Statistical Elucidation of Latent Structures via Copulas

Radu Craiu

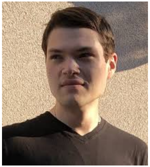
Department of Statistical Sciences
University of Toronto

MEMOTEF, Sapienza - April 7, 2025

Outline

Copulas for serially correlated data with hidden structures

This project was done in collaboration with



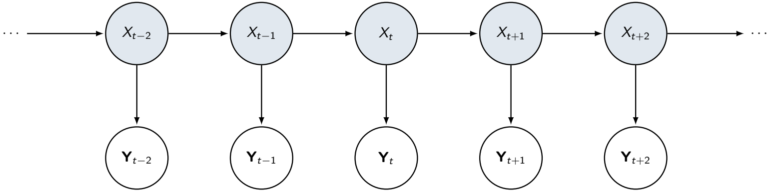
Robert Zimmerman (Imperial College London)

Vianey Leos Barajas (Toronto)

Paper: Copula Modelling of Serially Correlated Multivariate Data with Hidden Structures (JASA, 2024).

Hidden Markov Models: Briefly put

- A hidden Markov model (HMM) pairs an observed time series $\{\mathbf{Y}_t\}_{t\geq 1} \subseteq \mathbb{R}^d$ with a Markov chain $\{X_t\}_{t\geq 1}$ on some state space $\mathcal{X}$, such that the distribution of $\mathbf{Y}_s \mid X_s$ is independent of $\mathbf{Y}_t \mid X_t$ for $s \neq t$:



- $\mathbf{Y}_{t,h}|\{X_t = k\} \sim f_{k,h}(\cdot|\lambda_{k,h})\ \forall h = 1,\ldots,d$
- $\{X_t\}$ is a Markov process (finite state space $\mathcal{X}$) with initial probability mass distribution $\{\pi_i\}_{i\in\mathcal{X}}$ and transition probabilities $\{\gamma_{i,j}\}_{i,j\in\mathcal{X}}$

Inferential aims for HMMs

- Typically, the chain $\{X_t\}_{t\geq 1}$ is partially or completely unobserved.
- The hidden states can correspond to a precise variable (occupancy data) or might be postulated (psychology, ecology, etc)
- Aim 1: Model the data generating mechanism Nasri et al. (2020)
- Aim 2: Decode (i.e., classify) or predict the X_t's from the observed data.

Fusion of Multiple Data Sources

- In real-world applications (sports, stock exchange, animal movement, etc), various sensors capture multiple streams of data, which are "fused" into a multivariate time series $\{\mathbf{Y}_t\}_{t\geq 1}$
- In such situations, the components of any $\mathbf{Y}_t = (Y_{t,1}, \ldots, Y_{t,d})$ cannot be assumed independent (even conditional on X_t)
- The corresponding assumption for HMMs – that of contemporaneous conditional independence Zucchini et al. (2017) – is often violated
- Instead, it is common to assume that $\mathbf{Y}_t$ follows a multivariate Gaussian distribution, but this places limits on marginals and dependence structures
- What if the strength of dependence between the components of $\mathbf{Y}_t$ could be informative about the underlying state X_t?

Occupancy Data

- The ability to detect whether a room is occupied using sensor data (such as temperature and CO_2 levels)
- Consider three publicly-available labelled datasets presented by Candanedo and Feldheim (2016) which contain multivariate time series of four environmental measurements (light, temperature, humidity, CO_2) and one derived metric (the humidity ratio), as well as binary indicators for whether the room was occupied or not at the time of measurement

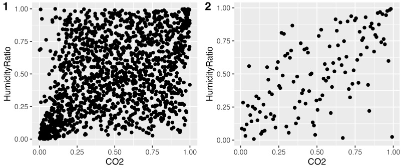


Figure: Pseudo-observations computed from unoccupied (Panel 1) and occupied (Panel 2) subsets.

