Journal of Applied Statistics
Vol. 00, No. 00, December 2012, 1–15

RESEARCH ARTICLE

Eliciting expert judgements about a set of proportions

Rita E. Zapata-Vázquez[abc*], Anthony O'Hagan[a] and Leonardo Soares Bastos[d]

[a] *Department of Probability and Statistics, University of Sheffield, UK*; [b] *Facultad de Medicina, Universidad Autónoma de Yucatán;* [c] *UMAE "Ignacio García Téllez", Instituto Mexicano del Seguro Social, México.*; [d] *Scientific Computational Program, Oswaldo Cruz Foundation, Brazil.*

(*v4.4 released December 2012*)

Eliciting expert knowledge about several uncertain quantities is a complex task when those quantities exhibit associations. A well-known example of such a problem is eliciting knowledge about a set of uncertain proportions which must sum to 1. The usual approach is to assume that the expert's knowledge can be adequately represented by a Dirichlet distribution, since this is by far the simplest multivariate distribution that is appropriate for such a set of proportions. It is also the most convenient, particularly when the expert's prior knowledge is to be combined with a multinomial sample since then the Dirichlet is the conjugate prior family.

Several methods have been described in the literature for eliciting beliefs in the form of a Dirichlet distribution, which typically involve eliciting from the expert enough judgements to identify uniquely the Dirichlet hyperparameters.

We describe here a new method which employs the device of over-fitting, i.e. eliciting more than the minimal number of judgements, in order to (a) produce a more carefully considered Dirichlet distribution and (b) ensure that the Dirichlet distribution is indeed a reasonable fit to the expert's knowledge. The method has been implemented in a software extension of the Sheffield Elicitation Framework (SHELF) to facilitate the multivariate elicitation process.

Keywords: elicitation; Dirichlet distribution; SHELF; over-fitting.

1. Introduction

There is a growing use of elicitation to express expert knowledge about uncertain quantities in the form of a probability distribution. One such use is for the formulation of prior distributions for parameters in a statistical model, to be combined with observed data from the model by Bayesian analysis. Another is to express uncertainty about inputs to a mechanistic model that is to be used to predict a complex real-world process such as climate or the performance of an aero-engine [14]. A third use is to express uncertainty about parameters in a decision analysis. A substantial literature on elicitation is spread across many disciplines, including statistics, psychology, economics, decision-making and various application fields [1, 6, 10, 11, 17]. Where the results of elicitation have significant value it is usual for a facilitator, who is knowledgeable in probability, statistics and the processes of elicitation, to work with the expert(s) to elicit a suitable probability distribution.

Eliciting knowledge about a single uncertain quantity in the form of a univariate probability distribution is a reasonably well studied and understood task. In particular the SHELF package [16] provides a number of templates and some supporting

*Corresponding author. Email: zvazquez@uady.mx

ISSN: 0233-1934 print/ISSN 1360-0532 online

DOI: 10.1080/0266476YYxxxxxxxx
http://www.informaworld.com

software to aid a facilitator in designing and conducting an elicitation with one or more experts. However, eliciting multivariate distributions is a complex and much less well understood task. Unless two or more uncertain quantities are judged to be mutually independent it is not enough to elicit the (marginal) distributions of each quantity separately. One approach is to elicit marginal distributions together with some specific measures of association (such as rank correlation coefficients), and then to use a copula to construct the multivariate distribution [12, 13]. However, the key challenge is to elicit judgements about associations between uncertain quantities, and even experts with a background in statistics have shown limited ability to assess correlations well. In the absence of other generic methods for multivariate elicitation, it is common for the facilitator to resort to problem-specific solutions such as methods based on the idea of elaboration [14].

An example of a multivariate elicitation task where the quantities are inevitably correlated is eliciting knowledge about a set of proportions that must sum to 1. The elicited distribution must satisfy this constraint, and by far the simplest and most widely used distributions with that property are the Dirichlet family. This makes the Dirichlet an attractive choice for expressing beliefs about a set of proportions, particularly when the task is to elicit a prior distribution for use in a Bayesian analysis because the Dirichlet is the conjugate family to the multinomial likelihood [3]. In Section 2 we review the Dirichlet distribution and methods previously proposed in the literature for elicitation. We propose a new method in Section 3, with an example presented in Section 3.4. A feature of our presentation in Section 3 is that the elicitation method is given in a way that is full and detailed but as non-technical as possible.

2. The Dirichlet distribution and elicitation methods

We consider eliciting expert beliefs about a set of k uncertain quantities $\pi = (\pi_1, \pi_2, \ldots, \pi_k)$ which are constrained to lie on the $(k-1)$-dimensional simplex, i.e. $\pi_i \geq 0$ for $i = 1, 2, \ldots, k$ and $\sum_{i=1}^{k} \pi_i = 1$. We shall think of π_i as the proportion of members of some population which belong to category i. An example with $k = 4$ is the proportion of adults in the UK who are classified (according to a criterion such as body mass index) as under-weight ($i = 1$), neither over-weight nor under-weight ($i = 2$), over-weight but not obese ($i = 3$) and obese ($i = 4$). We will suppose for ease of presentation that the elicitation will seek the views of a single expert, although the method will work equally well with a group of experts who jointly agree on their judgements.

The result of the elicitation should be a probability distribution on the simplex representing the expert's knowledge about π. The simplest class of distributions defined on the simplex is the Dirichlet family, and we suppose that the result of the elicitation should be a Dirichlet distribution unless this is not a reasonable representation of the expert's knowledge.

