

# Open Research Online

---

The Open University's repository of research publications and other research outputs

## Eliciting Dirichlet and Gaussian copula prior distributions for multinomial models

### Journal Item

How to cite:

Elfadaly, Fadlalla G. and Garthwaite, Paul H. (2017). Eliciting Dirichlet and Gaussian copula prior distributions for multinomial models. *Statistics and Computing*, 27(2) pp. 449–467.

For guidance on citations see [FAQs](#).

© 2016 The Authors

Version: Version of Record

Link(s) to article on publisher's website:  
<http://dx.doi.org/doi:10.1007/s11222-016-9632-7>

---

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online's data [policy](#) on reuse of materials please consult the policies page.

---

[oro.open.ac.uk](http://oro.open.ac.uk)

# Eliciting Dirichlet and Gaussian copula prior distributions for multinomial models

Fadlalla G. Elfadaly<sup>1,2</sup> · Paul H. Garthwaite<sup>1</sup>

Received: 19 June 2015 / Accepted: 26 January 2016 / Published online: 19 February 2016  
© The Author(s) 2016. This article is published with open access at Springerlink.com

**Abstract** In this paper, we propose novel methods of quantifying expert opinion about prior distributions for multinomial models. Two different multivariate priors are elicited using median and quartile assessments of the multinomial probabilities. First, we start by eliciting a univariate beta distribution for the probability of each category. Then we elicit the hyperparameters of the Dirichlet distribution, as a tractable conjugate prior, from those of the univariate betas through various forms of reconciliation using least-squares techniques. However, a multivariate copula function will give a more flexible correlation structure between multinomial parameters if it is used as their multivariate prior distribution. So, second, we use beta marginal distributions to construct a Gaussian copula as a multivariate normal distribution function that binds these marginals and expresses the dependence structure between them. The proposed method elicits a positive-definite correlation matrix of this Gaussian copula. The two proposed methods are designed to be used through interactive graphical software written in Java.

**Keywords** Dirichlet distribution · Elicitation method · Gaussian copula elicitation · Interactive graphical software · Multinomial model · Prior distribution

## 1 Introduction

Bayesian statistical methods provide a formal mechanism for taking into account prior knowledge. In many circumstances, prior knowledge is based on historical data that are only recorded in the form of the personal experience of experts. Then expert opinion must be quantified as a prior distribution if the information is to be used. This can be done through an elicitation process which is defined as asking the expert for information that is then used as a basis for encoding the parameters of a subjective prior distribution. The method of quantifying opinion must be designed for the sampling model of interest and the form of prior distribution that is to be used.

One important sampling model that has attracted limited attention in the elicitation context is the multinomial model. This model is used in many scientific and industrial applications. For example, it is frequently applied to voting behavior in political science (e.g. Chamberlin and Featherston 1986), to the compositions of rocks in geology (e.g. Weltje 2002) and to patterns of consumer selection preferences in micro-economics (e.g. Manski 2007). The model is so common that good elicitation methods for quantifying opinion about its parameters are clearly needed. For example, in their recent applied research, Fu et al. (2012) and Conn et al. (2013) stated that an elicited informative prior for the multinomial model would be a possible alternative.

Here we are concerned with the most common case of the multinomial model, where each observation has the same probability of falling into any specified category and observations are independent of each other. Then observations can be modelled to follow a multinomial distribution with, say, probability  $p_i$  that an observation falls in the  $i$ th category. To quantify an expert's opinions about the  $p_i$ , we assume some distributional form to represent her opinion and uncertainty. Instead of directly asking the expert to assess the param-

---

✉ Fadlalla G. Elfadaly  
f.elfadaly@open.ac.uk

<sup>1</sup> Department of Mathematics and Statistics, The Open University, Milton Keynes, UK

<sup>2</sup> Department of Statistics, Faculty of Economics and Political Science, Cairo University, Giza, Egypt

ters of the distribution, we ask her to make assessments that determine appropriate values for the parameters of that distribution, which are referred to as its hyperparameters.

The Dirichlet distribution, which is the conjugate prior distribution for the parameters of a multinomial distribution, is one natural choice for representing an expert's opinion. A number of elicitation methods that follow this approach have been proposed in the literature. For example, [Bunn \(1978\)](#) encoded a Dirichlet prior from assessed beta marginal distributions through hypothetical future samples. [Chaloner and Duncan \(1987\)](#) based their elicitation method on assessing a sample size and modal values of the Dirichlet variates. [Dorp and Mazzuchi \(2004\)](#) introduced a numerical algorithm for encoding the Dirichlet parameters from a number of quantile assessments that is equal to the number of parameters. Recently, two elicitation methods have been proposed for the Dirichlet distribution by reconciling the assessed parameters of its univariate beta conditional distributions ([Elfadaly and Garthwaite 2013](#)) and of its univariate marginal beta distributions ([Zapata-Vázquez et al. 2014](#)).

The first method proposed in the current paper elicits the hyperparameters of the Dirichlet distribution from those of its marginal beta distributions through forms of reconciliation that use least-squares techniques. The usage of least-squares techniques for reconciliation has been discussed in [Albert et al. \(2012\)](#). Our proposed method differs from those of [Bunn \(1978\)](#) and [Chaloner and Duncan \(1987\)](#) as we require only medians and quartiles to be assessed. We assess more quartiles than [Dorp and Mazzuchi \(2004\)](#) and then reconcile them to encode hyperparameters. Our method encodes a Dirichlet distribution using its beta marginal distributions rather than its conditional distributions as in [Elfadaly and Garthwaite \(2013\)](#). Our proposed reconciliation methods are different from those that have been independently developed in [Zapata-Vázquez et al. \(2014\)](#) for reconciling beta marginal distributions into a Dirichlet distribution for a set of proportions.

However, while the standard Dirichlet distribution offers tractability and mathematical simplicity, it has been criticized as too inflexible to represent a broad range of prior information about the parameters of multinomial models ([Aitchison 1986](#); [O'Hagan and Forster 2004](#)). Specifically, the Dirichlet variates are always negatively correlated and the Dirichlet distribution contains the same number of parameters as the number of categories. With  $k$  categories,  $k - 1$  parameters are required to describe the proportion in each category, leaving only one parameter to describe their dependence structure. This may sometimes be insufficient to give a reasonable representation of an expert's opinions.

In a Dirichlet prior distribution, the dependence structure that is imposed by only one parameter is a pattern of negative correlations between the multinomial probabilities of all categories. This negative correlation pattern is not appropriate

in many practical situations. For example, if the categories represent the political parties in some country, with the corresponding multinomial probabilities reflecting the voting behavior in the next elections, a high probability of voting for a specific radical party might be accompanied by rather high probabilities of voting for other radical parties. A suitable prior in this case should imply a positive correlation pattern between parties of the same political views.

In the case of multinomial models, a particular difficulty is to elicit assessments that satisfy all the constraints of mathematical coherence. For example, the probabilities of each category must be non-negative and sum to one, which we refer to as the unit-sum constraint. Other constraints are implicit and less intuitive. We believe that a well-designed elicitation method should lead to coherent assessments without the expert having to be conscious of coherence constraints. This is not straightforward, especially with dependent variables in multivariate distributions. The expert has no built-in prior distribution to assess directly. Instead, she must be asked to quantify her opinions through questions she can both comprehend and answer. We believe it is important that the expert should be able to focus solely on her opinions when giving assessments.

To date, perhaps due to the difficulty discussed above, elicitation methods for multinomial sampling have mostly been proposed for modelling opinion by a Dirichlet distribution. An exception is the work by [Elfadaly and Garthwaite \(2013\)](#), who model opinion by a Connor–Mosimann distribution. However, this form of generalized Dirichlet distribution provides a structure that is only slightly richer than the standard Dirichlet distribution. Methods that yield a more flexible distribution than the Dirichlet distribution are long overdue. The present paper focuses on elicitation methods that yield a much more flexible distribution than the Dirichlet prior distribution.

Motivated by the deficiencies of the standard Dirichlet distribution, several authors have constructed new families of distributions for proportions that allow more general types of dependence structures. These distributions include, for example, the generalized Dirichlet distribution ([Connor and Mosimann 1969](#)), the multivariate normal distribution of log contrasts ([Leonard 1975](#)), and the multivariate logistic normal distribution ([Aitchison 1986](#); [Elfadaly and Garthwaite 2016](#)).

A possible general multivariate distribution, that can serve as a prior distribution for multinomial models, is constructed through using a multivariate copula function. A copula is a multivariate function that represents a joint multivariate cumulative distribution function (CDF) in terms of one-dimensional marginal CDFs. Hence, it joins marginal distributions into a multivariate distribution that has those marginals. The theoretical foundation of copula functions was introduced in [Sklar \(1959\)](#), where he gave Sklar's Theo-

rem which states that any joint distribution can be written in a copula form. The marginal distributions can thus be chosen independently from the dependence structure that is represented by the copula function. For an introduction to copulas, see for example Joe (1997), Frees and Valdez (1998) or Nelsen (1999).

The use of copula functions to elicit multivariate distributions has been considered in the literature; see Jouini and Clemen (1996) or Clemen and Reilly (1999), among others. In general, the joint distribution can be elicited by first assessing each marginal distribution. Then the dependence structure is elicited through the copula function. Different families and classes of copula functions have been defined for both bivariate and multivariate distributions. Jouini and Clemen (1996) used bivariate and multivariate Archimedean families and Frank's families of copulae to aggregate multiple experts' opinions about a random quantity. The book by Kurowicka and Cooke (2006) lists and compares different copula functions and discusses many other methods of modelling high-dimensional dependence. However, the simplest and most intuitive family of copulae is the inversion copula (Nelsen 1999). This has the form

$$C[G_1(x_1), \dots, G_k(x_k)] = F_{(1, \dots, k)} \left\{ F_1^{-1}[G_1(x_1)], \dots, F_k^{-1}[G_k(x_k)] \right\}, \quad (1)$$

where  $G_i$  are the known marginal CDFs of  $x_1, \dots, x_k$ ,  $F_{(1, \dots, k)}$  is their assumed multivariate CDF, and its marginals are  $F_i$ . Hence, the marginal functions  $G_1, \dots, G_k$  are coupled through  $F_{(1, \dots, k)}$ , via the quantile functions evaluated for each marginal, into a new multivariate distribution given by the copula function  $C$ . It has to be mentioned here that the arguments of the inversion copula function in (1) are the quantile functions  $F_i^{-1}[\cdot]$  ( $i = 1, \dots, k$ ). It is therefore appealing to exploit the inversion copula in our elicitation method in which we ask the expert to assess only medians and quartiles. Experimental work in the literature suggests that the assessment of quantile functions is often the best way to elicit information from experts. See for example Garthwaite et al. (2005) and the references therein. Specifically, we believe that medians and quartiles are easier for an expert to assess than other quantile functions as they can be obtained by the first two steps of equally likely subdivisions using the bisection method (Garthwaite et al. 2005).

The distribution  $F_{(1, \dots, k)}$  in (1) has sometimes been taken as a multivariate  $t$ -distribution (Demarta and McNeil 2005), or even as a Dirichlet distribution (Lewandowski 2008). Most commonly though, it is selected as a multivariate normal distribution with a joint CDF,  $F_{(1, \dots, k)} = \Phi_{k, \mathbf{R}}$ , and standard normal marginal CDFs,  $F_i = \Phi$ . This gives a Gaussian copula (Clemen and Reilly 1999) that is parameterized by the

correlation matrix  $\mathbf{R}$  of a multivariate normal distribution. To elicit  $\mathbf{R}$ , Clemen and Reilly (1999) suggest that a pairwise rank-order correlation between each  $X_i$  and  $X_j$ , such as Spearman's  $\rho_{i,j}$  or Kendall's  $\tau_{i,j}$ , should be assessed. Then properties of the multivariate normal distribution are used to transform them into the product-moment Pearson correlation  $r_{i,j}$  as follows:

$$r_{i,j} = 2 \sin(\pi \rho_{i,j}/6), \quad \text{or} \quad r_{i,j} = \sin(\pi \tau_{i,j}/2). \quad (2)$$

The product-moment correlation matrix  $\mathbf{R}$  is then formed from the elements  $r_{i,j}$ . In practice, however, it may be preferable to elicit correlations on a numerical scale rather than rank-order correlations. In this paper, we use quartile assessments to obtain the product-moment correlation matrix  $\mathbf{R}$ .

Clemen and Reilly (1999) favour eliciting rank-order correlations, not product-moment Pearson correlation, on the grounds that the latter cannot necessarily be transformed through the function  $\Phi^{-1}[G_i(\cdot)]$ . In contrast, rank-order correlations are transformation respecting regardless of the choice of the marginal CDF  $G_i(\cdot)$ , i.e. they are invariant under strictly monotone increasing transformations of the form  $\Phi^{-1}[G_i(\cdot)]$ , for any  $G_i(\cdot)$ . To elicit these correlations, Clemen and Reilly (1999) mention three methods that can be used either separately or together. However, none of the methods is guaranteed to yield a positive-definite correlation matrix. On the other hand, the elicitation method of Kadane et al. (1980) has been designed to encode a positive-definite covariance matrix of a multivariate  $t$ -distribution as a conjugate prior for the hyperparameters of a normal multiple linear regression model. Their method can be useful in a variety of multivariate elicitation problems that require the assessment of positive-definite matrices (Garthwaite et al. 2005, 2013). These clearly include the problem addressed here, so that we modify their method to form a Gaussian copula.

