Approximate Bayesian (un)conditional copula

Clara Grazian

University of New South Wales

c.grazian@unsw.edu.au

Joint work with Luciana Dalla Valle and Brunero Liseo

Algorithms & Computationally Intensive Inference seminar

Warwick University

18 September 2020

Sklar's Theorem

A copula model is a way of representing the joint distribution of a random vector $\mathbf{Y} = (Y_1, \ldots, Y_d)$. Given a d-variate cumulative distribution function F which depends on some parameter λ, it is possible to show [Sklar, 1959] that there always exists an d-variate function $C_\theta : [0,1]^d \to [0,1]$, such that

$$F(y_1, \ldots, y_d; \lambda_1, \cdots, \lambda_d, \theta) = C_\theta(F_1(y_1; \lambda_1), \ldots, F_d(y_d; \lambda_d))$$

where F_j is the marginal CDF of Y_j.

Therefore, in case that the multivariate distribution has a density $\mathbf{f}$, and this is available, it holds further that

$$\mathbf{f}(y_1, \ldots, y_d) = c(F_1(y_1), \ldots, F_d(y_d)) \cdot f_1(y_1) \cdots f_d(y_d)$$

Example of copula functions

- Clayton copula: $C_\theta(u,v) = [\max\{u^{-\theta} + v^{-\theta} - 1; 0\}]^{-\frac{1}{\theta}}$ for $\theta \in [-1, 1)$
- Gumbel copula: $C_\theta(u,v) = \exp[-((-\log(u)^\theta + (-\log(v))^\theta))^{\frac{1}{\theta}}]$ for $\theta \in [1, \infty]$
- Frank copula $C_\theta(u,v) = -\frac{1}{\theta}\log\left[1 + \frac{(\exp(-\theta u)-1)(\exp(-\theta v)-1)}{\exp(-\theta)-1}\right]$ for $\theta \in \mathbb{R} \setminus 0$

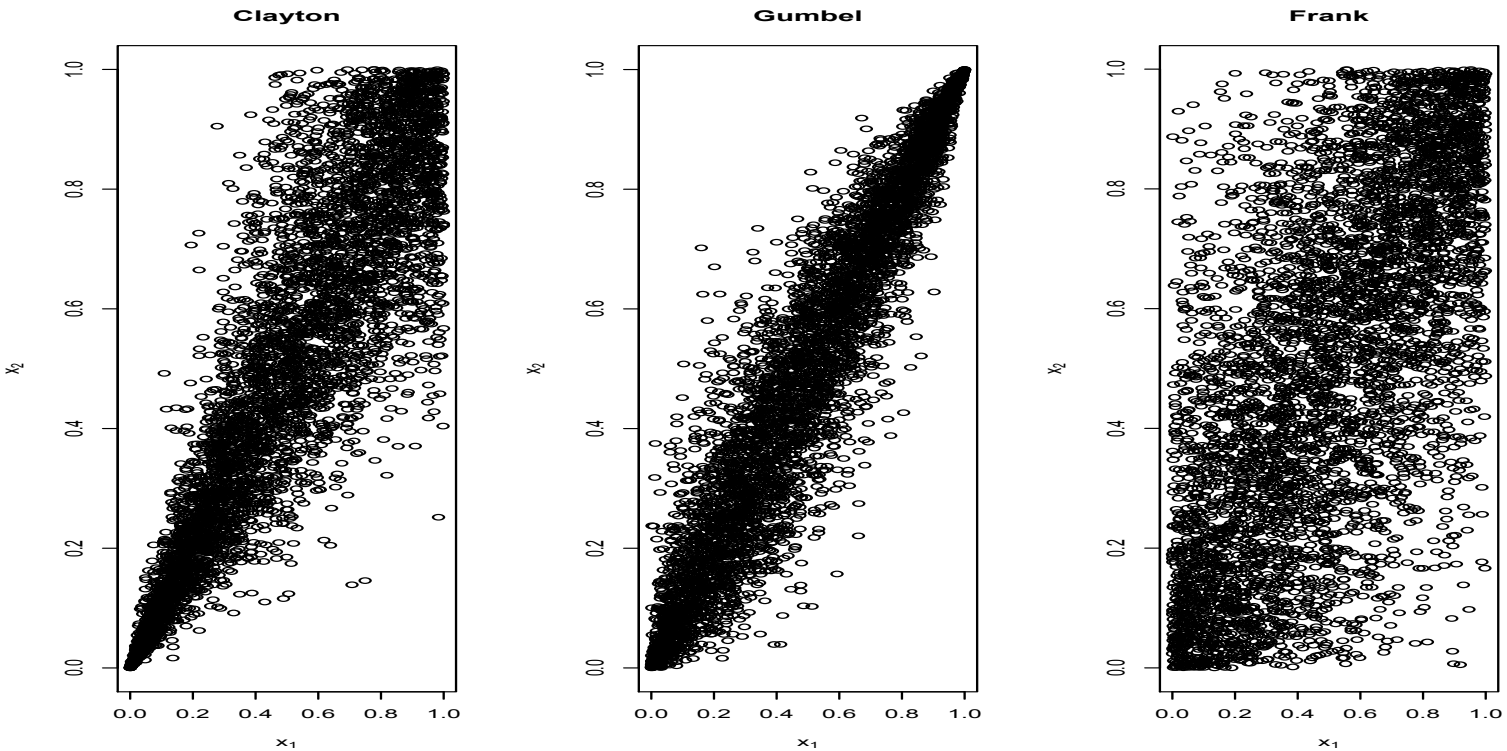


Figure: Simulations from different copula functions

Estimation methods for copula models

- **Frequentist methods:**
  - *Inference from the margins* [Joe, 2015]
  - Method of moments [Oh and Patton, 2013]
- **Bayesian methods:** [Smith, 2013]
  - multivariate discrete data [Smith and Khaled, 2012]
  - conditional copulae [Craiu and Sabeti, 2012]
  - vine-copulae [Min and Czado, 2010]

Likelihood function

The likelihood function is complicated!

Example: Clayton copula

$$Ł(\theta; u, v) = \prod_{i=1}^{n} (\theta+1)(u_i v_i)^{-(\theta+1)} (u_i^{-\theta} + u_i^{-\theta} - 1)^{-\frac{2\theta+1}{\theta}}$$

If the interest is in a functional of the dependence, the likelihood function is even more complicated, because the relationship between θ and the particular function may be difficult or analytically unavailable.

Example: Clayton copula

$$\tau = \frac{\theta}{\theta+2} \qquad \textit{and} \qquad \lambda_L = 2^{-\frac{1}{\theta}} \qquad \textit{but} \qquad \rho = \cdots$$

Approximate Bayesian computation

There is a full set of algorithms under the class called "approximate Bayesian computation" (ABC):

- rejection ABC: simulate parameters from the prior distribution and discard all the values for which $\rho(\eta(y), \eta(y_{sim})) > \varepsilon$ [Marin et al., 2016]
- approximating the likelihood by approximating the distribution of the distance measure $\hat{L}_{\mathscr{K}} = \hat{\Pr}(\Delta(\eta_0, \eta_{sim}) \leq \varepsilon)$ [Gutmann and Corander, 2016]
- approximating the likelihood: synthetic likelihood [Price et al., 2018], empirical likelihood [Mengersen et al., 2013]

Empirical Likelihood

Empirical likelihood is a way of producing a nonparametric likelihood for a quantity of interest [Owen, 2001]. Schennach [2005] proposes a **Bayesian exponentially tilted empirical likelihood**.

Consider a given set of generalized **moment conditions**

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

where $h(\cdot)$ is a known function, and φ is the quantity of interest.

$L_{BEL}(\varphi; x)$ is defined as the solution of

$$\max_{(p_1,\ldots,p_n)} \sum_{i=1}^{n} (-p_i \log p_i)$$

under constraints

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

The Bayesian use of the empirical likelihood I

We are interested in a function φ and in its posterior

$$\pi(\phi|y) \propto \int_N p(y|\nu,\phi)\pi(\nu|\phi)\pi(\phi)d\nu$$

or

$$\pi(\phi|y) \propto \lim_{N\to\infty} \int_N p(y|\xi_N,\phi)\pi(\xi_N|\phi)\pi(\phi)d\xi_N$$

Then the distribution C can be represented as

$$C = (\varphi, C^*)$$

where C^* belongs to an infinite dimensional metric space (H, d_H).
L_{BEL} may be seen as the derivation of the integrated likelihood for ϕ

$$L_{BEL}^{(\lambda)}(\phi;y) = \int_{\Xi} L(\phi,\xi;y)d\Pi(\xi)$$

where $\Pi(\xi)$ is the prior process implicitly induced by L_{BEL}.

