# Gaussian Process Vine Copulas for Multivariate Dependence

David Lopez-Paz[1,⋆], Jose Miguel Hernández-Lobato[2,⋆], Zoubin Ghahramani[2]

[1]Max Planck Institute for Intelligent Systems; [2]University of Cambridge; ⋆Equal contributors.

## Abstract

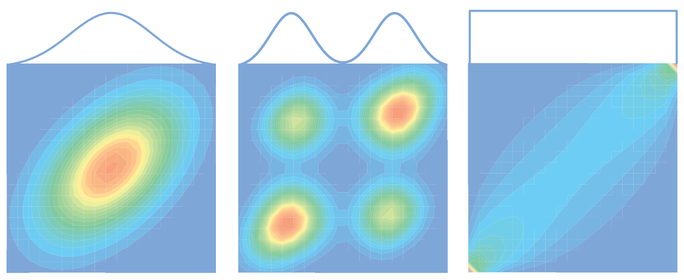
*Vine factorizations* ease the learning of high-dimensional densities by constructing a hierarchy of parametric conditional bivariate copulas. However, to simplify inference, it is common to assume that each of these conditional bivariate copulas is independent from its conditioning variables. In this work, we relax this assumption by discovering the latent functions that specify the shape of a parametric conditional copula given its conditioning variables. We follow a Bayesian approach based on sparse Gaussian processes with expectation propagation for scalable, approximate inference.

## Copulas

The copula $c(\mathbf{u})$ of a given density $p(\mathbf{x})$ describes the dependencies among the rr.vv. $x_1, \ldots, x_d$, but contains no information about their marginal distributions, since $P_X(X) \sim U(0,1)$ for any r.v. $X$ [8]:

$$p(\mathbf{x}) = \left[\prod_{i=1}^{d} p_i(x_i)\right] \underbrace{c(P(x_1), ..., P(x_d))}_{\text{copula}}.$$

1. Copulas separate the learning of univariate marginal distributions from the learning of their multivariate dependence [5]. Different multivariate models (left, middle) may share the same copula function (right):



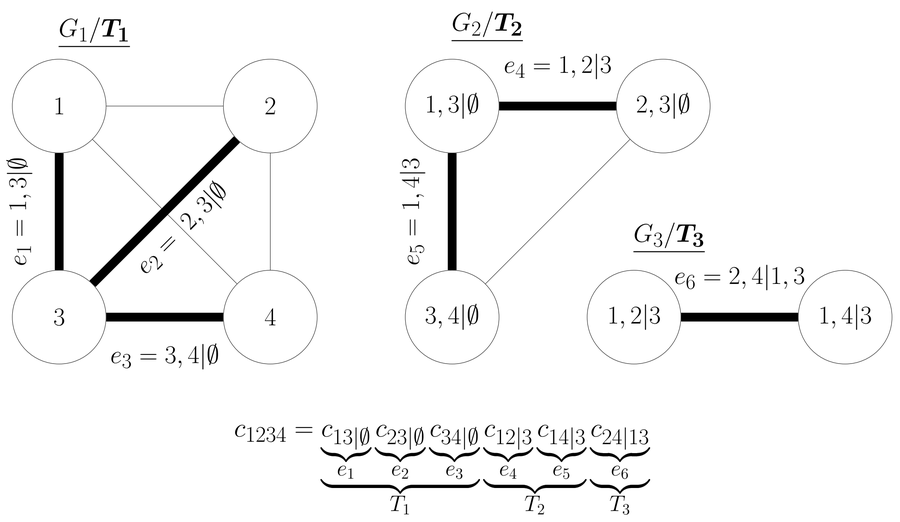
2. There exist many parametric models for two-dimensional copulas. However, for more than two dimensions, the number and expressiveness of families of parametric copulas is more limited. A solution to this problem is given by the so-called *vine copula distributions* [4, 3, 6].

