Vines in Prose

Roger M. Cooke

5-22-2015

(Based on Cooke, Roger M., Joe, H. and Aas, K. (2010) Vines Arise, in Kurowicka and Joe (eds) Dependence Modeling: Handbook on Vine Copulae, World Scientific, Singapore, 978-981-4299-87-9, 981-4299-87-1, pp43-84.)

Introduction

A vine is a graphical tool for labeling constraints in high-dimensional distributions. A regular vine is a special case for which all constraints are two-dimensional or conditional two-dimensional. Regular vines generalize trees, and are themselves specializations of Cantor trees (Bedford and Cooke 2002) . Combined with bivariate copulae, regular vines have proven to be a flexible tool in high-dimensional dependence modeling. Copulae (Joe 1997, Nelson 2006) are multivariate distributions with uniform univariate margins. Representing a joint distribution as univariate margins plus copulae allows us to separate the problems of estimating univariate distributions from problems of estimating dependence. This is handy in as much as univariate distributions in many cases can be adequately estimated from data, whereas dependence information is rough hewn, involving summary indicators and judgment (Kraan and Cooke 2000, Ale et al 2009). Although the number of parametric multivariate copula families with flexible dependence is limited, there are many parametric families of bivariate copulae. Regular vines owe their increasing popularity to the fact that they leverage from bivariate copulae and enable extensions to arbitrary dimensions. Sampling theory and estimation theory for regular vines are well developed (Kurowicka and Cooke 2007, Aas et al 2009), and model inference has left the post (Kurowicka and Cook 2006, Kurowicka et al 2007, Aas et al 2009). Regular vines have proven useful in other problems such as (constrained) sampling of correlation matrices (Joe 2005, Lewandowski 2008, Lewandowski et al 2009), building non-parametric continuous Bayesian belief nets (Hanea 2008, Hanea et al 2010), and characterizing the set of rank correlation matrices (Joe 2006) . Software for estimating and sampling regular vines, literature and event notices are available at http://www.statistics.ma.tum.de/en/research/vine-copulamodels. Recent reference publications are (Kurowicka and Joe 2010 and Joe 2014).

Historical Origins

The first regular vine, avant la lettre, was introduced by (Joe 1994). The motive was to extend the bivariate extreme value copula to higher dimensions. To this end he introduced what would later be called the *D-vine*. Joe (Joe 1996) was interested in a class of n-variate distributions with given one dimensional margins, and $n(n-1)$ dependence parameters, whereby $n-1$ parameters correspond to bivariate margins, and the others correspond to conditional bivariate margins. In the case of multivariate normal distributions, the parameters would be $n-1$ correlations and $(n-1)(n-2)/2$ partial correlations, which were noted to be algebraically independent in $(-1, 1)$. Implicit in this remark is the observation that partial correlations on what is now called D-vine provide an algebraically independent parametrization of the set of positive definite correlation matrices.

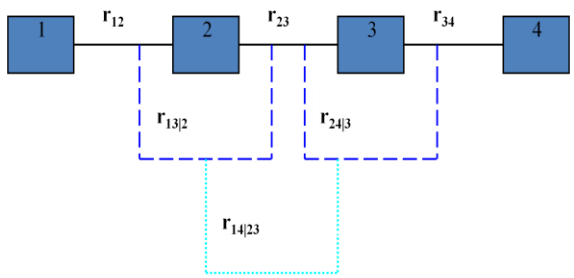
An entirely different motivation underlay the first formal definition of vines in (Cooke 1997). Uncertainty analyses of large risk models, such as those undertaken for the European Union and the US Nuclear Regulatory Commission for accidents at nuclear power plants, involve quantifying and propagating uncertainty over hundreds of variables (Goossens et al 2000, Harper et al 1994) . Dependence information for such studies had been captured with *Markov trees*, (Whittaker, 1990) which are trees constructed with nodes as univariate random variables and edges as bivariate copulae. For n variables, there are at most $n-1$ edges for which dependence can be specified. New techniques at that time involved obtaining uncertainty distributions on modeling parameters by eliciting experts' uncertainties on other variables which are predicted by the models. These uncertainty distributions are pulled back onto the model's parameters by a process known as probabilistic inversion (Goossens et al 2000, Kurowicka and Cooke 2006) The resulting distributions often displayed a dependence structure that could not be captured as a Markov tree. Graphical models called *vines* were introduced in (Cooke 1997 , Bedford and Cooke 2002 and Kurowicka and Cooke 2006).

