On Eliciting Logistic Normal Priors for Multinomial Models

Fadlalla G. Elfadaly & Paul H. Garthwaite
Department of Mathematics and Statistics, The Open University, UK

1 Introduction

Multinomial models arise when there is a set of complementary and mutually exclusive categories and each observation falls into one of these categories. Such models are used in many scientific and industrial applications. For example, they are frequently applied to the compositions of rocks in geology, to patterns of consumer selection preferences in microeconomics, and to voting behavior in political science. Other examples arise in medicine, psychology and biology.

Here we are concerned with the simplest case, where each observation has the same probability of falling into any specified category and observations are independent of each other. Then observations follow a multinomial distribution with, say, probability p_i that an observation falls in the ith category. We suppose that an expert's opinions about the p_i are to be quantified as a prior distribution for use in a Bayesian analysis. To quantify the expert's knowledge, the approach we adopt is to specify a distribution to represent her opinion and then ask her to make assessments that determine appropriate values for the parameters of that distribution.

It might seem natural to use the Dirichlet distribution to represent an expert's opinion, as this is the conjugate prior distribution for the parameters of a multinomial distribution. Elicitation methods that follow this approach have been proposed. O'Hagan *et al.* (2006) discussed two main available methods for Dirichlet elicitation, the method of Dickey *et al.* (1983) and that of Chaloner and Duncan (1987). 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 [e.g. Aitchison (1986), O'Hagan and Forster (2004)]. Its main drawback is that it has a limited number of parameters – if there are k categories then the corresponding Dirichlet distribution has only k parameters. Typically they will be too few to represent an expert's opinions about the means, variances and covariances of the p_i. In particular, the dependence structure between Dirichlet variates cannot be determined independently of its mean values; Dirichlet variates are always negatively correlated, which may not represent prior belief.

Motivated by these deficiencies, several authors have constructed new families of distributions for proportions that allow more general types of dependence structures [e.g. Leonard (1975), Tian *et al.* (2010)]. Some of these new distributions are direct generalizations of the standard Dirichlet distribution [e.g. Connor and Mosimann (1969)]. Generalized, nested or mixed forms of the Dirichlet distribution have been introduced and proposed as suitable prior distributions. For more details on possible prior distributions for multinomial models see, for example, O'Hagan and Forster (2004). Here we model expert opinion by a logistic normal distribution [Aitchison (1986)]. This has a large number of parameters and gives a prior distribution with a much more flexible dependence structure.

Eliciting parameters of multivariate distributions is not, in general, an easy task, especially if variates are not independent [O'Hagan *et al.* (2006)]. In the case of multinomial models, a particularly difficulty is to elicit assessments that satisfy all the constraints of mathematical coherence. Some of these constraints are obvious; the probabilities of each category must be non-negative and sum to one, for example. Others are less obvious. For example, if there are only two categories, the lower quartile for one category and the upper quartile of the other category must add to one. As the number of categories increases the constraints that must be satisfied increase and become less intuitive. For example, suppose there are three categories. If the median of one category is 0.6 and the lower quartile of the second category is 0.8, what values could be taken by the lower quartile of the third category if mathematical coherence is to be satisfied? (The answer is any non-negative value less than $0.13\dot{3}$, though this range shrinks under assumptions that impose smoothness on the pdf.) Partly because of these difficulties, no doubt, elicitation methods for multinomial sampling seem to have been constructed only for modelling opinion by a Dirichlet distribution.

We propose a method for quantifying opinion in the form of a logistic normal prior distribution. The method uses assessment tasks and a task structure that lead to coherent assessments without the expert having to be conscious of coherence constraints. The method has been implemented in interactive graphical user-friendly software developed in Java. In the next section of this paper we define the logistic normal prior to be used and justify the assumptions. The required assessments and their use to elicit the prior and obtain feedback is given in Section 3 and 4. We finish with concluding comments in Section 5.

