Approximate Bayesian Conditional Copulas

Brunero Liseo

Sapienza Università di Roma

brunero.liseo@uniroma1.it

Joint work with Luciana Dalla Valle and Clara Grazian

12° edizione delle Giornate del dipartimento MEMOTEF

31 Maggio, 2022

Outline

Main goal:
The implementation of Approximate Semiparametric Inference from a Bayesian perspective.

Ingredients:

- Approximate Bayesian Computation
- Empirical Likelihood
- Copulas.
- Gaussian Processes

Inference on dependencies

- One of the goals of a statistical analysis is often to understand how different phenomena are related and how to describe or to model such relations.
- Even in the simplest cases, eye impression can be misleading ...
- Linear correlation (ρ) is not able to describe, rank correlation (τ)does a better job

Some examples

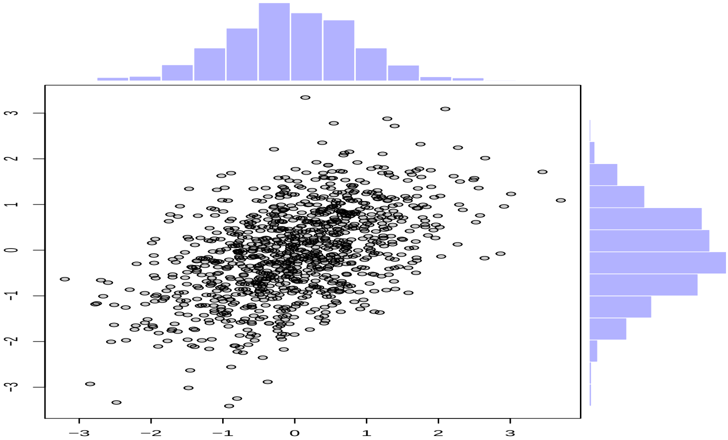


Figure: A bivariate Gaussian: $\rho = .505$, $\tau = .342$

Some examples

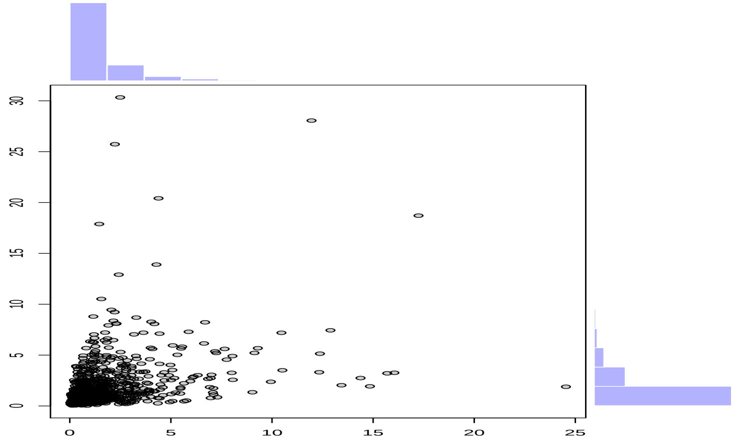


Figure: Different marginals: $\rho = .361$, $\tau = .342$

Some examples

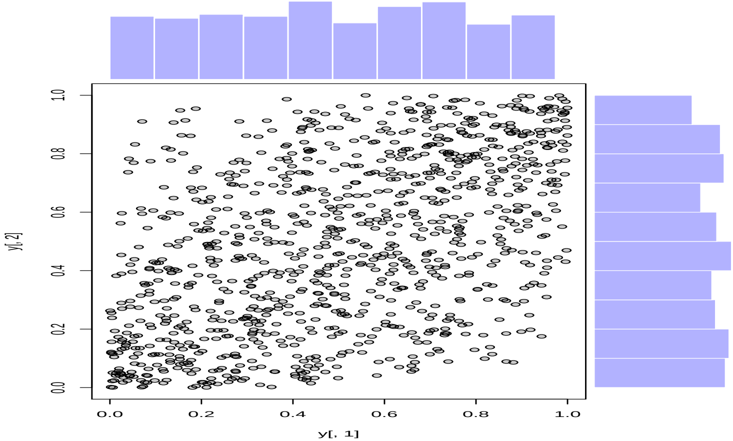


Figure: Uniform marginals: $\rho = .496$, $\tau = .342$

Some examples

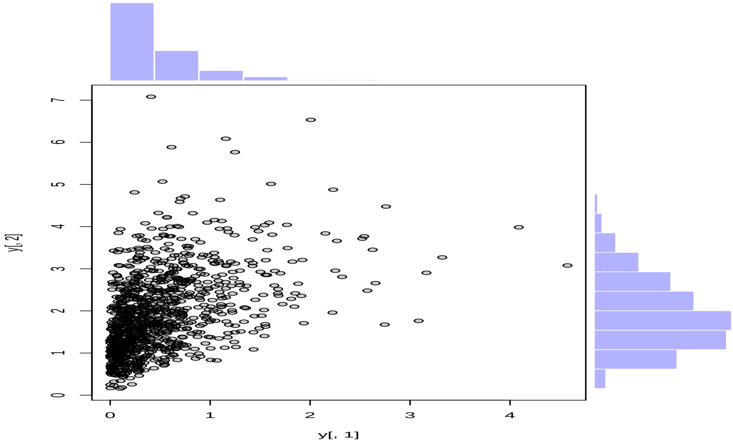


Figure: Gamma marginals: $\rho = .441$, $\tau = .342$

Conditional dependence

First message: Understanding the relation among X and Y is complex, and often depends on the way they are measured Things

become even more difficult when

- the number of variables increase
- the intensity of the relation varies according to a set of other variables

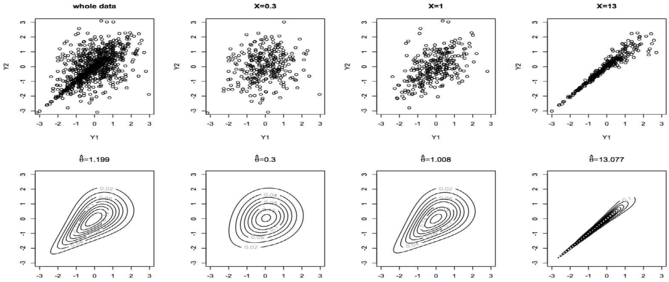
In other words, in the presence of a Dynamic Regime of Dependence

Motivating Application 1: Visual Acuity Data

- Visual acuity is measured on a group of 2810 children together with age
- As children grow, their sight changes
- Bivariate data $z_i = (x_i, y_i) =$ (right visual acuity, left visual acuity) and a covariate $w_i =$ age from each child (unit) indexed by $i = 1, \ldots, n$

Motivating Examples 2: Blood Pressure

It is known that there is a dependence between blood pressure (BP) and body mass index (BMI). What if the dependence varies with subject's age?



