Expectation Propagation for the Estimation of Conditional Bivariate Copulas

José Miguel Hernández-Lobato[1], David López-Paz[2] and Zoubin Ghahramani[1]

[1]Department of Engineering, Cambridge University

[2]Max Planck Institute for Intelligent Systems

December 16, 2011

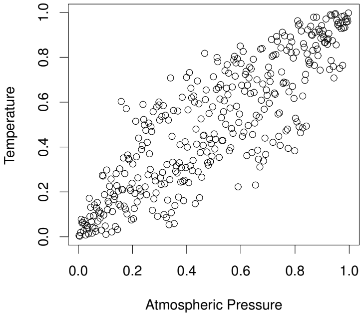
Copulas and Covariates

Copulas are useful tools for the construction of **multivariate** models.
They allow us to **separate** the modelling of marginals and dependence.

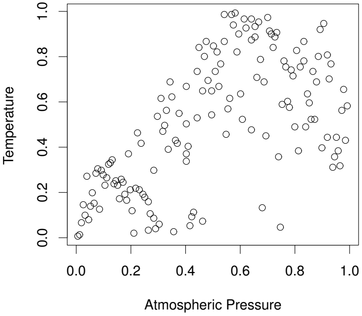
Sklar's Theorem:
Let F be a bivariate distribution with continuous marginals F_X and F_Y. Then, there exists a unique copula C such that $F(x, y) = C[F_X(x), F_Y(y)]$.

However, C (and F_X and F_Y) may depend on the effect of a **covariate** Z.

Wind Velocity Low



Wind Velocity High



Conditional Copulas

We can adjust for the effect of Z by modelling the conditional copula.

Conditional Copula: [Patton 2006]
Let $F_{X|Z}$ and $F_{Y|Z}$ be the conditional marginals of X and Y given Z.
The conditional copula C_Z is the distribution of $F_{X|Z}(X|Z)$ and $F_{Z|Y}(Y|Z)$.

Extension of Sklar's Theorem: [Patton 2006]
Let F_Z be the conditional distribution with continuous marginals $F_{X|Z}$ and $F_{Y|Z}$.
Then, there is a unique conditional copula C_Z such that
$F_Z(x, y|z) = C_Z[F_{X|Z}(x|z), F_{Y|Z}(y|z)|z]$.

The estimation of F_Z from a sample $\{X_i, Y_i, Z_i\}_{i=1}^n$ can then be done by

1. Estimating $F_{X|Z}$ and $F_{Y|Z}$ using our favorite method.
2. Mapping $\{X_i, Y_i\}_{i=1}^n$to $[0, 1]^2$ using the estimates for $F_{X|Z}$ and $F_{Y|Z}$.
3. Estimating C_Z using the data from the step 2. **We need a model for C_Z**.

A Semi-parametric Model for Conditional Copulas

We describe C_Z using a **parametric model** specified in terms of Kendall's tau.

Kendall's tau: [Joe 1997]
Let C be a bivariate copula and let (U_1, V_1) and (U_2, V_2) be two independent samples from C. Then, Kendall's tau is given by

$$\tau = \mathcal{P}[(U_1 - U_2)(V_1 - V_2) > 0] - \mathcal{P}[(U_1 - U_2)(V_1 - V_2) < 0]$$

and satisfies:

- ⋆ U and V are independent $\rightarrow \tau = 0$.
- ⋆ U and V have perfect positive dependence $\rightarrow \tau = 1$.
- ⋆ U and V have perfect negative dependence $\rightarrow \tau = -1$.

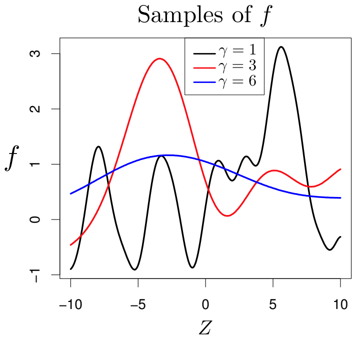
Most parametric bivariate copulas are **fully determined given $\boldsymbol{\tau}$**.
For example, in the Gaussian copula, $\rho = \sin(\tau\pi/2)$.

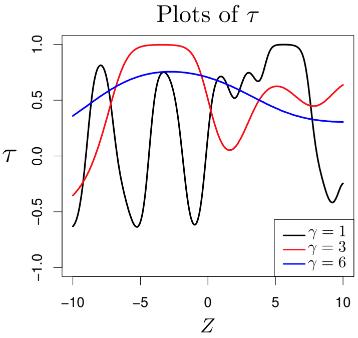
The dependence of C_Z on Z is captured by the relationship $\tau = \sigma[f(Z)]$, where f is an arbitrary non-linear function and $\sigma(x) = 2\Phi(x) - 1$ is a sigmoid function.