ABSCop algorithm

Goal: *estimating a functional of the dependence (Spearman's ρ, Kendall's τ, tail dependence coefficients λ_L and λ_U, etc.)*

- Select a quantity of interest ϕ and a prior $\pi(\phi)$

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

  with $\pi(\rho) \sim \mathscr{U}(-1,1)$.

- Select a (nonparametric) estimators ϕ_n

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

- Compute the empirical likelihood of ϕ based on its estimate
- Derive via simulation the posterior distribution $\pi(\phi; \mathbf{x})$

Conditional Copulas

Patton [2006] extends the definition of copula function in the presence of covariates, such that:

$$F(Y_1, \ldots, Y_p \mid X) = C_x(F_{1|x}(Y_1 \mid X), \ldots, F_{p|X}(Y_p \mid X))$$

where $X \in \mathbb{R}^q$, $C_x(\cdot)$ is the conditional copula and $F_{j|X}$ is the conditional cdf of Y_j.

Estimation procedures:

- introducing a parametric model for the copula and assuming that the parameter of the copula varies as a function of X, such that $C_x = C_{\theta(x)}$
  - parametrically [Genest et al., 1995]
  - nonparametrically [Craiu and Sabeti, 2012]
- Simplifying Assumption [Gijbels et al., 2015]
- nonparametric estimate of the conditional copulas [Dalla Valle et al., 2018]

Conditional measures

Since the above dependence indices can be directly defined through their copula function, the conditional versions of τ and ρ can be easily derived in terms of conditional copulas:

$$\tau(x) = 4 \int \int_{[0,1]^2} C_x(u,v) dC_x(u,v) - 1, \quad \rho(x) = 12 \int \int_{[0,1]^2} C_x(u,v) du dv - 3.$$

The most common estimators of measures of conditional dependence are expressed in terms of estimators of the conditional copula

$$C_{x;h}(y_1, y_2) = \sum_{i=1}^{n} w_{n;i}(x, h_n) \mathbb{I}[Y_1 \leq y_1, Y_2 \leq y_2]$$

where $\{w_{ni}(x, h_n)\}$ is a sequence of weights that smooth over the covariate space (for example, the Nadaraya-Watson or the local-linear weights) and $h_n > 0$ is a bandwidth which is assumed to vanish as the sample size increases.

How to choose the weights?

Common choices of the weights are

- Nadaraya-Watson

$$w_{ni}(x, h_n) = \frac{\mathscr{K}\left(\frac{X_i - x}{h_n}\right)}{\sum_j \mathscr{K}\left(\frac{X_j - x}{h_n}\right)}$$

- Local linear

$$w_{ni}(x, h_n) = \frac{\frac{1}{nh_n}\mathscr{K}\left(\frac{X_i - x}{h_n}\right)\left(S_{n,2} - \frac{X_i - x}{h_n} S_{n,1}\right)}{S_{n,0}S_{n,2} - S_{n,1}^2}$$

where $S_{n,j} = \frac{1}{nh_n}\sum_{i=1}^{n}\left(\frac{X_i - x}{h_n}\right)^j \mathscr{K}\left(\frac{X_i - x}{h_n}\right)$

How to choose the kernel?

Common choices of the kernel are

- triweight

$$\mathscr{K}(x) = \frac{35}{32}(1-x^2)^3 \mathbb{I}(|x|<1)$$

- Gaussian

$$\mathscr{K}(x) = \frac{\exp\left(-\frac{x^2}{2}\right)}{\sqrt{2\pi}}$$

Pitfall of the estimator of the conditional copula

Although the copula estimator C_{xh} just extends the unconditional copula estimator in presence of covariates, there are some disadvantages

- Suppose that Y_1 and Y_2 are conditionally independent given $X = x$, but that their ocnditional distributions are both stochastically increasing with x. $\longrightarrow$ Larger values of Y_1 will occur together with larger values of Y_2 only because of the same trend in the covariate x, creating an artificial dependence

Pitfall of the estimator of the conditional copula

Although the copula estimator C_{xh} just extends the unconditional copula estimator in presence of covariates, there are some disadvantages

- Suppose that Y_1 and Y_2 are conditionally independent given $X = x$, but that their ocnditional distributions are both stochastically increasing with x. $\longrightarrow$ Larger values of Y_1 will occur together with larger values of Y_2 only because of the same trend in the covariate x, creating an artificial dependence
- The estimator is biased!

Pitfall of the estimator of the conditional copula

Although the copula estimator C_{xh} just extends the unconditional copula estimator in presence of covariates, there are some disadvantages

- Suppose that Y_1 and Y_2 are conditionally independent given $X = x$, but that their ocnditional distributions are both stochastically increasing with x. $\longrightarrow$ Larger values of Y_1 will occur together with larger values of Y_2 only because of the same trend in the covariate x, creating an artificial dependence
- The estimator is biased!
- Gijbels, Veraverbeke, and Omelka (2010) proposed a new estimator which is still biased by can reduce the bias in some situations.

Conditional Spearman's ρ

Suppose that, for $h = 1, \ldots, k$, we know the values of a covariate $x_1, x_2, \ldots, x_k$.

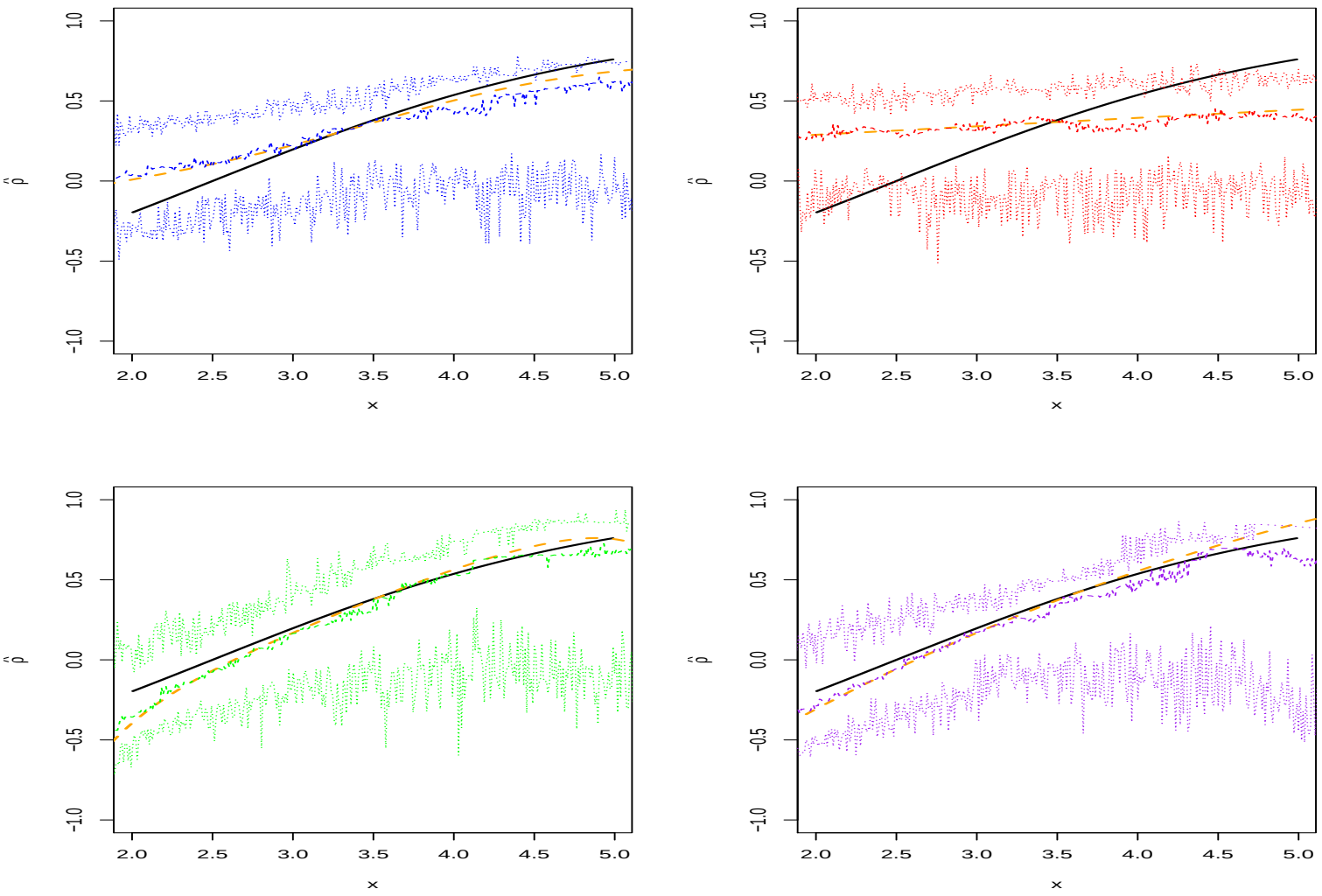
For each $h = 1, \ldots, k$ we also know n_h replications of $\{U, V | X = x_h\}$