Copulas: The Joys

- Copulas are mathematical devices used to model dependence between random variables regardless of their marginals.
- Copulas are useful for data fusion/integration as they lead to coherent joint models, even when the marginals are in different families or of different types.
- Copulas unlock information contained in the dependence part of the distribution (second-order) that complements the information in the marginals.
- Copulas extend statistical methods beyond the use of a multivariate Gaussian or Student.

At the root of it all, a theorem

- Copulas are distribution functions on $[0,1]^d$ that model dependence between continuous random variables.
- Sklar's Theorem: If $Y_1, Y_2, \dots Y_d$ are continuous r.v.'s with distribution functions (df) $F_1, \dots, F_d$, there exists an unique copula function $C : [0,1]^d \to [0,1]$ such that

$$H(t_1, \dots, t_d) = \mathbb{P}(Y_1 \le t_1, \dots, Y_d \le t_d) = C(F_1(t), \dots, F_d(t_d)).$$

- The copula bridges the marginal distributions of $Y_1, \dots, Y_d$ with the joint distribution. It corresponds to a distribution on $[0,1]^d$ with uniform margins.
- This can be extended to conditional distributions and copulas:

$$\mathbb{P}(Y_1 \le t_1, \dots, Y_d \le t_d | X) = C(F_1(t|X), \dots, F_d(t_d|X)|X).$$

Copulas Within HMMs

- Here we consider a HMM $\{(\mathbf{Y}_t, X_t)\}_{t\geq 1} \subseteq \mathbb{R}^d \times \mathcal{X}$ in which the state-dependent distributions use copulas:

$$\mathbf{Y}_t \mid (X_t = k) \sim H_k(\cdot) = \underbrace{C_k\Big(F_{k,1}(\cdot\,;\lambda_{k,1}), \ldots, F_{k,d}(\cdot\,;\lambda_{k,d}) \Big| \theta_k\Big)}_{\text{depends on the hidden state value } k}.$$

- $C_k(\cdot, \ldots, \cdot \mid \theta_k)$ is a d-dimensional parametric copula
- $\{X_t\}_{t\geq 1}$ is a Markov process on finite state space $\mathcal{X} = \{1, 2, \ldots, K\}$ and K is known.
- In this model, all aspects of the state-dependent distributions are allowed to vary between states

Information in the dependence

- For a range of $\theta \in [0, 100)$, we simulated a bivariate time series of length $T = 100$ from the 2-state HMM

$$\mathbf{Y}_t \mid (X_t = k) \sim C_{\text{Frank}} \left(\mathcal{N}(0,1), \mathcal{N}(0,1) \mid (-1)^k \cdot |\theta| \right), \quad k = 1, 2$$

and then separately assessed the accuracy of a standard decoding algorithm, first assuming independent margins and then the true model:

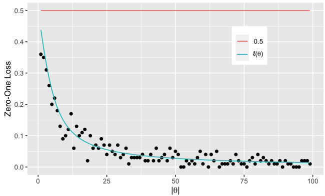


Figure: Zero-one losses for independent margins (red dots) and true model (blue dots)

Estimation with missing data

- Data consist in observed $\mathbf{Y}_{1:T}$ and missing $X_{1:T}$
- Parameters are $\eta = \{\lambda_{h,k}\}_{\substack{h=1:d \\ k=1:T}} \cup \{\theta_k\}_{k=1:T} \cup \{\gamma_{i,j}\}_{\substack{i=1:K \\ j=1:K}}$ $\cup\{\pi_j\}_{j=1:K}$.
- The complete-data log-likelihood for one trajectory of the copula HMM is given by

$$\ell_{\text{com}}\left(\boldsymbol{\eta} \mid \mathbf{y}_{1:T}, X_{1:T}\right) = \pi_{X_1} + \sum_{t=2}^{T} \log \gamma_{X_{t-1},X_t} + \sum_{h=1}^{d} \log f_{X_t,h}(y_{t,h}; \lambda_{X_t,h})$$
$$+ \sum_{t=1}^{T} \log c_{X_t}\left(F_{X_t,1}(y_{t,1}; \lambda_{X_t,1}), \ldots, F_{X_t,1}(y_{t,d}; \lambda_{X_t,d}) \mid \theta_{X_t}\right). \tag{1}$$