2 The additive logistic normal prior

A well-known multivariate distribution for proportions is the logistic normal distribution, in which the proportions are transforming to variables that (by assumption) follow a multivariate normal distribution (Aitchison, 1986). Different multivariate logistic transformations are given in the literature, the additive logistic transformation being the most widely used [see, for example, Aitchison (1986)]. In this work, we develop a method of eliciting an additive logistic normal prior distribution for the multinomial model.

We suppose there are k categories. Let $\mathbf{p} = (p_1, \ldots, p_k)$ and let $\mathbf{Y}_{k-1} = (Y_2, \ldots, Y_k)'$. The additive logistic transformation from $\mathbf{Y}_{k-1}$ to $\mathbf{p}$ is defined by

$$p_1 = \frac{1}{1 + \sum_{j=2}^{k} \exp(Y_j)}, \qquad p_i = \frac{\exp(Y_i)}{1 + \sum_{j=2}^{k} \exp(Y_j)}, \quad i = 2, 3, \cdots, k, \tag{1}$$

with the inverse transformation

$$Y_i = \log(p_i/p_1) = \log(p_i/1 - p_2 - \cdots - p_k), \quad i = 2, 3, \cdots, k. \tag{2}$$

The vector $\underline{p}$ is said to have a logistic normal distribution if

$$\mathbf{Y}_{k-1} \sim \text{MVN}(\underline{\mu}_{k-1}, \Sigma_{k-1}). \tag{3}$$

In the above formulae, p_1 seems to be treated differently to the other p_i variables. However, the additive logistic normal distribution is permutation invariant. That is, whatever the ordering of the elements of the vector $\mathbf{p}$, the density function given above is invariant. For a theoretical proof of this property see Aitchison (1986). Hence, any order of the elements of $\mathbf{p}$ can be considered. With the representation given in (1) – (3), $p_1 = 1 - p_2 - \ldots - p_k$ is referred to as the fill-up variable. Its choice is arbitrary, in principle, though we believe that our assessment tasks are easier for the expert if categories are ordered so that p_1 is the probability of the most common category.

We assume that the prior opinion about $\mathbf{Y}_{k-1}$ can be represented by the multivariate normal distribution in (3). The elicitation method also requires a variable Y_1, defined as

$$Y_1 = \log(p_1/1 - p_1). \tag{4}$$

Now the components of $\mathbf{p}$ must sum to 1, which we refer to as the unit sum constraint on $\mathbf{p}$. Hence, from the normality assumption of $\mathbf{Y}_{k-1}$ in (3), the random variable e^{-Y_1} can be represented as a sum of $k-1$ lognormally distributed random variables: $e^{-Y_1} = (1-p_1)/p_1 = \sum_{i=2}^{k}(p_i/p_1)$. Although the sum of lognormal random variables has no exact distribution, there is an accumulated body of evidence indicating that the distribution of the sum of a finite number of lognormal random variables is well-approximated, at least to first order, by another lognormal distribution [Schwartz and Yeh (1982)]. Estimating the parameters of this latter distribution is challenging, and various approximations have been proposed [see for example, Beaulieu and Xie (2004)].

Hence it is reasonable to assume that e^{-Y_1} has a lognormal distribution, approximately, so that the distribution of Y_1 is approximately normal. We do not require any approximations for the distribution parameters, as they will be determined exactly from the expert's assessments. The main assumption we make is that $\mathbf{Y}_k$ follows a singular multivariate normal distribution,

$$\mathbf{Y}_k = (Y_1, \quad Y_2, \quad \cdots, \quad Y_k)' \sim \text{MVN}(\underline{\mu}_k, \Sigma_k). \tag{5}$$

The unit sum constraint on $\mathbf{p}$ will always lead to a singular variance matrix Σ_k. However, we assume that deleting the ith row and ith column of Σ_k, for any i $(i = 1, \ldots, k)$, will result in a positive-definite matrix.