Regular Vines (R-Vines)

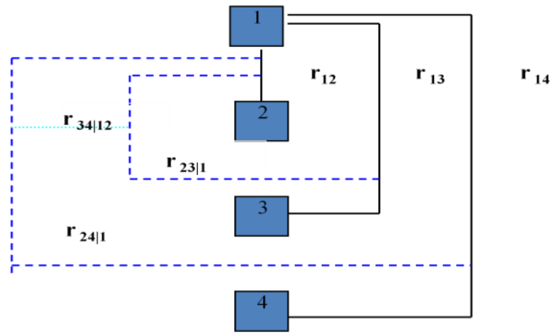
A *vine V* on n variables is a nested set of connected trees where the edges in the first tree are the nodes of the second tree, the edges of the second tree are the nodes of the third tree, etc. A *regular vine* or *R-vine* on n variables is a vine in which two edges in tree j are joined by an edge in tree $j+1$ only if these edges share a common node, $j = 1, \ldots, n-2$. The nodes in the first tree are univariate random variables. The edges are constraints or conditional constraints explained as follows.

Recall that an edge in a tree is an unordered set of two nodes. Each edge in a vine is associated with a *constraint set*, being the set of variables (nodes in first tree) reachable by the set membership relation. For each edge, the constraint set is the union of the constraint sets of the edge's two members called its component constraint sets (for an edge in the first tree, the component constraint sets are empty). The constraint associated with each edge is now the symmetric difference of its component constraint sets conditional on the intersection of these sets. One can show that for a regular vine, the symmetric difference of the component constraint sets is always a doubleton and that each pair of variables occurs exactly once as constrained variables. In other words, all constraints are bivariate or conditional bivariate constraints.

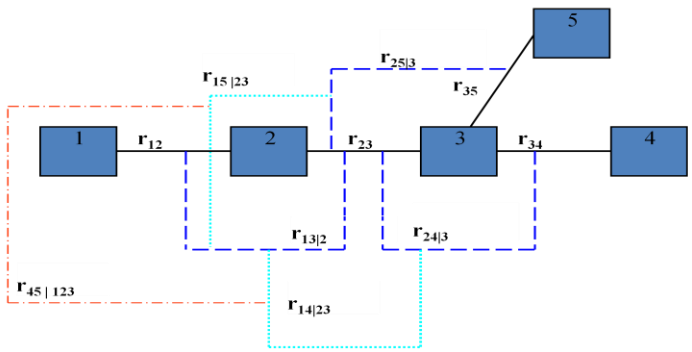
The degree of a node is the number of edges attaching to it. The simplest regular vines have the simplest degree structure; the D-Vine assigns every node degree 1 or 2, the C-Vine assigns one node in each tree the maximal degree. The following figure shows a C and D vine on 4 variables with constraints, and a regular vine on five variables which is neither. For larger vines, other graphical representations are employed.



D-Vine on 4 variables



C-Vine on 4 variables



Regular vine on 5 variables

The number of regular vines on *n* variables grows rapidly in *n*: there 2^{n-3} ways of extending a regular vine with one additional variable, and there are $n(n-1)(n-2)!2^{(n-2)(n-3)/2}/2$ labeled regular vines on *n* variables (Morales et al 2008, Cooke et al 2015).

The constraints on a regular vine may be associated with *partial correlations* or with *conditional bivariate copula*. In the former case, we speak of a *partial correlation vine*, and in the latter case of a *vine copula.*

Partial Correlation Vines

(Bedford and Cooke 2002) show that any assignment of values in the open interval (-1, 1) to the edges in any partial correlation vine is consistent, the assignments are algebraically independent, and there is a one-to-one relation between all such assignments and the set of correlation matrices. In other words, partial correlation vines provide an algebraically independent parametrization of the set of correlation matrices, whose terms have an intuitive interpretation. Moreover, the determinant of the correlation matrix is the product over the edges of $(1 - \rho^2_{ik\,|\,D(ik)})$, where $\rho_{ik\,|\,D(ik)}$ is the partial correlation assigned to the edge with conditioned variables *i*,*k* and conditioning variables *D(ik)*. (A similar decomposition characterizes the mutual information, which generalizes the determinant of the correlation matrix, (Cooke 1997)). These features have been used in constrained sampling of correlation matrices (Joe 2005, Lewandowski 2008, Lewandowski et al 2009), building non-parametric continuous Bayesian belief nets (Hanea 2008, Hanea et al 2010), characterizing the set of rank correlation matrices (Joe 2006), and addressing problem or extending partially specified matrices to positive definite matrices (Kurowicka and Cooke 2003, 2006a).

