

# Dependent Dirichlet Processes and Copulas

Radu V. Craiu<sup>\*</sup> and Luis E. Nieto-Barajas<sup>†</sup>

## Introduction

We congratulate Professor MacEachern on the publication of his paper [10] that has shaped a quarter century of Bayesian nonparametrics while circulating as a preprint. Its influence is visible in the ANOVA-DDP of De Iorio et al. [5], in the support results of Barrientos et al. [3], and in the many dependent processes built for time series and longitudinal data [13, 14, 15, among others]. Our discussion focuses on the connections between copula models and the dependent Dirichlet process (DDP). We believe this can potentially guide the application of DDP and its extensions.

## Connections between DDP and copulas

MacEachern (2026, Definition 2.3) builds the location process from a stochastic process  $Z_{\mathcal{X}}$  through

$$\theta_{ix} = F_{0,x}^{-1}(G_x(Z_{ix})), \quad x \in \mathcal{X}, \quad (1)$$

where  $G_x$  is the continuous marginal distribution of  $Z_x$ . Fix distinct covariate levels  $x_1, \dots, x_d$ . The map (1) is increasing in each coordinate, so the joint distribution of  $(\theta_{ix_1}, \dots, \theta_{ix_d})$  has margins  $F_{0,x_1}, \dots, F_{0,x_d}$  and copula equal to the copula of  $(Z_{x_1}, \dots, Z_{x_d})$ , which we denote  $C_{x_1, \dots, x_d}^Z$ . By Sklar [21],

$$\Pr(\theta_{ix_1} \leq t_1, \dots, \theta_{ix_d} \leq t_d) = C_{x_1, \dots, x_d}^Z(F_{0,x_1}(t_1), \dots, F_{0,x_d}(t_d)), \quad (2)$$

and (1) is the constructive half of Sklar's theorem applied to the finite-dimensional laws of a process. MacEachern (2026, Section 4) takes  $Z_{\mathcal{X}}$  to be a Gaussian process with exponential variogram, so that  $C_{x,x'}^Z$  is the Gaussian copula with correlation  $\rho(x, x')$ .

The Gaussian copula has zero tail dependence: for  $\rho < 1$ ,

$$\lambda_L = \lim_{u \downarrow 0} \frac{C_{x,x'}^Z(u, u)}{u} = 0, \quad \lambda_U = \lim_{u \uparrow 1} \frac{1 - 2u + C_{x,x'}^Z(u, u)}{1 - u} = 0.$$

Translated through (2), atoms that fall in the tails of  $F_{0,x}$  are asymptotically independent of the corresponding atoms at  $x'$ , however close  $x'$  is to  $x$  and however large  $\rho(x, x')$  is short of one. A possible implication is that while  $F_x$  and  $F_{x'}$  may agree closely in their central regions, they are potentially free to disagree in their tails. For the church data of MacEachern (2026, Section 4) this is innocuous. For the uses the paper anticipates in its conclusion, in particular for regression aimed at predictive quantiles far from the center, it is consequential since the DDP with a Gaussian  $Z_{\mathcal{X}}$  may let the extremes drift apart.

---

arXiv: 2010.00000

<sup>\*</sup>Department of Statistical Sciences, University of Toronto, [radu.craiu@utoronto.ca](mailto:radu.craiu@utoronto.ca)

<sup>†</sup>Department of Statistics, ITAM, Mexico City, [luis.nieto@itam.mx](mailto:luis.nieto@itam.mx)

Alternatives are available within the paper’s framework. MacEachern (2026, Theorems 3.1, 3.3 and 3.6) use separability, absolute continuity of the finite-dimensional laws and continuity of sample paths; Gaussianity enters only through the choice made in MacEachern (2026, Section 4). Any process with continuous margins and continuous paths may replace it in (1), and the copula it induces characterizes the dependence. Zero tail dependence is a property of the Gaussian copula alone within the elliptical class. The simplest alternative is a *Student*–*t* process with  $\nu$  degrees of freedom, whose finite-dimensional copulas are *t* copulas [6] with

$$\lambda_L = \lambda_U = 2t_{\nu+1}\left(-\sqrt{\frac{(\nu+1)(1-\rho)}{1+\rho}}\right),$$