## Regular Vine Copula Distributions

High-dimensional copulas can be factorized as a collection of conditional bivariate copulas, in a so-called Regular Vine Copula Distribution or Decomposition [1]:

$$p(\mathbf{x}) = \prod_{i=1}^{d} p_i(x_i) \prod_{i=1}^{d-1} \prod_{e(j,k)\in E_i} c_{jk|D(e)}(P_{j|D(e)}(x_j|D(e)), P_{k|D(e)}(x_k|D(e))) \quad (1)$$

These models are formed by a collection of trees, in which each edge represents a bivariate copula:




$$c_{1234} = \underbrace{c_{13|\emptyset}}_{e_1}\underbrace{c_{23|\emptyset}}_{e_2}\underbrace{c_{34|\emptyset}}_{e_3}\underbrace{c_{12|3}}_{e_4}\underbrace{c_{14|3}}_{e_5}\underbrace{c_{24|13}}_{e_6}$$
$$T_1 \qquad\qquad T_2 \qquad T_3$$

### Regular Vine Details

1. The deeper a bivariate copula is located into the vine hierarchy, the more variables it will be conditioned on.
2. Each bivariate copula forming the vine can belong to a different parametric family.
3. The election of each spanning tree is done by maximizing the sum of the absolute empirical Kendall's $\tau$'s associated with each edge.
4. Conditional c.d.f.s at tree $i$ are expressed as partial derivatives of copulas from tree $i-1$ [1]:

$$P(u|\mathbf{v}) = \frac{\partial\, C_{u,v_j|\mathbf{v}_{-j}}(P_{u|\mathbf{v}_{-j}}(u|\mathbf{v}_{-j}), P_{v_j|\mathbf{v}_{-j}}(v_j|\mathbf{v}_{-j}))}{\partial P_{v_j|\mathbf{v}_{-j}}(v_j|\mathbf{v}_{-j})}. \quad (2)$$

5. **To simplify inference, it is often assumed that the bivariate conditional copulas are independent from their conditioning variables.**
6. **We relax the latter assumption by discovering the latent functions that specify the shape of the parametric conditional copula given its conditioning variables.**

## Estimation of Bivariate Parametric Conditional Copulas

Most parametric bivariate copulas can be specificied in terms of their rank correlation parameter $\tau \in [-1, 1]$ (Kendall's $\tau$). The dependence of the copula on a vector of variables $\mathbf{z} = (z_1, \ldots, z_d)^{\mathrm{T}}$ is captured by specifying the relationship $\tau = \sigma(f(\mathbf{z}))$, where $\sigma(x) = 2(\Phi(x) - 1)$, $\Phi$ is the cdf of a standard Gaussian distribution and $f$ is a non-linear function.

Let $\mathcal{D} = \{\mathcal{D}_{u,v} = \{u_i, v_i)\}_{i=1}^n, \mathcal{D}_{\mathbf{z}} = \{\mathbf{z}_i\}_{i=1}^n\}$ be a sample from $C_{u,v|\mathbf{z}}$ where $\mathcal{D}_{\mathbf{z}}$ are the values of $\mathbf{z}$ that were used to generate $\mathcal{D}_{u,v}$. We place a Gaussian process prior on $f$ and compute the posterior of this function given $\mathcal{D}$.

Let $\mathbf{f} = (f(\mathbf{z}_1), \ldots, f(\mathbf{z}_n))^{\mathrm{T}}$ and $p(\mathbf{f}|\mathcal{D}_{\mathbf{z}}) = \mathcal{N}(\mathbf{f}|\mathbf{m}, \mathbf{K})$. The posterior for $\mathbf{f}$ given $\mathcal{D}$ is

$$p(\mathbf{f}|\mathcal{D}_{u,v}, \mathcal{D}_{\mathbf{z}}) = \frac{p(\mathcal{D}_{u,v}|\mathbf{f})p(\mathbf{f}|\mathcal{D}_{\mathbf{z}})}{p(\mathcal{D}_{u,v}|\mathcal{D}_{\mathbf{z}})}, \quad (3)$$

where $p(\mathcal{D}_{u,v}|\mathbf{f}) = \prod_{i=1}^n c(u_i, v_i|\tau = 2\Phi(f_i) - 1)$.