Interestingly, no other elicitation method for Gaussian copulas seems to encode a positive-definite correlation matrix using quartile assessments. To fill this gap, the second proposed method in this paper elicits a Gaussian copula function with beta marginal distributions as a prior distribution for multinomial models. Our approach simultaneously overcomes two problems of the method of Clemen and Reilly (1999). First, we transform the assessed conditional quartiles of  $X_i$  and  $X_j$ , through  $\Phi^{-1}[G_i(\cdot)]$ , which enables product-moment correlations to be computed on the normal scale without the need of rank-order correlations. Second, the conditional quartiles are assessed according to the structural elicitation procedure of Kadane et al. (1980), which guarantees that the elicited correlation matrix is positive-definite.

The paper is organized as follows. In Sect. 2, we discuss our choice of distribution, the Dirichlet prior and the

Gaussian copula prior, for representing expert uncertainty relevant to proportions of a categorical variable. We show how this can be used to specify a prior for the probabilities of a multinomial distribution. In Sect. 3 we describe the assessment tasks that an expert performs, and in Sect. 4 we derive hyperparameters of both forms of prior distributions from the elicited assessments. Sections 3 and 4 also describe an interactive graphical interface that has been developed to implement the two proposed elicitation methods. In Sect. 5, these methods and the implementing software are used by an environmental ecologist to quantify his opinion about the future proportions of different phenotypes of a specific snail species in the UK. Concluding comments are given in Sect. 6.

## 2 Two multivariate prior distributions for a multinomial model

### 2.1 The multinomial likelihood

Let the random vector  $\mathbf{W} = (W_1, \dots, W_k)$  be multinomially distributed with  $k$  categories,  $n$  trials and a vector of probabilities  $\mathbf{p} = (p_1, \dots, p_k)$ , so that

$$f(w_1, w_2, \dots, w_k) = \frac{n!}{w_1!w_2! \dots w_k!} p_1^{w_1} p_2^{w_2} \dots p_k^{w_k}, \quad (3)$$

where  $\sum w_i = n$ ,  $0 \leq w_i \leq n$ ,  $0 \leq p_i \leq 1$  and  $\sum p_i = 1$ .

### 2.2 The Dirichlet prior distribution

A conjugate prior for the parameter vector  $\mathbf{p}$  is the Dirichlet distribution of the form

$$\pi(p_1, p_2, \dots, p_k) = \frac{\Gamma(N)}{\Gamma(a_1)\Gamma(a_2) \dots \Gamma(a_k)} p_1^{a_1-1} p_2^{a_2-1} \dots p_k^{a_k-1}, \quad (4)$$

where  $\Gamma(\cdot)$  is the gamma function defined by  $\Gamma(a) = \int_{t=0}^{\infty} t^{a-1} e^{-t} dt$ ,  $N = \sum a_i$  and  $a_i$  are the hyperparameters representing the relative size of each category, so  $a_i > 0$ .

The expectation and variance of  $p_i$  are  $a_i/N$  and  $a_i(N - a_i)/[N^2(N + 1)]$  and the covariance between  $p_i$  and  $p_j$  is  $-a_i a_j / [N^2(N + 1)]$ , where  $i = 1, \dots, k$ ;  $j = 1, \dots, k$  and  $i \neq j$ . The marginal distribution of each  $p_i$  ( $i = 1, \dots, k$ ) is a beta distribution (e.g. Wilks 1962),

$$g_i(p_i) = \frac{\Gamma(\alpha_i + \beta_i)}{\Gamma(\alpha_i)\Gamma(\beta_i)} p_i^{\alpha_i-1} (1 - p_i)^{\beta_i-1}, \quad (5)$$

where

$$\alpha_i = a_i \quad \text{and} \quad \beta_i = \sum_{j \neq i} a_j. \quad (6)$$

To elicit the vector of hyperparameters  $\mathbf{a} = (a_1, \dots, a_k)$ , we will exploit this direct relationship between the Dirichlet distribution and its *marginal* beta distributions.

### 2.3 Constructing a Gaussian copula prior distribution

Rather than assuming that the marginal beta distributions in (5) stem from a Dirichlet distribution, we would like to allow a more flexible dependence structure via another joint distribution, with the aim of better representing the expert's opinion. A flexible tool for this task is given by the copula function, as discussed in Sect. 1. This allows us to choose the marginal distributions independently from the dependence structure between them.

The best-known example of the inversion copula in (1) is the Gaussian copula (Clemen and Reilly 1999). For any random vector  $\mathbf{X} = (X_1, \dots, X_k)$ , the Gaussian copula is defined at the point  $(x_1, \dots, x_k)$  as

$$C[G_1(x_1), \dots, G_k(x_k)] = \Phi_{k,\mathbf{R}} \left\{ \Phi^{-1}[G_1(x_1)], \dots, \Phi^{-1}[G_k(x_k)] \right\}. \quad (7)$$

Here  $\Phi_{k,\mathbf{R}}$  is the CDF of a  $k$ -variate normal distribution with zero means, unit variances, and a correlation matrix  $\mathbf{R}$  that reflects the desired dependence structure. The function  $\Phi$  is the marginal standard univariate normal CDF, and  $G_i$  ( $i = 1, \dots, k$ ) are any chosen marginal CDFs. The marginal mean and variance of each  $x_i$  are expressed through its marginal CDF,  $G_i(x_i)$ , hence, no other parameters are required for  $\Phi_{k,\mathbf{R}}$  beyond the correlation matrix  $\mathbf{R}$ . To be used as a prior distribution, the density function of the Gaussian copula is needed. Fortunately, since  $\Phi_{k,\mathbf{R}}$  and  $\Phi$  are differentiable, the Gaussian copula density function is easily obtained by differentiating (7) with respect to  $x_i$  ( $i = 1, \dots, k$ ); see for example Clemen and Reilly (1999). This gives

$$f(x_1, \dots, x_k | \mathbf{R}) = \frac{\prod_{j=1}^k g_j(x_j)}{|\mathbf{R}|^{1/2}} \exp \left\{ -\frac{1}{2} \mathbf{y}'_k (\mathbf{R}^{-1} - \mathbf{I}_k) \mathbf{y}_k \right\}, \quad (8)$$

where  $\mathbf{y}'_k = (\Phi^{-1}[G_1(x_1)], \dots, \Phi^{-1}[G_k(x_k)])$ ,  $g_i(\cdot)$  is the density function corresponding to the choice of  $G_i(\cdot)$  ( $i = 1, \dots, k$ ), and  $\mathbf{I}_k$  is the identity matrix of order  $k$ .

To construct a Gaussian copula function in the case of a multinomial model, it might seem natural to assume that each marginal distribution of  $p_i$  ( $i = 1, \dots, k$ ) is a beta distribution, as in (5), and try to construct a Gaussian copula function for the multivariate distribution of  $\mathbf{p}$ . However, applying the normalizing transformations  $Y_i = \Phi^{-1}[G_i(p_i)]$  directly on the multinomial probabilities  $p_i$ , where  $G_i(\cdot)$  is the assessed beta CDF, would raise three problems. First, encoding marginal beta distributions directly on the elements of  $\mathbf{p}$  might



result in mathematically incoherent marginal distributions with mean values that do not sum to one. Second, the inverse normalizing transformations  $p_i = G_i^{-1}[\Phi(Y_i)]$ , while preserving the marginal domain of each probability,  $0 \leq p_i \leq 1$ , do not necessarily fulfil the unit-sum constraint  $\sum_{i=1}^k p_i = 1$ . Third, under the unit-sum constraint on  $\mathbf{p}$ , the normally distributed vector  $\mathbf{y}'_k = (Y_1, \dots, Y_k)$  contains a redundant variable and hence leads to a singular multivariate normal distribution.

To overcome these problems, we define new variables  $Z_1, \dots, Z_k$ , where

$$Z_1 = p_1, \quad Z_i = \frac{p_i}{1 - \sum_{j=1}^{i-1} p_j}, \quad \text{for } i = 2, \dots, k-1, \quad \text{and} \quad Z_k = 1. \tag{9}$$

The inverse transformations are given by

$$p_1 = Z_1, \quad p_i = Z_i \prod_{j=1}^{i-1} (1 - Z_j), \quad \text{for } i = 2, \dots, k. \tag{10}$$

Instead of assuming a beta marginal distribution of the form (5) for each  $p_i$  ( $i = 1, \dots, k$ ), we assume that each  $Z_i$  ( $i = 1, \dots, k-1$ ) has a marginal beta distribution with parameters  $\alpha_i^\#$  and  $\beta_i^\#$  to be encoded as in Sect. 4.1. Then we construct a Gaussian copula function for the multivariate distribution of  $\mathbf{z}_{k-1} = (Z_1, \dots, Z_{k-1})$ . The beta marginal distributions preserve the domain of each  $Z_i$  ( $i = 1, \dots, k-1$ ) on  $[0, 1]$ , while the Gaussian copula accounts only for their dependence structure. Here, no other constraints need to be fulfilled on  $\mathbf{z}_{k-1}$ , while the unit-sum constraint on  $\mathbf{p}$  is always attained by virtue of the transformations in (9) and (10).

Using the Gaussian copula function, the dependence structure of the multivariate distribution of  $Z_i$ , and consequently of  $p_i$ , will have high flexibility rather than the limited dependence structure imposed by the Dirichlet distribution. In the underlying multinomial model, each multinomial parameter  $p_i$  represents the probability that an observation falls in the  $i$ th category. Hence,  $Z_i$  ( $i = 1, \dots, k-1$ ) can be defined as the probability that an observation falls in category  $i$ , given that the first  $i-1$  categories have been eliminated and the available categories are only the last  $k-i+1$  categories.

Connor and Mosimann (1969) argued that if  $Z_i$  ( $i = 1, \dots, k-1$ ) are mutually independent and each  $Z_i \sim \text{beta}(\alpha_i^\#, \beta_i^\#)$ , then  $\mathbf{p}$  has a generalized Dirichlet distribution with parameters  $(\alpha_1^\#, \dots, \alpha_{k-1}^\#, \beta_1^\#, \dots, \beta_{k-1}^\#)$ . Thus the generalized Dirichlet distribution can be regarded as a special case of the Gaussian copula in which  $Z_1, \dots, Z_{k-1}$  are independent. If, moreover,  $\beta_i^\# = \alpha_{i+1}^\# + \beta_{i+1}^\#$ , for  $i = 1, \dots, k-2$ , then  $\mathbf{p}$  has a standard Dirichlet distribution with parameters  $(\alpha_1^\#, \dots, \alpha_{k-1}^\#, \beta_{k-1}^\#)$ , so the standard Dirichlet distribution is also a special case of the Gaussian

copula prior. This confirms that the Gaussian copula function imposes a prior distribution on  $\mathbf{p}$  with a dependence structure that is more flexible than that of the standard and generalized Dirichlet distributions.

The prior distribution for  $\mathbf{p}$  is defined in terms of the prior distribution of  $\mathbf{z}_{k-1}$ . Let  $G_i(Z_i)$  be the CDF of the beta distribution of  $Z_i$  with hyperparameters  $\alpha_i^\#$  and  $\beta_i^\#$  ( $i = 1, \dots, k-1$ ) to be elicited in Sect. 4.1. We assume that the joint density of  $\mathbf{z}_{k-1}$  is given by a Gaussian copula density,

$$f(z_1, \dots, z_{k-1} | \mathbf{R}) = \frac{\prod_{j=1}^{k-1} g_j(z_j)}{|\mathbf{R}|^{1/2}} \exp \left\{ -\frac{1}{2} \mathbf{y}'_{k-1} (\mathbf{R}^{-1} - \mathbf{I}_{k-1}) \mathbf{y}_{k-1} \right\}, \tag{11}$$

where  $\mathbf{y}'_{k-1} = (\Phi^{-1}[G_1(Z_1)], \dots, \Phi^{-1}[G_{k-1}(Z_{k-1})])$ , and  $g_i(\cdot)$  is the beta density of  $Z_i$  ( $i = 1, \dots, k-1$ ). The correlation matrix  $\mathbf{R}$  and the parameters of the beta densities are the hyperparameters of the Gaussian copula prior distribution.

