

Specification of Informative Prior Distributions for Multinomial Models Using Vine Copulas

Kevin James Wilson*

Abstract. We consider the specification of an informative prior distribution for the probabilities in a multinomial model. We utilise vine copulas: flexible multivariate distributions built using bivariate copulas stacked in a tree structure. We take advantage of a specific vine structure, called a D-vine, to separate the specification of the multivariate prior distribution into that of marginal distributions for the probabilities and parameter values for the bivariate copulas in the vine. We provide guidance on each of the choices to be made in the prior specification and each of the questions to ask the expert to specify the model parameters within the context of an engineering application. We then give full details of the approach for the general problem.

Keywords: prior specification, vines, subjective Bayes, multinomial, elicitation.

1 Introduction

The specification of an informative prior distribution over multiple unknowns, for example the probabilities in a multinomial distribution, is a challenging task. When the unknowns are dependent in the prior beliefs of the expert, then the most flexible approaches currently used necessitate the specification of a prior covariance or correlation matrix. The restriction of this matrix of positive definiteness adds further complexity as it restricts the dependencies to be specified by the expert. Vine copulas offer no such restrictions. All correlations to be specified are algebraically free between -1 and $+1$. Thus, vines offer a promising approach to the specification of informative prior distributions.

One of the first approaches to the specification of a prior distribution for multinomial models was that of Chaloner and Duncan (1987) who proposed an extension of a method to elicit binomial probabilities in which the specifications of the expert, in the form of modes, were converted to the parameters of a Dirichlet distribution. While this is conjugate to the multinomial distribution, it is a very restrictive form for the dependency structure between the probabilities (O’Hagan et al., 2006). Dickey (1983) also proposed the Dirichlet distribution as a prior for multinomial models, using hypothetical future samples to specify the parameters. The method adjusted the expert’s specifications. Other approaches based on the Dirichlet distribution and ordered Dirichlet distribution are given in van Dorp and Mazzuchi (2003); Zapata-Vázquez et al. (2014).

Elfadaly and Garthwaite (2013, 2016) considered the Dirichlet distribution as a prior for multinomial models and proposed two extensions: the Connor-Mosimann distribution, which is also conjugate to the multinomial distribution, and the multivariate Gaussian copula. The Connor-Mosimann distribution has more parameters than the

*School of Mathematics, Statistics & Physics, Newcastle University, UK, kevin.wilson@ncl.ac.uk

Dirichlet distribution and so provides more flexibility for prior specification. However, some restrictive structure remains, such as forcing all of the probabilities to be negatively correlated with the first probability. The Gaussian copula relaxes the restrictions still further on the dependency structure, while still requiring a positive definite covariance matrix. Elfadaly and Garthwaite (2016) developed a procedure to ensure this condition holds.

Vines offer an extension to multivariate copulas in which a multivariate distribution can be built from the marginal distributions of the unknowns and a series of bivariate copulas encoding dependence in a tree structure (Joe, 1996; Bedford and Cooke, 2002). Kurowicka and Cooke (2007); Cooke et al. (2015) considered sampling from, conditioning in and searching through vines. Haff et al. (2010); Stöber et al. (2013) considered simplified vines, in which conditional dependence is assumed constant. Schepsmeier (2015); Dißmann et al. (2013) proposed methods for model selection in vines, which have been extended to Bayesian approaches by Gruber and Czado (2015); Min and Czado (2010b,a). Bayesian inference for D-vines has been considered by Czado and Min (2011).

Example 2 in Bedford et al. (2016) considered prior specification of a bivariate copula within a vine. All other applications to date have been concerned with fitting vines to data. This represents a missed opportunity for vines, whose sequential specification and completely flexible dependence structure makes them an ideal tool for prior specification. In this paper we describe how to operationalise vines as prior distributions for multinomial models. We describe the implementation of the prior specification in detail through an extended application, illustrating each step explicitly, indicating how each of the various necessary choices is to be made and providing specific elicitation questions to ask the expert. This is followed by a discussion of the general problem, providing the mathematical details of the approach.

The context of the application is a group of engineers responsible for the safety of a complex structure. In engineering applications such as this, interest typically lies in the tails of the joint distributions, as this is where safety critical events will manifest. By assuming relatively light-tailed forms for joint distributions such as the Gaussian copula, there is a risk that the probability of safety critical events will be underestimated. By allowing more flexibility in the specification of joint distributions, we are able to provide more accurate, and more conservative, estimates.

In Section 2 we provide the engineering application, detailing the specification of vines as prior distributions for multinomial models within an example. In Section 3 we give details of the approach for the general problem of the specification of a prior distribution for multinomial probabilities. In Section 4 we give a summary and areas for future work.

2 Application

2.1 Background and structure

We consider a desensitised industrial case of an ageing complex engineering structure as described in Wilson et al. (2013). The problem under consideration was to assess

the underlying condition of the structure. There were several possible tests which could be carried out to assess the condition, and each would give information about specific aspects of the condition. The tests varied in their costs, use of time and resources and invasiveness. A Bayesian network was developed to assess the potential reduction in uncertainty which would result from the tests. It is given in Figure 1.

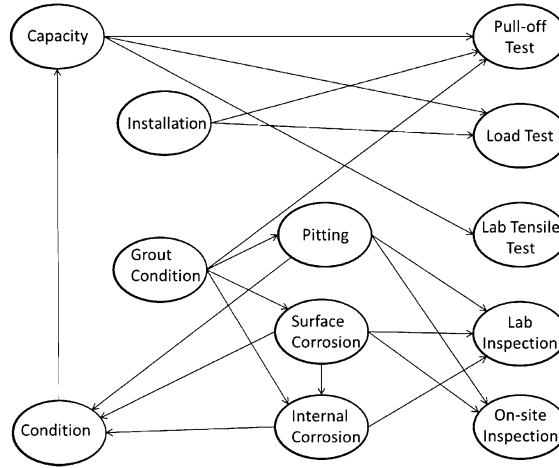


Figure 1: The Bayesian network representing the relationships between the condition of the engineering structure and the possible tests.

Each condition variable, such as grout condition and pitting, was represented by a multinomial distribution categorised by the condition of that element of the structure. For example, grout condition was categorised into fully effective, partially effective and not effective. In each node in the network, a prior distribution was required on the probabilities of belonging to each category, which would be updated on observation of results from the chosen test.

Consider the variable “capacity”. It has four possible states: original level, acceptable reduced capacity, unacceptable reduced capacity and failed. There are a large number of locations across the structure where we are interested in the state of the capacity. The multinomial observation is therefore the number of locations where capacity is in each of its states. Prior to the elicitation, full definitions of capacity and each of its states were agreed with the expert.