$$[(u_{h,1}, v_{h,1}), \ldots, (u_{h,n_h}, v_{h,n_h})],$$

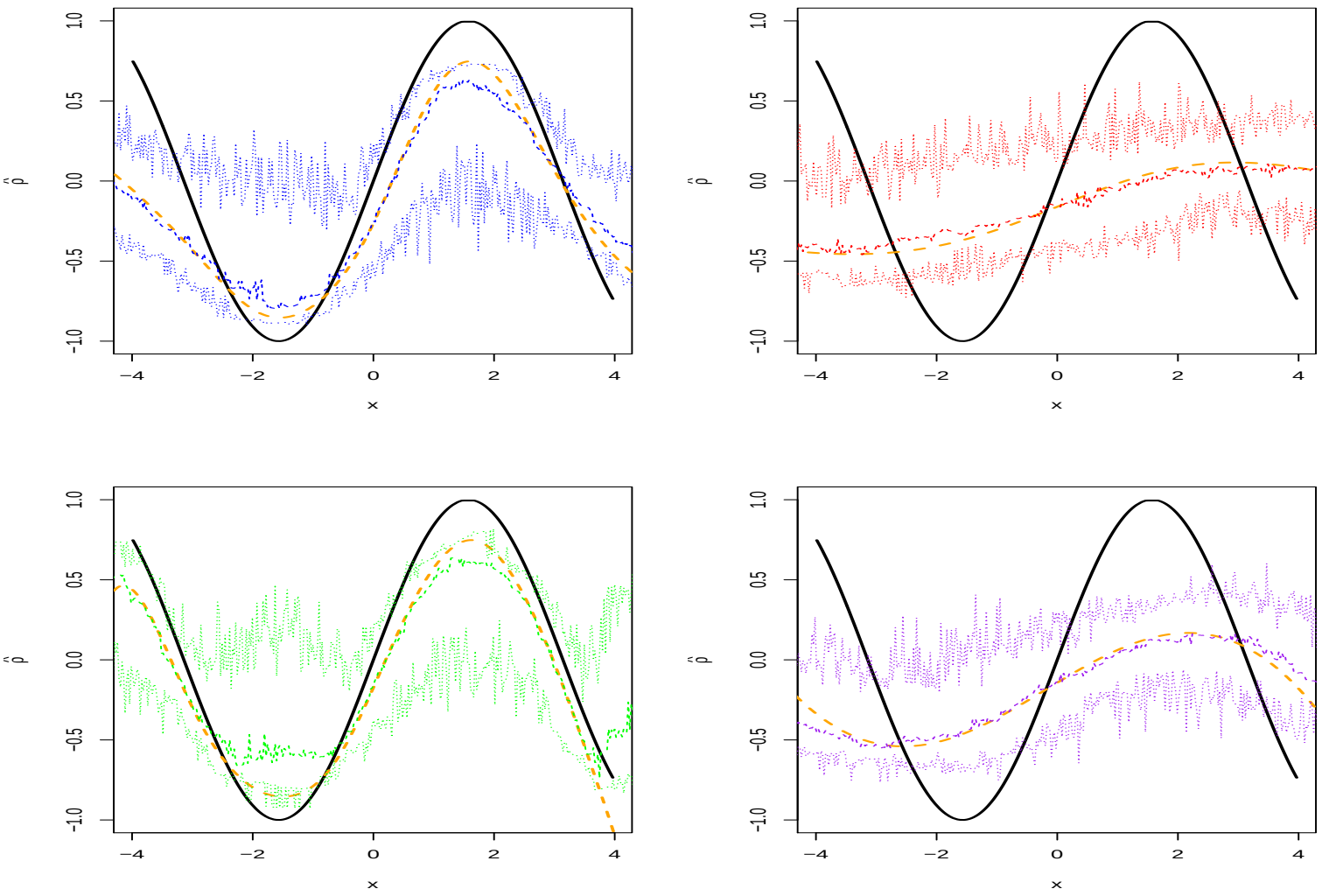
such that $\sum_{h=1}^{k} n_h = n$.

Our goal is to estimate $\rho(u, v|x) = \rho(x)$ where ρ is the conditional Spearman's index.

ABSCop for $\rho(x)$ as a linear function



ABSCop for $\rho(x)$ as a sin function



Gaussian Processes

To encourage normality, we consider

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

and we observe its sample version

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

where $\hat{\rho}(x)$ is a the unconditional Spearman's ρ computed for repetitions of observations at covariate level x

Pior distributions

First, we assume that $Z(x)$ follows - *a priori* - a Gaussian process (GP)

$$Z(x) \sim GP\left(\psi(x)^T\theta, \sigma^2\mathscr{K}(x,x';\xi)\right).$$

The location parameter of the Gaussian process is

$$\mathbb{E}[Z(x)] = \psi(x)^T\theta,$$

where $\psi(x) = (\psi_1(x), \ldots, \psi_q(x))^T$ is a set of known functions, $x \in \mathbb{R}^p$ and $\theta \in \mathbb{R}^q$.

Common choices for the basis function $\psi(x)$ are

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

Prior distributions

$\mathscr{K}(x, x'; \xi)$ is a generic correlation kernel depending on a parameter ξ, and σ^2 is a positive scale parameter, so that

$$\mathbb{C}ov(Z(x), Z(x')) = \sigma^2 \mathscr{K}(x, x'; \xi).$$

Without loss of generality, we will consider the squared exponential kernel

$$\sigma^2 \mathscr{K}(x, x'; \xi) = \sigma^2 \exp\left(-\frac{1}{2}\sum_{j=1}^{p} \frac{d(x_j, x_j')}{\xi_j}\right) = \sigma^2 \exp\left(-\frac{1}{2}\sum_{j=1}^{p} \frac{(x_j - x_j')^2}{\xi_j}\right).$$

Additional noise

We assume that each set of observations generates a noisy version of $Z(x)$, say $W(x)$, which depends on a statistic evaluated at location x; in practice $W(x)$ is a noisy observation of the signal $Z(x)$.

It is possible to account for the assumption on the type of error through a model. For instance, the case of compensating errors can be modelled through a Gaussian distribution

$$W(x_h) = Z(x_h) + \varepsilon_h \qquad h = 1, \ldots, k$$

where $\varepsilon_h \sim \mathscr{N}(0, \tau_h^2)$, with $\tau_h^2 = \frac{\tau^2}{n_h}$ and $\varepsilon_h \perp \varepsilon_{h'}$.
Therefore, the observations follow a normal distribution

$$W(x_h) \sim \mathscr{N}\left(\psi(x_h)^T \theta, \sigma^2 + \tau_h^2\right).$$

The likelihood is available

It follows that the likelihood associated to the observations related to locations $x_1, \ldots, x_k$, with a number of observations $n_1, \ldots, n_k$ each, is

$$L(\theta, \sigma^2, \xi, \tau^2) = \mathscr{N}\left(\psi(\mathbf{x})^{\mathsf{T}}\theta, \sigma^2\boldsymbol{\Sigma} + \tau^2\tilde{\mathbf{I}}\right)$$

where $\tilde{\mathbf{I}} = \text{diag}\left(\frac{1}{n_1}, \ldots, \frac{1}{n_k}\right)$, (indicating a diagonal matrix with element $\left(\frac{1}{n_1}, \ldots, \frac{1}{n_k}\right)$ on the diagonal) and

$$\boldsymbol{\Sigma} = \begin{pmatrix} \mathscr{K}(x_1, x_1) & \cdots & \mathscr{K}(x_1, x_k) \\ \vdots & \ddots & \vdots \\ \mathscr{K}(x_k, x_1) & \cdots & \mathscr{K}(x_k, x_k) \end{pmatrix}.$$