Bayesian Inference

Given a sample $\mathcal{D}_{UV} = \{U_i, V_i\}_{i=1}^n$ from C_Z with corresponding covariate values $\mathcal{D}_Z = \{Z_i\}_{i=1}^n$, we want to identify the value of f that generated the data.

We assume that f follows *a priori* a **Gaussian process** with zero mean and covariance function $k[f(Z_i), f(Z_j)] = \exp\{-0.5\gamma^{-2}(Z_i - Z_j)^2\}$.





Posterior and Predictive Distributions

Given $\mathcal{D}_{UV} = \{U_i, V_i\}_{i=1}^n$ and $\mathcal{D}_Z = \{Z_i\}_{i=1}^n$, the **posterior distribution** for $\mathbf{f} = (f_1, \ldots, f_n)^\mathsf{T}$, where $f_i = f(Z_i)$, is obtained by Bayes rule:

$$\mathcal{P}(\mathbf{f}|\mathcal{D}_{UV}, \mathcal{D}_Z) = \frac{\prod_{i=1}^n \mathcal{P}(U_i, V_i|\tau = \sigma[f_i])\mathcal{P}(\mathbf{f}|\mathcal{D}_Z)}{\mathcal{P}(\mathcal{D}_{UV}|\mathcal{D}_Z)}\,,$$

where $\mathcal{P}(\mathbf{f}|\mathcal{D}_Z) = \mathcal{N}(\mathbf{f}|\mathbf{0}, \mathbf{K})$ is the **GP prior** with $k_{ij} = k(Z_i, Z_j)$, $\mathcal{P}(U_i, V_i|\tau = \sigma[f_i])$ is the **likelihood** of the parametric copula model and $\mathcal{P}(\mathcal{D}_{UV}|\mathcal{D}_Z)$ is the **model evidence**.

Given Z_{n+1}, the **predictive distribution** for U_{n+1} and V_{n+1} is

$$\mathcal{P}(u_{n+1}, v_{n+1}|Z_{n+1}, \mathcal{D}_{UV}, \mathcal{D}_Z) = \int \mathcal{P}(u_{n+1}, v_{n+1}|\tau = \sigma[f_{n+1}]) \\ \mathcal{P}(f_{n+1}|\mathbf{f}, Z_{n+1}, \mathcal{D}_Z)\mathcal{P}(\mathbf{f}|\mathcal{D}_{UV}, \mathcal{D}_Z)d\mathbf{f}\,,$$

These computations are infeasible. We have to use **approximate inference**.

Expectation Propagation [Minka 2001]

EP approximates the posterior $\mathcal{P}(\mathbf{f}|\mathcal{D}_{UV}, \mathcal{D}_Z)$ by a simpler distribution $\mathcal{Q}(\mathbf{f}) = \mathcal{N}(\mathbf{f}|\mathbf{m}, \mathbf{V})$, where

$$\mathcal{P}(\mathbf{f}|\mathcal{D}_{UV},\mathcal{D}_Z) \propto \overbrace{\boxed{\mathcal{P}(U_1, V_1|\tau = \sigma[f_1])}}^{q_1(f_1)} \cdots \overbrace{\boxed{\mathcal{P}(U_n, V_n|\tau = \sigma[f_n])}}^{q_n(f_n)} \boxed{\mathcal{N}(\mathbf{f}|\mathbf{0}, \mathbf{K})}$$

$$\Big\downarrow \text{EP} \qquad\qquad \Big\downarrow \text{EP} \qquad\qquad \Big\downarrow \text{Exact}$$

$$\mathcal{Q}(\mathbf{f}) \propto \underbrace{\boxed{k_1\mathcal{N}(f_1|\hat{m}_1, \hat{v}_1)}}_{\hat{q}_1(f_1)} \cdots \underbrace{\boxed{k_n\mathcal{N}(f_n|\hat{m}_n, \hat{v}_n)}}_{\hat{q}_n(f_n)} \boxed{\mathcal{N}(\mathbf{f}|\mathbf{0}, \mathbf{K})}$$

EP tunes $\hat{m}_i$ and $\hat{v}_i$ by minimizing $\mathrm{KL}[q_i(f_i)\mathcal{Q}(\mathbf{f})[\hat{q}_i(f_i)]^{-1}||\mathcal{Q}(\mathbf{f})]$, $i = 1, \ldots, n$.
Similar to EP for Gaussian process classification [Rasmussen and Williams 2006].

We fix the kernel length-scale by maximizing the EP approx. of $\mathcal{P}(\mathcal{D}_{UV}|\mathcal{D}_Z)$.
The total cost is $\mathcal{O}(n^3)$.