Note that the marginal distributions of this joint density are still the same beta marginal distributions. After eliciting the hyperparameters of the beta distribution for each  $Z_i$  ( $i = 1, \dots, k-1$ ) as detailed in Sect. 4.1, the prior distribution is totally known except for the matrix  $\mathbf{R}$ . Here,  $\mathbf{R}$  is the correlation matrix of the multivariate normally distributed vector  $\mathbf{y}_{k-1}$ ; it needs to be encoded effectively and must be a positive-definite matrix. The aim is to choose  $\mathbf{R}$  so as to model the expert’s opinion about the dependence between the  $p_i$ s. In Sect. 4.3, we introduce a method, inspired by Kadane et al. (1980), to encode the correlation matrix  $\mathbf{R}$ . Although the density in (11) is neither a multivariate normal for  $\mathbf{p}'_{k-1} = (p_1, \dots, p_{k-1})$  nor  $\mathbf{z}'_{k-1} = (Z_1, \dots, Z_{k-1})$ , and the correlation matrix of  $\mathbf{z}_{k-1}$  is not an explicit function in  $\mathbf{R}$ , the dependence structure between the elements of  $\mathbf{p}_{k-1}$ , and consequently that of  $Z_{k-1}$ , is indirectly reflected by  $\mathbf{R}$ . Moreover, we can still use the multivariate normal properties to form a positive-definite matrix  $\mathbf{R}$  by considering the following normalizing transformations,

$$Y_i = \Phi^{-1}[G_i(Z_i)], \quad i = 1, \dots, k-1. \tag{12}$$

Under the main assumption of the Gaussian copula construction, and from (12), the vector  $\mathbf{y}'_{k-1} = (Y_1, \dots, Y_{k-1})$  has a multivariate normal distribution with zero means, unit variances and a correlation matrix  $\mathbf{R}$ , so  $\mathbf{y}_{k-1} \sim \text{MVN}(\mathbf{0}, \mathbf{R})$ . In our proposed approach, the matrix  $\mathbf{R}$  is encoded as a covariance and a correlation matrix of the multivariate normal random vector  $\mathbf{y}_{k-1}$ . This is achieved by utilizing the monotone increasing property of the transformations in (12). We assess conditional quartiles of  $\mathbf{p}_{k-1}$ , then transform them into those of  $\mathbf{z}_{k-1}$  and  $\mathbf{y}_{k-1}$  using (9) and (12), respectively.

Correlation coefficients between the elements of  $\mathbf{y}_{k-1}$  can then be estimated using their conditional quartiles and utilizing properties of the multivariate normal distribution. This is described in Sects. 3.2.2, 3.2.3 and 4.3.

### 3 Assessment tasks

With both the Dirichlet and Gaussian copula prior distributions, marginal distributions are each of the beta type. In both cases, the first set of assessments determines the hyperparameters of the marginal beta distributions. For the standard Dirichlet distribution, as detailed in Sect. 3.1 the expert is asked to assess marginal medians and quartiles of the probability of each category. These quartiles are the median and quartiles of the expert's subjective probability distribution for each  $p_i$ ; the distance between the quartiles reflects her uncertainty about the true value of  $p_i$ . The marginal median and quartile assessments are obtained in Sect. 3.1.1. They are used to encode the marginal beta distribution of each  $p_i$  ( $i = 1, \dots, k$ ). In the second part of the elicitation process, detailed in Sect. 3.2, we obtain the required assessments to construct a Gaussian copula prior distribution. Specifically, in Sect. 3.2.1, we obtain the marginal median and quartile assessments that are required to encode the marginal beta distributions of  $Z_i$  ( $i = 1, \dots, k - 1$ ). Further sets of conditional medians and quartiles are obtained in Sects. 3.2.2 and 3.2.3; they will be used to encode the correlation matrix  $\mathbf{R}$  of the Gaussian copula prior.

To ensure that the expert understands the request to assess her median estimate of a proportion, she is told: “You should consider it equally likely that the true proportion is above the median or below the median. For example, suppose you assess the median as 0.6, you should think it equally likely that the proportion will be above 0.6 as that it will be below 0.6. Hence, if you were asked ‘Would I rather bet on the proportion being above 0.6 or below 0.6?’, you should find it hard to decide.” For the lower (upper) quartile assessment, the expert is told: “The lower (upper) quartile divides the probability below (above) the median in half. Suppose your assessed median is 0.3 and your assessed lower quartile is 0.2. Then your probability that the proportion is below 0.2 should equal your probability that it is between 0.2 and 0.3.”

#### 3.1 Assessments required for the standard Dirichlet distribution

##### 3.1.1 Obtaining marginal assessments for $\mathbf{p}$

For each category in turn, the expert is asked to assess a lower quartile, a median and an upper quartile for the probability

$p_i$ ,  $i = 1, \dots, k$ , say  $\bar{L}_i$ ,  $\bar{m}_i$  and  $\bar{U}_i$ , respectively. In the practical evolutionary example detailed in Sect. 5, the expert assessed his median, lower and upper quartiles for the future proportion of each of the 6 different phenotypes of the most common snail species in the UK, see Table 1. Here we adopt a marginal approach, so that the categories are interchangeable and their order does not matter. In Sect. 4.1, the marginal medians and quartiles assessed here are used to encode the parameters  $\alpha_i$  and  $\beta_i$  of each marginal beta distribution of  $p_i$  for the standard Dirichlet distribution.

#### 3.2 Assessments required for the Gaussian copula distribution

##### 3.2.1 Obtaining marginal assessments for $\mathbf{z}_{k-1}$

For category  $i$  ( $i = 2, \dots, k - 1$ ), the expert is asked to assume that the first  $i - 1$  categories have been eliminated from the analysis and that she should only consider the remaining  $k - i + 1$  categories, namely categories  $i, i + 1, \dots, k$ . That is, the expert should assume that the observation does not fall in one of the first  $i - 1$  categories, but which of the remaining category it falls in is unknown. Given this situation, the expert is asked to assess the median of the probability of each of the categories  $i, i + 1, \dots, k$ . Only the assessment for the first of these categories is used and we denote it  $m_{i,0}^\#$ . The other assessments are elicited to improve internal consistency; assessments of means should sum to one so the assessments of the medians should be close to one.

In fact, these assessments represent the marginal unconditional medians of each  $Z_i = p_i / (1 - \sum_{j=1}^{i-1} p_j)$ , for  $i = 2, \dots, k - 1$ . After these medians have been assessed, the expert assesses the marginal unconditional lower and upper quartile of each  $Z_i$ . We denote these assessments as  $L_{i,0}^\#$  and  $U_{i,0}^\#$  ( $i = 2, \dots, k - 1$ ). Clearly, since  $Z_1 = p_1$ , we have  $L_{1,0}^\# = \bar{L}_1, m_{1,0}^\# = \bar{m}_1$  and  $U_{1,0}^\# = \bar{U}_1$  and these are assessed as in Sect. 3.1.1. Moreover, by definition,  $m_{k,0}^\# = 1$ , since  $Z_k = 1$ .

If a Gaussian copula prior distribution is to be encoded, we define  $m_{1,0}^*$  as the median of  $p_1$ , and  $m_{i,0}^*$  as the median of  $(p_i | p_1 = m_{1,0}^*, \dots, p_{i-1} = m_{i-1,0}^*)$  for  $i = 2, \dots, k$ . But we do not directly assess  $m_{i,0}^*$  ( $i = 1, \dots, k$ ). Instead, we compute their values from the corresponding marginal medians,  $m_{i,0}^\#$ , of  $Z_i$  ( $i = 1, \dots, k - 1$ ) assessed above. This can be done by utilizing Lemma 1 (below) to obtain  $m_{i,0}^*$  in terms of  $m_{i,0}^\#$ . The former will be used as the conditioning values in Sects. 3.2.2 and 3.2.3.

**Lemma 1** *Under the unit-sum constraint on  $\mathbf{p}$ , and the multivariate normality of  $\mathbf{y}_{k-1}$ , we have*

$$m_{1,0}^* = m_{1,0}^\# \quad \text{and} \quad m_{i,0}^* = m_{i,0}^\# \prod_{j=1}^{i-1} (1 - m_{j,0}^\#),$$

for  $i = 2, \dots, k$ .

See the Appendix for a proof of this lemma.

Assessments obtained in this section are used in Sect. 4.1 to encode the hyperparameters  $\alpha_i^\#$  and  $\beta_i^\#$  of the marginal beta distribution of  $Z_i$  ( $i = 1, \dots, k - 1$ ) for the Gaussian copula prior distribution.

### 3.2.2 Assessing conditional quartiles

After finishing the marginal assessment tasks in Sect. 3.2.1, the expert is asked to give conditional quartile assessments for specified categories under the assumption that the initial assessments are the correct values for the other categories. Specifically, the expert is asked to assume that  $p_1$  equals its initially assessed median value, i.e.  $p_1 = m_{1,0}^*$ , and then gives a lower quartile  $L_2^*$  and an upper quartile  $U_2^*$  for  $p_2$ . For each remaining  $p_j$ ,  $j = 3, \dots, k - 1$ , the expert assesses the two quartiles  $L_j^*$  and  $U_j^*$  conditional on  $p_1 = m_{1,0}^*, \dots, p_{j-1} = m_{j-1,0}^*$ .

For the last category, the lower (upper) quartile  $L_k^*$  ( $U_k^*$ ) of  $p_k$  is automatically shown to the expert once she assesses the upper (lower) quartile  $U_{k-1}^*$  ( $L_{k-1}^*$ ) of  $p_{k-1}$ . The two quartiles  $L_k^*$  and  $U_k^*$  are shown to the expert as a guide to help her choose  $L_{k-1}^*$  and  $U_{k-1}^*$ . When  $p_1 = m_{1,0}^*, \dots, p_{k-2} = m_{k-2,0}^*$ , from the unit-sum constraint it follows that  $L_{k-1}^* + U_k^* = U_{k-1}^* + L_k^* = 1 - m_{1,0}^* - \dots - m_{k-2,0}^*$ .

In practice, these conditional assessments constitute an easy task for the experts to perform. They normally think of successive category assessments in a conditional way given their already assessed values for the preceding categories. For conditional assessments, the order of the categories is important. We recommend ordering them from the most probable category to the least probable, unless there is a natural order to the categories. This ordering leads to assessment tasks that involve less skewed distributions. See [Elfadaly and Garthwaite \(2013\)](#) for more detail and examples on this recommended category ordering.

The conditional quartile assessments of  $p_i | p_1, \dots, p_{i-1}$  obtained in this section, namely  $L_i^*$  and  $U_i^*$  ( $i = 2, \dots, k - 1$ ), are transformed to the corresponding conditional quartiles of  $Z_i | Z_1, \dots, Z_{i-1}$ , say  $L_i^\#$  and  $U_i^\#$  ( $i = 2, \dots, k - 1$ ) as follows.

$$L_i^\# = \frac{L_i^*}{1 - \sum_{j=1}^{i-1} m_{j,0}^*} \quad \text{and} \quad U_i^\# = \frac{U_i^*}{1 - \sum_{j=1}^{i-1} m_{j,0}^*}.$$

$L_i^\#$  and  $U_i^\#$  are used to estimate the correlation matrix  $\mathbf{R}$  of the Gaussian copula prior, as will be discussed in Sect. 4.3.

### 3.2.3 Assessing conditional medians

First the expert is asked to assume that the true value of  $p_1$  is  $m_{1,1}^*$ , while her median assessment of  $p_1$  had been  $m_{1,0}^*$ . The value specified for  $m_{1,1}^*$  does not equal  $m_{1,0}^*$  and we let  $\eta_1^* = m_{1,1}^* - m_{1,0}^*$ . Given the information that  $p_1 = m_{1,1}^*$ , the expert is asked to change her previous medians  $m_{j,0}^*$  of each  $p_j$ , and her new (conditional) median is denoted by  $m_{j,1}^*$  ( $j = 2, \dots, k$ ). In practice, this question can be phrased to the expert as: ‘‘Suppose that the proportion in the first category were actually changed from your original assessment to the value given on the interactive software graph. If that were the case, what would be your median estimates of the proportions in the other categories?’’ Here, the expert is asked to sequentially assume that each conditional median assessment she gives at this stage is the true value and that she should condition on each of them to be true while she is assessing the remaining conditional medians. For example, when the expert is asked to assess  $m_{4,1}^*$ , she does not only assume that  $p_1 = m_{1,1}^*$ , but she also assumes that  $p_2 = m_{2,1}^*$  and  $p_3 = m_{3,1}^*$ .

In each successive step  $i$ , for  $i = 2, \dots, k - 2$ , the expert is asked to suppose that the true value of  $p_i$  is  $m_{i,i}^* = m_{i,0}^* + \eta_i^*$  (while her initially calculated median of  $p_i$  had been  $m_{i,0}^*$ ) and that the true values of  $p_1, \dots, p_{i-1}$  are equal to their initially calculated medians,  $m_{1,0}^*, \dots, m_{i-1,0}^*$ , respectively. Given this information, the expert is asked to revise her initial median values  $m_{j,0}^*$  of  $p_j$ , and the new assessments are denoted by  $m_{j,i}^*$  ( $j = i + 1, \dots, k$ ). Again, as explained in the case of  $i = 1$ , these assessments are also sequentially conditional assessments for  $i = 2, \dots, k - 2$ , i.e.  $m_{j,i}^*$  is assessed under the condition that  $p_1 = m_{1,0}^*, \dots, p_{i-1} = m_{i-1,0}^*, p_i = m_{i,i}^*, p_{i+1} = m_{i+1,i}^*, \dots, p_{j-1} = m_{j-1,i}^*$ , for all  $j = i + 1, \dots, k$ .