Computation for HMMs Via the EM Algorithm

- Without copula, the estimation is done via the EM algorithm (aka Baum-Welch)
- The complete-data log-likelihood is written in terms of the state membership indicators $U_{k,t} = \mathbb{1}_{X_t=k}$ and $V_{j,k,t} = \mathbb{1}_{X_{t-1}=j,X_t=k}$
- In the **E-Step**, these indicators are estimated by the conditional probabilities $\hat{u}_{k,t} = \mathbb{P}(X_t = k \mid \mathbf{Y}_{1:T} = \mathbf{y}_{1:T})$ and $\hat{v}_{j,k,t} = \mathbb{P}(X_{t-1} = j, X_t = k \mid \mathbf{Y}_{1:T} = \mathbf{y}_{1:T})$, which are computed based on current parameter estimates
- This only requires evaluating the state-dependent densities at each of the observations $\mathbf{y}_1, \ldots, \mathbf{y}_T$ (this is "OK")

The M-Step Is Hard

- In the **M-Step**, the resulting complete-data log-likelihood is maximized with respect to all parameters in the model simultaneously
  - Only for the simplest univariate models do the state-dependent MLEs exist in closed form; otherwise, one must resort to numerical methods (this is hard and unstable!)
  - Evaluating a copula density $c_k(\cdot, \dots, \cdot \mid \theta_k)$ in high dimensions is slow
  - When the state-dependent distributions in an HMM are copulas, performing the M-Step directly requires the evaluation of

$$\underset{\{\theta_k\},\{\lambda_{k,h}\}}{\operatorname{argmax}} \left\{ \sum_{k=1}^{K} \sum_{t=1}^{T} \hat{u}_{k,t} \left[\log c_k \left(F_{k,1}(y_{t,1}; \lambda_{k,1}), \dots, F_{k,d}(y_{t,d}; \lambda_{k,d}) \,\middle|\, \theta_k \right) + \sum_{h=1}^{d} \log f_{k,h}(y_{t,h}; \lambda_{k,h}) \right] \right\}$$

  - This is very unstable (and slow)

Inference Functions for Margins

- Likelihood-based inference for copulas is easier when the goal is to estimate θ alone in the presence of known margins
- Why not perform inference on the marginal distributions first, and then on the copula itself?
- In the context of iid data, this is exactly the inference functions for margins (IFM) approach of Joe and Xu (1996):
  - First estimate each λ_h by its "marginal MLE" $\hat{\lambda}_h$ given $\{Y_{t,h}\}_{t\geq 1}$, for $h \in \{1, \ldots, d\}$
  - Then estimate θ assuming fixed marginals $F_1(\cdot; \hat{\lambda}_1), \ldots, F_d(\cdot; \hat{\lambda}_d)$
- One can show that the IFM estimator is consistent and asymptotically normal (although relatively less efficient than the MLE)

New problems

- The EIFM algorithm is not an GEM algorithm

$$\sum_{t=1}^{T} \hat{u}_t \cdot \log\left(f_h(y_{t,h}; \lambda_h^{(s)})\right) \leq \sum_{t=1}^{T} \hat{u}_t \cdot \log\left(f_h(y_{t,h}; \lambda_h^{(s+1)})\right), \qquad h \in \{1, \dots, d\} \quad (2)$$

does not imply

$$\sum_{t=1}^{T} \hat{u}_t \cdot \log\left(c\left(F_1(y_{t,1}; \lambda_1^{(s)}), \dots, F_d(y_{t,d}; \lambda_d^{(s)}) \,\middle|\, \theta^{(s)}\right)\right)$$
$$\leq \sum_{t=1}^{T} \hat{u}_t \cdot \log\left(c\left(F_1(y_{t,1}; \lambda_1^{(s+1)}), \dots, F_d(y_{t,d}; \lambda_d^{(s+1)}) \,\middle|\, \theta^{(s)}\right)\right).$$

- The EIFM algorithm will converge (to a local or global maximum).
- The estimator is consistent and asymptotically normal (under mild regularity conditions).
- EIFM as a version of the ES algorithm of Elashoff and Ryan (2004).
- Use asymptotic theory of M-estimators for HMMs Jensen (2011).