Vine Copulae
Under suitable differentiability conditions, any multivariate density may be represented in closed form as a product of univariate densities and (conditional) copula densities on any R-vine (Bedford and Cooke 2001). The conditional copulas in this representation depend on the cumulative conditional distribution functions of the conditioned variables and, potentially, on the values of the conditioning variables. When the conditional copulas do not depend on the values of the conditioning variables, one speaks of the *simplifying assumption* of constant conditional copulas. Though most applications invoke this assumption, exploring the modeling freedom gained by discharging this assumption has begun (Hobaek-Haff et al 2010, Acar et al 2012, Acar et al 2013, Stoeber et al 2013) .

Sampling and Conditionalizing
R packages for sampling regular vines are available (Brechmann and Schepsmeier 2013, Schepsmeier et al 2014).

A sampling order for n variables is a sequence of conditional densities in which the first density is unconditional, and the densities for other variables are conditioned on the preceding variables in the ordering. A sampling order is implied by a regular-vine representation of the density if each conditional density can be written as a product of copula densities in the vine and one dimensional margins (Cooke et al 2015).

An implied sampling order is generated by a nested sequence of subvines where each sub-vine in the sequence contains one new variable not present in the preceding sub-vine. For any regular vine on *n* variables there are 2^{n-1} implied sampling orders. Implied sampling orders are a small subset of all *n*! orders but they greatly facilitate sampling. Conditionalizing a regular vine on values of an arbitrary subset of variables is a complex operation. However, conditionalizing on an initial sequence of an implied sampling order is trivial, one simply plugs in the initial conditional values and proceeds with the sampling. A general theory of conditionalization does not exist at present.

Parameter Estimation
Aas et al (2009) develop a maximum likelihood procedure to estimate parameters in copulae for D- and C-vines. The procedure can be extended to arbitrary regular vines and is the basis for most applications in mathematical finance (Heinen and Valdesogo 2008, Aas and Berg 2009, Czado et al 2009, Jaworski et al 2012,Fischer et al 2009, Low et al 2013), geostatistics (Kolbjornsen O. and Stien M.,2008,), and for most of the 672 hits (as of 5-22-2915) for Regular Vine or Vine Copula on Google Scholar. The procedure invokes the simplifying assumption (though this could be discharged) and utilizes the tree hierarchy of regular vines. Bivariate copulae are fit to the nodes in the first tree. With these, the conditional cumulative distribution functions are available to fit conditional copulae in the second tree, and so on. R code implementing these procedures in available (Schepsmeier et al 2014).

Model Inference
Model inference relates to the problem of choosing a regular vine to model a multivariate data set. If the conditional copulae are not constant, then any regular vine can be used to describe any multivariate distribution. Following Joe (1993), the motive underlying the vine copula approach

to modeling is to have a flexible low parameter set of models. In the first instance, this has led to the restriction to constant conditional copulae. When a joint distribution is defined by one particular regular vine with constant conditional copulae, these copulae will not in general remain constant when a different regular vine is used. Various strategies and heuristics have been deployed to find the best 'constant conditional copula representation', such as ’most dependence in the lowest trees’, ’most independence in the lowest trees’ and ’most dependence in the least number of nodes’. Because the number of regular vines grows rapidly in the number of variables, finding good search heuristics and model selection strategies are open and active research topics (Czado et al 2013).

Websites

http://www.statistics.ma.tum.de/en/research/vine-copula-models/#c662
http://www.birs.ca/events/2013/5-day-workshops/13w5146
http://www.cias-cufe.org/dependence/
http://rogermcooke.net/
http://www.ewi.tudelft.nl/en/the-faculty/departments/applied-mathematics/applied-probability/education/risk-analysis/

References

Aas K. and Berg D., (2009), Models for construction of multivariate dependence — a comparison study, *European J. Finance*, 15:639–659.

Aas K., Czado C., Frigessi A. and Bakken H., (2009), Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics*, 44(2):182–198.

Acar E. F., Genest C. and Nešlehová J. (2012). Beyond simplified pair-copula constructions. Journal of Multivariate Analysis, 110, 74-90

Acar,E.F., Radu V. Craiu, and Fang Yao (2013) Statistical testing of covariate effects in conditional copula models, Electronic Journal of Statistics, Volume 7 (2013), 2822-2850

Ale B.J.M., Bellamy L.J., van der Boom R., Cooper J., Cooke R.M., Goossens L.H.J., Hale A.R., Kurowicka D., Morales O., Roelen A.L.C. and Spouge J., (2009), Further development of a causal model for air transport safety (cats); building the mathematical heart. *Reliability Engineering and System Safety Journal*.

