U.P.B. Sci. Bull., Series A, Vol. 70, No. 3, 2008 ISSN 1223-7027

BAYESIAN INFERENCE FOR COPULA MODELS

Mariana CRAIU[1], Radu V CRAIU[2]

Prezentam o metoda generală de aplicare a inferenței Bayesiene pentru repartiții bidimensionale definite de copule. In această lucrare considerăm cazul în care ambele repartiții marginale sunt Weibull sau exponențiale dar metoda poate fi extinsă şi la alte repartiții. Pentru determinarea repartiției posterioare se folosesc tehnici de simulare Markov chain Monte Carlo.

We present a general methodology for performing Bayesian inference on copula models. Here we discuss the case in which each marginal distribution is Weibull or Exponential but the approach can be generalized to other distributions. We solve the computational problem associated with sampling from the posterior distribution using Markov chain Monte Carlo. We illustrate the method with simulated data in order to assess its efficiency.

Key words: Copula models, Posterior distribution, Gamma and Weibull distributions, Bayesian statistics.

1. Introduction

The term *copula* was first introduced by Sklar (1959) following some initial ideas by Hoffding (1940). Copulas can flexibly "couple" fixed marginal continuous distributions into a multivariate distribution. There exists a vast literature on connections between dependence concepts and various families of copulas but for reference we recommend Joe (1997) and Nelsen (2006).

The multivariate function $C:[0,1]^p \to [0,1]$ is called a *copula* if it is a continuous distribution function and each marginal is a uniform distribution function on [0,1] so that $C(u_1,...,u_p) = P(U_1 \le u_1,...,U_p \le u_p)$. If p=2 and if X, Y are continuous random variables with distribution functions (df) F and G, respectively, we specify the joint df using the copula C:[0,1] x [0,1] $\to$ [0,1] such that

$$H(F^{-1}(u),G^{-1}(v)) = P(X \le F^{-1}(u), Y \le G^{-1}(v)) = C(u,v). \qquad (1)$$

Equation (1) illustrates the way in which the copula function "bridges" the marginal and the joint df's. The existence of such a map C is guaranteed by Sklar's Theorem (Sklar, 1959). The uniqueness of C once we fix, F, G and H is guaranteed as long as the random variables are continuous.

---

[1] Prof., Department of Mathematics III, University "Politehnica" of Bucharest, ROMANIA
[2] Associate professor, University of Toronto, CANADA

In this paper we consider the case in which F and G are Weibull with unknown parameters (α_1,β_1), and (α_2,β_2) respectively, or Exponential with parameters a λ_1 and λ_2 The marginals are coupled using Clayton's copula (Clayton, 1978) with parameter θ so that

$$C_\theta(u,v) = (u^{-\theta} + v^{-\theta} - 1)^{-\frac{1}{\theta}}. \tag{2}$$

The copula density corresponding to (2) is

$$c_\theta(u,v) \propto (1+\theta)u^{-\theta-1}v^{-\theta-1}(u^{-\theta} + v^{-\theta} - 1)^{-\frac{1+2\theta}{\theta}}. \tag{3}$$

In the next section we discuss Bayesian inference methodology for estimating the parameters $\alpha_1,\beta_1,\alpha_2,\beta_2$ and θ.

2. Prior and posterior distributions

We start our discussion in the case of Weibull marginals. The two-parameter Weibull density is

$$f(x\,|\,\alpha,\beta) = 1_{[0,\infty)}(x)\alpha\beta^{-\alpha}x^{\alpha-1}\exp[-(x/\beta)^\alpha]\ , \tag{4}$$

