Bayesian Inference for Conditional Copulas using Gaussian Process Single Index Models

Radu Craiu

Department of Statistical Sciences
University of Toronto

Joint with Evgeny Levi (Toronto)

ISNPS, Salerno
June 2018

Conditional Copula

- Consider a random sample $\{x_i \in \mathbf{R}^d, y_{1i} \in \mathbf{R}, y_{2i} \in \mathbf{R}\}_{1 \le i \le n}$ and suppose $F_X(y_1)$ and $G_X(y_2)$ are the unknown marginal conditional cdf's.
- The bivariate conditional copula (CC) of $(Y_1, Y_2)|X = x$, is the conditional joint distribution function of $U = F_x(Y_1)$ and $V = G_x(Y_2)$ given $X = x$ (Patton, Int'l Econ. Rev. '06)

$$H_x(t, s) = C_x(F_x(t), G_x(s))$$

- The parametric bivariate CC model assumes there is a parametric family $\mathcal{C} = \{C_\theta : \theta \in \Theta\}$ s.t.

$$C_x(F_x(t), G_x(s)) = C_{\theta(x)}(F_x(Y_1), G_x(Y_2)).$$

- The simplifying assumption:

$$C_x(F_x(y_1), G_x(y_2)) = C(F_x(y_1), G_x(y_2)).$$

Why CC?

- We are interested in understanding the covariate effect on the dependence pattern between responses.
- Joint models for multivariate data: if $U_1, U_2, U_3 \sim \text{Uniform}(0,1)$ then the joint pdf

$$c(u_1, u_2, u_3) = c_{12}(u_1, u_2)c_{23}(u_2, u_3)c_{\theta(u_2)}(F(u_1|u_2), G(u_3|u_2)).$$

- Regression-based prediction: if

$$h_x(y_1, y_2) = f_x(y_1)g_x(y_2)c_{\theta(x)}(F_x(y_1), G_x(y_2)), \text{ then}$$

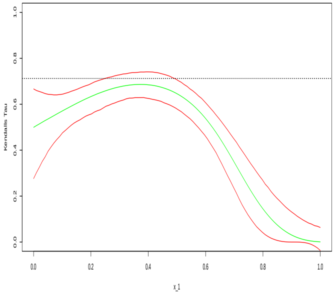
$$h_x(y_1|y_2) = f_x(y_1)c_{\theta(x)}(F_x(y_1), G_x(y_2)).$$

Why CC? - Model misspecification effects

- Marginals
  - $f_1(x) = 0.6\sin(5x_1) - 0.9\sin(2x_2)$
  - $f_2(x) = 0.6\sin(3x_1 + 5x_2)$
  - $\sigma_1 = \sigma_2 = 0.2$, $X_1 \perp X_2$.
- Copula: $\tau(x) = 0.71$
- Model:
  - Fit nonparametric model for marginals and CC with only x_1.

Why CC? - Model misspecification effects

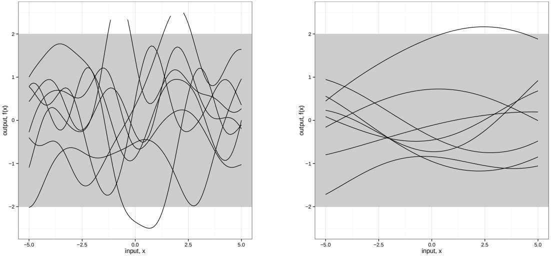
- Marginals
  - $f_1(x) = 0.6\sin(5x_1) - 0.9\sin(2x_2)$
  - $f_2(x) = 0.6\sin(3x_1 + 5x_2)$
  - $\sigma_1 = \sigma_2 = 0.2$, $X_1 \perp X_2$.
- Copula: $\tau(x) = 0.71$
- Model:
  - Fit nonparametric model for marginals and CC with only x_1.



Gaussian Process Prior

- GP prior for smooth f without specifying the form of f.
- For $x \in [-5, 5]^n$, consider $f \sim N_n(0, K(x, x))$ where $K_{ij}(x, x) = k(x_i, x_j)$ and $f_i = f(x_i)$
- $\text{Cov}(f(x_i), f(x_j)) = k(x_i, x_j) = \exp\{-0.5 * \frac{|x_i - x_j|}{L}^2\}$.