2.1 *Definition and properties of the Dirichlet distribution*

We say that π has the Dirichlet distribution with parameter vector $\mathbf{d} = (d_1 \ldots, d_k)$, denoted by $\pi \sim Di(\mathbf{d})$, if it has probability density function

$$f(\boldsymbol{\pi}|\boldsymbol{d}) = c(\mathbf{d}) \prod_{i=1}^{k} \pi_i^{d_i - 1} \tag{1}$$

(for $\pi_i > 0$ and $\sum_{i=1}^k \pi_i = 1$, otherwise the density is zero), where the normalising constant is

$$c(\mathbf{d}) = \Gamma\left(\sum_{i=1}^k d_i\right) / \prod_{i=1}^k \Gamma(d_i) \ .$$

The means and variances of $\pi_1, \ldots, \pi_k$, are given, respectively, by:

$$E(\pi_i \mid d_i) = \frac{d_i}{n}, \qquad \mathrm{Var}(\pi_i \mid d_i) = \frac{d_i(n-d_i)}{n^2(n+1)}, \qquad \text{where } n = \sum_{i=1}^k d_i \ . \qquad (2)$$

The constraint that $\sum_{i=1}^k \pi_i = 1$ means that the π_is are inevitably correlated, as shown by the covariance

$$\mathrm{Cov}(\pi_i, \pi_j \mid d_i) = -\frac{d_i d_j}{n^2(n+1)} \ .$$

Two properties of the Dirichlet distribution, which we will refer to as the marginal and conditional properties, are useful for elicitation. The first property concerns marginal distributions. Let the first $m < k-1$ elements of π be denoted by $\pi^m = (\pi_1, \pi_2, \ldots, \pi_m)$, and let $\pi^*_{m+1} = \sum_{i=m+1}^k \pi_i = 1 - \sum_{i=1}^m \pi_i$. Then $\pi^{m+} = (\pi^m, \pi^*_{m+1}) = (\pi_1, \ldots, \pi_m, \pi^*_{m+1})$ takes values in the m-dimensional simplex. The marginal property is that the distribution of π^{m+} is $Di(d_1, d_2, \ldots, d_m, d^*_{m+1})$ where $d^*_{m+1} = \sum_{i=m+1}^k d_i = n - \sum_{i=1}^m d_i$.

A special case of the marginal property says that the marginal distribution of π_i is the beta distribution with parameters d_i and $n - d_i$, i.e. $\pi_i \sim Be(d_i, n-d_i)$. Indeed, the moments (2) follow directly from this fact.

The second property concerns conditional distributions. For $i = m+1, \ldots, k$ let

$$\pi'_i = \pi_i / (1 - \pi_1 - \pi_2 - \ldots - \pi_m) \ .$$

Notice that $\pi' = (\pi'_{m+1}, \ldots, \pi'_k)$ satisfies the conditions to lie on the $(k-m-1)$-dimensional simplex. The conditional property is that the conditional distribution of π' given π^m (or π^{m+}) is $Di(d_{m+1}, \ldots, d_k)$. By repeatedly using this property we can decompose the Dirichlet distribution into a sequence of $k-1$ beta conditional distributions.

It is worth noting an implication of the two conditions. The sum of the parameters of the Dirichlet conditional distribution of π' given π^{m+} is equal to d^*_{m+1}, which is the last parameter of the marginal Dirichlet distribution of π^{m+}.

2.2 *Elicitation methods*

The constraint that $\sum_{i=1}^k \pi_i = 1$ means that although there are k individual quantities π_i only $k-1$ are uncertain, since once any $k-1$ are specified the k-th is determined by the constraint. Instead of the k parameters d_i, the Dirichlet family is sometimes parametrised with $p_i = d_i/n$ $(i = 1, 2, \ldots, k)$ and n. Although there now appear to be $k+1$ parameters there is a constraint that $\sum_{i=1}^k p_i = 1$, so that there are in effect only k. In this parametrisation we see that the p_is control the means (or other location measures) of the π_is, while n controls the overall amount of uncertainty. We could therefore ask the expert to provide a judgement of the expected value of each π_i together with one further judgement concerned with uncertainty to identify n.

Methods that have been proposed in the literature for eliciting a Dirichlet distribution can mostly be viewed as suggesting alternative kinds of judgement of location to identify the p_is and/or alternative kinds of judgement of uncertainty to identify n. For instance, to elicit a value for p_i the evidence from the psychology literature is that people do not judge expectations well, so we could ask the expert for a median value of π_i because the judgement of equal probability (above and below the median) is generally made more accurately. We could alternatively ask for the expert's mode, which is convenient because unlike the median it has the closed form expression $(d_i - 1)/(n - 2)$ provided both numerator and denominator are positive.

Some other approaches are based on predictive distributions. If we consider N draws from the population, then the expectation of the number N_j to be drawn from category i is p_i. So we could ask the expert to specify their mean of this number for each i (constrained to sum to N) [4]. Chaloner and Duncan [5] ask the expert to specify the most probable set of counts $(N_1, N_2, \dots, N_k)$, i.e. the joint mode of the predictive distribution. It is worth noting that we can also elicit the p_is using simple probabilities by considering the case $N = 1$. Then p_i is simply the probability that a single draw from the population is found to be in category i.

One way to identify n is simply by eliciting a measure of uncertainty about a single π_i. Suppose for instance that the expert is asked for a judgement of the variance of π_1. If p_i has also been specified then from the equations (2) we can deduce n by

$$n = \{p_1(1 - p_1)/\mathrm{Var}(\pi_1 \mid \mathbf{d})\} - 1 \ .$$

Any other specification of uncertainty about π_1 (such as the interquartile range) might be used in a similar way to fix n.