We wish to define a prior distribution on $(p_1, p_2, p_3, p_4)'$, the probabilities that capacity is in each of its states, ordered as above. However, these probabilities have a unit sum constraint, $\sum_{i=1}^4 p_i = 1$, and if we ask an expert about the probabilities directly, they are likely to give us beliefs which are incoherent. Therefore, instead we choose to elicit a prior distribution over $(\theta_1, \theta_2, \theta_3, \theta_4)'$, the probabilities that capacity is in each of its possible states conditional on not having been in any of its previous states. We can then recover $(p_1, p_2, p_3, p_4)'$ via

$$p_1 = \theta_1, \quad p_2 = \theta_2(1 - \theta_1),$$

$$p_3 = \theta_3(1 - \theta_2)(1 - \theta_1), \quad p_4 = \theta_4(1 - \theta_3)(1 - \theta_2)(1 - \theta_1).$$

By asking experts about the conditional probabilities, this allows them a free choice and ensures coherence of the resulting probability distribution. By defining the prior distribution over the conditional probabilities, this means we only need to consider $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)'$, as $\theta_4 = 1$. Thus, the ordering of the states is an important consideration. We choose to make the original level state 1, and so failed is state 4. We could have chosen the reverse order. However, complex engineering structures are typically very reliable. This means that the engineers will have lots of experience of locations where capacity is at its original level, and relatively little experience of locations where it is failed. It is therefore more reasonable to ask them questions about the original level, acceptable reduced level and unacceptable reduced level than the failed and two reduced levels.

We represent the prior distribution for $\boldsymbol{\theta}$ using a D-vine, and so it takes the form

$$\begin{aligned} f^{(0)}(\boldsymbol{\theta}) &= f_1^{(0)}(\theta_1)f_2^{(0)}(\theta_2)f_3^{(0)}(\theta_3)c_{1,2}(F_1^{(0)}(\theta_1), F_2^{(0)}(\theta_2)) \\ &\quad \times c_{2,3}(F_2^{(0)}(\theta_2), F_3^{(0)}(\theta_3))c_{1,3}(F_{1|2}^{(0)}(\theta_1 | \theta_2), F_{3|2}^{(0)}(\theta_3 | \theta_2)), \end{aligned} \quad (1)$$

where $f_i^{(0)}(\theta_i)$, $F_i^{(0)}(\theta_i)$ are the prior Probability Density Function (PDF) and Cumulative Distribution Function (CDF) of θ_i , $c_{i,j}(\cdot, \cdot)$ is a bivariate copula and $F_{i|j}^{(0)}(\theta_i | \theta_j)$ is the prior conditional CDF of $\theta_i | \theta_j$. From this, we see that, to fully define the prior distribution for $\boldsymbol{\theta}$, we need to specify the marginal distributions for θ_1, θ_2 and θ_3 and the dependence between θ_1 and θ_2 , θ_2 and θ_3 and θ_1 and θ_3 conditional on θ_2 . The vine structure is particularly suitable for prior specification in this case because it allows us to consider each of the specifications in isolation, reducing the elicitation burden on the expert. The vine, like multivariate copulas, preserves the marginal distributions of the parameters and uses the bivariate copulas to encode the dependencies between the parameters.

2.2 Marginal distributions

Consider the marginal distribution for θ_1 , the probability that capacity is in its original state. To elicit the marginal distribution for θ_1 , the expert is asked three questions:

Q1, Consider the proportion of locations in the structure in which capacity is at its original level.

- (a) What is the value for which you think that the true proportion would be equally likely to be above or below this value? *Call this $q_{0.5,1}$.*
- (b) What is the value for which you think it is equally likely that the true proportion would be between 0 and this value and this value and $q_{0.5,1}$?
- (c) What is the value for which you think it is equally likely that the true proportion would be between $q_{0.5,1}$ and this value and this value and 1?

We choose to ask the expert about proportions to avoid asking for probabilities of probabilities, and to link the questions more closely to observable quantities. Part (a) gives the expert’s median for θ_1 and parts (b) and (c) give their lower and upper quartiles respectively. For θ_2 and θ_3 parts (a), (b) and (c) would remain the same (using the values given by the expert for $q_{0.5,2}$ and $q_{0.5,3}$ respectively) but the populations the expert was to consider would be

- Q2, Consider all of the locations where capacity is not at its original level. We are interested in the proportion of these locations in which capacity is at a reduced acceptable level.
- Q3, Consider all of the locations where capacity is not at its original level or an acceptable reduced level. We are interested in the proportion of these locations in which capacity is at a reduced unacceptable level.

Suppose that the elicited quantiles for an expert are given in Table 1.

State	$q_{L,i}$	$q_{0.5,i}$	$q_{U,i}$
Original level	0.7	0.75	0.8
Acceptable reduced capacity	0.6	0.7	0.75
Unacceptable reduced capacity	0.2	0.25	0.3
Failed	1	1	1

Table 1: The quantiles specified by an expert for the conditional probabilities of capacity being in its possible states.

We see that, in the beliefs of this expert, the capacity is most likely to be at its original level. Conditional on capacity not being at its original level, it is most likely to be at an acceptable reduced level. Of course, once we know that it is not in any of the first three categories, it will be failed with probability one.

For each pair of quantiles, we can calculate the values of the parameters of the equivalent beta distribution. For three quartiles, there is no exact beta distribution which matches them in general. We choose a beta distribution which approximately fits all three quantiles by matching the first two moments. Full details are given in the general problem section. By fitting a two parameter distribution to three elicited quantiles, this allows us to provide feedback to the expert as to the consistency of their quantile specifications. If the beta distributions implied by the three pairs of quantiles are very different, then this could imply that the expert may wish to revise one or more of their quantiles.

From the elicited quantiles in Table 1, we find exact beta distribution parameter values $(a_{i,j}, b_{i,j})$, where $j = 1, 2, 3$ represent the pairs of quantiles, $(q_{0.25,i}, q_{0.5,i})$, $(q_{0.5,i}, q_{0.75,i})$ and $(q_{0.25,i}, q_{0.75,i})$ respectively, and the parameter values which approximately match all three quantiles, (a_i, b_i) . They are given in Table 2 for θ_1, θ_2 and θ_3 .