Although no density function can be defined for the singular multivariate normal distribution, it is reasonably straightforward to use and it has found application in multivariate regression [Khatri (1968)] and Bayesian dynamic models [West and Harrison (1997)]. Its properties are also well-studied. In particular Rao (2002) derived various properties and characterizations of a multivariate normal distribution without using its pdf. He showed that the singular normal distribution is a special case of the usual normal distribution with similar conditional properties, though the usual inverse of the covariance matrix must be replaced by its generalized inverse. Conditional properties of the singular normal distribution are used extensively in the work reported here.

3 The elicitation method

Using the interactive software, the expert is required to assess conditional medians and quartiles for the elements of the probability vector $\mathbf{p}$. The expert does not need to be conscious of the constraints on her assessments. Instead, through the software we suggest coherent values that are close to her initial assessments, which she may accept or modify. From (2) and (4), the transformation from p_i to Y_i is strictly monotonic increasing $(i = 1, \ldots, k)$, so if we transform, say, an upper quartile assessment of p_i, we obtain the upper quartile of Y_i. Similarly for any other quantile. Exploiting the normality of $\mathbf{Y}_k$, its conditional medians and quartiles can be transformed into expectations, variances and covariances. For example, under the normality assumptions, the mean of Y_i can be equated to the median of Y_i; the latter is the transformed median assessment of p_i. The required assessments are detailed below.

3.1 Eliciting a mean vector $\underline{\mu}_{k-1}$ using median assessments

The expert is asked to assess the median values, $m_{j,0}$, of p_j, for $j = 1, 2, \ldots, k$. These assessed values are shown by the five outer (blue) bars in Figure 1.

The unit sum constraint on $\mathbf{p}$ implies that the means of its elements must sum to one, while medians do not necessarily have a unit sum, as the marginal distributions of the p_i are not symmetric. However, using the normality assumption of $\mathbf{Y}_k$ and the unit sum constraint on $\mathbf{p}$, we have shown that the median values must satisfy $\sum_{j=1}^{k} m_{j,0} = 1$. The importance of this interesting result is that it tells us whether an expert's conditional medians are statistically coherent, and we need to elicit median assessments from the expert rather than assessments of means, because conditional medians of $\mathbf{p}$ give conditional medians of $\mathbf{Y}_k$, through the monotonic transformations in (2) and (4). Clearly, means could not be used instead, as they do not transform.

To attain mathematical coherence, we normalize the expert's assessments to values that sum to one; these values are the five inner (yellow) bars in Figure 1. The expert is asked if they are a reasonable representation of her opinions, but can keep changing her assessments until she is happy with the normalized values that result.

The median assessments determine the hyperparameter $\underline{\mu}_{k-1} = (\mu_2, \cdots, \mu_k)$, as follows. Let $M(.)$ denote the median function. Then from (2), (4) and (5), we have, for $i = 2, 3, \cdots, k$,

$$\begin{aligned} \mu_i = E[Y_i] &= E[Y_i | Y_1 = E(Y_1)] = M[Y_i | Y_1 = M(Y_1)] \\ &= M[\log(p_i) - \log(p_1) | p_1 = m_{1,0}] = \log(m_{i,0}) - \log(m_{1,0}). \end{aligned} \tag{6}$$

The equivalence between the two events $Y_1 = M(Y_1)$ and $p_1 = m_{1,0}$ results from the monotonic function in (4). Conditioning on p_1 is essential for the medians in (6) to be fully transformable. This is the reason for adding the extra variable Y_1.