...another example in Nina Deliu's talk tomorrow morning

Copulas

- Copula functions are used to model dependence between continuous random variables.
- (Sklar,'59) If Y_1, Y_2 are continuous r.v.'s with distribution functions (df) F_1, F_2, there exists a unique copula function $C : [0,1] \times [0,1] \to [0,1]$, such that

$$F_{12}(t,s) = Pr(Y_1 \leq t, Y_2 \leq s) = C(F_1(t), F_2(s)).$$

- C is a distribution function on $[0,1]^2$ with uniform margins.
- The copula bridges the marginal distributions with the joint distribution.

Conditional Copulas

- The conditional copula of $(Y_1, Y_2)|X = x$ is the conditional joint distribution function of $U = F_{1|X}(Y_1|x)$ and $V = F_{2|X}(Y_2|x)$ given $X = x$ (Patton, Int'l Econ. Rev. 2006).
- Consider a random sample $(x_i, y_{1i}, y_{2i}); 1 \leq i \leq n$ and suppose $F_{1|X}$ and $F_{2|X}$ are the unknown marginal conditional cdf's. The conditional copula model assumes

$$(Y_{1i}, Y_{2i})|X = x_i \sim C(F(Y_{1i}|x_i), F(Y_{2i}|x_i)|\theta(x_i)).$$

- Marginals and copula are conditional on the same variables.

A statistical model

The general expression of the likelihood of a single observation is then

$$f(x,y|z) = g_1(x|z,\theta_x)g_2(y|z,\theta_y)c_Z\left(G_1(x|z,\theta_x), G_2(y|z,\theta_y);\psi\right)$$

- g_1, g_2 are much easier than $c_Z(\cdot,\cdot)$
- Here, our quantity of interest is mainly $c_Z(\cdot,\cdot)$ or a functional T of it, maybe $\tau(c_Z(\cdot,\cdot))$

An Approximate Bayesian alternative

ABC is a computational technique that only requires being able to sample from the likelihood $p(\cdot|\omega)$ This technique stemmed from population genetics models, about 20 years ago (Marjoram, 2003, PNAS).

- A stands for approximate (in several ways . . .)
- B stands for Bayesian
- C stands for computation (it produces a posterior sample)

An Approximate Bayesian alternative

Basic ABC algorithm. Given a sample $\boldsymbol{x}$,

1. Generate θ^* from the prior $\pi(\cdot)$
2. Generate another dataset $\boldsymbol{z} \sim f(\boldsymbol{z}|\theta^*)$ where f is our working likelihood
3. If $\rho\left(\eta(\boldsymbol{z}), \eta(\boldsymbol{x})\right) < \varepsilon$ accept θ^*, otherwise return to step 1.