Given $\mathbf{z}^\star$, we make predictions using

$$p(u^\star, v^\star|\mathbf{z}^\star) = \int c(u^\star, v^\star|\tau = 2\Phi(f^\star) - 1)p(f^\star|\mathbf{f}, \mathbf{z}^\star, \mathcal{D}_{\mathbf{z}})p(\mathbf{f}|\mathcal{D}_{u,v}, \mathcal{D}_{\mathbf{z}})\, d\mathbf{f}\, df^\star, \quad (4)$$

$p(f^\star|\mathbf{f}, \mathbf{z}^\star, \mathcal{D}_{\mathbf{z}}) = \mathcal{N}(f^\star|\mathbf{k}^{\mathrm{T}}\mathbf{K}^{-1}\mathbf{f}, k - \mathbf{k}^{\mathrm{T}}\mathbf{K}^{-1}\mathbf{k})$, $\mathbf{k} = (\mathrm{Cov}(f(\mathbf{z}^\star), f(\mathbf{z}_1)), \ldots, \mathrm{Cov}(f(\mathbf{z}^\star), f(\mathbf{z}_n)))^{\mathrm{T}}$ and $k = \mathrm{Cov}(f(\mathbf{z}^\star), f(\mathbf{z}^\star))$.

Approximate inference is performed using expectation propagation [7]. To reduce computational costs, we use the FITC approximation for Gaussian Processes [9].

### Related Work: The Maximum Local-Likelihood Method (MLLVINE)

Acar, Genest and Neslehova [2] approximate linearly $f$ at any point $z$ where it needs to be evaluated. They do so by solving the optimization problem:

$$(b_0^\star, b_1^\star) = \arg\max_{(b_0,b_1)} \left\{ \sum_{i=0}^{N} k_h(z - z_i) \log c(u_i, v_i|\tau = 2\Phi[b_0 + b_1(z - z_i)] - 1) \right\}, \quad (5)$$

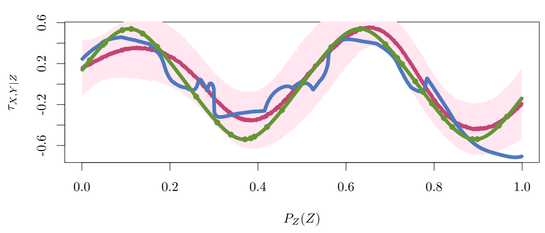
where the neighborhood of $z$ is determined by the Epanechnikov kernel $k_h(x) = \frac{3}{4h}\max(0, 1 - (\frac{x}{h})^2)$ of bandwidth $h$. An estimate of $f(z)$ is then obtained as $f(z) \approx b_0^\star$.

Some disadvantages of this method are:

1. It can only condition on a single scalar variable.
2. The optimization problem (5) needs to be solved for each prediction.
3. Since it is a local-based method, it can lead to poor performance when the data is sparse.