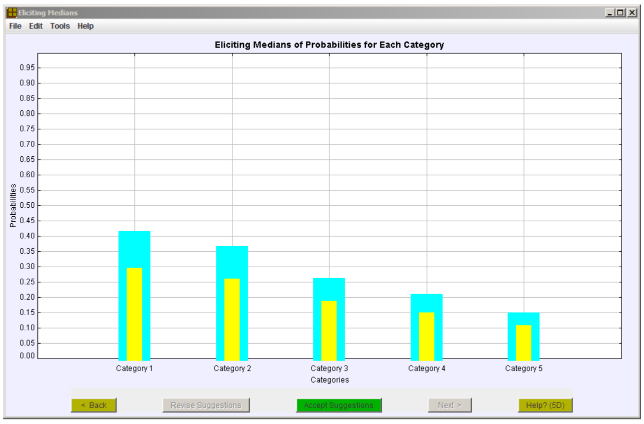

Figure 1: Assessing probability medians

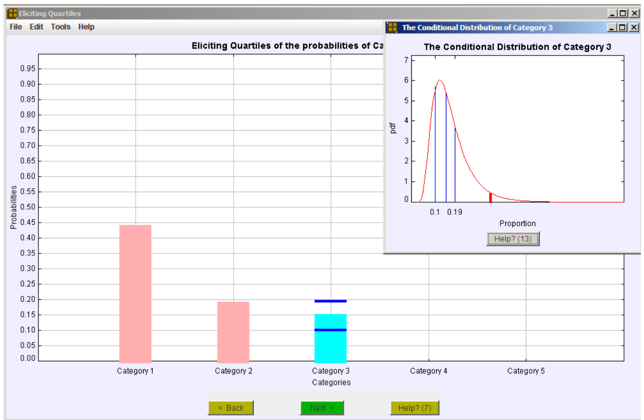

Figure 2: Assessing conditional quartiles

3.2 Eliciting a positive-definite variance matrix Σ_{k-1} using conditional assessments

3.2.1 Assessing conditional quartiles

The expert assesses a lower quartile L_1 and an upper quartile U_1 for p_1. Then, for each remaining p_j, $j = 2, \cdots, k-1$, she assesses two quartiles L_j and U_j given that $p_1 = m_{1,0}, \cdots, p_{j-1} = m_{j-1,0}$. See Figure 2, where the expert has assessed the two quartiles of p_3 conditional on the median values of p_1 and p_2 as given by the two leftmost (red) bars. To help the expert, the software presents an interactive graph showing the pdf curve of the lognormal distribution of $(p_j | p_1 = m_{1,0}, \cdots, p_{j-1} = m_{j-1,0})$, for $j = 2, \cdots, k-1$, see Figure 2. The expert is able to change her assessed conditional quartiles of p_j until the conditional pdf curve forms an acceptable representation of her opinion.

It is very tricky to find all the constraints that the conditional quartiles should satisfy. However, one clear constraint is that any upper quartile must not exceed the complement of the conditioning medians sum. With the aid of the lognormal curve, the expert is advised to make sure that her assessed interquartile range gives an almost zero probability of p_j exceeding $1 - \sum_{i=1}^{j-1} m_i$. This boundary is given by the shortest (red) vertical line on the pdf graph.

3.2.2 Assessing conditional medians

Here, the expert is asked to assume that the median of p_1 has been changed from $m_{1,0}$ to $m_{1,1}$. Given this information, she will be asked to change her previous medians $m_{j,0}$ of each p_j into a new assessment $m_{j,1}$, for $j = 2, \cdots, k$.

Then, in each successive step i, for $i = 2, \cdots, k-2$, the expert will be asked to suppose that the median values of $p_1, \cdots, p_i$ are $m_{1,1}, \cdots, m_{i,i}$, respectively, shown as red bars in Figure 3. Given this information, she will be asked to change her assessed medians of the most recent previous step $m_{i+1,i-1}, \cdots, m_{k,i-1}$, shown by the horizontal (black) lines on Figure 3, respectively. Her new medians, as the three outer (blue) bars on Figure 3, are denoted, $m_{i+1,i}, \cdots, m_{k,i}$, where $m_{j,i}$ is the median of $(p_j|p_1 = m_{1,1}, \cdots, p_i = m_{i,i})$.

For mathematical coherence, these conditional medians must also sum to one, i.e. $\sum_{j=1}^{i} m_{j,j} + \sum_{j=i+1}^{k} m_{j,i} = 1$, for $i = 1, \cdots, k-2$. The software suggests a normalized set that satisfies this condition, shown as the three inner (yellow) bars on Figure 3. The expert can change her assessments until she feels that the suggestions give the best representation of her opinion.