ρ is a distance; η is a (non)-sufficient statistic.
The vector of accepted values of θ^*'s can be considered as - approximately - generated from $\pi(\theta \mid \boldsymbol{x})$.
This is the basic algorithm. Many improvements can be done, from several perspectives

Empirical likelihood (EL)

Dataset $\boldsymbol{x}$: n independent replicates $\boldsymbol{x} = (x_1, \ldots, x_n)$ of a r.v. $X \sim F$.

Induce a pseudo-model in terms of generalized moment conditions, i.e.

$$E_F\left(h(X, \varphi)\right) = 0,$$

where $h(\cdot)$ is a known function, and φ an unknown parameter.

[Example for the mean of F: $h(X, \varphi) = X_i - \varphi$ ...]

The resulting empirical likelihood is

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

for all $\boldsymbol{p}$ such that $0 \leq p_i \leq 1$, $\sum_{i=1}^{n} p_i = 1$, and

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

Owen, 1988, BKA, & Empirical Likelihood, 2001

BETEL

Bayesian Exponentially Tilted Empirical Likelihood [?][1]
As before, but

$$L_{BEL}(\varphi; x) = \max_{(p_1, \ldots, p_n)} \sum_{i=1}^{n} \left(-p_i \log p_i\right),$$

under constraints

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

[1] BETEL has an interesting Bayesian nonparametric interpretation, since it has the well-known property of the Dirichlet prior used in the Bayesian bootstrap of providing posteriors that assign probability one to distributions supported on the sample

BC_{el}: Bayesian Computation with Empirical Likelihood

Mengersen et al. (PNAS, 2013) use the EL approximation to the true "unknown" likelihood within a Bayesian framework:

1. Pretend (BET)EL is an exact likelihood
2. `for` $i = 1, \dots B$
   - generate $\varphi^{(i)}$ from the prior distribution $\pi(\cdot)$
   - set the weight $w_i = (BET)EL(\varphi^{(i)}; \boldsymbol{x})$

   end for
3. `resample w/r from` $(\varphi^{(i)}, w_i)$, $i = 1, \dots, B$.

Performance evaluated through the so-called effective sample size

$$ESS = 1 \bigg/ \sum_{b=1}^{B} \left\{ w_b \bigg/ \sum_{r=1}^{B} w_r \right\}$$

(... the larger the better ...)
More advanced algorithms can be used.

Mixing the ingredients

Grazian and L. (2017) propose a semi-parametric Bayesian approach for estimating a functional ψ of a multivariate distribution by using the following ingredients:

- Copula representation of a multivariate CDF (Sklar, 1959)
- (BET) Empirical Likelihood of ψ
- Approximate Bayesian Computation method BC_{EL}

Different approaches

Here we discuss three different ways of tackling the conditional problem

- Gaussian Processes (*Levi and Craiu, CSDA, 2018*)
- Conditional BETEL
- B-splines (*Stander, L. et al. 2020, Stat in Med.*)

1. Gaussian processes

To encourage normality, we consider the Fisher transform

$$Z(x) = \log \frac{1-\rho(x)}{1+\rho(x)}$$

and consider its sample version

$$W(x) = \log \frac{1-\hat{\rho}(x)}{1+\hat{\rho}(x)}$$

where $\hat{\rho}(x)$ is the conditional Spearman's ρ restricted to observations at covariate level x

1. Gaussian processes

Assume $Z(x)$ is - *a priori* - a GP

$$Z(x) \sim \mathcal{GP}\left(\mathbf{g}(x)^T\boldsymbol{\beta}, \sigma^2\mathcal{K}(x,x';\xi)\right).$$

Tthe location parameter is $\mathbb{E}[Z(x)] = \mathbf{g}(x)^T\boldsymbol{\beta}$, where $\mathbf{g}(x) = (g_1(x),\ldots,g_q(x))^T$ is a set of known functions, $x \in \mathbb{R}^p$ and $\boldsymbol{\beta} \in \mathbb{R}^q$.
Common choices for the basis function $\mathbf{g}(x)$ are

- $\mathbf{0}$
- $(1,\mathbf{x})$
- $(1,\mathbf{x},\mathbf{x}^2)$

Additional noise

In practice, $W(x)$ is a noisy observation of the signal $Z(x)$.

It is possible to explicitly model the noise through some parametric assumption. For instance, the case of compensating errors can be modelled through a Gaussian distribution

$$W(x_\ell) = Z(x_\ell) + \varepsilon_\ell \qquad \ell = 1, \dots, k$$

where $\varepsilon_\ell \sim \mathcal{N}(0, \tau_\ell^2)$, independently, with $\tau_\ell^2 = \tau^2 / n_\ell$;
then, one gets

$$W(x_\ell) \sim \mathcal{N}\left(\mathbf{g}(x_\ell)^T \beta, \sigma^2 + \tau_\ell^2\right).$$