where  $t_\nu(\cdot)$  is the cdf of a Student-*t* with  $\nu$  df.  $\lambda_L$  is strictly positive for every  $\nu < \infty$  and every  $\rho > -1$ , and tends to zero only in the Gaussian limit  $\nu \rightarrow \infty$ . At the correlation  $\rho = 0.8$  used in MacEachern (2026, Section 4), the coefficient equals 0.54, 0.45 and 0.29 for  $\nu = 3, 5, 10$ . Extreme atoms at neighbouring  $x$  then move together with a strength that  $\nu$  controls and that the practitioner can elicit, and the Gaussian DDP is recovered as a limiting case. A max-stable process gives an extreme-value copula, the Hüsler–Reiss family being the closest relative of the Gaussian case [8], and is the natural choice when the DDP is used to model distributions of maxima indexed by a covariate. The construction of Wilson and Ghahramani [22], who call this device a copula process, shows that computation with non-Gaussian  $Z_{\mathcal{X}}$  is practical. The same considerations apply within a  $p$ -dimensional atom, where the paper couples coordinates through “collections of stochastic processes”: the copula linking  $\theta_{x,1}, \dots, \theta_{x,p}$  at a fixed  $x$  determines whether large values of the coordinates of a location co-occur, and a Gaussian choice again rules this out.

A reader used to Archimedean copulas will ask whether the Clayton copula, with  $\lambda_L = 2^{-1/\theta}$ , can play the same role, and the answer separates the two cases of  $\mathcal{X}$  that the paper treats. Every Archimedean copula has a frailty representation [11, 12]: with  $E_1, E_2, \dots$  independent unit exponentials,  $V$  a positive random variable with Laplace transform  $\psi$ , and  $U_j = \psi(E_j/V)$ , the vector  $(U_1, \dots, U_d)$  has the Archimedean copula with inverse generator  $\psi$ ; Clayton corresponds to  $V \sim \text{Gamma}(1/\theta, 1)$  and  $\psi(t) = (1+t)^{-1/\theta}$ . Setting  $Z_{ix} = \psi(E_{ix}/V_i)$  in (1) therefore does induce a Clayton coupling of the atoms across  $x$ , with lower tail dependence  $2^{-1/\theta}$  between every pair of covariate levels and a strictly positive density on  $(0, 1)^d$ , so the support conditions of MacEachern (2026, Theorem 3.3) hold. The price is exchangeability: a common frailty gives the same dependence to every pair  $(x, x')$ , whatever their distance. For the finite  $\mathcal{X}$  of the multi-centre trial in MacEachern (2026, Section 2) that is acceptable, and a nested Archimedean copula [7] with a stronger Clayton within groups of centres and a weaker one between groups gives a tree-structured, still tail-dependent, alternative. For an interval  $\mathcal{X}$ , property 4 in MacEachern (2026, Section 1) imposes that the copula of  $(\theta_x, \theta_{x_0})$  must tend to  $M$ , the Fréchet–Hoeffding upper bound, as  $x \rightarrow x_0$ , and away from it as  $|x - x_0|$  grows, and no single Archimedean copula does both. The natural attempt is to let the frailty itself be a process,  $Z_{ix} = \psi(E_{ix}/V_{ix})$  with  $V_{\mathcal{X}}$  a positive Markov process such as the gamma process of Nieto-Barajas and Walker [16]; the resulting

copula is positively dependent and decays with distance, but whether it retains tail dependence depends on whether small values of  $V_x$  force small values of  $V_{x'}$ , which the usual innovation-based constructions do not guarantee. We regard the existence of a covariate-indexed process with distance-decaying, tail-dependent, Archimedean-type finite-dimensional copulas as an open question raised by the paper, while the elliptical and max-stable routes above are available today.

### DDP helps with flexible copula modelling

We end our discussion by emphasizing that the DDP can help enlarge the scope of copula models. For instance, the mixture of Archimedean copulas of Pan et al. [17] uses a Pitman–Yor process [19] on the copula parameter with atoms that ignore covariates.

Conditional copula models are designed to allow the dependence patterns to vary with covariates, as seen in [1, 2, 4, 9, 18, 20], among others. Replacing the atoms  $\theta_k$  by processes  $\theta_{k,x}$  through (1), with  $F_{0,x}$  the centering measure on the copula parameter space, gives a covariate-dependent mixture copula  $C_x(\mathbf{u}) = \sum_k \omega_k C(\mathbf{u} \mid \theta_{k,x})$  in which clusters are shared across covariate levels while the strength, and for the Frank family the sign, of dependence evolves with  $x$ . The single- $p$  version keeps the Pólya urn computations of Pan et al. [17] intact. We regard this as a direct gain brought by [10], and we are excited by the opportunities it opens.