Conditional medians and quartiles are used to elicit a positive-definite variance matrix Σ_{k-1}. We modify the method of Kadane *et al.* (1980) to show that the conditional assessments are sufficient to elicit a positive-definite matrix Σ_{k-1}. The main idea is that, under the normality assumption, the quartiles of $\mathbf{Y}_k$ are transformed into variances using the well-known formula of the interquartile range IQR_i and the variance σ_i^2 of any normal variate Y_i, namely, $\sigma_i^2 = (\text{IQR}_i/1.349)^2$. Covariances $\sigma_{i,j}$ are also elicited from the formula $E(Y_i|Y_j = y_j) = E(Y_i) - \sigma_{i,j}\sigma_j^{-2}[E(Y_j) - y_j]$. The structural assessing way we adopt here guarantees the required rank of the resulting variance matrix. As mentioned before, Σ_k is elicited as a singular matrix of rank $k-1$; removing its first row and column leads to the desired positive-definite matrix Σ_{k-1}. We are interested in Σ_{k-1} as it is the variance of $\mathbf{Y}_{k-1}$, which is the hyperparameter of the prior distribution of $\mathbf{p}$.

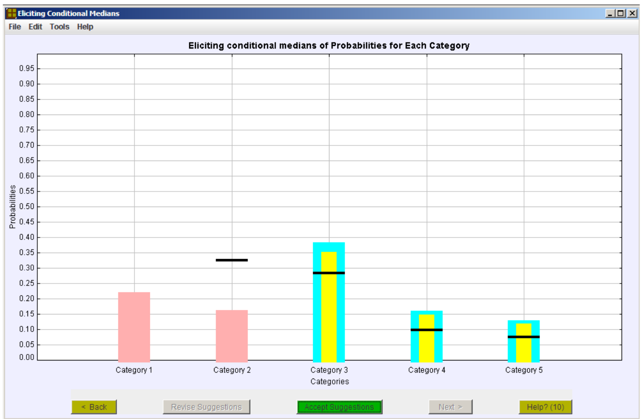

Figure 3: Assessing conditional medians

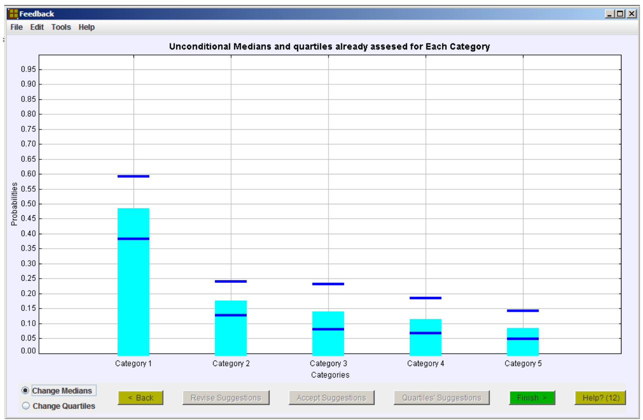

Figure 4: Feedback screen

4 Feedback using marginal quartiles

Conditional assessment tasks are generally harder to perform than unconditional tasks. Most expert assessments have been conditional assessments, so it is important to inform the expert of the marginal quartiles that the prior distribution implies and give her the option of changing them if she wants. Hence, after eliciting the hyperparameters, we calculate marginal medians and quartiles of the probability of each category and show these as feedback to the expert; see Figure 4.

To add this feedback to the software, a method is required for estimating marginal quartiles from the elicited hyperparameters $\underline{\mu}_{k-1}$ and Σ_{k-1}. Unfortunately, a closed form method for estimating the marginal moments, or quartiles, of the logistic normal distribution does not seem to exist. Aitchison (1986) suggested using Hermitian numerical integration methods to obtain marginal moments.