Bedford T.J. and Cooke R.M., (2001), Probability density decomposition for conditionally dependent random variables modeled by vines. *Annals of Mathematics and Arti_cial Intelligence*, 32:245–268.

Bedford T.J. and Cooke R.M., (2002), Vines — a new graphical model for dependent random variables. *Ann. of Stat.*, 30(4):1031–1068.

Besag J., (1974), Spatial interaction and the statistical analysis of lattice systems. *J. Royal. Stat. Soc. B*, 34:192–236.

Biller B.(2009), Copula-based multivariate input models for stochastic simulation *Operations Research*, 57:878–892.

Brechmann, E.C. and Schepsmeier, U. (2013) . Modeling dependence with c- and d-vine copulas: The R package CDVine. Journal of Statistical Software, 52(3):1–27.

Chollete L., Heinen A. and Valdesogo A., (2009), Modeling international financial returns with a multivariate regime switching copula, Journal of Financial Econometrics, 2009, Vol. 7, No. 4, 437–480.

Cooke R.M., (1997), Markov and entropy properties of tree and vinesdependent variables. In *Proceedings of the ASA Section of Bayesian Statistical Science*.

Cooke, R.M., Kurowicka, D. and Wilson, K. (2015) Sampling, conditionalizing, counting, merging, searching regular vines, Journal of Multivariate Analysis, Available online 14 February 2015, ISSN 0047-259X, http://dx.doi.org/10.1016/j.jmva.2015.02.001

Cooke, Roger M., Joe, H. and Aas, K. (2010) Vines Arise, in Kurowicka and Joe (eds) Dependence Modeling: Handbook on Vine Copulae, World Scientific, Singapore, 978-981-4299-87-9, 981-4299-87-1, pp43-84.

Cowell R.G., Dawid A.P., Lauritzen S.L. and Spiegelhalter D.J., (1999), *Probabilistic Networks and Expert Systems.* Statistics for Engineering and Information Sciences. Springer- Verlag, New York.

Czado C., Min A., Baumann T. and Dakovic R., (2009), Pair-copula constructions for modeling exchange rate dependence. Technical report, Technische Universit¨at M¨unchen.

Czado, C., Brechmann, E.C., Gruber, L. (2013) Selection of Vine Copulas, Copulae in Mathematical and Quantitative Finance Lecture Notes in Statistics Volume 213, 2013, pp 17-37.

Fischer M., K¨ock C., Schl¨uter S. and Weigert F., (2009), Multivariate copula models at work. Quantitative Finance, 9(7): 839–854.

Goossens L.H.J., Harper F.T., Kraan B.C.P. and Metivier H., (2000), Expert judegement for a probabilistic accident consequence uncertainty analsis. *Radiation Protection Dosimetry*, 90(3):295–301, 2000.

Goossens L.H.J., Kraan B.C., Cooke R.M., Jones J.A., Brown J., Ehrhardt J., Fischer F. and Hasemann I., (2001), Methodology and processing techniques.Directorate-General for Research EUR 18827 EN, European Commission, Luxembourg.

Hanea A.M., (2008), *Algorithms for Non-parameteric Bayesian Belief Nets*. PhD thesis, Delft Institute of Applied Mathematics, Delft University of Technology.

Hanea A.M., Kurowicka D., Cooke R.M. and Ababei D.A.,(2010), Mining and visualising ordinal data with non-parametric continuous BBNs, *Computational Statistics and Data Analysis*, 54: 668–687.

Harper F., Goossens L.H.J., Cooke R.M., Hora S., Young M., Pasler-Ssauer J., Miller L., Kraan B.C.P., Lui C., McKay M., Helton J. and Jones A., (1994), Joint USNRC CEC consequence uncertainty study: Summary of objectives, approach, application, and results for the dispersion and deposition uncertainty assessment. Technical Report VOL. III, NUREG/CR-6244, EUR 15755 EN, SAND94-1453.

Heinen A. and Valdesogo A., (2008), Canonical vine autoregressive model for large dimensions. Technical report.

Hobaek Haff,I., Aas, K. and Frigessi, A..(2010) On the simplified pair-copula construction - simply useful or too simplistic? Journal of Multivariate Analysis, 101:1296–1310.

Jaworski, P. Durante,F., Härdle, W.K., (2012) Copulae in Mathematical and Quantitative Finance: Proceedings of the workshop held in Cracaw, July 10-11, 2012, Lecture Notes in Statistics 213, Springer .

Joe H., (1993), Multivariate dependence measures and data analysis. *Comp. Stat. and Data Analysis*, 16:279–297.

Joe H., (1994), Multivariate extreme-value distributions with applications in environmental data. *The Canadian Journal of Statistics*, 22:47–64.