for $\alpha,\beta > 0$

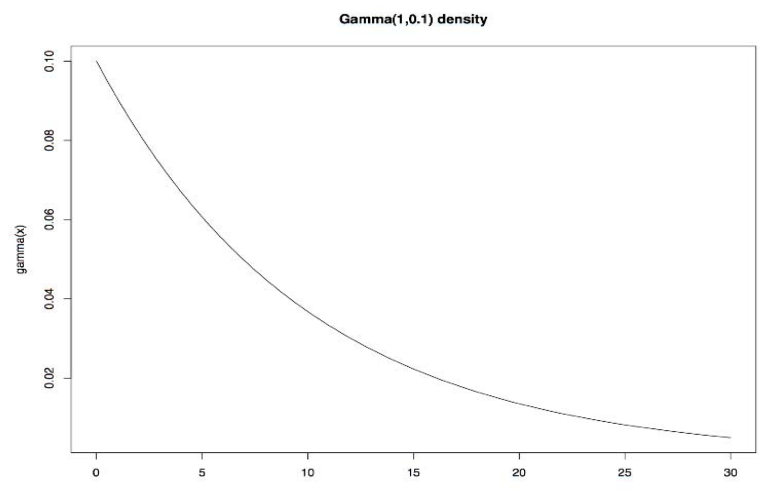

Fig. 1: Density of the prior Gamma (1,0.1).

We assume for each parameter of the Weibull distribution a diffuse Gamma distribution prior with parameters 1 and 0.1 as shown in Figure 1. The parameter of the Clayton family is restricted to the positive real axis so we consider for it the same prior.

Assuming we have available n pairs of observations from X and Y and if we denote $\xi = (\alpha_1, \beta_1, \alpha_2, \beta_2, \theta)$ then the posterior distribution is

$$\pi(\xi) \propto \pi(\alpha_1)\pi(\beta_1)\pi(\alpha_2)\pi(\beta_2)\pi(\theta) \times \\ \times \prod_{i=1}^{n} f(x_i \mid \alpha_1, \beta_1) f(y_i \mid \alpha_2, \beta_2) c_\theta(F(x_i \mid \alpha_1, \beta_1), F(y_i \mid \alpha_2, \beta_2)) \quad (5)$$

where c is the copula density (3), f has the form given in (4), and F, G are the corresponding cumulative distribution functions (cdf). The cdf of the Weibull distribution is known as

$$F(x \mid \alpha, \beta) = 1 - \exp(-(x/\beta)^\alpha)$$

In the case of Exponential marginals we use, for λ_1 and λ_2, the same prior as the one described above.
In both the Weibull and the exponential cases we encounter computational difficulties when we try to determine the normalizing constant for the posterior distribution. Therefore we have to compute the characteristics of the posterior using Markov chain Monte Carlo (MCMC) algorithms.

3. Metropolis algorithms within MCMC

Many problems arising in Bayesian statistics involve calculation of integrals of the form

$$I = \int f(x)\pi(x)dx, \quad (6)$$

where π is a (posterior) density known up to th**e** **n**ormalizing constant. The Monte Carlo method produces an approximation of I in (5) using an i.i.d. sample $X_1, ..., X_n$ from π for the *Monte Carlo estimator*

$$\varphi = \sum_{i=1}^{n} h(x_i)/n$$

Unfortunately, in most cases, sampling independently from π is impossible. Markov chain Monte Carlo (MCMC) methods makes it possible to obtain numerical approximations of expectations such as (6) in cases when direct independent sampling from π is not available. Note that once a sample from π is available we can compute (6) for any integrable function h.

The underlying principle of an MCMC algorithm is the construction of an ergodic Markov chain whose stationary distribution has density π. More precisely, suppose that we generate a sequence of random variables $X_0, X_1, ...$ such that at each time $t \geq 0$ X_t is sampled from a distribution P(| X_t) which depends only on the current state of the chain, X_t. As t increases, the chain gradually ``forgets'' its initial state and the distribution of X_t is closer and closer to the stationary distribution of the chain. With the samples obtained we can approximate I using the *ergodic average*

$$\varphi = \frac{1}{n} \sum_{i=m+1}^{m+n} h(X_i) \tag{7}$$

Note that we allow in (7) for first m samples to be simply discarded. The set of samples not used in the estimation is known as burn-in. The length of what constitutes an appropriate burn-in for a given problem remains an active area of research.