Other approaches elicit n as a measure of the amount of information that the expert has. Through the Dirichlet distribution's role as the conjugate family for sampling with replacement from the population [2, 15], n can be thought as defining the strength of the expert's information in terms of an equivalent number of observations from the population. Another way to think of the strength of information is as determining the extent to which the expert's judgements would change when presented with additional information. Thus, if the expert were to observe one draw from the population, and that item was found to be in category i, then (using the Dirichlet conjugacy) the expert's expected value of π_i would change from $p_i = d_i/n$ to $p_i' = (d_i + 1)/(n + 1)$. So if the expert were asked to imagine making that observation and to specify what his/her expectation for π_i (or some equivalent location judgement) would then be, we can deduce n as

$$n = (1 - p_i')/(p_i' - p_i) \ .$$

This is known as the device of imaginary observations [4, 9]. Chaloner and Duncan [5] more explicitly link n to a measure of uncertainty by asking the expert to specify the probability for a range of values around the joint mode of $(N_1, N_2, \dots, N_k)$.

2.3 *Beyond the Dirichlet*

The fact that the Dirichlet distribution has only one parameter to control overall uncertainty is a serious limitation. It means that as soon as we have specified n, for instance through the variance of one of the π_is then the variances of all the π_is are now determined by n and the p_is. If the expert's uncertainty about different π_is did not agree reasonably well with that system of constraints then their beliefs could not be adequately described by a Dirichlet distribution. A number of generalisations of the Dirichlet distribution have been proposed to provide more flexibility in representing an expert's knowledge. One very flexible family is the finite mixtures of Dirichlet distributions [7], but mixtures introduce a very large number of additional parameters and thereby demand a correspondingly large number of additional judgements to be elicited.

An early generalisation of the Dirichlet distribution was introduced by Dickey [8]. It is based on the conditional property of the Dirichlet, but removes the constraint noted at the end of Subsection 2.1. Thus Dickey allows the conditional distribution of each π'_{m+1} conditional on $\pi_1, \ldots, \pi_m$ to be any beta distribution. There are now $2(d-1)$ parameters in this generalised Dirichlet distribution, and these can clearly be specified by asking the expert for judgements about these conditional distributions. The disadvantages of this approach are that it is dependent on the order in which we place the categories, and that it is difficult for the expert to make judgements about quantities such π'_{m+1}.

3. An elicitation method for expert beliefs about proportions

The discussion of methods in Section 2 is not intended to be an exhaustive review but instead to exemplify the range of methods already in the literature for eliciting judgements about a set of proportions in the form of a Dirichlet or generalised Dirichlet distribution. The papers cited all date from a time when the limitations of Bayesian computation meant that analysis would be extremely difficult unless the prior distribution were specified as a member of the relevant conjugate family. So the literature of the time was expressed in terms of eliciting a Dirichlet distribution, rather than as eliciting expert beliefs about proportions. With modern computing methods, conjugate distributions are still convenient and widely used but Bayesian analysts should use them only when they genuinely reflect expert knowledge adequately. Another difference since the time of those papers is that elicitation methodology has moved on. The purpose of our article is to present a novel approach to eliciting expert beliefs about a set of proportions in a way that reflects modern elicitation methodology and also facilitates the assessment of whether a Dirichlet distribution is a suitable choice in the light of the expert's judgements. The outcome of this approach is either a Dirichlet distribution that is an acceptable representation of the expert's knowledge, or else a conclusion that the expert's knowledge is such that no Dirichlet distribution would represent it adequately.

3.1 *Overview*

Our approach has two key features.

(1) We make use of the device known as *over-fitting*. That is, we elicit more judgements from an expert than are needed to fit a Dirichlet distribution. Over-fitting has two principal benefits. First, in practice the expert's judgements will not all fit perfectly a Dirichlet distribution with any $\mathbf{d}$, because the expert's knowledge may not be perfectly represented by a Dirichlet distribution and because the expert's judgements are necessarily themselves imprecise. The degree to which a Dirichlet distribution with a well chosen $\mathbf{d}$ can fit the expert's elicited values allows us to assess whether the Dirichlet distribution is an adequate representation. Second, in the same way that

more data allows more accurate estimation in statistical analysis generally, more elicited quantities allow us to more accurately identify a good choice of **d**.

(2) We use the Sheffield Elicitation Framework (SHELF). SHELF is a package of documents, templates and software providing structured elicitation protocols that conform to good modern elicitation practice [16]. It is available for download free of charge at `http:www.tonyohagan.co.uk/shelf/index.html`. The software routines are implemented in the free R software (`http:www.r-project.org`) for statistical computing and graphics. Over-fitting is a feature of the SHELF elicitation protocols.

SHELF is of course just one framework for eliciting a distribution for each π_i and an experienced facilitator might well choose to use another method. However, we recommend the use of SHELF for less experienced facilitators and our presentation in the next section is based on this. Furthermore, we have developed add-ons for SHELF to implement our method. Although the official SHELF package currently only contains procedures for eliciting univariate distributions, we hope that our add-ons will be included in the next official release. They are otherwise available at `http://tonyohagan.co.uk/shelf/dirichlet.html`.

It should be noted that over-fitting has been a part of previously proposed elicitation schemes for a Dirichlet distribution. For instance, Chaloner and Duncan [5] suggest using their method of specifying n with different values of the number N of imaginary future observations. Similarly, in methods which specify n using uncertainty in a single π_i, such as Dickey et al. [9], this can be done with each π_i to obtain various values of n. We adopt a similar approach, but with a clear procedure to obtain a 'best' n and to assess adequacy of fit.

A simple outline of our method is as follows. First, we use SHELF to elicit the expert's beliefs about each π_i in turn. If the expert's elicited judgements can be adequately represented by a beta distribution in each case, then SHELF identifies the parameters of the respective beta distributions. We then adjust these to conform to the constraint that $\sum_{i=1}^{k} \pi_i = 1$. Each beta distribution will in general imply a different value of n, so the next step is to find a best fitting value for n. Finally, we assess whether this value adequately represents all the expert's judgements.

