2$ matrix $\mathbf{U}^{(s)}$ whose generic element is given by $u_{ij}^{(s)} = F_j(x_{ij}; \lambda_j^{(s)}) \in [0, 1]$, $i = 1, \dots, n$; $j = 1, 2$. Then we take ranks $(R_i^{(s)}, S_i^{(s)})$ of the original values and compute $\rho_n^{(s)}$ for each single $\lambda^{(s)}$. Also we are able to evaluate the empirical likelihood of ρ for a given value of ρ_n as $\max_{\rho_i} EL(\rho; \lambda^{(s)}) = \prod_{i=1}^n n p_i(\rho)$ under the constraints $\sum_{i=1}^n p_i = 1$, and $\sum_{i=1}^n p_i \left(\frac{12 R_i^{(s)} S_i^{(s)}}{n^2 - 1} - 3 \frac{n+1}{n-1} - \rho \right) = 0$. From general results on empirical likelihood, one has

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

where η is the Lagrange multiplier which can be obtained from $\sum_{i=1}^n \frac{g(R_i^{(s)}, S_i^{(s)}; \rho)}{1 + \eta g(R_i^{(s)}, S_i^{(s)}; \rho)} = 0$, where $g(R_i^{(s)}, S_i^{(s)}; \rho) = \frac{12 R_i^{(s)} S_i^{(s)}}{n^2 - 1} - 3 \frac{n+1}{n-1} - \rho$. We can then use Algorithm 2 to produce a posterior sample for the quantity of interest ρ .

References

- Chaussé, P. (2010): Computing Generalized Method of Moments and Generalized Empirical Likelihood with R, *Journal of Statistical Software*, 34, 11, 1–35.
- Hoff, P. (2007): Extending the rank likelihood for semiparametric copula estimation, *The Annals of Applied Statistics*, 1, 1; 265–283.
- Marin, J.M.; Pudlo, P.; Robert, C.P.; Ryder, R.J. (2012): Approximate Bayesian computational methods, *Statistics and Computing*, 22(6), 1167–1180.
- Mengersen, K., Pudlo, P., Robert, C.P. (2013): Bayesian computation via empirical likelihood, *Proc. of the National Academy of Sciences*, 110 (4), 1321–1326.
- Min, A., Czado, C. (2010): Bayesian Inference for Multivariate Copulas using Pair-copula Constructions, *Journal of Financial Econometrics*, 8(4), 511–546.
- Owen, A.B. (2010): *Empirical Likelihood*, Chapman & Hall/CRC Press, New York.
- Schennach, S.M. (2005): Bayesian exponentially tilted empirical likelihood, *Biometrika*, 92, 1, 31–46.
- Shih, J.H.; Louis, T.A. (1995): Inferences on the association parameter in copula models for bivariate survival data, *Biometrics* 51 (1995), no. 4, 1384–1399.
- Smith, M.S., Gan, Q., Kohn, R.J. (2012): Modelling dependence using skew- t copulas: Bayesian inference and applications, *Journal of Applied Econometrics*, 27, 3, 500–522.
- Wu, J., Wang, X., Walker, S.G. (2014): Bayesian Nonparametric Inference for a Multivariate Copula Function, *Methodol. Comput. Appl. Probab.* (in press).