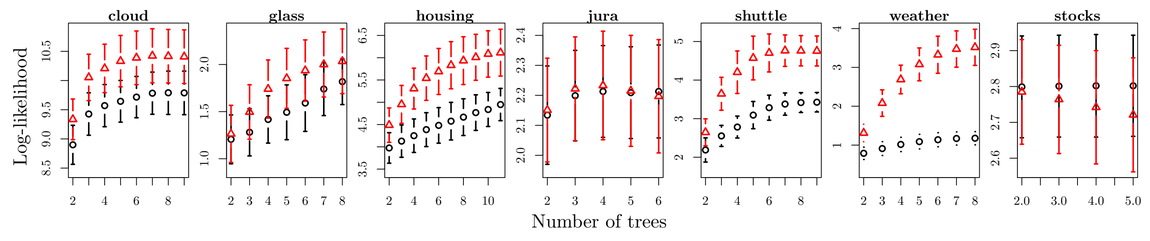
## Experimental Results

### Synthetic Data



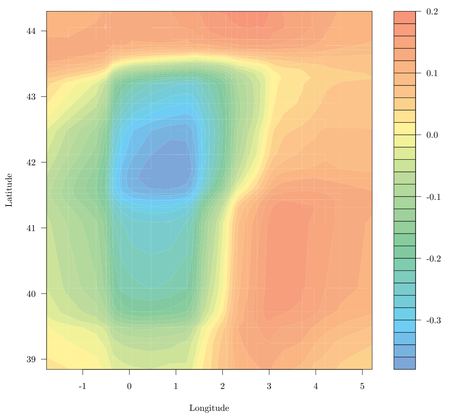

We consider a two-dimensional Gaussian, in which its two random variables $(X, Y)$ have correlation $\rho(Z) = 3/4\,\sin(Z)$. **Green:** Kendall's $\tau$ correlation function of the distribution, with 30 training points. **Blue:** MLLVINE approximation. **Red**: GPVINE approximation, including the uncertainty in the prediction of the Gaussian process.

### Real-World Data




Test log-likelihoods achieved by the proposed method (GPVINE, red triangles) and the simplified vine model (SVINE, red circles), as the number of trees forming the models increase.

| Data | SVINE | MLLVINE | GPVINE |
|---|---|---|---|
| Synthetic | $-0.005 \pm 0.012$ | $0.101 \pm 0.162$ | $\mathbf{0.298 \pm 0.031}$ |
| Uranium | $0.006 \pm 0.006$ | $0.016 \pm 0.026$ | $\mathbf{0.022 \pm 0.012}$ |
| Cloud | $8.899 \pm 0.334$ | $9.013 \pm 0.600$ | $\mathbf{9.335 \pm 0.348}$ |
| Glass | $1.206 \pm 0.259$ | $0.460 \pm 1.996$ | $\mathbf{1.264 \pm 0.303}$ |
| Housing | $3.975 \pm 0.342$ | $4.246 \pm 0.480$ | $\mathbf{4.487 \pm 0.386}$ |
| Jura | $2.134 \pm 0.164$ | $2.125 \pm 0.177$ | $\mathbf{2.151 \pm 0.173}$ |
| Shuttle | $2.552 \pm 0.273$ | $2.256 \pm 0.612$ | $\mathbf{3.645 \pm 0.427}$ |
| Weather | $0.789 \pm 0.159$ | $0.771 \pm 0.890$ | $\mathbf{1.312 \pm 0.227}$ |
| Stocks | $\mathbf{2.802 \pm 0.141}$ | $2.739 \pm 0.155$ | $2.785 \pm 0.146$ |




**Left:** Test log-likelihoods achieved by GPVINE, SVINE and MLLVINE, when using two trees. **Right:** Kendall's $\tau$ correlation between *atmospheric pressure* and *cloud percentage cover* (color scale) when conditioned to *longitude* and *latitude*, for the dataset *Weather*. The blue region in the plot corresponds to the Pyrenee mountains.

## References

[1] K. Aas, C. Czado, A. Frigessi, and H. Bakken. Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics*, 44(2):182–198, 2006.

[2] E. F. Acar, C. Genest, and J. Neslehova. Beyond simplified pair-copula constructions. *Journal of Multivariate Analysis*, 110:74–90, 2012.

[3] T. Bedford and R. M. Cooke. Vines–a new graphical model for dependent random variables. *The Annals of Statistics*, 30(4):1031–1068, 2002.

[4] H. Joe. Families of $m$-variate distributions with given margins and $m(m-1)/2$ bivariate dependence parameters. *Distributions with Fixed Marginals and Related Topics*, 1996.

[5] H. Joe. Asymptotic efficiency of the two-stage estimation method for copula-based models. *Journal of Multivariate Analysis*, 94(2):401–419, 2005.

[6] D. Kurowicka and R. Cooke. *Uncertainty Analysis with High Dimensional Dependence Modelling*. Wiley Series in Probability and Statistics, 1st edition, 2006.

[7] T. P. Minka. Expectation Propagation for approximate Bayesian inference. *Proceedings of the 17th Conference in Uncertainty in Artificial Intelligence*, pages 362–369, 2001.

[8] R. Nelsen. *An Introduction to Copulas*. Springer Series in Statistics, 2006.

[9] E. Snelson and Z. Ghahramani. Sparse Gaussian processes using pseudo-inputs. *Proceedings of the 20th Conference in Advances in Neural Information Processing Systems*, pages 1257–1264, 2006.