The assumed conditioning true value  $m_{i,i}^*$  should be (a) reasonably different from the initial value  $m_{i,0}^*$ , so that the expert revises the previously calculated assessments at the other categories by a measurable change, but (b) not too different from  $m_{i,0}^*$ , so that it is still a plausible value according to the expert’s belief. Moreover, as will be shown later, for mathematical coherence the arbitrary values  $\eta_i^*$ s have to be chosen such that  $\eta_i^* \neq 0$  ( $i = 1, \dots, k - 2$ ). To fulfil these constraints, the lower quartile of each category is taken as the value for  $m_{i,i}^*$ . Hence, we put  $m_{1,1}^* = L_{1,0}^*$  and  $m_{i,i}^* = L_i^*$ , so  $\eta_1^* = L_{1,0}^* - m_{1,0}^*$  and  $\eta_i^* = L_i^* - m_{i,0}^*$  ( $i = 2, \dots, k - 2$ ).

The expert uses the software to make her assessments on an interactive graph. An example is illustrated in Fig. 1, where the categories represent 6 phenotypes of snails: yellow-banded/ yellow-unbanded/ pink-banded/ pink-unbanded/ brown-banded/ brown-unbanded. The conditioning set of median values are the three right hand side (red) bars at the first three categories. The expert is asked to assess how her new median values, shown as the three right hand side (one



**Fig. 1** Software suggestions for conditional medians. The following message is shown to the expert: “The grey boxes show your original median assessments. Suppose that the proportion in the first three categories were actually the value given by the red boxes, what would be your median estimates of the proportions in the other categories? Your assessments appear as blue and orange boxes. For your assessment to be internally consistent, the values given by the boxes must add up to 1. The computer suggests values (in yellow) that are close to your assessments and that add up to one.” The grey boxes are the  $m_{i,0}^*$ , the blue bar is  $m_{1,4}^*$  and the orange bars are  $m_{1,5}^*$  and  $m_{1,6}^*$



blue and two orange) bars at the last three categories, will change based on the new conditioning set. The initially calculated medians of all categories are shown as the left hand side (grey) bars.

For mathematical coherence, the expert’s assessments must fulfill the first statement of Lemma 2 below. The second statement of Lemma 2 gives the conditional medians of  $Z_j$ , say  $m_{j,i}^\#$ , in terms of the corresponding conditional medians of  $p_j$ ,  $m_{j,i}^*$  ( $j = i + 1, \dots, k$ ). The former are used in Sect. 4.3 to estimate the correlation matrix  $\mathbf{R}$  of the Gaussian copula prior distribution.

**Lemma 2** Under the unit-sum constraint on  $\mathbf{p}$ , and the multivariate normality of  $\mathbf{y}_{K-1}$ , for  $i = 1, \dots, k - 2$ , we have

1.  $\sum_{j=1}^i m_{j,0}^* + \eta_i^* + \sum_{j=i+1}^k m_{j,i}^* = 1$ ,
2.  $m_{j,i}^\# = m_{j,i}^* / (1 - \sum_{r=1}^i m_{r,0}^* - \eta_i^* - \sum_{r=i+1}^{j-1} m_{r,i}^*)$ , for  $j = i + 1, \dots, k$ .

See the Appendix for a proof of this lemma.

By scaling their values, the software normalizes the values of  $m_{i+1,i}^*, \dots, m_{k,i}^*$  so that they satisfy Lemma 2. Specifically, the normalized value of  $m_{j,i}^*$  is  $[(1 - \sum_{r=1}^i m_{r,0}^* - \eta_i^*) / \sum_{r=i+1}^k m_{r,i}^*] m_{j,i}^*$ , for  $j = i + 1, \dots, k$ . The expert is told that the conditional medians must satisfy a mathematical equation and that a good representation of her assessments has been calculated. The software displays the normalized conditional medians as the thinner (yellow) bars in Fig. 1. The expert has the option of changing these medians until she feels that the suggested normalized set  $m_{i+1,i}^*, \dots, m_{k,i}^*$

gives an acceptable representation of her opinion. In Fig. 1, the expert’s conditional assessments are the three right-most (blue and orange) bars for the last three categories. These are reasonably coherent, so the suggested values in the thinner (yellow) bars are almost equal to these assessments.

The current assessment task stops at step  $k - 2$ , as we do not ask for any conditional assessments for the last remaining category  $p_k$ . Since the condition of summing to one should always be fulfilled, conditioning on specific values of all  $p_1, \dots, p_{k-1}$  gives a fixed value for  $p_k$ .

## 4 Encoding the hyperparameters

### 4.1 Encoding the hyperparameters of the beta marginal distributions

To encode a standard Dirichlet prior distribution, the expert has already assessed a lower quartile,  $\bar{L}_i$ , a median,  $\bar{m}_i$ , and an upper quartile,  $\bar{U}_i$ , for  $p_i$  ( $i = 1, \dots, k$ ), as discussed in Sect. 3.1.1. But, to encode a Gaussian copula prior distribution, the expert has already assessed a lower quartile,  $L_{i,0}^\#$ , a median,  $m_{i,0}^\#$ , and an upper quartile,  $U_{i,0}^\#$ , for  $Z_i$  ( $i = 1, \dots, k - 1$ ), as discussed in Sect. 3.2.1. The quartile assessments  $\bar{L}_i$ ,  $\bar{m}_i$  and  $\bar{U}_i$  are used here to encode the hyperparameters  $\alpha_i$  and  $\beta_i$  of the marginal beta distribution of each  $p_i$  ( $i = 1, \dots, k$ ). Similarly,  $L_{i,0}^\#, m_{i,0}^\#$  and  $U_{i,0}^\#$  are used to encode the hyperparameters  $\alpha_i^\#$  and  $\beta_i^\#$  of the marginal beta distribution of each  $Z_i$  ( $i = 1, \dots, k - 1$ ). The same method of encoding beta hyperparameters is used in

both cases, so we denote a general set of quartile assessments by  $\tilde{L}_{i,0}$ ,  $\tilde{m}_{i,0}$  and  $\tilde{U}_{i,0}$  and show how we use them to encode the beta hyperparameters denoted by  $\tilde{\alpha}_i$  and  $\tilde{\beta}_i$ .

The method proposed in [Elfadaly and Garthwaite \(2013\)](#) is used to encode the two parameters  $\tilde{\alpha}_i$  and  $\tilde{\beta}_i$  of each marginal beta distribution. In their approach, the three assessed quartiles  $\tilde{L}_{i,0}$ ,  $\tilde{m}_{i,0}$  and  $\tilde{U}_{i,0}$  are substituted in a normal approximation of the beta distribution to obtain initial values of the beta parameters. These are then used as the initial values in a numerical least-squares optimization which encodes the beta parameters as non-negative values  $\tilde{\alpha}_i$  and  $\tilde{\beta}_i$  that minimize

$$Q = [G(\tilde{L}_{i,0}; \tilde{\alpha}_i, \tilde{\beta}_i) - 0.25]^2 + [G(\tilde{m}_{i,0}; \tilde{\alpha}_i, \tilde{\beta}_i) - 0.5]^2 + [G(\tilde{U}_{i,0}; \tilde{\alpha}_i, \tilde{\beta}_i) - 0.75]^2, \tag{13}$$

where  $G(q; \tilde{\alpha}_i, \tilde{\beta}_i)$  is the CDF of a beta distribution with parameters  $\tilde{\alpha}_i$  and  $\tilde{\beta}_i$  at the point  $q$ . The least-squares optimization enables fast and accurate computation. See, for example, [Albert et al. \(2012\)](#) and the references therein.

If a Dirichlet prior is to be encoded, the parameters  $\alpha_i = \tilde{\alpha}_i$  and  $\beta_i = \tilde{\beta}_i$  will be directly reconciled to obtain the Dirichlet hyperparameters in Sect. 4.2. But for encoding a Gaussian copula prior, we take  $\alpha_i^\# = \tilde{\alpha}_i$  and  $\beta_i^\# = \tilde{\beta}_i$ ; these encoded beta marginal distributions of  $Z_1, \dots, Z_{k-1}$  have no constraint to satisfy—under the transformations in (9) and (10), the unit-sum constraint on  $\mathbf{p}$  is always fulfilled for any encoded beta marginal distributions of  $Z_1, \dots, Z_{k-1}$ .

### 4.2 Encoding the Dirichlet hyperparameters

After beta marginal distributions have been assessed for each category’s probability  $p_i$  (c.f. Sect. 4.1), the beta parameters are then adjusted to estimate the Dirichlet hyperparameter vector as follows. First, note that the system of equations in (6) does not guarantee a consistent solution for  $\mathbf{a}$ . This is because the expert will seldom be fully consistent and give marginal assessments that are exactly mathematically coherent. From (6), each marginal step of the elicitation process provide two estimates for  $a_i$  and  $N_i$ , namely, for  $i = 1, \dots, k$ , we get  $a_i = \alpha_i$  and  $N_i = \alpha_i + \beta_i = \sum_{j=1}^k a_j$ . Moreover, the estimated hyperparameters must fulfill the unit-sum constraint on the expected values of the probabilities. Specifically, if the expected values are denoted by  $\mu_i$  (where  $\mu_i = a_i/N_i$ , for  $i = 1, \dots, k$ ) they must satisfy  $\sum_{i=1}^k \mu_i = 1$ . If the assessments are not coherent and do not satisfy these constraints, which is usually the case, initial assessments will need to be reconciled. Even if the expert were very consistent and gave reasonably coherent assessments, reconciliation would be needed for correcting small rounding errors.

[Lindley et al. \(1979\)](#) investigated the reconciliation of assessments that are inconsistent with the laws of probabilities (incoherent). They developed least-squares procedures as reconciliation tools that may be used for any expert’s incoherent assessments. In the spirit of their approach, we propose the following three options to reconcile incoherent estimates of the  $\mu_i$  and  $N$  into mathematically coherent estimates  $\mu_i^*$ ,  $N^*$ , respectively.

- (I) Normalize  $\mu_i$ ’s into  $\mu_i^*$ ’s that add up to one, i.e. put  $\mu_i^* = \mu_i / \sum_{j=1}^k \mu_j$ , for  $i = 1, \dots, k$ .
- (II) Minimize the sum of squares of differences between  $\mu_i^*$  and  $\mu_i$ , subject to the constraints  $0 < \mu_i^* < 1$  and  $\sum_{i=1}^k \mu_i^* = 1$ . So, the  $\mu_i^*$  are taken as the values that minimize

$$S_1 = \sum_{i=1}^k (\mu_i^* - \mu_i)^2 + \lambda \left( \sum_{i=1}^k \mu_i^* - 1 \right),$$

where  $\lambda$  is a Lagrange multiplier.

- (III) The precision of each  $p_i$ , i.e. the inverse of its variance, can be used as a weight to reflect the expert’s confidence about each of her assessments ([Lindley et al. 1979](#)). These weights are used in a weighted least-squares procedure. Hence we determine the  $\mu_i^*$ s that minimize

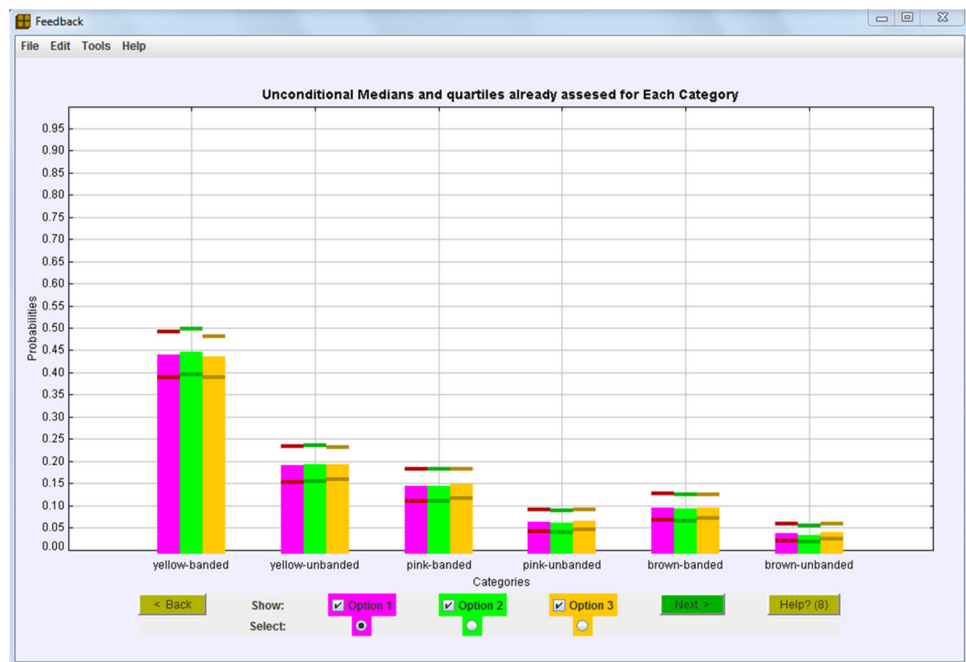
$$S_2 = \sum_{i=1}^k w_i (\mu_i^* - \mu_i)^2 + \lambda \left( \sum_{i=1}^k \mu_i^* - 1 \right),$$

where  $w_i = [\alpha_i \beta_i / (\alpha_i + \beta_i + 1)(\alpha_i + \beta_i)^2]^{-1}$ , for  $i = 1, \dots, k$ .