Reparametrization

Let's set

$$\sigma^2\Sigma + \tau^2\tilde{I} = \sigma^2\left(\Sigma + \frac{\tau^2}{\sigma^2}\tilde{I}\right) = \sigma^2\left(\Sigma + \lambda\tilde{I}\right) = \sigma^2 M,$$

where M depends on ξ and λ.
The likelihood could then be written as

$$L(\theta, \sigma^2, \lambda, \xi) = \frac{1}{\sigma^k |M|^{1/2}} \exp\left(-\frac{1}{2\sigma^2}(W - X\theta)^T M^{-1}(W - X\theta)\right),$$

where $X = \psi(\mathbf{x})^{\mathbf{T}}$ is a $k \times q$ matrix of known constants.
Integrating L with respect to θ (using $\pi(\theta) \propto 1$) and with respect to σ^2 (using a inverse gamma prior $IG(\alpha, r)$) we obtain the integrated likelihood

$$L^I(\xi, \lambda) \propto |M|^{-1/2} |X^T M^{-1} X|^{-1/2} \frac{1}{\left(\tilde{S}_\xi^2 + r\right)^{\frac{n-q}{2} + \alpha}}$$

Posterior distribution

The posterior sample of (ξ, λ) is then used to approximate the posterior predictive distribution of W at new locations x^*

$$\begin{aligned} f(w^*|x^*, \mathbf{W}) &= \int_{\xi,\lambda} f(w^*, \xi, \lambda | x^*, \mathbf{W})\, d\xi\, d\lambda \\ &= \int_{\xi,\lambda} f(w^*|x^*, \xi, \lambda)\, f(\xi, \lambda | \mathbf{W}) d\xi\, d\lambda \end{aligned}$$

This approximation can be computed for each value x^* defined over a grid. A natural summary value for the predictive distribution of W^*at w is the expected value $\mathbb{E}(W|x^*) \approx \bar{W}^*(x^*) = \sum w^* f(w^*|x^*)$.
Finally, the estimation of the functional of interest φ can be obtained as

$$\hat{\varphi}(x^*) = \frac{\exp\{2\bar{W}^*(x^*)\} - 1}{\exp\{2\bar{W}^*(x^*)\} + 1}.$$

Gaussian Processes - Simulation study

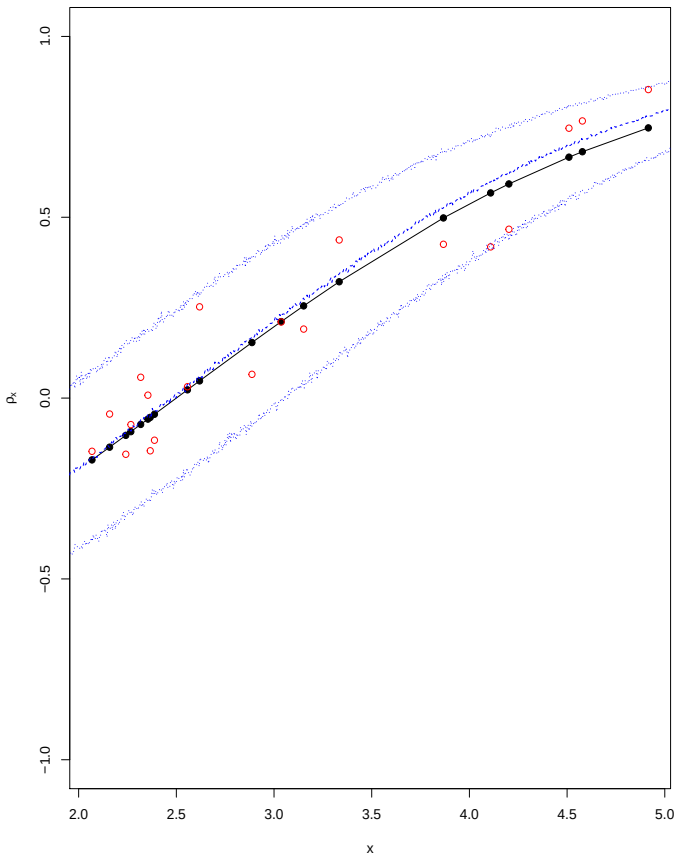


Figure: $\rho(x)$ as a linear function

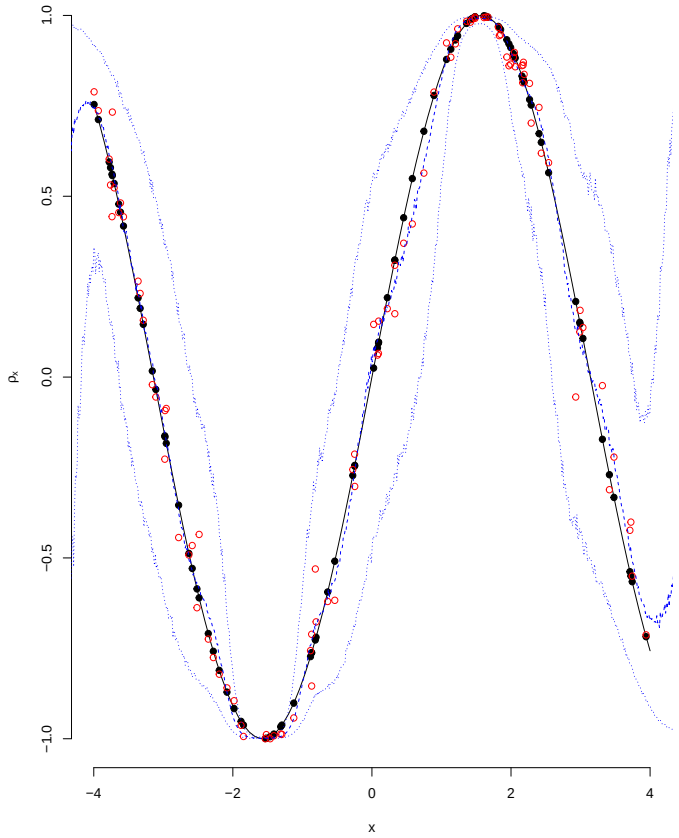


Figure: $\rho(x)$ as a sin function

Empirical likelihood approach

The assumption of normality of the Fisher's transform of the copula functionals may be too strict.

A fully nonparametric approach can also be implemented by using the empirical likelihood approximation, along the line of Grazian and Liseo (2017).

The Fisher's transform of the observations can be defined as a function of the covariates, through a Taylor's expansion in terms of a polynomial of degree ℓ. Assume that $W(\cdot)$ is differentiable ℓ times in the neighbourhood of an interior point x_0; then

$$W(x_h) \approx W(x_0) + W(x_0)'(x_h - x_0) + \ldots + \frac{W(x_0)^{(\ell)}}{p!}(x_h - x_0)^p \equiv \mathbf{x}_{h,x_0}^{*T}\beta$$

where $\mathbf{x}_{h,x_0}^{*T} = (1, (x_h - x_0), \ldots, (x_h - x_0)^p)$ and $\beta = (\beta_0, \beta_1, \ldots, \beta_\ell)$.

Empirical likelihood approach

Alternatively, it is possible to use a spline approximation, along the line of Craiu and Sabeti (2012), using a cubic spline as a model for the calibration function of the copula parameter

$$W(x_h) \approx \sum_{j=1}^{3} \alpha_j x_h^j + \sum_{s=1}^{S} \psi_s (x_h - \gamma_s)_+^3$$

where $a_+ = \max(0, a)$ and $\{\gamma_s\}_{s=1}^S$ is a set of knots.

It is then possible to derive consistent estimators of the coefficient of the spline function which can be used to define the moment condition in the definition of the empirical likelihood.