3.2 *Detailed procedures*

A description of the steps involved in the proposed method for eliciting a Dirichlet distribution follows. The method is applicable for eliciting from multiple experts, but for ease of exposition we assume there is only one expert, and that the expert is female.

I. *Preparation and training*

As in any elicitation process, it is important to prepare carefully. There is advice on these matters in the SHELF package and much additional information in *O'Hagan, Buck and cols.* [17]. We simply note here that it is important for the expert to be committed to providing her best judgements, to train the expert in the meaning of personal probability judgements and distributions, and to familiarise her with the specific elicitation method that will be used, ideally through the use of a practice exercise.

II. *Elicit beta distributions for each π_i using SHELF.*

For each π_i in turn, the SHELF framework is followed to elicit a Beta distribution.

Within SHELF there are several methods available to elicit a distribution, involving different judgements to be elicited from the expert. The example in Section 3.4 uses the quartile method. The example also illustrates how to use the SHELF R functions to fit a beta distribution to the expert's elicited judgements.

Having fitted a beta distribution, it is important for the facilitator to verify that the expert acknowledges that the fitted distribution is a reasonable representation of her beliefs about π_i. SHELF methods all incorporate an element of over-fitting, because only two judgements are required to fit a unique beta distribution but we always elicit at least three judgements. The fitted distribution is necessarily a compromise because in general no beta distribution will exactly fit all the elicited values. The software selects the beta distribution which fits the expert's judgements as closely as possible. The facilitator now feeds back various summaries of the fitted distribution, including showing the density function and reporting how closely this distribution fits the expert's judgements. This is an opportunity for the expert to say that this is not an adequate match to her beliefs. In that case the facilitator will explore the nature of the expert's dissatisfaction, perhaps eliciting additional judgements. If no beta distribution can be adequate (for instance if the expert insists that her beliefs should be represented by a bimodal density function), then this procedure terminates with the conclusion that no Dirichlet distribution can represent the expert's beliefs about the whole π vector.

III. *Check and adjust means.*

Assuming that beta distributions can adequately represent the expert's beliefs about each π_i separately, the next step is to check consistency with the constraint that $\sum \pi_i = 1$. Suppose that the elicited distribution for π_i is $Be(d_i, e_i)$. Then this implies $E(\pi_i) = \frac{d_i}{d_i+e_i}$. The constraint now requires that

$$\sum_{i=1}^{k} \frac{d_i}{d_i + e_i} = 1 \ . \tag{3}$$

In practice, it is unlikely that the separate elicitations will lead to d_i and e_i values satisfying this equation. In the common situation when all of the elicited median values are less than 0.5, the beta distributions will all be positively skewed. Even if the implication of this is explained and fully understood by the expert, the temptation is to specify median values which sum to 1. As a result we may often find that the sum of expected values exceeds 1. Whether greater or less than 1, it is necessary to adjust the elicited distributions so that (3) is satisfied. If the discrepancy is large, outside the range $[0.9, 1.1]$ for instance, then it may be necessary to review the entire elicitation to resolve the expert's apparent misunderstanding.

A small discrepancy can simply be corrected by a mechanical adjustment to the d_i and e_i values. If the sum on the left hand side of (3) is $r \neq 1$, then new values d_i^* and e_i^* are given by

$$d_i^* = d_i / r \ , \qquad e_i^* = d_i + e_i - d_i^* \ . \tag{4}$$

The facilitator might choose to give new feedback to the expert on how closely the new fitted, mean-adjusted Beta distributions match her originally elicited judgements. However, in practice it is unlikely that the quality of fit will be changed appreciably when r is close to 1.

IV. *Finding a compromise value of n.*

At this point you should have an acceptable beta distribution for each hyperparameter of the Dirichlet distribution of interest. The separate beta distributions correspond to a Dirichlet distribution if the values of $n_i = d_i^* + e_i^*$ are all equal. In practice this is unlikely to happen. If a Dirichlet distribution is to be found that is an acceptable representation of the expert's beliefs then we seek a compromise value of n to act as a common value to replace the disparate n_is. Given any proposed n we define the corresponding Dirichlet distribution to be

$$Di(\mathbf{d}(n)) \ , \tag{5}$$

where the parameter vector $\mathbf{d}(n)$ has i-th element

$$d_i(n) = n \, \frac{d_i^*}{d_i^* + e_i^*} \ .$$

What value of n would produce a distribution (5) that best reflects the expert's knowledge? In general, a higher n implies stronger information and an appropriate compromise value should lie between $n_{\min} = \min_i \{n_i\}$ and $n_{\max} = \max_i \{n_i\}$. Ideally we should choose the n value that makes the Dirichlet fit all of the expert's elicited judgements as accurately as possible. However, this entails a very large computation in practice and is surely unnecessary because a value chosen by simpler or more approximate methods can be expected to provide an essentially equivalent fit. The following might be considered.

- *Use a compromise value, somewhere around the middle of the range from $n_{\min}$ to $n_{\max}$.* The strict middle value is $n_{mid} = (n_{\min} + n_{\max})/2$, but we might equally consider the mean $\bar{n} = \sum n_i/k$ or the median n_{med} of the n_is. Any one of these might be expected to approximate the ideal value n_{opt}.
- *Use a simplified optimisation.* Create a simple objective criterion $F(n)$, and choose n to optimise this criterion. A criterion $F(n)$ that we have found useful in practice is based on how closely the Dirichlet distribution $Di(\mathbf{d}(n))$ matches the standard deviations of the separately elicited beta distributions. In the Appendix, this criterion is formulated and shown to yield the optimised value

$$n_{opt} = \left(\frac{\sum_{i=1}^k v_i^* \, (n_i + 1)}{\sum_{i=1}^k v_i^* \sqrt{n_i + 1}} \right)^2 - 1 \ , \tag{6}$$

  where

$$v_i^* = \frac{d_i^*(n_i - d_i^*)}{n_i^2(n_i + 1)}$$

  is the variance of the i-th fitted, mean-adjusted beta distribution $Be(d_i^*, n_i - d_i^*)$, using (2).