With (I) and (II),  $N^*$  is set equal to the average of the  $N_i$ ’s, while with (III) it is set equal to their weighted average as  $N^* = \sum_{i=1}^k w_i N_i / \sum_{j=1}^k w_j$ . Estimating  $\mu_i^*$  and  $N^*$ , using any of the options listed above, makes it easy to estimate  $a_i$  by  $a_i^*$ , where  $a_i^* = \mu_i^* N^*$  for  $i = 1, \dots, k$ . For options (II) and (III), the software computationally estimates  $\lambda$  as zero up to a tolerance of  $10^{-4}$ .

The software elicits three hyperparameter vectors of the Dirichlet distribution, one vector for each of the above options. Each vector is then used to compute the corresponding pairs of marginal beta parameters as given in (6). Three quartiles of each beta marginal distribution are computed numerically for each Dirichlet hyperparameter vector. The three sets of quartiles are then displayed to the expert and she is asked to select the set of quartiles that best represents her opinion. See Fig. 2: the middle three button groups at the bottom of the screen offer the options that the expert can choose for her quartile assessments. (The other buttons are for navigation to the previous screen or the next screen.) The vector with the selected set of quartiles will be taken

**Fig. 2** A feedback screen for Dirichlet elicitation. The following message is shown to the expert: The probabilities of the different categories must add to one. The quartile assessments for the different categories must also meet certain requirements in order to be internally consistent. This program gives three options for reconciling your assessments to meet these requirements. Select the option that best matches your opinion and then click ‘Next’



as the final elicited hyperparameter vector of the Dirichlet prior.

The expert is still able, however, to modify any or all of the selected set of quartiles, in which case beta parameters are computed again and the final Dirichlet hyperparameter vector is computed through averaging, as in option (I) above.

### 4.3 Encoding the Gaussian copula hyperparameters

The normalizing one-to-one functions in (12) are used to transform the assessed conditional quartiles of  $\mathbf{z}_{k-1}$  into conditional quartiles of  $\mathbf{y}_{k-1}$ , yielding conditional expectations, variances and covariances of the multivariate normal variables. To form a positive-definite correlation matrix  $\mathbf{R}$ , let  $\mathbf{y}_i = (Y_1, \dots, Y_i)$  and  $\mathbf{R}_i = \text{Var}(\mathbf{y}_i)$ , for  $i = 1, \dots, k-1$ , where  $\mathbf{R}_1 = \text{Var}(Y_1) = 1$  and the final matrix  $\mathbf{R} = \mathbf{R}_{k-1}$ . Supposing that  $\mathbf{R}_{i-1}$  has been estimated as a positive-definite matrix, an encoding method, based on that of Kadane et al. (1980), is used to form a positive-definite correlation matrix  $\mathbf{R}_i$  ( $i = 2, \dots, k-1$ ). The mathematical details of the method are shown in the Appendix. Since  $\mathbf{R}_1 = 1 > 0$ , by mathematical induction the full correlation matrix  $\mathbf{R} = \mathbf{R}_{k-1}$  is guaranteed to be positive-definite.

We have to note that, according to this method of elicitation, the variances on the main diagonal of  $\mathbf{R}$ , say  $r_{i,i}$ ,  $i = 1, \dots, k-1$ , cannot be guaranteed to each equal one, except for the first element  $r_{1,1}$ . It is easy, however, to transform  $\mathbf{R}$  into  $\mathbf{R}^* = \mathbf{A}\mathbf{R}\mathbf{A}$ , where  $\mathbf{A}$  is a diagonal matrix with  $a_{1,1} = 1$  and  $a_{i,i} = 1/\sqrt{r_{i,i}}$ , for  $i = 2, \dots, k-1$ . Then  $\mathbf{R}^*$  can be used as the correlation matrix in the Gaussian copula function

and satisfies both the unit variances and positive-definiteness requirements. The unit variances in the correlation matrix  $\mathbf{R}^*$  guarantee that each marginal distribution  $G_i(Z_i)$  is still a beta distribution with the same hyperparameters  $\alpha_i^\#$  and  $\beta_i^\#$  that were encoded, for  $i = 1, \dots, k-1$ , in Sect. 4.1.

### 4.4 Evaluating the encoded priors

A good elicitation method encodes a prior distribution that captures the expert’s knowledge and beliefs faithfully and as closely as possible. Thus, a reasonable way to compare the performance of the two encoded prior distributions is to check which of them produces values that are closer to the expert’s original assessments. The aim of the Gaussian copula prior is to capture the correlation structure better than the Dirichlet prior, so the conditional medians computed from each encoded prior distribution can be compared to the expert’s original assessments of the conditional medians to check whether this aim has been achieved. We implement this approach at the end of the evolution example in Sect. 5 to evaluate the encoded priors.

A different quality of probability assessments is *calibration*, which measures the agreement between an expert’s assessed probabilities and reality. Events that are given a prior probability of occurring of, say, 0.3 should ideally occur about 30 % of the time. Similarly, a set of subjective 50 % prediction intervals should contain the target event half the time. A body of psychological research has examined the calibration of probability assessments in varied contexts; see (O’Hagan et al. 2006, chapter 4) for a review. How-

ever, as noted by (O’Hagan et al. 2006, p. 161), “comparing the elicited probabilities or distributions with the true values of the events or variables in question does not evaluate the success of elicitation; it does not compare the elicited probabilities with the expert’s underlying opinions.”

## 5 Example: Evolution of the snails in the UK

Environmental scientists are interested in the effect of climate change on the evolutionary responses of different species. The brown-lipped banded snail (*Cepaea nemoralis*) is known to have a short generation time and tends to sensitively adapt its shell colour and banding pattern according to its thermal environment. Silvertown et al. (2011) tested the evolutionary changes in this type of snail by comparing a recent dataset with historical datasets from the last century. Data on the frequency of each phenotype of this species were collected by volunteers in 15 European countries through the citizen science project Evolution MegaLab (Worthington et al. 2012). Thousands of public volunteers contributed data to the project by recording the shell colour and banding pattern of the snails they had observed in each country.

An environmental expert who has worked with these data is interested in the modelling of ecosystems under future climate change. In this example, the expert (who is male) used the elicitation software to quantify his opinion about the expected proportions of each phenotype of *Cepaea nemoralis* snails in the UK in 2050 under the assumption that only a little is done to reduce climate change (a scenario called Representative Climate Forcing 6, under which greenhouse gas emissions continue to rise without strong mitigation). He advised that three main phenotypes can be considered: yellow, pink or brown; each of them can be banded or unbanded. Hence, any single snail can be categorized in one of six different phenotypes, so the problem can be formulated as a multinomial model with six categories. The probabilities  $p_1, \dots, p_6$  are the proportions of each phenotype that might be observed in the UK. Our method and software were used to quantify the expert’s opinion about his uncertainty regarding the parameters of this multinomial model, representing his opinion by both a Dirichlet prior distribution and a Gaussian copula prior.

After initializing the software and defining the model, the expert assessed his medians,  $\bar{m}_i$  ( $i = 1, \dots, 6$ ), of the proportion of each of the following 6 phenotypes of *Cepaea nemoralis* snails: yellow-banded/ yellow-unbanded/ pink-banded/ pink-unbanded/ brown-banded/ brown-unbanded. Then the expert assessed lower and upper quartiles ( $\bar{L}_i$  and  $\bar{U}_i$ ) for the proportion of each phenotype. His assessed medians and quartiles are given in Table 1.

These assessments were used to encode a marginal beta prior distribution, with hyperparameters  $\alpha_i$  and  $\beta_i$ , for the

**Table 1** Expert’s assessments of medians and quartiles

Phenotype		Lower quartile ( $\bar{L}_i$ )	Median ( $\bar{m}_i$ )	Upper quartile ( $\bar{U}_i$ )
Yellow-banded	$p_1$	0.351	0.449	0.551
Yellow-unbanded	$p_2$	0.154	0.198	0.240
Pink-banded	$p_3$	0.127	0.146	0.182
Pink-unbanded	$p_4$	0.048	0.069	0.092
Brown-banded	$p_5$	0.056	0.092	0.140
Brown-unbanded	$p_6$	0.025	0.042	0.058

**Table 2** The encoded hyperparameters of the Dirichlet prior distribution

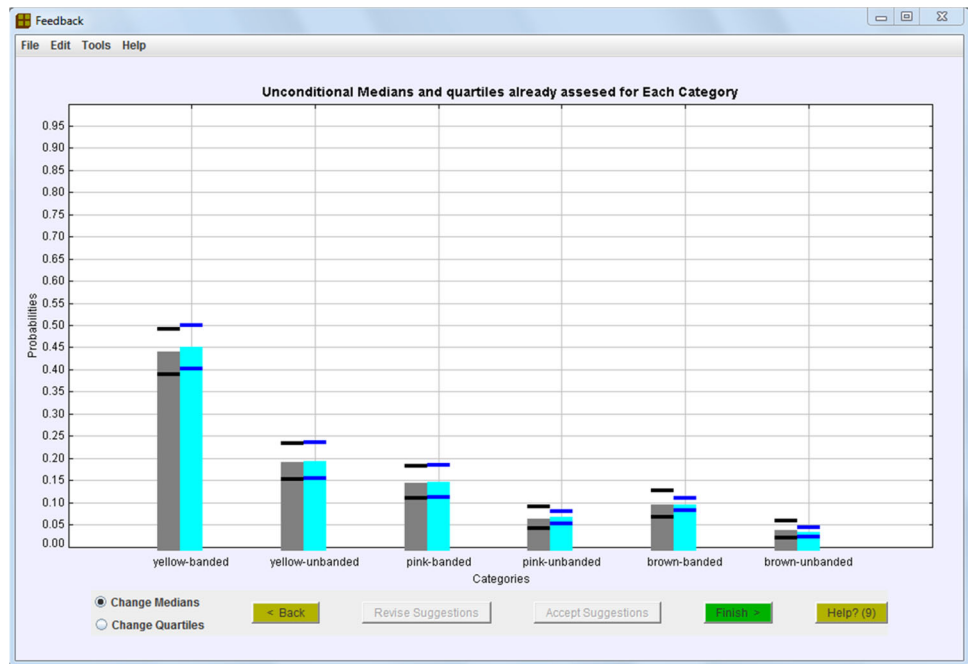
$a_1^*$	$a_2^*$	$a_3^*$	$a_4^*$	$a_5^*$	$a_6^*$	$N^*$
18.33	8.12	6.27	2.96	4.12	1.81	41.60

proportion of phenotypes in each category using equation (13) in Sect. 4.1. These can then be used to determine the hyperparameters of a Dirichlet distribution using any of the three least-squares options detailed in Sect. 4.2. The expert chose option (II) where the values of the Dirichlet hyperparameters were encoded as given in Table 2.

The median values and quartiles of the coherent marginal beta distributions of the Dirichlet variates were computed and presented to the expert as feedback in Fig. 3. These are shown as the left hand side (grey) bars for each category. During this feedback stage he was invited to accept or revise these quantities by clicking on the right hand side (blue) bars of each category in Fig. 3. The initial median values,  $\bar{m}_i$  ( $i = 1, \dots, 6$ ), as assessed by the expert have a sum that is nearly equal to one, so the coherent medians suggested by the software in Fig. 3 were close to his median assessments as shown in Table 1. However, the expert felt that some of the suggested quartiles were too far apart to be a reasonable representation of his opinion, and he chose to revise the suggested quartile values for the last three categories. These were then used to encode another Dirichlet distribution with hyperparameters given in Table 3 and this is taken as the expert’s Dirichlet prior distribution. Table 4 shows the final coherent set of medians and quartiles after the expert’s revisions computed from the Dirichlet hyperparameters in Table 3.

The expert’s initial assessments in Table 1 give hyperparameter values  $\alpha_i$  and  $\beta_i$  of the marginal beta distributions that result in different estimates of the Dirichlet hyperparameter  $N$ . Specifically, they give  $N_i = \alpha_i + \beta_i$  ( $i = 1, \dots, 6$ ) that range between 11.75 and 73.27. The variability of these values can be used as a diagnostic check to determine whether the Dirichlet prior distribution might be a reasonable representation of the expert’s knowledge. As concluded

**Fig. 3** The coherent marginal quartiles for the Dirichlet distribution given as the *left (grey) boxes*, with the expert’s revised values, given as the *right (blue) boxes*. The following message is shown to the expert: This screen shows the medians and quartiles that the software has determined for your prior distribution. You should change them if they do not form a reasonable representation of your opinion. The *grey (left) boxes* are there to show the coherent quartiles



**Table 3** The encoded hyperparameters of the revised Dirichlet prior distribution

$a_1^*$	$a_2^*$	$a_3^*$	$a_4^*$	$a_5^*$	$a_6^*$	$N^*$
51.46	22.63	17.06	7.70	10.89	4.29	114.03

**Table 4** Medians and quartiles of the revised marginal beta distributions

Phenotype		Lower quartile	Median	Upper quartile
Yellow-banded	$p_1$	0.401	0.451	0.501
Yellow-unbanded	$p_2$	0.156	0.194	0.236
Pink-banded	$p_3$	0.111	0.144	0.182
Pink-unbanded	$p_4$	0.050	0.065	0.082
Brown-banded	$p_5$	0.083	0.095	0.107
Brown-unbanded	$p_6$	0.027	0.036	0.046

in [Elfadaly and Garthwaite \(2013\)](#), a substantial variation in the values of  $N_i$  ( $i = 1, \dots, k$ ) indicates that a Dirichlet prior does not capture the expert’s opinions well. This suggests that here the Dirichlet prior may not be an adequate representation of the expert’s opinion, in which case the Gaussian copula prior is to be preferred.

