Copula Bayesian Networks

Gal Elidan
Department of Statistics
Hebrew University
Jerusalem, 91905, Israel
galel@huji.ac.il

Multivariate real-valued distributions are of paramount importance in a variety of fields ranging from computational biology and neuro-science to economics to climatology. Choosing and estimating a useful form for the marginal distribution of each variable in the domain is often a straightforward task. In contrast, aside from the normal representation, few univariate distributions have a convenient multivariate generalization. Indeed, modeling and estimation of flexible (skewed, multi-modal, heavy tailed) high-dimensional distributions is still a formidable challenge. In this work we present a novel multivariate density model that is a marriage of the copula and Bayesian networks frameworks. Our construction offers great flexibility in modeling high dimensional distributions and results in consistent generalization advantages in varied domains. In addition, our model gives rise to an efficient mean-field like approximate inference procedure, facilitating practical structure learning in non-linear domains.

Copulas [11, 15] are functions that link given (or estimated) univariate marginals into a joint distribution. This allows us to robustly estimate marginals (e.g. using a non-parametric approach), and then use only few parameters to capture the dependencies. The resulting model is typically less prone to over-fitting than a fully non-parametric one, while at the same time avoiding the limitations of a fully parameterized distribution. This, in turn, often leads to significant generalization benefits. Accordingly, interest in copulas has grown rapidly with applications ranging from mainstream financial risk assessment (e.g., Embrechts et al. [5]) to off-shore engineering (e.g., Accioly and Chiyoshi [2]). Unfortunately, constructing high-dimensional copulas is difficult and, despite many innovations (e.g., [1, 3]), applications are almost always limited to few (< 10) variables, or build on specific structures or copulas (e.g., [8, 10]).

Bayesian networks (BNs) [13] offer a markedly different approach for density estimation that relies on a graph that encodes independencies which imply a decomposition of the joint density. This facilitates efficient computations, making the framework amenable to high-dimensional domains. However, the expressiveness of BNs is hampered by practical considerations that almost always lead to simple parametric forms. Specifically, non-parametric variants of BNs (e.g., [6, 14]) involve elaborate training setups with a running time that grows unfavorably with the number of samples and local graph connectivity. Furthermore, aside from the case of the normal distribution, the form of the univariate marginal is neither under control nor is it typically known.

Topic: learning algorithms, graphical models
Preference: oral

We present Copula Bayesian Networks, a multivariate real-valued distribution model that combines the strengths of both worlds.[1] As in BNs, we make use of a graph to encode independencies. Differently, we rely on local copula functions along with an explicit *g*lobally shared parameterization of the univariate densities. At the heart of our approach is a novel reparameterization of a conditional density using a copula quotient. With this construction, we prove a parallel to the BN composition and decomposition theorems for joint copulas. The result is a multivariate model that facilitates highly flexible structure and parameter learning of non-linear high-dimensional models. Importantly, as we demonstrate for three varied real-life domains, our model leads to consistent and significant generalization benefits.

The unique characteristics of our construction also give rise to a novel variational-like inference procedure that is significantly more efficient than the similar mean-field method [7]. This allow us to tackle the computationally daunting task of structure learning in the face of partial information. Concretely, we learn Copula Bayesian Networks that generalize well in domains where learning even simple non-linear Bayesian networks is computationally prohibitive.

References

[1] K. Aas, C. Czado, A. Frigessi, and H. Bakken. Pair-copula constructions of multiple dependencies. *Insurance: Mathematics and Economics*, 44:182–198, 2009.

[2] R. Accioly and F. Chiyoshi. Modeling dependence with copulas: a useful tool for field development decision process. *J of Petroleum Science and Engineering*, 44:83–91, 2004.

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

[4] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*, B 39:1–39, 1977.

[5] P. Embrechts, F. Lindskog, and A. McNeil. Modeling dependence with copulas and applications to risk management. *Handbook of Heavy Tailed Distributions in Finance*, 2003.

[6] N. Friedman and I. Nachman. Gaussian Process Networks. In *UAI*, 2000.

[7] M. Jordan, Z. Ghahramani, T. Jaakkola, and L. Saul. An introduction to variational approximations methods for graphical models. In *Learning in Graphical Models*, 1998.

[8] S. Kirshner. Learning with tree-averaged densities and distributions. In *NIPS*, 2007.

[9] D. Koller and N. Friedman. *Probabilistic Graphical Models: Principles and Techniques*. The MIT Press, 2009.

[10] H. Liu, J. Lafferty, and L. Wasserman. The nonparanormal: Semiparametric estimation of high dimensional undirected graphs. *Journal of Machine Learning Research*, 2010.

[11] R. Nelsen. *An Introduction to Copulas*. Springer, 2007.

[12] E. Parzen. On estimation of a probability density function and mode. *Annals of Mathematical Statistics*, 33:1065–1076, 1962.

[13] J. Pearl. *Probabilistic Reasoning in Intelligent Systems*. Morgan Kaufmann, 1988.

[14] A. Schwaighofer, M. Dejori, V. Tresp, and M. Stetter. Structure Learning with Nonparametric Decomposable Models. In *International Conf. on Artificial Neural Networks*, 2007.

[15] A. Sklar. Fonctions de repartition a n dimensions et leurs marges. *Publications de l'Institut de Statistique de L'Universite de Paris*, 8:229–231, 1959.

[1] This abstracts subsumes the author's two publications on the topic in the last UAI and NIPS conferences