We see that there are differences between the beta distributions resulting from the different pairs of quantiles, though in general these are relatively small. We can also

Probability	$a_{i,1}$	$b_{i,1}$	$a_{i,2}$	$b_{i,2}$	$a_{i,3}$	$b_{i,3}$	a_i	b_i
θ_1	27.85	9.50	23.22	7.96	25.53	8.63	25.40	8.65
θ_2	8.59	3.43	24.84	10.85	12.04	5.88	11.89	5.45
θ_3	8.00	23.30	9.52	27.85	8.63	25.53	8.67	25.43

Table 2: The beta distribution parameter values resulting from the elicited quantiles from the expert.

view the densities of the marginal beta prior distributions in each case, and these are given in Figure 2. The colours represent $(a_{1,i}, b_{1,i})$ (black), $(a_{2,i}, b_{2,i})$ (red), $(a_{3,i}, b_{3,i})$ (green) and (a_i, b_i) (blue).

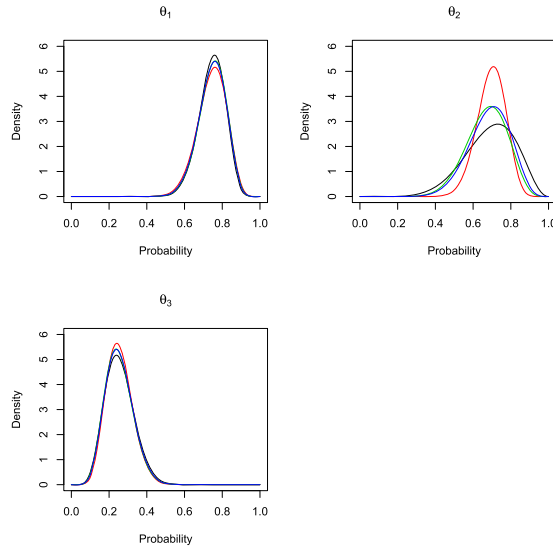


Figure 2: The marginal prior beta distributions based on each pair of elicited quantiles and the final distribution for $(\theta_1, \theta_2, \theta_3)$. The colours represent $(a_{1,i}, b_{1,i})$ (black), $(a_{2,i}, b_{2,i})$ (red), $(a_{3,i}, b_{3,i})$ (green) and (a_i, b_i) (blue).

We see that the overall densities in blue represent a reasonable aggregation of the other densities for θ_1, θ_2 and θ_3 . For θ_2 , one of the elicited quantiles appears to be inconsistent with the other two. We may wish to ask the expert to reconsider their quantile specifications in light of this. We also see the differences in the variances of the prior distributions, with the largest uncertainty in the value of θ_2 and the smallest in the value of θ_1 .

2.3 Bivariate copulas

We need to specify the dependence structure between the elements of $\boldsymbol{\theta}$. This is equivalent to specifying the copulas $c_{1,2}(\cdot, \cdot)$, $c_{2,3}(\cdot, \cdot)$ and $c_{1,3}(\cdot, \cdot)$. The elicitation in this stage

asks the expert to condition on the value of one of the probabilities and provide revised beliefs about another probability in light of this. It would be extremely cognitively challenging to ask the expert to condition on the value of a conditional probability and then give their updated beliefs about a further conditional probability. Therefore, instead, we return to $\mathbf{p} = (p_1, p_2, p_3)'$, the unconditional probabilities that capacity is at its original level, a reduced acceptable level and a reduced unacceptable level respectively.

To specify the two unconditional copulas in the vine, we ask the expert questions about $p_2 \mid p_1$ and $p_3 \mid p_2$ respectively. In particular, we choose to condition on the probabilities taking their prior median values. We could condition on any value of the probabilities, but by choosing the median we ensure that the expert is considering values that are not inconsistent with their prior beliefs. The questions we ask the expert to elicit $c_{1,2}(\cdot, \cdot)$ are

Q4, In Q1, you identified that you think that the proportion of locations where capacity is at its original level is equally likely to be above or below 0.75. Suppose that this is correct. In light of this information:

- (a) Are you still comfortable that it is equally likely that the proportion of locations where the capacity is at an acceptable reduced level is above or below 0.175? *If so,*
- (b) What is the value for which you think it is equally likely that the true proportion is between 0 and this value and this value and 0.175?
- (c) What is the value for which you think it is equally likely that the true proportion is between 0.175 and this value and this value and 1?

This provides $(q'_{0.25,2}, q'_{0.5,2}, q'_{0.75,2})$, the lower quartile, median and upper quartile of $p_2 \mid p_1$. The same three quantiles of $p_3 \mid p_2$ can be elicited via:

Q5, In Q2, you identified that you think that the proportion of locations where capacity is at an acceptable reduced level is equally likely to be above or below 0.175. Suppose that this is correct. In light of this information:

- (a) Are you still comfortable that it is equally likely that the proportion of locations where the capacity is at an unacceptable reduced level is above or below 0.01875? *If so,*
- (b) What is the value for which you think it is equally likely that the true proportion is between 0 and this value and this value and 0.01875?
- (c) What is the value for which you think it is equally likely that the true proportion is between 0.01875 and this value and this value and 1?

We notice a strength of vines which makes them an ideal structure for the prior distribution here. Vines are not symmetric in the sense that they do not treat the relationships between all of the variables equally, as opposed to a Gaussian copula for example. The

only unconditional relationships to be specified in the vine are between adjacent nodes. Similarly, in this application, the multinomial categories have a specific structure. They are ordered. Thus by choosing the ordering as in (1), we can elicit both of the dependencies in the first tree of the vine by only conditioning on adjacent categories. It would be a much more cognitively challenging task for an expert to revise their quartiles for one category on observation of a non-adjacent category only. The vine structure allows us to avoid this.

Returning to the example, we saw in Q4, that the prior median for p_1 was 0.75. Suppose that, in Q4, the expert chooses to accept the conditional median of $p_2 | p_1 = 0.75$ and gives revised quartiles of $q'_{0.25,2} = 0.165$ and $q'_{0.75,2} = 0.185$. This compares to unconditional quartiles of $q^{\#}_{0.25,2} = 0.15$ and $q^{\#}_{0.75,2} = 0.188$. We can transform these to quartiles of $\theta_2 | \theta_1$, i.e., $q^*_{0.25,2} = 0.66$ and $q^*_{0.75,2} = 0.74$ using

$$q^*_{0.25,2} = \frac{q'_{0.25,2}}{1 - q^{\#}_{0.5,1}}, \quad q^*_{0.75,2} = \frac{q'_{0.75,2}}{1 - q^{\#}_{0.5,1}}.$$

We see that the upper and lower quartiles reflect reduced uncertainty in light of the extra information. The expert further identifies that they would expect the dependence to be positive.