The expert gave further assessments to encode a Gaussian copula prior as well, which gains much greater flexibility in the correlation structure of the parameters of his multinomial model. To encode a Gaussian copula prior, the expert was asked to assess a median value for the proportion of the yellow-banded phenotype of snails in the UK in 2050. He assessed this as  $m_{1,0}^\# = 0.446$ , the median of  $Z_1 = p_1$ .

As mentioned before in Sect. 3.2.1,  $m_{1,0}^\# \simeq \bar{m}_1 = 0.449$ . To improve the internal consistency between this assessment and the forthcoming ones, the expert was also asked at this stage to assess the median values of the proportions of all other phenotypes. These were assessed as 0.198, 0.147, 0.076, 0.091, 0.041. Clearly, these assessments are quite consistent with the experts’s assessments  $\bar{m}_i$  in Table 1.

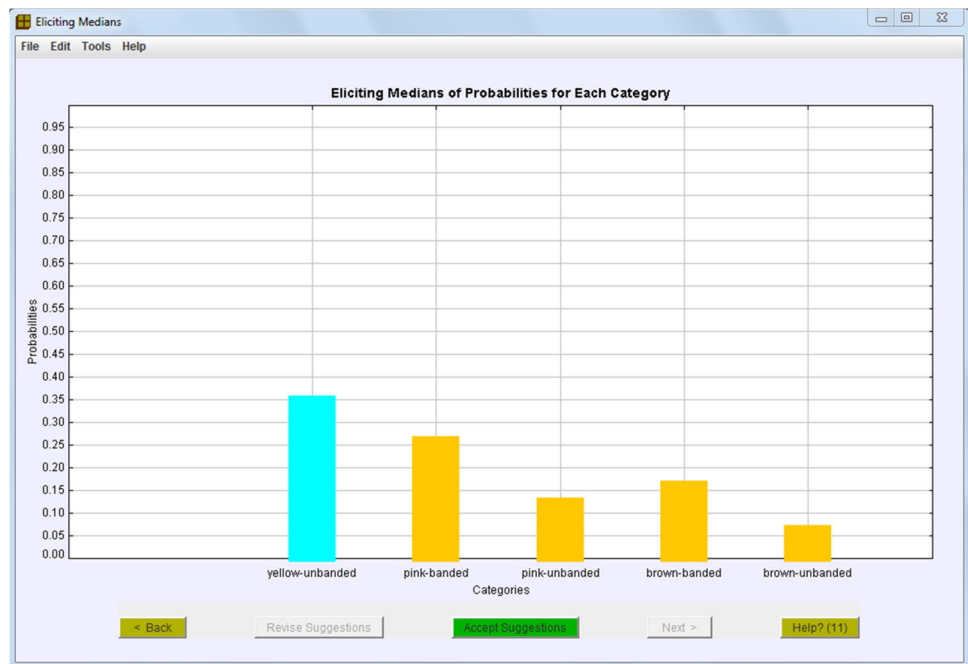
The expert was then asked to assume that a randomly selected snail is not yellow-banded. For this snail he was asked to assess the median probability,  $m_{2,0}^\#$ , that it belongs to the yellow-unbanded category. To help the expert, the software uses the median values assessed at the previous stage to suggest a value for the median of  $Z_2 = p_2/(1 - p_1)$  as  $m_{2,0}^\# = 0.198/(1 - 0.446) = 0.358$ . Suggested median values for the remaining categories were also computed and presented to the expert (See Fig. 4). He accepted the value of  $m_{2,0}^\# = 0.358$  and all the other suggested median values as a reasonable representation of his opinion.

Then at each step  $i$  ( $i = 3, 4, 5$ ), assuming that a selected snail does not belong to the first  $i - 1$  phenotypes, the expert accepted the suggested value of  $m_{i,0}^\#$  as the median of  $Z_i = p_i/(1 - \sum_{j=1}^{i-1} m_{j,0}^\#)$ , together with the remaining median values at each step.

Similarly, the expert was asked to assess a lower and an upper quartile ( $L_{i,0}^\#$  and  $U_{i,0}^\#$ ) for the probability  $Z_1 = p_1$  and for  $Z_i$  ( $i = 2, \dots, 5$ ) as the probability of belonging to the  $i$ th phenotype given that the snail could not be categorized in the first  $i - 1$  categories. The full set of median and quartile assessments obtained at this stage are shown in Table 5. These assessments were used in the optimization equation (13) in Sect. 4.1 to encode the hyperparameters  $\alpha_i^\#$  and  $\beta_i^\#$  of the



**Fig. 4** Unconditional median assessment for the second category ( $m_{2,0}^\# = 0.358$ ). The following message is shown to the expert: Assessments are only required for the categories that are labelled and have blue or orange boxes. Assume that you know an item falls in one of these categories. Given that information, what is the probability that the item will be in the second category? the computer’s suggestion for this category,  $m_{2,0}^\# = 0.358$ , is given by the left-most (blue) box. The computer’s suggestions for other categories are given by the other (orange) boxes



**Table 5** Expert’s assessments for the medians and quartiles of the  $Z_i$  variables

Phenotype		Lower quartile ( $L_{i,0}^\#$ )	Median ( $m_{i,0}^\#$ )	Upper quartile ( $U_{i,0}^\#$ )
Yellow-banded	$Z_1$	0.349	0.446	0.545
Yellow-unbanded	$Z_2$	0.253	0.358	0.449
Pink-banded	$Z_3$	0.330	0.413	0.495
Pink-unbanded	$Z_4$	0.296	0.366	0.438
Brown-banded	$Z_5$	0.557	0.688	0.802

**Table 6** The encoded hyperparameters of marginal beta distributions

	$Z_1$	$Z_2$	$Z_3$	$Z_4$	$Z_5$
$\alpha^\#$	5.353	3.906	6.819	7.838	4.531
$\beta^\#$	6.577	6.934	9.607	13.344	2.227

beta distribution of each  $Z_i$  ( $i = 1, \dots, 5$ ). The values of these hyperparameters are given in Table 6.

To encode a correlation matrix for the Gaussian copula prior, the expert gave conditional assessments that quantified his opinion about the dependence structure between the marginal beta distributions. Specifically, he assessed conditional quartile values,  $L_i^*$  and  $U_i^*$ , for each  $i$ th ( $i = 2, \dots, 5$ ) category, under the condition that the assessed medians for the previous  $j$  ( $j = 1, \dots, i - 1$ ) categories were actually their true values. The expert’s five pairs of assessments for the conditional lower and upper quartiles are given in Table 7. The quartiles for the last category in Table 7 were automatically computed by the software when the expert assessed

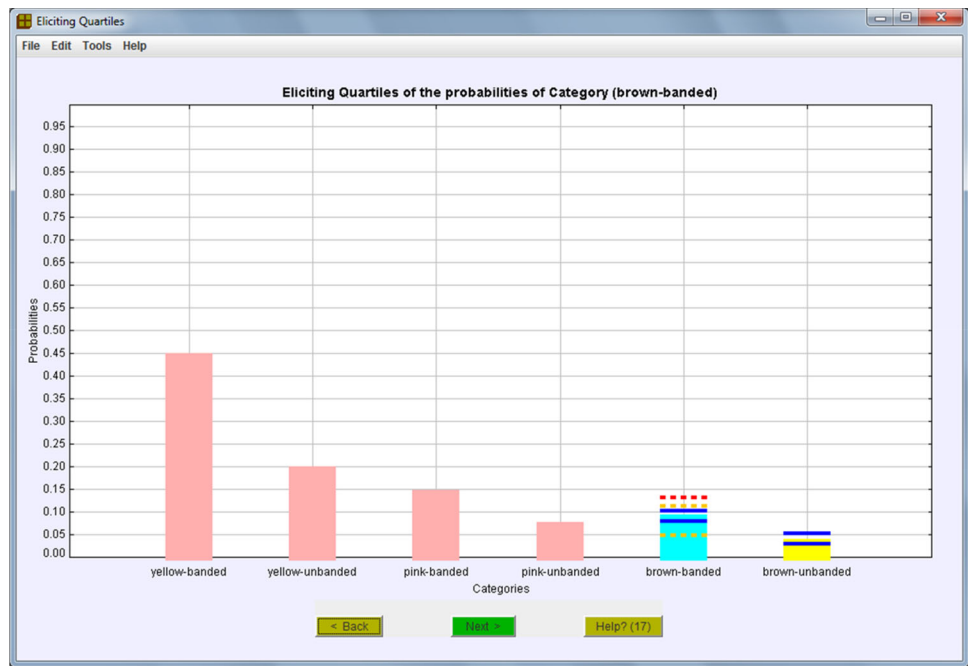
**Table 7** Expert’s assessments of conditional quartiles

	$p_2$	$p_3$	$p_4$	$p_5$	$p_6$
Conditional lower quartile ( $L_i^*$ )	0.163	0.132	0.067	0.081	0.032
Conditional upper quartile ( $U_i^*$ )	0.230	0.163	0.086	0.100	0.051

two quartiles for the fifth category. This is also illustrated in Fig. 5.

Next, conditional on the proportion for the first category being  $m_{1,1}^* = 0.349$  (his assessment of its lower quartile,  $L_{1,0}^*$ ), the expert gave conditional median assessments,  $m_{j,1}^*$  ( $j = 2, \dots, 6$ ), for the proportions of the five remaining categories. The number of conditions was then increased in stages. Table 8 gives all the conditional median assessments,  $m_{j,i}^*$  ( $i = 1, \dots, 4; j = i + 1, \dots, 6$ ), where the underlined values constitute the conditioning set at each stage. The assessments reported at each row of Table 8 were sequentially obtained under the assumption that the previously assessed

**Fig. 5** Assessing conditional quartiles for the last two categories ( $L_5^* = 0.081$  and  $U_5^* = 0.100$ ). The following message is shown to the expert: Give your lower and upper quartile assessments on category five, assuming that your median assessments for the first four categories with (red) boxes are correct. The implied quartiles for the last category—the right-most (yellow) box—are also shown. To be logical, your assessment must be below the upper-most short horizontal (red) dotted line and should be between the other two short (orange) horizontal dotted lines



**Table 8** Expert’s assessments of conditional medians. The conditional median assessment in the  $i$ th row and  $j$ th column is  $m_{j,i}^*$  ( $i = 1, \dots, 4; j = i + 1, \dots, 6$ )

$p_1$	$p_2$	$p_3$	$p_4$	$p_5$	$p_6$
<u>0.349</u>	0.279	0.162	0.084	0.088	0.039
<u>0.446</u>	<u>0.163</u>	0.167	0.090	0.092	0.043
<u>0.446</u>	<u>0.198</u>	<u>0.132</u>	0.093	0.089	0.041
<u>0.446</u>	<u>0.198</u>	<u>0.147</u>	<u>0.067</u>	0.094	0.048

The underlined values represent the conditioning set in each row, where the underlined values in the  $i$ th row are  $m_{j,0}^*$  ( $j = 1, \dots, i - 1$ ) and  $m_{i,i}^*$  for  $i = 1, \dots, 4$

conditional medians at the same line are the true values, not only the underlined values.

For example, Fig. 1 shows the expert’s assessments of conditional medians  $m_{1,4}^*$ ,  $m_{1,5}^*$  and  $m_{1,6}^*$  for the last three categories, given that  $p_1 = m_{1,1}^* = 0.349$ , where the conditional median assessments  $m_{1,2}^* = 0.279$  and  $m_{1,3}^* = 0.162$  are also assumed to be the true values of  $p_2$  and  $p_3$ , respectively. This is one of the sub-stages required to obtain the assessments in the first row of Table 8. However, the expert chose not to change his conditional median assessments at different sub-stages. He only revised his assessments when the underlined conditioning values were changed, i.e. from one row to the next in Table 8.

This was the last assessment task, after which the software output the encoded hyperparameters,  $\alpha_i^\#$  and  $\beta_i^\#$ , of the marginal beta prior distributions of  $Z_i$  ( $i = 1, \dots, 5$ ) as in Table 6. These were computed as detailed in Sect. 4.1 using Eq. (13). The dependence structure between these beta

**Table 9** The encoded correlation matrix,  $\mathbf{R}^*$ , of the Gaussian copula prior

	$Y_1$	$Y_2$	$Y_3$	$Y_4$	$Y_5$
$Y_1$	1.000	−0.742	−0.416	−0.386	−0.079
$Y_2$	−0.742	1.000	0.145	0.031	0.099
$Y_3$	−0.416	0.145	1.000	−0.296	0.076
$Y_4$	−0.386	0.031	−0.296	1.000	0.145
$Y_5$	−0.079	0.099	0.076	0.145	1.000

marginals was quantified as a multivariate Gaussian copula function with the encoded correlation matrix,  $\mathbf{R}^*$ , given in Table 9. This has been encoded as detailed in Sect. 4.3.