A quite different approach is:

- *Use a conservative value.* Choose $n = n_{\min}$ because it does not express more knowledge about any element π_i than was elicited.

In using $n_{\min}$ we choose not to seek a Dirichlet distribution to represent adequately the expert's knowledge. Instead, we are using a Dirichlet distribution for convenience and choose to play safe by implying no more knowledge about the π_is as a whole than the expert has expressed about any one of them.

V. *Proceed to feedback.*

If we have computed a representative central value like n_{mid} or $\bar{n}$, or if we have found a simplified optimal value n_{opt}, it is now important to present feedback again to the expert on the implications of the fitted Dirichlet distribution. In particular, this will entail looking at the implied marginal density for each π_i and seeing how closely it matches the original elicited values. If the expert finds the fitted Dirichlet distribution acceptable, then the procedure terminates with this elicited distribution as its conclusion.

On the other hand, particularly if the original n_i values were not reasonably similar, we may find that the fit is now too poor for the Dirichlet to adequately represent the expert's knowledge. In this case the procedure terminates with the conclusion that no Dirichlet distribution would represent a suitable joint distribution for π.

Note that if we have chosen to use the conservative value $n_{\min}$ then no useful purpose is served by this feedback step. If we were to carry it out we might find the not-unexpected result that it does not fit the expert's judgements well, whereas it was never intended to. The procedure concludes with the chosen Dirichlet distribution $Di(\mathbf{d}(n_{\min}))$.

3.3 *New SHELF materials*

We have developed additional materials for the SHELF package to facilitate the use of this procedure, available for download from `http://tonyohagan.co.uk/shelf/dirichlet.html`. These include a new template for recording the elicitation, new R functions for doing various computations and guidance notes on how to use these materials.

3.4 *Example*

This example concerns the efficacy of a new antibiotic in patients who are hospitalised in the Pediatric Intensive Care Unit (PICU) and who are severely infected by pneumococci (which is associated with pneumonia, meningitis, and septicaemia, among other conditions). The possible results after the infection are: to survive in good condition, to have a sequel, or to die. An expert is asked to provide judgements about the proportions of patients who will have each of these possible results. Denoting these proportions by π_1, π_2 and π_3, these form a set of proportions that must sum to 1.

I. *Preparation and training.* The facilitator chooses to use the SHELF package, and so has downloaded the necessary materials. He has also installed the *rpanel* library in order to use the SHELF R functions, as described in the instructions. SHELF offers a variety of specific protocols for eliciting knowledge about a single distribution, and the facilitator chooses to use the Quartile method.

The purpose of the elicitation is explained to the expert, including the nature of the antibiotic and the classification of patients according to three possible results after pneumococcal infection. The facilitator also gives training in probability judgements, including a practice exercise to familiarise the expert with the Quartile method. The facilitator completes the "SHELF 1 (Context)" form to record basic details such as date, time, participants, purpose of the elicitation and what training was given.

II. *Elicit beta distributions for each π_i using SHELF.* The facilitator now follows the Quartile procedure as set out in the "SHELF 2 (Distribution) Q" form. One important point to note is that when SHELF fits a beta distribution it will in general fit a scaled beta over the range $[L, U]$, where L and U are lower and upper plausible bounds for the quantity of interest that are judgements elicited from the expert. For the purposes of the method developed here, it is important that unscaled beta distributions are fitted. Therefore, whichever SHELF protocol is used, instead of asking the expert for judgements of L and U these are fixed at $L = 0$ and $U = 1$. The facilitator enters these values into the SHELF software.

The expert was first asked for his judgements about the proportion π_1 of patients in the PICU that he believes will survive in good condition after receiving the new drug. He gave a median value of 0.55 for π_1, confirming that he judged it equally likely that π_1 would be less than 0.55 or greater than 0.55. Following appropriate questioning from the facilitator, as described in the SHELF guidance, the expert also gave values of 0.60 for the upper quartile and 0.50 for the lower quartile. These values are shown in Figure 1 (at left), which are screen shots from the SHELF software. The fitted beta distribution has parameters $d_1 = 25.4, e_1 = 20.8$. This beta distribution was shown to the expert and feedback is given in the form of the 10-th and 90-th percentiles of the $Be(25.4, 20.8)$ distribution, respectively 0.46 and 0.64. The expert agreed that according to his judgement there was a probability of approximately 10% in each case that π_1 would lie below 0.46 or above 0.64. Furthermore, the beta distribution $Be(25.4, 20.8)$ fitted the expert's judgements well, since its lower quartile, median and upper quartile are 0.50, 0.55 and 0.60 respectively, agreeing to two decimal places with the expert's elicited quartiles. The expert confirmed that the fitted distribution was a reasonable representation of his knowledge about π_1 (Table 1).