Does This Work?

- For $T \in \{100, 1000, 5000\}$ and $d \in \{2, 5, 10\}$, we simulated a d-dimensional time series of length T from the 2-state HMM

$$\mathbf{Y}_t \mid (X_t = 1) \sim C_{\text{Frank}} \left((\mathcal{N}(\mu_{1,h} = -h, 1))_{h=1}^{d} \mid \theta_1 = 3 \right)$$
$$\mathbf{Y}_t \mid (X_t = 2) \sim C_{\text{Clayton}} \left((\mathcal{N}(\mu_{2,h} = h, 1))_{h=1}^{d} \mid \theta_2 = 3 \right)$$

  and estimated $\boldsymbol{\eta} = (\mu_{1,1}, \ldots, \mu_{2,d}, \theta_1, \theta_2)$ using both approaches

- Applied to the basic EM algorithm, R's `optim` with `L-BFGS-B` (i.e., quasi-Newton with box constraints) typically fails as soon as $d \geq 3$
  - The procedure is extremely sensitive to initial values and requires $\hat{\boldsymbol{\eta}}^{(0)} \approx \boldsymbol{\eta}$ just to avoid overflow
  - This kind of tuning is very tedious or impossible in high dimensions

Does This Work?

- We keep track of the time (in seconds) until the algorithm converges, and the L_2 error of the resulting estimate, $\epsilon = \|\boldsymbol{\eta} - \hat{\boldsymbol{\eta}}\|_2$
  - We used the `lbfgsb3c` package, which is more stable than `optim`

	$d = 2$	$d = 5$	$d = 10$
$T = 100$	111.9 s, $\epsilon = 0.14$	123.4 s, $\epsilon = 299.98$	111.8 s, $\epsilon > 10^9$
$T = 1000$	166.6 s, $\epsilon = 0.63$	169.5 s, $\epsilon > 10^{11}$	418.23 s, $\epsilon = 725.06$
$T = 5000$	?	?	?

Table: EM Algorithm

	$d = 2$	$d = 5$	$d = 10$
$T = 100$	5.1 s, $\epsilon = 0.29$	3.0 s, $\epsilon = 0.94$	4.2 s, $\epsilon = 0.58$
$T = 1000$	34.4 s, $\epsilon = 0.57$	22.9 s, $\epsilon = 0.60$	34.4 s, $\epsilon = 0.80$
$T = 5000$	172.6 s, $\epsilon = 0.13$	106.2 s, $\epsilon = 0.12$	168.7 s, $\epsilon = 0.19$

Table: EFM Algorithm

Numerical Experiment I

- Generative model:

$$\mathbf{Y}_i \mid (X_i = k) \sim C_k \left(SN(\cdot; \xi_{k,1}, \omega_{k,1}, \alpha_{k,1}), SN(\cdot; \xi_{k,2}, \omega_{k,2}, \alpha_{k,2}) \mid \tau_k \right),$$

for $k \in \{1, \ldots, 4\}$.

State	Copula family	τ_k	$\xi_{k,1}$	$\omega_{k,1}$	$\alpha_{k,1}$	$\xi_{k,2}$	$\omega_{k,2}$	$\alpha_{k,2}$
1	Clayton	0.2	-4	1	5	-1	1	-3
2	B4	0.4	-2	1	3	2	1	-3
3	Gaussian	0.6	0	1	5	3	1	-5
4	$t_{(\nu=5)}$	0.8	2	1	3	4	1	-5

Table: True parameters for the state-dependent distributions.