Joe H., (1996), Families of *m*-variate distributions with given margins and*m*(*m*□1)=2 bivariate dependence parameters. In L. R¨uschendorf, B. Schweizer and M. D. Taylor, editor, *Distributions with Fixed Marginals and Related Topics*, volume 28, pages 120–141. IMS Lecture Notes.

Joe H., (1997), *Multivariate Models and Dependence Concepts*. Chapman & Hall, London.

Joe H., (1999), Relative entropy measures of multivariate dependence. *J. Amer. Stat. Assoc*, 84(405):157–164.

Joe H., (2005), Generating random correlation matrices based on partial correlations. *J. of Multivariate Analysis*, 97:2177–2189.

Joe H., (2006), Range of correlation matrices for dependent random variables with given marginal distributions. In N. Balakrishnan, E. Castillo and J. M. Sarabia, editor, *Advances in Distribution Theory, Order Statistics and Inference, in honor of Barry Arnold*, pages 125–142. Birkhauser, Boston.

Joe H., Li H. and Nikoloulopoulos A.K., (2010), Tail dependence functions and vine copulas. *J. of Multivariate Analysis*, 101: 252–270.

Joe, H (2014) Dependence modeling with Copulas, Chapman Hall, CRC, isbn 978-1-4665-8322-1

Kolbjornsen O. and Stien M., (2008), The D-vine creation of non-Gaussian random fields. In *GEOSTATS*.

Kraan B.C.P. and Cooke R.M., (2000), Processing expert judgements in accident consequence modeling. *Radiation Protection Dosimetry*, 90(3).

Kurowicka and Joe (eds) (2010) Dependence Modeling: Handbook on Vine Copulae, World Scientific, Singapore, 978-981-4299-87-9, 981-4299-87-1, pp43-84.

Kurowicka D. and Cooke R.M., (2003), A parametrization of positive definite matrices in terms of partial correlation vines. *Linear Algebra and its Applications*, 372:225–251.

Kurowicka D. and Cooke R.M., (2004), Distribution-free continuous Bayesian belief nets. In *Mathematical Methods in Reliability*.

Kurowicka D. and Cooke R.M., (2006), *Uncertainty Analysis with High Dimensional Dependence Modelling*. Wiley.

Kurowicka D. and Cooke R.M., (2006a), Completion problem with partial correlation vines. *Linear Algebra and Its Applications*, 418(1):188–200.

Kurowicka D. and Cooke R.M., (2007), Sampling algorithms for generating joint uniform distributions using the vine-copula method. *Computational Statistics and Data Analysis*, 51:2889–2906.

Kurowicka D., Cooke R.M. and Callies U., (2007), Vines inference. *Brazilian Journal of Praobablitity and Statistics*.
Lewandowski D., (2008), *High Dimensional Dependence. Copulae, Sensitivity, Sampling*. PhD thesis, Delft Institute of Applied Mathematics, Delft University of Technology.
Lewandowski D., Kurowicka D. and Joe H., (2009), Generating random correlation matrices based on vines and extended onion method, *J. Mult. Anal.*,100:1989–2001.
Low, R.K.Y., Alcock, J., Robert Faff, R. , Brailsford, T., (2013) Canonical vine copulas in the context of modern portfolio management: Are they worth it? Journal of Banking & Finance 37 (2013) 3085–3099
Min A. and Czado C., (2008), Bayesian inference for multivariate copulas using pair copula constructions. *Submitted for publication*.
Morales Napoles O., Cooke R.M. and Kurowicka D., (2008), The number ofvines and regular vines on *n* nodes. Technical report, Delft Institute of Applied Mathematics, Delft University of Technology.
Nelsen R.B., (2006), *An Introduction to Copulas*, 2nd ed. Springer, New York.
R Core Team.(2014.) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria,
Schepsmeier, U., Stoeber, J., Brechmann, E.C. and Graeler, B. .(2014) Vine Copula:Statistical inference of vine copulas, R package version 1.3.
Schirmacher D. and Schirmacher E., (2008), Multivariate dependence modeling using pair-copulas. Technical report, Presented at The 2008 ERM Symposium, Chicago.
Stoeber,J. Joe, H. and Czado, C.(2013) Simplified pair copula constructions, limitations and extensions. Journal of Multivariate Analysis, 119:101 – 118.
Whittaker J., (1990), *Graphical Models in Applied Multivariate Statistics*.Wiley, Chichester.
Xu J. J., (1996), *Statistical Modelling and Inference for Multivariate and Longitudinal Discrete Response Data*. Ph.D. thesis, Department of Statistics, University of British Columbia.