We need to use the three revised quantiles for $\theta_2 | \theta_1$ to specify the bivariate copula $c_{1,2}(\cdot, \cdot)$. We could do so by fitting a non-parametric copula which exactly matches the quantiles, for example using minimum information methods (Bedford et al., 2016). This in general would be under-constrained. Alternatively, we could impose a specific parametric copula, for example the Gaussian copula, for each bivariate copula in the vine and choose the parameter values to most closely match the elicited quantiles. This is more restrictive, but offers a more convenient form for inference.

Instead, for each bivariate copula we will consider a number of parametric copulas suitable for representing dependence: Gaussian, Frank, Clayton, Gumbel and t-, fit each to the three quantiles given by the expert using least squares and then choose the copula which provides the best fit of these, again using least squares. This approach offers more flexibility than assuming a single form for all copulas in the vine and the resulting vine is less complex than that when using non-parametric copulas. An alternative approach would be to choose the copula based on the expert's knowledge of the dependencies and how this relates to properties of different copulas. We will assess the impact of this approach compared to assuming a Gaussian copula for all bivariate relationships in the next section.

In Table 3 we provide the parameter values which most closely match the revised quantiles and the fit, in terms of the sum of squared differences, for each of the candidate copulas for $\theta_2 | \theta_1$.

From the table, we see that the Clayton copula is providing the best fit to the elicited judgements of the expert, though the Frank, Gaussian and t-copula also appear to fit the quantiles well. The parameter value chosen for the Clayton copula is $\lambda = 3.61$ and so we see positive dependence between θ_1 and θ_2 .

Copula	λ	Fit
Gaussian	0.85	0.00019
Frank	7.48	0.00019
Clayton	3.61	0.000057
Gumbel	2.98	0.00048
t-	(0.76,2.1)	0.00019

Table 3: A comparison of each of the candidate copulas for the prior dependence between (θ_1, θ_2) .

We can see the differences between the fitted copulas, and the fits of the copulas to the elicited quantiles, in Figure 3, which plots the conditional CDF of $\theta_2 \mid \theta_1 = 0.75$ for the Gaussian copula (black), Frank copula (red), Clayton copula (green), Gumbel copula (light blue) and t-copula (pink), as well as the three elicited quantiles (dark blue).

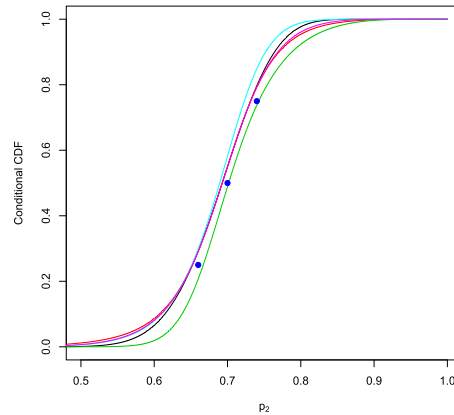


Figure 3: The conditional CDF of θ_2 given $\theta_1 = q_{0.5,1}$ for the Gaussian copula (black), Frank copula (red), Clayton copula (green), Gumbel copula (light blue) and t-copula (pink), as well as the three elicited quantiles (dark blue).

We see from the figure that most of the copulas produce relatively similar conditional distributions, with all except the Gumbel and Clayton copulas appearing very similar. The Clayton copula provides a conditional distribution closest to the median and upper quartile. We choose the Clayton copula to represent the dependence between θ_1 and θ_2 . This gives a Kendall’s Tau value of $\tau = 0.64$, which indicates moderate correlation.

For the copula between θ_2 and θ_3 , based on the quantiles elicited in Q5, we can use the same approach. In this case, the Clayton copula again provides the best fit with $\lambda = 3.06$ and a value of Kendall’s Tau of $\tau = 0.61$.

In the second tree of the vine, we require the copula between θ_1 and θ_3 conditional on θ_2 . We again choose to ask questions concerning \mathbf{p} as they are cognitively simpler for the expert. In this case, we condition on p_1 and p_2 both taking their prior median

values $q_{0.5,1}^{\#} = 0.75$ and $q_{0.5,2}^{\#} = 0.175$, and ask the expert for revised quantiles for p_3 . The specific questions are:

Q6, Suppose that the proportions of locations where capacity is at its original level and at an acceptable reduced level are 0.75 and 0.175 respectively. In light of this information:

- (a) Are you still comfortable that it is equally likely that the proportion of locations where the capacity is at an unacceptable reduced level is above or below 0.01875? *If so,*
- (b) What is the value for which you think it is equally likely that the true proportion is between 0 and this value and this value and 0.01875?
- (c) What is the value for which you think it is equally likely that the true proportion is between 0.01875 and this value and this value and 1?

This provides revised quantiles for $p_3 \mid p_1, p_2$ which we convert to revised quantiles of $\theta_3 \mid \theta_1, \theta_2$ using the same approach as the unconditional copulas. This provides information about $F_{3|1,2}^{(0)}(\theta_3 \mid \theta_1, \theta_2)$ which, together with $F_{1|2}^{(0)}(\theta_1 \mid \theta_2)$ and $F_{3|2}^{(0)}(\theta_3 \mid \theta_2)$, which we obtain using the copulas in the first tree of the vine, allows us to specify $c_{1,3}(\cdot, \cdot)$. Full details are given in the general problem section.

Suppose that the quantiles resulting from Q6, are $(q'_{0.25,3}, q'_{0.5,3}, q'_{0.75,3}) = (0.018, 0.01875, 0.0195)$. The revised quantiles for $\theta_3 \mid \theta_1, \theta_2$ are then $(0.24, 0.25, 0.26)$. Fitting each of the candidate copulas to these three quantiles results in the Frank copula providing the best fit. The parameter value which most closely matches the quantiles specified by the expert is $\lambda = 0.98$ and the Kendall's Tau value is $\tau = 0.11$. We see that the correlation between θ_1 and θ_3 conditional on θ_2 is weaker than either of those in the first tree of the vine. This is partly as a result of θ_1 and θ_3 representing non-adjacent categories, i.e., original level and unacceptable reduced level. Thus, it is more suitable to consider this relationship in the second tree of the vine. This fully defines the prior distribution.