## References

- [1] Acar, E. F., Craiu, R. V., and Yao, F. (2011). “Dependence calibration in conditional copulas: a nonparametric approach.” *Biometrics*, 67: 445–453. [3](#)
- [2] — (2013). “Statistical testing of covariate effects in conditional copula models.” *Electronic Journal of Statistics*, 7: 2822–2850. [3](#)
- [3] Barrientos, A. F., Jara, A., and Quintana, F. A. (2012). “On the support of MacEachern’s dependent Dirichlet processes and extensions.” *Bayesian Analysis*, 7: 277–310. [1](#)
- [4] Craiu, R. V. and Sabeti, A. (2012). “In mixed company: Bayesian inference for bivariate conditional copula models with discrete and continuous outcomes.” *Journal of Multivariate Analysis*, 110: 106–120. [3](#)
- [5] De Iorio, M., Müller, P., Rosner, G. L., and MacEachern, S. N. (2004). “An ANOVA model for dependent random measures.” *Journal of the American Statistical Association*, 99: 205–215. [1](#)
- [6] Demarta, S. and McNeil, A. J. (2005). “The  $t$  copula and related copulas.” *International Statistical Review*, 73: 111–129. [2](#)
- [7] Hofert, M. and Pham, D. (2013). “Densities of nested Archimedean copulas.” *Journal of Multivariate Analysis*, 118: 37–52. [2](#)
- [8] Hüsler, J. and Reiss, R.-D. (1989). “Maxima of normal random vectors: between independence and complete dependence.” *Statistics & Probability Letters*, 7: 283–286. [2](#)

- [9] Levi, E. and Craiu, R. V. (2018). “Bayesian inference for conditional copulas using Gaussian process single index models.” *Computational Statistics and Data Analysis*, 122: 115–134. [3](#)
- [10] MacEachern, S. N. (2026). “Dependent Dirichlet Processes (with Discussion).” *Bayesian Analysis*, doi:10.1214/26-BA1610. [1](#), [3](#)
- [11] Marshall, A. W. and Olkin, I. (1988). “Families of multivariate distributions.” *Journal of the American Statistical Association*, 83: 834–841. [2](#)
- [12] McNeil, A. J. and Nešlehová, J. (2009). “Multivariate Archimedean copulas,  $d$ -monotone functions and  $\ell_1$ -norm symmetric distributions.” *Annals of Statistics*, 37: 3059–3097. [2](#)
- [13] Nieto-Barajas, L. E. (2021). “A class of dependent Dirichlet processes via latent multinomial processes.” *Statistics*, 55: 1169–1179. [1](#)
- [14] Nieto-Barajas, L. E. and Contreras-Cristán, A. (2014). “A Bayesian nonparametric approach for time series clustering.” *Bayesian Analysis*, 9: 147–170. [1](#)
- [15] Nieto-Barajas, L. E., Müller, P., Ji, Y., Lu, Y., and Mills, G. B. (2012). “A time-series DDP for functional proteomics profiles.” *Biometrics*, 68: 859–868. [1](#)
- [16] Nieto-Barajas, L. E. and Walker, S. G. (2002). “Markov beta and gamma processes for modelling hazard rates.” *Scandinavian Journal of Statistics*, 29: 413–424. [2](#)
- [17] Pan, R., Nieto-Barajas, L. E., and Craiu, R. V. (2025). “Bayesian nonparametric mixtures of Archimedean copulas.” *Journal of Agricultural, Biological and Environmental Statistics*. DOI: 10.1007/s13253-025-00703-8. [3](#)
- [18] Patton, A. J. (2006). “Modelling asymmetric exchange rate dependence.” *International Economic Review*, 47: 527–556. [3](#)
- [19] Pitman, J. and Yor, M. (1997). “The two-parameter Poisson–Dirichlet distribution derived from a stable subordinator.” *Annals of Probability*, 25: 855–900. [3](#)
- [20] Sabeti, A., Wei, M., and Craiu, R. V. (2014). “Additive models for conditional copulas.” *Stat*, 3: 300–312. [3](#)
- [21] Sklar, A. (1959). “Fonctions de répartition à  $n$  dimensions et leurs marges.” *Publications de l’Institut de Statistique de l’Université de Paris*, 8: 229–231. [1](#)
- [22] Wilson, A. G. and Ghahramani, Z. (2010). “Copula processes.” In *Advances in Neural Information Processing Systems 23*, 2460–2468. [2](#)