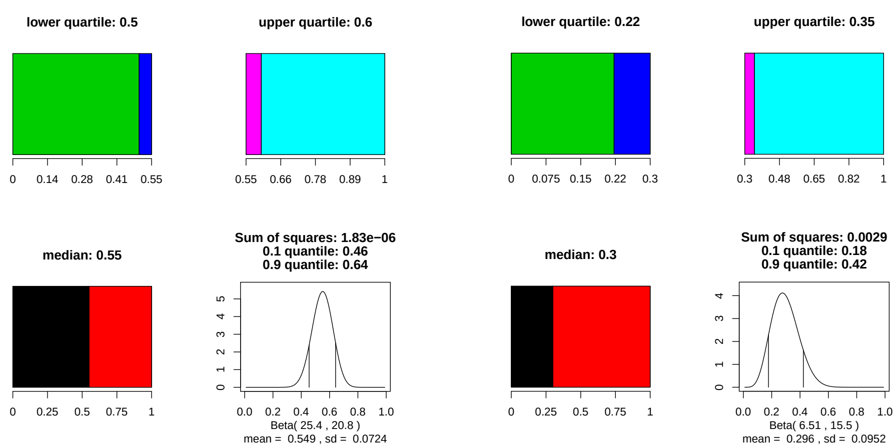

Figure 1. SHELF windows. The quartiles, median and beta distribution for outcomes: at left good outcome π_1; and at right, the quartiles, median and the beta distribution for patients with sequels π_2.

The facilitator next asked the expert to think about π_2, the proportion of patients who would have sequels. However, the expert asked to deal with π_3 first, the proportion who would die, saying that he felt better able to judge this proportion. His judgements were based on his experience with a similar antibiotic. Following the Quartile elicitation protocol again, he gave values of 0.15 for the median, 0.11 for the lower quartile and 0.20 for the upper quartile. The fitted beta distribution is shown in Table 1, and through feedback and comparison of the fitted quartiles with the elicited values was confirmed as a reasonable representation of the expert's knowledge about π_3.

The expert was then asked to think about π_2, although he admitted to feeling least sure about this proportion. Having stated median values of 0.55 and 0.15 for π_1 and π_3, respectively, he chose a median of 0.30 for π_2, but gave quite wide quartile bounds; his lower quartile is 0.22 and his upper quartile is 0.35. The fitted beta distribution is $Be(6.51, 15.5)$. Table 1 shows the elicited and the fitted values and the facilitator pointed out that the upper quartile of the fitted distribution was a little higher than the elicited value, while the median was a little lower. The facilitator asked whether the expert would (a) accept the fitted distribution (and implicitly change his median and quartiles to those of the fitted distribution), (b) rethink and adjust his judgements in a different way, or (c) hold to his original judgements. The expert chose option (a). Had he chosen option (b), a fresh beta distribution would have been fitted to his new judgements and fresh feedback given. Had he chosen option (c) the elicitation according to the present procedure would have had to terminate with the conclusion that the expert's knowledge could not be represented adequately with a Dirichlet distribution.

Table 1. Elicitation and fitted values for the example

	Good Outcome	Sequel	Dead
The elicited median and quartiles	0.55 (0.50 - 0.60)	0.30 (0.22 - 0.35)	0.15 (0.11 - 0.20)
The fitted beta parameters $Be(d_i, e_i)$	$Be(25.4, 20.8)$	$Be(6.51, 15.5)$	$Be(4.46, 23.6)$
The fitted 10th and 90th percentiles	0.46 - 0.64	0.18 - 0.42	0.078 - 0.25
The fitted median and quartiles	0.55 (0.50 - 0.60)	0.29 (0.22 - 0.36)	0.15 (0.11 - 0.20)
The fitted mean and standard deviation	0.550 (0.072)	0.296 (0.095)	0.159 (0.068)
The sum of the parameters, $n_i = d_i + e_i$	46.2	22.01	28.06

III. *Check and adjust the means.* Applying equation (3), the sum of the mean values (given in Table 1) is $r = 1.004$. This is so close to 1 that the adjustment indicated in equation (4) would have negligible effect in this case. Therefore no adjustment was applied and we set $d_i^* = d_i$, $e_i^* = e_i$.

IV. *Finding a compromise value of n.* We have three n_i values shown in Table 1. Computing the various derived n values defined in Section 3.2 we find the six figures in the first column of Table 2.

These computations may be done using the new SHELF R software mentioned in Section 3.3.

The four compromise values $n_{med}, n_{mid}, \bar{n}$ and n_{opt} cover a substantial range, with n_{mid} being 20% larger than n_{med}. The second column of Table 2 shows the corresponding values of the simple optimality criterion. We see that the two extreme values $n_{\max}$ and $n_{\min}$ produce much higher values of $F(n)$ than the compromise values, but it is noticeable that $\bar{n}$ is very close to optimal. Figure 2 was also produced using the new SHELF R functions and shows the optimal $n_{opt} = 30.975$ but also shows that n values between 30 and 32 are all very close to optimal.

V. *Proceed to feedback.*

Table 2. The criterion $F(n)$ for different values of n

n	F(n)
$n_{max} = 46.2$	$(0.072 - 0.072)^2 + (0.066 - 0.095)^2 + (0.053 - 0.068)^2 = 0.001038$
$n_{min} = 22.01$	$(0.104 - 0.072)^2 + (0.095 - 0.095)^2 + (0.076 - 0.068)^2 = 0.001050$
$n_{med} = 28.06$	$(0.092 - 0.072)^2 + (0.085 - 0.095)^2 + (0.068 - 0.068)^2 = 0.000505$
$n_{mid} = 34.105$	$(0.084 - 0.072)^2 + (0.077 - 0.095)^2 + (0.062 - 0.068)^2 = 0.000499$
$\bar{n} = 32.09$	$(0.086 - 0.072)^2 + (0.079 - 0.095)^2 + (0.063 - 0.068)^2 = 0.000466$
$n_{opt} = 30.975$	$(0.088 - 0.072)^2 + (0.081 - 0.095)^2 + (0.065 - 0.068)^2 = 0.000461$

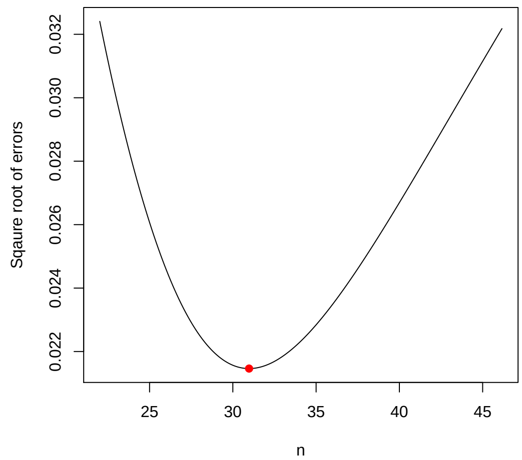

Figure 2. n values versus $\sqrt{F(n)}$, with the minimum shown at $n = n_{opt}$.