2.4 Impact of using a range of copulas

Let us suppose that there are safety implications for the structure if more than a quarter of the locations have a capacity in the unacceptable reduced or failed states. Then the engineers are interested in $\Pr(p_3 \cup p_4 > 0.25)$. We simulate from our prior distribution to obtain samples $\boldsymbol{\theta}^{(j)} = (\theta_1^{(j)}, \theta_2^{(j)}, \theta_3^{(j)}, \theta_4^{(j)})'$, for $j = 1, \dots, 100,000$ and transform these to samples from $\boldsymbol{p}^{(j)} = (p_1^{(j)}, p_2^{(j)}, p_3^{(j)}, p_4^{(j)})'$ via

$$\begin{aligned} p_1^{(j)} &= \theta_1^{(j)}, \\ p_2^{(j)} &= \theta_2^{(j)}(1 - \theta_1^{(j)}), \\ p_3^{(j)} &= \theta_3^{(j)}(1 - \theta_2^{(j)})(1 - \theta_1^{(j)}), \\ p_4^{(j)} &= \theta_4^{(j)}(1 - \theta_3^{(j)})(1 - \theta_2^{(j)})(1 - \theta_1^{(j)}). \end{aligned}$$

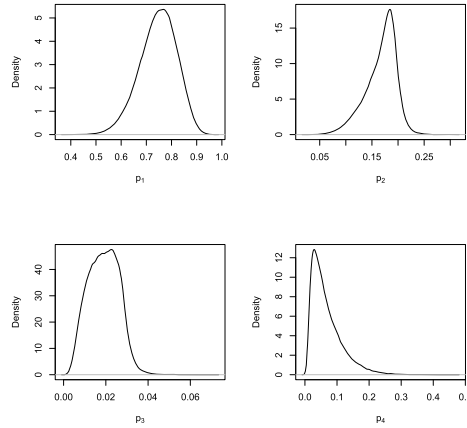


Figure 4: The prior marginal densities of p_1 , p_2 , p_3 and p_4 .

We plot the resulting densities for \mathbf{p} in Figure 4.

We calculate the proportion of samples in which $p_3 + p_4 > 0.25$, which provides our estimate of $\Pr(p_3 \cup p_4 > 0.25)$. We obtain 0.0123.

We can also fit a vine to the specifications of the expert in which we make each bivariate copula a Gaussian copula. If we do so, then the correlation parameter ρ takes the values 0.85, 0.79 and 0.04 for $c_{1,2}(\cdot, \cdot)$, $c_{2,3}(\cdot, \cdot)$ and $c_{1,3}(\cdot, \cdot)$ respectively. Under this copula, $\Pr(p_3 \cup p_4 > 0.25) = 0.0102$. Thus, by allowing flexibility in the choice of copulas in the vine, and fitting the parametric copula which most closely matches the expert’s specifications, the probability of safety implications for the structure increases by 21%. This demonstrates the value of taking the extra effort to investigate each of the candidate copulas.

We can see the differences between the joint densities under the flexible choice of copulas and the vine built only using the Gaussian copula in Figure 5, where we have plotted the joint prior density between p_3 and p_4 from the simulations.

In particular, we see stark differences in the tail behaviour under the two vines.

3 General problem

3.1 D-vine prior distribution structure

Suppose that we have counts in $(m + 1)$ categories, $\mathbf{Y} = (Y_1, \dots, Y_{m+1})'$ and, conditional on the probabilities of each of the categories, $(p_1, \dots, p_{m+1})'$, they follow a multinomial distribution $\mathbf{Y} \mid (p_1, \dots, p_{m+1})' \sim \text{MN}(N, (p_1, \dots, p_{m+1}))$, for some number of observations N . The probabilities are constrained to sum to one and so we set $\mathbf{p} = (p_1, \dots, p_m)'$ and then $p_{m+1} = 1 - \sum_{i=1}^m p_i$. To perform inference we require a prior distribution $f^{(0)}(\mathbf{p})$ representing the beliefs of an expert. It will typically be the case that $p_i \not\perp p_j$ for $i \neq j$. We wish to define a flexible prior distribution with this property.

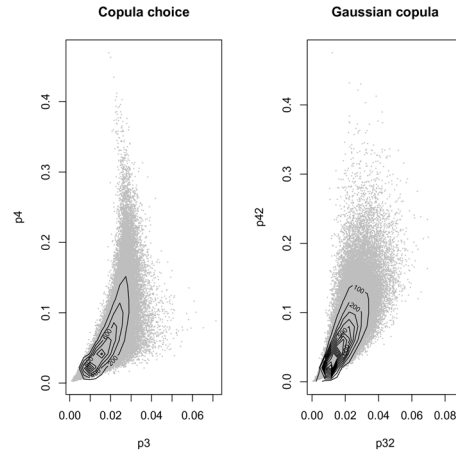


Figure 5: The prior joint density between p_3 and p_4 using the best fitting parametric bivariate copulas (left) and using only Gaussian copulas (right).

We will use a D-vine to construct such a distribution. A background on copulas and vines is given in the Supplementary Material (Wilson, 2017).

As a result of the unit sum constraint on $(p_1, \dots, p_{m+1})'$, we propose the transformation (Elfadaly and Garthwaite, 2016),

$$\theta_i = \frac{p_i}{1 - \sum_{j=1}^{i-1} p_j},$$

for $i = 2, \dots, m + 1$, where $\theta_1 = p_1$. We then specify a prior distribution for $\boldsymbol{\theta} = (\theta_1, \dots, \theta_m)'$, where θ_i is the conditional probability that an observation falls into category i given that it did not fall into categories $1, \dots, i - 1$. We can recover \mathbf{p} using the inverse transformation $p_i = \theta_i \prod_{j=1}^{i-1} (1 - \theta_j)$, for $i = 2, \dots, m + 1$, where $\theta_{m+1} = 1$. Elicitation of conditional probabilities has been successfully carried out in practice (e.g. Quigley and Walls (2010)).

We will use a D-vine to represent the prior distribution $f^{(0)}(\boldsymbol{\theta})$. The vine is fully defined by the ordering of the variables in the first tree of the vine. Suppose this ordering is $\tilde{\boldsymbol{\theta}} = (\tilde{\theta}_1, \dots, \tilde{\theta}_m)'$. Then the structure of the vine is given in Figure 6.

In some cases, there will be a natural ordering of the categories which should be maintained in $\tilde{\boldsymbol{\theta}}$. This was the case in the application and will always be the case if the multinomial variable represents the partitioning of a continuous univariate variable. If there is no natural ordering, the ordering $\tilde{\boldsymbol{\theta}}$ should be chosen to assess the strongest dependencies in the first tree of the vine. This is consistent with advice in Aas et al. (2009); Min and Czado (2011).