The encoded matrix in Table 9 does not give covariances between the beta distributed variables,  $Z_1, \dots, Z_5$ . Instead, it gives the covariances between the transformed normal variates,  $Y_1, \dots, Y_5$ . The Gaussian copula multivariate distribution is parameterized by both the marginal beta parameters in Table 6 and the correlation matrix in Table 9. These determine the experts’s Gaussian copula prior distribution. The software produces a file containing the WinBUGS model (Spiegelhalter et al. 2007) that uses the Gaussian copula prior distribution. This is given at the end of the Appendix.

To examine the assessed prior distributions, we compute the conditional medians from each of the two encoded prior distributions and compare them to the expert’s assessments in Table 8. The aim is to check whether the Gaussian copula has captured the correlation structure of the multinomial probabilities better than the Dirichlet distribution. Table 10 gives the conditional medians computed from the Dirichlet distribution, with the encoded hyperparameters in Table 3, together

**Table 10** Conditional medians as assessed by the expert (denoted by A), as computed from the encoded Dirichlet distribution (denoted by D) and as computed from the encoded Gaussian copula prior distribution (denoted by C)

	$p_1$	$p_2$	$p_3$	$p_4$	$p_5$	$p_6$
A	<u>0.349</u>	0.279	0.162	0.084	0.088	0.039
D		0.234 (0.16 %)	0.177 (0.09%)	0.079 (0.06%)	0.116 (0.32%)	0.044 (0.13%)
C		0.278 (0.00 %)	0.166 (0.02 %)	0.082 (0.02 %)	0.093 (0.06 %)	0.038 (0.03 %)
A	<u>0.446</u>	<u>0.163</u>	0.167	0.09	0.092	0.043
D			0.166 (0.01 %)	0.074 (0.18 %)	0.109 (0.18 %)	0.041 (0.05 %)
C			0.168 (0.01 %)	0.087 (0.03 %)	0.094 (0.02 %)	0.043 (0.00 %)
A	<u>0.446</u>	<u>0.198</u>	<u>0.132</u>	0.093	0.089	0.041
D				0.074 (0.20 %)	0.109 (0.22 %)	0.041(0.00 %)
C				0.087 (0.06 %)	0.096 (0.08 %)	0.043 (0.05 %)
A	<u>0.446</u>	<u>0.198</u>	<u>0.147</u>	<u>0.067</u>	0.094	0.048
D					0.103 (0.10 %)	0.039 (0.19 %)
C					0.094 (0.00 %)	0.046 (0.04 %)

The *underlined values* represent the conditioning set in each row. The discrepancies between the conditional medians of each prior and the expert’s assessments are given as bracketed percentages

with the corresponding conditional medians computed from the Gaussian copula prior, with the encoded hyperparameters in Tables 6 and 9. The original conditional medians assessed by the expert are given again in Table 10 to facilitate the comparison. For both prior distributions, the discrepancies of the computed conditional medians from the expert’s assessments are also presented in the same table as percentages of the assessments.

Comparing the conditional median values with the corresponding expert’s original assessments it can be seen that the conditional medians computed from the proposed Gaussian copula prior are much closer to the expert’s assessments than are those computed from the Dirichlet distribution. This illustrates the potential advantage from using the Gaussian copula prior rather than the Dirichlet prior distribution.

The elicitation process took about an hour to complete. The expert stressed the importance of the convenient order of categories when giving conditional assessments, commenting that ordering the categories in a suitable sequence made it easier for him to think about the conditions.

## 6 Concluding comments

Our aim in this paper was to propose methods for eliciting expert opinion about the hyperparameters of both the Dirichlet distribution and the Gaussian copula distribution. So as to simplify the separate tasks that an assessor must perform, the elicitation process for a multivariate prior has been decomposed into a number of processes for eliciting univariate beta distributions. The hyperparameters of these beta distributions are estimated from marginal median and quartile assessments obtained from a field expert. Two novel methods have been proposed for reconciling the beta mar-

ginal distributions into a multivariate prior distribution. The first method reconciles the beta distributions as marginals of the well-known standard conjugate Dirichlet distribution, giving a prior distribution that is easily applied in practice.

In the second method, instead of assuming a Dirichlet prior, we use a Gaussian copula function with beta marginals to model the joint prior distribution of multinomial probabilities. The proposed Gaussian copula prior assumes that the dependence structure between the transformed probabilities can be represented by a multivariate normal distribution, where the marginal prior distribution of each of them is still expressed as a beta distribution. This requires further conditional quartile assessments to describe the correlation structure between these probabilities. The monotonicity of the Gaussian copula transformation allows conditional quartiles of the multinomial probabilities to be transformed into normal quartiles. The latter are used to obtain product-moment correlations for normal variates. This powerful technique of transforming quartiles avoids the difficulties encountered when transforming product-moment correlations. Structural assessments of the conditional quartiles are used to ensure that the encoded correlation matrix is positive-definite. Although the encoding of a Gaussian copula requires more assessments than the Dirichlet distribution, it gives a more flexible multivariate prior distribution. A practical example where the methods are used was given and the two encoded prior distributions compared.

The two methods have been implemented in a single user-friendly software package that is freely available at <http://statistics.open.ac.uk/elicitation>. The user has the option of eliciting either of the two priors or both of them simultaneously. For the Dirichlet prior distribution, the available software outputs the encoded hyperparameter vector  $\mathbf{a}^*$ . The expectations and variances of each  $p_i$  are also given. For the

Gaussian Copula prior, the output includes the encoded pairs of beta parameters  $\alpha_i^\#$  and  $\beta_i^\#$  ( $i = 1, \dots, k - 1$ ), together with the encoded correlation matrix  $\mathbf{R}^*$ .

**Acknowledgements** The work reported here was supported by a studentship from The Open University. We are very grateful to Dr. Yoseph Araya (a plant ecologist and a lecturer in Environmental Geography at Birkbeck College, University of London, UK) whose opinion was quantified in the example of the phenotypes of *Cepaea nemoralis* snails.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

### Appendix

#### Proof of Lemma 1

By definition,  $m_{1,0}^* = m_{1,0}^\#$ . From the normal distribution of  $Y_{k-1}$ , for  $i = 2, \dots, k - 1$ , we have

$$E(Y_i) = E(Y_i|Y_1 = E(Y_1), \dots, Y_{i-1} = E(Y_{i-1})).$$

Replacing the expected value by the median function  $M(\cdot)$ , using equation (12), we get

$$M\left\{\Phi^{-1}[G_i(Z_i)]\right\} = M\left\{\Phi^{-1}[G_i(Z_i)]|Z_1 = m_{1,0}^\#, \dots, Z_{i-1} = m_{i-1,0}^\#\right\}. \tag{14}$$

The monotonicity of the CDFs  $\Phi(\cdot)$  and  $G_i(\cdot)$  in (14), using (9) and (10), gives

$$m_{i,0}^\# = \frac{M\left(p_i|p_1 = m_{1,0}^\#, \dots, p_{i-1} = m_{i-1,0}^\# \prod_{j=1}^{i-2} (1 - m_{j,0}^\#)\right)}{\prod_{j=1}^{i-1} (1 - m_{j,0}^\#)}. \tag{15}$$

Then, the statement of Lemma 1 follows from (15) by induction, for any  $i = 2, \dots, k - 1$ . Moreover, under the unit-sum constraint,  $m_{k,0}^* = 1 - \sum_{i=1}^{k-1} m_{i,0}^* = \prod_{j=1}^{k-1} (1 - m_{j,0}^\#)$ . Since, by definition,  $m_{k,0}^\# = 1$ , Lemma 1 is also valid for  $i = k$ .

#### Some definitions and notation

For  $i = 1, \dots, k - 2$ , and  $j = i + 1, \dots, k - 1$ , let  $m_{j,i}^\# = (m_{i,0}^* + \eta_i^*) / (1 - \sum_{j=1}^{i-1} m_{j,0}^*)$ , and put  $m_{j,i} = E(Y_j|Z_1 = m_{1,0}^\#, \dots, Z_{i-1} = m_{i-1,0}^\#, Z_i = m_{i,i}^\#)$ , i.e.  $m_{j,i} = E(Y_j|p_1 = m_{1,0}^*, \dots, p_{i-1} = m_{i-1,0}^*, p_i = m_{i,0}^* +$

$\eta_i^*)$ . Then

$$m_{j,i} = \Phi^{-1}[G_j(m_{j,i}^\#)], \tag{16}$$

where  $m_{j,i}^\#$  is defined as the median of  $(Z_j|Z_1 = m_{1,0}^\#, \dots, Z_{i-1} = m_{i-1,0}^\#, Z_i = m_{i,i}^\#)$ . Similarly, for  $i = 1, \dots, k - 1$ , let  $m_{i,0} = \Phi^{-1}[G_i(m_{i,0}^\#)]$ . Since  $m_{i,0}^\#$  is the median of the CDF  $G_i(\cdot)$ ,  $m_{i,0} = 0$  ( $i = 1, \dots, k - 1$ ). For  $i = 1, \dots, k - 2$ , define  $\eta_i$  by letting  $Y_i = m_{i,0} + \eta_i = \eta_i$  when  $Z_i = m_{i,i}^\#$ , i.e. when  $p_i = m_{i,0}^* + \eta_i^*$ . Then

$$m_{j,i} = E(Y_j|Y_1 = 0, \dots, Y_{i-1} = 0, Y_i = \eta_i), \tag{17}$$

for  $j = i + 1, \dots, k - 1$ ,

and

$$\eta_i = \Phi^{-1}\left[G_i\left(\frac{m_{i,0}^* + \eta_i^*}{1 - \sum_{j=1}^{i-1} m_{j,0}^*}\right)\right] = \Phi^{-1}[G_i(m_{i,0}^\# + \eta_i^\#)], \tag{18}$$

where  $\eta_i^\# = \eta_i^* / (1 - \sum_{j=1}^{i-1} m_{j,0}^*)$ .

Moreover, define the conditional variance  $V_{j,i} = \text{Var}(Y_j|Y_1 = 0, \dots, Y_i = 0)$ , so, using the relationship between the variance of the normal distribution and its quartiles, we get, for  $j = 2, \dots, k - 1$ ,

$$V_{j,j-1} = [(U_j - L_j) / 1.349]^2, \tag{19}$$

where

$$U_j = \Phi^{-1}[G_j(U_j^\#)], \quad L_j = \Phi^{-1}[G_j(L_j^\#)].$$

#### Proof of Lemma 2

Since  $m_{k,i}^*$  is the conditional median of  $p_k$  given that  $p_1 = m_{1,0}^*, \dots, p_{i-1} = m_{i-1,0}^*, p_i = m_{i,0}^* + \eta_i^*, p_{i+1} = m_{i+1,0}^*, \dots, p_{k-1} = m_{k-1,i}^*$ , then, under the unit-sum constraint on  $\mathbf{p}$ , we have

$$m_{k,i}^* = 1 - \sum_{j=1}^i m_{j,0}^* - \eta_i^* - \sum_{j=i+1}^{k-1} m_{j,i}^*. \tag{20}$$

This proves the first statement of Lemma 2.

Using the conditional expectation of the multivariate normal distribution, for  $i = 1, \dots, k - 2$  and  $j = i + 1, \dots, k - 1$ ,

$$E[Y_j|Y_1 = 0, \dots, Y_{i-1} = 0, Y_i = \eta_i] = E[Y_j|Y_1 = 0, \dots, Y_{i-1} = 0, Y_i = \eta_i, Y_{i+1} = E(Y_{i+1}|Y_1 = 0, \dots, Y_{i-1} = 0, Y_i = \eta_i), \dots, Y_{j-1} = E(Y_{j-1}|Y_1 = 0, \dots, Y_{i-1} = 0, Y_i = \eta_i)],$$

then, from Eq. (17), replacing the expected values by the median function,  $M(\cdot)$ , we get

$$M(Y_j|Y_1 = 0, \dots, Y_{i-1} = 0, Y_i = \eta_i) \\ = M(Y_j|Y_1 = 0, \dots, Y_{i-1} = 0, Y_i = \eta_i, \\ Y_{i+1} = m_{i+1,i}, \dots, Y_{j-1} = m_{j-1,i}),$$

Hence

$$M \left\{ \Phi^{-1}[G_j(Z_j)]|Z_1 = m_{1,0}^\#, \dots, \right. \\ \left. Z_{i-1} = m_{i-1,0}^\#, Z_i = m_{i,i}^\# \right\} \\ = M \left\{ \Phi^{-1} \left[ G_j \left( \frac{p_j}{1 - \sum_{r=1}^{j-2} p_r} \right) \right] | p_1 = m_{1,0}^*, \dots, \right. \\ \left. p_{i-1} = m_{i-1,0}^*, p_i = m_{i,0}^* + \eta_i^*, \right. \\ \left. p_{i+1} = m_{i+1,i}^*, \dots, p_{j-1} = m_{j-1,i}^* \right\},$$

which, utilizing monotonicity, gives

$$m_{j,i}^\# = \frac{m_{j,i}^*}{1 - \sum_{r=1}^i m_{r,0}^* - \eta_i^* - \sum_{r=i+1}^{j-1} m_{r,i}^*}, \\ \text{for } j = i + 1, \dots, k - 1. \tag{21}$$

From (20), statement (21) is also valid for  $j = k$ . This coincides with the definition of  $Z_k = 1$ , and completes the proof of the second statement of Lemma 2.