$L = 1$ $L = 5$



- Random functions f generated from a GP prior when $n = 100$

Gaussian Process Estimation

- Observe $\{y_i : 1 \le i \le n\}$ noisy realizations of $f(x_i)_{i=1,n}$, $y_i = f(x_i) + \epsilon_i$, $\epsilon_i \sim N(0, \sigma^2)$.
- When interested in predicting $f^* = (f(x_j^*))_{j=1,q}$ use

$$\begin{pmatrix} y \\ f^* \end{pmatrix} \sim N_{n+q}\left(\mathbf{0}, \begin{bmatrix} K(x,x) + \sigma^2 \mathbf{I}_n & K(x,x^*) \\ K(x,x^*) & K(x^*,x^*) \end{bmatrix}\right)$$

- The conditional distribution of f^* is Gaussian with

$$\mathrm{E}(f^*|y) = K(x^*,x) \overbrace{[K(x,x) + \sigma^2 \mathbf{I}_q]^{-1}}^{\text{expensive for large } n} y$$

$$V(f^*|y) = K(x^*,x^*) - K(x^*,x) \underbrace{[K(x,x) + \sigma^2 \mathbf{I}_q]^{-1}}_{\text{expensive for large } n} K(x,x^*)$$

Sparse GP-SIM

- When n is large the computation effort is prohibitive so we adopt a sparse GP approach (Snelson & Ghahramani 2005; Quiñonero-Candela & Rasmussen 2005)
- The information about f in the data is funnelled using a smaller sample of size $m << n$ of inducing (or latent) variables $\tilde{x}_g$, $1 \leq g \leq m$.
- We consider the SIM model (Choi et al. 2011; Gramacy & Lian 2012)
$$f(X) = f(\beta^T X).$$
- GP-SIM model is invariant to nonlinear one-to-one transformations $\tau(\theta)$.

Proof of concept

Sc1 $f_1(x) = 0.6\sin(5x_1) - 0.9\sin(2x_2),$
$f_2(x) = 0.6\sin(3x_1 + 5x_2),$
$\tau(x) = 0.7 + 0.15\sin(15x^T\beta)$
$\beta = (1,3)^T/\sqrt{10},\ \sigma_1 = \sigma_2 = 0.2\ \ n = 400$

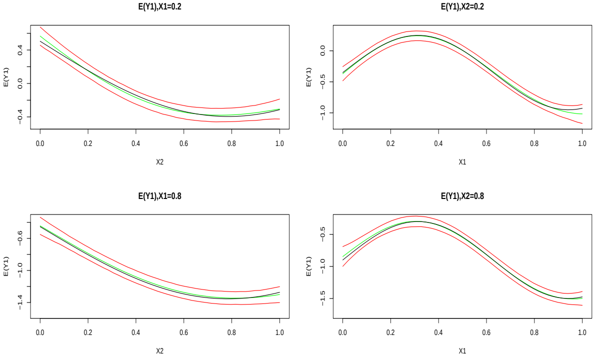
	Clayton			Frank			Gaussian			Clayton SA		
Scenario	$\sqrt{\text{IBias}^2}$	$\sqrt{\text{IVar}}$	$\sqrt{\text{IMSE}}$	$\sqrt{\text{IBias}^2}$	$\sqrt{\text{IVar}}$	$\sqrt{\text{IMSE}}$	$\sqrt{\text{IBias}^2}$	$\sqrt{\text{IVar}}$	$\sqrt{\text{IMSE}}$	$\sqrt{\text{IBias}^2}$	$\sqrt{\text{IVar}}$	$\sqrt{\text{IMSE}}$
Sc1	0.0231	0.0531	0.0579	0.1264	0.0322	0.1304	0.1434	0.0557	0.1539	0.0416	0.0579	0.0713

Integrated error for the estimator of $\tau(x)$.

Prediction performance

- If $y_i|x \sim N(\mu_i(x), \sigma_i^2)$, $i = 1, 2$ then

$$E_x[Y_1|Y_2 = y_2] = \mu_1(x) + \sigma_1 \int_0^1 \Phi^{-1}(z) c_{\theta(x)} \left(z, \Phi\left(\frac{y_2 - \mu_2(x)}{\sigma_2} \right) \right) dz.$$



Model Selection Problems

- Choice of copula family.
- Choice of calibration
  - Simplifying Assumption or not?
- Covariate selection.

CV Marginal Likelihood (CVML)

- Calculates the average (over parameter values) prediction potential for model $\mathcal{M}$ via

$$\text{CVML}(\mathcal{M}) = \sum_{i=1}^{n} \log\left(P(Y_{1i}, Y_{2i}|\mathcal{D}_{-i}, \mathcal{M})\right),$$

- $\mathcal{D}_{-i}$ is the data set from which the ith observation has been removed.

CV Marginal Likelihood (CVML)

- Estimate CVML using

$$E_\pi\left[P(Y_{1i}, Y_{2i}|\omega, \mathcal{M})^{-1}\right] = P(Y_{1i}, Y_{2i}|\mathcal{D}_{-i}, \mathcal{M})^{-1}$$

where ω represents the vector of all the parameters and latent variables in the model.