The prior distribution is given by the vine distribution representing the D-vine in

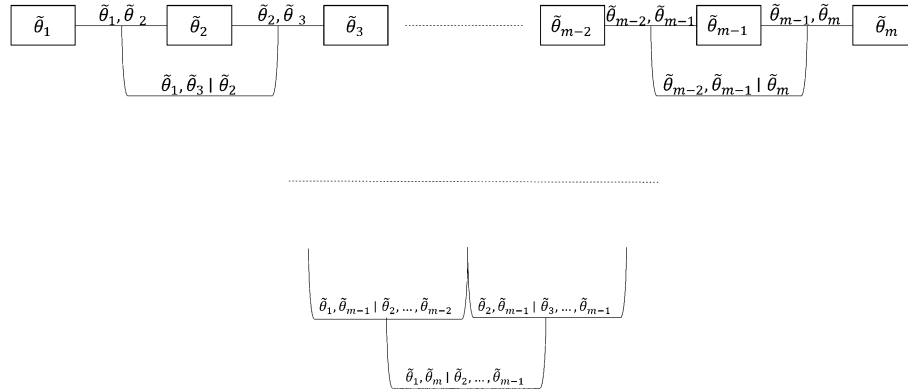


Figure 6: The structure of a general D-vine in m dimensions.

Figure 6. This is

$$\begin{aligned}
 f^{(0)}(\tilde{\theta}) &= \prod_{i=1}^m \left[f_i^{(0)}(\tilde{\theta}_i) \right] \times \prod_{i=1}^{m-1} \left[c_{i,i+1} \left(F_i^{(0)}(\tilde{\theta}_i), F_{i+1}^{(0)}(\tilde{\theta}_{i+1}) \right) \right] \\
 &\times \prod_{i=1}^{m-2} \prod_{j=i+2}^m \left[c_{i,j} \left(F_{i|i+1,\dots,j-1}^{(0)}(\tilde{\theta}_i | \tilde{\theta}_{i+1}), F_{j|i+1,\dots,j-1}^{(0)}(\tilde{\theta}_j | \tilde{\theta}_{i+1}, \dots, \tilde{\theta}_{j-1}) \right) \right],
 \end{aligned}$$

where $f_i^{(0)}(\tilde{\theta}_i)$ is the marginal prior PDF of $\tilde{\theta}_i$, $c_{i,j}(\cdot, \cdot)$ is a bivariate copula PDF, $F_i^{(0)}(\tilde{\theta}_i)$ is the marginal prior CDF of $\tilde{\theta}_i$ and $F_{i|i+1,\dots,j-1}^{(0)}(\tilde{\theta}_i | \tilde{\theta}_{i+1}, \dots, \tilde{\theta}_{j-1})$ is the prior conditional CDF of $\tilde{\theta}_i | \tilde{\theta}_{i+1}, \dots, \tilde{\theta}_{j-1}$. In order to write the prior density in this form we have made the assumption of a simplified vine. Full details of this are given in the Supplementary Material (Wilson, 2017).

In order to fully specify the D-vine we need to specify the marginal distributions of each of the probabilities, $f_i^{(0)}(\tilde{\theta}_i)$, the unconditional copulas in the first tree of the vine, $c_{i,i+1}(\cdot, \cdot)$, and the conditional copulas in trees 2 to $m - 1$ of the vine, $c_{i,j}(\cdot, \cdot)$. We do not need to specify the prior conditional distributions $F_{i|i+1,\dots,j-1}^{(0)}(\cdot | \cdot)$, as they can be calculated from the other specifications. For example, if we require $F_{i|i+1}^{(0)}(\tilde{\theta}_i | \tilde{\theta}_{i+1})$ then we can find

$$f_{i|i+1}^{(0)}(\tilde{\theta}_i | \tilde{\theta}_{i+1}) = c_{i,i+1} \left(F_i^{(0)}(\tilde{\theta}_i), F_{i+1}^{(0)}(\tilde{\theta}_{i+1}) \right) f_i^{(0)}(\tilde{\theta}_i),$$

and integrate over $\tilde{\theta}_i$. These prior conditional distributions from the second tree of the vine can then be used to calculate the prior conditional distributions in the third tree of the vine, for example $F_{i|i+1,i+2}^{(0)}(\tilde{\theta}_i | \tilde{\theta}_{i+1}, \tilde{\theta}_{i+2})$ via

$$f_{i|i+1,i+2}^{(0)}(\tilde{\theta}_i | \tilde{\theta}_{i+1}, \tilde{\theta}_{i+2}) = c_{i,i+2} \left(F_{i|i+1}^{(0)}(\tilde{\theta}_i | \tilde{\theta}_{i+1}), F_{i+2|i+1}^{(0)}(\tilde{\theta}_{i+2} | \tilde{\theta}_{i+1}) \right) \\ \times c_{i,i+1} \left(F_i^{(0)}(\tilde{\theta}_i), F_{i+1}^{(0)}(\tilde{\theta}_{i+1}) \right) f_i^{(0)}(\tilde{\theta}_i),$$

and so on.

3.2 Elicitation and specification

For each conditional probability $\tilde{\theta}_i$, we require a marginal prior distribution with support on $[0, 1]$. A flexible and much used prior distribution is the beta distribution. We define $\tilde{\theta}_i \sim \text{beta}(a_i, b_i)$, $i = 1, \dots, m$, and so we need to elicit information to specify (a_i, b_i) .

We choose to elicit three quantiles to specify these two parameters, which provides a consistency check on the quantiles (Garthwaite et al., 2005). We elicit the median $q_{0.5,i}$, a quantile below the median, $q_{L,i}$ and a quantile above the median, $q_{U,i}$. In the example we chose to elicit the lower and upper quartiles, $q_{0.25,i}$ and $q_{0.75,i}$, which allowed us to use the bisection method. Other common choices are $(q_{0.33,i}, q_{0.67,i})$ and $(q_{0.05,i}, q_{0.95,i})$. There are no exact values for (a_i, b_i) in general which match all three quantiles. We can, however, find exact beta parameter values for each pair of quantile specifications from the expert.