However, under the normality assumption of $\mathbf{Y}_{k-1}$ and the unit sum constraint, we proved that the conditional medians $m_{j,0}$, for $j = 2, \cdots, k$, equal the marginal unconditional medians of p_j, for $j = 2, \cdots, k$, respectively, provided the lognormal sum is adequately approximated by another lognormal random variable. Moreover, these results make it possible to estimate marginal lower and upper quartiles for each p_j, for $j = 2, \cdots, k$. The main idea is to change the fill-up probability from p_1 to each other probability p_j, $j = 2, \cdots, k$, making use of the permutation invariance property of the logistic normal distribution. Treated as the fill-up probability, each p_j is the sum of lognormal variates and follows an approximate lognormal marginal distribution, from which the unconditional quartiles of p_j can be computed.

On Figure 4, the expert can change both the marginal medians and quartiles of some or all categories. Then she gets coherent suggestions showing the impact of any change on the rest of the marginal assessments. The expert may keep changing her assessments until she is happy with the suggested marginal values. On completion of the elicitation, the software outputs the elicited hyperparameters of the logistic normal prior distribution, $\underline{\mu}_{k-1}$ and Σ_{k-1}, in a suitable format for further Bayesian analysis in WinBUGS.

5 Concluding comments

The proposed method of eliciting a logistic normal prior for multinomial models has the potential to be extended to more general contexts, such as multinomial logistic regression models and contingency tables. The method described here has been implemented in interactive software that is freely available on the web.

References

Aitchison, J. (1986). *The Statistical Analysis of Compositional Data*. Chapman and Hall, London.

Beaulieu, N. C. and Xie, Q. (2004). An optimal lognormal approximation to lognormal sum distributions. *IEEE Transactions on Vehicular Technology*, **53**, 479–489.

Chaloner, K. and Duncan, G. T. (1987). Some properties of the Dirichlet-multinomial distribution and its use in prior elicitation. *Communications in Statistics-Theory and Methods*, **16**, 511–523.

Connor, R. J. and Mosimann, J. E. (1969). Concepts of independence for proportions with a generalization of the Dirichlet distribution. *Journal of the American Statistical Association*, **64**, 194–206.

Dickey, J. M., Jiang, J. M., and Kadane, J. B. (1983). Bayesian methods for multinomial sampling with noninformatively missing data. Technical Report 6/83 - #15, State University of New Yourk at Albany, Department of Mathematics and Statistics.

Kadane, J. B., Dickey, J. M., Winkler, R., Smith, W., and Peters, S. (1980). Interactive elicitation of opinion for a normal linear model. *Journal of the American Statistical Association*, **75**, 845–854.

Khatri, C. G. (1968). Some results for the singular normal multivariate regression models. *Sankhyā*, **30**, 267–280.

Leonard, T. (1975). Bayesian estimation methods for two-way contingency tables. *Journal of the Royal Statistical Society, Series B*, **37**, 23–37.

O'Hagan, A. and Forster, J. (2004). B*ayesian Inference*, volume 2B of *Kendall's Advanced Theory of Statistics*. Arnold, London, second edition.

O'Hagan, A., Buck, C. E., Daneshkhah, A., Eiser, J. R., Garthwaite, P. H., Jenkinson, D. J., Oakley, J. E., and Rakow, T. (2006). *Uncertain Judgements: Eliciting Expert Probabilities*. John Wiley, Chichester.

Rao, C. R. (2002). *Linear Statistical Inference and its Applications*. John Wiley & Sons, Inc., New York, second edition.

Schwartz, S. C. and Yeh, Y. S. (1982). On the distribution function and moments of power sums with log-normal components. *The Bell System Technical Journal*, **61**, 1441–1462.

Tian, G. L., Tang, M. L., Yuen, K. C., and Ng, K. W. (2010). Further properties and new applications of the nested Dirichlet distribution. *Computational Statistics and Data Analysis*, **54**, 394–405.

West, M. and Harrison, J. (1997). B*ayesian Forecasting and Dynamic Models*. Springer-Verlag, New York, second edition.