The facilitator proposed the value of $n_{opt} = 30.975$ and further feedback was now provided to the expert in order to check whether the compromise Dirichlet distribution $Di(\mathbf{d}(n_{opt}))$ would adequately represent the expert's knowledge. In this case, the three separately elicited values of 46.2, 22.01 and 28.06 covered a substantial range, so that any compromise distribution would imply probabilities for the individual π_is that might differ from the expert's initial judgements in ways that he would not find acceptable. In particular, $n = 30.975$ implies more uncertainty about π_1 (because $n_1 = 46.2$ is considerably larger) and less uncertainty about π_2 and π_3 than the expert originally specified.

Table 3. Original fitted values and optimum fitted values to feed back to the expert for the example

Original values	Good Outcome	Sequel	Dead
The elicited median and quartiles	0.55 (0.50 - 0.60)	0.30 (0.22 - 0.35)	0.15 (0.11 - 0.20)
The fitted beta parameters $Be(d_i, e_i)$	$Be(25.4, 20.8)$	$Be(6.51, 15.5)$	$Be(4.46, 23.6)$
The fitted 10th and 90th percentiles	0.46 - 0.64	0.18 - 0.42	0.078 - 0.25
The fitted median and quartiles	0.55 (0.50 - 0.60)	0.29 (0.22 - 0.36)	0.15 (0.11 - 0.20)
The fitted mean and standard deviation	0.550 (0.072)	0.296 (0.095)	0.159 (0.068)
The sum of the parameters, $n_i = d_i + e_i$	46.2	22.01	28.06
the optimum value $n = 30.975$ **implies:**			
The beta parameters $Be(d_i, e_i)$	$Be(17.03, 13.945)$	$Be(9.162, 21.813)$	$Be(4.923, 26.052)$
The 10th and 90th percentiles	0.44 - 0.66	0.19 - 0.40	0.082 - 0.25
The median and quartiles	0.54 (0.50 - 0.62)	0.30 (0.23 - 0.34)	0.15 (0.11 - 0.20)
The mean and standard deviation	0.551 (0.0879)	0.296 (0.0811)	0.16 (0.0643)
The sum of the parameters, $n_i = d_i + e_i$	30.975	30.975	30.975

The facilitator again invited the expert to respond to the feedback by (a) accepting that the medians, quartiles and percentiles implied by the fitted Dirichlet distribution were adequate representations of his knowledge and beliefs (despite the differences from his original judgements), (b) revising one or more of his original judgements which now seem to have been ill-judged, or (c) insist that his original judgements should stand and that the fitted Dirichlet distribution implies values that are not close enough to those judgements. The expert chose response (a), feeling that the differences from his original judgements were within the range of accuracy of those judgements. Had he chosen response (b), the elicitation process would have gone back to step II for refitting and feedback on individual beta distributions using the revised judgements. Had he chosen response (c), the facilitator would have attempted to identify an alternative compromise value of n that would be acceptable, and failing that the procedure would have ended with the conclusion that no Dirichlet distribution could adequately represent the expert's knowledge.

The Dirichlet distribution with parameters $\mathbf{d} = (17.03, 9.162, 4.923)$ was agreed as the outcome of the elicitation.

4. Discussion

In this article, we address a multivariate elicitation problem, the elicitation of a joint distribution for a set of proportions which must logically sum to 1.The key features of the contribution presented here are as follows.

- In most contexts, the most convenient distribution for a set of proportions is a Dirichlet distribution. Our method delivers an elicited distribution *provided* that such a distribution can adequately represent the expert's knowledge and judgements.
- Our method includes specific steps of overfitting and feedback which allow the judgement of whether a Dirichlet distribution can indeed represent the expert's knowledge adequately, and has clear rules to identify when it cannot.
- The elicitation protocol builds on established good practice in elicitation of univariate distributions, and in particular we recommend the use of the SHELF system for eliciting the marginal beta distributions.
- We have provided freely-available templates and software to augment the standard SHELF package, in order to assist facilitators in applying our method in practice.

The whole area of multivariate elicitation has been only sparsely studied in the literature. Although there are existing methods for eliciting a Dirichlet distribution, we believe that the above features make our method unique.

Multivariate elicitation is difficult. One reason is that for any given context there are few standard multivariate distribution to use to represent the expert's knowledge. For a set of proportions, there are alternatives to the Dirichlet, but they would be much more complex to use in practice. One direction for future research is to consider how to employ alternative distributions when our procedure determines that a Dirichlet distribution is not adequate. Another future research direction is to tackle other kinds of multivariate elicitation and develop methods with the same kind of key features.

4.1 *Acknowledgements*

The first named author is grateful for support from PROMEP and IMSS to carry out this work.