Suppose that the exact parameter values are $(q_{L,i}, q_{0.5,i}) \Rightarrow (a_{i,1}, b_{i,1})$, $(q_{0.5,i}, q_{U,i}) \Rightarrow (a_{i,2}, b_{i,2})$ and $(q_{L,i}, q_{U,i}) \Rightarrow (a_{i,3}, b_{i,3})$. We can find the means and variances of each of these beta distributions, $\mu_{i,j} = a_{i,j}/(a_{i,j} + b_{i,j})$ and $\sigma_{i,j}^2 = a_{i,j}b_{i,j}/[(a_{i,j} + b_{i,j})^2(a_{i,j} + b_{i,j} + 1)]$. We specify the prior distribution for $\tilde{\theta}_i$ by setting its mean and variance to be weighted averages of $(\mu_{i,j}, \sigma_{i,j}^2)$, $j = 1, 2, 3$:

$$\mu_i = w_{i,1}\mu_{i,1} + w_{i,2}\mu_{i,2} + w_{i,3}\mu_{i,3} \\ \sigma_i^2 = \frac{1}{w_i} (w_{i,1}^2\sigma_{i,1}^2 + w_{i,2}^2\sigma_{i,2}^2 + w_{i,3}^2\sigma_{i,3}^2),$$

where $\sum_{j=1}^3 w_{i,j} = 1$ and $w_i = \sum_{j=1}^3 w_{i,j}^2$. The beta distribution parameters which satisfy this mean and variance are $a_i = \mu_i[\{\mu_i(1 - \mu_i)\}/\sigma_i^2 - 1]$ and $b_i = (1 - \mu_i)[\{\mu_i(1 - \mu_i)\}/\sigma_i^2 - 1]$.

We also need to specify each of the bivariate copulas in the vine, which represent the dependencies between the probabilities. To do so, we condition on the values of some of the probabilities and ask for revised quantiles for a further probability. For this to be a manageable task for the expert, we choose to elicit quantiles of the elements of $\tilde{\mathbf{p}}$, the ordered unconditional probabilities, then convert them to quantiles of the elements of $\tilde{\boldsymbol{\theta}}$.

Initially, we consider the specification of the copulas involving $\tilde{\theta}_1$. Consider the copula $c_{1,j}(\cdot, \cdot)$. To specify this, we ask the expert to suppose that we have observed the values of $\tilde{p}_1, \dots, \tilde{p}_{j-1}$ at their prior medians, $\tilde{p}_i = q_{0.5,i}^\#$, calculated approximately (Elfadaly and Garthwaite, 2016) from the marginal specifications as $q_{0.5,i}^\# = q_{0.5,i} \prod_{k=1}^{i-1} (1 - q_{0.5,k})$. We then elicit $(q'_{L,j}, q'_{U,j})$, the revised quantiles below and above the median of $\tilde{p}_j | \tilde{p}_1 =$

$q_{0.5,1}^\#, \dots, \tilde{p}_{j-1} = q_{0.5,j-1}^\#$. These are converted to quantiles from $\tilde{\theta}_j \mid \tilde{\theta}_1, \dots, \tilde{\theta}_{j-1}$ via

$$q_{L,j}^* = \frac{q'_{L,j}}{1 - \sum_{k=1}^{j-1} q_{0.5,k}^\#}, \quad q_{U,j}^* = \frac{q'_{U,j}}{1 - \sum_{k=1}^{j-1} q_{0.5,k}^\#}.$$

These are quantiles of $F_{j|1,\dots,j-1}^{(0)}(\tilde{\theta}_j \mid \tilde{\theta}_1, \dots, \tilde{\theta}_{j-1})$. We can evaluate these quantiles for any given bivariate copula with given parameter values using the relationship

$$F_{j|1,\dots,j-1}^{(0)}(\tilde{\theta}_j \mid \tilde{\theta}_1, \dots, \tilde{\theta}_{j-1}) = \frac{\partial C_{1,j} \left(F_{1|2,\dots,j-1}^{(0)}(\tilde{\theta}_1 \mid \tilde{\theta}_2, \dots, \tilde{\theta}_{j-1}), F_{j|2,\dots,j-1}^{(0)}(\tilde{\theta}_j \mid \tilde{\theta}_2, \dots, \tilde{\theta}_{j-1}) \right)}{\partial F_{1|2,\dots,j-1}^{(0)}(\tilde{\theta}_1 \mid \tilde{\theta}_2, \dots, \tilde{\theta}_{j-1})}.$$

This allows us to evaluate the same quantiles, for any bivariate copula with any parameter values, as those specified by the expert. Suppose, for a copula with parameter values $\lambda \in \Lambda$, they are given by $(q_{L,j}^\lambda, q_{0.5,j}^\lambda, q_{U,j}^\lambda)$. Then we can fit the copula to the expert’s revised quantiles using least squares by choosing λ to satisfy

$$\hat{\lambda} = \min_{\lambda \in \Lambda} \{ (q_{L,j}^\lambda - q_{L,j}^*)^2 + (q_{0.5,j}^\lambda - q_{0.5,j}^*)^2 + (q_{U,j}^\lambda - q_{U,j}^*)^2 \}.$$

There are many different parametric bivariate copulas which could be used as $c_{1,j}(\cdot, \cdot)$. We advocate choosing the bivariate copula which minimises the sum of squares overall with the expert’s revised quantiles out of a number of candidates. Suitable candidates for both positive and negative dependence are the Gaussian, t- and Frank copulas, for positive dependence only are the Clayton and Gumbel copulas and for negative dependence only are the rotated Clayton and rotated Gumbel copulas. Details of each of these copulas are given in the Supplementary Material (Wilson, 2017).

This procedure will specify each of the bivariate copulas on the very left of each tree of the vine. To specify the subsequent copulas in each tree, for example the k ’th copulas, we ask the expert to suppose that an observation hasn’t fallen into the first $(k - 1)$ categories. This redefines the remaining unconditional probabilities as

$$\begin{aligned} \tilde{p}_k &= \tilde{\theta}_k \\ \tilde{p}_{k+\ell} &= \tilde{\theta}_{k+\ell} \prod_{r=k}^{k+\ell-1} (1 - \tilde{\theta}_r). \end{aligned}$$

The method above will then specify all copulas of the form $c_{k,j}(\cdot, \cdot)$, for $j > k$.

4 Summary

In this paper we have considered the problem of specifying a prior distribution for probabilities in a multinomial model. We have proposed a method which is based on a structure called a D-vine in which the specification of the multivariate prior distribution

reduces to that of the marginal distributions, the ordering of the probabilities and the conditional and unconditional bivariate copulas. This provides a flexible structure and avoids issues of underparameterisation associated with conjugate priors such as the Dirichlet distribution and the direct specification of a positive definite correlation matrix.

While this paper focuses on multinomial models, there is much scope to utilise this approach to specify informative prior distributions more generally. The vine structure separates the marginal and dependency specifications and allows the Bayesian statistician to relate questions on observable quantities for elicitation to the parameters in the vine. With the recent advances in Bayesian inferential techniques for vines, the use of vines as a structure for multivariate prior distributions incorporating dependency is now very appealing.

Supplementary Material

Supplementary Material to: Specification of informative prior distributions for multinomial models using vine copulas (DOI: [10.1214/17-BA1068SUPP](https://doi.org/10.1214/17-BA1068SUPP); .pdf).