Numerical Experiment I

T:		500	1000	2500	5000
Stopping Rule Tolerance:	0.01	14	24	23	15
	0.001	17	26	25	17
	0.0001	36	59	62	39
	0.00001	230	115	460	269
Classifier:	k-means	0.9020	0.9090	0.9200	0.9196
	Local state decoding	0.9640	0.9640	0.9696	0.9732

Table: For each $T \in \{500, 1000, 2500, 5000\}$: (Top rows) Number of iterations taken by the EIFM algorithm applied to $\mathbf{Y}_{1:T}$ before stopping using L_1-norm tolerances in $\{0.01, 0.001, 0.0001, 0.00001\}$. (Bottom rows) Classification accuracy of initial k-means clustering and local decoding with parameter estimates provided by the EIFM algorithm.

Occupancy Data

- Several families of parametric copulas were tried

Family	AIC (State 1)	AIC (State 2)
Gauss	-370.786	-59.281
t	-437.542	-71.291
Clayton	-474.836	-66.087
Gumbel	-484.252	-72.437
Frank	-273.613	-66.497
Joe	-490.103	-64.995
Galambos	-295.549	-71.031
Hüsler-Reiss	-282.452	-61.547
BB1	-497.013	-70.509
BB6	-495.401	-70.435
BB7	**-518.923**	-68.207
BB8	-490.223	-66.738
Tawn (type 1)	-411.847	**-76.976**
Tawn (type 2)	-422.268	-55.382

Table: AICs for unoccupied (state 1) and occupied (state 2) classifications of the occupancy data.

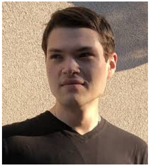
Occupancy Data

Classifier	Train	Test 1
k-means clustering	0.865	0.818
Independence copulas within HMM	0.895	0.846
BB7/Tawn copulas within HMM	0.900	0.852

Table: Overall state classification accuracy for the training dataset and the test dataset, using *k*-means clustering and local decoding via the HMM with independent margins and the copula-within-HMM model.

Latent Variable Models with Copulas

This project is currently developed in collaboration with



Robert Zimmerman (Imperial College London)

Latent Variables (LV)

- The variable of interest W is sometimes impossible to measure directly
  - State of the economy
  - Traffic in a city
  - State of your health
  - State of a complex disease
- Instead, one measures
  - $\mathbf{Y} = (Y_1, \ldots, Y_k)^T$ whose components are surrogates of W and each provide partial information about W
  - Covariate $\mathbf{X} \in \mathbb{R}^p$
- We are often interested in the explanatory power of $\mathbf{X}$ for W.

An example

- Cardiotocography (CTG) is a medical procedure that monitors the fetal heart rate.
- The LV is the fetus' underlying state of health during birth, W.
- Our surrogate response is the bivariate vector (Q, Y) where
  - Q is the number of peaks (acceleration followed by a deceleration of heart beats) for the signal recorded by the CTG
  - Y is the log of mean short-term "beat-to-beat" variability (MSTV) where the short-term variability (STV) is obtained by measuring the time between successive R waves (cardiac systoles) of the fetus' electrocardiogram.
- The covariates are `FM` (fetal movement) and `UC` (uterine contraction), two continuous variables monitored during birth.

Conditional independence LV model

- A canonical LV model, given $W_i = X_i\beta + \epsilon$, is

$$
\begin{aligned}
&Y_i \perp Q_i | W_i \\
&Y_i \sim N(\mu_c + \lambda_c W_i, \sigma^2) \\
&Q_i \sim Poisson(\exp(\mu_d + \lambda_d W_i))
\end{aligned}
$$

- This implies that the two marginal regressions share a common random effect so they are marginally dependent (and conditionally independent)

- The induced dependence is not analytically available.