Empirical likelihood approach

For the parameters β it is common to define weakly informative priors, for instance, $\mathcal{N}(0,100)$, and combine them with the weights defined by the empirical likelihood approach described in Grazian and Liseo (2017) in order to obtain a sample of size G of $(\beta_0,\ldots,\beta_\ell|w_1,\ldots,w_k)$ such that

$$\begin{gathered}\beta_0^{(1)},\ldots,\beta_\ell^{(1)},\boldsymbol{\omega}^{(1)}\\ \vdots\\ \beta_0^{(G)},\ldots,\beta_\ell^{(G)},\boldsymbol{\omega}^{(G)}\end{gathered}$$

where ω_j, with $j=1,\ldots,G$ represents the weights of $(\beta_0^{(j)},\ldots,\beta_\ell^{(j)})$ induced by the empirical likelihood.

Empirical likelihood approach

The approximate posterior distribution β can be used to approximate the posterior predictive distribution

$$\begin{aligned} f\left(w^*|x^*,\mathbf{W}\right) &= \int_{\mathbf{B}} f\left(w^*,\beta|x^*,\mathbf{W}\right) d\beta \\ &= \int_{\mathbf{B}} f\left(w^*|x^*,\beta\right) f(\beta|\mathbf{W}) d\beta \\ &\approx \sum_{g=1}^{G} f\left(w^*|x^*,\beta^{(g)}\right) \bar{\omega}^{(g)} \end{aligned}$$

where $\bar{\omega}^{(g)} = \frac{\omega^{(g)}}{\sum_{g=1}^{G} \omega^{(g)}}$ is the normalized weight for the g-th iteration.

Empirical likelihood - Simulation study

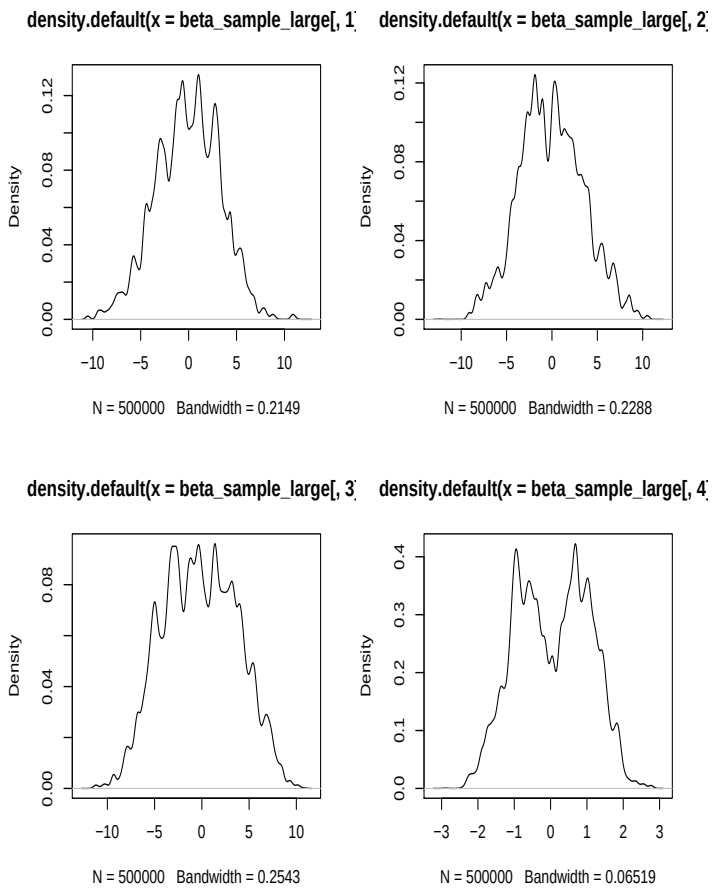


Figure: $\rho(x)$ as a linear function

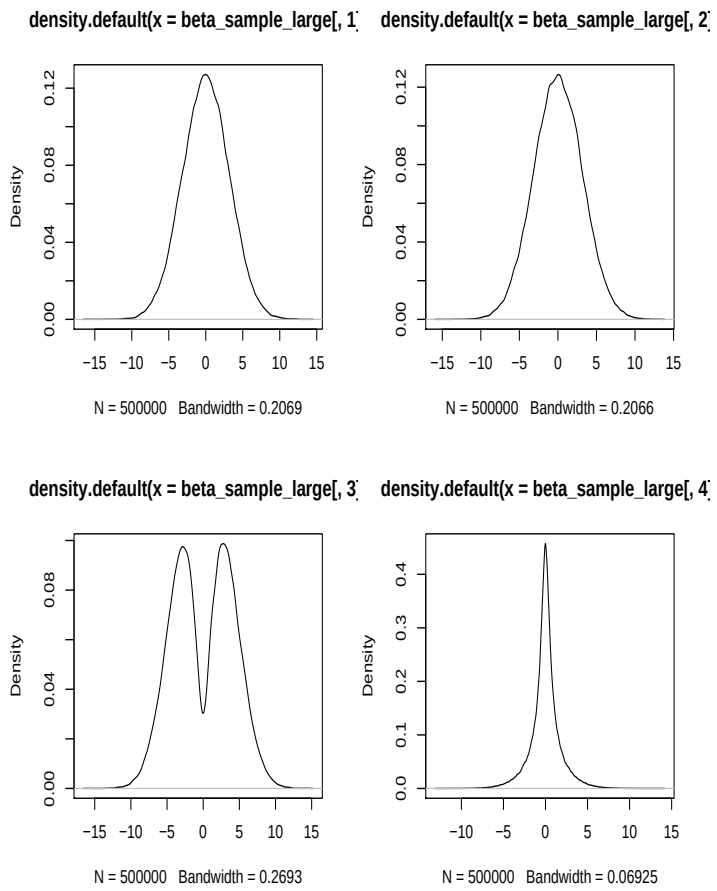


Figure: $\rho(x)$ as a sin function

A last consideration

Choosing the copula is difficult!

Clayton	Frank	Gumbel	Gaussian
0.600	0.250	0.316	0.200

Table: Relative frequencies of correct selection of the copula model using leaving-one-out cross-validation.

Energy performance of buildings

Recently, there is more and more concern about the energy waste and its impact on environment.

In efficient building design, two variables are considered to determine the specifications of heating and cooling equipment: heating load (HL, Y_1) and cooling load (CL, Y_2).

In order to forecast building energy consumption, several covariates are considered.

We focus out attention on dependence (through the Kendall's τ), between HL and CL when varying the **relative compactness (RC)**.

The morphology and the shape of a building are two important factors that could influence the increase/decrease of energy required to heat or coll a space. The relative compactness of a shape is defined by comparing its volume to surface ratio to that of the most compact shape with the same volume.

Energy data

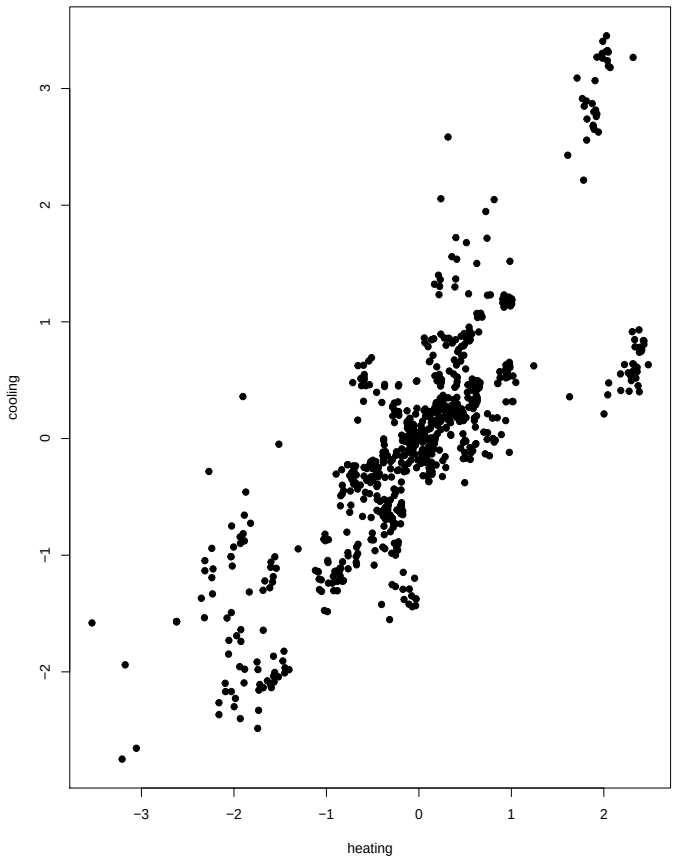


Figure: Scatterplot of heating and cooling loads.