References

- Aas, K., Czado, C., Frigessi, A., and Bakken, H. (2009). “Pair-copula constructions of multiple dependence.” *Insurance: Mathematics and Economics*, 44: 182–198. MR2517884. doi: <https://doi.org/10.1016/j.insmatheco.2007.02.001>. 760
- Bedford, T. and Cooke, R. (2002). “Vines – a new graphical model for dependent random variables.” *Annals of Statistics*, 30: 1031–1068. MR1926167. doi: <https://doi.org/10.1214/aos/1031689016>. 750
- Bedford, T., Daneshkhah, A., and Wilson, K. (2016). “Approximate Uncertainty Modeling in Risk Analysis with Vine Copulas.” *Risk Analysis*, 36: 792–815. 750, 756
- Chaloner, K. and Duncan, G. (1987). “Some properties of the Dirichlet-multinomial distribution and its use in prior elicitation.” *Communications in Statistics-Theory and Methods*, 16: 511–523. MR0886361. doi: <https://doi.org/10.1080/03610928708829384>. 749
- Cooke, R., Kurowicka, D., and Wilson, K. (2015). “Sampling, Conditionalizing, Counting, Merging, Searching Regular Vines.” *Journal of Multivariate Analysis*, 138: 4–18. MR3348830. doi: <https://doi.org/10.1016/j.jmva.2015.02.001>. 750
- Czado, C. and Min, A. (2011). *Handbook on Vines*, chapter Bayesian Inference for D-vines: Estimation and Model Selection. World Scientific. MR2856977. 750
- Dickey, J. (1983). “Multiple hypergeometric functions: probabilistic interpretations of statistical uses.” *Journal of the American Statistical Association*, 78: 1615–1627. MR0721212. 749

- Dißmann, J., Brechmann, E., Czado, C., and Kurowicka, D. (2013). “Selecting and estimating regular vine copulae and application to financial returns.” *Computational Statistics and Data Analysis*, 59: 52–69. MR3000041. doi: <https://doi.org/10.1016/j.csda.2012.08.010>. 750
- Elfadaly, F. and Garthwaite, P. (2013). “Eliciting Dirichlet and Connor-Mosimann prior distributions for multinomial models.” *Test*, 22: 628–646. MR3122326. doi: <https://doi.org/10.1007/s11749-013-0336-4>. 749
- Elfadaly, F. and Garthwaite, P. (2016). “Eliciting Dirichlet and Gaussian copula prior distributions for multinomial models.” *Statistics and Computing*, 27: 449–467. MR3599683. doi: <https://doi.org/10.1007/s11222-016-9632-7>. 749, 750, 760, 762
- Garthwaite, P., Kadane, J., and O’Hagan, A. (2005). “Statistical methods for eliciting probability distributions.” *Journal of the American Statistical Association*, 100: 680–700. MR2170464. doi: <https://doi.org/10.1198/016214505000000105>. 762
- Gruber, L. and Czado, C. (2015). “Sequential Bayesian Model Selection of Regular Vine Copulas.” *Bayesian Analysis*, 10: 937–963. MR3432245. doi: <https://doi.org/10.1214/14-BA930>. 750
- 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. MR2595309. doi: <https://doi.org/10.1016/j.jmva.2009.12.001>. 750
- Joe, H. (1996). *Distributions with Fixed Marginals and Related Topics*, chapter Families of m -variate distributions with given margins and $m(m-1)/2$ bivariate dependence parameters. Institute of Mathematical Statistics. MR1485527. doi: <https://doi.org/10.1214/lnms/1215452614>. 750
- Kurowicka, D. and Cooke, R. (2007). “Sampling algorithms for generating joint uniform distributions using the vine-copula method.” *Computational Statistics and Data Analysis*, 51: 2889–2906. MR2345612. doi: <https://doi.org/10.1016/j.csda.2006.11.043>. 750
- Min, A. and Czado, C. (2010a). “Bayesian model selection for D-vine pair-copula constructions.” *Canadian Journal of Statistics*, 39: 239–258. 750
- Min, A. and Czado, C. (2010b). “Bayesian model selection for multivariate copulas using pair-copula constructions.” *Journal of Financial Econometrics*, 8: 511–546. 750
- Min, A. and Czado, C. (2011). “Bayesian model selection for D-vine pair-copula constructions.” *The Canadian Journal of Statistics*, 39: 239–258. MR2839479. doi: <https://doi.org/10.1002/cjs.10098>. 760
- O’Hagan, A., Buck, C., Daneshkhah, A., Eiser, J., Garthwaite, P., Jenkinson, D., Oakley, J., and Rakow, T. (2006). *Uncertain Judgements: eliciting expert probabilities*. Chichester: John Wiley. 749
- Quigley, J. and Walls, L. (2010). “Reconciling experts opinion concerning the value of

- testing using Bayesian networks: a bridge too far?" In *5th International ASRANet Conference*. 760
- Schepsmeier, U. (2015). "Efficient information based goodness-of-fit tests for vine copula models with fixed margins." *Journal of Multivariate Analysis*, 138: 34–52. MR3348832. doi: <https://doi.org/10.1016/j.jmva.2015.01.001>. 750
- Stöber, J., Joe, H., and Czado, C. (2013). "Simplified Pair Copula Constructions – Limitations and Extensions." *Journal of Multivariate Analysis*, 119: 101–118. MR3061418. doi: <https://doi.org/10.1016/j.jmva.2013.04.014>. 750
- van Dorp, J. and Mazzuchi, T. (2003). "Parameter specification of the beta distribution and its Dirichlet extensions utilizing quantiles." *Beta Distribution and Its Applications*, 29: 1–37. 749
- Wilson, K., Quigley, J., Bedford, T., and Walls, L. (2013). "Bayes linear Bayes graphical models in the design of optimal test strategies." *International Journal of Performance Engineering*, 9: 715–728. 750
- Wilson, K. J. (2017). "Supplementary Material to: Specification of informative prior distributions for multinomial models using vine copulas." *Bayesian Analysis*. doi: <https://doi.org/10.1214/17-BA1068SUPP>. 760, 761, 763
- Zapata-Vázquez, E., O'Hagan, A., and Bastos, L. (2014). "Eliciting expert judgements about a set of proportions." *Journal of Applied Statistics*, 41: 1919–1933. MR3292618. doi: <https://doi.org/10.1080/02664763.2014.898131>. 749

Acknowledgments

We would like to thank John Quigley and Lesley Walls for their original work which formed the basis of the application presented in the paper. We are grateful to two reviewers, an associate editor and the editor for advice which transformed the paper.