APPENDIX

Analytical solution to optimisation of n

We have k categories, with elicited beta distributions having adjusted parameters d_i^* and e_i^*. The categories have different values of $n_i = d_i^* + e_i^*$ and we seek a compromise n to minimise a criterion $F(n)$ which we define to be the sum of squared differences between the standard deviations implied by the individual $Be(d_i^*, e_i^*)$ distributions and the compromise $Di(\mathbf{d}(n))$ distribution, whose parameters are

$$d_i(n) = d_i^* n / n_i \ .$$

The variance of $Be(d_i^*, n_i - d_i^*)$ is

$$v_i^* = \frac{d_i^*(n_i - d_i^*)}{n_i^2(n_i+1)} \ ,$$

whereas the variance of the i-th parameter in the $Di(\mathbf{d}(n))$ distribution is

$$\begin{aligned} v_i &= \frac{d_i(n)(n - d_i(n))}{n^2(n+1)} = \frac{(d_i^* n/n_i)(n - d_i^* n/n_i)}{n^2(n+1)} \\ &= \frac{d_i^*(n_i - d_i^*)}{n_i^2(n+1)} = v_i^* \ \frac{n_i+1}{n+1} \ . \end{aligned}$$

So the problem is to minimise

$$F(n) = \sum_{i=1}^{k} \left(\sqrt{v_i^*} - \sqrt{v_i}\right)^2 = \sum_{i=1}^{k} v_i^* \left(1 - \sqrt{\frac{n_i+1}{n+1}}\right)^2$$

with respect to n. We can do this simply by solving $\frac{d}{dn}F(n) = 0$. We find

$$\begin{aligned} \frac{dF(n)}{dn} &= \sum_{i=1}^{k} 2v_i^* \left(1 - \sqrt{\frac{n_i+1}{n+1}}\right) \frac{1}{2}\sqrt{\frac{n_i+1}{n+1}} . \frac{1}{n+1} \\ &= \frac{1}{n+1} \sum_{i=1}^{k} v_i^* \left(1 - \sqrt{\frac{n_i+1}{n+1}}\right) \sqrt{\frac{n_i+1}{n+1}} \ , \end{aligned}$$

which equals zero if

$$\begin{aligned} \sum_{i=1}^{k} v_i^* \left(\sqrt{n+1} - \sqrt{n_i+1}\right) \sqrt{n_i+1} = 0 \\ \therefore \sqrt{n+1} \sum_{i=1}^{k} v_i^* \sqrt{n_i+1} = \sum_{i=1}^{k} v_i^* \left(n_i+1\right) \\ \therefore n = \left(\frac{\sum_{i=1}^{k} v_i^* \left(n_i+1\right)}{\sum_{i=1}^{k} v_i^* \sqrt{n_i+1}}\right)^2 - 1 \ . \end{aligned}$$

It is easy to see that $F(n)$ has just one minimum, and so this is the solution that is denoted by n_{opt} in (6).

References

[1] M.C. Baddeley, A. Curtis, and R.A. Wood, *An introduction to prior information derived from probabilistic judgements: Elicitation of knowledge, cognitive bias and herding*, Geol Soc. 239 (2004), pp. 15–27.
[2] J.O. Berger, *Statistical decision theory and Bayesian analysis*, 2nd ed., Springer-Verlag, New York, 1985.
[3] J.M. Bernardo and A.F. Smith, *Bayesian theory*, J Wiley & Son, England, 1994.
[4] D. Bunn, *Estimation of subjective probability distributions in forecasting and decision making*, Technological Forecasting and Social Change. 14 (1979), pp. 205–16.
[5] K.M. Chaloner and G.T. Duncan, *Some properties of the Dirichlet-multinomial distribution and its use in prior elicitation*, Commun. Statist.-Theory Meth. 16(2) (1987), pp. 511–523.
[6] K.M. Chaloner and F.S. Rhame, *Quantifying and documenting prior beliefs in clinical trials*, Stat Med. 20 (2001), pp. 581–600.
[7] S. Dalal and W. Hall, *Approximating priors by mixtures of natural conjugate priors*, Journal of the Royal Statistical Society Series B. 45 (1983), pp. 278–86.
[8] J.M. Dickey, *Multiple hypergeometric functions: probabilistic interpretations and statistical uses*, J. Am. Statist. Assoc. 78 (1983), pp. 628–37.
[9] J.M. Dickey, J.M. Jiang, and J.B. Kadane, *Bayesian methods for multinomial sampling with noninformatively missing data*, Tech. Rep. SUNYA Res. 15, Departament of Mathematics and statistics, State University of New York at Albany, Albany, NY, 1222, 1983.
[10] P.H. Garthwaite, J.B. Kadane, and A. O'Hagan, *Statistical methods for eliciting probability distributions*, J. Am. Statist. Assoc. 100(470) (2005), pp. 680–700.
[11] J. Gill and L. Walker, *Elicited priors for bayesian model specifications in political science research*, The Journal of Politics. 67(3) (2005), pp. 841–72.
[12] M.N. Jouini and R.T. Clemen, *Copula models for aggregating expert opinions*, Operations Research. 44(3) (1996), pp. 444–457.
[13] D. Kurowicka and R. Cooke, *Uncertainty analysis with high dimensional dependence modelling*, Wiley, Chichester, 2006.
[14] A. O'Hagan, *Probabilistic uncertainty specification: Overview, elaboration techniques and their application to a mechanistic model of carbon flux*, Environmental Modelling and Software. 36 (2012), pp. 35–4.
[15] A. O'Hagan and J. Forster, *Kendalls advanced theory of statistics*, 2nd ed., Vol. 2B, Arnold, London, 2004.
[16] A. O'Hagan and J.E. Oakley, *The Sheffield Elicitation Framework (SHELF), version (2.0)*, University of Sheffield. Sheffield, UK. 2010 URL http://www.tonyohagan.co.uk/shelf/.
[17] A. O'Hagan, C.E. Buck, A. Daneshkhah, J.R. Eiser, P.H. Garthwaite, D.J. Jenkinson, J.E. Oakley, and T. Rakow, *Uncertain judgements*, 1st ed., John Wiley & Sons, Ltd, England, 2006.