Conditional independence is a Copula LV

- The copula alternative is, conditional on W_i,

$$H(Y_i, Q_i|W_i) = C_{\theta_i}(F_Y(Y_i|W_i), F_Q(Q_i|W_i)), \;\; \theta_i = \kappa^{-1}(\xi_0 + \xi_1 W_i)$$
$$Y_i \sim N(\mu_c + \lambda_c W_i, \sigma^2); \; Q_i \sim Poisson(\exp(\mu_d + \lambda_d W_i))$$

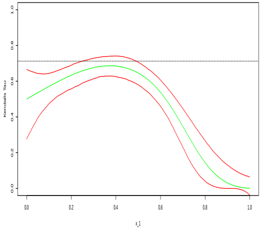
- The whole joint distribution of (Y, Q) is varying with W not just the marginals.
- The copula captures the residual dependence on W after the marginal effects have been accounted for.
- The previous model is obtained when the copula is the independence copula.

Why the Conditional Copula?

- $Y_i|x \sim N(f_i(x), \sigma_i)$ $x \in \mathbb{R}^2$
- True marginal means:
  - $f_1(x) = 0.6\sin(5x_1) - 0.9\sin(2x_2)$
  - $f_2(x) = 0.6\sin(3x_1 + 5x_2)$
  - $\sigma_1 = \sigma_2 = 0.2$, $X_1 \perp X_2$.
- Copula: Frank with $\theta(x) = 0.71$
- Suppose x_2 is not observed so inference is based only on x_1

Why the Conditional Copula?

- $Y_i | x \sim N(f_i(x), \sigma_i)$ $x \in \mathbb{R}^2$
- True marginal means:
  - $f_1(x) = 0.6 \sin(5x_1) - 0.9 \sin(2x_2)$
  - $f_2(x) = 0.6 \sin(3x_1 + 5x_2)$
  - $\sigma_1 = \sigma_2 = 0.2$, $X_1 \perp X_2$.
- Copula: Frank with $\theta(x) = 0.71$
- Suppose x_2 is not observed so inference is based only on x_1



(Levi and Craiu, 2018)

CTG: The LV Copula Model

- $(Q_i, Y_i)|W_i$ has joint density

$$f_{(Q,Y)}(q,y) = f_c(y) \cdot \left[C_{d|c}\left(F_d(q), F_c(y)\right) - C_{d|c}\left(F_d(q-), F_c(y)\right)\right],$$

where

$$C_{d|c}(u_d, u_c) = \frac{\partial}{\partial u_c} C(u_d, u_c).$$

- Data Augmentation: Introduce latent variable Z such that

$$Q \stackrel{d}{=} F_d^-(F_Z(Z)),$$

- The copula between (Y, Z) is the same as the copula between (Y, Q)
- We can choose the distribution of Z to help the computation.
- For instance if we use a Gaussian copula, it helps to have $Z \sim N(0,1)$
- Craiu and Sabeti (2012); Smith and Khaled (2012).

CTG: The Augmented LV Copula Model

- The dependence between Y, Z and Q is defined by their joint conditional distribution

$$f_{(Q,Z,Y)|W}(q,z,y \mid w) = h(z,y \mid w, \mu_c, \lambda_c, \boldsymbol{\psi}_c, \boldsymbol{\xi}) \cdot \mathbb{1}_{F_Z^{-1}(F_d(q-|\varphi_d(\mu_d,\lambda_d,w))) \leq z < F_Z^{-1}(F_d(q|\varphi_d(\mu_d,\lambda_d,w)))}.$$

- Let $\boldsymbol{\xi} = (\xi_0, \xi_1) \in \mathbb{R}^2$ and $A(w) = \xi_0 + \xi_1 \cdot w$. Then we set

$$\theta(w, \boldsymbol{\xi}) = \frac{e^{A(w)} - e^{-A(w)}}{e^{A(w)} - e^{-A(w)}}$$

  as the correlation parameter of the bivariate Gaussian conditional copula of $(Y,Z)|W = w$.

- Parameters are a priori independent

Model Selection: WAIC

- The WAIC is defined as

$$\mathrm{WAIC}(\mathcal{M}) = -2\mathrm{fit}(\mathcal{M}) + 2\mathrm{p}(\mathcal{M}), \tag{3}$$

where the model fitness is

$$\mathrm{fit}(\mathcal{M}) = \sum_{i=1}^{n} \log\left(\mathbb{E}\left[\Pr(y_i, q_i|\omega, \mathcal{M})\right]\right) \tag{4}$$

and the penalty

$$\mathrm{p}(\mathcal{M}) = \sum_{i=1}^{n} \mathrm{Var}\left(\log\left(\Pr(y_i, q_i|\omega, \mathcal{M})\right)\right), \tag{5}$$

where ω contains all the parameters and latent variables in the model.