Experiments with Synthetic and Real-world Data

We choose a **Gaussian copula** for the likelihood function $\mathcal{P}(u, v|\tau = \sigma[f(z)])$.

Synthetic Data:

1000 points. Z sampled uniformly from $[0, \pi]$.
U and V sampled from a Gaussian copula with $\tau = 0.5\cos(3z)$.

Meteorological Data:

522 points. Atmospheric pressure (X), Temperature (Y) and Wind Velocity (Z).
Conditional marginals $F_{X|Z}$ and $F_{Y|Z}$ estimated using kernels [Hall et al. 2004].
These estimates are used to map X and Y into U and V.

Experimental Protocol:

Training sets with 100 observations.
40 repetitions. Test log-likelihood (**TLL**) used as a measure of performance.

Benchmark Methods:

- ⋆ Parametric Gaussian copula with no dependence on Z.
- ⋆ Non-parametric maximum local likelihood estimator of f [Acar et al. 2011].

Results

EP-CC: The EP method for the estimation of conditional copulas.
MLL: Non-parametric maximum local likelihood estimator.
GC: Parametric Gaussian copula with constant τ.

Dataset	Method	Avg. TLL	p-value[1]
Synthetic	**EP-CC**	**0.13±0.021**	-
	MLL	0.06±0.150	0.002
	GC	-0.01±0.017	$< 2.2 \cdot 10^{-16}$
Meteorological	**EP-CC**	**0.49±0.019**	-
	MLL	0.42±0.479	0.35
	GC	0.47±0.029	$7.03 \cdot 10^{-5}$

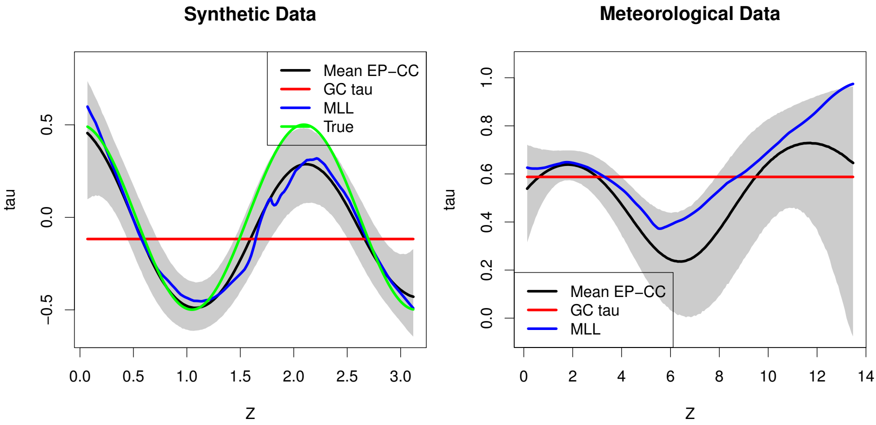
1: *p*-value of a paired Student's *t* test between EP-CC and the other methods.

EP-CC is **more robust** than MLL.

The meteorological data have a copula that **depends** on the wind velocity.

Plots of f for each Dataset

Predictions for f on each dataset when the training set contains 200 points.



Conclusions and Future Work

- The copula of a bivariate distribution may depend on a **covariate** Z.
  The **conditional copula** incorporates the effect of Z.
- We have proposed a **semi-parametric** model for conditional copulas.
  The model employs a **Gaussian process** prior.
  Expectation propagation can be used for efficient inference.
- The performance of this method (EP-CC) has been illustrated in experiments with **synthetic** and **meteorological** data.
- In these experiments, EP-CC is more **robust** and can perform better than other approaches that maximize the local likelihood.
- As future work, we propose to
  1. Evaluate EP-CC for **selecting** among different parametric copulas.
  2. Increase the number of **parameters** in the Gaussian process prior.
  3. Extend this method to **more than two dimensions**.

References

- A. J. Patton. Modelling Asymmetric Exchange Rate Dependence International, Economic Review, 2006, 47, 527-556.
- H. Joe. Multivariate models and dependence concepts. Chapman & Hall, 1997.
- T. Minka. A family of algorithms for approximate Bayesian inference. Massachusetts Institute of Technology, 2001.
- C. E. Rasmussen and C. K. I. Williams. Gaussian Processes for Machine Learning. MIT Press, 2006.
- P. Hall, J. S. Racine and Q. Li. Cross-Validation and the Estimation of Conditional Probability Densities, Journal of the American Statistical Association, 2004, 99, 1015-1026.
- E. F. Acar, R. V. Craiu and F. Yao. Dependence Calibration in Conditional Copulas: A Nonparametric Approach, Biometrics, 2011, 67, 445-453.

Thank you for your attention!