- Numerically

$$\begin{aligned}\text{CVML} &= \sum_{i=1}^{n} \log\left\{\frac{1}{M}\sum_{t=1}^{M}\frac{1}{\sigma_1^{(t)}}\phi\left(\frac{y_{1i}-f_{1i}^{(t)}}{\sigma_1^{(t)}}\right)\frac{1}{\sigma_2^{(t)}}\phi\left(\frac{y_{2i}-f_{2i}^{(t)}}{\sigma_2^{(t)}}\right)\times\right.\\ &\times \left. c_{\theta_i^{(t)}}\left[\Phi\left(\frac{y_{1i}-f_{1i}^{(t)}}{\sigma_1^{(t)}}\right),\Phi\left(\frac{y_{2i}-f_{2i}^{(t)}}{\sigma_2^{(t)}}\right)\right]\right\}.\end{aligned}$$

- Add scenario **S2** where SA is true:

$$f_1(x) = 0.6\sin(5x_1) - 0.9\sin(2x_2)$$
$$f_2(x) = 0.6\sin(3x_1 + 5x_2)$$
$$\tau(x) = 0.5$$
$$\sigma_1 = \sigma_2 = 0.2$$

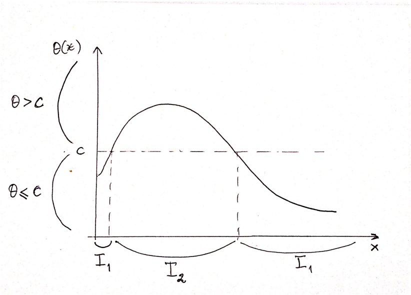
Calibration Selection - Results

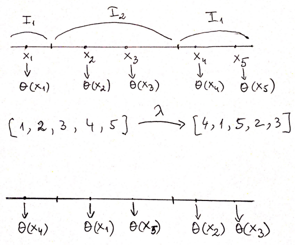
Scenario	Clayton		Frank		Gaussian	
	CVML	CCVML	CVML	CCVML	CVML	CCVML
Sc2 (SA is true)	58%	62%	100%	100%	100%	100%

A permutation based diagnostic for SA

- Randomly partition the data into a training set $\mathcal{D}_* = \{y_{1i}, y_{2i}, x_i\}_{i=1,\ldots,n_1}$ (66% of sample) and a test set $\mathcal{D}^* = \{y^*_{1i}, y^*_{2i}, x^*_i\}_{i=1,\ldots,n_2}$ (34% of sample).
- Fit marginal models using GP and a nonconstant calibration on $\mathcal{D}$.

A permutation based diagnostic for SA





- Use training data to fit the calibration curve
- Use the test data to verify support for SA.
- Permutation influences only the copula factor.

A permutation based diagnostic for SA

- Consider J permutations of $\{1\dots n_2\}$ which we denote as $\lambda_1,\dots,\lambda_J:\{1,\dots,n_2\}\to\{1,\dots,n_2\}$
- Compute J permuted CVMLs as:

$$\begin{aligned}\mathrm{CVML}_j &= \sum_{i=1}^{n_2}\log\left\{\frac{1}{M}\sum_{t=1}^{M}\frac{1}{\sigma_1^{(t)}}\phi\left(\frac{y_{1i}^*-f_{1i}^{*(t)}}{\sigma_1^{(t)}}\right)\frac{1}{\sigma_2^{(t)}}\phi\left(\frac{y_{2i}^*-f_{2i}^{*(t)}}{\sigma_2^{(t)}}\right)\times\right.\\ &\times \left. c_{\theta^{*(t)}_{\lambda_j(i)}}\left[\Phi\left(\frac{y_{1i}^*-f_{1i}^{*(t)}}{\sigma_1^{(t)}}\right),\Phi\left(\frac{y_{2i}^*-f_{2i}^{*(t)}}{\sigma_2^{(t)}}\right)\right]\right\}.\end{aligned}$$

- If calibration is constant then CVML_{obs} and CVML_j should be similar
- Define the evidence

$$\mathrm{EV}=2\times\min\left\{\frac{\sum_{j=1}^{J}\mathbf{1}_{\{CVML_{obs}<CVML_j\}}}{J},\frac{\sum_{j=1}^{J}\mathbf{1}_{\{CVML_{obs}>CVML_j\}}}{J}\right\}. \quad (1)$$

A permutation based diagnostic for SA

Scenario	Perm CVML	Perm CCVML	CVML	CCVML
Sc1	98%	96%	94%	94%
Sc2	92%	90%	58%	62%
Sc3	100%	100%	100%	100%

Papers available at
`http://www.utstat.toronto.edu/craiu/Papers/index.html`