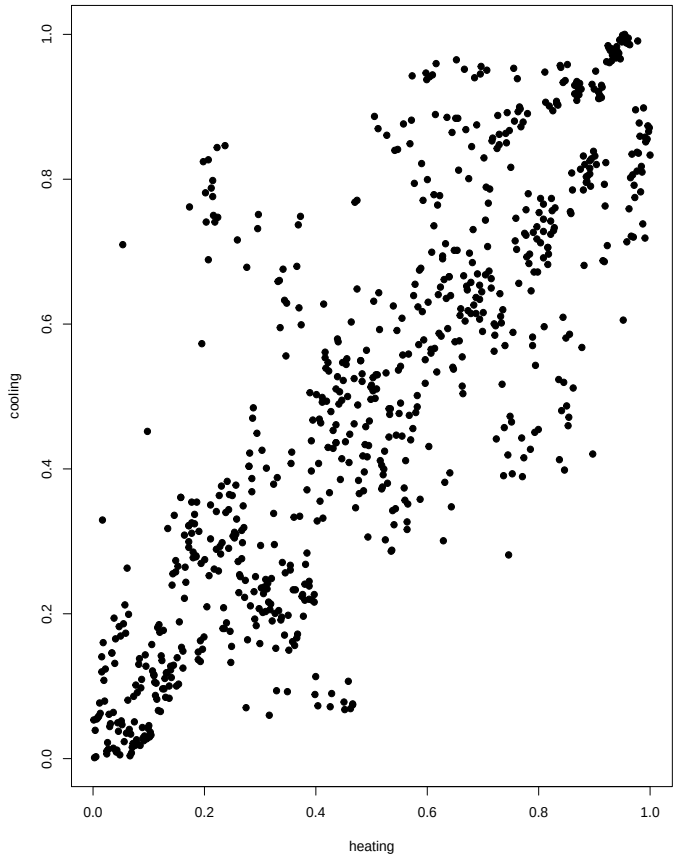


Figure: Pseudo-data

Energy data - Results

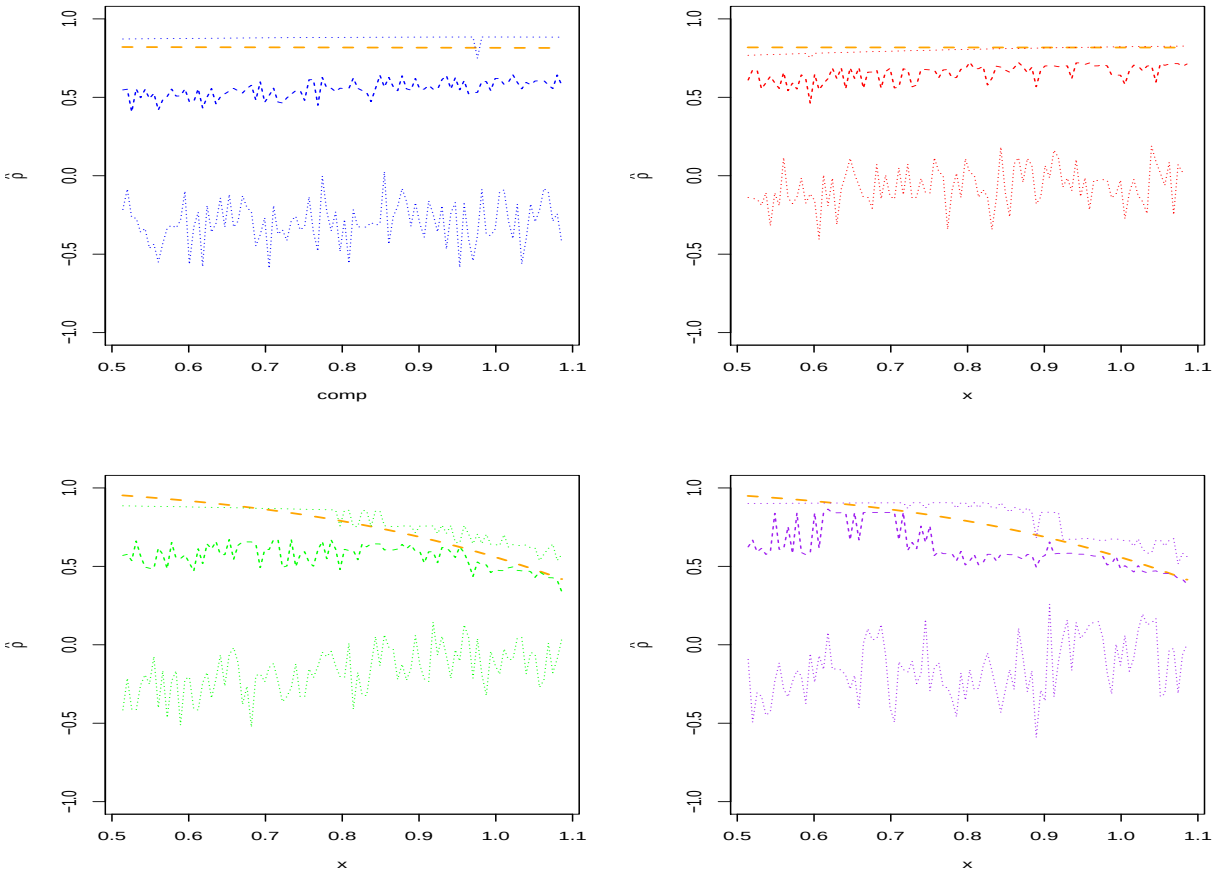


Figure: Estimation of the ρ with ABSCop

Energy data - Results

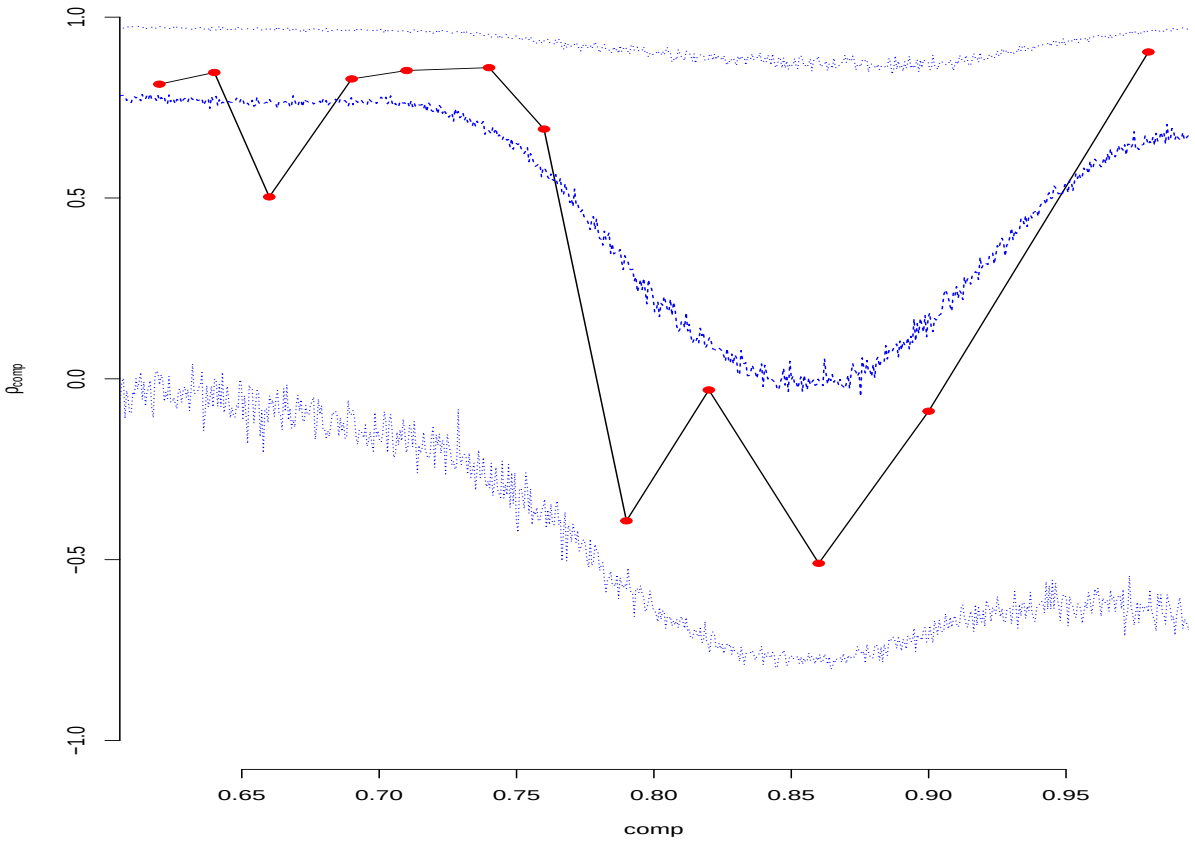


Figure: Estimation of the ρ with GP

Energy data - Results