Spotlight on dependence: A conditional WAIC

- We use the following two conditional WAICs (Levi and Craiu, 2018)

$$\mathrm{CWAIC}_{Y|Q}(\mathcal{M}) = -2\sum_{i=1}^{n}\log\left(\mathbb{E}\left[\Pr(y_i|q_i,\omega,\mathcal{M})\right]\right) + \\ +2\sum_{i=1}^{n}\mathrm{Var}\left(\log\left(\Pr(y_i|q_i,\omega,\mathcal{M})\right)\right),$$

$$\mathrm{CWAIC}_{Q|Y}(\mathcal{M}) = -2\sum_{i=1}^{n}\log\left(\mathbb{E}\left[\Pr(q_i|y_i,\omega,\mathcal{M})\right]\right) + \\ +2\sum_{i=1}^{n}\mathrm{Var}\left(\log\left(\Pr(q_i|y_i,\omega,\mathcal{M})\right)\right),$$

- $\frac{1}{2}(\mathrm{CWAIC}_{1|2} + \mathrm{CWAIC}_{2|1})$ is asymptotically equivalent to the following CCV for the marginal (conditional) distribution

$$\mathrm{CCV}(\mathcal{M}) = \frac{1}{2}\left\{\sum_{i=1}^{n}\log\left(\Pr(y_i|q_i,\mathcal{D}_{-i},\mathcal{M})\right) + \sum_{i=1}^{n}\log\left(\Pr(q_i|y_i,\mathcal{D}_{-i},\mathcal{M})\right)\right\}.$$

Simulation Experiment

- Generate data using a Gaussian copula

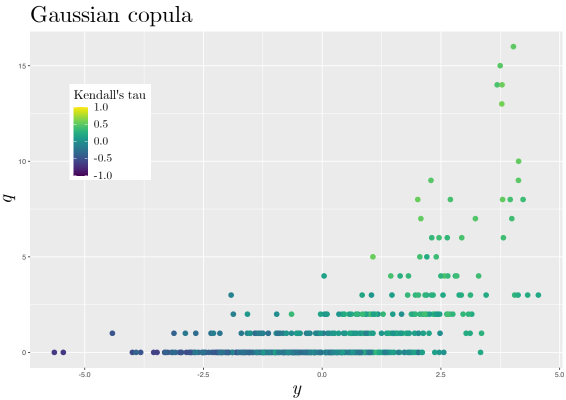


Figure: Bivariate scatterplot of the generated data with Gaussian copula, and Poisson and normal marginals

Simulation Experiment

- $\text{CWAIC}_{Y|Q}$ and $\text{CWAIC}_{Q|Y}$ selection criteria

Criteria\Copula	Gaussian	Frank	Gumbel	Clayton	Indep
$\text{CWAIC}_{Y\|Q}$	1627.36	1642.36	2395.17	1637.17	1606.31
$\text{CWAIC}_{Q\|Y}$	950.71	982.42	1673.57	976.05	997.43
Average	1289.04	1312.39	2034.37	1306.61	1301.87