The most widely used MCMC algorithm is the one proposed by Hastings (1970), as a generalization of the sampler designed by Metropolis et al. (1953). For the *Metropolis-Hastings algorithm*, at each step t the next state X_{t+1} is chosen by first sampling a candidate draw y from a proposal distribution q($|\, x_t$). Note that the proposal distribution is allowed to depend on the current state of the chain; if it does not, the algorithm is also known *as independent Metropoli*s. Assuming that the target density is π, the candidate sample y is retained with probability $r(x_t, y)$, where

$$r(x_t, y) = \min\left(1, \frac{\pi(y) q(y \,|\, x_t)}{\pi(x_t) q(x_t \,|\, y)}\right). \tag{8}$$

If y is accepted then X_{t+1}=y, otherwise X_{t+1}= x_t. Note that the calculation of normalizing constant for π is bypassed in (8) since it cancels out between the numerator and the denominator.

For further theoretical and methodological developments related to MCMC sampling we refer to Robert and Casella (2004) and Liu (2001).

4. Sampling from the posterior distribution

We use an independent Metropolis algorithm to construct a five-dimensional Markov chain that has the posterior density (5) as its stationary distribution (in an slight abuse of notation we use interchangeably π to denote both the stationary density and distribution) . All the parameters are restricted to be positive so we use disperse gamma distributions for proposing new moves. Instead of updating all components of the parameter vector simultaneously, we perform a Metropolis algorithm for one-at-a-time update of each coordinate. This implies that for each iteration we perform five Metropolis updates. The advantage is that we can get a better hold of the proposal's parameters which are crucial to obtaining reasonable acceptance rates. In particular, one could use the adaptive method of Gasemyr (2003) to find appropriate values for the proposals' parameters. In the simulations performed for this paper we used proposals generated from Gamma (2,0.75) distribution.

We present in fig. 2, the scatter plot of the data consisting of 500 pairs. The two marginals are Weibull (1,1), for X, and Weibull(2,2) for Y. The copula parameter has value 4. The Uniform (0,1) random variates corresponding to a Clayton copula are generated using the algorithm of Devroye (1986) and using the inverse cdf are transformed into Weibull variates.

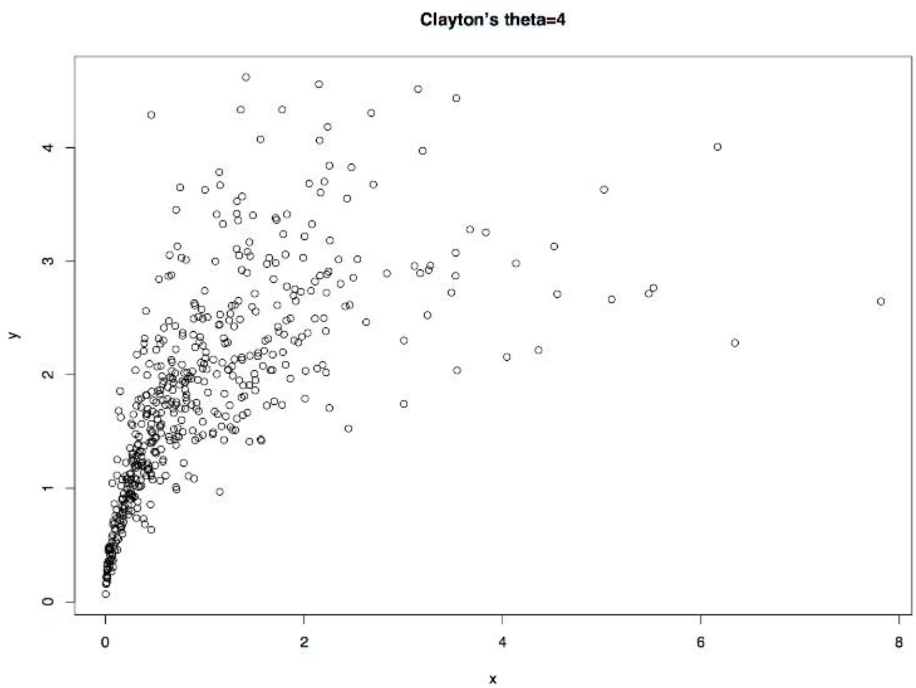

Fig. 2: Scatter plot of 500 realizations from Weibull(1,1) and Weibull(2,2) correlated using a Clayton copula with parameter 4.