**Encoding the correlation matrix  $\mathbf{R}_i$  ( $i = 2, \dots, k - 1$ )**

Let  $\mathbf{R}_i$  be partitioned as follows

$$\mathbf{R}_i = \begin{bmatrix} \mathbf{R}_{i-1} & \mathbf{R}_{i-1}\mathbf{r}_i \\ \mathbf{r}_i'\mathbf{R}_{i-1} & \sigma_i^2 \end{bmatrix}, \tag{22}$$

where  $\mathbf{R}_{i-1}\mathbf{r}_i = \text{Cov}(\mathbf{y}_{i-1}, Y_i)$ , and  $\sigma_i^2 = \text{Var}(Y_i)$ . Although the Gaussian copula function imposes the constraint that  $\text{Var}(Y_i) = 1$ , we will find another estimate for  $\sigma_i^2$  using the conditional variance of  $Y_i$  elicited in (19). The reason for that, as will be shown later, is to follow the approach of Kadane et al. (1980), which will ensure the positive-definiteness of the matrix  $\mathbf{R}_i$ . In what follows, we use the conditional median assessments to estimate  $\mathbf{r}_i$ .

Using the partition in (22), we have

$$E(Y_i|\mathbf{y}_{i-1}) = \mathbf{y}'_{i-1}\mathbf{r}_i, \tag{23}$$

since  $E(\mathbf{y}) = \mathbf{0}$ . Moreover, for  $j = 1, \dots, i - 1$ , taking the conditional expectation of both sides of (23), given that  $\mathbf{y}_j = \mathbf{y}_j^o$ , where  $\mathbf{y}_j^o = (m_{1,0}, \dots, m_{j-1,0}, m_{j,0} + \eta_j)$ , gives

$$E \left[ E(Y_i|\mathbf{y}_{i-1})|\mathbf{y}_j = \mathbf{y}_j^o \right] = E(\mathbf{y}'_{i-1}|\mathbf{y}_j = \mathbf{y}_j^o) \mathbf{r}_i, \tag{24}$$

i.e.

$$E(Y_i|\mathbf{y}_j = \mathbf{y}_j^o) = (\mathbf{y}_j^{o'}, E(Y_{j+1}|\mathbf{y}_j), \dots, E(Y_{i-1}|\mathbf{y}_j)) \mathbf{r}_i. \tag{25}$$

From (17) and (25), and since  $m_{j,0} = 0$  ( $j = 1, \dots, k - 1$ ), we get

$$m_{i,j} = (0, \dots, 0, \eta_j, m_{j+1,j}, \dots, m_{i-1,j}) \mathbf{r}_i.$$

This holds for  $j = 1, \dots, i - 1$ , so we have a system of  $i - 1$  equations of the form

$$\mathbf{u}_i = \mathbf{Q}_{i-1}\mathbf{r}_i, \tag{26}$$

where  $\mathbf{u}'_i = (m_{i,1}, \dots, m_{i,i-1})$  and

$$\mathbf{Q}_{i-1} = \begin{bmatrix} \eta_1 & m_{2,1} & \cdots & m_{i-1,1} \\ 0 & \eta_2 & \ddots & \vdots \\ \vdots & 0 & \ddots & m_{i-1,i-2} \\ 0 & \cdots & 0 & \eta_{i-1} \end{bmatrix}.$$

Provided that  $\eta_j \neq 0$  ( $j = 1, \dots, i - 1$ ), the upper diagonal matrix  $\mathbf{Q}_{i-1}$  is non-singular, and hence  $\mathbf{r}_i = \mathbf{Q}_{i-1}^{-1}\mathbf{u}_i$ . The existence of this solution for  $\mathbf{r}_i$  is guaranteed by choosing a non-zero value for  $\eta_j^*$  ( $j = 1, \dots, i - 1$ ) when eliciting conditional medians in Sect. 3.2.3. It can be seen from the relationship in (18) that  $\eta_j = 0$  if and only if  $\eta_j^* = 0$  ( $j = 1, \dots, i - 1$ ).

Since  $\text{Var}(Y_i|\mathbf{y}_{i-1}) = \sigma_i^2 - \mathbf{r}'_i\mathbf{R}_{i-1}\mathbf{r}_i$ , we can use the assessed conditional variance given by  $V_{i,i-1}$  in (19) to estimate the unconditional variance as  $\sigma_i^2 = V_{i,i-1} + \mathbf{r}'_i\mathbf{R}_{i-1}\mathbf{r}_i$ . Then, using the Schurr complement, the matrix  $\mathbf{R}_i$  is positive-definite if and only if  $\sigma_i^2 - \mathbf{r}'_i\mathbf{R}_{i-1}\mathbf{r}_i > 0$ , which is guaranteed from (19), since  $V_{i,i-1} > 0$ .

**The WinBUGS model**

The following is the WinBUGS model for the example in Sect. 5.



```

Model {
  # Priors
  for (i in 1:5){mean[i] <- 0}
  y[1:5] ~ dnorm(mean[], PREC[,])
  a[1] <- 5.353; a[2] <- 3.906; a[3] <- 6.819; a[4] <- 7.838;
  a[5] <- 4.531; b[1] <- 6.577; b[2] <- 6.934; b[3] <- 9.607;
  b[4] <- 13.344; b[5] <- 2.227
  PREC[1:5,1:5] <- inverse(R[,])
  C <- 100000; zero <- 0; zero ~ dpois(phi) # zeros-trick
  prod[1] <- 1
  for (j in 1:5){
    z[j] ~ dbeta(a[j],b[j]); p[j] <- z[j]*prod[j]
    prod[j+1] <- prod[j]*(1-z[j])
    log.beta[j] <- -loggam(a[j]+b[j])+loggam(a[j])+loggam(b[j])
      -((a[j]-1)*log(z[j]))-((b[j]-1)*log(1-z[j]))
    ss[j] <- pow(y[j],2)*(PREC[j,j]-1)
    for (i in 1:5){st[j,i] <- step(j-i-0.5)*y[j]*y[i]*PREC[j,i]}
  }
  p[6] <- prod[6]
  phi <- sum(log.beta[1:5])-0.5*logdet(R[,])+0.5*sum(ss[1:5])
    +sum(st[1:5,1:5])+C
  # Likelihood
  x[1:6] ~ dmulti(p[], N) # N should equal x[1] + ... + x[6]
}
# Data should appear here as x[1] ... x[6]
# Prior Data
list( R = structure( .Data=c(
  1.000, -0.742, -0.416, -0.386, -0.079,
-0.742, 1.000, 0.145, 0.031, 0.099,
-0.416, 0.145, 1.000, -0.296, 0.076,
-0.386, 0.031, -0.296, 1.000, 0.145,
-0.079, 0.099, 0.076, 0.145, 1.000), .Dim=c(5,5))

```

## References

- Aitchison, J.: *The Statistical Analysis of Compositional Data*. Chapman and Hall, London (1986)
- Albert, I., Donnet, S., Guihenneuc-Jouyaux, C., Low-Choy, S., Mengersen, K., Rousseau, J.: Combining expert opinions in prior elicitation. *Bayesian Anal.* **7**, 503–532 (2012)
- Bunn, D.W.: Estimation of a Dirichlet prior distribution. *Omega* **6**, 371–373 (1978)
- Chaloner, K., Duncan, G.T.: Some properties of the Dirichlet-multinomial distribution and its use in prior elicitation. *Commun. Stat.-Theory. Methods* **16**, 511–523 (1987)
- Chamberlin, J.R., Featherston, F.: Selecting a voting system. *J. Polit.* **48**, 347–369 (1986)
- Clemen, R.C., Reilly, T.: Correlations and copulas for decision and risk analysis. *Manag. Sci.* **45**, 208–224 (1999)
- Conn, P.B., McClintock, B.T., Cameron, M.F., Johnson, D., Moreland, E.E., Boveng, P.L.: Accommodating species identification errors in transect surveys. *Ecology* **94**, 26072618 (2013)
- Connor, R.J., Mosimann, J.E.: Concepts of independence for proportions with a generalization of the Dirichlet distribution. *J. Am. Stat. Assoc.* **64**, 194–206 (1969)
- Cook, L., Cameron, R., Dodd, M., McConway, K., Worthington, J., Skelton, P., Anton, C., Bossdor, O., Baur, B., Schilthuisen, M., Fontaine, B., Sattmann, H., Bertorelle, G., Correia, M., Oliveira, C., Pokryszko, B., Ożgo, M., Stalažs, A., Gill, E., Üllar, Rammu, Sólymos, P., Féher, Z., Juan, X.: Citizen science reveals unexpected continental-scale evolutionary change in a model organism. *PLoS One* **6**, e18927 (2011)
- Demarta, S., McNeil, A.J.: The t copula and related copulas. *Int. Stat. Rev.* **73**, 111–129 (2005)
- Elfadaly, F.G., Garthwaite, P.H.: Eliciting Dirichlet and Connor-Mosimann prior distributions for multinomial models. *Test* **22**, 628–646 (2013)
- Elfadaly, F.G., Garthwaite, P.H.: On eliciting logistic normal priors for multinomial models. In preparation. <http://mcs-brains.open.ac.uk/elicitation/Logistic%20Paper.pdf> (2016)
- Frees, E.W., Valdez, E.A.: Understanding relations using copulas. *North Am. Actuar. J.* **2**, 125 (1998)
- Fu, H., Qu, Y., Zhu, B., Huster, W.: A Bayesian approach to the statistical analysis of device preference studies. *Pharm. Stat.* **11**, 149–156 (2012)
- Garthwaite, P.H., Kadane, J.B., O’Hagan, A.: Statistical methods for eliciting probability distributions. *J. Am. Stat. Assoc.* **100**, 680–701 (2005)

- Garthwaite, P.H., Al-Awadhi, S.A., Elfadaly, F.G., Jenkinson, D.J.: Prior distribution elicitation for generalized linear and piecewise-linear models. *J. Appl. Stat.* **40**, 59–75 (2013)
- Joe, H.: *Multivariate Models and Dependence Concepts*. Chapman & Hall, London (1997)
- Jouini, M.N., Clemen, R.T.: Copula models for aggregating expert opinions. *Oper. Res.* **44**, 444–457 (1996)
- Kadane, J.B., Dickey, J.M., Winkler, R., Smith, W., Peters, S.: Interactive elicitation of opinion for a normal linear model. *J. Am. Stat. Assoc.* **75**, 845–854 (1980)
- Kurowicka, D., Cooke, R.: *Uncertainty Analysis with High Dimensional Dependence Modelling*. Wiley Series in Probability and Statistics. Wiley, Chichester (2006)
- Leonard, T.: Bayesian estimation methods for two-way contingency tables. *J. R. Stat. Soc. Ser. B* **37**, 23–37 (1975)
- Lewandowski, D.: *High Dimensional Dependence: Copulae, Sensitivity, Sampling*. Ph.D. thesis, Delft University of Technology, Netherlands (2008)
- Lindley, D.V., Tversky, A., Brown, R.V.: On the reconciliation of probability assessments (with discussion). *J. R. Stat. Soc. Ser. A* **142**, 146–180 (1979)
- Manski, C.F.: Partial identification of counterfactual choice probabilities. *Int. Econ. Rev.* **48**, 1393–1410 (2007)
- Nelsen, R.B.: *An Introduction to Copulas*. Lecture Notes in Statistics, vol. 139. Springer, New York (1999)
- O’Hagan, A., Forster, J.: *Bayesian Inference*, volume 2B of Kendall’s *Advanced Theory of Statistics*, 2nd edn. Arnold, London (2004)
- O’Hagan, A., Buck, C.E., Daneshkhah, A., Eiser, J.R., Garthwaite, P.H., Jenkinson, D.J., Oakley, J.E., Rakow, T.: *Uncertain Judgements: Eliciting Expert Probabilities*. Wiley, Chichester (2006)
- Sklar, A.: Fonctions de répartition à  $n$  dimensions et leurs marges. *Publ. Inst. Stat. Univ. Paris* **8**, 229–231 (1959)
- Spiegelhalter, D., Thomas, A., Best, N., Lunn, D.: *WinBUGS Version 1.4.3 User Manual*. Medical Research Council Biostatistics Unit, Cambridge (2007)
- van Dorp, J.R., Mazzuchi, T.A.: Parameter specification of the beta distribution and its Dirichlet extensions utilizing quantiles. In: Gupta, A.K., Nadarajah, S. (eds.) *Handbook of Beta Distribution and Its Applications*. Marcel Dekker Inc, New York (2004)
- Weltje, G.J.: Quantitative analysis of detrital modes: statistically rigorous confidence regions in ternary diagrams and their use in sedimentary petrology. *Earth-Sci. Rev.* **57**, 211–253 (2002)
- Wilks, S.S.: *Mathematical Statistics*. Wiley, New York (1962)
- Worthington, J.P., Silvertown, J., Cook, L., Cameron, R., Dodd, M., Greenwood, R.M., McConway, K., Skelton, P.: Evolution megalab: a case study in citizen science methods. *Methods Ecol. Evol.* **3**, 303–309 (2012)
- Zapata-Vázquez, R.E., O’Hagan, A., Bastos, L.S.: Eliciting expert judgements about a set of proportions. *J. Appl. Stat.* **41**, 1919–1933 (2014)