Simulation Experiment

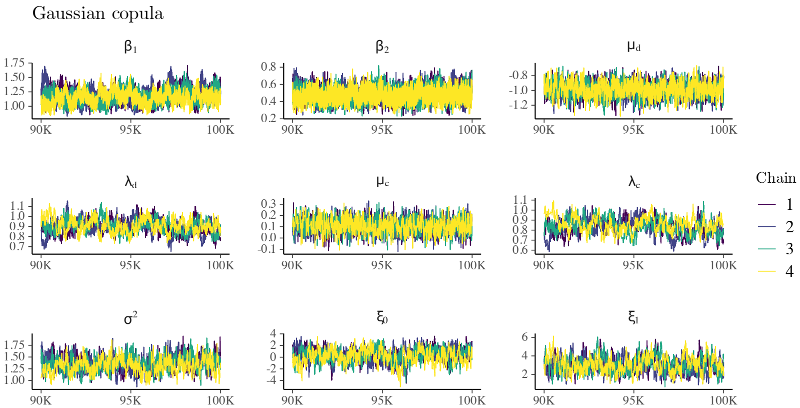
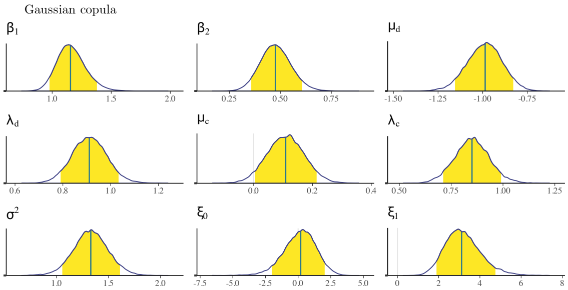


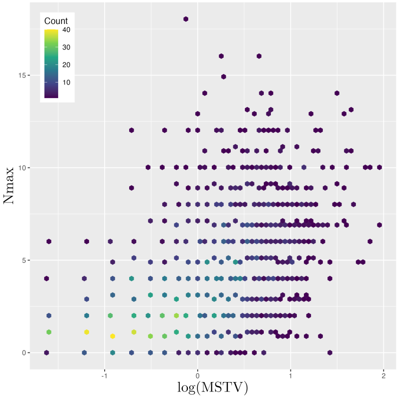
Figure: Traceplots for η's components.

Simulation Experiment



	β_1	β_2	λ_d	λ_c	ξ_1
Mean	1.18	0.48	0.90	0.84	3.10
True	1	0.5	1	1	3

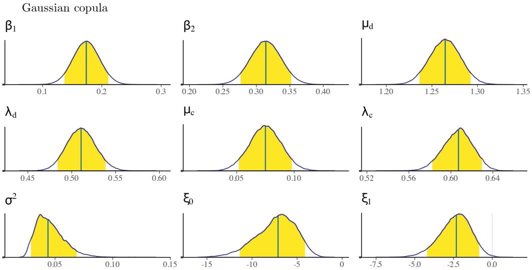
CTG: The data



CTG: Estimates

- $WAIC$, $WAIC_{Y|Q}$ and $WAIC_{Q|Y}$ all point to the Gaussian copula (over Gumbel, Frank, Clayton, Independence).
- The posterior means

	β_1 (FM)	β_2 (UC)	λ_d	λ_c	ξ_1
Mean	0.1744	0.3147	0.5101	0.6038	-2.3401



References

Candanedo, L. M. and Feldheim, V. (2016). Accurate occupancy detection of an office room from light, temperature, humidity and co2 measurements using statistical learning models. *Energy and Buildings* **112** 28–39.

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

Elashoff, M. and Ryan, L. (2004). An EM algorithm for estimating equations. *J. Comput. Graph. Stat.* **13** 48–65.

Jensen, J. L. (2011). Asymptotic normality of m-estimators in nonhomogeneous hidden Markov models. *J. App. Probab.* **48** 295–306.

Joe, H. and Xu, J. J. (1996). The estimation method of inference functions for margins for multivariate models. Tech. Rep. 166, Department of Statistics, University of British Columbia.

Levi, E. and Craiu, R. V. (2018). Bayesian inference for conditional copulas using Gaussian process single index models. *Computational Statistics & Data Analysis* **122** 115–134.

Nasri, B. R., Remillard, B. N. and Thioub, M. Y. (2020). Goodness-of-fit for regime-switching copula models with application to option pricing. *Can. J. Stat.* **48** 79–96.

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

Zucchini, W., MacDonald, I. L. and Langrock, R. (2017). *Hidden Markov models for time series: an introduction using R*. CRC press.

Papers available on my homepage: raducraiu.com