In fig. 3 we show the traces of four out of five paths produced by the MCMC algorithm. One can see that after a while the path stabilizes around the true value marked with a horizontal solid line.

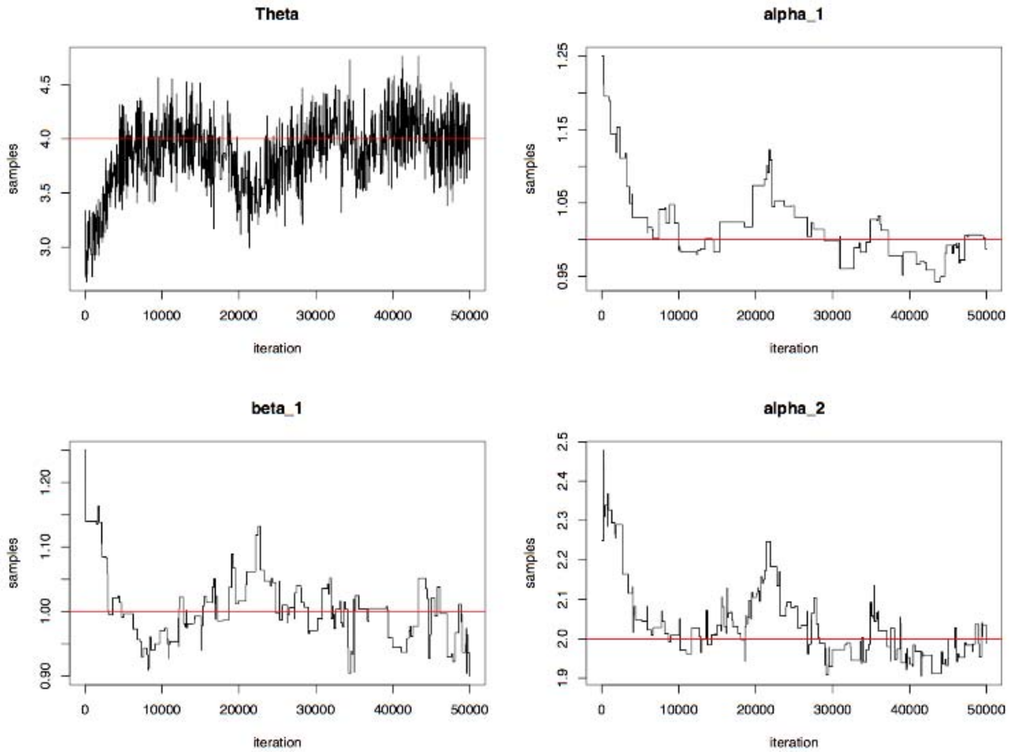

Fig. 3: Traces of the paths produced by the Metropolis algorithm for $\theta, \alpha_1, \beta_1, \alpha_2$. True values of the parameters are shown with horizontal lines.

Finally, in Table 1 we summarize the samples obtained with MCMC. Since the data's sample size is n = 500 one can fairly assume that the influence of the prior is minimal.

Table 1

Mean and SD (between brackets) for the MCMC estimates in the Weibull example

α_2	λ_1	λ_2	θ	θ
-0.011 (0.031)	0.017 (0.052)	0.003 (0.053)	0.033 (0.091)	-0.003 (0.056)

Similar graphics are produced for the situation with Exponential marginals. We consider a sample of 300 pairs where X is distributed Exp(2.5) and Y has distribution Exp(1.75). In fig. 5 we show the data scatterplot. Clayton's copula parameter is 2. The proposal distribution used in the independent Metropolis algorithm is Gamma(2,0.85).