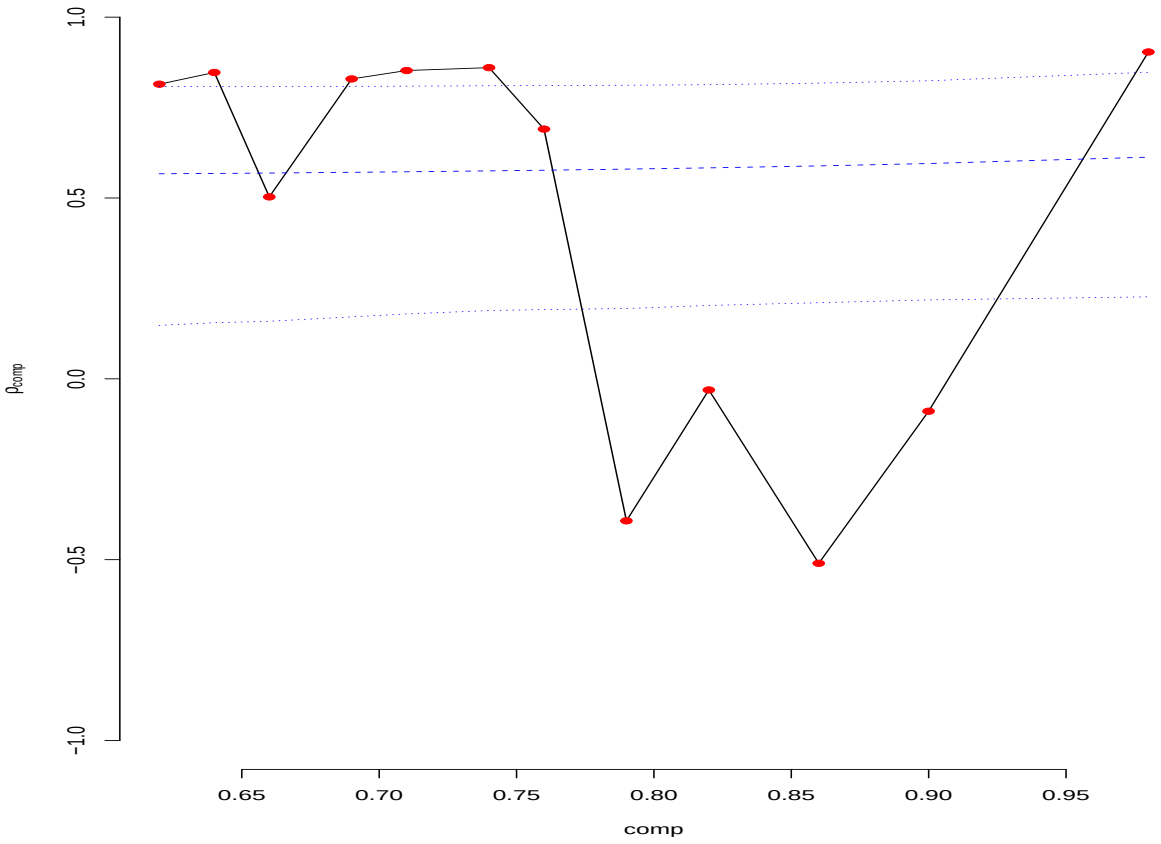


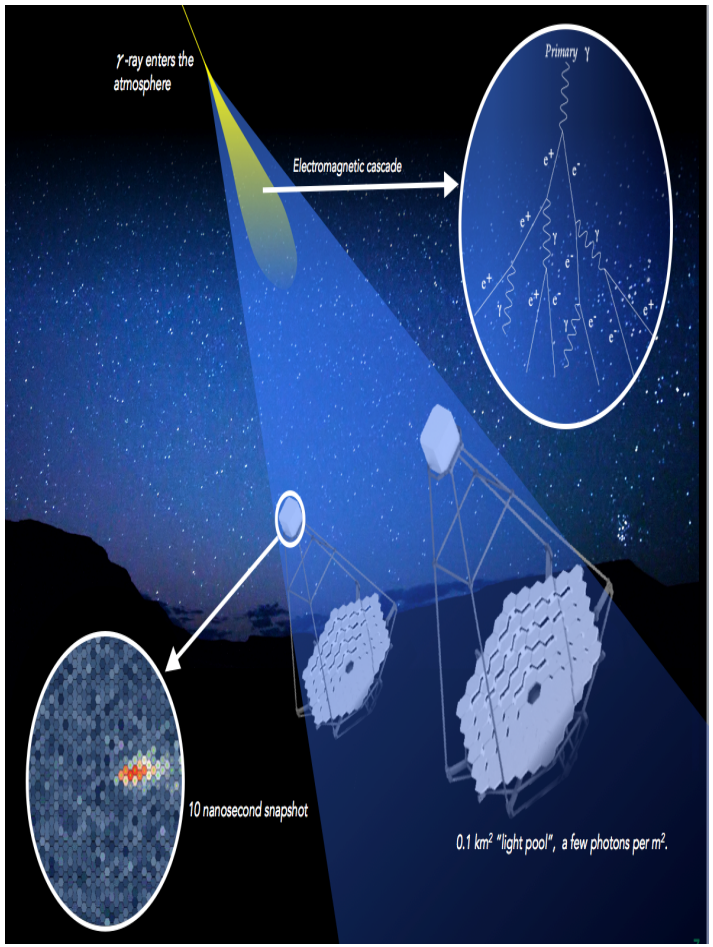
Figure: Estimation of the ρ with splines

The MAGIC dataset

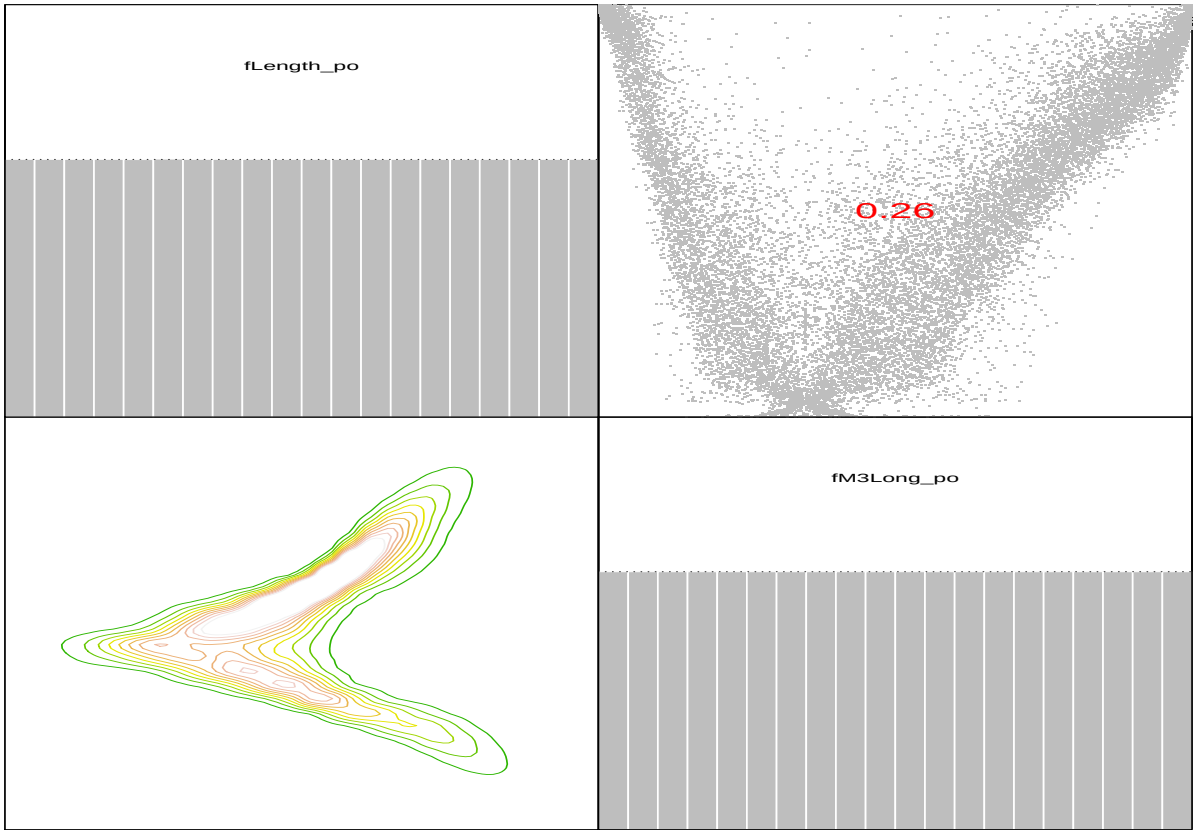
The data are MC generated to simulate registration of high energy gamma particles in a ground-based atmospheric **Cherenkov gamma telescope**.