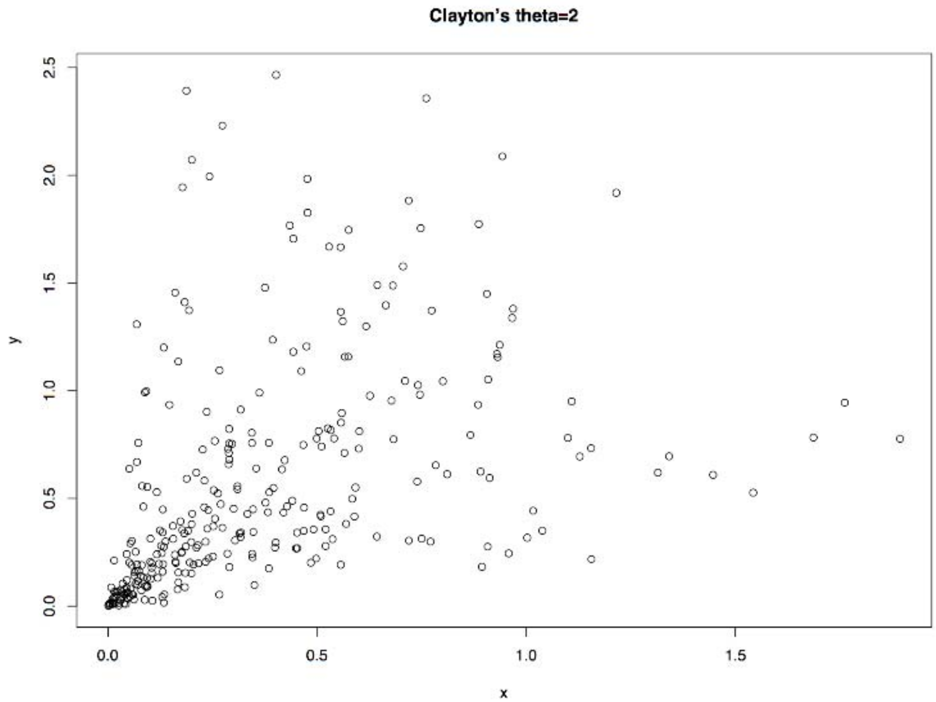

Fig. 4: Scatter plot of 300 realizations from Exp(2.5) and Exp(1.75) correlated using a Clayton copula with parameter 2.

In Table 2 we summarize the samples obtained after 10,000 iterations. One can see that the the performance of the Monte Carlo estimator remains good.

Table 2

Means and SD for the MCMC estimates in the Exponential example

λ_1	λ_2	θ
-0.044 (0.101)	0.033 (0.105)	0.102 (0.097)

6. Conclusions

We perform two Bayesian analyses of copula models. The distributions and class of copulas used in this paper are commonly used in reliability and medical studies. We studied the performance of the estimators with simulated data. However, advanced computational tools such as MCMC algorithms are needed in order to finalize the analysis.

R E F E R E N C E S

[1] *D.G. Clayton,* A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. Biometrika, **65,** 141--151, 1978.

[2] *J. Gasemyr*. On an adaptive version of the Metropolis-Hastings algorithm with independent proposal distribution. *Scand. J. Statistics,* **30,** 159—173, 2003.

[3] *L. Devroye,* Non-uniform Random Variate Generation. Springer, 1976.

[4] *W. K. Hastings,* Monte Carlo Sampling Methods using Markov chains and Their Applications. Biometrika, **57,** 97—109, 1970.

[5] *W. Hoffding* , Masztabinvariante Korrelationstheorie. Schr. Math. Inst. u. Inst. Angew. Math. Univ. Berlin, **5**, 181—233, 1940.

[6] *H. Joe,* Multivariate models and dependence concepts. Chapman and Hall,1997.

[7] *J.S. Liu* , Monte Carlo Strategies in Scientific Computing. Springer, 2001.

[8] *N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller and E. Teller* , Equations of state calculations by fast computing machines. *J. Chem. Phys.* **21,** 1087—1092, 1953.

[9] *R. B. Nelsen* , An Introduction to Copulas. Springer,2006.

[10] *C.P. Robert and G. Casella,* Monte Carlo statistical Methods., Springer, 2004.

[11] *M. Sklar* , Fonctions de repartition a n dimensions et leurs marges. Publ. Inst. Statist. Univ. Paris, **8**, 229—231,1959.