Cherenkov gamma telescope observes **high energy gamma rays**, taking advantage of the radiation emitted by charged particles produced inside the electromagnetic showers initiated by the gammas.

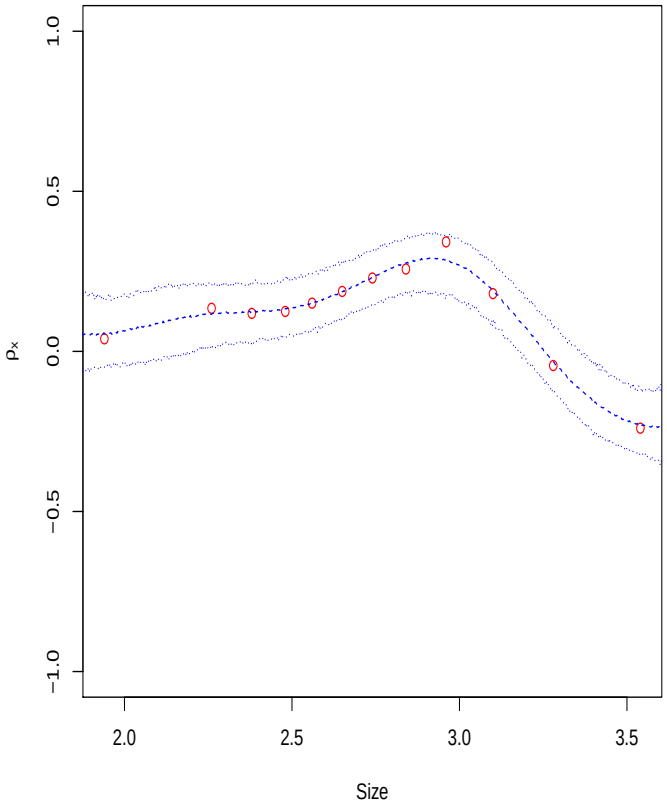
The goal is to discriminate statistically those caused by primary gammas (signal) from the images of hadronic showers initiated by cosmic rays in the upper atmosphere (background).

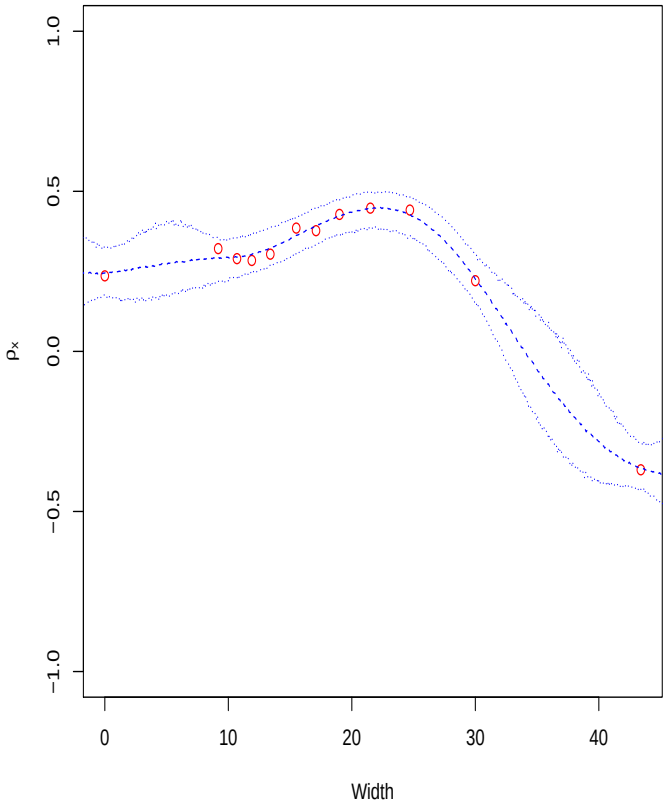


The MAGIC dataset



The MAGIC dataset





Conclusions

The proposed approaches has the following advantages:

- it is nonparametric in terms of joint distribution
- it may be easily extended to a general dimension d (of the observations or the covariates)
- it may be generalized to any functional of the dependence
- it may consider both unconditional and conditional distributions

However,

- the choice of the moment conditions is essential for ABSCop
- the implementation of GP needs replications

Bibliography I

Craiu, V. R. and Sabeti, A. In mixed company: Bayesian inference for bivariate conditional copula models with discrete and continuous outcomes. *J. Multivariate Anal.*, 110: 106–120, 2012.

Dalla Valle, L., Leisen, F., and Rossini, L. Bayesian nonparametric conditional copula estimation of twin data. *Journal of the Royal Statistical Society: Series C*, 67(3), 523-548, 2018.

Genest, C., Ghoudi, K., and Rivest, L. P. A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika*, 82(3), 543-552, 1995.

Gijbels, I., Omelka, M., and Veraverbeke, N. Estimation of a copula when a covariate affects only marginal distributions. *Scandinavian Journal of Statistics*, 42(4), 1109-1126, 2015.

C. Grazian, and B. Liseo. Approximate Bayesian inference in semiparametric copula models. *Bayesian Analysis*, 12(4), 991-1016, 2017.

Gutmann, M. U., and Corander, J. Bayesian optimization for likelihood-free inference of simulator-based statistical models. *The Journal of Machine Learning Research*, 17(1), 4256-4302, 2016.

Bibliography II

H. Joe. *Dependence modeling with copulas*, volume 134 of *Monographs on Statistics and Applied Probability*. CRC Press, Boca Raton, FL, 2015.

Marin, J. M., Pudlo, P., Robert, C. P., and Ryder, R. J. Approximate Bayesian computational methods. *Statistics and Computing*, 22(6), 1167-1180, 2016.

K. Mengersen, P. Pudlo, and C.P. Robert. Bayesian computation via empirical likelihood. *Proc. of the National Academy of Sciences*, 4(110):1321–1326, 2013.

Min, A. and Czado, C. Bayesian inference for multivariate copulas using pair-copula constructions. *Journal of Financial Econometrics*, 8(4): 511–546, 2010.

Oh, D. H. and Patton, A. J. Simulated method of moments estimation for copula-based multivariate models. *Journal of the American Statistical Association*, 108(502): 689–700, 2013

Owen, A. B. *Empirical likelihood*. CRC press, 2001.

Patton, A. J. Modelling asymmetric exchange rate dependence. *International economic review*, 247(2), 527-556, 2006.

Paulo, R. Default priors for Gaussian processes. *The Annals of Statistics*, 33(2), 556-582.

Price, L. F., Drovandi, C. C., Lee, A., and Nott, D. J. Bayesian synthetic likelihood. *Journal of Computational and Graphical Statistics*, 27(1), 1-11, 2018.

Bibliography III

Schennach, S. M. Bayesian exponentially tilted empirical likelihood. *Biometrika*, 92(1): 31–46, 2005.

M. Sklar. Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris*, 8:229–231, 1959.

Smith, M. S. (2013). Bayesian approaches to copula modelling. In *Bayesian theory and applications*, 336–358. Oxford Univ. Press, Oxford, 2013.

Smith, M. and Khaled, M. Estimation of copula models with discrete margins via Bayesian data augmentation. *Journal of the American Statistical Association*, 107(497): 290–303, 2012